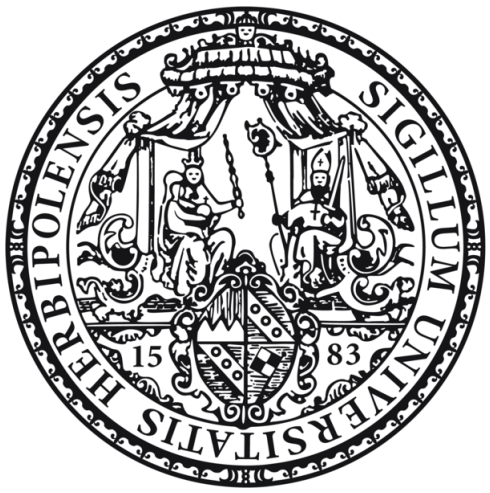
Mechanistic Insights into the Inhibition of Cathepsin B and Rhodesain with Low-Molecular Inhibitors



Dissertation zur Erlangung

des naturwissenschaftlichen Doktorgrades

der Julius-Maximilians-Universität Würzburg

vorgelegt von

Anna Heilos

aus Würzburg

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

_____

Inhaltsverzeichnis

1 Summary

Cysteine proteases play a crucial role in medical chemistry concerning various fields reaching from more common ailments like cancer and hepatitis to less noted tropical diseases, namely the so called African Sleeping Sickness (Human Arfican Trypanosomiasis). Detailed knowledge about the catalytic function of these systems is highly desirable for drug research in the respective areas. In this work, the inhibition mechanisms of the two cysteine proteases cathepsin B and rhodesain with respectively one low-molecular inhibitor class were investigated in detail, using computational methods. In order to sufficiently describe macromolecular systems, molecular mechanics based methods (MM) and quantum mechanical based method (QM), as well as hybrid methods (QM/MM) combining those two approaches, were applied. Chapter 5 provides information about the theoretical background of the utilized methods. In Chapter 7 and 8, the results of this work are presented and discussed. They are structured as follows: After a short introduction of the respective subject, a brief summary of all relevant results will be given followed by detailed discussions.
Chapter 7 deals with the reaction behaviour of carbamate-based inhibitors in cathepsin B. The various reaction pathways of the inhibitor in the active site of the enzyme were evaluated *via* thermodynamic calculations. In combination with docking studies and consideration of experimental data concerning the inhibition potential of different derivatives performed in the group of Prof. Guetschow in Bonn,[1] only one reaction path remained feasible. The path contains an E1cB elimination reaction, followed by a carbamylation of the active site amino acid Cys29 (see Firgue 1.1).

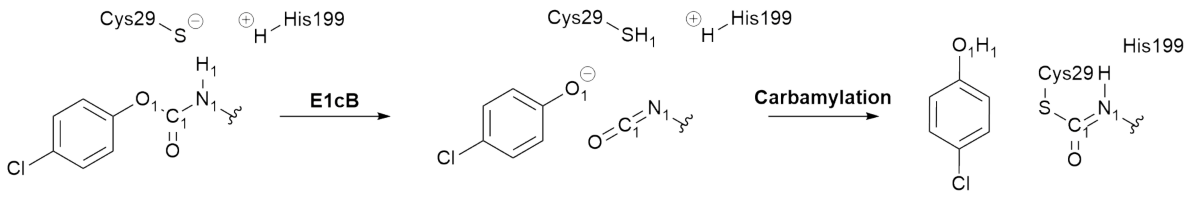


Figure 1.1: Proposed mechanism for the reaction of a carbamate-based inhibitor in cathepsin B.

The two reaction steps were investigated separately. Concerning the first step (elimination), different approaches at the QM/MM level of theory were applied in order to find the energetically most favourable reaction path. Based on the computed reaction profiles and potential energy surfaces, two conclusions were drawn. First, the phenolate anion, formed during the reaction, has to be stabilized *via* a proton transfer through Cys29 (see Firgue 1.1). Second, the results indicated that the mechanism is a complex process involving a number of reaction coordinates. Therefore, "Nudged Elastic Band" (NEB) calculations (QM/MM level) were performed, showing reaction barriers of nearly 40 kcal/mol. As the inhibition is unlikely to occur under these conditions, a new approach was applied. Based on a molecular mechanics simulation, the active site of the enzyme holds enough space for the inhibitor and an additional water molecule, placed in a favourable position for a water-bridged proton transfer.

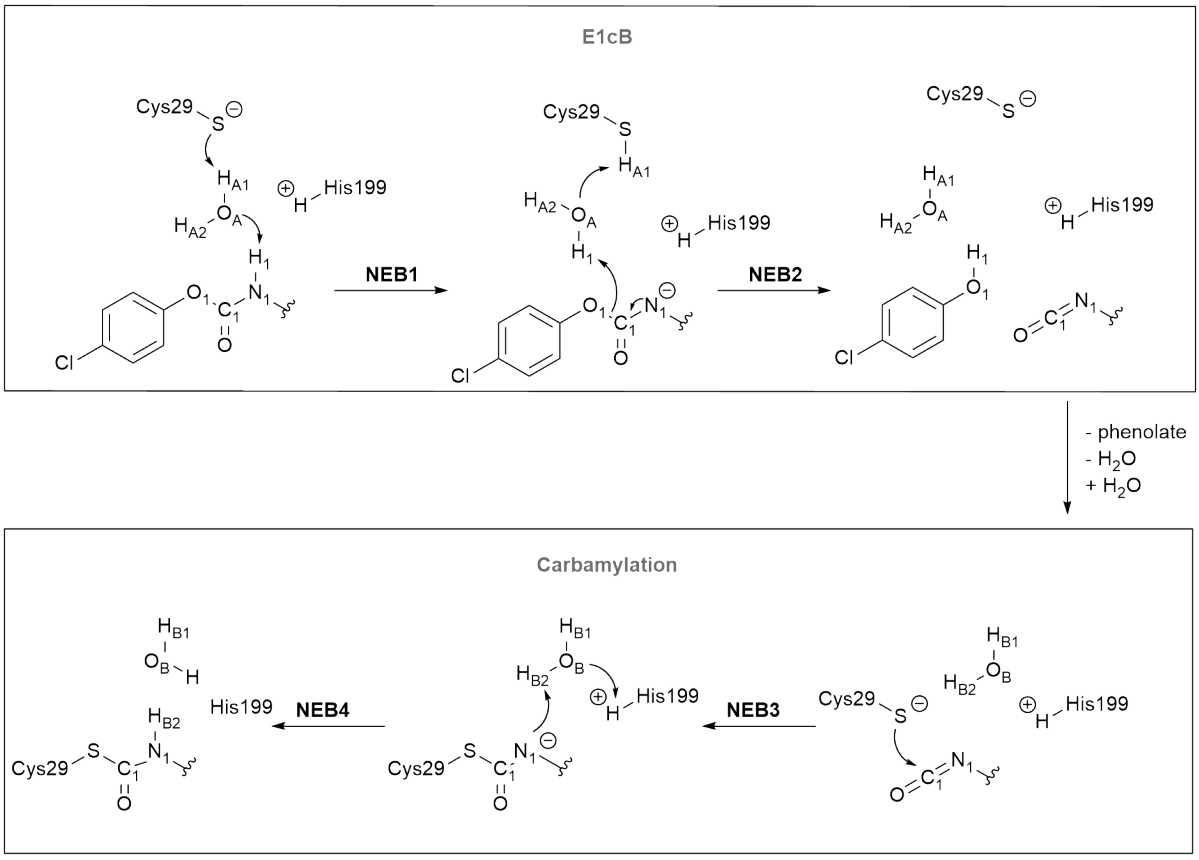


Figure 1.2: Reaction scheme of the water-mediated mechanism for the inhibition of cathepsin B using carbamate-based inhibitors.

The computed potential energy surfaces (QM/MM) predict reaction barriers of around 20 kcal/mol, whereas the NEB calculations (QM/MM) revealed barriers of only

12 kcal/mol. Based on these results, a water-mediated E1cB elimination was assumed. As the NEB calculations did not converge for the whole E1cB elimination (and also for the carbamylation reaction) the mechanism was divided into four distinctive steps shown in Figure 1.2. The second step of the investigated inhibition mechanism (carbamylation) was also calculated with help of several reaction profiles and NEB calculations at the QM/MM level of theory. The resulting path predicts a reaction barrier of around 5 kcal/mol for the nucleophilic attack.

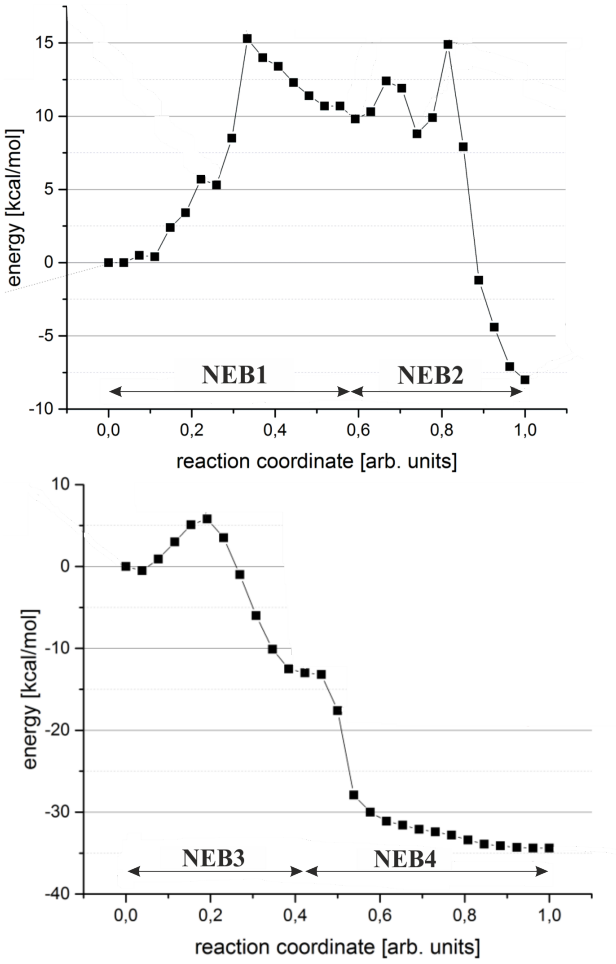


Figure 1.3: Computed NEB pathways of the reaction scheme shown in Figure 1.2. QM/MM calculations with B3LYP/cc-pVDZ as QM method; for the utilized MM method, it is referred to Chapter 6.

The calculations of the subsequent protonation of the inhibitor through His199 did again indicate the participation of a water molecule in the mechanism. The inclusion of this water molecule resulted in a nearly barrier free water-bridged proton transfer. The resulting reaction path of the scheme in Figure 1.2 is depicted in Figure 1.3. In

conclusion, the investigated mechanism seems to be a complex reaction that depends on a number of coordinates and is partly concerted. Furthermore, the participation of a water-bridged proton transfer reaction seems to be crucial for the reaction.

The second class of systems investigated in this work contain the cysteine protease rhodesain as well as nitroalkene-based inhibitors. These are discussed in Chapter 8. The computations entailed three derivatives of a nitroalkene-based inhibitor, differing in one residue at the α position of the nitro alkene (see Figure 1.4). Nitroalkenes represent a promising inhibitor class concerning the covalent-reversible inhibition of rhodesain. Compared to the irreversible vinylsulfon inhibitor K11777, the inhibition with nitroalkenes was shown to have higher binding affinities in combination with a reversible reaction mechanism.[2]

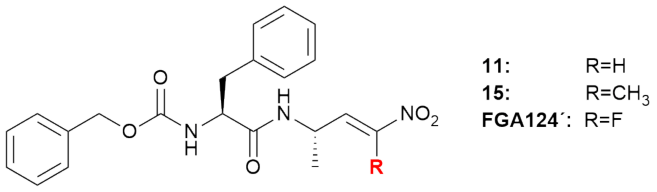


Figure 1.4: Investigated nitroalkene inhibitors.

In Chapter 8.2, the inhibition behaviour of fluorinated and unfluorinated nitroalkenes is discussed. The fluorinated nitroalkenes showed an interesting behaviour in the experiment, performed by the group of Prof. Schirmeister.[3] While for the unfluorinated systems a ”one-step” mechanism was measured, the experiments showed a ”two-step” mechanism for the fluorinated compounds. The terms ”one-step”/”two-step” mechanism do, however, not refer to an actual mechanistic course of the reaction, but to the scheme shown in Figure 1.5. The mechanism of a covalent inhibition contains the two reaction steps, depicted. Enzyme (E) and inhibitor (I) form a non-covalent complex ($E\cdots I$) in the first step, which subsequently reacts to a covalent product (E-I). The second step of this reaction may contain several sub-steps.

$$E + I \rightleftharpoons E\cdots I \rightleftharpoons E\text{-}I$$

Figure 1.5: Scheme of a two step mechanism for a covalent reversible inhibition.

For the reaction of the unfluorinated nitroalkenes, only one value quantifying the inhibition potential was measured (K_i). This value thus contains information about all mechanistic steps that occur during the reaction. Assuming the inhibition reaction only includes the formation of the non-covalent complex $E^{\cdots}I$, K_i would exclusively describe this single step. In case of a rapid reaction from $E + I$ to $E\text{-}I$, K_i contains information about all steps. There are thus several possibilities for the inhibition by the unfluorinated molecules: The reaction could stop at the non-covalent complex. Alternatively, enzyme and inhibitor rapidly react to the covalent product or any intermediate structure. For the fluorinated compounds, in contrast, two separated steps were identified by experimental measurements, which renders the respective inhibitors so called "slow-binders". This implies that at some point of the mechanism, the reaction is slowed down. In order to find the element that slows the reaction of the fluorinated systems down, the two systems were compared *via* various computational methods. Thermodynamic calculations within a model system on a QM level of theory with a number of (un)fluorinated nitroalkene and vinylsulfone derivatives did not provide an explanation for the measured phenomenon. Within the NEB calculations of the systems, based on preceding MD simulations of the non-covalent enzyme-inhibitor complexes, no differences were found either. Both inhibitors showed a similar behaviour during the MD simulation of the non-covalent enzyme-inhibitor complex and the respective reaction barriers for the formation of the covalent product (step A and step B in Figure 1.6) were also in a similar range. The results would speak for a rapid reaction of the unfluorinated compound to the covalent product, as both reaction barriers are in a similar range and the reaction enthalpy is in a range around -10 kcal/mol. The reactants would thus react so rapidly that the measurement could only "resolve" one reaction step.

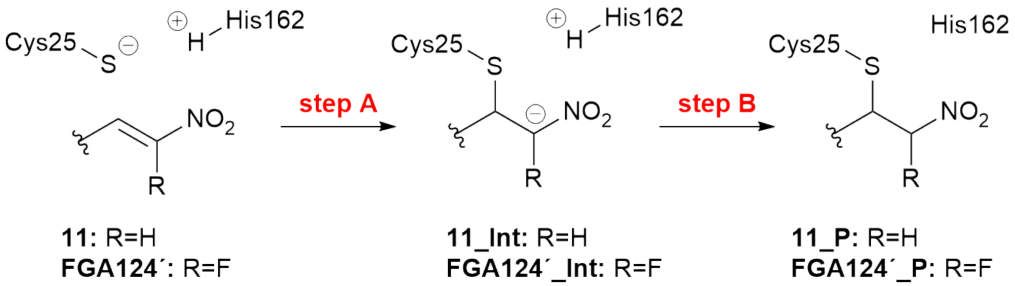


Figure 1.6: Reaction scheme for the investigation of the barriers and the performance of MD simulations.

In order to gain insight in a potential influence of the environment on the reaction, MD simulations were performed with the intermediate structures (**11_Int** and **FGA124_Int**) of both systems. The unfluorinated nitroalkene inhibitor stayed in a stable position inside the active site during the simulation with no significant structural changes. For the system with the fluorinated compound, the active site amino acid histidine (His), which normally stabilizes the NO_2 group and protonates the double bond in the second reaction step, rotates out of the active site, away from the electrophilic part of the inhibitor (also known as warhead). The reason for the rotation could be connected to the reduced electron density in the warhead, with respect to the unfluorinated molecule. This movement could cause a longer residence time inside the active site for the fluorinated intermediate, as either histidine has to rotate back, or the protonation would have to take place from the unfavourable orientation. The extended residence time of the fluorinated compounds would be a plausible explanation for the measurement of a two-step mechanism, as the reaction would rapidly proceed to the covalent intermediate **FGA124'_INT** (see Figure 1.6) and then slow down due to the rearrangement of the system. In order to quantify the influence of this rearrangement, reaction profiles for the second step of the mechanism (protonation through histidine, see Figure 1.6) were compared for the unfluorinated system and both orientations (His rotated and His unrotated) of the fluorinated systems. The unfluorinated system has a reaction barrier of +8.3 kcal/mol (QM/MM NEB calculations). For the rotated histidine ring, the barrier exceeded 30 kcal/mol (QM/MM reaction profile calculations) due to the unfavourable orientation. The fluorinated nitroalkene intermediate **FGA124'_INT** (see Figure 1.6) with the histidine ring in its original orientation has a barrier of +6.8 kcal/mol for the second step of the reaction. In conclusion, the decreased electron density of the fluorinated nitroalkenes is likely to cause weaker interactions between enzyme and inhibitor inside the active site. The resulting rotation of histidine out of the active site causes a longer residence time of the fluorinated derivatives, which can be used to explain the "slow-binder" mechanism of the compounds.

Chapter 8.3 discusses the different behaviour of compound **11** and **15** (see Figure 1.4). Experiments performed by the Schirmeister group showed that compound **11** (R = H) has a significantly lower K_i value for the formation of the non-covalent bond compared to compound **15** (R = CH_3). Thermodynamic calculations and docking studies of the systems did not show crucial differences in the behaviour of the two inhibitors. However, MD simulations of the non-covalent enzyme-inhibitor complexes revealed that due to steric hindrance, the methyl residue of inhibitor **15** changes its orienta-

tion, which leads to a rotation of the warhead out of the active site. Inhibitor **11**, in contrast, stays in a relatively stable position in favour of a reaction with the active site amino acids. The simulations could thus explain the different behaviour of the two compounds. In conclusion, the worse inhibition potential of inhibitor **15** compared to inhibitor **11** can be traced back to sterical hinderance due to the methyl substituent. For inhibitor **15**, no "two-step" mechanism was measured and previous experiments showed that without the presence of the warhead, no inhibition takes place at all. It can thus be assumed, that the inhibition mechanism with the methylated nitroalkene only contains the reversible formation of the non-covalent enzyme-inhibitor complex, whereas the unmethylated inhibitor forms a covalent bond with the enzyme.

2 Zusammenfassung

Cysteinproteasen spielen eine wichtige Rolle in der medizinischen Chemie. Nicht nur im Bereich bekannterer Krankheiten wie Krebs oder Hepatitis, sondern auch bezüglich weniger verbreiteter, tropischer Krankheiten wie der sogenannten afrikanischen Schlafkrankheit (Afrikanische Trypanosomiasis) haben diese Enzyme eine große Bedeutung. Im Bereich der Wirkstofffindung ist ein detailliertes Wissen über die katalytische Funktion der an einer Krankheit beteiligten Enzyme unabdingbar. In der vorliegenden Arbeit wurden die Inhibitionsmechanismen der beiden Cysteinproteasen Cathepsin B und Rhodesain in Verbindung mit zwei niedermolekularen Inhibitorklassen anhand theoretischer Berechnungen untersucht. Um die makromolekularen Systeme ausreichend genau beschreiben zu können, wurden neben molekularmechanischen (MM) und quantenmechanischen (QM) Ansätzen auch Hybridmethoden verwendet, welche beide Ansätze (QM/MM) verbinden. Der genaue mathematische Hintergrund dieser Methoden ist Gegenstand von Kapitel 5. In den Kapiteln 7 und 8 werden die Ergebnisse der Arbeit vorgestellt und diskutiert. Die Kapitel sind dabei wie folgt aufgebaut: Nach einer kurzen Einleitung zu den jeweiligen Themen folgt eine Zusammenfassung aller relevanten Ergebnisse. Danach werden die einzelnen Ergebnisse detailliert diskutiert.
Kapitel 7 befasst sich mit dem Reaktionsverhalten Carbamat-basierter Inhibitoren in Cathepsin B. Mithilfe thermodynamischer Berechnungen wurden die verschiedenen Reaktionsmechanismen des Inhibitors in der aktiven Tasche des Enzyms evaluiert. Die Erkenntnisse dieser Rechnungen in Kombination mit Dockingstudien und experimentellen Daten bezüglich des Inhibitionspotentials, welche im Arbeitskreis von Prof. Guetschow in Bonn[1] durchgeführt wurden, erlauben nur einen möglichen Reaktionsweg. Dieser beinhaltet eine E1cB Eliminierung als ersten Schritt, gefolgt von einer Carbamylierung der Aminosäure Cys29 (siehe Abbildung 2.1). Bezüglich des ersten Reaktionsschrittes (der Eliminierung) wurden Berechnungen mittels der QM/MM Methode durchgeführt, um den energetisch günstigsten Reaktionspfad zu finden.

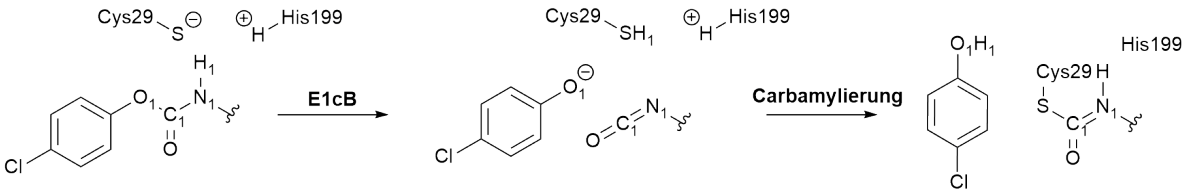


Abbildung 2.1: Vorgeschlagener Mechanismus für die Hemmung von Cathepsin B mit carbamat-basierten Inhibitoren.

Aufgrund der resultierenden Reaktionsprofile und Potentialflächen konnten zwei wichtige Schlussfolgerungen gezogen werden. Erstens muss das Phenolat-Anion, welches während der Reaktion gebildet wird, durch eine Protonierung von Cys29 stabilisiert werden (siehe Abbildung 2.2). Zweitens deuten die theoretischen Ergebnisse darauf hin, dass es sich bei dem vorliegenden Mechanismus um einen komplexen Vorgang handelt, der von einer Vielzahl an Reaktionskoordinaten abhängt. Daher wurden "Nudged Elastic Band"(NEB) Rechnungen (mit der QM/MM Methode) angefertigt, welche eine Reaktionsbarriere von nahezu 40 kcal/mol vorhersagen. Da die Reaktion unter diesen Bedingungen nicht real ablaufen würde, wurde ein weiterer Ansatz der theoretischen Modellierung getestet. Aufgrund einer vorangegangenen MD-Simulation des Systems stellte sich heraus, dass in der aktiven Tasche genug Platz für den Inhibitor und ein zusätzliches Wassermolekül ist. Dieses Wassermolekül diffundierte während einer Simulation in die aktive Tasche und lag in einer günstigen Position für einen wasser-verbrückten Protonentrasfer (siehe Abbildung 2.2). Die berechneten Potentialflächen für dieses neue System (QM/MM) sagen Reaktionsbarrieren von circa 20 kcal/mol voraus, wohingegen die NEB Rechnungen sogar die höchste Barriere bei nur 12 kcal/mol ansetzt. Basierend auf diesen Ergebnissen wird ein wasser-verbrückter Mechanismus angenommen. Die NEB-Rechnungen des gesamten E1cB Mechanismus als auch der gesamten Carbamylierung konvergieren nicht, wodurch keine validen Ergebnisse erhalten wurden. Daher wurden die Reaktionen in vier Schritte, wie sie in Abbildung 2.2 gezeigt sind, unterteilt. Der zweite Reaktionsschritt (Cabamylierung) wurde ebenfalls mithilfe mehrerer Potentialflächen und NEB-Pfade auf QM/MM Niveau berechnet.

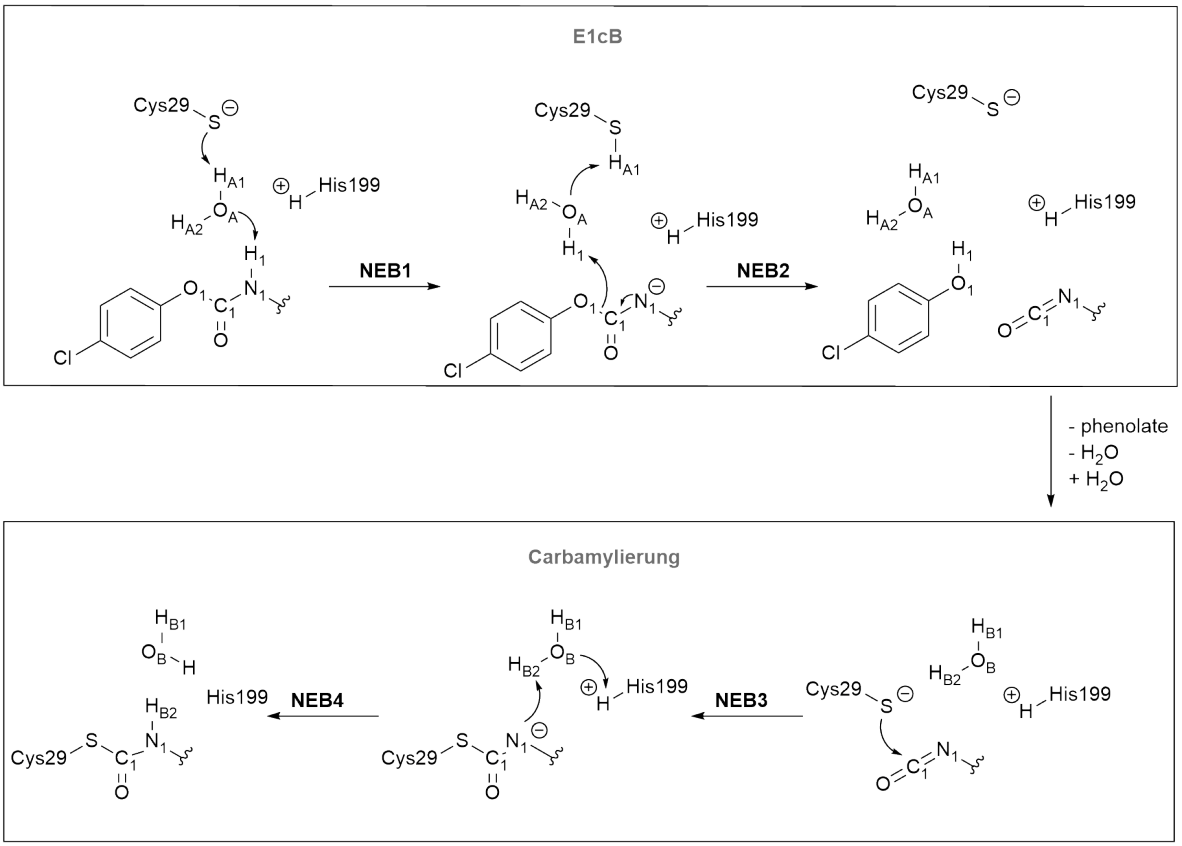


Abbildung 2.2: Reaktionsschema des Wasser-verbrückten Mechanismus.

Der energetisch günstigste Reaktionspfad hat eine Barriere von 5 kcal/mol für den nukleophilen Angriff von Cys29 an den Inhibitor (siehe Abbildung 2.3). Die Berechnungen des nachfolgenden Protonentransfers von His199 zum Inhibitor deuteten erneut auf einen wasser-verbrückten Mechanismus hin. Unter Einbezug des zusätzlichen Wassermoleküls resultierte ein nahezu barrierefreier Protonentransfer als letzter Schritt der Reaktion (siehe Abbildung 2.3). Der NEB Reaktionspfad des in Abbildung 2.2 gezeigten Schemas ist in Abbildung 2.3 dargestellt. Zusammenfassend handelt es sich bei dem untersuchten Mechanismus um eine komplexe Reaktion, deren Reaktionsschritte zum Teil konzertiert ablaufen. Außerdem scheinen wasser-verbrückte Protonentransferreaktionen entscheidend zum Reaktionsverlauf beizutragen.

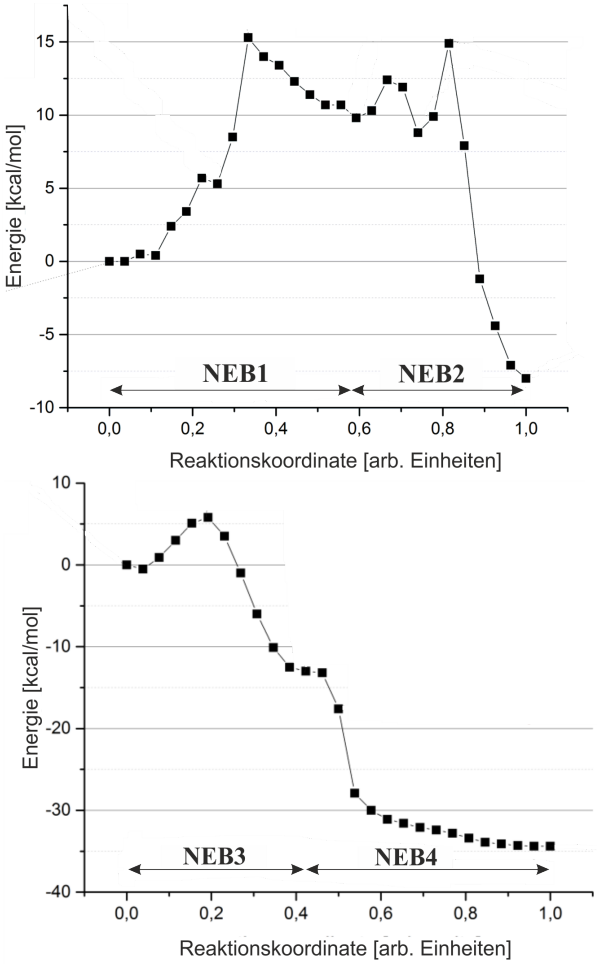


Abbildung 2.3: Berechneter NEB-Pfad (QM/MM mit B3LYP/cc-pVDZ als QM Methode) des Reaktionsschemas aus Abbilung 2.2.

Das zweite System, welches im Rahmen dieser Arbeit untersucht wurde, beinhaltet die Cysteinprotease Rhodesain zusammen mit Nitroalken-basierten Inhibitoren und wird in Kapitel 8 diskutiert. Es wurden drei Derivate eines Inhibitors untersucht, die sich jeweils in einer Modifikation am α-Kohlenstoff des Nitroalkens unterscheiden (siehe Abbildung 2.4). Nitroalkene stellen eine vielversprechende Inhibitorklasse bezüglich kovalent-reversibler Hemmung von Rhodesain dar.[2] Im Vergleich zum bereits bekannten, irreversibel bindenden Vinylsulfoninhibitor K11777 haben Nitroalkene höhere Bindungsaffinitäten in Kombination mit einem reversiblen Inhibitionsmechanismus. In Kapitel 8.2 wird das Inhibitionsverhalten fluorierter und nicht-fluorierter Nitroalkene in Rhodesain untersucht. In den Experimenten der Arbeitsgruppe von Prof. Schirmeister[3] zeigten die fluorierten Nitroalkene

ein interessantes Verhalten. Während für die unfluorierten Systeme ein "Ein-Schritt-Mechanismus"gemessen wurde, konnte bei den fluorierten Derivaten ein "Zwei-Schritt-Mechanismus" festgestellt werden. Die Begriffe beziehen sich hier allerdings nicht auf den tatsächlichen mechanistischen Ablauf, sondern auf das Reaktionschema, welches in Abbildung 2.5 gezeigt ist. Der Mechanismus einer kovalenten Inhibitionsreaktion beinhaltet die zwei in der eben erwähnten Abbildung gezeigten Schritte.

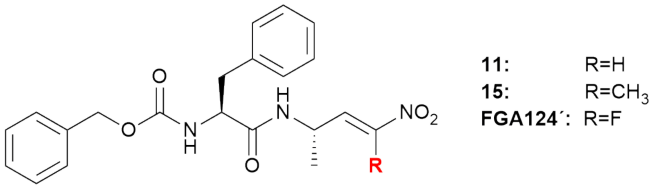


Abbildung 2.4: Untersuchte Nitroalken-Inhibitoren.

Enzym (E) und Inhibitor (I) bilden im ersten Schritt einen nicht-kovalenten Komplex ($E^{\cdots}I$), der dann im zweiten Schritt zu einem kovalenten Produkt (E-I) weiter reagiert. Dieser zweite Schritt kann je nach Reaktionsmechanismus selbst wieder aus mehreren Unterschritten bestehen.

$$E + I \rightleftharpoons E\cdots I \rightleftharpoons E\text{-}I$$

Abbildung 2.5: Reaktionsschema für einen zweischrittigen Inhibitionsmechanismus einer kovalent reversiblen Inhibition.

Für die Reaktion der unfluorierten Nitroalkene in Rhodesain wurde lediglich ein Wert gemessen, der das Inhibitionspotential quantifiziert (K_i). Dieser Wert enthält folglich Informationen über alle beteiligten Schritte. Würde die Hemmung nur aus der Bildung des nicht-kovalenten Komplexes $E^{\cdots}I$ bestehen, würde K_i ausschließlich diesen Schritt beinhalten und quantifizieren. Im Fall einer schnellen Reaktion von $E + I$ zu E-I, enthält K_i hingehen Informationen zu allen beteiligten Schritten. Es gibt daher mehrere Möglichkeiten, den Wert für die unfluorierten Nitroalkene zu interpretieren: Die Reaktion könnte nach der Bildung des nicht-kovalenten Komplexes stoppen oder Enzym und Inhibitor könnten so schnell zu dem kovalenten Produkt (oder jedem stabilen Intermediat) reagieren, dass experimentell nur ein Wert gemessen werden kann.

Für die fluorinierten Verbindungen wurden im Gegensatz dazu zwei separate Schritte gemessen, weshalb die entsprechenden Inhibitoren als "slow-binders"(sinngemäß "langsame Binder") bezeichnet werden. Diese Messungen implizieren, dass die Reaktion an irgendeinem Punkt verlangsamt wird. Um den Ursprung der Verlangsamung zu finden wurden die beiden Systeme (fluoriert: **FGA124'**; unfluoriert: **11**, siehe Abbildung 2.6) anhand diverser theoretischer Modell-Rechnungen untersucht und verglichen. Thermodynamik-Rechnungen mit Modellsystemen auf QM Niveau lieferten keine Erklärung für die gemessenen Werte. Auch NEB Rechnungen der Systeme auf QM/MM Niveau in relaxierter Umgebung (im Anschluss an eine MD Simulation des nicht-kovalent gebundenen Komplexes) konnten das unterschiedliche Verhalten nicht erklären. Beide Verbindungen zeigten ein ähnliches Verhalten im nicht-kovalenten Enzym-Inhibitor Komplex während der MD Simulation, sowie bei der Reaktion zu den Intermediaten (**11_Int** und **FGA124'_Int** in Abbildung 2.6) und den Produkten (**11_P** und **FGA124'_P** in Abbildung 2.6). Die Ergebnisse sprechen für eine schnelle Reaktion der unfluorierten Nitroalkene hin zum kovalent gebunden Produkt, da beide Reaktionsbarrieren in einem energetisch ähnlichen Bereich sind und die Reaktionsenthalpie bei -10 kcal/mol liegt. Die Reaktanden würden daher so schnell miteinander reagieren, dass bei der Messung nur ein Schritt aufgelöst werden kann.

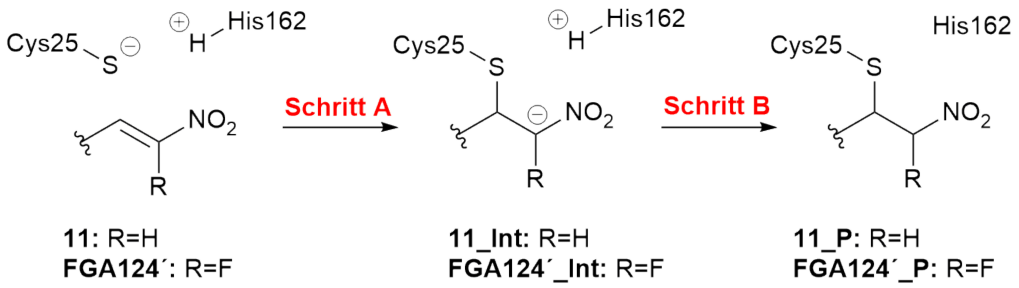


Abbildung 2.6: Reaktionsschema der (un)fluorierten Nitroalkene in Rhodesain für die Untersuchung der Reaktionsbarrieren und MD-Simulationen.

Um einen potentiellen Einfluss der Umgebung auf den Reaktionsversverlauf zu untersuchen, wurden ausgehend von Intermediat-Strukturen erneut MD-Simulationen von beiden Systeme angefertigt. Das unfluorierte Nitroalken blieb dabei stabil in der aktiven Tasche des Enzyms, ohne signifikante strukturelle Veränderungen des Systems. Bei der Simulation des fluorierten Intermediates dagegen drehte sich das Histidin des aktiven Zentrums, welches normalerweise die NO_2-Gruppe stabilisiert und im zweiten Reaktionsschritt die Doppelbindung protoniert, aus der Enzymtasche heraus.

Der Grund für die Drehung kann durch eine schwächere Wechselwirkung zwischen Histidin und der NO_2-Gruppe des Inhibitors erklärt werden, die wiederum durch eine verringerte Elektronendichte im "warhead"(electrophiler Bereich des inhibitors) der Verbindung, aufgrund des Fluorsubstituenten, zurückgeführt werden kann. Die daraus folgende, ungünstige Orientierung des Inhibitor-Intermediates zum Histidin würde eine verlängerte Residenzzeit des Inhibitors in der aktiven Tasche nach sich ziehen, da die Protonierung entweder aus der ungünstigen Orientierung statt finden müsste oder eine erneute Rotation des Histidin-Rings erforderlich wäre. Durch diese verlängerte Residenzzeit kann auch die Messung des Zwei-Schritt-Mechanismus im Experiment erklärt werden, wenn man annimmt, dass die Reaktion bis zum kovalenten Intermediat **FGA124'_INT** schnell durchläuft und dann aufgrund der Umorientierung des Histidin verlangsamt. Um den Einfluss der Drehung des Histidins zu quantifizieren, wurden Reaktionsprofile des zweiten Schrittes der Reaktion (Protonierung durch Histidin, siehe Abbildung 2.6) für beide Systeme (das fluorierte System in beiden Orientierungen des Histidins) berechnet. Das unfluorierte System zeigte bei den NEB-Rechnungen auf QM/MM Niveau eine Reaktionsbarriere von +8.3 kcal/mol auf. Die Barriere für das fluorierte System mit dem gedrehten Histidin-Ring war aufgrund der ungünstigen Orientierung über 30 kcal/mol hoch (Berechnung über Reaktionsprofil). Für die günstigere Orientierung des Histidins mit dem fluorierten Intermediat wurde anhand eines Reaktionsprofils eine Barriere von +6.8 kcal/mol berechnet. Zusammenfassend scheint aufgrund der verringerten Elektronendichte der fluorierten Nitroalkene eine schwächere Wechselwirkung zwischen Inihibitor und Enzym im aktiven Zentrum stattzufinden. Die darin begründete Drehung des Histidin-Rings weg vom "warhead"des Inhibitors zieht eine verlängerte Residenzzeit des fluorierten Nitroalkens nach sich, was wiederum den "slow-binder"-Mechanismus der fluorierten Systeme erklärt.

Kapitel 8.3 beschäftigt sich mit der theoretischen Modellierung von in der Arbeitsgruppe von Prof. Schirmeister durchgeführten Experimenten zu den Inhibitoren **11** und **15**.[2] Inhibitor **11** (R = H in Abbildung 2.4) zeigt ein signifikant höheres Inhibitionspotential in Rhodesain im Vergleich zu Verbindung **15** (R = CH_3 in Abbildung 2.4). Da sich bei thermodynamischen Berechnungen als auch in Dockingstudien keine Unterschiede zwischen den beiden Systemen zeigten, wurden MD-Simulationen der nicht-kovalenten Enzym-Inhibitor-Komplexe durchgeführt. Im Rahmen dieser Simulationen bewegte sich die Methylgruppe von Inhibitor **15** aufgrund sterischer Hinderung in eine andere Orientierung, wodurch der "warheadäus

dem aktiven Zentrum herausgedreht wurde. Da bei Inhibitor **11** keine solche sterische Komponente vorliegt, blieb dieser Inhibitor in einer relativ stabilen Position, von der aus er gut mit dem aktiven Zentrum des Enzyms reagieren könnte. Den Untersuchungen zufolge kann das bessere Inhibitionspotential von Verbindung **11** im Verlgeich zu Verbindung **15** also auf einen höheren sterischen Anspruch des methylierten Inhibitors **15** zurückgeführt werden. Da für Inhibitor **15** kein "Zwei-Schritt-Mechanismus"gemessen wurde und vorhergegangene Experimente zeigten, dass ohne "warhead"keine inhibition stattfindet, kann davon ausgegangen werden, dass der Inhibitionsmechanismus ausschließlich aus der Bildung des nicht-kovalenten Komplexes besteht, wohingegen der unmethylierte Inhibitor bis zum kovalenten Produkt durchreagiert.

3 Introduction

3.1 Cysteine Proteases

3.1.1 Cysteine Proteases in Medicine and Research

The human body contains over 500 proteolytic enzymes, also known as proteases. Their main task is to catalyze the cleavage of peptide bonds, which leads to the degradation of proteins. Roughly 2% of the human genome encodes proteases, whereas for bacteria and viruses it is up to 5%.[4,5] The role of proteases in pharmaceutical research has dramatically changed in the last decades. In the beginning, they were primary considered to be important for processes like digestion and intracellular turnover.[6] In the 1950s, Davie *et al.* (together with a number of others) discovered that limited proteolysis also takes place in protein activation of blood clotting.[7,8] Since then, proteases have also been considered as regulative proteins, making them an interesting target for drug research. Today, they are known to participate in physiological key processes like DNA replication, cell death or immune responses.[9] Cysteine proteases in particular make up a big part of all known proteolytic enzymes. They can be found in mammals, plants, fungi, protozoa viruses and bacteria. Their diverse occurrence in different organisms has made them an interesting target for medical research. The involvement of cysteine proteases in many different physiological and pathogenic processes offers a great number of opportunities for the development of pharmaceutical compounds. Regarding the human organism, cathepsins, which belong to the papain-like cysteine protease family, play a crucial role. Diseases related to cathepsins include Alzheimer's disease,[10] Osteoporosis[11] and cancer.[12,13] They occur from a breakdown in the control of protease activity which leads to undesired and unregulated proteolysis.[14] Parasitic cysteine proteases are most often associated with tropical diseases. They participate in a number of physiological key steps of the parasite and are essential for their survival.[11,15,16] Diseases like Malaria, Chagas or African Trypanosomiasis originate from parasites. In 2017, 216 million people were

infected with Malaria, from which 445.000 have died.[17] Viral infections connected with cysteine proteases are Hepatitis A or SARS.[11] In viruses, there is often only one essential cysteine protease for the replication of the virus and there are no structural related human enzymes, making them a much easier target than for example parasitic cysteine proteases.[18] Table 3.1 gives examples for different cysteine-protease related diseases and the involved enzymes.[9,11]

Tabelle 3.1: A number of pharmaceutical relevant cysteine proteases and their function in living organisms.

Occurrence	**Protease**	**Function**	**Disease**
Human			
Homo sapiens	Cathepsin B	Cell Degradation of Pathogens	Cancer, Rheumatoid Arthritis
Homo sapiens	Cathepsin K	Tumor Spreading	Breast Cancer
Parasites			
T. b. rhodesiense	Rhodesain	Cell Replication of the Parasite	African Trypanosomiasis
P. falcilarum	Falcipain-2	Hemoglobin Degradation in Trophozoites	Malaria
Viruses			
Hepatitis A Virus	Picormaine 3C	Viral Replication and Transcription	Hepatitis A
SARS Coronavirus	SARS-CoV M^{pro}	Viral Replication and Transcription	SARS

T. b.: Trypanosoma brucei; P. falcilarum: Plasmodium falciparum; SARS: Severe Acute Respiratory Syndrom.

Due to structural similarities of many cysteine proteases, the design of specific inhibitors is a real challenge. Proteases were once described as "Mother Nature's Swiss Army Knife. They have many different functions, even though they often have nearly identical structures."[19] Concerning cysteine proteases, there are many evolutionary relations between the different clans and families, making it a very challenging task to address one specific enzyme.

3.1.2 Classification, Structure and Catalytic Mechanism

While all proteases cleave peptide bonds *via* hydrolysis, they differ in a number of aspects, for example regarding their catalytic mechanism, structure and recognition patterns. A number of models arose from different attempts to classify proteases.
One of these models parts the proteases in two categories, namely endo- and exopeptidases. Endopeptidases cleave peptide bonds of non-terminal amino acids, whereas exopeptidases cleave terminal peptide bonds. There is a further categorization of exopeptidases into amino and carboxy-peptidases, depending on whether the peptide bond is cleaved from the carboxy or amino end.
In 1999, N.D. Rawlings and A.J Barrett developed a database for the categorization of proteases, called MEROPS.[20] Their classification is based on the structural relations between the proteases and includes evolutionary information.

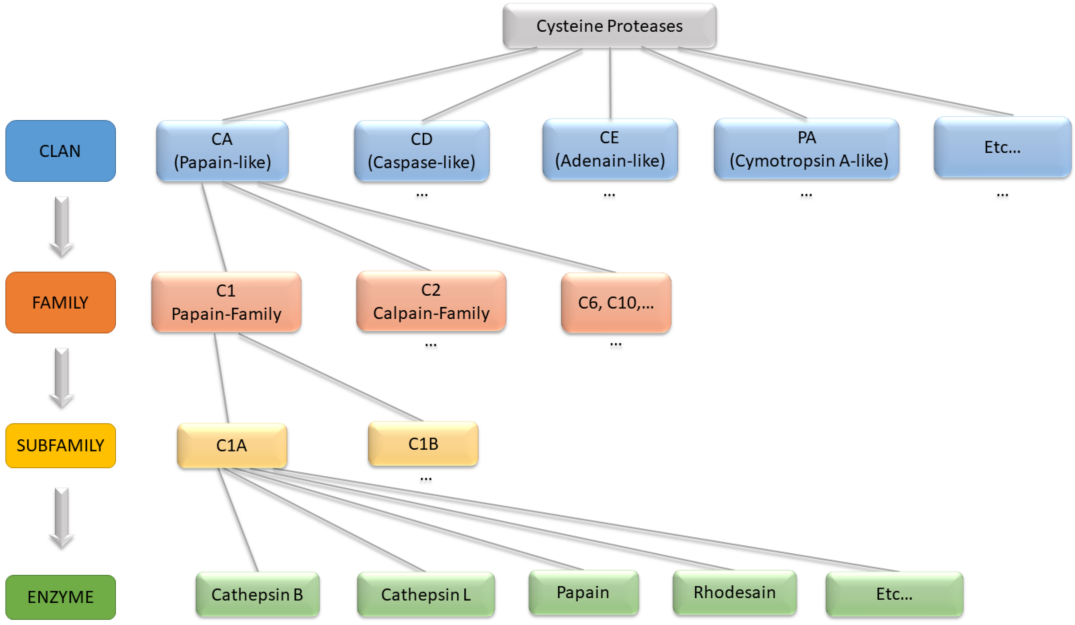


Abbildung 3.1: Schematic categorization of cysteine proteases on the basis of the MEROPS database with selected examples.

MEROPS includes six main categories: The Cysteine-, Serine-, Aspartic-, Threonine-, Glutamine- and Metalloproteases, based on the active-site amino acid or additional compounds. Furthermore, there is a seventh category for unknown proteases that are mixtures of the above mentioned. The proteases inside a category are further divided into clans and families. All proteolytic enzymes inside a clan have the same evolutionary origin. They show similarities in their tertiary structures or amino acid

sequences around the active site.[20] A clan can include one or more families, whereas a family is assigned *via* similarities of amino acid sequences.
Cysteine proteases are divided into a number of clans. The three biggest are the papain-like (CA), the caspase-like (CD) and the adenain-like clan (CE). Furthermore, there are cymotrypsin-like cysteine proteases (PA), which also include serine proteases.[20,21] The majority of cysteine proteases are members of the papain-like clan, rendering this category the subject of intensive research. The clan includes over 40 families from which the papain-like family C1 and the calpain-like family C2 play a crucial role in parasitic proteases.[20,22] The human cathepsin B and the parasitic rhodesain, which are both subjects of this work, belong to the subfamily C1A of family C1. Figure 3.1 shows a schematic categorization of cysteine proteases with selected examples on the basis of the MEROPS data base.[20] As the investigated systems of this work are papain-like cysteine proteases, the following sections will focus on this subject.

All papain-like proteases share similar features in their structure. They consist of two domains, the L-domain (left side) and the R-domain (right side). In between them lies the active site cleft with a cysteine and a histidine residue as active site amino acids.[23] The histidine is located in the R-domain, which also includes a short α-helical passage and larger passages of β-sheets (orange in Figure 3.2). The domain has a shape similar to a vessel. The active site cysteine is part of the L-domain, which is dominated by a large α-helix (green in Figure 3.2).[9,23,24] Figure 3.2 illustrates the structures of three papain-like cysteine proteases with the characteristics described above.

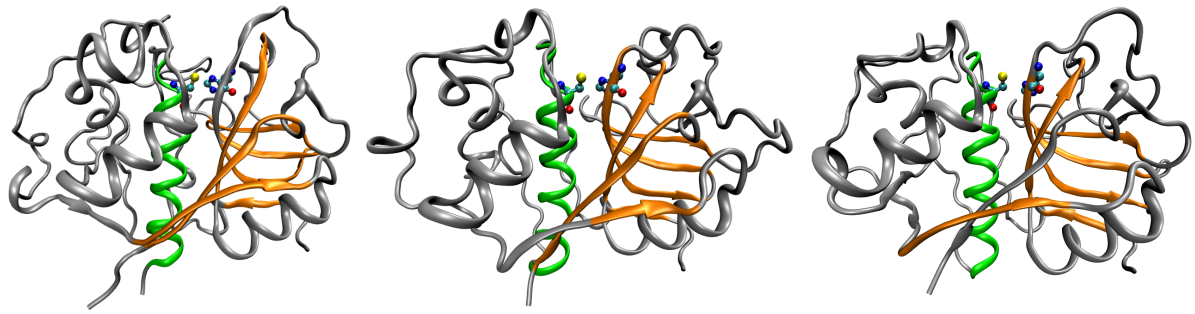


Abbildung 3.2: Crystal structures from the RCSB data-base[25] of cathepsin B (1HUC),[26] papain (1PPN)[27] and rhodesain (2P7U).[28] Active site amino acids are depicted as ball and stick.

Even though the tertiary structures and a number of amino acid sequences are similar inside a family, the different enzymes can be highly selective. The structure and orientation of the binding site has a strong influence on this selectivity. The binding site of proteases can be described *via* subsites or bonding pockets, depending on the orientation of the ligand in the catalytic center. Berger and Schechter introduced this scheme in 1967.[29] The binding site is divided into several binding pockets (see Figure 3.3). In the center of the binding site lies the cleavage site. The non-primed side is located towards the amino-end of the ligand (N-terminus). The binding pockets of the enzyme on this side are numbered as S_n with $n = 1,2,3,...$ starting with the smallest number in the center of the binding site. The respective positions of the substrate are named P_n, where P_1 and S_1 can interact. Regarding the carboxy-end of the ligand (C-terminus), the same principle is used, whereas the side is called primed side with binding pockets $S_n^{'}$ of the enzyme and areas $P_n^{'}$ of the ligand.

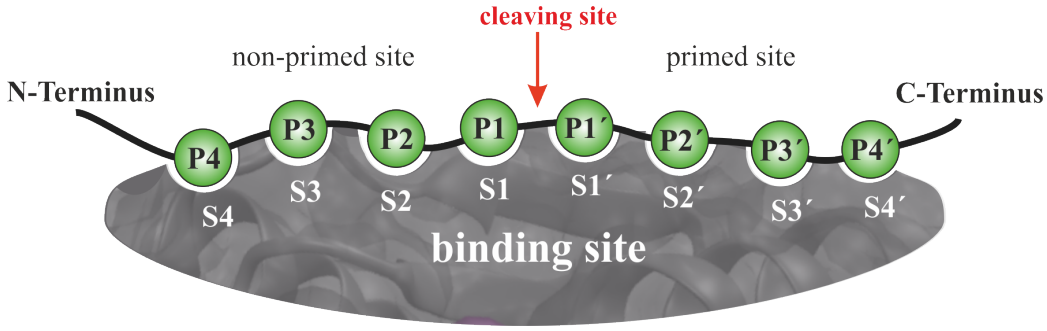


Abbildung 3.3: Berger-Schechter Scheme of the binding site of a protease with a substrate.

The highly conserved active site of the papain-like cysteine proteases lies, as mentioned above, in the cleft between L- and R-domain. It contains the two amino acids Cys25 and His159 (papain-numbering) and additionally holds an asparagine residue Asn175, which stabilizes the His159 *via* hydrogen bonding. The glutamine residue Gln19 is part of the oxyanion-hole, an area which contains a number of amino acids that form hydrogen bonds with the negative charged oxygen-atom of the substrate-intermediate (see Figure 3.4).

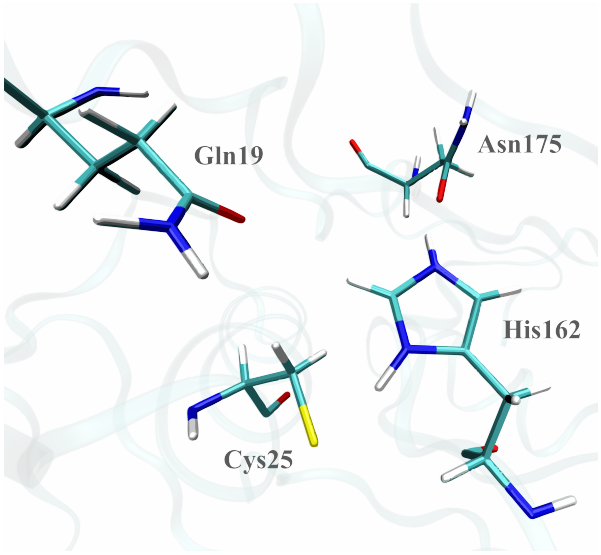


Abbildung 3.4: Active site amino acids of an exemplary cysteine protease (rhodesain). The underlying structure is taken from the RCSB data base.[28]

Cysteine proteases use covalent catalysis for the cleavage of peptide bonds. The cysteine residue acts as a nucleophile, when it is in a deprotonated form. The protonation states of the cysteine and histidine residue inside the active site differs for the particular enzymes. The protonation state may furthermore be pH-dependent and is still a controversial and debated issue for some species.[30,31,32,33] In the zwitter-ionic form of the catalytic dyade, the cysteine attacks the peptide bond at the carbonyl carbon atom, forming a tertiary intermediate. The negative charge at the oxygen atom is stabilized through Gln19 inside the oxyanion hole of the enzyme. As typical for a nucleophilic substitution, the carboxylic double bond is rebuilt, releasing the amino group of the substrate, which is then protonated through His159. The primed side of the substrate exits the reaction and the starting point of the catalysis is reached again with the hydrolysis of the thioester and the release of the non-primed side of the substrate. This mechanism is shown schematically in Figure 3.5.

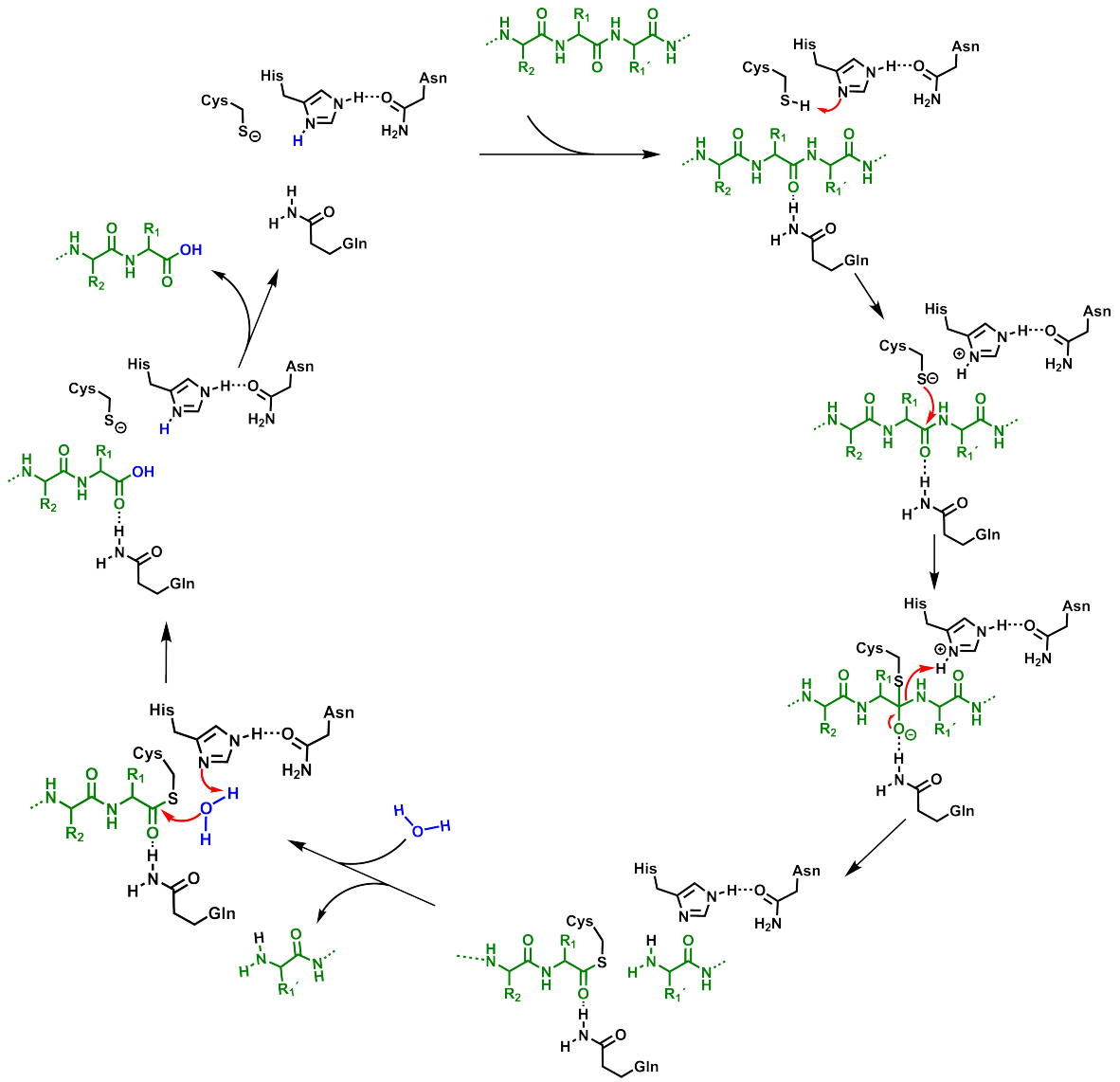


Abbildung 3.5: Mechanism of peptide bond cleaving in the active site of cysteine proteases.

3.2 Covalent Reversible Inhibition

Covalent inhibition of enzymatic reactions has not been a focus of pharmaceutical research throughout the last years, even though well known covalent inhibitors like penicilins or acetylsalicylic acid exist.[34] The reasons for the seeming unpopularity of these agents are the possible severe side effects that may occur due to the strong covalent bonding.[35,36,37] A promising research area addressing this issue is covalent reversible inhibition.[2,34,38,39,40] Covalent reversible inhibitors hold the advantages of covalent inhibition, namely higher binding affinities, longer drug target residence time

or lower sensitivity against pharmacokinetic aspects, while the reversibility reduces the risk of undesired side effects.[38]

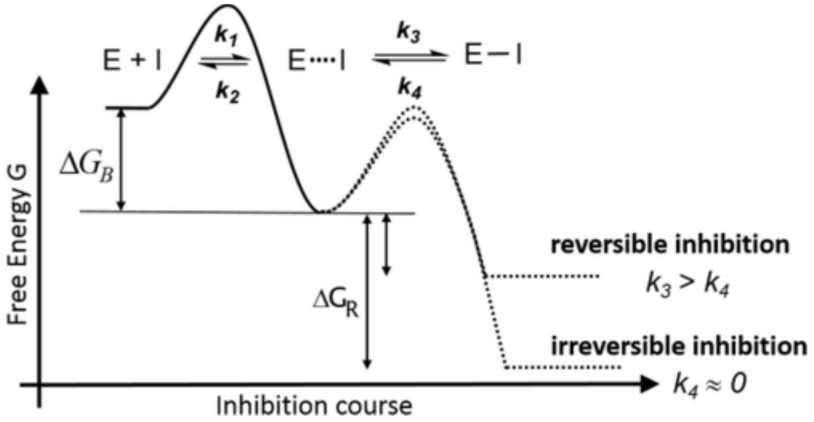


Abbildung 3.6: Energy diagram of a covalent inhibition mechanism. ΔG_B is the enthalpy for the formation of the non-covalent complex and ΔG_R is the enthalpy of reaction for the formation of the covalent bond.[38]

Figure 3.6 illustrates the mechanism's underlying covalent inhibition. Due to the second step of the reaction mechanism, covalent inhibition is a more complex task compared to non-covalent inhibition. As a result, the development of covalently reversible binding compounds is much more challenging. A covalent inhibitor can bind irreversibly to an enzyme or form a reversible bond. In case of a reversible bond formation, an additional rate constant (k_4 in Figure 3.6) for the dissociation of the reaction product *E-I* is introduced. For both, a reversible and irreversible covalent inhibition, the inhibitor enters the active site of the enzyme in the first step and forms a non-covalent enzyme-inhibitor complex *E···I*. The stability of the complex *E···I* (ΔG_B in Figure 3.6) depends on the arrangement of the inhibitor in the active site and the resulting interactions. The corresponding dissociation constant K_i can be obtained experimentally and is often used as a representative value for the inhibition potential of a component. As the constant only holds information about the first, non-covalent step of the mechanism, it is a common quantity for not only covalent reversible, but also irreversible and non-covalent reversible inhibition. In the following step, a covalent bond is formed between the enzyme and the inhibitor, resulting in the product *E-I*. The size of the exothermicity of this step is crucial for the reversibility of the reaction. In Figure 3.7 a scheme of the reaction steps is depicted, where k_n are the respective rate constants for the formation and dissociation of the non-covalent enzyme-inhibitor complex *E···I* and the covalent product *E-I*.

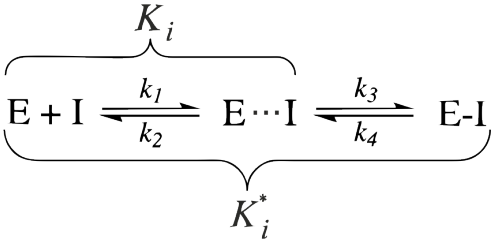


Abbildung 3.7: Scheme of a two step mechanism for a covalent reversible inhibition.

K_i is the dissociation constant of the first step of the mechanism and K_i^* corresponds to the entire two-step mechanism. The energetic profile of the second step can be controlled by the choice of the inhibitor warhead.
A popular example in this area is the inhibition of rhodesain with the peptidyl vinylsulfon compound K11777 (see chapter 3.4). The original inhibitor binds to the enzyme in a covalent irreversible reaction.[28] The halogenation of the inhibitor at the alpha position of the double bond with bromine, however, leads to the reversibility of the inhibition.[38] This example shows that rational design of covalent inhibitors opens manifold possibilities in drug-research, combining the advantages of covalent and reversible inhibition. In the following section, the two papain-like proteases cathepsin B (human) and rhodesain will be discussed in detail.
The dissociation constants can be obtained by

$$K_i = \frac{k_2}{k_1} \tag{3.1}$$

and

$$K_i^* = K_i \frac{k_4}{k_3 + k_4}. \tag{3.2}$$

In the first step of the mechanism, the non-covalent inhibitor-enzyme complex $E\cdots I$ is formed. If k_3 is by far larger than k_4, or the rate constant of the reverse reaction is nearly zero, the second step, and with it the whole mechanism, becomes irreversible. For the quantization of inhibitor potentials, the K_i and IC_{50} value are typically used for competitive inhibition.[41] The latter gives the concentration of the inhibitor that is needed to halve the enzyme activity. For irreversible inhibition, it depends on the substrate concentration and the residence time of the inhibitor in the active site compared to the substrate.[42] The K_i value is connected with the law of mass action of the dissociation of the non-covalent inhibitor-enzyme complex $E\cdots I$. The smaller the K_i value, the stronger the inhibition.

3.3 Cathepsin B

Human cathepsin B contains around 250 amino acids and has a molecular weight of 30 kDa.[26] It is ubiquitously expressed in the human organism and is involved in a number of physiological processes, such as apoptosis (programmed cell death), wound healing or remodelling of extracellular matrices .[43,44,45] Due to a special structural feature (the occluding loop), cathepsin B can act both as an endo- and an exopeptidase.[43] The 18-residue-long loop contains two histidine residues and has a crucial impact on the enzyme activity.[43,46,47,48,49] It is located on the primed side of the enzyme and can cover the active site cleft (see Figure 3.8). As the loop is flexible, it controls the enzymatic activity as endo- or exopeptidase.[43] Montcourrier *et al.* identified a pH dependency of the loop formation.[50] When the loop covers the active site cleft, the two histidine residues can interact with the C-terminal of the substrate. The histidines are responsible for the carboxy peptidase activity of the enzyme. The catalytic dyade in the active site, Cys29 and His199, form an ion pair due to the conserving environment.[43]

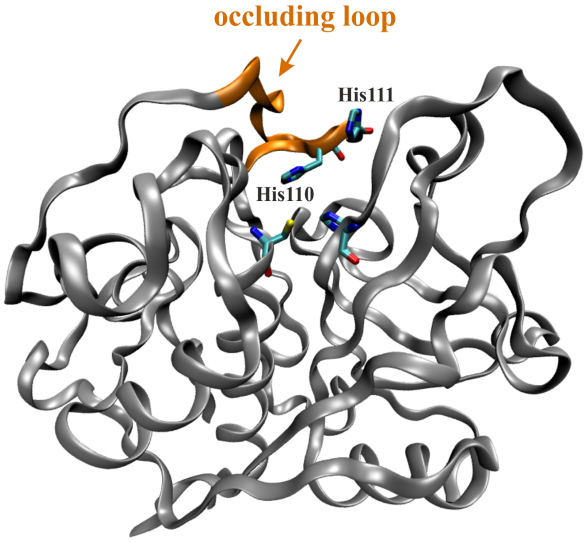


Abbildung 3.8: Structure of human cathepsin B (1HUC[26] from RCSB data base[25]) with the active site dyade Cys29 and His199 as well as the two histidine residues His110 and His111 of the occluding loop (cathepsin B numbering) highlighted.

Role in Pathogenic Processes

Cathepsin B is associated with a number of diseases, including invasive and metastatic cancer, Alzheimers Disease and rheumathiod arthritis.[11,51,52] The origin and symptoms of these pathogenic disorders do not have a systematical connection, which shows the variety of physiological impact of cathepsin B on the human organism. A severe symptom of rheumatoid arthritis is the erosion of articular cartilage.[11,53] Cathepsins, including cathepsin B, were found to be expressed in cartilage tissue.[54,55,56,57] The destruction of this tissue is caused by excessive proteolytic degradation.[53] It has been shown that cysteine protease inhibitors such as peptidyl vinyl sulfones or fluoromethyl ketones reduce this degradation.[58,59,60]
Another pathogenic field, on which cathepsin B has an impact, is cancer. The term cancer sums up a number of different disorders with a variety of origins. Tumor progression and metastasis depend on proteolysis in the respecting areas.[11] There is a number of complex cascades, responsible for the pathogenic process. The overexpression of cathepsin B is connected with prostate, bladder or colon carcinomas.[61,62,63,64]
A third intensive research area in medicine concerns the Alzheimer's Disease, a brain degrading disorder. Infected patients suffer, among other symptoms, from memory loss and speech disorder. In a diseased brain, amyloid plaques deposit between neuron cells.[65] Amyloid is a protein that is cleaved and degraded in a healthy brain. In Alzheimer's Disease, the precursor protein of amyloid, APP, is altered, leading to a defect in the degradation of amyloid. As a consequence, the protein agglutinates in the brain.[65] Cathepsin B is known to be involved in the processing of APP and therefore presents a potential target for a regulation of the pathogenic progression.[43,65]

Potential Inhibitors

Finding potent inhibitors for cathepsin B is a challenging task, as it requires selectivity and efficiency. Besides the endogenous protein inhibitors from the cystatin superfamily, recent drug-related research concentrates on the development of low-molecular compounds based on information gained from crystal-structures.[43,46] Those molecules consist of an electrophilic part, the warhead, which interacts with the Cys29 anion, and a peptidyl backbone to fit inside the active site pockets.[43] Examples include peptidyl-aldehydes, -vinylsulfones, -disulfides and -halomethyl ketones.[66]

3.4 Rhodesain

The parasitic cysteine protease rhodesain represents an important target in the field of tropical diseases. It contains around 219 amino acids and has a molecular weight of 40 kDa.[28] The name rhodesain origins from the protozoan parasite *trypanosoma brucei rhodesiense* (*t.b. rhodesiense*).[41] This parasite causes the human african trypanosomiasis (HAT), also known as african sleeping sickness.[15,41,67] HAT is transmitted by the bite of an infected tsetse fly. The parasite lives in the blood stream and tissue fluids of infected mammalian hosts.[67]

Human Arfican Trypanosomiasis

According to the WHO, over 70% of the reported cases of HAT over the last ten years occurred in the Democratic Republic of Congo.[68] Two forms of the disease exist, whereas the acute, rapid-onset variant transmitted from *t.b. rhodesiense* is widely spread in southern and eastern Africa.[69,70] This form represents the minority of the reported cases, but develops rapidly and is a lethal disease, if it is not treated.[41,68] The disease is characterized by two main symptomatic stages. In the first (haemolyphatic) stage, the parasite lives mainly in the bloodstream of the mammalian host, but can migrate to further areas like the lymphatic system, spinal fluids and spleen.[41,71] The symptoms in this stage are fever, muscle aches, itching or rash.[68,72] The second, neurological stage, is indicated by confusion, mental deterioration, extreme lethargy in combination with sleep disturbance (the origin of the disease's eponym) and eventually ends in death. In this stage, the parasite has crossed the blood brain barrier of the host and causes damage in the nervous system.[68,71,72,73]
The papain-like cysteine protease rhodesain appears in all life-cycle stages of *t.b. rhodesiense*, and thus plays a crucial role in a number of essential processes.[15,74] It is required for the nutrition of the parasite and is involved in the invasion of the host.[75,76,77] Furthermore, rhodesain is implicated in the crossing of the blood-brain barrier, which leads to the second stage of the disease.[78] The inhibition of this essential protease would thus cause the parasite's death.[15,74]

Inhibition of Rhodesain

There are several problems regarding the inhibition of rhodesain, and therefore the treatment of HAT. The parasite's ability to invade the host's immune system makes it a challenging task to find an effective treatment. There is no effective vaccine on the market and the drugs currently available are in some cases extremely toxic.[41,67,79,80] There are only four drugs against HAT on the market, from which three were developed more than 60 years ago and exhibit severe side effects. One even has a mortality rate of 1-5%, making the development of new drugs an urgent task.[41]

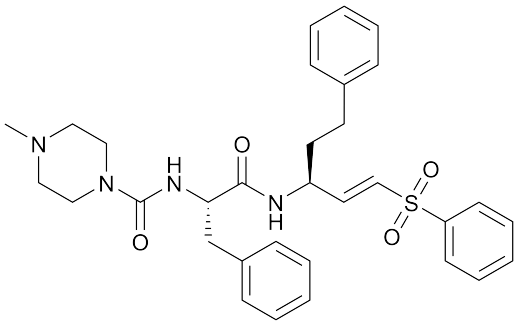


Abbildung 3.9: Structure of the peptidyl-based vinylsulfon inhibitor K11777.[28]

There are different kinds of potential inhibitors, like peptidyl-based inhibitors with warheads, similar to the cathepsin B inhibitors, and non-peptidyl agents.[41] Peptidyl vinyl-sulfones turned out to be a very promising molecule class regarding the inhibition of rhodesain. The compound K11777[28] (see Figure 3.9), in particular, was shown to perform covalent irreversible inhibition of the protease.[28]

4 Aim of this work

The development of new, effective drugs with little side effects is the desired goal in all areas of pharmaceutical research. In the case of disease-related enzymes, a detailed understanding of the inhibition mechanism is mandatory for the development of new selective drugs. Experimental data in this field provides information about the efficiency of different pharmaceutical components and allows for conclusions concerning the inhibition mechanism. However, the possibilities of such experiments have their limitations and can reach a point where further insights are required. Theoretical investigations can provide additional information by computer simulations of the respective systems. The combination of experiment and theory is therefore a promising connection, especially with regards to the examination of the exact reaction mechanism between inhibitor and enzyme.
In this work, two biochemical systems have been investigated *via* computational methods. In both cases, the examinations are based on preceding experiments and are aimed at the elucidation of the inhibition mechanism between different low-molecular components and the respective enzyme. The first topic concerns nitroalkene-based inhibitors in rhodesain, an enzyme that is connected with the African sleeping sickness. Experimental results showed that small structural changes like a methylation or halogenation of the inhibitor warhead leads to changes in the corresponding inhibition potentials and the mechanistic course. The inhibition potentials of the components are represented by their K_i and K_i^* values (depicted in Figure 4.1).

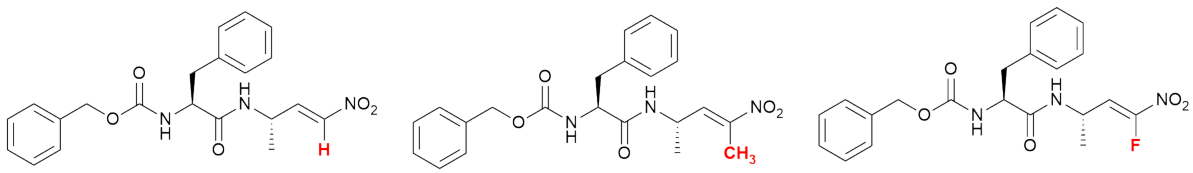


Abbildung 4.1: K_i and K_i^* values in [nM] for the evaluated nitroalkenes.[2,3]

A smaller value implies increased inhibition.For a detailed explanation and discussion, it is referred to the chapters 3.2 and 8.1. The aim of this work is thus to find the origin of this unexplained behaviour in order to provide information for the development of new and more effective compounds.
The second topic concerns the inhibition of cathepsin B with carbamate-based components. Until now, the mechanism of this inhibition has not been determined. Experimental data indicates that a reaction including an elimination and carbamylation could take place. However, there is no evidence for this. To investigate whether this interpretation is sound, the corresponding reaction courses are computed. All other possible reaction pathways are analyzed using thermodynamic considerations.

5 The Treatment of Biosystems in Computational Chemistry

5.1 Quantum Mechanical / Molecular Mechanical Hybrid Methods

The treatment of biomolecules in theoretical chemistry has always been a challenge, due to the size of the systems. In 1976, A. Warshel and M. Levitt published their idea of treating only small parts of a large system on a quantum mechanical (QM) level, while the rest is computed based on molecular mechanical (MM) methods. For those hybrid schemes, the phrase "QM/MM"has prevailed.[81] Taking enzymes as an example, the "important"part, where the catalysis takes place, is limited to the active site and its adjacent amino acids in most cases. Before QM/MM was established, molecular modeling on a QM basis was actually not possible, as the computational methods were restricted to a few hundred atoms.[82] By now, the areas of application not only include biosystems, but also solid-state[83,84] as well as inorganic and organometallic systems[85,86]. Figure 5.1 illustrates the concept of QM/MM hybrid methods.
For the successful performance of a QM/MM calculation, several decisions concerning the calculation setup have to be made. First of all, there are different possibilities to compute the total energy of a system. As the QM and MM regions interact, the energy cannot simply be computed by adding up the energies of the two isolated subsystems, especially when bond breaking is involved. Another challenging task is the computation of the covalent and non-covalent interactions between the regions, which also bears different possibilities.

Calculation of the Total Energy

As mentioned above, the calculation of the total energy of a system on a QM/MM level requires some thought. For this task, two main models have emerged, the *additive* and

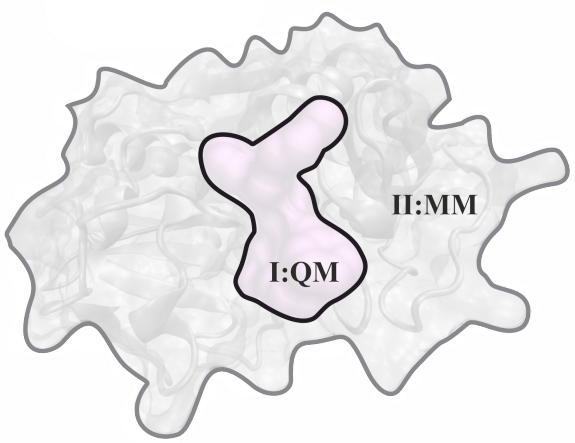


Abbildung 5.1: Scheme of a system with a QM/MM setup with an inner region I and an outer region II.

the *subtractive* scheme.[82,87] For the latter, an MM calculation of the whole system ($E_{MM}(I+II)$) and the inner region ($E_{MM}(I)$) is needed. Additionally, a QM calculation of the inner subsystem ($E_{QM}(I)$) is required. The total energy is then given by

$$E_{QM/MM}(I+II)^{sub} = E_{MM}(I+II) - E_{MM}(I) + E_{QM}(I). \tag{5.1}$$

The biggest advantage of this method is its simplicity. No explicit QM/MM coupling terms are needed and the different calculations can be performed independently. This means there is no need for special implementation. Secondly, there are no artifacts from link-atoms (see section 5.1), as long as the parametrization of the QM part is done carefully. However, there are some severe drawbacks of the method. Most importantly, the coupling between the QM and MM region is purely carried out on an MM level. This is especially problematic for the calculation of non-bonded electrostatical interactions between the two regions.[82]

The second way to calculate the total energy of the system is the additive scheme. Here, the outer system (II) is calculated at the MM level of theory($E_{MM}(II)$), while a QM calculation is performed on the inner system ($E_{QM}(I)$). Additionally, a QM/MM coupling term $E_{coupling}(I,II)$ is introduced. This term contains van der Waals, electrostatic and covalent interactions in the form of an explicit QM/MM method:

$$E_{QM/MM}(I+II)^{sub} = E_{QM}(I) + E_{MM}(II) + E_{coupling}(I,II). \tag{5.2}$$

The drawback of this, compared to the subtractive scheme, is the additional effort in computation and implementation. Nevertheless, it has become the model of choice in most QM/MM schemes.[88] In recent years, new methods have been developed that go beyond the classification of the standard definition of subtractive or additive schemes. Those multilayer methods introduce additional areas around the QM region into the system. A popular example for this is the substractive multilayer scheme ONIOM, a model where MM charges are included into the QM Hamiltonian.[87,89]
The coupling term $E_{coupling}(I,II)$ contains the bonded and non-bonded interactions, such as van der Waals and electrostatics. There are different embedding schemes to manage this task.

Embedding Schemes

The coupling between the QM charge density and the MM point charges requires accurate evaluation using a sophisticated method. Concerning electrostatic interaction, three main models have been developed, which differ in the level of mutual polarization.[88]
In mechanical embedding, the non-covalent interactions are treated solely at the MM level of theory (see section 5.4). This means that the MM part accounts for the QM part only as a region of rigid point charges. This scheme does not require additional effort during the computation and can be implemented easily into existing software. Those advantages, however, entail drawbacks in the quality of the calculations. If all interaction between the two subsystems is based on MM, the logic consequence is that the outer MM region cannot interact with the QM density, as the latter is only represented by rigid point charges. The MM region is thus not polarized by the QM density.
The more common electrostatical embedding includes the MM point charges into the QM Hamiltonian *via* one-electron terms. In this model, the QM subsystem is polarized by the MM area. The major drawback of this model is an overpolarization, which is obtained for atoms in close proximity to the QM charge distribution.[82] Additionally, the inclusion of an MM point charge into the Hamiltonian poses difficulties, as the parametrization of a force field is not automatically suitable for the representation of a QM charge.[82] Still, electrostatic embedding is one of the most popular embedding schemes in use today, especially for biomolecular applications.
The computationally most demanding, and for that matter also the most realistic embedding scheme, is called polarizable embedding. It includes a flexible MM charge

distribution that can react to the QM density and is thus polarized by it. Polarizable embedding is further divided into two models. In the first one, only the MM part is polarizable, whereas in the second one, a mutual polarization between the two subsystems takes place. In principle, the latter one would be the method of choice for a QM/MM calculation, but besides the higher effort in computation, there is only a small number of biomolecular force fields with these capabilities available today. However, the promising idea of the model causes progression in this research area.[90,91,92,93,94] Van der Waals interactions are most commonly treated solely on the MM level of theory, using the Lennard-Jones potential (equation 5.58).[82,95] An improvement in this area would be for example an optimization of parameters for B3LYP/AMBER.[96]

Treatment of Covalent Bonds

In some cases, the QM/MM border cuts through a covalent bond. This situation requires a certain theoretical treatment. The cleaved bond must in some way be capped, as an unsaturated bond would not describe the electronic character of the system properly. There are three common schemes available to deal with this problem, the link-atom scheme, boundary atoms and frozen localized orbitals.[82] This work will concentrate on the most widely used link-atom scheme. The cleaved bonds are saturated with the help of an additional atomic center, most commonly a hydrogen atom. These link-atoms are not part of the real system (see Figure 5.2). They are used for the computation of the inner QM subsystem, in order to correctly describe the capped atom Q1, but are not considered in the actual calculation of the total system energy.

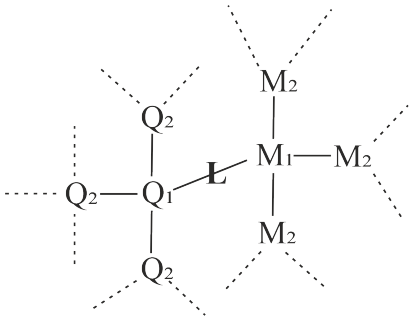


Abbildung 5.2: QM/MM boundary area with atom names according to the relative position. L is the linker atom and QM or MM designates atoms in the QM or MM area.

As previously mentioned, the main problem within this scheme is the overpolarization. Since the two adjacent atoms M1 and Q1 are extremely close, their non-bonded

interaction will be overestimated. A widely used approach for this issue is the shifting scheme.[88] Here, the point charge of the adjacent M1 atom is distributed to its neighbouring atoms M2 (see Figure 5.2).
QM/MM has become the method of choice for the calculation of large systems, especially in biochemistry. The number of possibilities for the setup of such a calculation requires a lot of thought and experience in order to obtain reproducible results and make realistic predictions. Of course, apart from the various embedding schemes, the QM and MM methods themselves are crucial for the performance of the hybrid calculation. The following sections will give an overview of QM and MM methods, as well as the technique of MD simulations.

5.2 Quantum Mechanical Methods

Parts of this section have already been published in the master thesis of the author.[97]

5.2.1 Hartree-Fock

The Hartree-Fock (HF) method provides the foundation of all wave function-based computational methods. It is an iterative approach using a single determinant as a wave function (equation 5.3):[98,99]

$$|\Psi_0\rangle = |\chi_1\chi_2...\chi_N\rangle. \tag{5.3}$$

The general idea behind the method is to minimize the energy of a system as a function of its spin orbitals $\{\chi_a\}$:[99]

$$\frac{\partial}{\partial\chi_a}\langle\Psi_0|\hat{H}_{el}|\Psi_0\rangle = 0. \tag{5.4}$$

Solving this equation by the utilization of Lagrangian multipliers leads to the Fock equation

$$\hat{f}|\chi_a\rangle = \epsilon_a|\chi_a\rangle, \tag{5.5}$$

where ϵ_a is the energy of spin orbital a in the mean field of all other electrons of the system and $\hat{f}$ is the effective one-electron Fock operator:[98,99]

$$\hat{f} = h(1) + \sum_{b \neq a} J_b(1) - \sum_{b \neq a} K_b(1). \tag{5.6}$$

In this equation, $h(1)$ is the core hamiltonian operator, which contains the kinetic energy and core interaction of one electron

$$h(1) = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}}. \tag{5.7}$$

J_b and K_b are the Coulomb- and Exchange operators:

$$J_b(1)\chi_a(1) = \left[\int dx_2 \chi_b^*(2)\frac{1}{r_{12}}\chi_b(2)\right]\chi_a(1) \tag{5.8}$$

$$K_b(1)\chi_a(1) = \left[\int dx_2 \chi_b^*(2)\frac{1}{r_{12}}\chi_a(2)\right]\chi_b(1) \tag{5.9}$$

For the computational description of molecules, the molecular orbitals ψ_i are often written as a linear combination of N atomic orbitals ϕ_{mu} (LCAO approach):

$$\psi_i = \sum_{\mu}^{N} c_{\mu i}\phi_\mu \tag{5.10}$$

Applying this approximation and converting the differential equations into the matrix formulation, the Fock equation becomes the Roothaan-Hall matrix equation:[100,101]

$$\mathbf{FC} = \mathbf{SCE}, \tag{5.11}$$

where **C** is the N x N matrix of the coefficients $c_{\mu i}$, **E** is a N x N diagonal matrix containing all orbital energies ϵ_i and **S** is the overlap matrix between two atomic orbitals ϕ_μ and ϕ_ν

$$S_{\mu\nu} = \langle \mu, \nu \rangle = \int \phi_{mu}^*(1)\phi_{nu}(1)d\tau_1. \tag{5.12}$$

The nontrivial solution of this eigenvalue problem provides a set of optimal spin orbitals. The HF energy is obtained as the expectation value of the Slater determinant built from these optimized spin orbitals.

However, the HF approach is problematic, because of the lack of electron correlation. The term correlation energy is defined by the difference between the Hartree-Fock energy and the actual energy of a system. Therefore, fundamental processes like

homolytic dissociation can not be described accurately by the HF approach.[99] To overcome this lack in description, post-Hartree-Fock methods (electron correlation methods) were developed. All post Hartree Fock methods build upon a Hartree Fock calculation, providing optimized spin orbitals. The wave function is then improved by taking more than one Slater determinant into account:

$$\Psi_{tot} = c_0\Phi_0 + \sum_{i,a} c_{i,a}\Phi_i^a + \sum_{ij,ab} c_{ij,ab}\Phi_{ij}^{ab} + ... \tag{5.13}$$

In this equation, the letters i and j represent indices for occupied orbitals, whereas a and b stand for virtual orbitals. Φ_i^a and Φ_{ij}^{ab} are excited state determinants including single and double excitations. There are various post HF methods that differentiate in their manner of determining the coefficients c_i.[99]

5.2.2 Møller-Plesset Perturbation Theory

Many Body Perturbation Theory

Many-Body Perturbation theory generally assumes that the problem to be solved (perturbed problem) only slightly differs from a solvable (unperturbed) problem.[99] This approach is applied to the Hamilton operator:

$$H = H_0 + \lambda H^{'}, \tag{5.14}$$

where H is the exact Hamilton operator, H_0 is the unperturbed operator and $H^{'}$ is the perturbation operator. λ is a parameter defining the strength of the perturbation. The perturbed Schrödinger equation is

$$H\Psi = W\Psi. \tag{5.15}$$

The wave function Ψ and energy W can be written in a taylor expansion in powers of the perturbation strength:

$$W = \lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + ... + \lambda^n W_n, \tag{5.16}$$

$$\Psi = \lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + ... + \lambda^n \Psi_n. \tag{5.17}$$

After substituting Ψ and W in equation 5.15 with the taylor expansion and sorting the terms by the power of λ, n perturbation equations of the n*th* order result.

Zeroth order ($\lambda^0 = 0$): $H_0\Psi_0 = W_0\Psi_0$.
First order (λ^1): $H_0\Psi_1 + H'\Psi_0 = W_0\Psi_1 + W_1\Psi_0$
Second order (λ^2): $H_0\Psi_2 + H'\Psi_1 = W_0\Psi_2 + W_1\Psi_1 + W_2\Psi_0$
n*th* order (λ^n): $H_0\Psi_n + H'\Psi_{n-1} = \sum_{i=0}^{n} W_i\Psi_{n-i}$

The resulting energies (up to the second order) are:

$$W_0 = \langle\Phi_0|H_0|\Phi_0\rangle \tag{5.18}$$

$$W_1 = \langle\Phi_0|H'|\Phi_0\rangle \tag{5.19}$$

$$W_2 = \sum_i c_i\langle\Phi_0|H'|\Phi_i\rangle = \sum_{i\neq 0} \frac{\langle\Phi_0|H'|\Phi_i\rangle\langle\Phi_i|H'|\Phi_0\rangle}{E_0 - E_i} \tag{5.20}$$

Møller-Plesset Perturbation Theory

In Møller-Plesset perturbation theory (MP), the unperturbed Hamilton operator is the sum of the Fock operators of all occupied electrons:[102]

$$H_0 = \sum_{i=1}^{occ} f_i \tag{5.21}$$

As the perturbation is defined as $H' = H - H_0$, H' contains the difference between the two-electron operator g_{ij} of H and the averaged two-electron operator $\langle g_{ij}\rangle$ of the Fock operator:

$$H' = H - H_0 = V_{ee} - \langle V_{ee}\rangle. \tag{5.22}$$

The unperturbed energy (MP0) then equals the sum of the occupied spin orbital energies ϵ_i:

$$W_0 = \sum_i \epsilon_i, \tag{5.23}$$

This is a worse approximation than the Hartree-Fock energy. The energy correction of first order W_1 is

$$W_1 = \langle\Phi_0|H'|\Phi_0\rangle = -\langle V_{ee}\rangle. \tag{5.24}$$

As a result, the MP1 energy equals the Hartree-Fock energy:

$$E(MP1) = W_0 + W_1 = E(HF) \tag{5.25}$$

The most common method in this field is MP2. Here, the energy correction is

$$W_2 = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\langle \Phi_0 | H' | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | H' | \Phi_0 \rangle}{E_0 - E_{ij}^{ab}}, \tag{5.26}$$

which gives the MP2 energy

$$E(MP2) = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{5.27}$$

The numerator of the term contains two-electron integrals of the coulomb and exchange interaction between the occupied orbitals $\phi_{i,j}$ and the virtual orbitals $\phi_{a,b}$. The denominator of equation 5.27 contains the orbital energies of the HOMO and the LUMO. As a consequence, the method fails for systems with a small HOMO-LUMO gap.

Spin Component Scaling (SCS)

Spin-Component-Scaled MP2 was proposed in 2003.[103] The idea of spin component scaling focuses on the fact that the correlation energy E_C between electron pairs with the same (SS) and opposite spin (OS) differ systematically. Physically speaking, the correlation of an SS electron pair refers to the Fermi hole (which results from the Pauli principle[104]), whereas the behaviour of OS pairs is connected with the Coulomb hole.

$$E_C = E_C^{SS} + E_C^{OS}. \tag{5.28}$$

The theoretical background of the method is based on the fact that in Hartree-Fock calculations, the correlation of SS pairs is already included through the anti-symmetric wave function and the resulting exchange terms. For OS pairs, no correlation is included yet. As a post-HF method, MP2 introduces OS correlation, but due to an original imbalance from HF, this contribution is still underestimated.[105] SCS-MP2 introduces empirical factors c to the correlation energy of a system, upscaling the OS correlation while the SS correlation is scaled down ($c_{OS} = 6/5$ and $c_{SS} = 1/3$).

$$E_C[SCS-MP2] = c_{SS}E_C^{SS}[MP2] + c_{OS}E_C^{OS}[MP2]. \tag{5.29}$$

In the original MP2 method, these factors are both equal to 1. There is no extra cost to the calculation, as the energy-contributions are simply scaled *via* the empirical factor. The performance of SCS-MP2 in reaction energy calculations was shown to be comparable to QCISD(T).[103] It is a trustworthy method for the computation of barrier heights[106] and geometries.[30,107,108] It is a very reliable method that corrects for many of the drawbacks from MP2, especially for ground state properties of most biologically relevant molecules. Regarding processes like the dissociation of molecules or spin contamination, the underlying weakness of HF gains influence on the performance of the method.[105] There is a number of variants, namely SOS-MP2, where c_{SS} is set to zero[109], MOS-MP2[110], that uses a long range correction in the correlation energy, or SCS-CC2 and SCS-CIS(D), which focus on excited state calculations[105]. As always in computational chemistry, the key step is to use the right method for the right system, which also accounts for the parametric scaling factors.

5.2.3 Density Functional Theory

Other than wave function-based methods, the density functional theory (DFT) describes a system by means of its electron density. This approach implies that the density contains all relevant information about a system, namely the number of electrons, the geometry and the nuclear charges. These quantities define the Hamilton operator of any system.[111,112] The number of electrons N can be obtained by integrating the density over all coordinates:

$$\int \rho(\vec{r})d\vec{r} = N \tag{5.30}$$

The poles of the electron density contain information about the geometry, as they correspond to the position of the nuclei. The charge of a nucleus is indicated by the mean density at its position.[112] Within DFT, several advantages arise, compared to wave function-based methods. The electron density is a property that can be measured experimentally. In addition, the dimension of the calculations decreases drastically, from a 4N dimensional problem for wave function-based methods (one spin and three spacial coordinates per electron) to a three-dimensional problem within DFT. This is the main reason why DFT is widely used for systems with a large number of elec-

trons. The application of DFT for quantum chemical problems arose in the 1990s.[111] Since then, the size of the investigated systems grew drastically and therefore had a great impact to computational research, especially in a biochemical context. Another remarkable aspect about DFT is that within this method, it is (theoretically) possible to get an exact solution for any many-electron system. However, so far this remains unfeasible. The following sections give an overview about basic ideas, theorems and variants of DFT.

Thomas-Fermi Model

The idea to use electron density as a descriptive representative of a system originated from Thomas in 1927.[113] He made the assumption that an electronic state of a homogeneous electron gas (HEG) is a solution of the Schrödinger equation (in its corresponding form). The Schrödinger equation was introduced one year before, in 1926.[114] Thomas also assumed that the nuclear-electron interaction depends exclusively on the position and charge of the nuclei. Based on that, he formulated the first kinetic energy functional together with Fermi:[113,115]

$$T^{TF} = C_F \int d^3\vec{r}\rho^{5/3}(\vec{r}), \tag{5.31}$$

where ρ is the electron density and the expression C_F corresponds to

$$C_F = \frac{3}{10}(3\pi^2)^{2/3}. \tag{5.32}$$

Within the Thomas-Fermi model, the nuclear-electron and electron-electron contributions were treated in a classical way, leading to the following energy expression:

$$E_{TF}[\rho(\vec{r})] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r})d\vec{r} - Z \int \frac{\rho(\vec{r})}{r} d\vec{r} + \frac{1}{2} \int \int \frac{\rho(\vec{r_1})\rho(\vec{r_2})}{r_{12}} d\vec{r_1} d\vec{r_2}. \tag{5.33}$$

This made it possible to calculate the energy of a system exclusively by means of its electron density. Nevertheless, the procedure had some severe drawbacks and did not yield reliable quantitative results.
In 1930, Dirac addressed this problem by introducing the first exchange functional (equation 5.34).[116] The phenomenon of exchange energy was formulated earlier in the same year by Fock.[117]

$$E_x^{LDA} = -\frac{3}{4}(\frac{3}{\pi})^{1/3} \int d^3\vec{r}\rho^{4/3}(\vec{r}) \tag{5.34}$$

Even though this step was an improvement, it was still not possible to describe fundamental properties like chemical bonding with it.

Hohenberg-Kohn Theorem

In 1964 the Hohenberg-Kohn theorem was formulated by Pierre Hohenberg and Walter Kohn.[118] The theorem provides the basis for density functional theory, as it addresses the question whether it is reasonable to take the electron density of a system as the uniquely defining quantity.

The first theorem states, that *"the external potential $V_{ext}(\vec{r})$ is (to within a constant) a unique functional of $\rho(\vec{r})$; since, in turn $V_{ext}(\vec{r})$ fixes $\hat{H}$ we see that the full many particle ground state is an unique functional of $\rho(\vec{r})$"*.[118] In other words, the external potential $V_{ext}(\vec{r})$ has an unique dependence on both the electron density and the wave function. This allows to take the electron density of a system as the system representative. The two functionals for the electron-electron interaction and the kinetic energy of the electrons are system-independent.
The theorem can be proven by *reductio ad absurdum*, thus considering that the ground state electron density is not an unique functional of the ground state wave function and the Hamiltonian. This assumption leads to a mathematical contradiction and therefore shows the unique connection between ground state density, wave function and Hamiltonian. Since the density also contains all information about the system, namely the number of electrons N, the charges and the position of the nuclei Z_A and R_A, the first Hohenberg-Kohn theorem can be formulated as[111]

$$\rho_0 \rightarrow \{N, Z_A, R_A\} \rightarrow \hat{H} \rightarrow \Psi_0 \rightarrow E_0. \tag{5.35}$$

Equation 5.35 shows that the ground state energy E_0 is a functional of the ground state electron density with E_0 being defined as

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0]. \tag{5.36}$$

As stated above, the external potential $V_{ext}(\vec{r})$ (represented through the functional $E_{Ne}[\rho_0]$ in equation 5.36) is the only system-dependent component of equation 5.36.

It is called external potential or electron nuclei potential. $T[\rho]$ and $E_{ee}[\rho]$ are universal functionals and can be combined in the Hohenberg-Kohn functional:

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho]. \tag{5.37}$$

If this functional would be known exactly, it would yield the exact energy and due to its system independence, molecules from a very small size up to vast biomolecules could be computed with absolute accuracy.

The second theorem addresses the question whether the observed density really is the ground state density. To achieve this, it has to have the lowest energy for the system. Hohenberg and Kohn showed that the variational principle is also a valid approach concerning the electron density.

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{Ne}[\tilde{\rho}] + E_{ee}[\tilde{\rho}]. \tag{5.38}$$

In this equation, $\tilde{\rho}$ stands for any trial density that represents a system. This means any density other than the ground state density will result in higher energies. Using the unique connection between the density and the Hamiltonian and thus the wave function, the second Hohenberg-Khon theorem can also be written as

$$\langle\tilde{\Psi}|\hat{H}|\tilde{\Psi}\rangle = T[\tilde{\rho}] + V_{ee}[\tilde{\rho}] + \int \tilde{\rho}(\vec{r})V_{ext}d\vec{r} = E[\tilde{\rho}] \geq E_0[\rho_0] = \langle\Psi_0|\hat{H}|\Psi_0\rangle. \tag{5.39}$$

Kohn-Sham Approach

In 1965, Kohn and Sham published a paper that revolutionized DFT. This eventually established a way for density functional theory to correctly describe molecules.[111,112,119] The problem older DFT approaches had regarding the proper description of covalent bonds occurs particularly due to the bad description of the kinetic energy functional of electrons. It is known that wave function-based methods, such as Hartree-Fock, describe the kinetic energy accurately. Kohn and Sham thus suggested to exactly calculate all known contributions and then put the remaining unknown parts into one approximate functional, known as the exchange-correlation functional.[111] The connection between a density-based and a wave function-based description was established *via* the so-called Kohn-Sham orbitals. For this purpose, Kohn and Sham constructed a non-interacting reference system, built from one-electron functions (i.e. the Kohn-Sham orbitals ϕ_i), which are moving in an effective potential

V_s. This system of non-interacting electrons is exactly described by one Slater determinant. For this system, the operator of the kinetic energy is known from HF-theory as

$$T_{HF} = -\frac{1}{2}\sum_{i}^{N}\langle\phi_i|\nabla^2|\phi_i\rangle, \tag{5.40}$$

which yields the exact kinetic energy.
Thus in Khon-Sham theory V_s is chosen in a way that the density of this non-interacting reference system $\rho_s(\vec{r})$ equals the density of the real system $\rho_o(\vec{r})$:

$$\rho_s(\vec{r}) = \sum_{i}^{N}\sum_{s}|\phi_i(\vec{r},s)|^2 = \rho_o(\vec{r}). \tag{5.41}$$

Within the Kohn-Sham approach, the respective functionals may furthermore be interpreted in a new way.
The functional for the kinetic energy $T[\rho]$ becomes $T_s[\rho]$, which gives the exact solution for the kinetic energy for a system of non-interacting electrons. Analogous, the exact functional for the electron-electron interaction $E_{ee}[\rho]$ becomes $J[\rho]$, which contains the Coulomb interaction for the electrons. As a result, the energy of a system can be described as

$$E[\rho] = T_s[\rho] + J[\rho] + E_{Ne}[\rho] + E_{xc}[\rho] \tag{5.42}$$

with

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]). \tag{5.43}$$

As stated above, the exact solution for the correlation-exchange functional would yield the exact energy for every system, regardless of its size. As the exact functional is being unknown, this created room for countless different approaches for this functional. These approaches have no hierarchy, so a systematical improvement is not possible. Nevertheless, they are often categorized with the Jacob's ladder of DFT.[120] The different rungs of the ladder are defined by the dependency of E_{xc} on the electron density.
In the first rung (LDA) the functional exclusively depends on the electron density. In the second rung, the derivative of the density is additionally considered (Generalized Gradient Approximation-GGA). The third rung (meta-GGA) additionally depends on

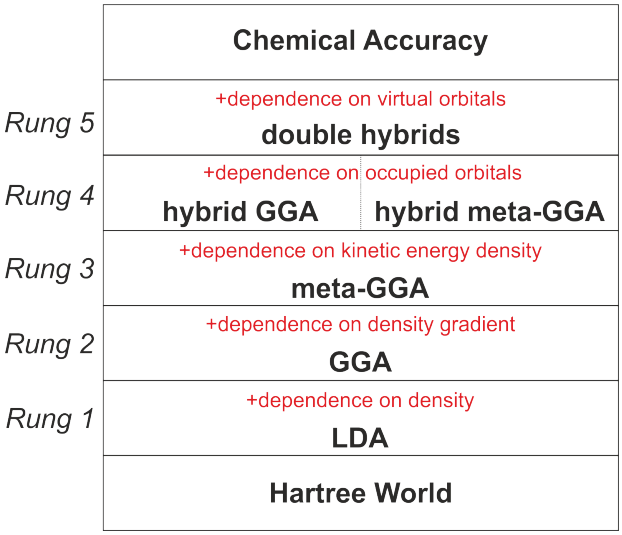


Abbildung 5.3: Jacobs ladder of DFT.[121]

the kinetic energy density of the system. The fourth rung has an additional dependence on occupied orbitals, whereas the fifth rung depends on occupied and virtual orbitals. This means, that in both cases (fourth and fifth rung), a certain amount of HF exchange is added, whereas the fifth rung contains contributions derived from perturbation theory (double hybrid). The functionals often also contain (semi-)empirical parameters which were often optimized for certain kind of systems or properties. As a result, the different functionals have to be adapted to the system and specific target properties of a calculation.

In this work, the functional B3LYP,[122] named after Becke, Lee, Yang and Parr, was widely used as the QM method in QM/MM calculations. Being a hybrid GGA functional, it is assigned to the third rung of the Jacob's ladder. It also contains semi-empirical parameterization and is a functional that is routinely used for many different kind of systems and calculations. There are more modern functionals like the Minnesota family (M06L,[123] M11,[124]...) or double hybrids such as B2PLYP[125] or PBE0-DH.[126] Especially for the fifth rung, the computation time increases drastically, but gives very reliable results if applied to the right kinds of system. Although DFT still has some drawbacks regarding its accuracy and is strongly dependent on which functional is used for which kind of system, it is widely employed in computational chemistry. The scaling factor $O(N^{3-4})$ is remarkable for this level of accuracy, which makes DFT the method of choice, especially for large systems.[112]

5.3 Semi-Empirical Approaches

The main problem with *ab initio* methods is their computational cost and basis set dependency. Semi-empirical methods try to conquer this problem by introducing a number of approximations. First of all, valence electrons are only explicitly treated with a minimal atomic orbital basis, while the core electrons are considered as a reduced nuclear charge. Additionally, some of the two-electron integrals of the Fock matrix (equation 5.11) are neglected. To compensate for these simplifications, the remaining integrals are modified by parametrization, based on empirical measurements or highly accurate *ab initio* calculations.[127]
In order to decide which integrals can be excluded from the calculation, it is necessary to consider the Fock matrix (equation 5.11) more closely. The matrix consists of a sum of the one-electron matrix **H** (core Hamiltonian) and two-electron matrix **G**. The matrix elements are

$$F_{\mu\nu} = H_{\mu\nu} + G_{\mu\nu}. \tag{5.44}$$

$H_{\mu\nu}$ contains the kinetic energy of the electrons and the nuclei-electron attraction for the atomic orbitals μ and ν:

$$H_{\mu\nu} = \langle\mu| - \frac{1}{2}\nabla^2|\nu\rangle - \sum_A \langle\mu|\frac{Z_A}{r_{iA}}|\nu\rangle. \tag{5.45}$$

The two-electron matrix $G_{\mu\nu}$ describes the electron-electron pair interactions. It contains the orbital coefficients in the form of the density matrix $P_{\lambda\sigma}$ and all two-electron integrals of the system.

$$G_{\mu\nu} = \sum_\lambda \sum_\sigma P_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)] \tag{5.46}$$

The density matrix is defined as:

$$P_{\lambda\sigma} = 2\sum_{i=1}^{N_{occ}/2} c_{\lambda i}c_{\sigma i}. \tag{5.47}$$

It is important to consider the two-electron integrals in order to decide which of the interaction terms are to be neglected. Equation 5.48 and 5.49 show the interaction terms in the basis of spacial orbitals.

They consist of a Coulomb term

$$(\mu\nu|\lambda\sigma) = \int dr_1 dr_2 \mu^*(1)\nu(1)\frac{1}{r_{12}}\lambda^*(2)\sigma(2) \tag{5.48}$$

and an exchange term

$$(\mu\sigma|\lambda\nu) = \int dr_1 dr_2 \mu^*(1)\sigma(1)\frac{1}{r_{12}}\lambda^*(2)\nu(2). \tag{5.49}$$

Zero Differential Overlap

The Zero Differential Overlap (ZDO) approximation is a central quantity for semi-empirical calculations. The differential overlap (DO) is defined as

$$DO(\mu\nu) = \mu(1)\nu(1)d\tau, \tag{5.50}$$

thus it contains information about the position of an electron, depending on the relative position of the orbitals μ and ν. Inserting this definition into the equations 5.48 and 5.49 yields

$$(\mu\nu|\lambda\sigma) = \int DO(\mu\nu)DO(\lambda\sigma)\frac{1}{r_{12}} \tag{5.51}$$

and

$$(\mu\sigma|\lambda\nu) = \int DO(\mu\sigma)DO(\lambda\nu)\frac{1}{r_{12}}. \tag{5.52}$$

Considierung all two-electron integrals, the computational cost of a calculation for a system with N atomic orbitals would have a scaling factor of N^4.[128] Taking the position of the atomic centers into account, the integrals can be separated into one-, two-, three- and four-center terms. The idea of semi-empirical methods is to reduce the scaling factor by neglecting some of those integrals, depending on how many different atom centers are involved.

The Zero Differential Overlap (ZDO) approximation was introduced in the late 1950s.[129,130] According to the ZDO, the differential overlap $DO(\mu\nu)$ of two atomic orbitals μ, ν is set to zero if they are located on different atomic centers μ_A and ν_B. This approximation zeros all three- and four center integrals. On the basis of this, three different models were developed: CNDO (Complete Neglect of Differential Overlap), INDO (Intermediate Neglect of Differential Overlap) and NDDO (Neglect of Diatomic

Differential Overlap), the latter being the most widely used model nowadays.[131] Depending on the model, the remaining one- and two-center integrals are further sorted out. The well known methods MNDO[132], AM1[133] and PM3[132,134] all belong to the family of NDDO. The difference between them lies in their parameterization.

5.4 Molecular Mechanics

In molecular mechanics (MM), molecules are described classically using a ball-and-spring model.[135] Atoms are represented by dots without expansion with a certain mass m and charge q. Bonds between the atoms are described as springs. Depending on the strength of the bond, the spring stiffness can be harder or softer. The biggest difference between molecular and quantum mechanics is the treatment of the electrons.[136] In quantum mechanical approaches, the electrons are treated explicitly in a three-dimensional distribution, whereas in molecular mechanics, they are not considered explicitly. Information about the geometry of a molecule is included using the angles Φ and dihedrals θ between the atoms. Non-covalent interactions like Coulomb- or van der Waals interactions are included with classical electrostatic interaction terms.

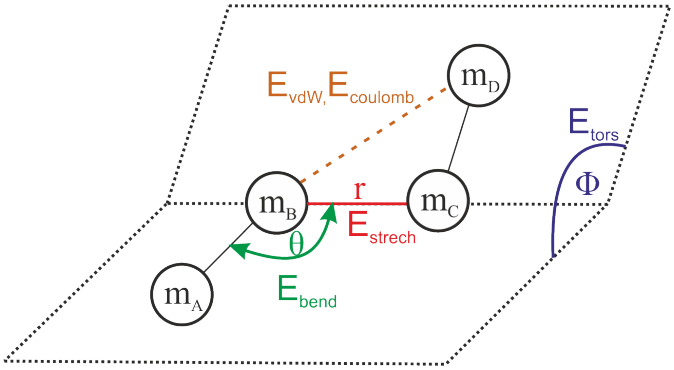


Abbildung 5.4: Parameters for molecular mechanics calculations.

In this model, the different energy contributions are assigned to classic mechanical energy terms. The terms are functions of the parameters given by the atom types and geometry of the system (see Figure 5.4). The energy of a system is described as a sum of different energy contributions.[99]

$$E = E_{strech} + E_{bend} + E_{tors} + E_{vdW} + E_{coulomb} + E_{cross}. \tag{5.53}$$

The energy E represents a snapshot of the system in one certain conformation. As soon as one bond or angle changes, the energy changes simultaneously.
The first three energy terms contribute to the bonded interactions, whereas E_{vdW} and $E_{coulomb}$ describe the non-bonded intermolecular forces. The last term E_{cross} contains possible couplings between the other terms, which depends on the respective force field.
E_{strech} describes the bond between two atoms A and B. It therefore contains a force constant k^{AB} as well as the distance of the atoms relative to their equilibrium distance r_0^{AB} and is summed up over the number of bonds in the system N:[99,137]

$$E_{strech} = \sum_{i=1}^{N} \frac{k^{AB}}{2}(r^{AB} - r_0^{AB})^2 \tag{5.54}$$

The second term E_{bend} contains information about the angle-dependent energy contributions. This sum is running over M, the number of directly bonded angles. Equally to E_{strech}, the term depends on a force constant k^{ABC} and the difference between an angle Φ and the equilibrium angle Φ_0.[99,137]

$$E_{bend} = \sum_{i=1}^{M} \frac{k^{ABC}}{2}(\Phi^{ABC} - \Phi_0^{ABC})^2 \tag{5.55}$$

Both of these energy expressions can be interpreted as a taylor series that is terminated at the second order. This implies the use of a harmonic potential. The last of the bonded energy contributions is E_{tors}. Here, the rotation around a bond is described with the spherical angle θ between four adjacent atoms. Unlike E_{strech} and E_{bend}, a Fourier series underlies this expression containing information about the rotational barrier V^{ABCD} as well as a parameter n, that describes the periodicity.[99]

$$E_{tors}^{ABDC} = V^{ABCD} cos(n^{ABCD}\theta) \tag{5.56}$$

For $n = 1$, the rotation is periodic by 360°, for n=2 it is periodic by 360°/2 and so on. The remaining non-bonding terms are E_{vdW} and $E_{coulomb}$. The latter describes the classic coulomb interaction between the partial charges q_1 and q_2 of a system. The potential depends on the inter-atomic distance r^{AB} and contains the dielectric constant ϵ.[99]

$$E_{coulomb}^{AB} = \frac{q_1 q_2}{\epsilon |r^{AB}|} \tag{5.57}$$

The last contribution includes all weak non-bonded interactions in the term E_{vdW}. The van der Waals energy expression contains dispersion and dipole-dipole interactions. The most common potential that is used for its description is the Lennard-Jones potential[95]:

$$E_{vdW}^{AB} = \frac{c_1}{(r^{AB})^{12}} - \frac{c_2}{(r^{AB})^6}. \tag{5.58}$$

The repulsive first term dominates the energy for short distances and represents the Pauli-repulsion. For medium distances, the second term outweighs the first one, leading to an attractive force. The Lennard-Jones potential is sometimes also referred to as the "12-6-potentialänd is widely used. Force field methods can be applied to large systems that cannot be computed with high level *ab initio* methods, due to their size. Especially for the description of biological systems, they provide the possibility for theoretical investigations. Nevertheless, it has to be kept in mind that these methods are highly dependent on the force field parameters used and that only ground-state properties can be calculated. Additionally, it is not possible to evaluate bond breaking with MM methods.

Molecular Dynamics

Molecular Dynamics (MD) simulations are an important tool in theoretical chemistry. They were first introduced in the late 1950s by B.J. Alder and T. W. Wainwright.[138,139] Regarding proteins, they can provide insight to inhibition mechanisms or protein folding. To perform a MD simulation, the energy expression E from equation 5.53 is needed. Additionally, an equation that introduces motion to the examined system is introduced. For the classic description of a system with low velocities (relative to the speed of light), Newton's equation of motion can be used.[140]

$$\frac{dV}{dr} = F(r) = m\frac{dv}{dt} = m\frac{d^2r}{dt^2}. \tag{5.59}$$

The equations provide information about how a particle with the constant mass m reacts to a force F. The "reactionöf the particle is described by the right side of the equation. It is expressed through the derivative of the velocity with respect to the time, which equals the second derivative of its position with respect to the time. In other words, the position changes with the time. In the left part of equation 5.59, the connection between the potential V and the force is defined, the latter being the

derivative of V with respect to the position. While performing a MD simulation, the computer solves the differential equations above with numerical integration as a series of a finite number of small steps. There are different logarithms for this calculation, a very common one being the Verlet algorithm.[141]

$$v(t + \Delta t) = v\left(t + \frac{1}{2\Delta t}\right) + \frac{f(t + \Delta t)}{2m}\Delta t. \tag{5.60}$$

The first step of each MD is the assignment of initial velocities for example with the help of the Maxwell-Boltzmann distribution for a given temperature. The initial coordinates are normally taken from the input data file. In the second step, the resulting force is calculated with a method of choice. In most cases, fore field methods are used, as the usage of *ab initio* methods would be extremely time-consuming. Although MD simulations are often performed on a low level of theory, they play an important role in computational analysis. In order to yield reliable results, sufficient sampling is required.

6 Computational Methods

MD Simulations

All MD simulations were performed using the AMBER program package.[142,143] The applied force fields were ff99SB[144] and ff14SB,[145] whereas GAFF[146,147] in combination with the program antechamber[147,148] was used for the parametrization of the ligands. An octahedral TIP3P[149] water shell with a radius of 30 Å was added around the parameterized system with help of the module AMBERTOOLS 1.4.[142,143] Spherical boundary conditions were applied to the system and the net charge was neutralized by the addition of sodium ions. A procedure containing three distinct calculation steps was applied to all systems in preparation for the actual MD run. First, a geometry optimization of only the water shell was performed, followed by an optimization of the whole system. After that, a short MD simulation, running over 400 ps with time steps of 1 fs was performed. The protein was kept within weak constraints during the simulation and the temperature was gradually increased from 0 K to 300 K. The simulation times of the subsequent MD calculations were adapted to the respective problem. The specific data is given in the respective sections. For temperature controlling, a Langevin thermostat was used. The SHAKE algorithm was applied to bonds that involved a hydrogen atom.

The simulation (QM/MM MD) discussed in chapter 7.3.2 was performed using the ChemShell 3.6 program.[150,151] The QM area was represented by an electrostatic potential, calculated with the semi-empirical method AM1. For the embedding scheme, the same settings as described below (see "QM/MM") were applied. The simulation ran over a time of 2 ns with fs time steps at 300 K, using a Bendersen thermostat. The applied force field was ff14SB.

QM Calculations

All quantum mechanical *ab-initio* calculations were performed, using the Turbomole program package (version 6.6 and 7.1[152,153,154]). For the simulation of a solvent surrounding, the implicit solvation model COSMO[155] was used. Semi-empirical calculations were performed using the MNDO 7.0 program package.[156]

QM/MM Calculations

The program ChemShell 3.6 was used as an interface within the performed QM/MM calculations. All systems calculated on a QM/MM level were taken from preceding MD calculations and therefore included the TIP3P water shell and the neutralizing sodium ions. The applied boundary treatment included electrostatic embedding in combination with the charge shift method and the link atom approach. All QM/MM optimizations were performed using the DL-FIND module[157] of ChemShell with an active area of 10 Å around the reaction center of the enzyme. The rest of the system was kept fixed during the calculations. For all QM/MM calculations, the MM part was calculated with the AMBER program package and the ff14SB force field. Therefore, the MM method is not explicitly mentioned in the respective chapters for every QM/MM calculation. The QM calculations were performed with the Turbomole 6.6 and 7.1 program package. For the NEB calculations, the same boundary treatment as described above was utilized. The calculations were performed with different counts of frames, reaching from 7 up to 17.

The programs used for the analysis and visualization of the results were OriginPro 2016,[158] VMD 1.9.1,[159] ChemDraw and CorelDraw 2018.

7 Inhibition Mechanisms of Carbamates in Cathepsin B

In this chapter, carbamate-based inhibitors for cathepsin B are discussed. The following sections are structured as follows: First, after a short introduction to the topic, a summary is provided. The detailed discussion of the various computational steps investigated is given subsequently. In addition to the Nudged Elastic Band (NEB) calculations performed for the analysis of certain mechanisms, potential energy surfaces and reaction profiles were computed. This proceeding is based on the fact that while NEB calculations can provide reliable results concerning the mechanistic process, they are computationally extremely demanding and their interative scheme does not converge in many cases. In order to gain maximal insight into the mechanism, both types of calculations were performed.

7.1 Background and Previous Experimental Results

Carbamates are known to be potent inhibitors for choline esterases. A popular example thereof is the inhibition of acetylcholine esterase (AChE).[160,161,162] AChE is responsible for the hydrolysis of the neurotransmitter Acetylcholin (ACh) in the nervous system.[161] The inhibition of the enzyme has been found to play a potential role in the treatment of Alzheimer's Disease.[163] Carbamate-based inhibitors carbamylate the active site amino acid of choline esterases (e.g. serine) to interfere their catalytic activity.[163] The carbamylation can occur with help of an isocyanate intermediate, formed in an elimination reaction of the carbamate according to an E1CB mechanism.[163,164,165,166] Alternatively, a covalent adduct can be formed through a direct nucleophilic attack at the carbamate carbon.[166,167,168,169] Experimental data regarding cathepsin B and carbamate-based inhibitors from the group of Prof. Guetschow in Bonn showed, that an acidic hydrogen atom at the carbamate nitrogen leads to a better inhibition potential (a detailed discussoin is given below). These results indi-

cate an E1cB elimination as the inhibition mechanism, as for such a reaction, an acidic hydrogen and a poor leaving group (LG), linked to two adjacent atoms, are necessary (see Figure 7.1).[170] The mechanism is initiated by the abstraction of the proton, forming an anion. The double bond is subsequently formed trough the expulsion of the LG.

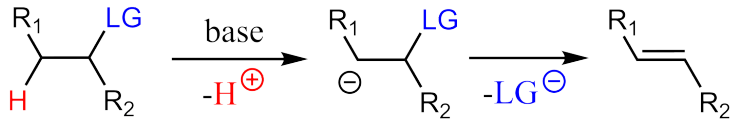


Abbildung 7.1: E1cB reaction mechanism.

Figure 7.2 schematically shows the elimination of carbamates by the degradation of the carbamate-based insecticide ethiofencarb, a pesticide used for the control of aphids on vegetables, corn and fruit.[171,172] Concerning carbamates, the acidic proton (red in Figure 7.2) is bound to the nitrogen atom. The alcohol, formed during the reaction (blue in Figure 7.2), is known to be a poor leaving group.

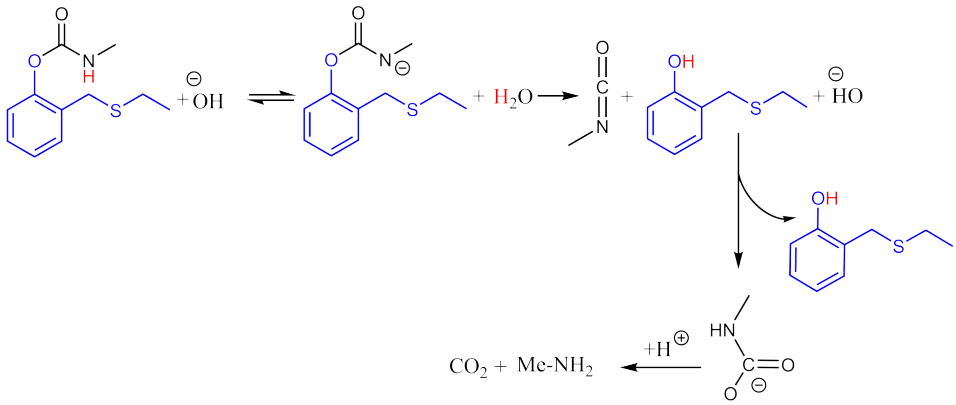


Abbildung 7.2: Degradation of ehtiofencarb in an alkaline medium.

The nitrogen atom is deprotonated in the first step of the reaction and forms an anion, which dissociates into an isocyanate and a phenolate derivative. The phenolate is protonated to an alcohol. In the subsequent step, a carbamic acid anion is formed through the nucleophilic attack of hydroxide. The unstable compound dissociates into carbon dioxide and a primary amine after the protonation of the nitrogen. The carbamate-based pesitcide ethiofenarb (popularly known as Croneton), undergoes this degradation mechanism when exposed to rough weather conditions or during storage periods.[165,171,172] In its function as an insecticide, ethiofencarb carbamylates the active site of AChE.[171,173]

Carbamates in Cysteine Proteases

Carbamate derivatives have been shown to be potential inhibitors of esterases and a number of cathepsins.[167,168,169,174,175] In the group of Prof. Guetschow, the four peptidylic carbamates displayed in Figure 7.3 are investigated (among others) as potential inhibitors for cathepsin B. Only two of the compounds (1 and 2) have an acidic hydrogen atom, making an E1cB mechanism possible. The other two compounds (3 and 4) are the respective derivatives with an α methyl group.

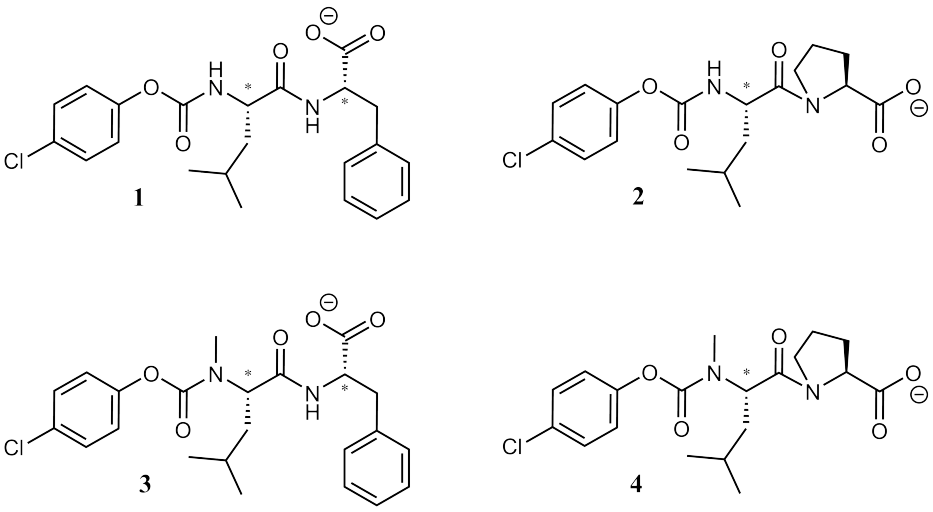


Abbildung 7.3: Investigated compounds from the group of Prof. Guetschow.

Based on the reaction behavior of carbamates reported in the literature, the initial step of the inhibition is assumed to be an elimination, according to an E1cB mechanism.[161,163,164,165,171] The measurements of the four compounds revealed an enzymatic rest activity of 2% and 4% for compound 1 and 2 inside cathepsin B.[1] In contrast, the activity of the inhibited enzyme with the compounds 3 and 4 were 85% and 98%, which indicates that the acidic α hydrogen is required for the inhibition.[1] All measured values refer to an inhibitor concentration of 20 μM. More detailed information can be found in the appendix of this work. Docking calculations of the carbamate-based inhibitor and its respective isocyanate shown in Figure 7.4 were performed by the group of Prof. Guetschow. The compounds were docked into the active site of cathepsin B (crystal structure 1HUC[26]). The resulting structures are taken as a basis for theoretical investigations of the proceeding of the proposed mechanism (see Figure 7.5).

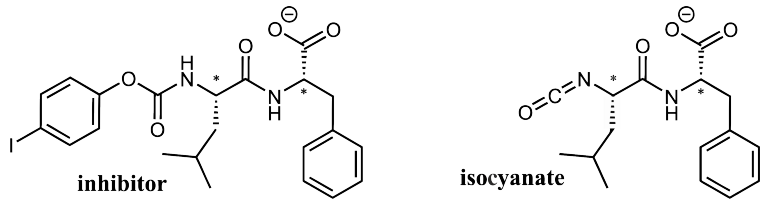


Abbildung 7.4: Investigated inhibitor and its respective isocyanate.

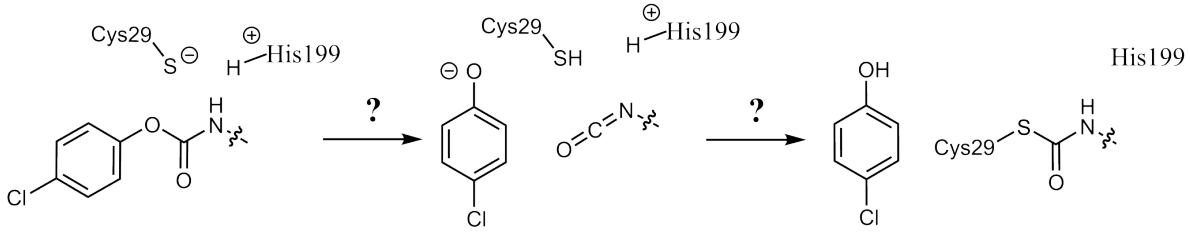


Abbildung 7.5: Reactant, intermediate and product structures of the proposed mechanism for a carbamate-based inhibitor in cathepsin B.

7.2 Summary

As already mentioned in the beginning of this chapter, a summary of all relevant results is given, before a detailed discussion of the respective topics follows. The reaction mechanisms of the system were investigated using different computational approaches including NEB calculations as well as reaction profiles and potential energy surfaces (PES). Even though NEB calculations provide more information compared to reaction profiles and PES, their iterative scheme hardly converge for complex reaction mechanisms. Therefore, other types of calculations were additionally performed. Figure 7.6 shows an overview of possible reaction mechanisms between carbamate-based inhibitors and the active site of cathepsin B.

Possible Reaction Mechanisms

In a first step, the various reaction pathways of the carbamate-based inhibitor in the active site of cathepsin B were evaluated *via* thermodynamic calculations (SCS-MP2/aug-cc-pVDZ with COSMO; ϵ=78.93). More detailed information about the applied methods is given in Chapter 6. For this purpose, a small model system was set up, containing only the side chains of histidine and cysteine (the backbone was replaced by a methyl residue) as well as the warhead of the inhibitor.

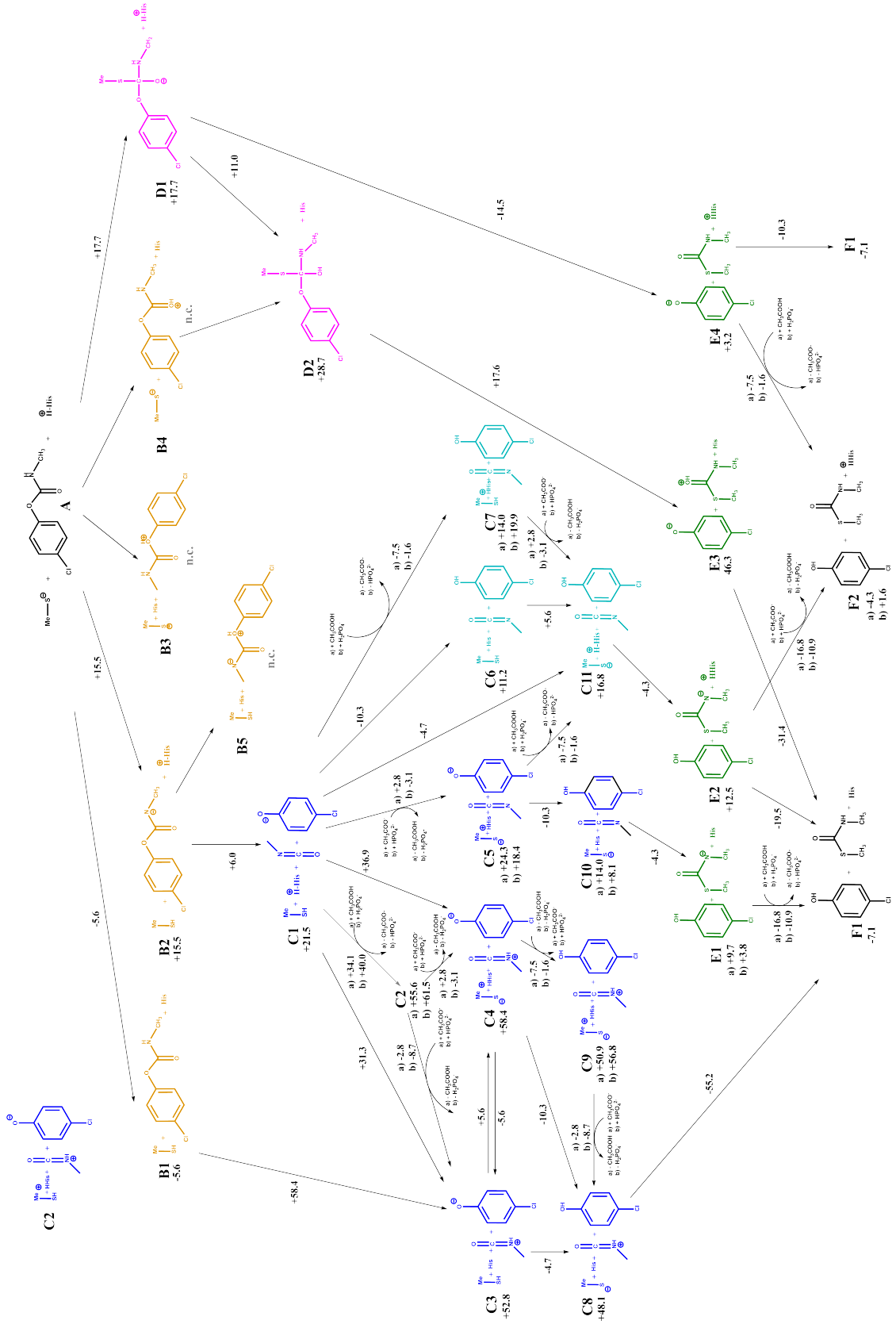


Abbildung 7.6: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; $\epsilon = 78.93$). Energies are given in kcal/mol; n.c.: not converged.

Calculations for the structures **B3**, **B4** and **B5** did not converge, which is in agreement with chemical intuition, as both of the protonated carbamate oxygen atoms are extremely weak bases. The reaction from **A** to **B1** is most exothermic compared to all other initiating steps. This mechanism describes a proton transfer from His199 to Cys29. However, as cathepsin B is known to have a zwitterionic active site that is conserved through the enzymatic environment, this pathway would be energetically unfavourable, when the whole enzyme is considered in the calculations.[43] Additionally, the reaction enthalpies of the subsequent steps exceed 50 kcal/mol, which excludes these pathways from further investigations.

For the remaining reactions, two main pathways arose that primarily differ in the function of Cys29 in the first step of the mechanism. For the mechanisms staring with **A** → **D1** (see Figure 7.6), cysteine initiates the reaction by a nucleophilic attack at the carbamate carbon. However, all pathways starting with **A** → **B2** (see Figure 7.6) contain an isocyanate intermediate, formed through the dissociation of the inhibitor. In these cases, the Cys29 acts as a base and deprotonates the carbamate group in the first step of the reaction. The subsequent dissociation of the inhibitor completes the E1cB elimination reaction, initiated by the deprotonation through Cys29. As the experimental data indicated an E1cB elimination reaction being the first step (see Chapter 7), all pathways leading over **D1** were also excluded from further investigations. Based on these considerations, the pathways containing **A** → **B2** → **C1** (see Figure 7.6) remain. Commencing from **C1**, there are seven possible reaction pathways, which all describe a proton transfer reaction. In three of them (**C2**, **C3** and **C4**) the isocyanate nitrogen is protonated forming a cation. All of those reactions have reaction enthalpies larger than 30 kcal/mol and are therefore not further investigated. The reactions from **C1** to **C5** (deprotonation of Cys29) and to **C7** (protonation of phenolate) include external buffer molecules. Thus, these pathways were also not further studied. Although a model system containing an acetate buffer is investigated in Chapter 7.4.1, at the QM level of theory, the buffer was removed for the computations using the QM/MM hybrid approach.

The remaining two reactions both entail a protonation of phenolate, either through His199 (**C6**) or Cys29 (**C11**). Histidine is a more potent proton donor, which is reflected in the lower reaction enthalpy compared to Cys29 as proton donor (around 4 kcal/mol). However, during MD simulations of the system, the arrangement of the active site amino acids and the inhibitor in cathepsin B showed a constantly smaller distance between Cys29 and the carbamate oxygen (around 2 - 4 Å) compared

to the distance between His199 and the oxygen atom O (5-8 Å) . Figure 7.7 shows a representative structure of the system, which was taken from a MD simulation and subsequently optimized (QM/MM with B3LYP/aug-cc-pVDZ as QM method).

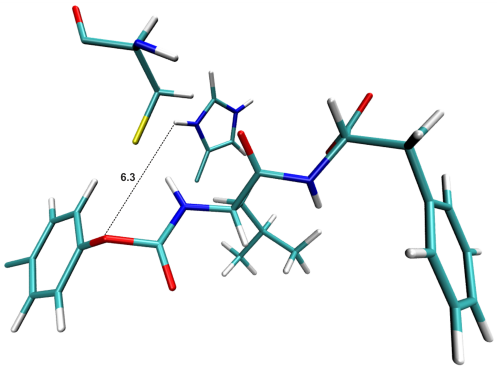


Abbildung 7.7: Structure of the inhibitor inside the active site of cathepsin B. The distance is given in Å.

After these preliminary investigations, the remaining feasible mechanism describes an elimination reaction according to an E1cB mechanism (**A** → **B2** → **C1**) with a proton transfer from cysteine to phenolate **(C1** → **C11**) and a subsequent carbamylation of cysteine (**C11** → **E2** → **F1**)(see emphasized scheme in Figure 7.8).
As this pathway fulfills all criteria and is therefore the only one the be reasonable, it alone will be investigated further. Therefore, the model system was enlarged to the whole enzyme including a docked inhibitor structure.

E1cB - Reaction Path

For the examination of the reaction with the real system, the reaction was parted into the two steps as shown in Figure 7.17. Starting with the elimination reaction, a potential energy surface was calculated on QM/MM level (PM3 [132,134] as QM method). The geometries of the resulting reaction path were used for single point calculations on a higher level of theory (B3LYP/aug-cc-pVDZ) [122,176] for the QM part. The energy barrier calculated using PM3 was 40.8 kcal/mol, whereas B3LYP predicted a barrier of only 22.0 kcal/mol. Additionally, the energetic course of the path calculated with B3LYP had abrupt changes in the energetic course of 5-8 kcal/mol between very similar structures. This indicates an insufficient description of the system by a pure treatment of the QM area with PM3. As a consequence, PM3 was henceforth only used for pre-optimizations and for qualitative estimations.

Step One: E1cB Elimination - Stabilization of Phenolate

Before further calculations concerning the mechanism were performed, the stabilization of the phenolate anion formed during the dissociation of the inhibitor was examined. The aim of these investigations was to find a path with a lower reaction barrier by stabilizing phenolate either through non-covalent interactions or protonation. A MD simulation with a frozen QM area containing the dissociated inhibitor (represented by its charge-density) was performed in order to see if any stabilizing interactions are formed with the negative charged phenolate. The enzymatic environment did not undergo significant changes, whereas a small water shell surrounded the anion. One water molecule was observed to diffuse into the active side. However, a quantification of the stabilizing effect was not possible, as the structural changes of the environment around the active site dominated the energy of the compared systems. A fixation of the surrounding amino acids lead to a destabilization of the structure with the water inside the active site by 10.8 kcal/mol, as the adjacent fixed molecules could not adapt to the water molecule. The promising structure with the water molecule inside the active site is examined further in a later part of this work (see "E1cB - Water-Mediated Proton Transfer" or Chapter 7.3.4).

Next, the protonation of the formed phenolate through Cys29 was investigated using DFT methods (no hybrid methods were applied). The deprotonation of the inhibitor was assumed to be the first step of the reaction with the subsequent dissociation and protonation of phenolate. The deprotonation was computed using a reaction profile (B3LYP/aug-cc-pVDZ with COSMO; ϵ=78.93). The dissociation of the inhibitor and protonation of phenolate were investigated with the help of a potential energy surface (B3LYP/aug-cc-pVDZ with COSMO; ϵ=78.93) in order to cover the possibility of a concerted mechanism. The resulting pathway has an energy barrier of 28.8 kcal/mol. In an attempt to investigate if a reaction path with a lower barrier exists, the possibility of the three steps (deprotonation, dissociation and protonation of phenolate) occurring simultaneously was investigated *via* multiple potential energy surfaces (B3LYP/aug-cc-pVDZ with COSMO; ϵ=78.93). However, the resulting reaction barrier still remained at 32.9 kcal/mol.

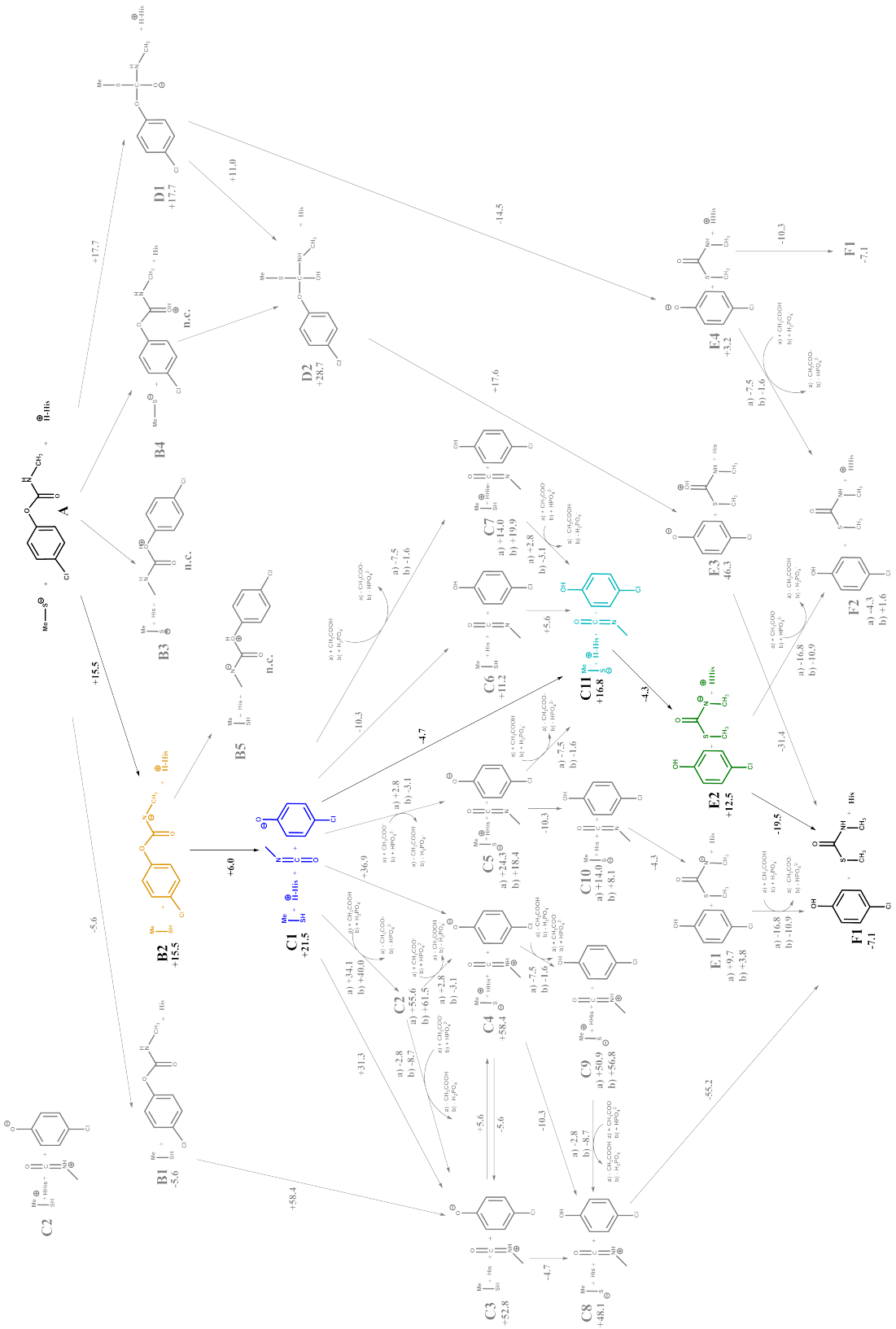


Abbildung 7.8: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; $\epsilon = 78.93$). Energies are given in kcal/mol; n.c.: not converged. The examined mechanism is indicated in bold.

Step One: E1cB Elimination - Nudged Elastic Band

The gained results indicate that the reaction depends on a number of variables and can therefore only be poorly described by reaction profiles and potential energy surfaces. As a solution, Nudged Elastic Band (NEB) calculations on QM/MM level (B3LYP/cc-pVDZ as QM method) were performed in order to find a reaction path between the structures. Two computations were performed, describing an E1cB elimination with and without protonation of the formed phenolate anion.

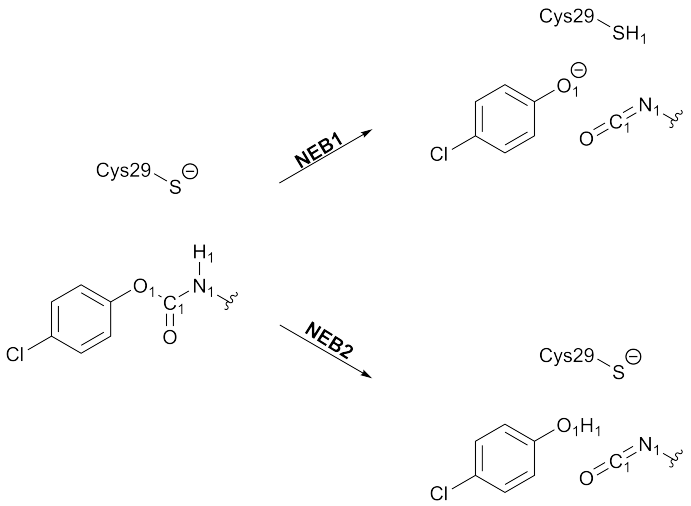


Abbildung 7.9: Reactant and product structures of the performed NEB calculations.

For the deprotonated phenolate (NEB1 in Figure 7.9), the reaction barrier as well as the product structure had an energy of 40.1 kcal/mol (see Figure 7.10a). This strongly indicates that a reaction is only possible if the alcoholate ion is protonated e.g. through cysteine. In the reaction pathway resulting from NEB2 (see Figure 7.9 and 7.10b), the product structure was energetically decreased compared to the highest point of the barrier. However, the predicted activation energy was still approximately 40 kcal/mol, which indicates that the reaction cannot take place.

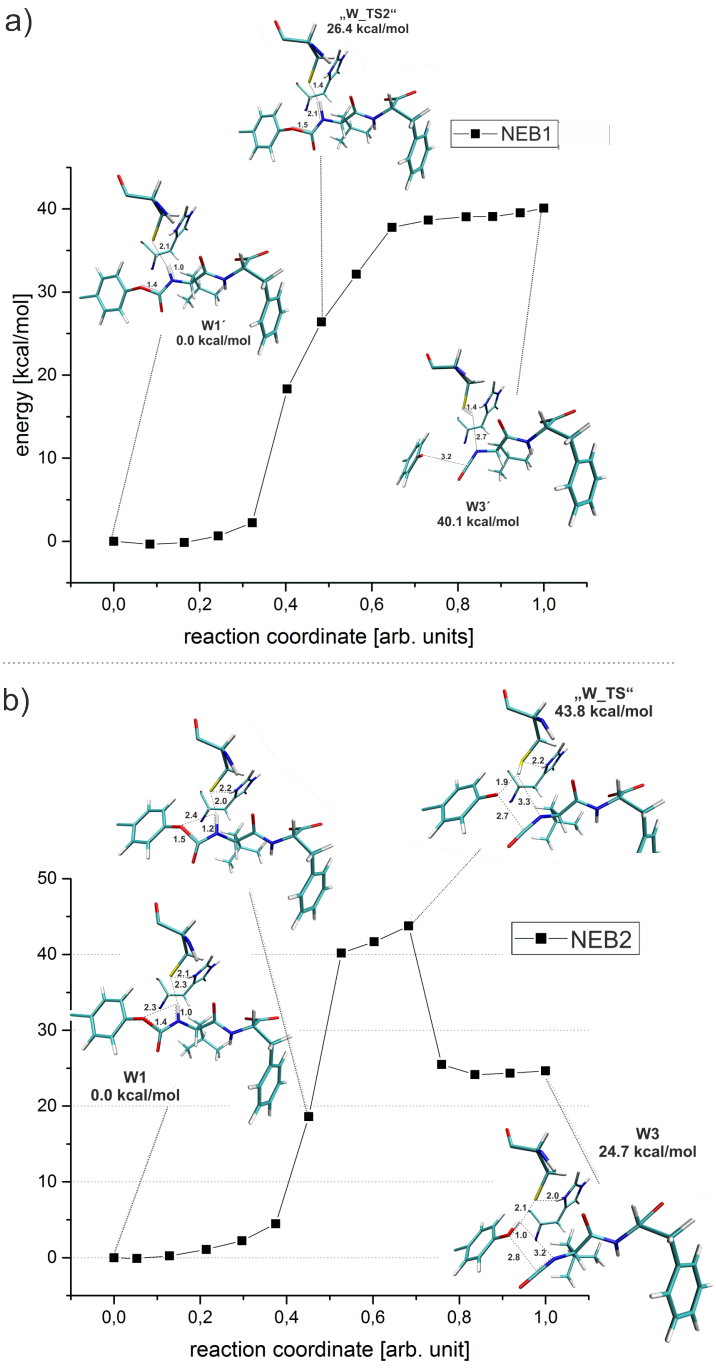


Abbildung 7.10: a) NEB pathway on QM/MM level (B3LYP/cc-pVDZ as QM method) of NEB1. b) NEB pathway on QM/MM level (B3LYP/cc-pVDZ as QM method) of NEB2.

Step One: E1cB Elimination - Reaction Path Water-Mediated Proton Transfer

Based on a recent publication on water-catalyzed proton transfer reactions in enzymes of the group of Prof. Engels in Würzburg, a water molecule was included in the system in order to simulate a water-bridged proton transfer reaction (see Figure 7.11).[177,178]

According to a MD simulation performed with the dissociated inhibitor structure, a water molecule can diffuse inside the active site of the enzyme (see "E1cB - Stabilization of Phenolate" or Chapter 7.3.2).

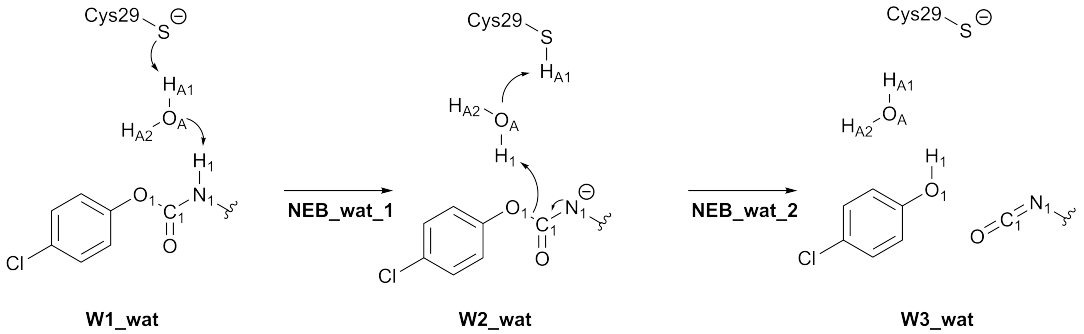


Abbildung 7.11: Reaction scheme of water-mediated E1cB reaction.

Therefore, a system with an additional water molecule was manually prepared for QM/MM calculations and optimized using B3LYP/cc-pVDZ as the QM method. As the number of relevant system variables increases by the inclusion of the water molecule, several PES (PM3 as QM method) were performed in order to gain qualitative insights. The most promising surfaces were combined in a reaction profile with subsequent single point calculations using QM/MM hybrid schemes with the QM methods B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ. The reaction barrier was significantly lower compared to the reaction without an additional water molecule (22.2 kcal/mol for SCS-MP2/aug-cc-pDVZ; Figure 7.12a). The structure **W12_TS'** is obtained from the calculation of a PES describing the deprotonation and dissociation of the inhibitor. It holds an isolated oxonium ion. This structure, however, does not appear in the NEB reaction path. In a final step, NEB calculations on the QM/MM level of theory (B3LYP/cc-pVDZ as QM method) were performed to examine whether a reaction path with an even lower barrier can be found. As the NEB calculation from structure **W1_wat** to **W3_wat** did not converge, the two NEB profiles from**W1_wat** to **W2_wat** and from **W2_wat** to **W3_wat** were combined in one reaction coordinate (see Figure 7.12b). The x-axis therefore contains equidistant marks, one for each frame of the calculation. The reaction barrier of this path is therefore the lowest one for the reaction (15.3 kcal/mol) with a reaction enthalpy of -8.1 kcal/mol. Based on these results, it can be concluded that an E1cB elimination including a catalytic water molecule can take place.

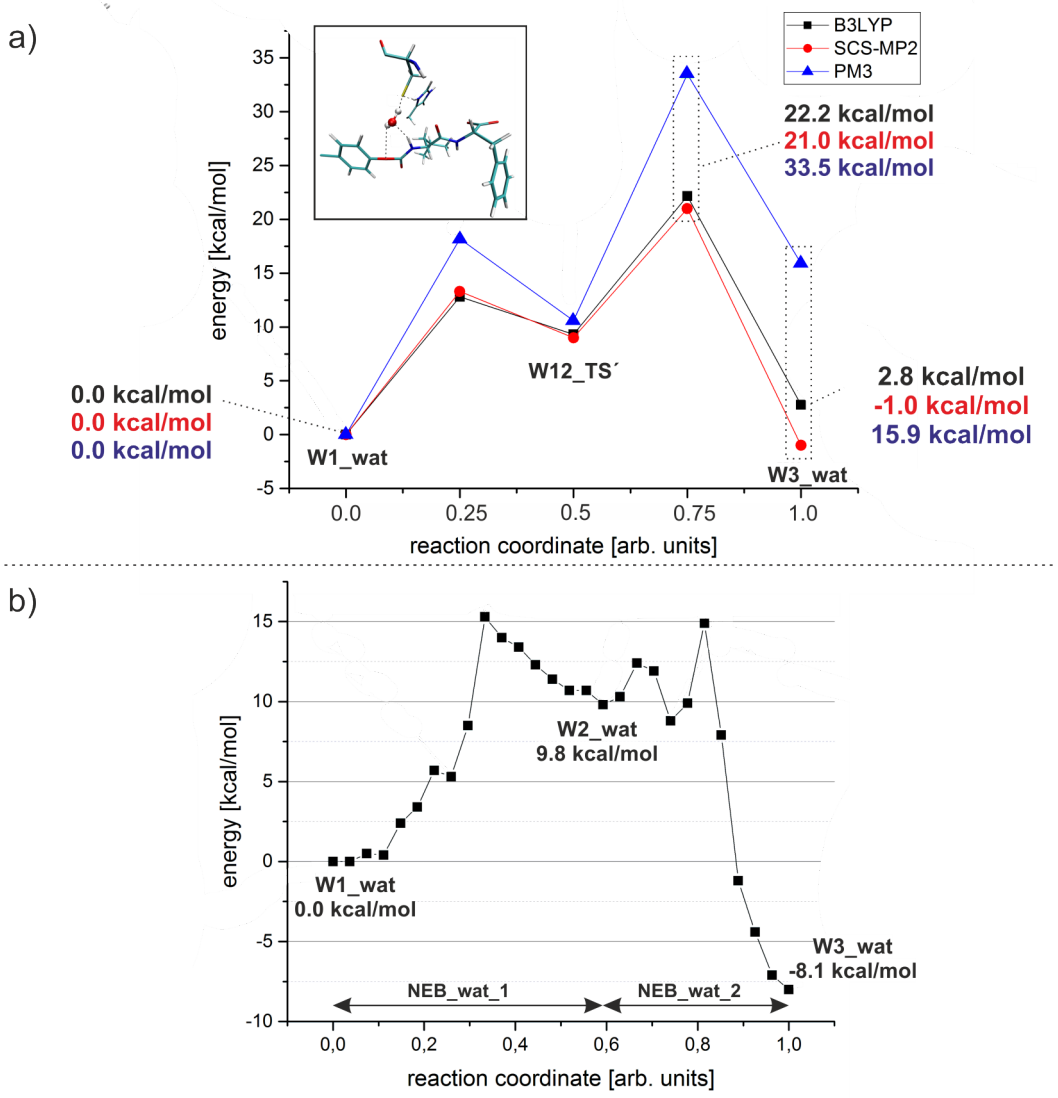


Abbildung 7.12: a) Reaction pathway resulting from the combination of several PES (PM3 as QM method) with single point calculations (B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ as QM methods). **W12_TS'** is the structure of the deprotonated inhibitor containing an isolated oxonium ion.
b) NEB reaction path (B3LYP/cc-pVDZ as QM method) of the reaction from **W1_wat** to **W2_wat** and from **W2_wat** to **W3_wat**. NEB profiles from **W1_wat** to **W2_wat** and from **W2_wat** to **W3_wat** were combined in one reaction coordinate, the x-axis therefore contains equidistant marks, one for each frame of the calculation. More detailed information about the respective structures is given in Figure 7.47 (see Chapter 7.3.4).

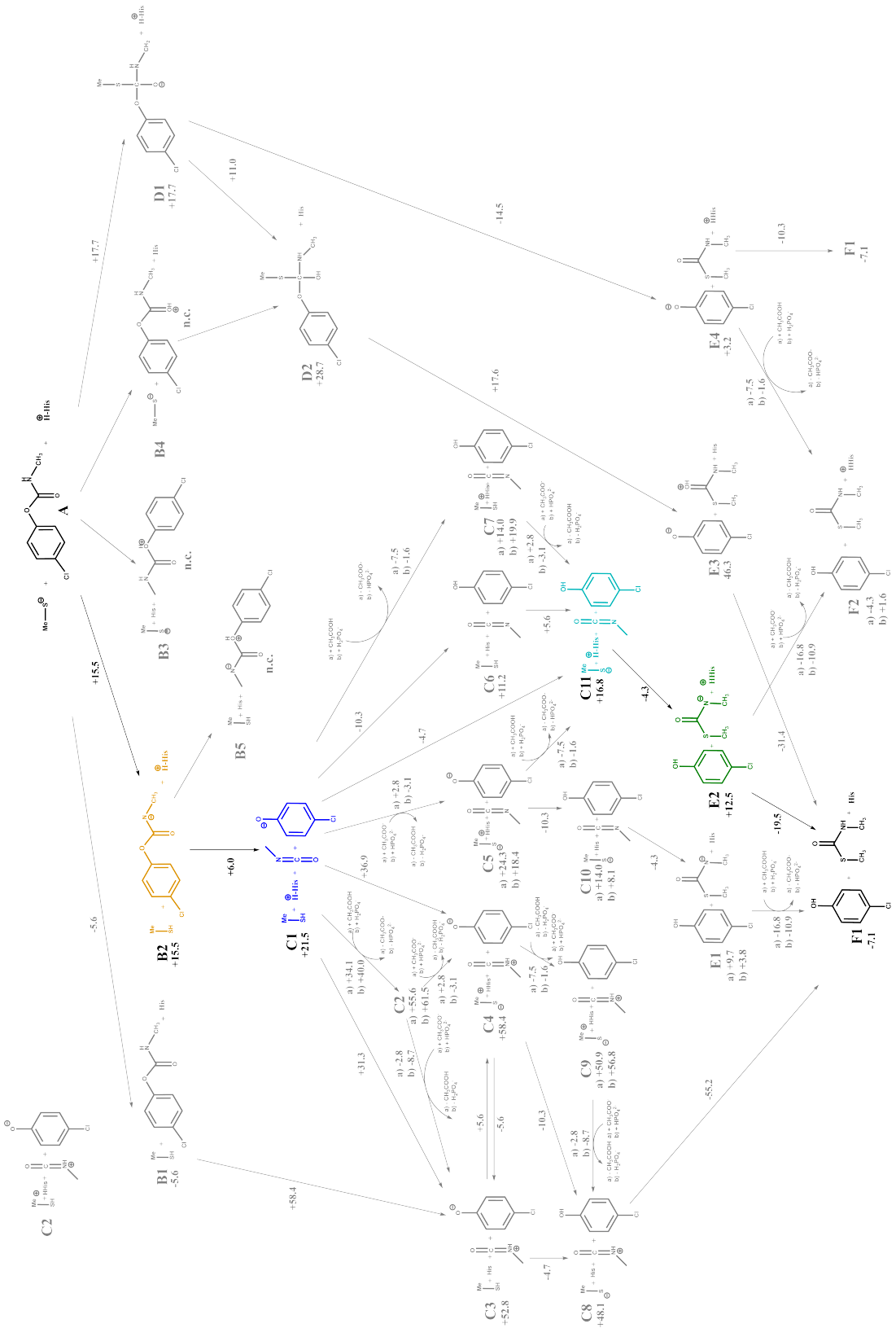


Abbildung 7.13: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; ϵ = 78.93). Energies are given in kcal/mol; n.c.: not converged. The examined mechanism is indicated in bold.

Step Two: Carbamylation - QM Calculations on a Model System
The subsequent reaction step of the inhibition (carbamylation) was investigated based on the aforementioned gained results. It includes a nucleophilic attack of Cys29 at the isocyanate carbon and a protonation at the nitrogen N_1 trough His199 (see Figure 7.14). In a first attempt, the small model system described above with an additional acetate buffer was used for DFT calculations. The calculations showed that an "*anti*" orientation of the deprotonated nucleophilic sulfur and the protonating H-atom of the buffer relative to the C-N double bond results in a barrier-free reaction.

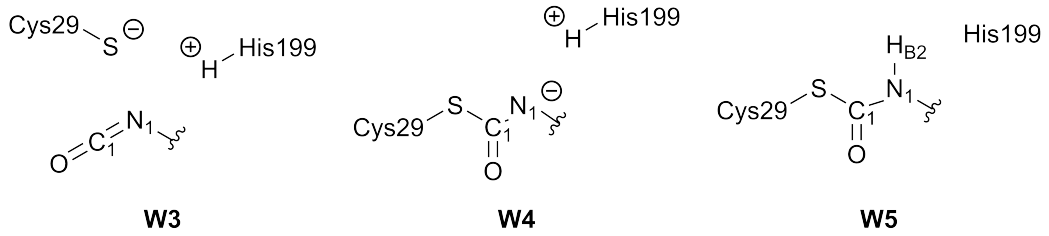


Abbildung 7.14: Lewis structures of **W3**, **W4** and **W5**.

Step Two: Carbamylation - Orientation of the Isocyanate
For further investigations using QM/MM methodologies, the acetate buffer and phenol were deleted from structure **W3_wat** (see Figure 7.11a). The remaining inhibitor (isocyanate) was in an unfavourable position for a reaction with the active site amino acids Cys29 and His199. In order to examine whether a reorientation of the warhead is possible, a MD simulation was performed. Throughout the simulation, the inhibitor left the active site several times but always diffused back inside in a different orientation. Therefore a rearrangement of the inhibitor does therefore not seem to be problematic.

Step Two: Carbamylation - Generation of a Suitable Starting Structure
In order to gain suitable structures for NEB calculations, several PES (PM3 as QM method) were calcua with the starting structure in order to gain a more favourable position of the inhibitor near Cys29. The resulting structure was used for the generation of **W4** and **W5** (see Figure 7.14) *via* several PES (PM3 as QM method). None of the three structures, **W3**, **W4** or **W5**, reached convergence within QM/MM geometry optimizations on a higher level of theory (B3LYP/cc-pVDZ as QM method). A closer examination of the adjacent amino acids and water shell lead to the need for additional MD simulations with the structures **W3** and **W5** in order to obtain geometries including a relaxed environment which can stabilize the systems. The

simulation of **W3** did not lead to further results. For **W5**, a water molecule diffused in a favourable position for a water-bridged proton transfer. Based on this structure, the reaction was calculated backwards *via* PES (PM3 as QM method) in order to gain converged structures of **W4** and **W3**. The structures converged for unconstrained geometry optimized with B3LYP/cc-pVDZ as QM method.

Step Two: Carbamylation - Nudged Elastic Band Calculations

The converged structures of **W3**, **W4** and **W5** were taken for NEB calculations of the whole reaction (**W3** to **W5**) and the single reaction steps (**W3** to **W4** and **W4** to **W5**). Due to the amount of structural changes during the reaction, the NEB calculation from **W3** to **W5** did not converge, in contrast to the other two computations. The final reaction path, gained through the combination of the two calculations, is shown in Figure 7.15.

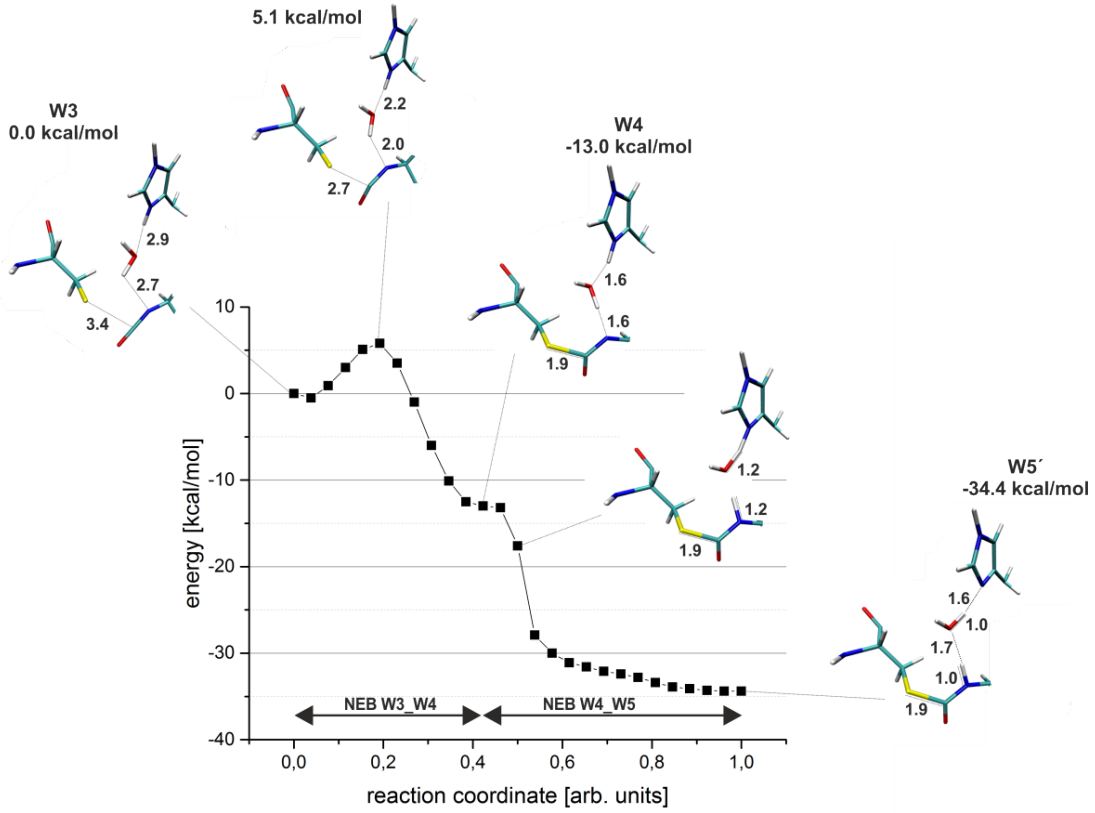


Abbildung 7.15: NEB path (B3LYP/cc-pVDZ as QM method) of the carbamylation of Cys29 through the isocyanate intermediate. NEB profiles from **W3** to **W4** and from **W4** to **W5** were combined in one reaction coordinate, the x-axis therefore contains equidistant marks, one for each frame of the calculation.

The x-axis contains equidistant marks, one for each frame of the calculation. The calculations predict a reaction barrier of 5.1 kcal/mol for the nucleophilic attack of sulfur and a nearly barrier-free water-bridged proton transfer between His199 and the inhibitor. The high exothermicity of the reaction could be traced back to the product **W5'**, a structure that emerged throughout the NEB calculation from **W4** to **W5** and entails strong hydrogen bonds between the reaction partners (see Figure 7.15).

Conclusions
Finally, the reaction mechanism of a carbamate-based inhibitor and cathepsin B was successfully investigated with help of MD simulations and different approaches on the QM and QM/MM level. The calculations showed that the water shell surrounding the active site has a significant impact on the mechanism. Several reaction steps would not take place without a catalytic water molecule partaking in the reaction. An energetic connection between the elimination reaction (first step) and the carbamylation (second step) could not be quantified, as the second step required a rearrangement of the participating reactants.

7.3 Step One: E1cB Elimination

There are several possibilities for the computational analysis of a mechanism, including reaction profiles, potential energy surfaces and NEB calculations. Within reaction profiles, the influence of one structural parameter on the reaction can be examined. Potential energy surfaces cover the mutual influence of two parameters while NEB calculations include the variation of an arbitrary number of parameters. The computational effort of these calculations rises with the number of varied parameters. The iterative scheme of NEB calculations hardly converge in difficult cases, depending on the amount of structural change during the mechanism. Thus, a combination of all three methods was utilized for the investigations in this work, depending on the respective task. As mentioned above, the investigated mechanism contains the two steps shown in Figure 7.17. Due to the complexity of the reaction path, the steps were investigated separately.

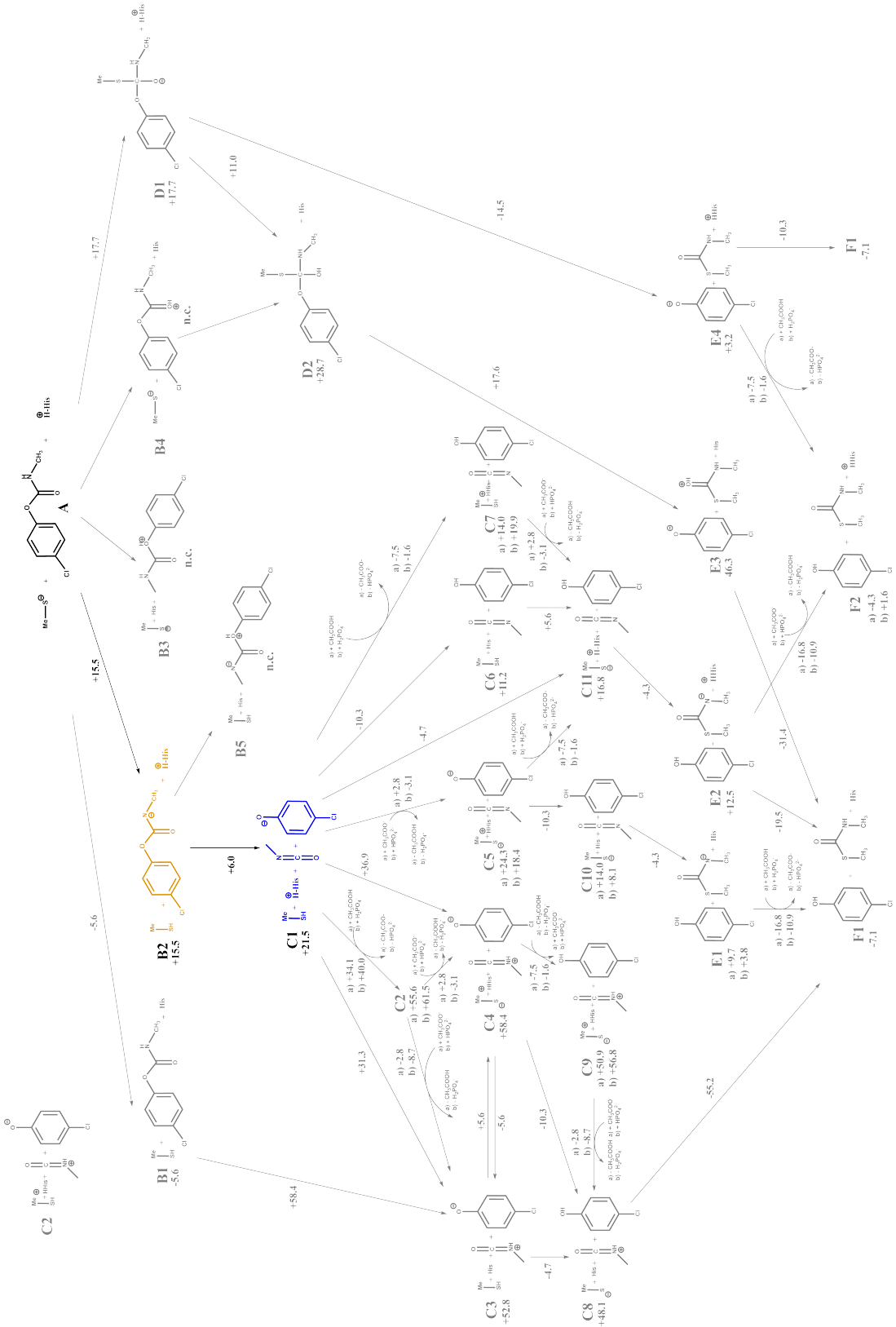


Abbildung 7.16: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; ϵ = 78.93). Energies are given in kcal/mol; n.c.: not converged. The examined mechanism is indicated in bold.

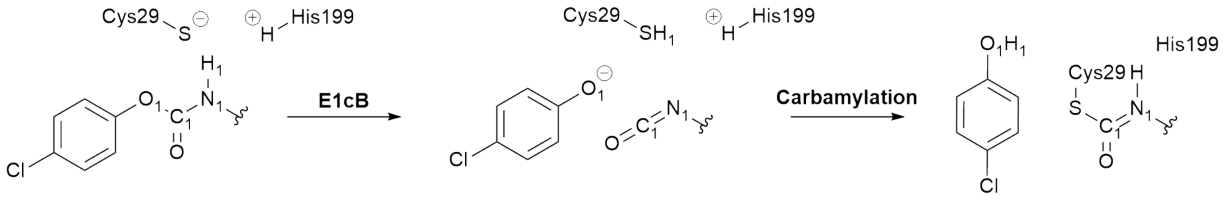


Abbildung 7.17: Proposed mechanism for the reaction of carbamate-based inhibitors in cathepsin B.

7.3.1 Reaction Path

In a first attempt, the initiating E1cB elimination is computed with the help of a potential energy surface, using the QM/MM method. The reaction corresponds to the path **A** → **B2** → **C1** of Figure 7.8 and 7.16. As QM method, the semi empirical PM3 approach is used. The reaction coordinates are defined as the distance between S and H_1 in order to describe the deprotonation and between C_1 and O_1 to simulate the dissociation of the inhibitor (see Figure 7.18).

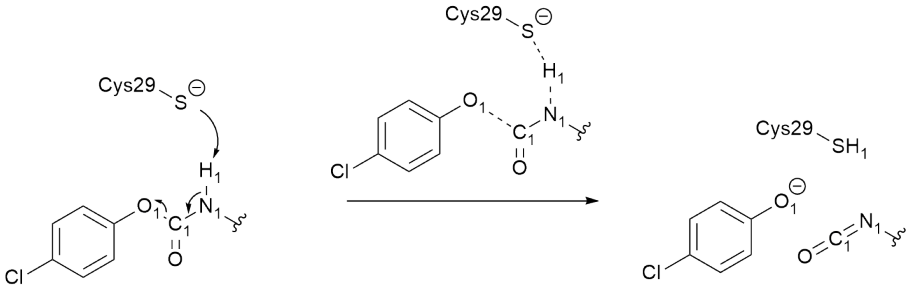


Abbildung 7.18: Reaction scheme of an E1cB elimination of the inhibitor inside the active site of cathepsin B.

The surface indicates, that the deprotonation of the inhibitor is the initiating step of the mechanism with the subsequent dissociation corresponding to an E1cB elimination reaction. The barrier of the resulting reaction path (grey line on surface in Figure 7.19) is high (40.8 kcal/mol), as is the reaction enthalpy of 34.7 kcal/mol. As semi empirical methods are mostly suitable for the qualitative assessment of a reaction, single point calculations on a higher level of theory for the QM area (B3LYP/aug-cc-pVDZ) were performed on the geometries of the reaction path, extracted from the surface (see Figure 7.20). Through the utilization of the diffuse basis set, the properties of the sulfur anion were supposed to be described more properly. The reaction barrier

decreases by 10.4 kcal/mol, compared to the semi empirical results. Moreover, there are abrupt changes in the energy (up to 8.0 kcal/mol), that can not be explained by the structurally small deviation of the respecting molecules. The results indicate, that the semi empirical PM3 method neglects fundamental effects of the reaction.

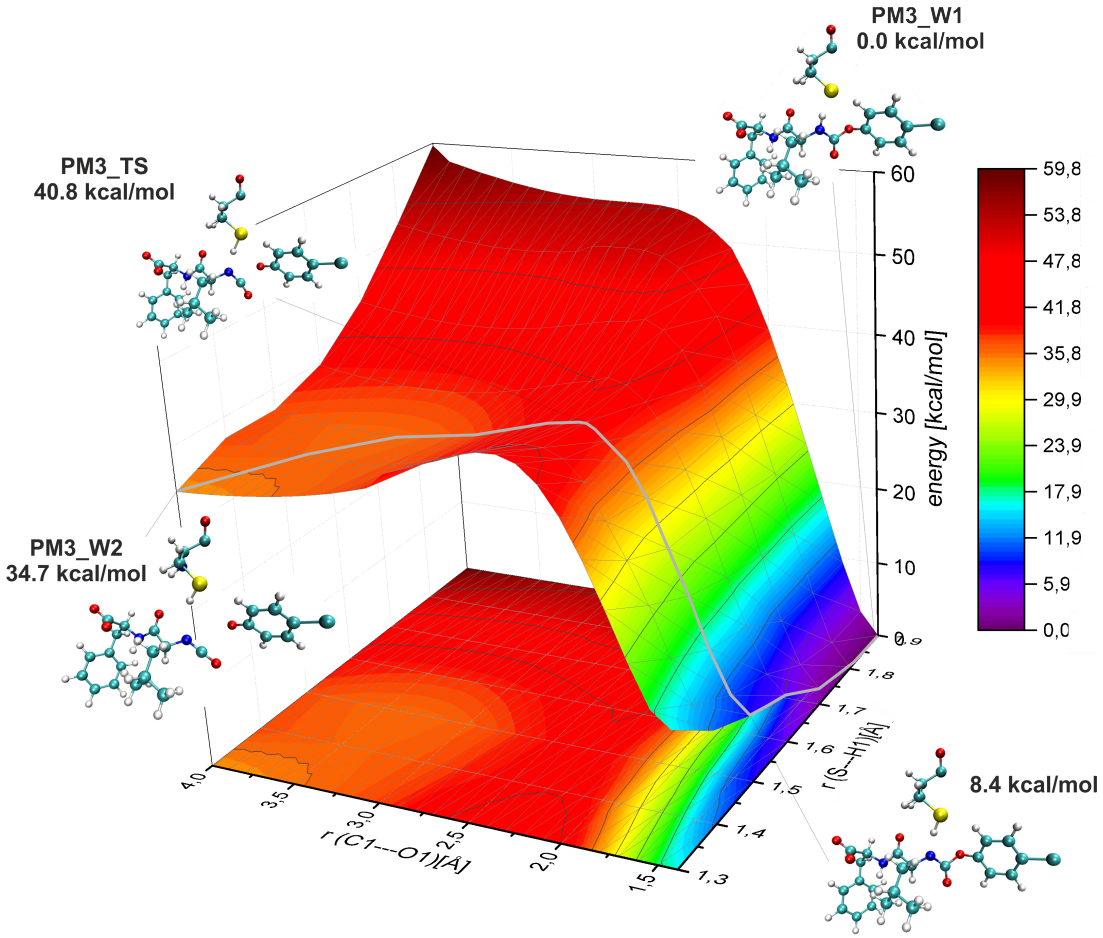


Abbildung 7.19: Potential energy surface calculated with the QM/MM method (PM3 as QM method) of the deprotonation and dissociation of the inhibitor. Structures **PM3_W1** and **PM3_W2** correspond to **A** and **C1** in Figure 7.16.

Several consequences are drawn from the results. Concerning the computational methods, PM3 can only be used for pre-optimizations. In order to save calculation time, all further QM/MM calculations will be pre-optimized with PM3 and subsequently re-optimized with B3LYP/cc-pVDZ as QM method. The choice of a non-diffuse basis set was based on the fact that geometry optimizations, including diffuse basis functions, are hard to converge. The augmented basis was therefore included in subsequent single point calculations. As the mechanism seems to be rather complex,

several following calculations were performed on a QM level of theory within a small model system in order to gain detailed insights. Based on the obtained conclusions from the model system, further QM/MM calculations with the original system could subsequently be performed more efficiently. However, before further calculations on the reaction mechanism were performed, the focus was directed to the phenolate ion that is formed by dissociation of the inhibitor.

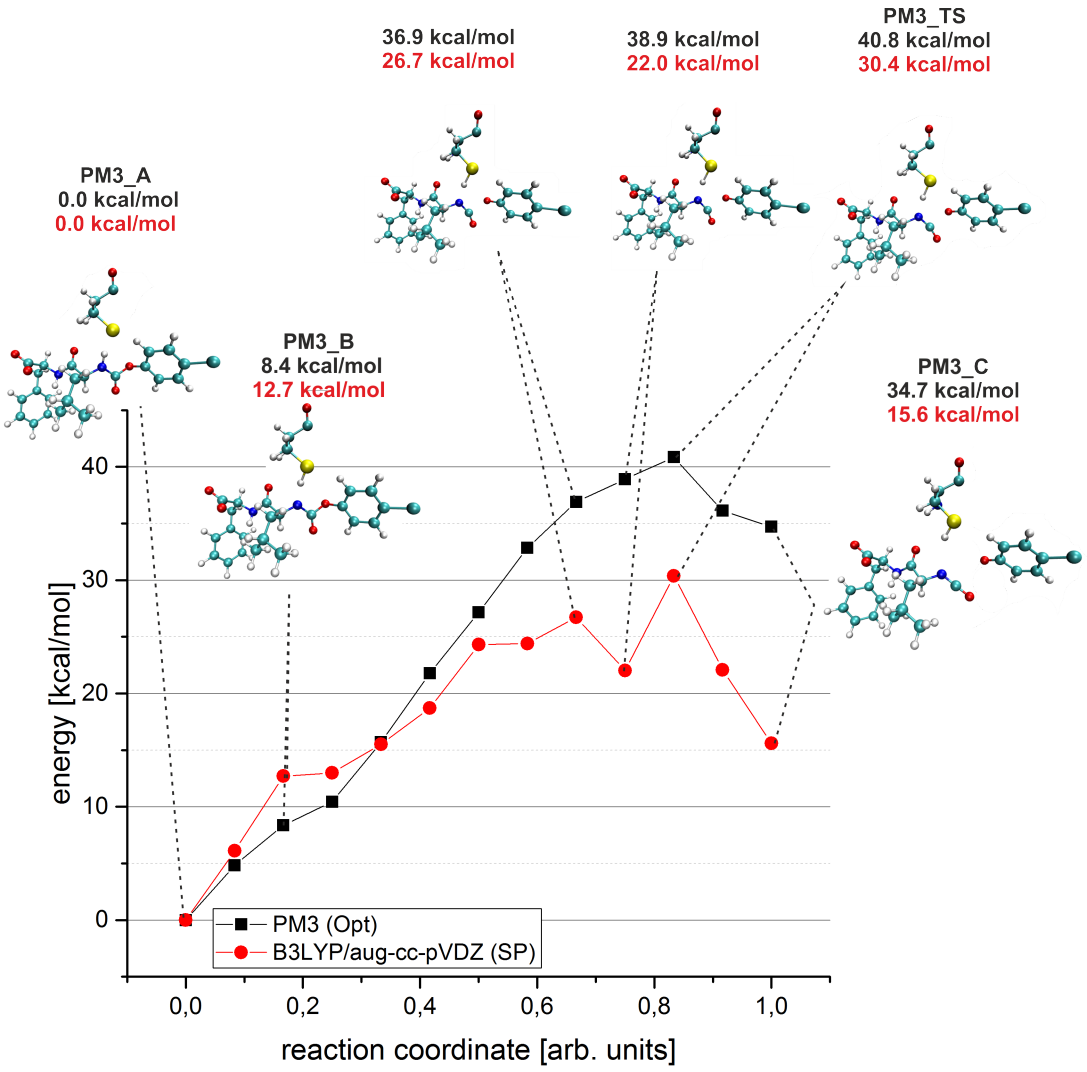


Abbildung 7.20: Energy profile of the deptrotonation and dissociation of the inhibitor. Levels of theory are single point QM/MM calculations (B3LYP/aug-cc-pVDZ as QM method; red) on optimized structures (PM3 as QM method; black).

7.3.2 Stabilization of Phenolate

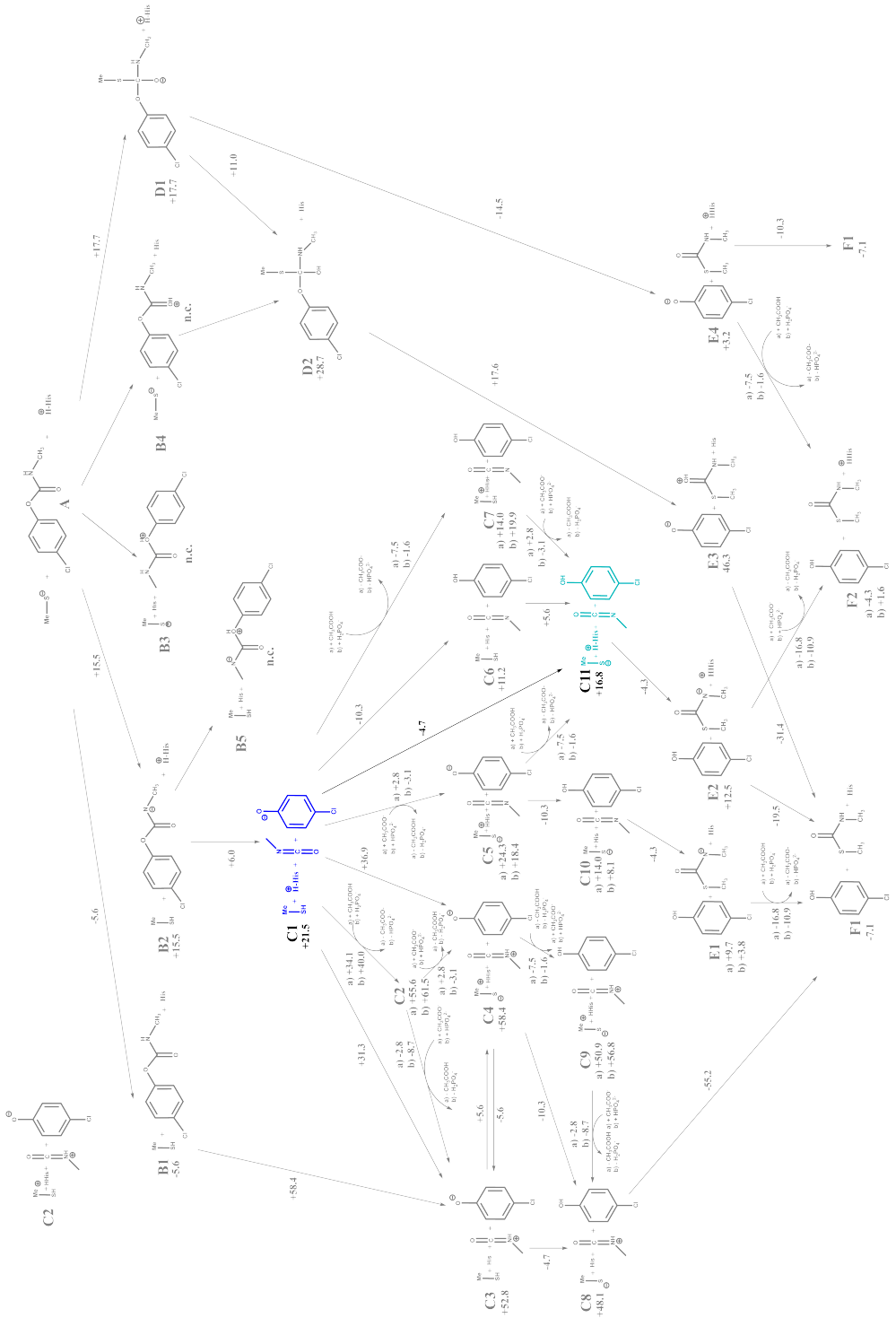


Abbildung 7.21: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; ϵ = 78.93). Energies are given in kcal/mol; n.c.: not converged. The examined mechanism is indicated in bold.

Based on the idea that a stabilization of the phenolate anion, formed after the dissociation (**C1** in Figure 7.21) could decrease the reaction barrier different realizations thereof were analyzed. Potential non-covalent interactions of the oxygen atom with either adjacent amino acids or water molecules were investigated. For this purpose, MD simulations with a structure of the dissociated inhibitor (**PM3_TS** in Figure 7.19) were performed. In a second step, an actual protonation of the alcoholate ion through cysteine was included in the reaction mechanism. As the inclusion of a third coordinate requires additional computational effort, the system was reduced to a model system, containing the warhead of the inhibitor and a methanethiol as a cysteine representative.

Non-Covalent Interactions

The environment around the phenolate anion was investigated in order to determine potential hydrogen bonding partners. Concerning adjacent amino acids, all relevant compounds inside a radius of 5 Å around the oxygen atom are shown in Figure 7.22.

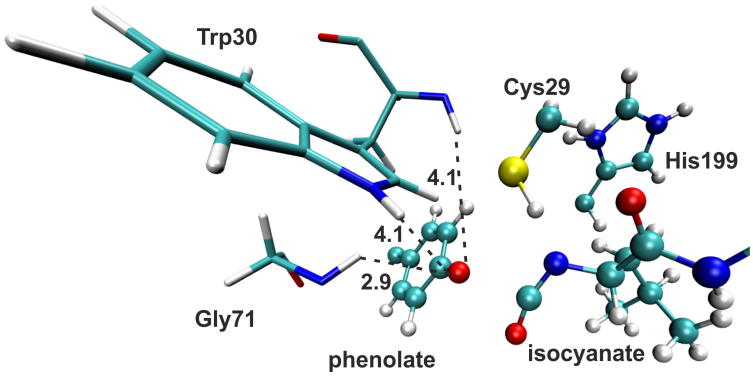


Abbildung 7.22: Dissociated inhibitor (ball and stick) with all adjacent amino acids in a radius of 5 Å around the negatively charged oxygen atom.

The structure taken for the investigations originates from the potential energy surface (PM3_TS Figure 7.19). It was subsequently optimized with B3LYP/cc-pVDZ as QM method. The smallest distance of a potential proton donor to the phenolate oxygen amounts to 3 Å and emerges with the backbone amino group of Gly71. Even though the position of the amino acids were altered during a MD simulation, their orientation relative to the inhibitor anion is not favourable for the formation of stabilizing

interactions. Additionally, the phenolate will presumably diffuse out of the active site, which makes a permanent stabilization of the phenolate by the amino acids unlikely. Similar to the adjacent amino acids, there are no water molecules near enough for a stabilization of the anion in the observed structure. As the small water molecules are more flexible in contrast to the amino acids, the possibility of a solvation of the anion was investigated with help of a MD simulation. Figure 7.23 shows the dissociated inhibitor inside the active site of the enzyme. The blue area marks a region with an increased number of water molecules.

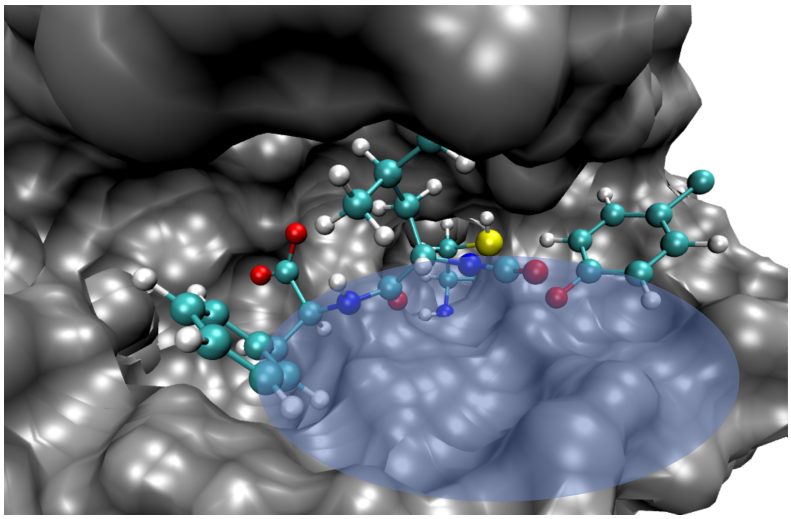


Abbildung 7.23: Dissociated inhibitor inside the active site of cathepsin B. The water-rich region is marked in blue.

In order to investigate the movement of the environment relative to a fixed inhibitor position, the following approach was applied: The QM area (containing Cys29 and the inhibitor) is described through an electrostatic potential (ESP), calculated with the semi empirical method AM1. As the calculation corresponds to a single structure, the QM area remains frozen throughout the MD simulation. The water molecules and enzymatic environment in contrast can freely move around the inhibitor. A simulation over a time of 2 ns was performed with the purpose of obtaining a relaxation of the environment around the dissociated inhibitor. The amino acid side chains and backbones remain in a stable position (with the exception of Trp30), whereas the water molecules migrate towards the phenolate anion. Different water molecules repeatedly stabilize the phenolate over time periods of at least 50 ps. Figure 7.24 shows the distance r between several water molecules (defined by the respectively closer

hydrogen atom) and the anion oxygen. One water molecule (Wat540 in Figure 7.25) even diffuses inside the active site and stays near the negative charge of the alcoholate, while Trp30 provides additional space by rotation of the side chain.

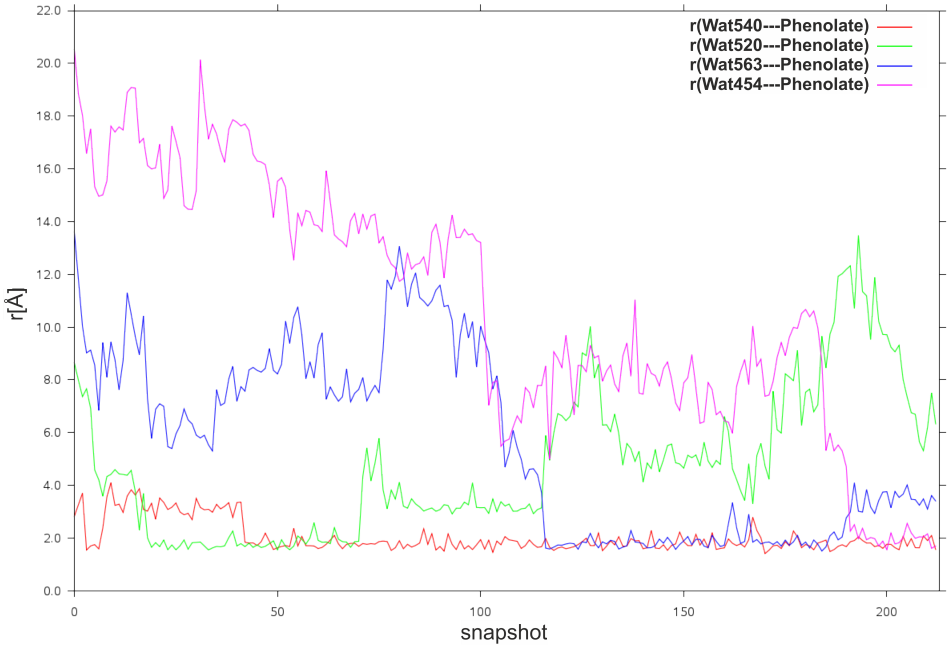


Abbildung 7.24: Distances between several water molecules (with regard to the closer hydrogen atom) and the phenolate oxygen. All distances are given in Å and the simulation time amounts to 213 ps in total.

The calculations show that a stabilization of the phenolate anion can be provided through non-covalent interactions with surrounding water molecules. The quantification of the stabilization is however a challenging task, as even small changes in a number of environmental compounds (water molecules and amino acids) can lead to an extreme difference in the energy of the system. In an attempt to do so, the left structure shown in Figure 7.25 was taken as a reference system. The water molecule Wat540 was positioned inside the active site, similar to the right structure in Figure 7.25. For both systems (Wat540 inside and Wat540 outside the active site), single point calculations were performed using B3LYP/aug-cc-pVDZ as QM method.

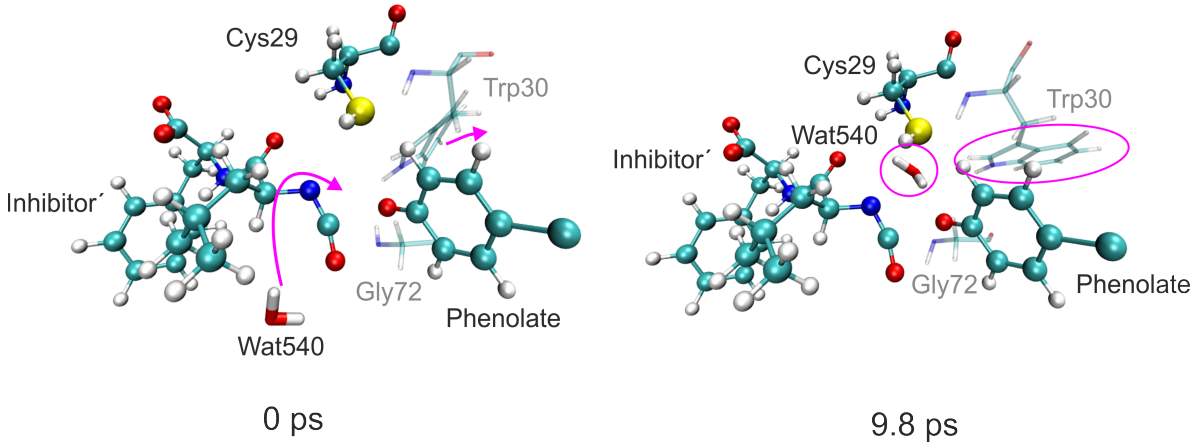


Abbildung 7.25: Snapshot of the MD simulation after 0.0 (left) and 9.8 (right) ps for the simulation of a relaxing environment with a frozen QM area (inhibitor after the dissociation).

The calculations show a destabilization of 10.8 kcal/mol for the system in which the water molecule sits inside the active site. However, this result has to be interpreted with caution, as the water was positioned manually and the structures were not optimized to avoid a big number of small structural changes in the environment. In conclusion, it can be stated that the negative charge of the phenolate molecule can be stabilized through adjacent water molecules. The promising structure with an additional water molecule inside the active site will be further studied in a later Chapter of this work (see Chapter 7.3.4).

Protonation of Phenolate on a Model System

As the proposed mechanism involves an actual protonation of the phenolate, the reaction is now extended to an additional step. Therefore, the system was reduced to a smaller model, containing the inhibitor warhead and a methanethiol as a representative for the active site amino acid Cys29 (see Figure 7.27). The reduced size allows for QM calculations at the B3LYP/aug-cc-pVDZ level of theory in cooperation with the implicit solvation model COSMO (ϵ=78.93). Initially, a potential energy surface, describing the dissociation of the inhibitor and protonation of phenolate through methanethiol, was calculated.

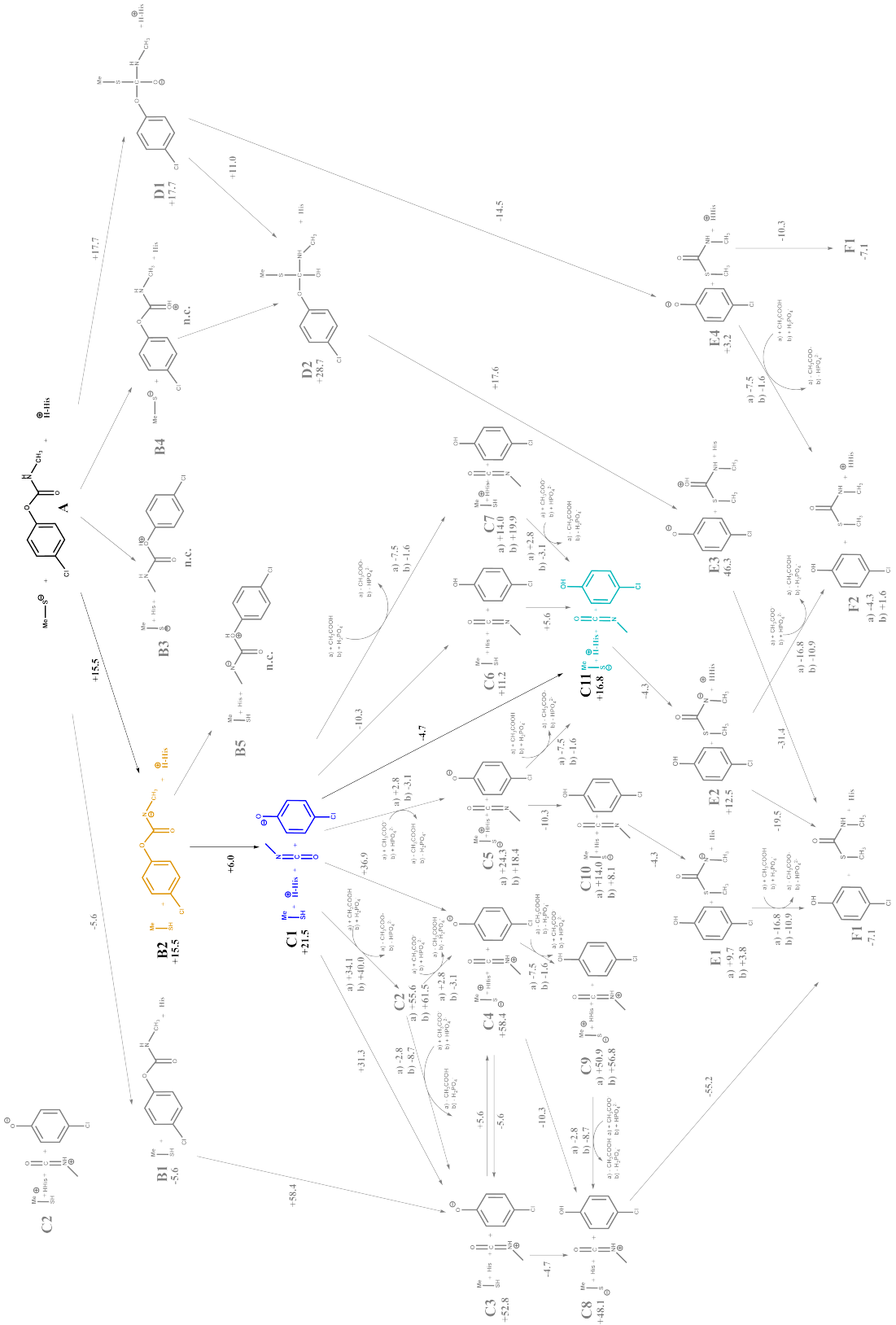


Abbildung 7.26: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; ϵ = 78.93). Energies are given in kcal/mol; n.c.: not converged. The examined mechanism is indicated in bold.

The starting point of the computations was a manually built structure of the deprotonated inhibitor (see Figure 7.27). In a subsequent calculation, the deprotonation of the inhibitor was calculated backwards, using the same starting structure.

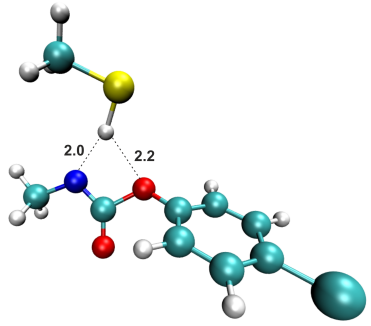


Abbildung 7.27: Structure of the deprotonated inhibitor and methanethiol, the starting point for the calculations of the PES. The structures **INT_1** and **INT_2** correspond to **B2** and **C11** in Figure 7.26. Distances are given in Å.

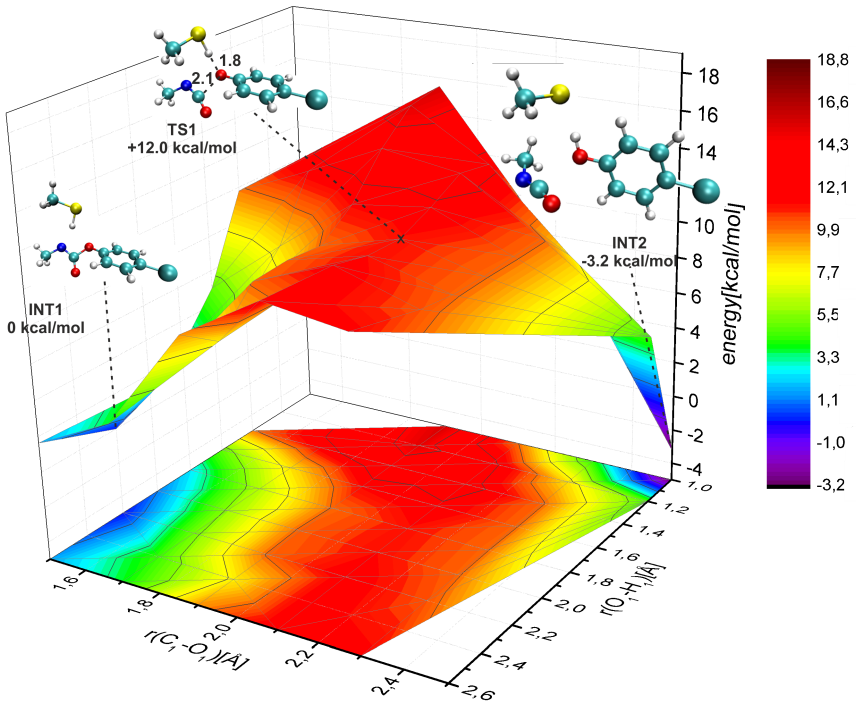


Abbildung 7.28: PES (B3LYP/aug-cc-pVDZ with COSMO) describing the dissociation of the inhibitor and the protonation osf phenolate through cysteine.

The PES shows a significantly lower reaction barrier for the dissociation of the inhibitor (12.0 kcal/mol) and indicates a concerted protonation of the alcoholate anion. As the

starting structure of the surface (deprotonated inhibitor; **INT1** in Figure 7.27 ≡ **B2** in Figure 7.26) was originally not the start of the reaction, its energy, relative to the protonated and undissociated inhibitor, is calculated with the help of the reaction profile mentioned above.

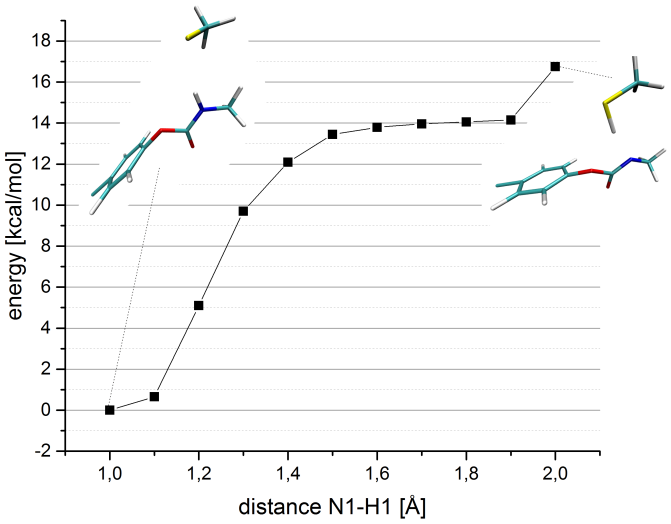


Abbildung 7.29: Energetic profile (B3LYP/aug-cc-pVDZ with COSMO) describing the deprotonation of the inhibitor.

The profile depicted in Figure 7.29 shows that the deprotonated form of the inhibitor is not a local minimum of the reaction mechanism, which is in agreement with the potential energy surface of the deprotonation and dissociation of the inhibitor (see Figure 7.19). The reaction barrier of 16.8 kcal/mol increases the barrier of the complete reaction to 28.8 kcal/mol, which is again unfavourable. The energetic difference of nearly 3 kcal/mol between the penultimate and last structure can be attributed to the change of the frozen reaction coordinates during the optimizations.

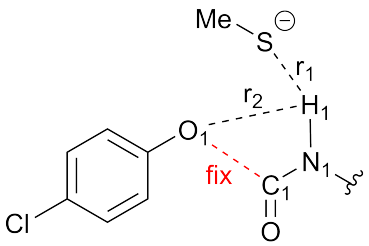


Abbildung 7.30: Reaction coordinates for the investigation of the concerted mechanism.

To this point, the deprotonation of the inhibitor was presumed to be the first step of the mechanism. In the following, a completely concerted reaction course, containing the deprotonation of the inhibitor (**A** → **B2** in Figure 7.26), its dissociation (**B2** → **C1** in Figure 7.26) and the protonation of the phenolate (**C1** → **C11** in Figure 7.26) was examined. For this purpose, the three reaction coordinates shown in Figure 7.30 were introduced.

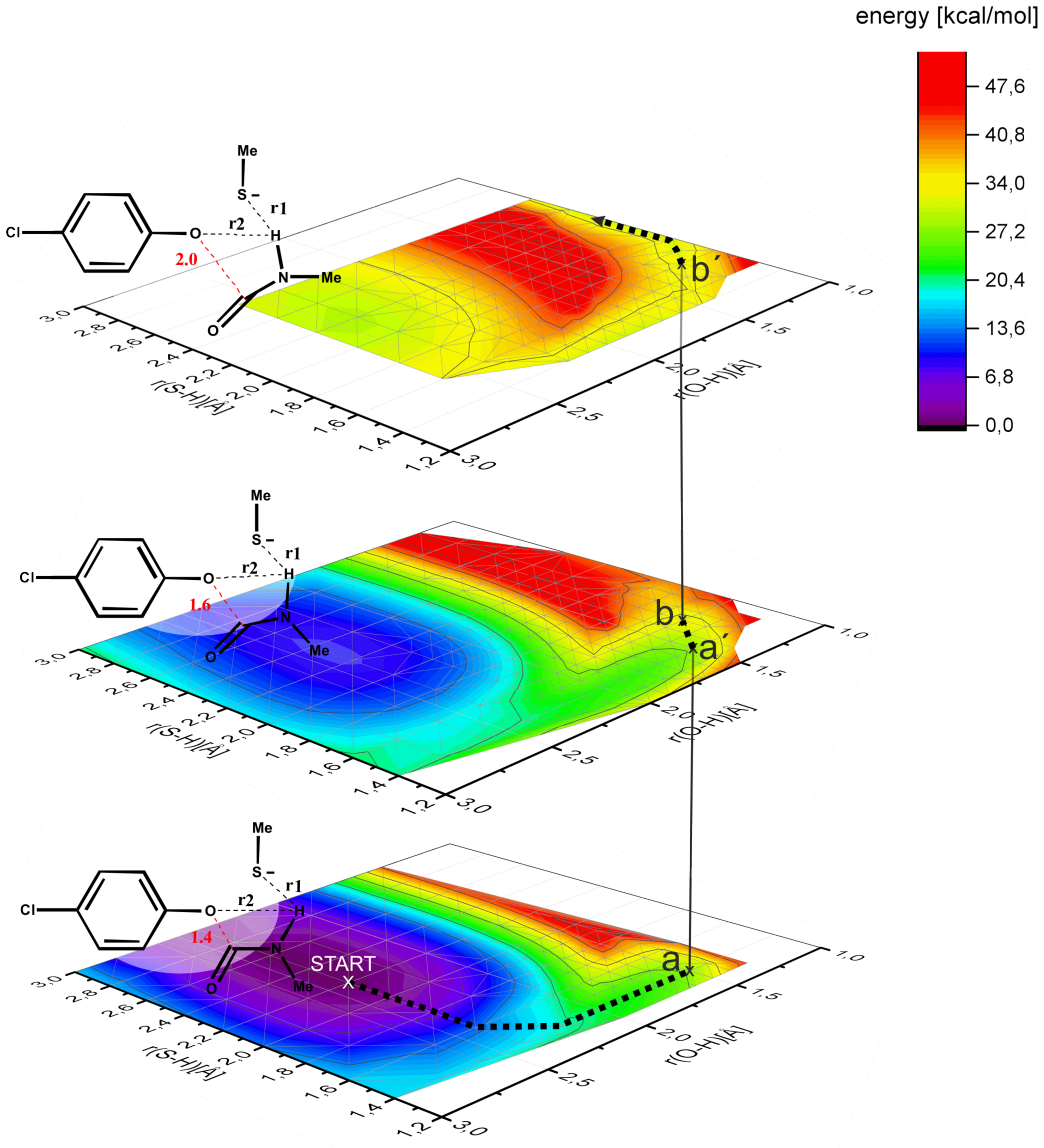


Abbildung 7.31: Potential energy surfaces (B3LYP/aug-cc-pVDZ) scanning three reaction coordinates by freezing one coordinate per PES and varying the other two.

As a potential energy surface only covers two reaction coordinates, several surfaces were calculated to describe the reaction. One of the coordinates (distance between C and O) was kept fixed throughout the calculation of one PES and was parametrically changed for each surface. Three surfaces were calculated with r(C-O) = 1.4 Å , 1.6 Å and 2.0 Å (see Figure 7.31). The resulting reaction path is shown in Figure 7.32. The path still shows a barrier of 32.9 kcal/mol. It also predicts the deprotonation to be the first step of the reaction mechanism. From that it could be concluded that the reaction would not take place under the present conditions. However, there are further parameters that have not been considered until now.

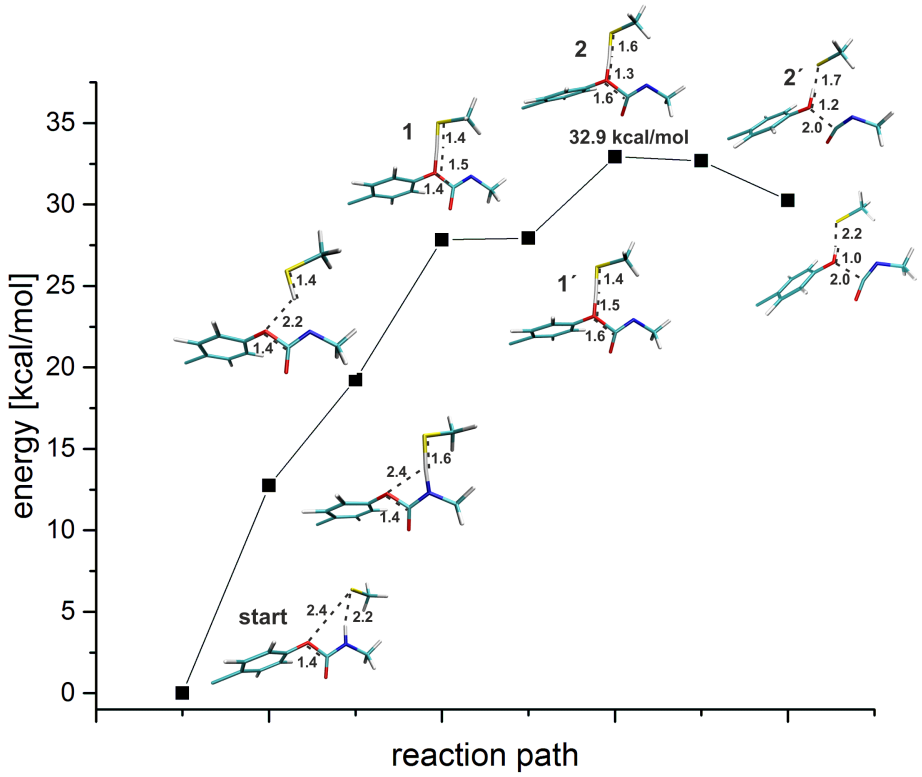


Abbildung 7.32: Reaction pathway (QM) for the concerted mechanism, resulting from the three PES (B3LYP/aug-cc-pVDZ) shown in Figure 7.31.

7.3.3 Nudged Elastic Band Calculations

The reaction under consideration seems to depend on a number of parameters. In order to include all relevant degrees of freedom into the reaction mechanism, Nudged Elastic Band (NEB) calculations were performed. For this purpose, the

system was enlarged to its original size and examined using QM/MM hybrid methods. The procedure entails an optimization of the desired start and end structure of the reaction, which were taken as input structures of the NEB calculation. The probability of an iterative NEB calculation to converge depends on the amount that the system changes during the reaction. Thus, not only the start and end structure of the mechanism, but also intermediate products were optimized (see Figure 7.33).

Geometry Optimizations

The key structures of the mechanism **W1**, **W2**, **W3'** and **W3** (see Figure 7.33) were optimized on QM/MM level using B3LYP/cc-pVDZ as the QM method. The calculations of the structure of the deprotonated inhibitor (**W2** ≡ **B2** in Figure 7.16) did not converge, but instead changed back to the reactant structure. The resulting structure will be called **W1'**. This result is in agreement with the previously described calculations, indicating that **W2** is not an intermediate partaking in the mechanism. As a consequence, only the NEB path of solely the deprotonation could not be calculated. Instead, the reaction path between**W1** and **W3** and from **W1** to **W3'** were computed.

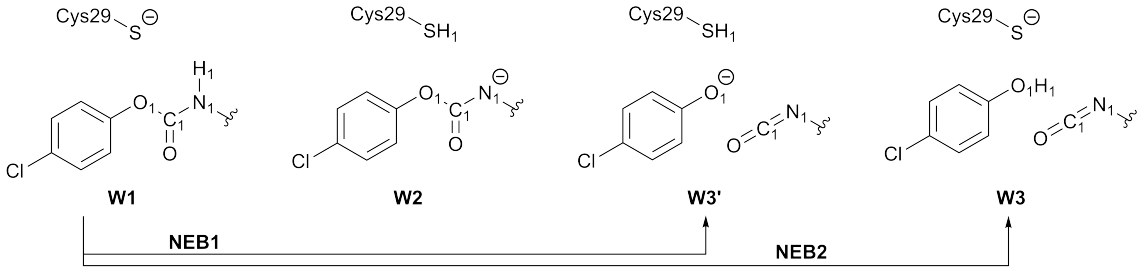


Abbildung 7.33: NEB reaction pathways for the system under consideration. All structures were optimized at the QM/MM level of theory with B3LYP/cc-pVDZ as QM method. **W1**, **W2**, **W3'** and **W3** ≡ **A**, **B2**, **C1** and **C11** respectively in Figure 7.16; n.c. = not converged.

Single point calculations with B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ as QM methods were subsequently performed for selected structures. The results are listed in Table 7.1. The comparison of the MM energy contributions reveal energy differences of 40 - 50 kcal/mol between **W1** and the remaining structures.

Tabelle 7.1: Energy differences between structures **W1**, **W1'**, **W3** and **W3'**. Single Point QM/MM energies (QM methods: B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ (bold)) on optimized geometries (B3LYP/cc-pVDZ as QM method). Energies are given in kcal/mol.

	W1 - W1'	**W1 - W3'**	**W1 - W3**	**W1' - W3**	**W1' -W3'**	**W3' - W3**
QM	1.9 / **2.2**	14.3 / **11.7**	23.8 / **22.1**	17.7 / **15.5**	27.2 / **25.8**	9.5 / **10.3**
MM	-46.5/ **-46.5**	-44.7 / **-44.7**	-40.9 / **-40.9**	3.3 / **3.3**	7.1 / **7.1**	3.8 / **3.8**
QM/MM	-48.4 / **-48.7**	-30.4 / **-32.9**	-17.1 / **-18.8**	20.9 / **18.7**	32.9 / **32.9**	13.3 / **14.1**

In order to determine the origin of this phenomenon, the two structures of **W1** and **W1'** were examined more closely, as they are structurally close, but still show an energetic difference of nearly 50 kcal/mol. Since this examination includes an enlargement of the QM area, a small benchmark was performed, in order to see if a computationally less demanding method than SCS-MP2/aug-cc-pVDZ is sufficient for these calculations. In Table 7.2, the difference of the single point QM/MM energies between **W1** and **W1'** are compared for different QM methods.

Tabelle 7.2: Energy differences between structure W1 and W1'. Single Point QM/MM energies on optimized geometry (B3LYP/cc-pVDZ as QM method). Energies are given in kcal/mol.

	PM3	B3LYP/cc-pVDZ	B3LYP/aug-cc-pVDZ	SCS-MP2/aug-cc-pVDZ
QM	-3.9	-2.3	-3.4	-3.7
MM	-47.9	-47.9	-47.9	-47.9
QM/MM	-51.9	-50.3	-51.3	-51.7

The results show that PM3 is sufficiently by accuracy for the usage in further investigations. The QM energy contribution only differs by 0.2 kcal/mol, compared to the SCS-MP2/aug-cc-pVDZ energy. As the same force field is used for all calculations, the MM contributions are identical. The QM area is now enlarged to radii of +3 Å, +5 Å and +7 Å, relative to the original QM area (QM_{norm}: inhibitor, Cys29 and His162). Additionally, a calculation excluding the water molecules from the MM area was performed. The results are summarized in Table 7.3. The enlargement of the QM area does not lead to significant changes in the energies. The neglect of the water molecules, in contrast, diminishes all differences in the MM energy contributions. The energy differences thus originate from the water shell. The structures of **W1** and **W1'** were aligned in order compare the arrangement of the water molecules for both systems (see Figure 7.34).

Tabelle 7.3: Energy differences between structure **W1** and **W1'**. Single Point QM/MM energies (PM3 as QM method) with different QM areas. Energies are given in kcal/mol.

	QM_{norm}	QM_{norm} +3 Å	QM_{norm} +5 Å	QM_{norm} +7 Å	QM_{norm} no water
QM	-3.9	-0.4	-0.5	-1.2	0.2
MM	-47.9	-49.2	-48.6	-48.4	0.0
QM/MM	-51.9	-48.8	-49.1	-49.6	0.2

The alignment shows that a large number of water molecules slightly differ in their respective position. Each deviation only contributes a small amount to the total energy. Nevertheless, the sum of all contributions can result in a significant energy difference. For a verification of this assumption, the two structures **W1** and **W1'** are compared to **W3'** (see Figure 7.35). As shown in Table 7.1, the MM energy contributions of **W3'** and **W1'** only have a small difference of 3.3 kcal/mol, whereas those of **W3'** and **W1** differ by 44.7 kcal/mol. If the assumption that the position of the water molecules is responsible for the MM energy difference is correct, the water shells of **W1'** and **W3'** should be nearly identical. At the same time **W1** and **W3'** should show the same deviations as the ones of **W1** and **W1'**, depicted in Figure 7.34.

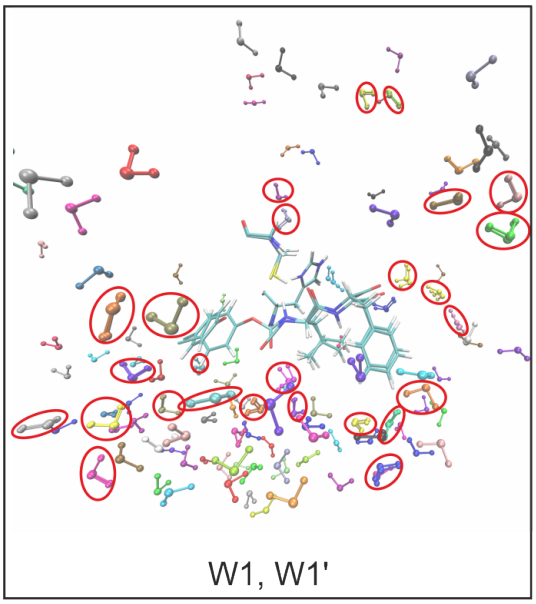


Abbildung 7.34: Structures **W1** and **W1'** with surrounding water molecules. The respectively identical water molecules of the systems have the same colour. Molecules with slightly different orientations are highlighted by circles.

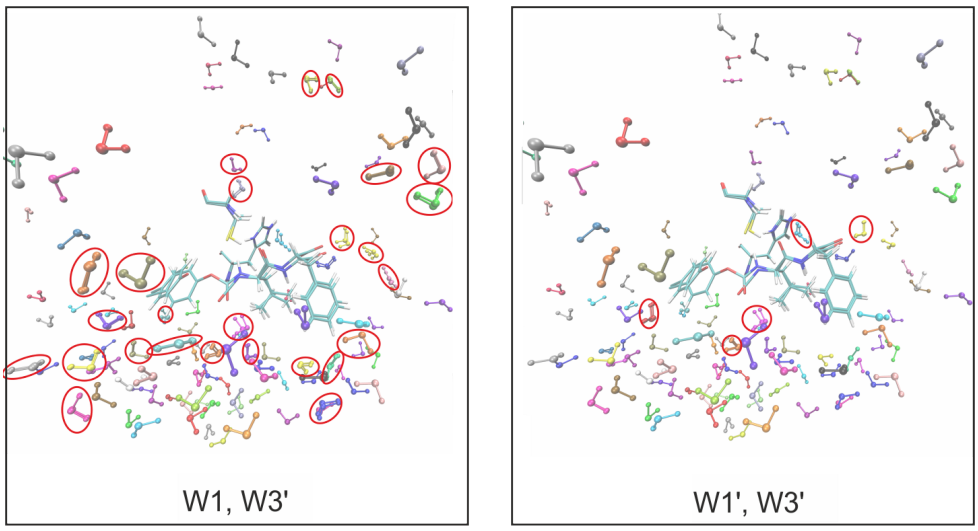


Abbildung 7.35: Comparison of the structures **W1**, **W3'** (left) and **W1'**, **W3'** (right). The respectively identical water molecules of the systems have the same colour. Molecules with slightly different orientations are highlighted by circles.

The comparison of the structures supports the assumption that the large energy differences between **W1'** and the remaining structures originate from the position of the water molecules. As a consequence, the structure **W1'** was from now on used as the reactant structure for further investigations. The single point energies (B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ) of **W1'**, **W3'** and **W3** are summarized in Table 7.4.

Tabelle 7.4: Energy differences between the structures **W1'**, **W3** and **W3'**. Single Point QM/MM energies (B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ in bold) as QM methods on optimized geometries (B3LYP/cc-pVDZ as QM methods). Results from the thermodynamic QM calculations with the model system from Chapter 7.2 are given for comparison. Energies are given in kcal/mol.

	W1' - W3	**W1' - W3'**	**W3' - W3**
QM	17.7 / **15.5**	27.2 / **25.8**	9.5 / **10.3**
MM	3.3 / **3.3**	7.1 / **7.1**	3.8 / **3.8**
QM/MM	20.9 / **18.7**	32.9 / **32.9**	13.3 / **14.1**
Model System	**16.8**	**21.5**	**4.7**

The Table additionally shows the reaction enthalpies (SCS-MP2/aug-cc-pDVZ) obtained from the thermodynamic calculations within the model system (see Chapter 7.2). Concerning the the energy difference between **W1'** and **W3**, the results are in a relatively good agreement between the QM contribution of the QM/MM

calculation (SCS-MP2) and the QM calculation of the model system (SCS-MP2) (~ 2 kcal/mol). The remaining results indicate that **W3'**, the intermediate structure with the deprotonated phenolate anion, is less stabilized within the QM/MM calculations. With a shift of around 10 kcal/mol, the negative phenolate anion is energetically more favoured inside the model system. This could be due to the fact that, in contrast to the protein environment, the implicit solvation contribution of COSMO is self-consistently optimized for each system. A possible stabilization of the phenolate through non-covalent interactions, as discussed in Chapter 7.3.2, could however lower the energy of **W3'**.

Nudged Elastic Band Calculations

Based on the introduced structures **W1'**, **W3'** and **W3**, NEB calculations were performed at the QM/MM level of theory with B3LYP/cc-pVDZ as QM method. The two resulting reaction pathways from **W1'** to **W3'** (**A** → **C1** in Figure 7.26) and to **W3** (**A** → **C11** in Figure 7.26) will be discussed in the following. The reaction path between **W1'** and **W3'** is depicted in Figure 7.36.

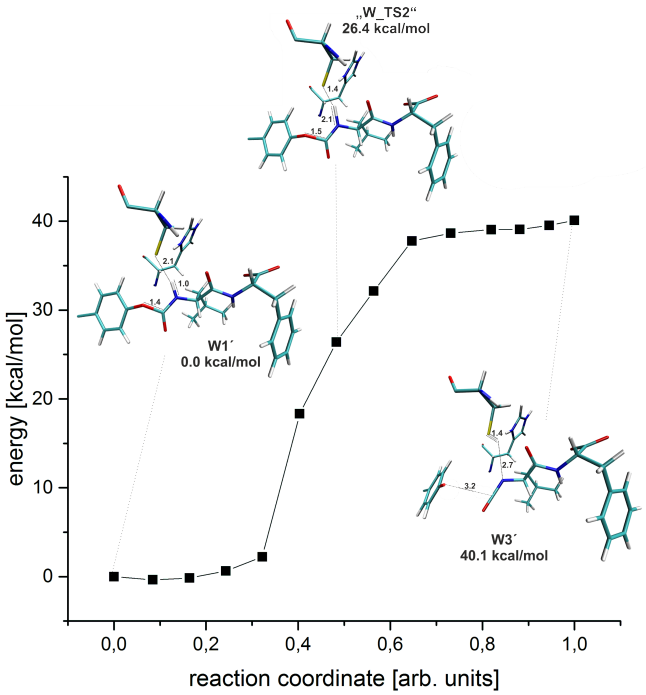


Abbildung 7.36: NEB path (B3LYP/cc-pVDZ as QM method) of the reaction from **W1'** to **W3'**.

The reaction has no local minimum at the structure of **W3′**. The reaction barrier is 40.1 kcal/mol, which indicates that the reaction will not take place in this setting. Figure 7.36 illustrates the progression of the relevant distances throughout the reaction. The deprotonation and dissociation of the inhibitor begin simultaneously at 0.4 on the reaction coordinate. However, the deprotonation is finished at 0.5 (r(H_1-S) = 1.4 Å), while the C-O bond is still at 1.5 Å, which means it is not dissociated completely at this point. The actual dissociation of the inhibitor occurs at 0.7 on the reaction coordinate.

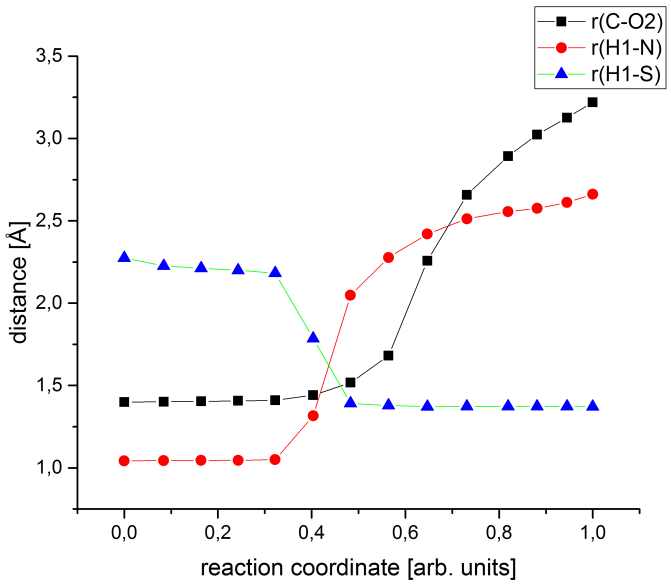


Abbildung 7.37: Changes of relevant distances during the reaction path from **W1′** to **W3′**.

Figure 7.38 shows the NEB reaction path between **W1′** and **W3**. In contrast to the previous reaction, the product structure is now a local minimum of the PES. The reaction barrier of 43.8 kcal/mol is in a similar range as the barrier between **W1′** and **W3′**. In accordance with the comparable energy barriers, the structure **W_TS2** in Figure 7.36 (highest point) practically corresponds to **W3′**. The progression of the relevant parameters throughout the reac-tion is shown in Figure 7.39. The deprotonation (r(H_1-N)) and dissociation (r(C-O_2)) are completely concerted and occur between 0.4 and 0.6 on the reaction coordinate. The protonation of the phenolate oxygen (r(H_1-O_2)) takes place subsequently.

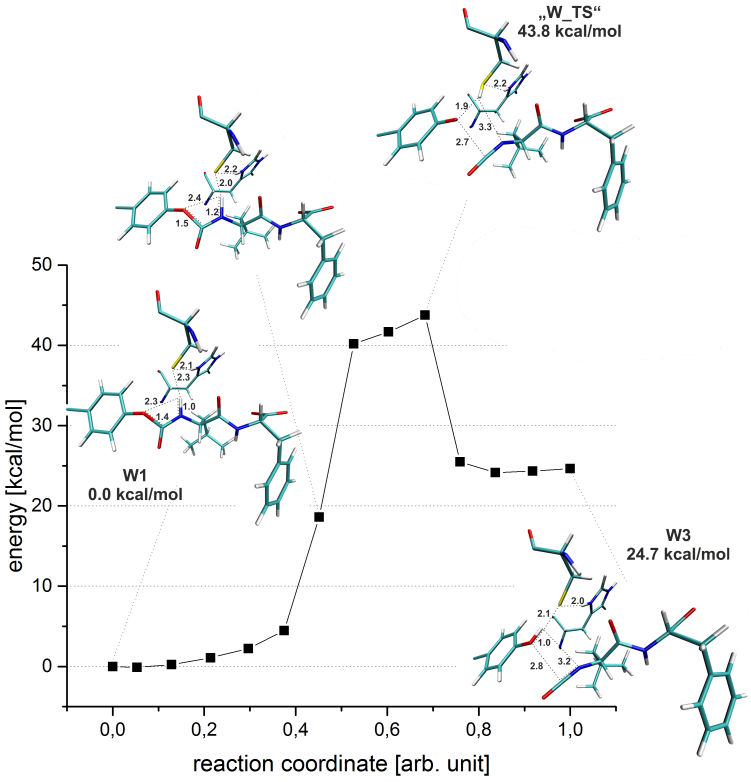


Abbildung 7.38: NEB path (B3LYP/cc-pVDZ as QM method) of the reaction from **W1′** to **W3**.

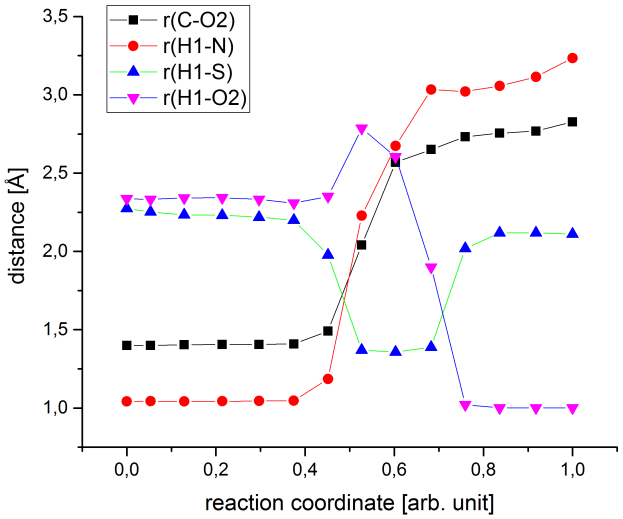


Abbildung 7.39: Changes of relevant distances during the reaction path from **W1′** to **W3**.

In conclusion, none of the applied approaches lead to an energy barrier of the reaction below 30 kcal/mol. An E1cB elimination of carbamate-based inhibitors would thus not take place in cathepsin B. However, further considerations based on a recent publication of the group of Prof. Engels in Würzburg lead to the idea of a water-bridged proton transfer.[34] The following Chapter concentrates on the implementation of this idea.

7.3.4 Water-Mediated Proton Transfer

During the simulation described in Chapter 7.3.2, a water molecule moved inside the active site of the enzyme. The position of the water molecules would allow for a water-mediated proton transfer between Cys29 and the inhibitor (see Figure 7.40).

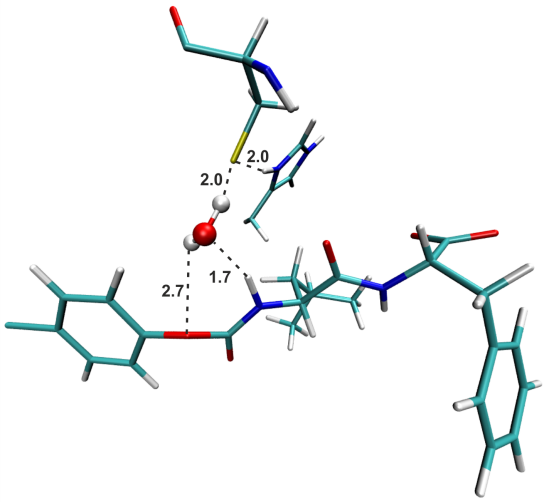


Abbildung 7.40: Orientation of the water molecule that moved into the active site during the MD simulation.

The reaction mechanism now entails a water molecule, that mediates both the deprotonation of the inhibitor and the protonation of phenolate, as shown in Figure 7.41. As this mechanism introduces further reaction coordinates, several potential energy surfaces were calculated (QM/MM level with PM3 as QM method) with different degrees of freedom that were kept fixed in order to obtain an overview of potential intermediate structures for the NEB calculations. The surfaces with the lowest reaction barriers are discussed in the following (the remaining surfaces are found in the appendix of this work). The PES of the water-mediated deprotonation of the inhibitor is shown

in Figure 7.42. With a barrier of 18.2 kcal/mol, the reaction is in a more favourable energetic area compared to the previous results without a participating water molecule. The calculations predict an intermediate structure with an isolated oxonium ion and the protonation of Cys29 in a separated second step. The surface contains three significant structures that may be utilized as starting points for the calculation of the inhibitor dissociation. Those structures are the transition state **W12_wat_TS**, the intermediate **W12_wat_INT** with an isolated hydronium ion and the structure **W2_wat** of the deprotonated inhibitor (see Figure 7.42). **W2_wat** will not be used as starting structure for further investigations, as the next reaction step includes a deprotonation of Cys29 and the reaction would thus proceed back to structure **W12_wat_INT**. For **W12_wat_TS** and **W12_wat_INT** the cysteine sulfur already interacts with H_{A1} but did not form a covalent bond that would have to be broken for the next reaction step.

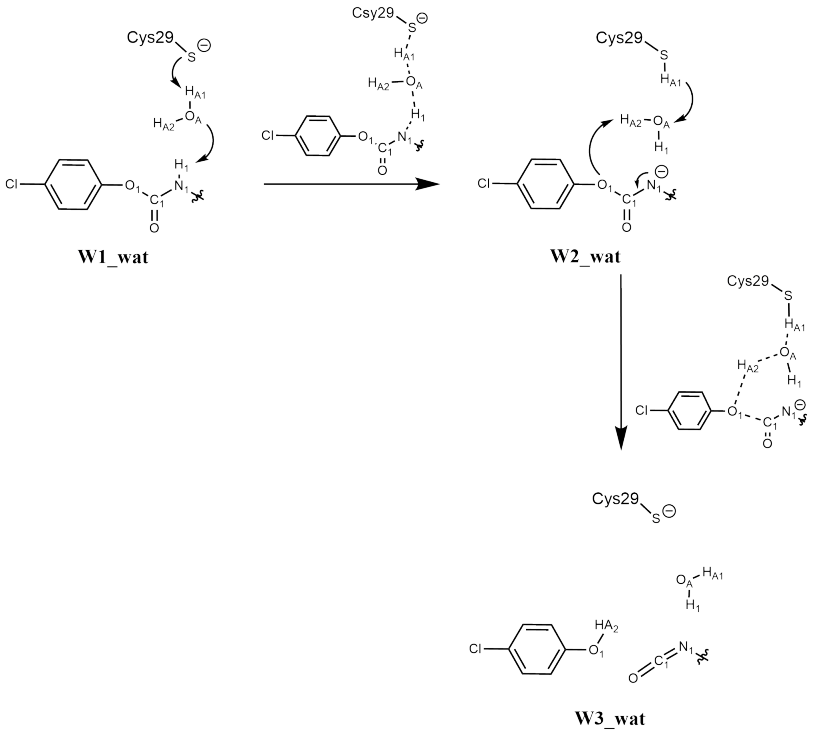


Abbildung 7.41: Reaction scheme of a water-mediated proton transfer between the inhibitor and Cys29.

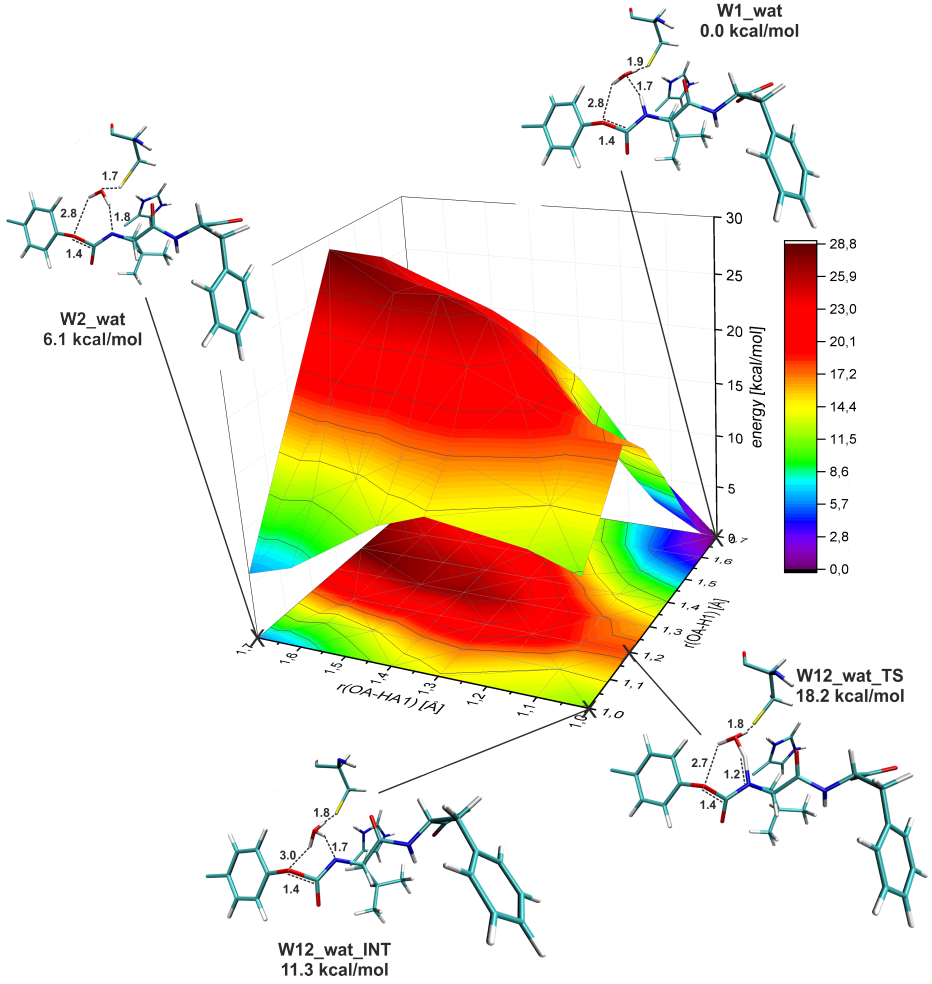


Abbildung 7.42: PES of the water bridged proton transfer (PM3 as QM method). **W1_wat** and **W2_wat** ≡ **A** and **B2** in Figure 7.26.

For the investigation of the following step (the dissociation of the inhibitor and the protonation of phenolate), three potential energy surfaces were calculated. Additionally to these evaluations, starting from **W12_wat_TS** and **W12_wat_INT**, a surface with **W1_wat** as starting structure is performed in order to investigate the possibility that Cys29 is not included in the mechanism at all. In this case, H_1 would be abstracted by the water molecule and subsequently be transferred to O_1 (phenolate oxygen) during the dissociation. Figure 7.43 shows the potential energy surface of this reaction. A second PES of this variant with another set of reaction variables was calculated that had an even higher barrier of 64.0 kcal/mol (see appendix of this work).

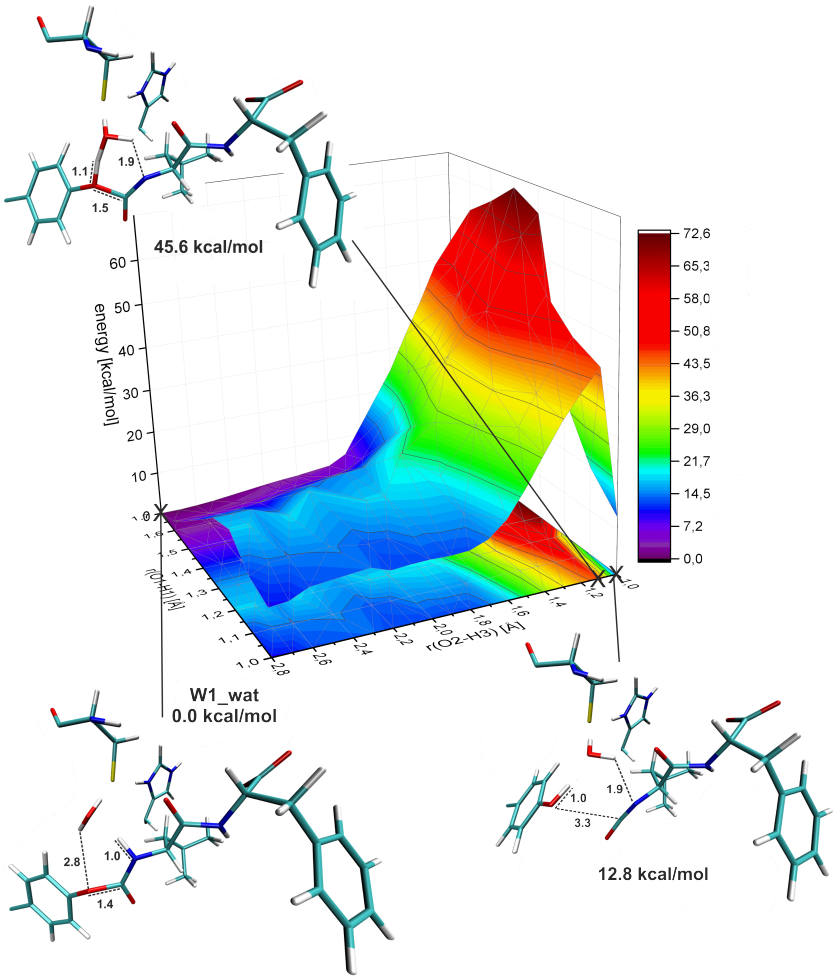


Abbildung 7.43: PES of the water bridged dissociation of the inhibitor with the simultaneous protonation of phenolate (PM3 as QM method) without the participation of Cys29.

The surface shows that the participation of Cys29 seems to be crucial for the reaction, as the surface excluding Cys29 as a reactive component has a reaction barrier of 45.6 kcal/mol. It is important to note that all calculations were performed with the semi empirical method PM3 and only give a qualitative assessment. The herein discussed mechanism without Cys29 is nevertheless energetically the most unfavourable one compared to the others and additionally excludes the key amino acid of the enzyme. Therefore it will not be further considered in this work. In comparison of the remaining two surfaces, the path with starting structure **W12_wat_TS** shows a lower reaction barrier and was therefore exclusively considered for further investigations. Figure 7.44 depicts the potential energy surface of the dissociation and protonation

of the inhibitor starting from structure **W12_wat_TS'** with an energy of 10.6 kcal/mol (gray values in Figure 7.44), relative to the reactant structure **W1_wat**. The structure **W12_wat_TS'** results from a constrained optimization of **W12_wat_TS** with the new reaction coordinates for the dissociation and protonation of the inhibitor.

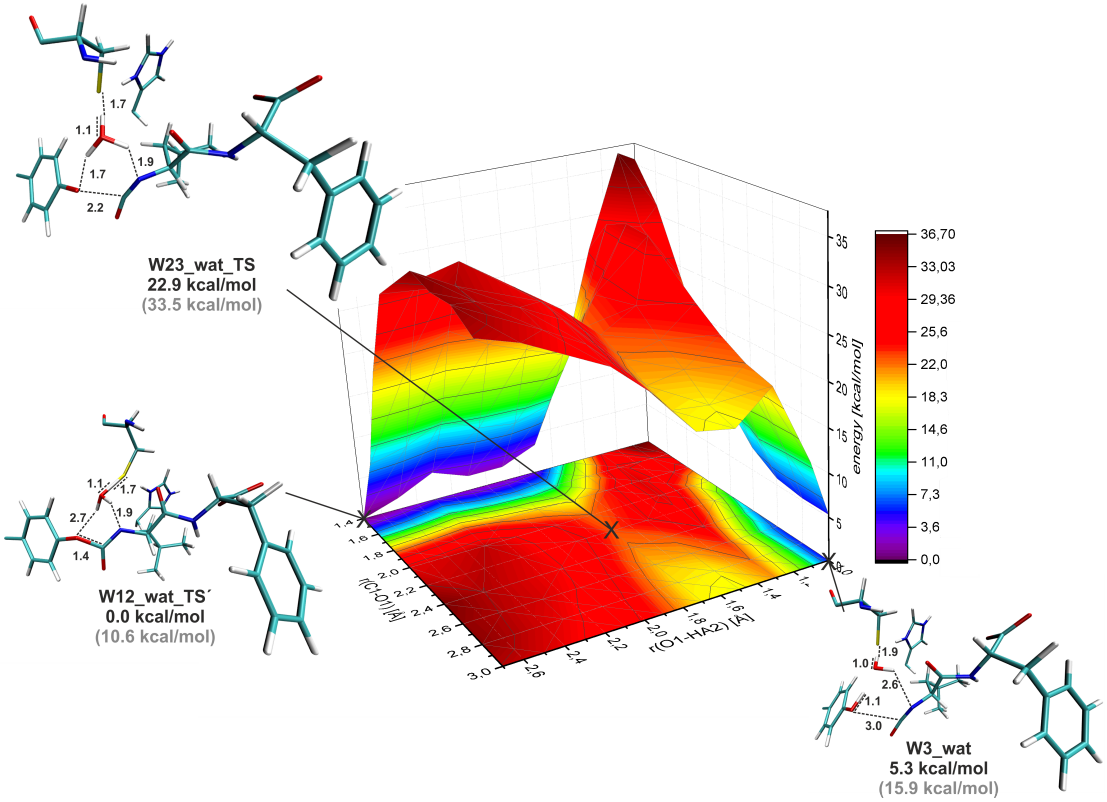


Abbildung 7.44: PES of the water bridged dissociation of the inhibitor with the simultaneous protonation of phenolate (PM3 as QM method) with starting structure **W_TS'**. Grey values are the energies of the respective structures relative to structure **W1_wat** in Figure 7.42.

The calculations predict the dissociation to be the initial step of the reaction with subsequent protonation of the phenolate. The reaction barrier is 33.5 kcal/mol (with PM3 as QM method) relative to the reactant structure **W1_wat** (gray values in Figure 7.44). A combination of the two surfaces depicted in Figure 7.42 and 7.44 gives a reaction profile of the water-mediated deprotonation and dissociation of the inhibitor with subsequent protonation of the phenolate anion. Figure 7.45 shows the relative energies of the key structures. The geometries calculated with PM3 as QM method

were used for QM/MM single point calculations with higher quantum mechanical levels for the QM area (B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pVDZ).

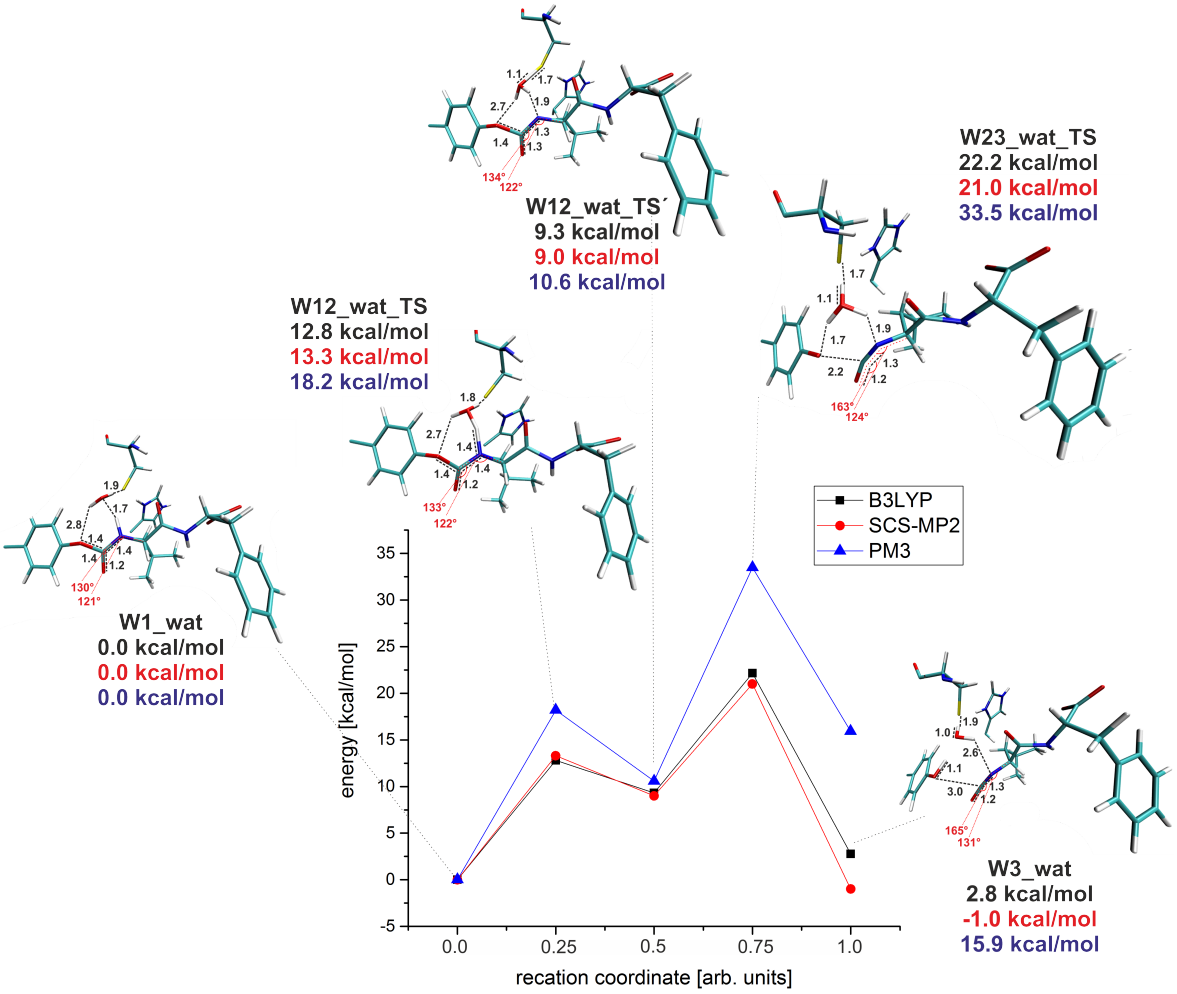


Abbildung 7.45: Reaction profile of the deprotonation and dissociation of the inhibitor. QM/MM optimizations were performed with PM3 as QM method. QM/MM single point calculations were performed with B3LYP and SCS-MP2 as QM method.

The highest reaction barrier of the profile is 21.0 kcal/mol (SCS-MP2/aug-cc-pVDZ as QM method), which is again in a more favourable energetic range compared to the calculated reaction pathways without a mediating water. In the following, different attempts were applied in order to find a path with an even lower reaction barrier. Several potential energy surfaces with different reaction coordinates, including the angles shown in Figure 7.45 in combination with a number of bonds were calculated.

The PES thereof are depicted in the appendix of this work. The approaches did not reveal new results. Comparable to the system without an additional water molecule, two reaction coordinates do not seem to be sufficient for an accurate description of the reaction. Therefore, NEB calculations were performed for the system.

Nuged Elastic Band Calculations

Similar to the procedure described in Chapter 7.3.3, three NEB calculations were performed for the system (see Figure 7.46). The calculation describing the whole E1cB elimination including the protonation of the phenolate (NEB12_wat) did not converge, as the structural changes during the reaction are too drastic. The two calculations NEB1_wat and NEB2_wat did however converge. Figure 7.47 depicts the resulting reaction path in one normalized coordinate with equidistant marks (one for each frame of the calculation).

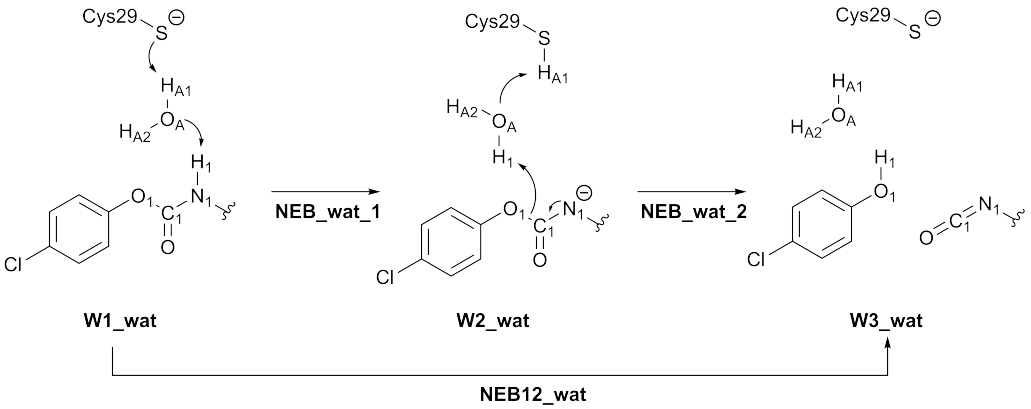


Abbildung 7.46: Reaction scheme of the performed NEB calculations with a water-mediated proton transfer.

The reaction barrier of the deprotonation (15.3 kcal/mol) is similar to the respective barrier calculated from the potential energy surfaces (see Figure 7.45). The barrier for the dissociation of the inhibitor is now significantly lower at 9.8 kcal/mol and the reaction path does not include an isolated hydronium ion. The small barriers at 0.2 and 0.7 on the reaction coordinate occur from a reorientation of the water molecule. Figure 7.48 shows the progression of all relevant distances during the reaction. The water-mediated proton transfer between the inhibitor and Cys29 occurs as a concerted reaction. The abstraction of the proton (r(O_A-H_1)) changes simultaneously to the

protonation of Cys29 (r(S-H_{A1})). After cysteine is in its protonated form, His199 rotates away from the sulfur atom towards the recognition unit of the inhibitor (r(S-His) in Figure 7.47). At 0.7-0.9 of the reaction coordinate, the protonation of phenolate (r(O_1-H_{A2})) takes place simultaneously to the dissociation of the inhibitor (r(C_1-O_1)).

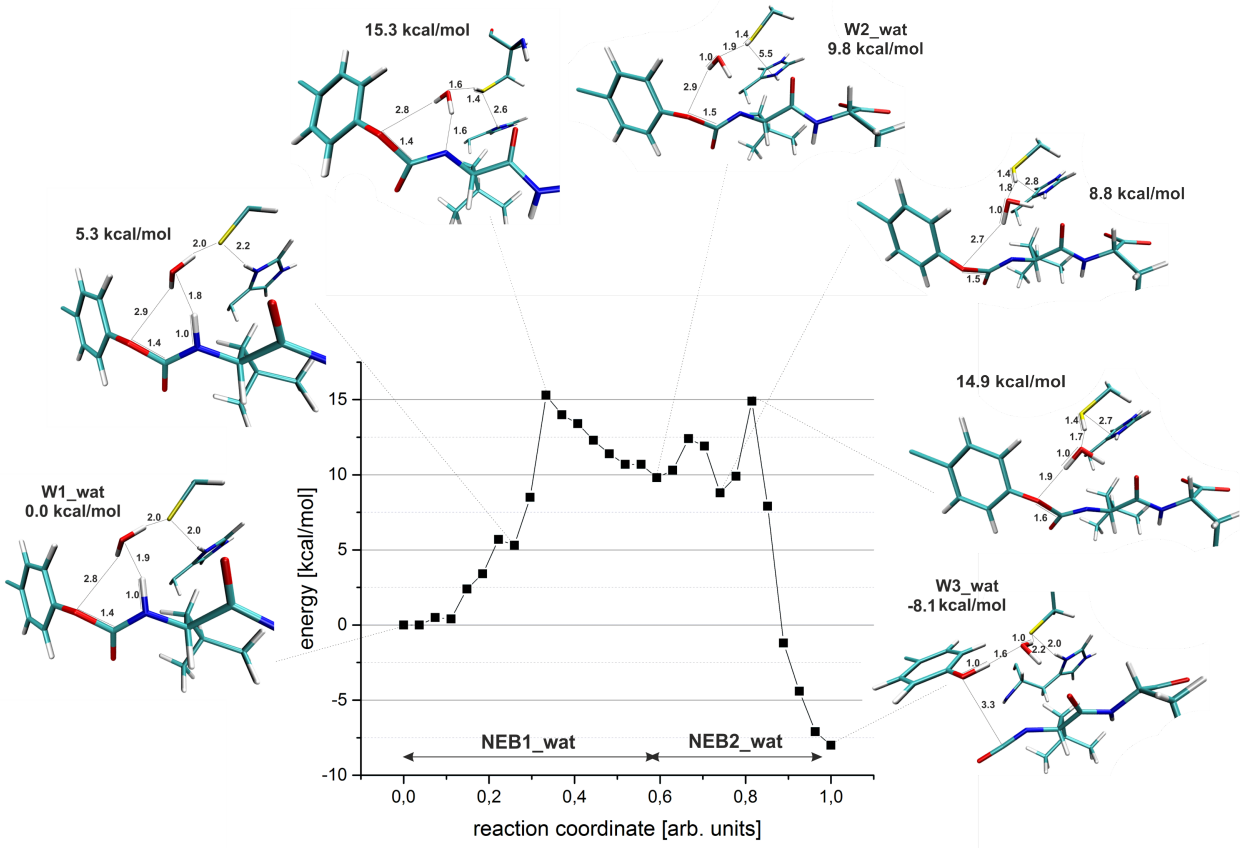


Abbildung 7.47: NEB path (B3LYP/cc-pVDZ as QM method) of the reaction from **W1_wat** to **W3_wat**. NEB profiles from **W1_wat** to **W2_wat** and from **W2_wat** to **W3_wat** were combined in one reaction coordinate, the x-axis therefore contains equidistant marks, one for each frame of the calculation.

The results of the NEB calculations show that the reaction depends on a number of different coordinates and is partially concerted. The inclusion of a water molecule in the active site leads to a significant reduction of the reaction barriers and results in an exothermic reaction with a barrier of only 15.3 kcal/mol. An inhibition of cathepsin B with a carbamate-based inhibitor, initiated by an E1cB mechanism, could thus take place.

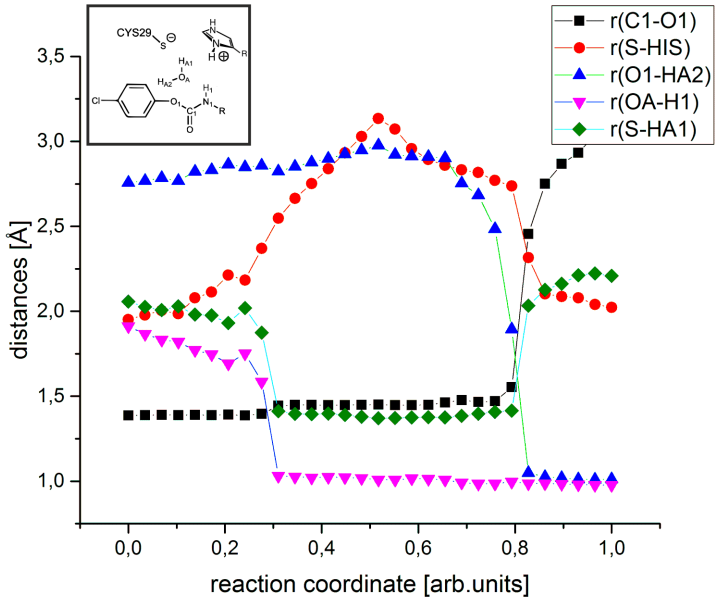


Abbildung 7.48: Changes of relevant distances during the NEB reaction path from **W1_wat** to **W3_wat**.

Proton Transfer *via* Two Water Molecules

As the reaction barrier was significantly reduced by the inclusion of a water molecule, the need to analyze the effect of a second water molecule inside the active site arose. The corresponding reaction scheme is depicted in Figure 7.49.

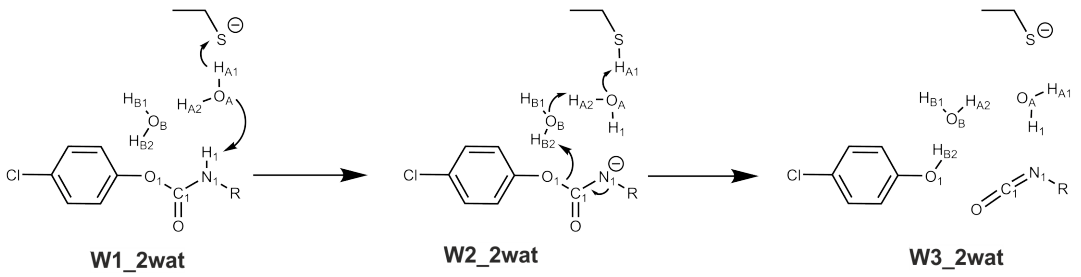


Abbildung 7.49: Inhibition scheme for carbamate-based inhibitors inside cathepsin B with two water molecules inside the active site.

The modified system was prepared by the manual insertion of a second water molecule inside the active site. Both the protonated and deprotonated inhibitor structures were used for this procedure. Several attempts were needed, as the second water

molecule repeatedly diffused outside the reactive center of the enzyme during geometry optimizations. With help of several constrained force field optimizations and subsequent unconstrained optimization at the QM/MM level of theory, both water molecules could be brought in a stable position between the deprotonated inhibitor and Cys29 (marked red in Figure 7.50). Unfortunately, the orientation of the molecules is unsuitable for the reaction mechanism. Therefore, they were rotated in the correct position (marked green in Figure 7.50) *via* a potential energy surface (PM3 as QM method). However, the two water molecules did not stay in their new orientation during an unconstrained optimization of the system. The conclusion is drawn that the active site of the enzyme does not appear to have enough space for two water molecules. The effect of two water molecules inside the active site on the mechanism was therefore not investigated any further.

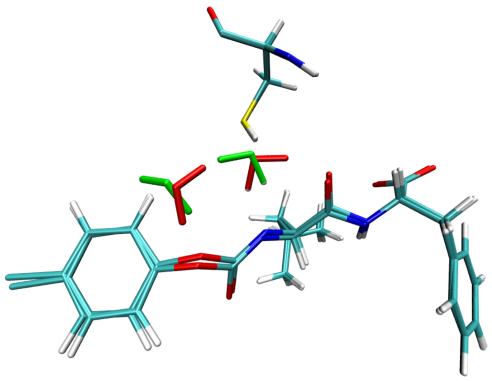


Abbildung 7.50: Deprotonated inhibitor inside the active site of cathepsin B with two different orientations of the water molecules (Inhibitor not shown).

7.4 Step Two: Carbamylation

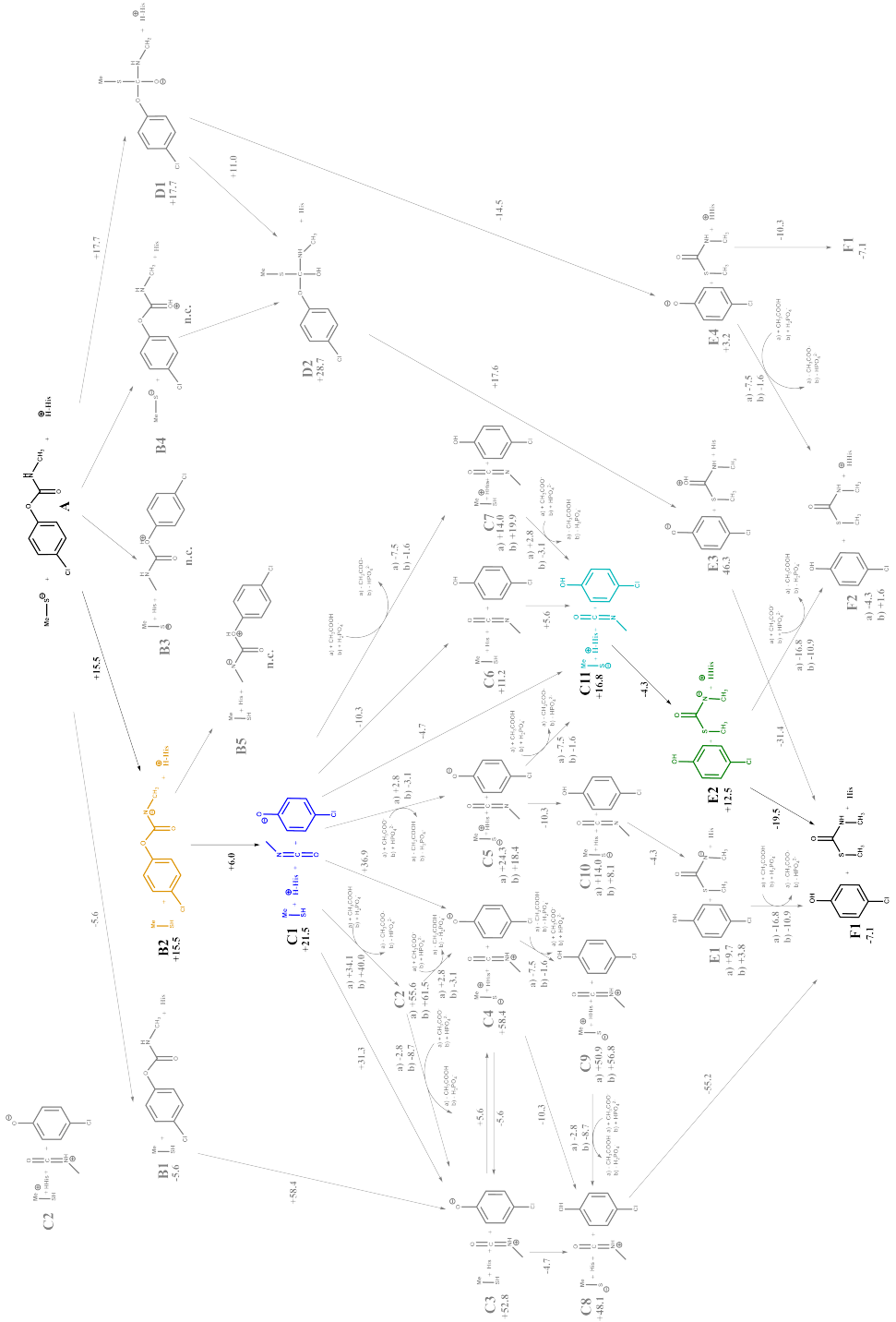


Abbildung 7.51: Possible reaction mechanisms for cathepsin B with a carbamate warhead. All structures are optimized using SCS-MP2/aug-cc-pVDZ including implicit solvation (COSMO; ϵ = 78.93). Energies are given in kcal/mol; n.c.: not converged. The examined mechanism is indicated in bold.

On the basis of the gained results, the carbamylation of Cys29 (see emphasized path in Figure 7.51 will be investigated in the following. Figure 7.52 illustrates the reaction mechanism between the compounds. The reaction includes two steps, the nucleophilic attack of the deprotonated Cys29 at the isocyanate carbon and the protonation of the nitrogen atom through His199 (see Figure 7.52).

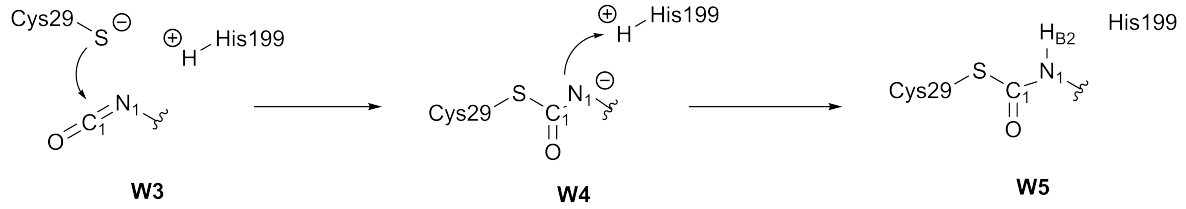


Abbildung 7.52: Reaction scheme of the carbamylation of Cys29 with an isocyanate and His199. **W3**, **W4** and **W5** ≡ **C11**, **E2** and **F1** in figure 7.51.

7.4.1 QM Calculations on a Model System

The influence of the protonation state of Cys29 on the reaction was investigated within a model system using a DFT approach. These investigations occur from an earlier attempt, where the protonation of the phenolate anion through Cys29 was not discussed yet. For an attack of the protonated cysteine, a *syn* orientation would occur. A deprotonated cysteine in contrast needs an external proton donor, that could be in an *anti* orientation relative to Cys29. These two possibilities were compared with help of potential energy surfaces.

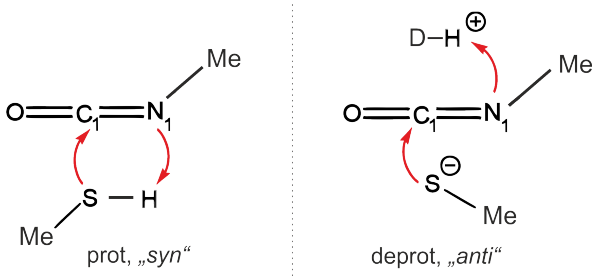


Abbildung 7.53: Scheme of the two reaction mechanisms for a carbamylation within a model system. Left: protonated Cys29 with a *syn* protonation; right: deprotonated Cys29 with an *anti* protonation.

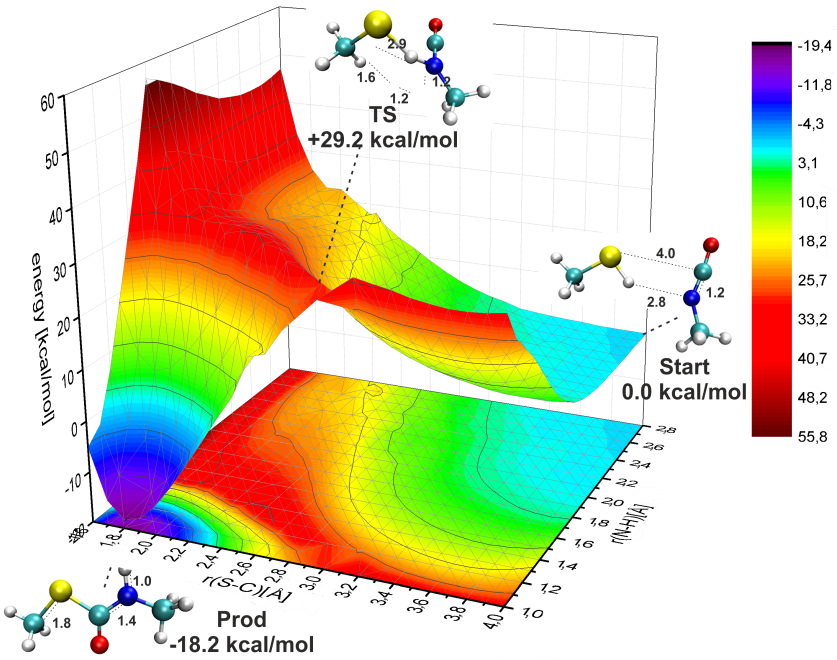


Abbildung 7.54: Potential energy surface (B3LYP/aug-cc-pVDZ with COSMO; ϵ=78.39) of the carbamylation of a protonated methanethiol.

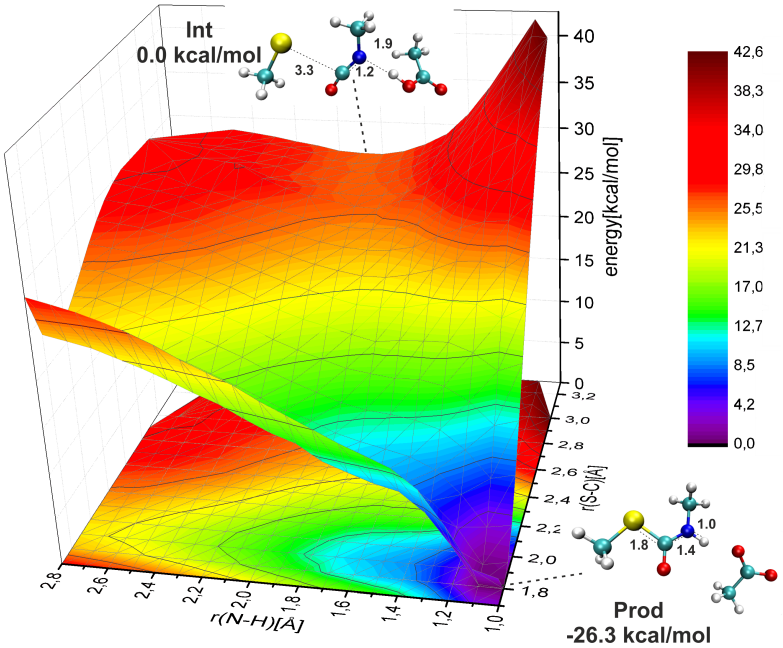


Abbildung 7.55: Potential energy surface (B3LYP/aug-cc-pVDZ with COSMO; ϵ=78.39) of the carbamylation of a deprotonated methanethiol with acetic acid as external proton donor.

The surfaces (see Figure 7.54 and 7.55) show a significant difference in the reaction barriers. While the reaction with a protonated thiol and a *syn* attack has a barrier of 29.2 kcal/mol, the deployment of a proton donor in *anti* orientation to the nucleophilic attack is barrier-free. This indicates that the reaction in the enzymatic environment would probably also be favoured if Cys29 attacks in a deprotonated form, while the protonation occurs *anti* to the nucleophilic attack from a proton donor. This is in accordance with previous results, which indicate a deprotonation of Cys29 before the carbamylation takes place.

7.4.2 Orientation of the Isocyanate

The structure arising from the calculations of the E1cB mechanism (**W3_wat** in Figure 7.47) was taken for the simulation of the final reaction steps on QM/MM level. Therefore, the phenolate was removed from the system, as it is assumed to diffuse out of the pocket and otherwise would provide sterical hindrance to the reaction. In the resulting structure, the distance between the isocyanate carbon and Cys29 exceeds 5 Å and the functional group faces in the opposite direction of the active site amino acid. This does thus not provide a good starting point for the reaction. Members of the group of Prof. Guetschow previously docked the structure of the isocyanate intermediate inside the active site of cathepsin B (crystal structure 1HUC[1,26]). This structure was used to investigate if the inhibitor maintains a stable position inside the active site or if a reorientation is possible. For this purpose, a MD simulation of the docked structure was performed. The simulation ran over 10 ns with time steps of 1 fs. Figure 7.56 illustrates the RMSD values of different parts inside the system relative to the first frame. Both amino acids of the catalytic dyade stay in a stable position during the simulation. Also, the entire enzymatic backbone does not show significant conformational changes (see Figure 7.56; protein backbone). The inhibitor, in contrast, changes its position and moves relatively free. It repeatedly leaves the active site, but always diffuses back inside, each time with a different orientation. From this it can be assumed that a reorientation of the inhibitor into a more favourable position for the reaction is possible.

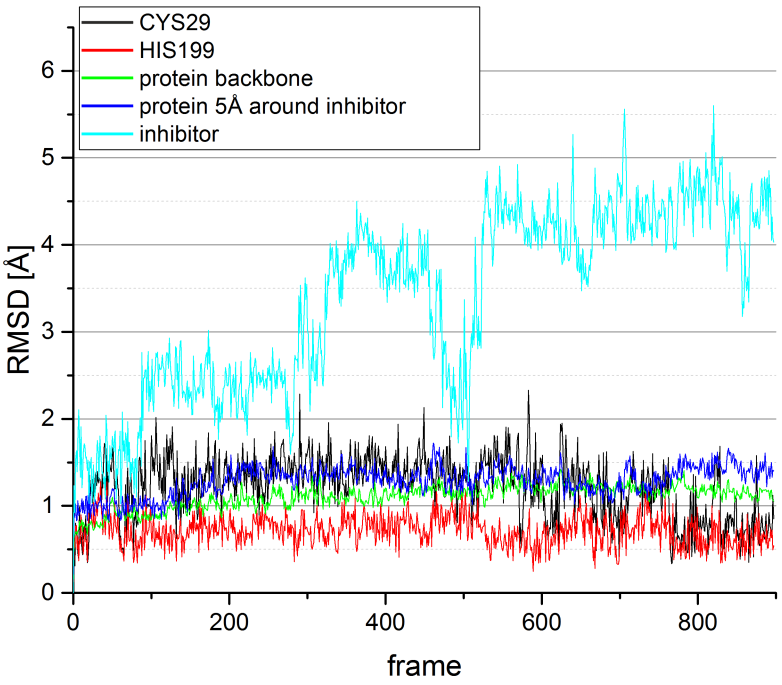


Abbildung 7.56: RMDS values from the MD simulation (10 ns) of the isocyanate intermediate inside the active site of Cathespin B.

7.4.3 Generation of a Staring Structure

Starting from structure **W3_wat** (see Figure 7.46), structure **W3** (see Figure 7.52) was formed though the removal of phenole from the system and the exclusion of the water molecule from the active site. The isocyanate was rotated and placed closer to the active site with the help of several PES (PM3 as QM method). However, the inhibitor moved back into a unfavourable position for unconstrained optimizations. In order to create the structures **W4** and **W5** (≡ **E2** and **F1** in Figure 7.51), shown in Figure 7.52, further PES were performed with the QM/MM approach, using PM3 as QM method. The negatively charged intermediate structure **W4** did not converge in an unconstrained optimization, but instead ran back to structure **W3** (see Figure 7.52). However, the structure remained in a stable position if the sulfur carbon distance is fixed (at 1.9 Å) during the optimization. The resulting structure is shown in Figure 7.57.

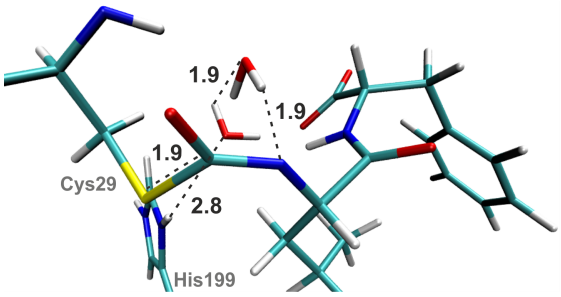


Abbildung 7.57: Structure of **W4** (see Figure 7.52) with two adjacent water molecules.

Two adjacent water molecules are in a favourable position for a water-bridged proton transfer from His199 to the inhibitor's nitrogen. The protonation was calculated by manual repositioning of the hydrogen atoms with subsequent optimizations using QM/MM (B3LYP/cc-pVDZ as QM method). Starting from this newly generated structure **W5**, a surface of the backwards reaction was calculated as a new attempt to create a converged structure **W4**. **W4** did not converge, but instead transformed back to structure **W3**. An energetic comparison of **W3** and **W5**, resulting from single point calculations with B3LYP/aug-cc-pVDZ as QM method, showed a destabilization of the product **W5** by 8.4 kcal/mol compared to the intermediate **W3**. The fact that the product structure lies energetically higher than the reactant is indicating that the enzymatic environment has an impact on the stabilization of the respective structures. A closer investigation of the adjacent compounds reveals that there are no potential interaction partners near the inhibitor to stabilize the warhead. Especially for the negatively charged adduct of structure **W4**, this could have a significant impact on the mechanism. A possible relaxation of the environment to the structures was therefore investigated *via* MD simulations with the isocyanate structure **W3** (resulting from the backwards reaction of W5) and the product structure **W5**. Before the simulation was started, several water molecules were added near the active site to support the formation of stabilizing interactions. Several frames from the simulation of the newly relaxed system **W3** are taken as starting points for QM/MM unconstrained optimizations (PM3 as QM method). None of the calculations converged in a stationary geometry. The structures taken from the simulations of **W5** in contrast, resulted in a converged product structure within a relaxed environment (see Figure 7.58; pre-optimized with PM3 as QM method, optimized using B3LYP/cc-pVDZ as QM method). Additionally, a water molecule diffused into the active site in a favourable position for a water-bridged proton transfer.

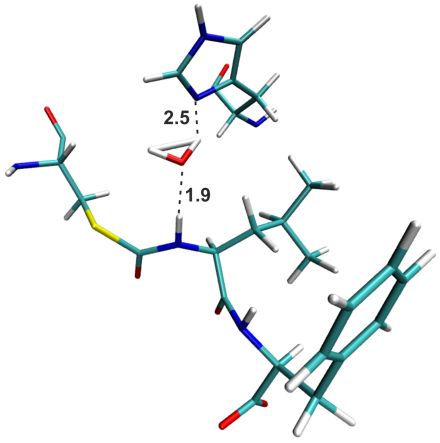


Abbildung 7.58: Product structure **W5** within a relaxed environment, including a water molecule in the active site. The distances are given in Å.

Starting from this structure, QM/MM potential energy surfaces (PM3 as QM method) of the backwards reaction were calculated in order to generate the structures **W3** and **W4** with the additional water molecule inside the active site. All structures converged in a subsequent unconstrained QM/MM optimization with B3LYP/cc-pVDZ as QM method. Figure 7.59 illustrates the structures that are in a following step taken for NEB calculations.

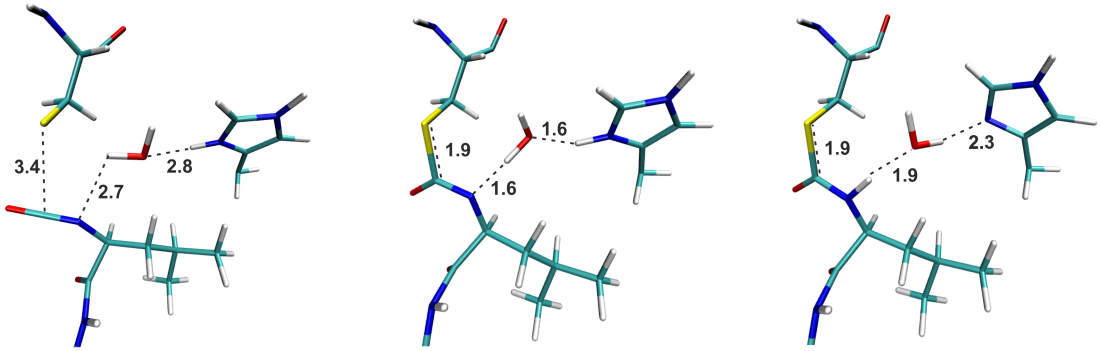


Abbildung 7.59: Converged structures **W3**, **W4** and **W5** with a water molecule in the active site.

7.4.4 Nudged Elastic Band Calculations and Potential Energy Surfaces

Three NEB calculations were performed at the QM/MM level of theory in order to investigate the carbamylation of Cys29 trough the isocyanate-based intermediate. One computation included the whole mechanism, starting from structure **W3** with structure **W5** as the reaction product. The calculation did, as some previous NEB calculations, not converge. As the tendency of a NEB calculation to converge is indirectly proportional to the amount of structural changes during the reaction, two calculations describing the reaction steps from **W3** to **W4** and from **W4** to **W5** were performed. Both calculations (B3LYP/cc-pVDZ as QM method) converged. The reaction path between **W4** and **W5** shows an interesting behaviour, as the structure **W5** is not predicted to be a local minimum. Suitable to the energies of the geometry optimizations (see Figure 7.59), **W5** lies 5.2 kcal/mol below **W4** in the NEB path. However, a structure similar to **W5** (His199 deprotonated, inhibitor protonated) emerges with an energy of -14.7 kcal/mol relative to **W4**. This structure will be called **W5'**.

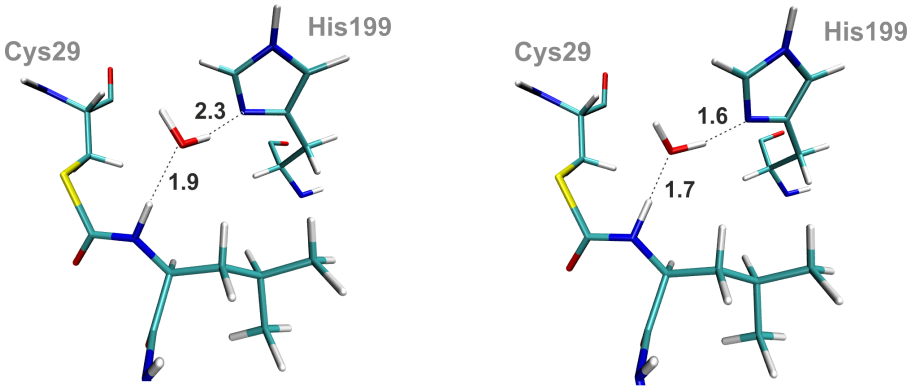


Abbildung 7.60: Structures of **W5** and **W5'**, the two product structures emerging from the NEB calculation from **W4** to **W5**.

A comparison of **W5** and **W5'** reveals a strong hydrogen bonding network between His199, the water and the amino group of the inhibitor (see Figure 7.60) for structure **W5'**. For **W5**, the respective compounds do also interact, but with increased distances. Additionally, small changes in a number of adjacent molecules contribute to the significant energy difference between the two structures of around 10 kcal/mol. Based

on these results, a second NEB calculation was performed, going from **W4** to **W5'**. The resulting reaction pathways (**W3** to **W4** and **W4** to **W5'**) were again combined in one normalized axis (see Figure 7.61) with equidistant marks (one per frame). The nucleophilic attack of sulfur at the isocyanate carbon has a reaction barrier of 5.1 kcal/mol. During this step, the water molecule already starts approaching to His199 (see structure one and two in Figure 7.61). The reaction has an enthalpy of -13.0 kcal/mol and segues nearly barrier-free into the next step, a water-bridged proton transfer. The deprotonation of His199 and the protonation of the inhibitor through a water molecule occurs simultaneously and results in the product structure **W5'**.

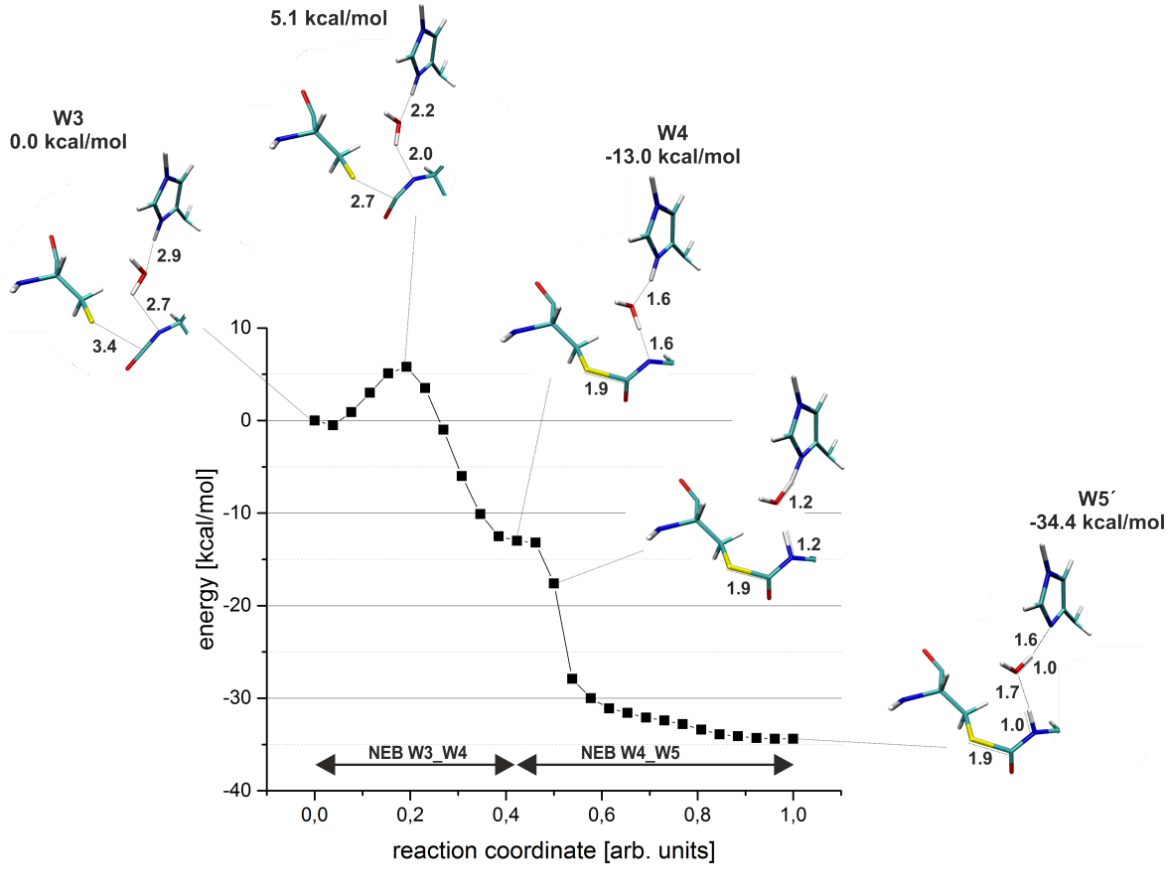


Abbildung 7.61: QM/MM NEB path (B3LYP/cc-pVDZ as QM method) of the carbamylation of Cys29 through the isocyanate intermediate. NEB profiles from **W3** to **W4** and from **W4** to **W5** (see Figure 7.52) were combined in one reaction coordinate. The x-axis therefore contains equidistant marks, one for each frame of the calculation.

The reaction enthalpy for the whole reaction is -34.4 kcal/mol. The calculations thus predict an irreversible inhibition. The significant decrease of the enthalpy for this calculation compared to the thermodynamic calculations of the model system seems to be mainly correlated to the adjacent water molecules. Beneath the stabilizing environmental water molecules around the reaction center, the water inside the active site stabilizes all reactants simultaneously.
In conclusion, the theoretical investigations showed that the inhibition mechanism of cathepsin B with carbamate-based inhibitors is a complex procedure. The protonation of the phenolate anion seems to be a crucial aspect, as well as several water-mediated proton transfer reactions. The environmental water shell and adjacent amino acids stabilize the reactive center of the system, especially concerning the carbamylation reaction. Based on the results, the examined mechanism could take place.

8 Nitroalkenes as Rhodesain Inhibitors

8.1 Background and Previous Experimental Results

This chapter focuses on nitroalkenes as potential covalent reversible inhibitors for rhodesain and provides a detailed overview of possible influencing factors on the inhibition. The following sections are structured similarly to chapter 7. A summary is placed directly after a short introduction to the topic. A detailed discussion of the investigations is given subsequently. Additionally to the Nudged Elastic Band calculations performed for the computation of certain mechanism steps, potential energy surfaces and reaction profiles were computed. This is due to the fact that while NEB calculations can provide reliable results concerning the mechanistic process, they are computationally extremely demanding and do not converge in many cases. In order to gain a detailed insight in the mechanism, both calculation types were performed.

The irreversibly binding vinylsulfone K11777[28] (Figure 8.1a) was taken as a basis for the development of an inhibitor class with a similar structure but higher binding affinities. Peptidyl nitroalkenes are potential compounds for this task, which was shown in a recent publication, testing a number of dipeptidyl nitroalkenes with rhodesain, cruzain and the cathepsins B and L.[2] The results showed that this molecule class is able to reversibly bind to proteases with high binding affinities. The structures of K11777 and a nitroalkene inhibitor are depicted in Figure 8.1a. The essential difference between the inhibitors lies in the electronic properties of the warheads. The inhibition mechanism for both compounds (Figure 8.1b) is initiated by a nucleophilic attack of the sulfur anion from deprotonated cysteine, creating a carbanion intermediate. Due to the pKa values of the corresponding acids (nitroalkanes: 17.2 pKa; sulfones: 29 pKa in DMSO),[170] the nitroalkene carbanion has a lower basicity than the one

from the vinylsulfones.[2] This could result in a less exothermic second reaction step for nitroalkenes, and might explain a reversible inhibition.

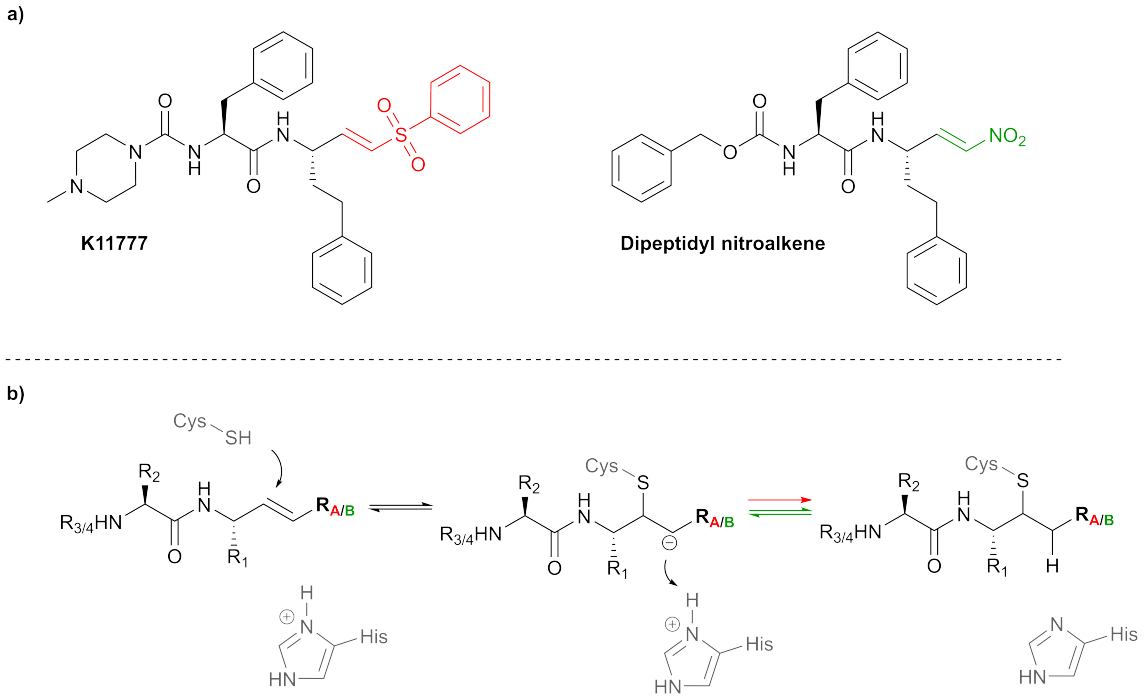
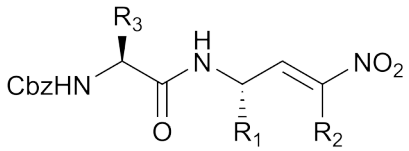


Abbildung 8.1: a) Structures of inhibitor K11777 and a dipeptidyl nitroalkene.
b) Inhibition mechanism of vinylsulfones and nitroalkenes in cysteine proteases; R_A = vinylsulfone, R_B = nitroalkene

In a collaboration with the groups of Prof. Schirmeister in Mainz[3] and Prof. González in Castelló (Spain),[179] several nitroalkene derivatives were synthesized and investigated experimentally. The compounds are shown in Figure 8.2 and 8.3 together with excerpts from recently published experimental data listed in Table 8.2.[2] The investigated compounds were compared to the irreversibly binding inhibtor K11777. Besides variations within the recognition unit of the inhibitors, the warhead was altered at the R_2 residue (see Figure 8.2). The data contains unaltered nitroalkenes (R_2 = H) methylated nitroalkenes (R_2 = CH_3) and fluorinated nitroalkenes (R_2 = F; see Figure 8.3). Dilution essays proved the reversibility of the inhibition for all compounds with the exception of K11777, which was already known to be a covalent irreversible inhibitor.



11: R_1=CH_3, R_2=H, R_3=Bn
12: R_1=*i*-Bu, R_2=H, R_3=Bn
13: R_1=Bn, R_2=H, R_3=Bn
14: R_1=CH_2CH_2Ph, R_2=H, R_3=Bn
15: R_1=CH_3, R_2=CH_3, R_3=Bn
16: R_1=CH_3, R_2=H, R_3=*i*-Bu
17: R_1=*i*-Bu, R_2=H, R_3=*i*-Bu
18: R_1=Bn, R_2=H, R_3=*i*-Bu
19: R_1=CH_2CH_2Ph, R_2=H, R_3=*i*-Bu
20: R_1=CH_3, R_2=CH_3, R_3=*i*-Bu

Abbildung 8.2: Dipeptidyl nitroalkenes that were tested as rhodesain inhibitors by the group of Prof. Schirmeister.[2]

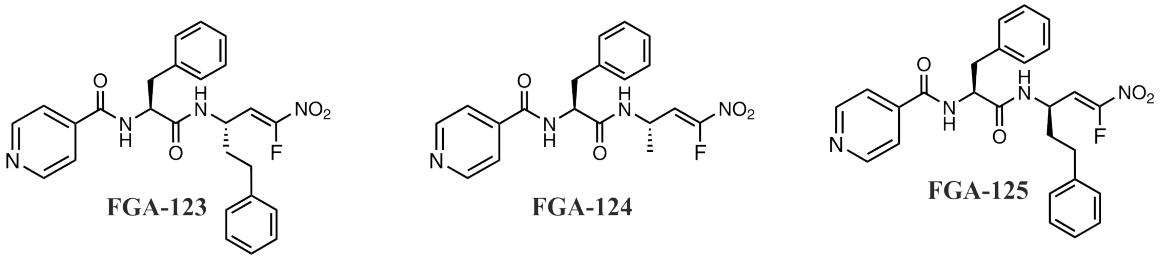


Abbildung 8.3: Structures of synthesized fluorinated nitroalkenes in the group of Prof. González.[179]

For components **11-20**, the inhibitors showed a large range in their inhibition potential. The K_i values differ significantly, depending on the peptidyl residues and the respective protease. The inhibition of the two parasitic cysteine proteases rhodesain and cruzain was more effective for all tested agents compared to the human cathepsin B and L. Nitroalkenes could therefore present an interesting molecule class for tropical diseases and parasitic infections. Compounds **15** and **20** both show very weak inhibition potentials for all proteases compared to the remaining nitroalkenes. Both components have a methyl group at the α position of the double bond. This phenomenon is even more interesting in inhibitor **11** and **15**, which only differ in this methyl residue (Figure 8.4) but showed drastic differences in their inhibition potential. While inhibitor **11** continuously has the lowest K_i values, inhibitor **15** shows poor values for the tested enzymes.

Tabelle 8.1: K_i values in [nM] for the evaluated nitroalkenes **11** to **20**, the vinylsulfone K11777 and for the fluorinated nitroalkenes **FGA123**, **FGA124** and **FGA125** (K_i^*/K_i).[180]

Inhibitor	Rhodesain	Cruzain	Cathepsin B	Cathepsin L
11	**0.49**	**0.44**	**8**	**11**
12	6.	16	68	30
13	29	130	310	110
14	4.2	11	34	19.6
15	**50**	**N.D.**	**6300**	**280**
16	30	130	310	200
17	18.4	25	210	46
18	18	28	710	49
19	14	17	72	27
20	544	N.D.	3200	2020
K11777	20	N.D.	390	80
FGA123	4.1* /16.3	-	-	-
FGA124	**2.9*/23.7**	-	-	-
FGA125	0.5*/4.9	-	-	-

N.D. = not determined; *: K_i^* value of the fluorinated nitroalkenes corresponds to K_i value of the unfluorinated compounds. K_i value of the fluorinated compounds is connected to a "sub-stepöf the reaction, measured in the experiments.

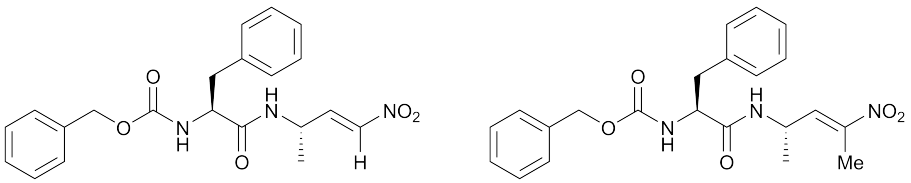


Abbildung 8.4: Structures of the dipeptidyl nitroalkene inhibitors **11** and **15**.

Concerning the fluorinated nitroalkenes, there are different kinds of K_i-values listed in Table 8.2, which can be explained by a peculiar aspect of the respective inhibition behaviour that was revealed through the experiments. For all unfluorinated nitroalkenes, one K_i value could be measured for each inhibitor. The information included in this parameter contains the formation of a non-covalent enzyme-inhibitor complex ($E^{\cdots}I$, see Figure 8.5) as well as the reaction to the covalent product $E\text{-}I$ (if it is formed; see Figure 8.5). Depending on the respective mechanism of the bond formation, this step can contain a number of reaction steps itself. In the case of the unfluorinated nitroalkenes, the whole reaction occurs so fast that only one K_i value was measured,

which represents the inhibition. This case will be referred to as an "one-step" mechanism. In case of the fluorinated nitroalkenes, two separate values K_i and K_i^* were measured. This case will be referred to as a "two-step" mechanism, assigning the K_i value to the first measured step, and the K_i^* to the whole reaction. Therefore, the K_i value of the unfluorinated nitroalkenes corresponds to the K_i^* value of the fluorinated alkenes. A "two-step" mechanism contains an element that slows down the reaction at some point. This element can be a longer residence time of the inhibitor in the active site of the enzyme in the non-covalent inhibitor-enzyme complex. In a more complex reaction mechanism where the second step (formation of the covalent product $E^{-}I$) itself consists of several steps, each of these "substeps"can cause the slowdown. This means that the K_i value of the fluorinated compounds is not necessarily assigned to the first step (shown in Figure 8.5). The respective inhibitors are called slow-binders. The K_i^* values of the fluorinated nitroalkenes are in a range between 0.5 to 4.1 nM, which is similar to the K_i values of the most effective unfluorinated compounds **11** and **14** (0.49 nM and 4.2 nM). The respective K_i value of the fluorianted nitroalkenes are in a higher range (4.9-23.7 nM).

$$\mathrm{E + I} \rightleftharpoons \mathrm{E \cdots I} \rightleftharpoons \mathrm{E\text{-}I}$$

Abbildung 8.5: Scheme of a "two-step"mechanism for a covalent reversible inhibition.

The experimental data discussed above raised questions concerning the different inhibition potentials and different behaviour of the compounds. Therefore, a computational study was performed in this work. In chapter 8.2, the unfluorinated nitroalkene **11** (see Figure 8.2) was compared to a fluorinated derivative in rhodesain in order to explain the connection between the slow-binding mechanism and the fluor substituent. The calculations focus on the mechanistic procedure of both compounds in order to find the element that slows down the reaction with the fluor-derivatives. This would thus define the part of the reaction that can be assigned to the K_i value of the "two-step" mechanism. In chapter 8.3 the inhibition potentials of the two inhibitors **11** and **15** were investigated in order to explain the quantitative difference in their inhibition behaviour.

8.2 Fluorinated Nitroalkenes

This chapter addresses the question why fluorinated nitroalkenes bind in a "two-step" mechanism to rhodesain, while unfluorinated compounds (represented by inhibitor **11**, see Figure 8.7) show an "one-step" mechanism. As mentioned, the term "one-step" mechanism includes a rapid reaction between enzyme and inhibitor. This entails the formation of the complex $E\cdots I$, as well as the possible reaction between the compounds to the product $E\text{-}I$. In the case of this mechanism, the second step consists of two stages it self. First, the sulfur of Cys25 binds in a nucleophilic attack to the β carbon of the warhead. The second step is the protonation at the α carbon through His197.

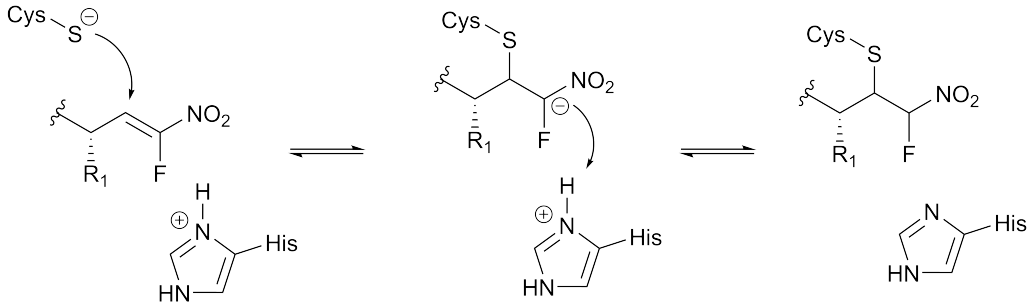


Abbildung 8.6: Inhibition mechanism of fluorinated nitroalkenes with cysteine proteases after the formation of the complex $E\cdots I$.

In case of the unfluorinated nitroalkene **11**, all steps occur rapidly and can not be measured as separate processes. However, the measured K_i value could also only include the formation of the non-covalent complex if the entire reaction stops at this point. In contrast, the mechanism of FGA123, FGA124 and FGA125 contains an element that slows the inhibition down. Thus, a "two-step"mechanism can be measured. In order to find this element, a computational study has been performed. The starting point of the theoretical investigations was a docking study of inhibitor **11** in rhodesain, performed by the group of Prof. Schirmeister. This situation refers to the formation of the non-covalent complex enzyme-inhibitor complex $E\cdots I$. The fluor atom was subsequently added into the system by hand. Hence, the two inhibitors shown in Figure 8.7 are the focus of the computations. **FGA124** and **FGA124'** differ in the functional group at the end of the recognition unit. These structural differences are not assumed to have a crucial impact on the behaviour of the compounds.

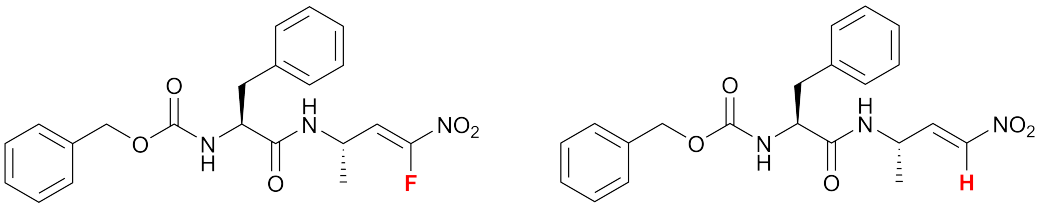


Abbildung 8.7: Comparison of the unfluorinated inhibitor **11** and the fluorinated inhibitor **FGA124'**.

8.2.1 Summary

As mentioned in the beginning of this Chapter, a summary of all relevant results is given, followed a detailed discussion of the respective topics.

Thermodynamics

In a first approach, thermodynamic calculations for several small model systems containing the warhead structures shown in Figure 8.8 as well as a methanethiol as a representative for cysteine were performed. In order to gain insight in a systematic influence of the warhead or substituent on the reaction enthalpy the investigated systems include vinylsulfones and nitroalkenes along with fluorinated and unfluorinated derivatives .

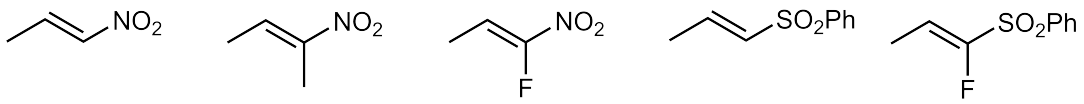


Abbildung 8.8: Structures of the investigated warheads used for thermodynamic calculations.

The calculations were performed on the SCS-MP2/aug-cc-pVDZ level of theory with the implicit solvent model COSMO (ϵ = 78.93). The comparison of the component classes showed the following order in reaction enthalpies, beginning with the most exothermic reaction:

$$fluorinated\ nitroalkenes < nitroalkenes \sim fluorinated\ vinylsulfones < vinylsulfones$$

The fluorination of the compounds leads to lower reaction enthalpies by 1 to 4 kcal/mol, depending on the respective warhead structure. In addition, the position of the formed thioester group was computed for both the α and the β position of

the double bond. The products, resulting from a nucleophilic attack at the β position, were approximately 1 to 5 kcal/mol lower in energy compared to their α derivatives. Based on the results of a bachelor thesis performed by Alexander Geis in the group of Prof. Engels, the idea of the mechanism ending after the nucleophilic attack was included in the calculations.[181] The stabilization of the resulting negatively charged adduct *via* hydrogen bonding was simulated through the protonation of the electron withdrawing group. The respective reaction enthalpies proved to be less exothermic (8 to 15 kcal/mol) compared to the other mechanisms. Although systematic differences in the reaction enthalpies, depending on the choice of the electron withdrawing group and substituent position could be revealed, there was no indication for a "two-step"mechanism within the fluorinated compounds compared to the unfluorinated systems.

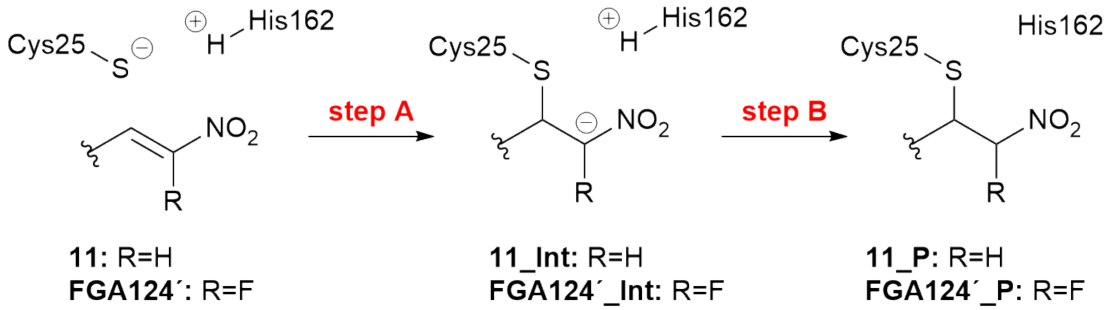


Abbildung 8.9: Reaction scheme for the investigation of the barriers and the performance of MD simulations.

Molecular Dynamic Simulations - Reactants

The behaviour of both compounds in the enzyme-inhibitor complex $E\cdots I$ was investigated with help of MD simulations. The simulations in the reactant environment did not reveal a different behavior of the two inhibitors **11** and **FGA124'**. Both inhibitors stayed in a similar position inside the active site of the environment.

Reaction Profiles - Reactants

Based on the MD simulations, reaction profiles for compound **11** were computed on a QM/MM level with different QM methods. The most promising procedure showed to be geometry optimizations with B3LYP/cc-pVDZ and subsequent SCS-MP2/aug-cc-pVDZ single point energies, which is used for all further investigations. From the structures of **11**, **11_Int** and **11_P** in the active site gained from the reaction profiles, the respective analogues **FGA124'**, **FGA124'_Int** and **FGA124'_P**

were built by the manual substitution of a the α hydrogen atom with fluor. NEB calculations of step one and two were performed for both systems in the relaxed environment of the reactant structures. The results showed a similar energy progression, with only approximately 1 kcal/mol deviation within the systems. The first barrier is around 9 - 10 kcal/mol and the second barrier is 6 - 7 kcal/mol. The energy of the fluorinated intermediate was 3 - 4 kcal/mol lower compared to **11_Int**. The results did not show a significant difference between the systems. Based on the idea that the intermediate structures could have a different residence time due to a better stabilization, MD simulations were performed for the intermediate systems.

Molecular Dynamic Simulations - Intermediates

For each system **11_Int** and **FGA124'_Int**, MD simulations were performed. The simulations revealed a different behavior of the inhibitors in the active site of the enzyme. The unfluorinated compound showed to be in a stabilized position during the simulation without significant changes. For **FGA124'_Int** in contrast, the active site's His162, which normally stabilizes the nitro compound of the warhead and is responsible for the protonation of the double bond, rotated away from the inhibitor and the active site Cys25. In this orientation, the protonation of the double bond through histidine is sterically hindered. The reason for the rotation could be linked to the reduced electron density in the warhead, caused by the fluorination. The rotation could imply a longer residence time inside the active site for the fluorinated intermediate, as either histidine has to rotate back or the protonation would have to take place from the unfavourable orientation. This behaviour is able to explain the "two-step"mechanism of the fluorinated inhibitor, as it is caused by a longer residence time of the intermediate inside the active site.

Reaction Profiles - Intermediates

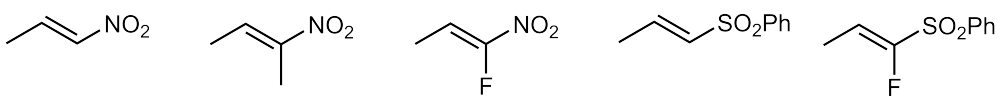
In order to quantize the energetic differences of the herein gained results, reaction profiles and NEB calculation were performed on a QM/MM level. As the starting point for the calculations are the intermediate structures, step A of the mechanism had to be calculated backwards. The profiles were computed with PM3 as QM method to obtained structures **11**, **FGA124'**, **11_P** and **FGA124'_P**. The structures were optimized with B3LYP/cc-pVDZ as QM method in order to perform NEB calculations. The NEB reaction path of compound **11** in both environments (reactant and intermediate) was compared. Logically, the intermediate structure **11_Int** was stabilized better in the intermediate environment. Additionally, the first reaction barrier is 4.3 kcal/mol lower

within the intermediate environment. As the NEB calculations for the fluorinated systems did not converge, reaction profiles were computed for both orientations of histidine with relaxed scans. For the sterically hindered system, the reaction barrier of the protonation (step B) exceeds 30 kcal/mol and would therefore not take place. The other orientation shows a barrier of 6.8 kcal/mol for the protonation, which is comparable to the unfluorinated system. In most cases during the simulations, His162 coordinated to the nitro group of the warhead.

Conclusions
The results explain the significantly different behaviour of the fluorinated intermediate system compared to **11_Int**. Due to the decreased electron density in **FGA124'_Int**, the non-covalent interactions of the warhead with the surrounding amino acids lead to a less fixed orientation of His162. The "two-step"mechanism of **FGA124'** can thus occur from longer residence time of the intermediate inside the active site, as the histidine ring has to rotate back in its original position before the protonation can take place. The NEB calculations together with the low K_i value of inhibitor **11** indicate that the inhibition reaction for this component includes both, the formation of the non-covalent enzyme-inhibitor complex and the formation of the reaction product.

8.2.2 Thermodynamics

The general influence of the fluor substituent on the electronic structure of the system was investigated by means of thermodynamic calculations. In order to use a more reliable and more computationally demanding QM method, the systems were scaled down. The inhibitors were reduced to their warhead and cysteine was represented by methanethiol. The computations did not only include fluorinated and unfluorinated nitroalkenes, but also vinylsulfones with and without a fluor substituent. A comparison of the different systems can give rise to systematical effects of the warheads and substituents. Five inhibitor warheads (shown in Figure 8.10) were included in the calculations.



wh_11 wh_15 wh_FGA124´ wh_VS wh_VS_F

Abbildung 8.10: Structures of the investigated warheads, used for thermodynamic calculations.

Based on results of Alexander Geis and Bernd Engels[181], the idea of an altered mechanism arose. Therein, the protonation in the second step would not take place. Instead, the electron density of the negatively charged intermediate, resulting from the nucleophilic attack, is stabilized through the mesomeric effect of the electron withdrawing group of the warhead and hydrogen bonding with adjacent amino acids and water molecules (see Figure 8.11). To simulate this mechanism within the framework of thermodynamic calculations, a proton was added to the NO_2-group (see Figure 8.12).

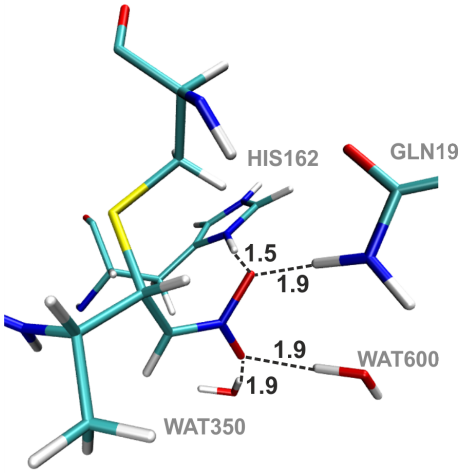


Abbildung 8.11: Structure of inhibitor **FGA124'** in rhodesain (crystal structure 2P7U[28]), taken from QM/MM calculations of Alexander Geis.[181]

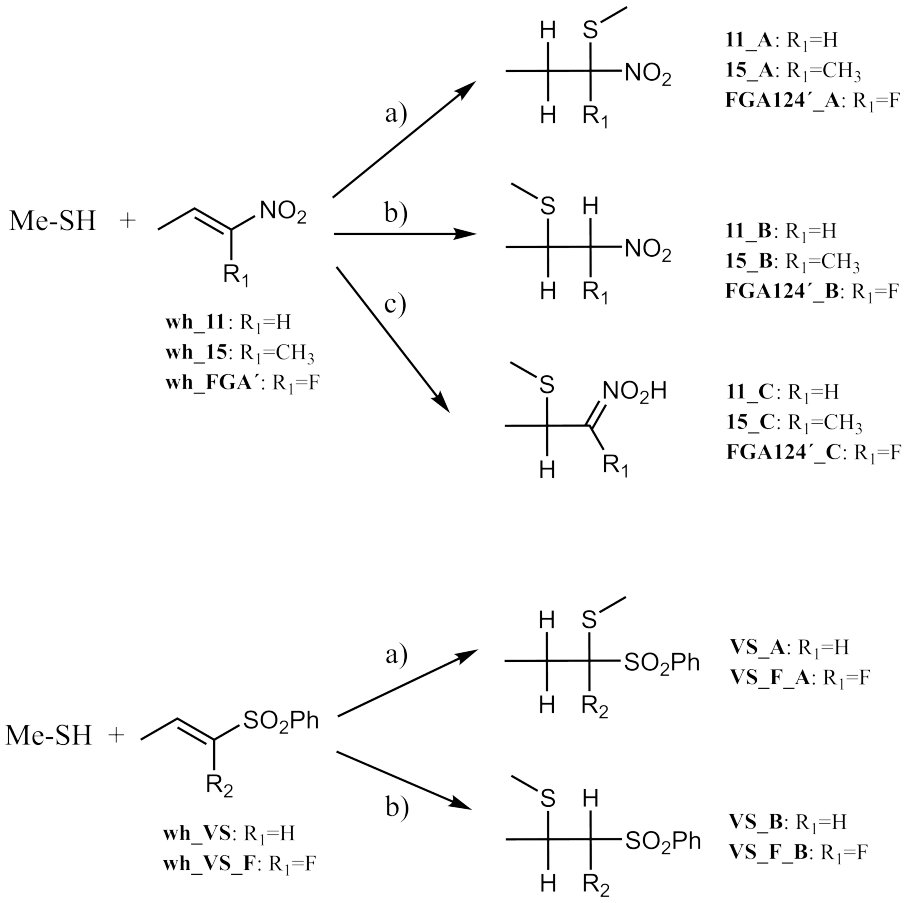


Abbildung 8.12: Reactant and product structures of the thermodynamic calculations within a model system, containing the warhead and methanethiole for a) the nucleophilic attack at the α carbon, b) the nucleophilic attack at the β carbon and c) the protonation at the nitro group of the warhead.

Tabelle 8.2: Reaction enthalpies of the model systems for the fluorinated and unfluorinated warheads. Calculations were performed on the SCS-MP2/aug-cc-pVDZ level, including an implicit solvent model to account for polarizable solvent effects; energies are given in kcal/mol.

system (reactants)	**11_A**	**11_B**	**11_C**	**15_A**	**15_B**	**15_C**
reaction enthalpy	-12.1	-15.1	-2.2	-10.4	15.0	-5.9

system (reactants)	**FGA124′_A**	**FGA124′_B**	**FGA124′_C**	**VS_A**	**VS_B**	**VS_F_A**	**VS_F_B**
reaction enthalpy	-16.3	-19.9	-2.2	-9.7	-10.1	-11.7	-10.6

The protonation of the electron withdrawing group is energetically not favoured compared to the protonation of the double bond. The respective reaction enthalpies are approximately 10 kcal/mol higher compared to the respective enthalpies of the structures without a protonated electron withdrawing group. From a chemical point of view, this outcome is unsurprising. Additionally, the proposed mechanism does not include an actual protonation of the electron withdrawing group. The systems with the protonation at α position are in all cases higher in energy (about 1 - 5 kcal/mol) than the β protonated systems. The fluorination of the warheads of inhibitor **11** and **15** lowers the reaction enthalpy approximately by 6.0 kcal/mol, except for the reaction pathways with **11_C** and **15_C**. For the vinylsulfones, the stabilization of the products by a fluorination is only around 1 kcal/mol. Compared to the respective nitroalkene systems, the vinylsulfones also have systematically higher reaction enthalpies.
Based on these thermodynamic results, the measured "two-stepmechanism of the fluor-substituted systems cannot be immediately explained. The fact that a fluor substituent lowers the reaction enthalpies would instead indicate a faster reaction due to a higher thermodynamic driving force. The concept of the negative charged intermediate being stabilized through hydrogen bonding could, however, entail a longer residence time of the covalent intermediate before the reaction continues. If the fluorinated and unfluorinated intermediates show a different behaviour in the active site of the enzyme, this could imply a difference in the inhibition and possibly explain the slow-binding mechanism of FGA123, FGA124 and FGA125. To address this issue, the systems have subsequently been enlarged and investigated in an enzymatic environment.

8.2.3 Molecular Dynamic Simulations - Reactants

The orientation and behaviour of the inhibitors **11** and **FGA124'** were investigated using MD simulations. The starting structure of inhibitor **11** was taken from docking studies from the group of Prof. Schirmeister in Mainz.[3] The structure of **FGA124'** was obtained by manual substitution of the α hydrogen with a fluor atom. Both structures were relaxed and prepared for the simulations as described in chapter 6. The simulations run over a time of 10 ns with time steps of 1 fs. A number of structural aspects including dihedral angels and distances, were compared for both inhibitors in order to investigate the orientation throughout the simulation. Neither revealed any differences of the compounds in the active site. As a representative example, the

distance between the Cys25 sulfur and the β carbon of the double bond is depicted in Figure 8.13.

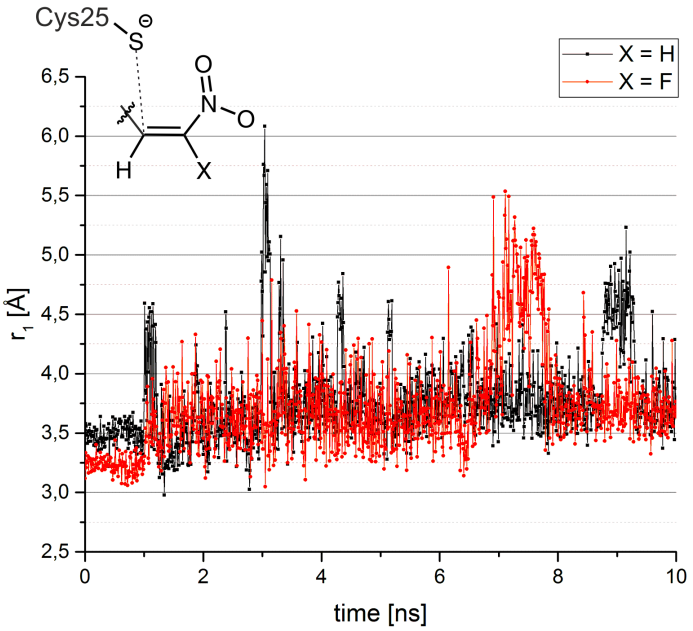


Abbildung 8.13: Distance between the active site sulfur S and the β carbon of the warheads throughout the MD simulations for inhibitor **11** and **FGA124'** in rhodesain.

As the simulations did not give a final answer to the question of the different behavior of the fluorinated and unfluorinated compound, the energetic progress of the reaction pathway is investigated as the next step.

8.2.4 Reaction Profiles - Reactants

In order to get insight into the energetic barriers of the reaction, the mechanisms were computed *via* relaxed scans and Nudged Elastic Band calculations. For the former, constrained optimizations were performed, which implies that selected degrees of freedom are to be frozen and varied for a defined set of values during the optimization. Depending on the chosen variables, different desired mechanisms can be simulated. NEB calculations automatically compute a reaction path between two structures. In order to perform this type of calculation, converged reactant and product structures are required. All calculations are performed using the QM/MM approach with different AMBER force fields. For more detailed information see chapter 6.
In the following, the procedure for the calculation of the reaction profiles and NEB calculations will be elucidated briefly. The investigation starts from a structure of com-

pound **11**, selected from MD simulations. This structure was taken for the computation of the reaction pathways of step A and B (see Figure 8.15) with different QM/MM methods in order to find the most promising method for further calculations. Based on the structures of inhibitor **11**, the intermediate **11_Int** and the product **11_P** (achieved from the relaxed scan calculations, the fluorinated analogues **FGA124'**, **FGA124'_Int** and **FGA124'_P**) were constructed by the manual substitution of the respective hydrogen atom with fluor. Unconstrained geometry optimizations were performed for all six systems in preparation of the NEB calculations. The final step is the performance of NEB calculations of step A and B for both systems. Although NEB calculations provide detailed information about the reaction mechanism of a system, the probability of convergence strongly depends on the amount of structural changes during the reaction. Therefore, both, reaction profiles and NEB calculations were performed and evaluated in this work.

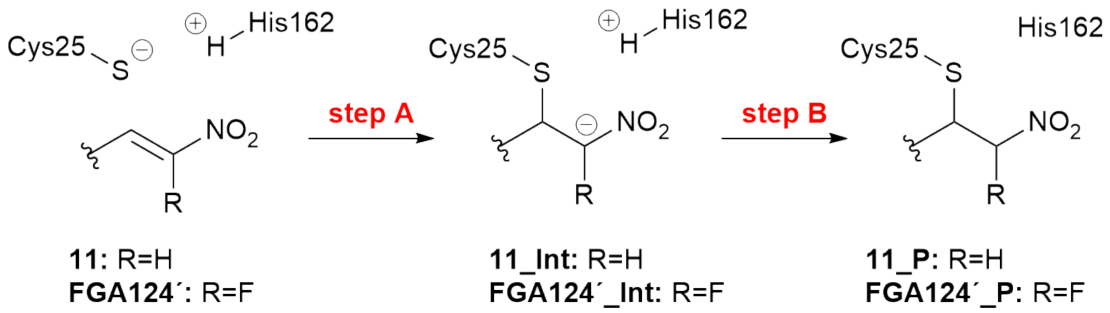


Abbildung 8.14: Reaction scheme of the mechanistic steps used for the calculations of energy profile and NEB calculations.

Reaction Profiles

Reaction profiles of the unfluorinated inhibitor **11** in rhodesain were computed. The starting structure was obtained from docking studies (performed by the group of Prof. Schirmeister). Subsequent MD Simulations were used to relax the system's environment. The structure which was selected from the simulations was optimized and subsequently varied with respect to the distance between the sulfur of Cys25 and the β carbon of the inhibitor (step A in Figure 8.14). The acquired structure of the intermediate, **11_Int**, was taken as a starting point for the simulation of the second mechanistic step. To this end, the distance between the α carbon of the double bond and the His162 hydrogen was used as reaction coordinate (step B in Figure 8.14). The resulting reaction profile combines both reaction coordinates in one axis. The QM/MM calculations

were performed using B3LYP/cc-pVDZ as QM method. QM/MM single point calculations were subsequently performed on the geometries using B3LYP/aug-cc-pVDZ and SCS-MP2/aug-cc-pDVZ as QM methods. The results are shown in Figure 8.15.

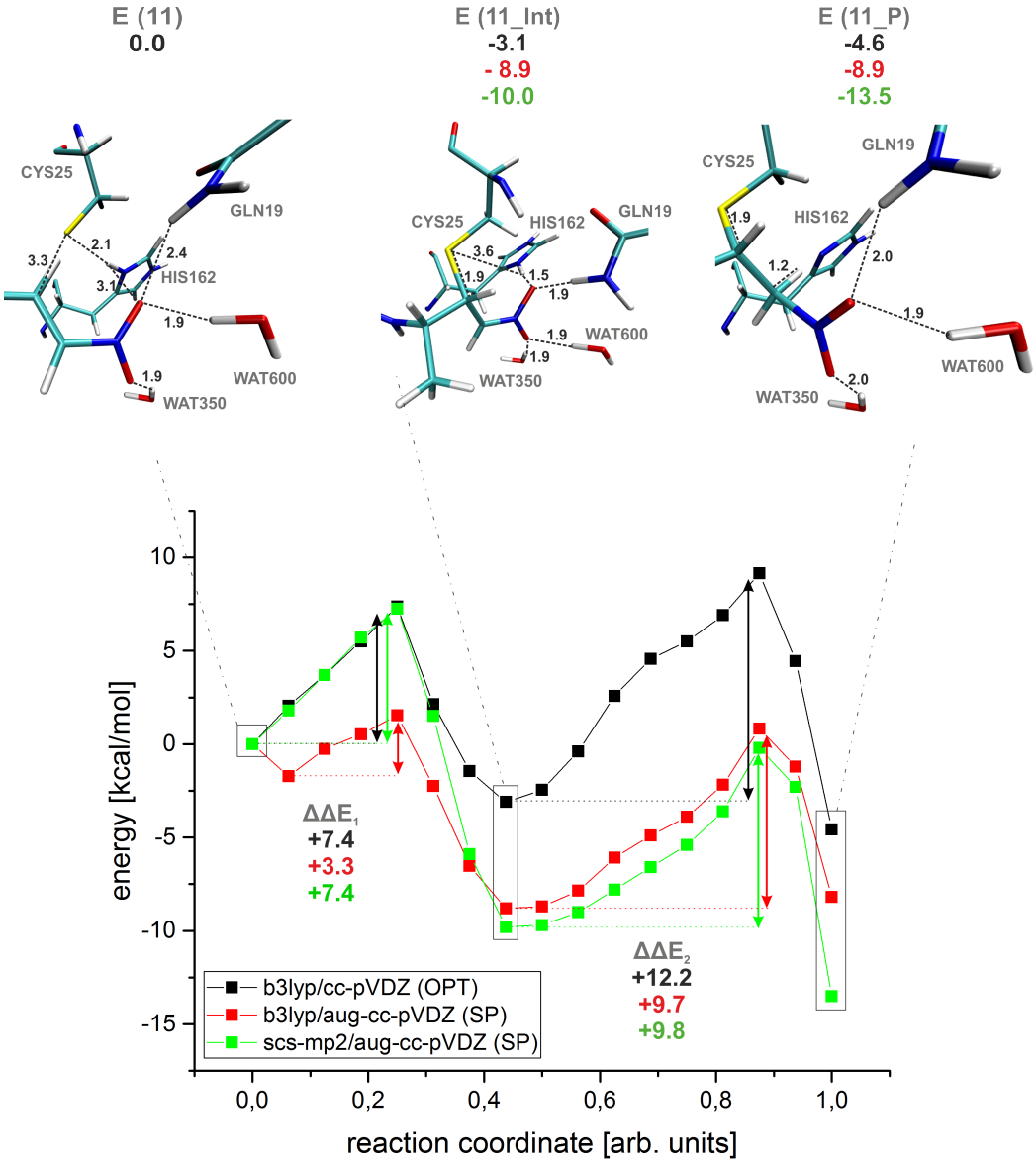


Abbildung 8.15: Reaction profile (relaxed scan) of the inhibition of rhodesaine by compound **11**. QM/MM calculations were performed with B3LYP/cc-pVDZ (optimization), B3LYP/aug-cc-pVDZ (single point) and SCS-MP2/aug-cc-pVDZ (singe point) as QM methods.

The nitro compound of the warhead continuously interacts with two water molecules as well as Gln19. The histidine ring rotates during the reaction. In the beginning, it

faces cysteine with a distance of 2.1 Å between hydrogen and sulfur (see Figure 8.15). After the nucleophilic attack at the double bond, the positively charged histidine turns towards the increased electron density at the NO_2 compound and stabilizes the negative charge along with glutamine and water. In step B of the reaction, the five-membered ring rotates back towards the double bond, where the protonation takes place. At the end of the reaction (reaction coordinate is 1.0), histidine has a orientation similar to the one in the beginning (reaction coordinate is 0.0).
The profiles shown in Figure 8.15 display differences in energies and barrier heights within the applied methods. With SCS-MP2/aug-cc-pVDZ assumed to be the most accurate method, the first reaction barrier $\Delta\Delta E_1$ is underestimated by around 4 kcal/mol within the B3LYP/aug-cc-pVDZ calculations. The second barrier only varies around 2 - 3 kcal/mol between the methods, whereas B3LYP/cc-pVDZ predicts the highest barrier. A more significant difference lies in the stabilization of the intermediate **11_Int**, where the energies of B3LYP with the non-diffuse basis set lie 5 - 6 kcal/mol above the other methods. The resulting reaction energies are in a broad range between -4.6 kcal/mol and -13.5 kcal/mol, which nevertheless allows for a covalent reversible inhibition in all cases. As the qualitative pathway of the profiles are in accordance, all further calculations follow the same pattern, where B3LYP/cc-pVDZ is taken for geometry optimizations, while the energies are calculated with SCS-MP2/aug-cc-pVDZ single point energies (B3LYP/cc-pVDZ//SCS-MP2/aug-cc-pVDZ).

NEB Calculations

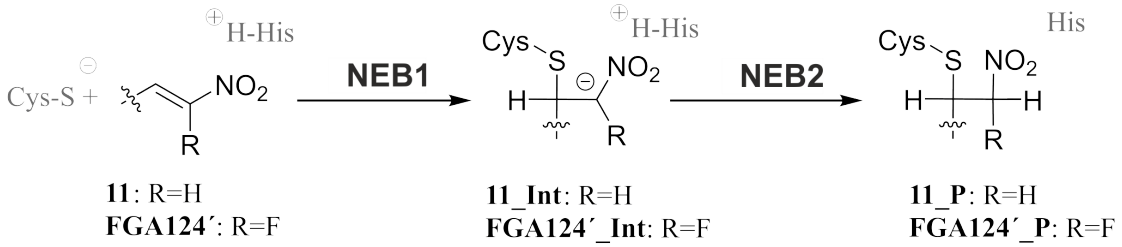


Abbildung 8.16: Reaction scheme

As mentioned above, the structures of inhibitor **11**, the intermediate **11_Int** and product **11_P** were used as a reference to construct the respective fluorinated systems **FGA124'**, **FGA124'_Int** and **FGA124'_P**. The reactant, intermediate and product structures of all systems were relaxed with unconstrained optimizations using B3LYP/cc-pVDZ as the QM method. The resulting structures were used as a starting

and end point for carrying out NEB calculations. Three calculations were performed for both systems. The first reaction step is calculated using **11** / **FGA124'** as starting and **11_Int** / **FGA124'_Int** as final structures. For the second step, the reaction starts with the intermediate structures and ends with the products **11_P** and **FGA124'_P**. Furthermore, NEB calculations of the entire reaction with the reactants as starting and products as end point were performed. The latter did non converge, as the structural changes inside the systems are too significant. The results of the individual reaction steps were combined in one normed reaction coordinate with one dot per frame for each system. They are depicted in Figure 8.17.

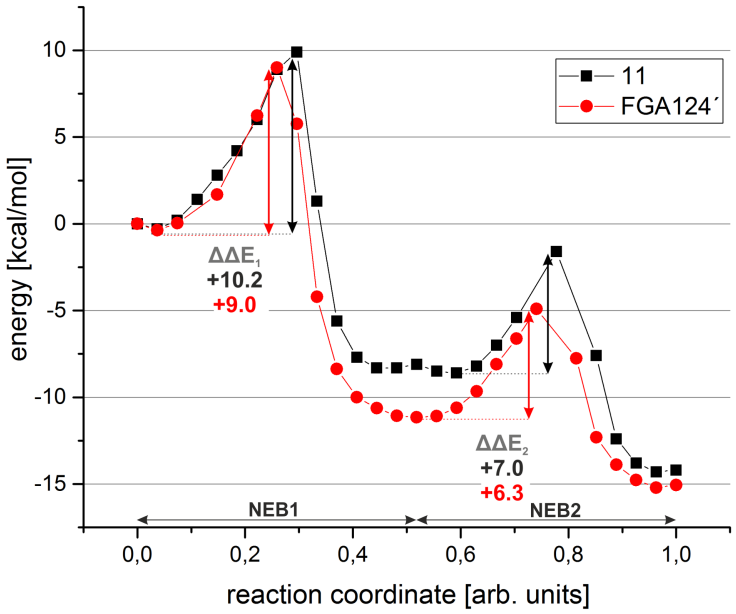


Abbildung 8.17: Resulting reaction path from NEB calculations of reaction step 1 and 2 for inhibitor **11** and **FGA124'**. The NEB calculations (QM/MM) were performed with B3LYP/cc-pVDZ as QM method. Subsequent single point calculations(QM/MM) were performed with SCS-MP2/aug-cc-pVDZ as QM method.

The results show that the fluorination of the system does not have a significant impact on the reaction barriers. They only differ by 0.7 - 0.9 kcal/mol. The intermediate structure of **FGA124'** is about 3 kcal/mol lower in energy compared to **11_Int**. The reaction enthalpies are in a similar range around -14 kcal/mol. In contrast to the computed relaxed scans (Figure 8.15), the first reaction barrier is lower than the second one (around 2 kcal/mol). The geometrical changes during the reaction are similar for

both systems. The comparison showed no significant differences. Figure 8.18 displays selected structures from the NEB calculation of inhibitor **11**. Besides the change of the coordinates involved in the reaction, the histidine ring performs a rotational movement, as already detected within the relaxed scan calculations. In the beginning, the positively charged NH group coordinates to Cys25. During the nucleophilic attack, it turns to the nitro compound of the warhead, indicating a strong interaction due to a hydrogen-oxygen distance of only 1.5 Å for structure **15**. The movement of the five-membered ring of His162 continues towards the opposite direction, back to Cys25, while a protonation of the double bond takes place. The electron withdrawing NO_2 compound is again stabilized by two water molecules and the amino acid Gln19.

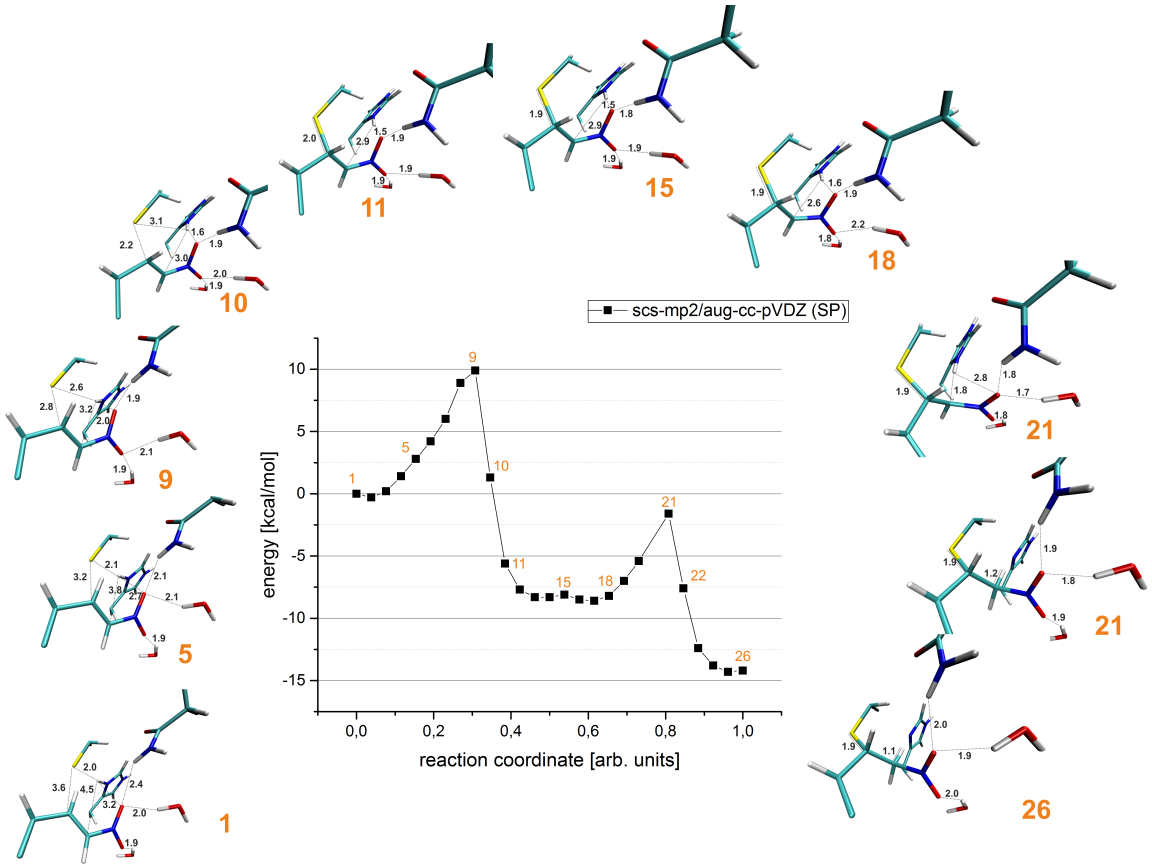


Abbildung 8.18: NEB path (QM/MM) of inhibitor **11** with exemplary structures. Calculations were performed with B3LYP/cc-pVDZ as QM method and the subsequent single point calculations with SCS-MP2/aug-cc-pVDZ as QM method.

As the structural progressions and energies are similar for both systems, the calculations do not indicate a significant difference between the inhibition potentials. Next, MD simulations of the intermediate systems **11_Int** and **FGA124'_Int** were performed.

Due to the fact that the calculation of an NEB pathway is based on one start and one final structure, a possible reorientation of the environment is neglected. In some cases, the influence of the environment can have a significant impact on the reaction energies and barriers. Based on this consideration, additional MD simulations were performed in order to relax the environment around the structures of **11_Int** and **FGA124'_Int**. In the aforementioned environment (reactant environment), the intermediates differ by up to 3.0 kcal/mol. The influence of the enzymatic environment and water molecules could enlarge this deviation or reveal a different behavior of the systems throughout the simulation.

8.2.5 Molecular Dynamic Simulations - Intermediates

MD simulations were performed with the intermediate structures inside the active site of rhodesain. The simulation time was 8 ns for each system with time steps of 1 fs. Several parameters were compared, out of which the distance between His162 and the α carbon and the dihedral between the five-membered ring and the double bond showed a significantly different behaviour between the systems. During the simulation of the fluorinated compound, His162 rotates out the active site and away from the double bond and Cys25. In this orientation, the function as a proton donor is sterically hindered.

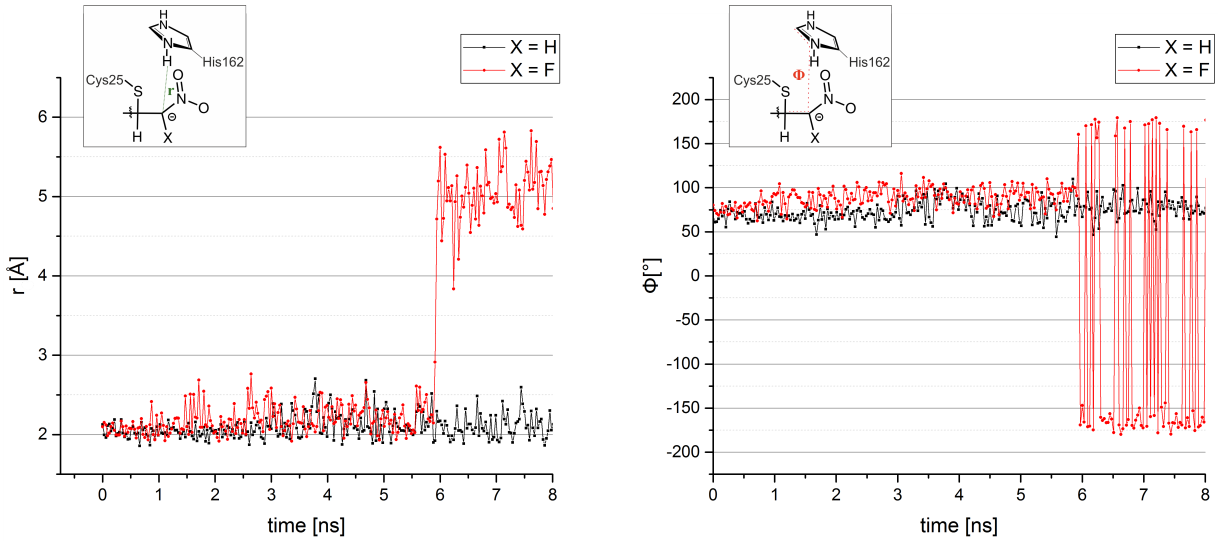


Abbildung 8.19: Dihedral angle Φ and distance r of the intermediates **11_Int** and **FGA124'_Int** during MD simulations.

Before the rotation, the inhibitors have a similar orientation in the active site of the enzyme. The distance between the α carbon and the histidine δ proton fluctuates around 2 Å for both systems. The same holds for the dihedral angle Φ between the ring and the double bond, which fluctuates by approximately 75°. The rotation of the histidine residue drastically changes these parameters for the fluorinated system.
In order to investigate the origin of this behavior, the active sites of both systems were compared, focusing on the arrangement and interactions of histidine (Figure 8.20). Within cysteine proteases, the active site's histidine is normally stabilized by the interaction with the active site cysteine and an asparagine residue *via* hydrogen bonding (see chapter 3.1.2 and Figure 8.20). As cysteine is covalently bound to the warhead, the former interaction does not take place within the intermediate structures. For **11_Int**, two major interactions contribute to the stabilization of histidine. The first one is between the ϵ amino group of histidine and asparagine. Additionally, the δ amino group of the positively charged histidine ring faces towards the nitro compound of the warhead creating another stabilizing interaction. In case of **FGA124'_inside**, the δ hydrogen of histidine does in contrast face towards the α atom of the double bond. As shown in Figure 8.19, the distance to the α carbon is, however, similar for both systems. The interaction with asparagine only takes place before the ring rotates. The inductive effect of the halogen substituent decreases the electron density of the warhead, including the NO_2 compound. As a result, the non-covalent interactions between the δ amino group of the histidine ring and the warhead are weaker compared to the unfluorinated system. This phenomenon could cause a rotation of the less stabilized ring. Concerning the experimental data, the two separately measured steps are the formation of the non-covalent enzyme-inhibitor complex together with the formation of the sulfur-carbon bond between Cys29 and the nitroalkene. The subsequent reorientation of the inhibitor causes the slowdown of the reaction.

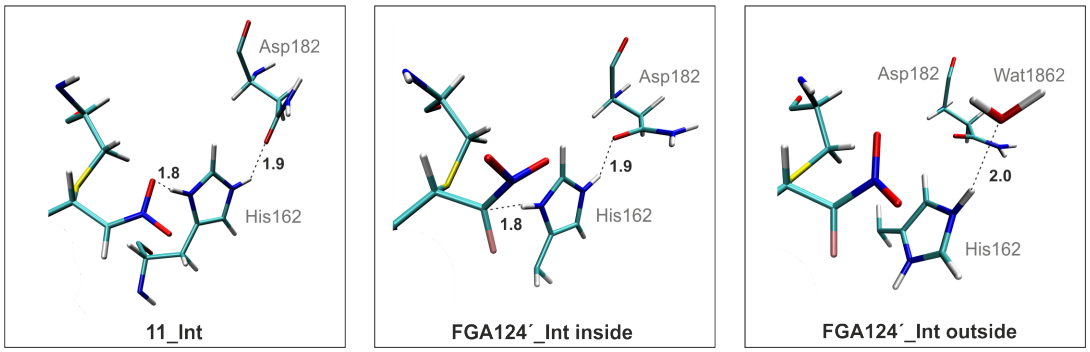


Abbildung 8.20: Exemplary structures of the orientations of intermediate **11_Int** and **FGA124'_Int** during the MD simulation. Distances are given in in Å.

8.2.6 Reaction Profiles - Intermediates

The effect on the reaction path by structural changes in the active site and the enzymatic environment was investigated with relaxed scans and NEB calculations in the relaxed environment of the intermediates. The starting structures for these investigations were taken from MD simulations as described in chapter 8.2.5. Within the simulations, containing the intermediate structures of the systems, the environment could adapt to the structural changes of the QM area (compared to the structural arrangement of the reactants). The following investigations are all based on the system containing this newly adapted environment. The starting structures are shown in Figure 8.20, whereas for the fluorinated system, two structures were investigated. In one structure, the histidine ring is in the orientation it had before the rotation and the other structure contains the ring after the rotation.
The unfluorinated system was investigated *via* NEB calculations in the newly adapted environment. Starting from the intermediate structure **11_Int** (depicted in Figure 8.20) two relaxed scans were performed on QM/MM level with PM3 as QM method. The goal thereof was to obtain the reactant and product structures of the reaction in this environment. The first reaction step (nucleophilic attack) had to be calculated backwards, as the computation started with an intermediate structure. The resulting structures of reactant **11**, intermediate **11_Int** and product **11_P** were subsequently optimized (QM/MM) with B3LYP/cc-pVDZ as the QM method to yield the starting and end structures for the subsequent NEB calculations. Two calculations were performed (one for each reaction step) and afterwards combined in one normed reaction coordinate with equidistant marks, one per frame. In Figure 8.21, the NEB pathways for inhibitor **11** inside the two relaxed environments are depicted. Compared to the reaction pathway of inhibitor **11** in the reactant environment, the reaction enthalpy is now 2.4 kcal/mol lower in energy. Better stabilization of the intermediate structure (4.3 kcal/mol) is an expected consequence of the MD simulation and the therefore relaxed environment around the intermediate structures. The second reaction barrier is similar for both environments, as is the barrier of the reaction from the intermediate to the reactant structure (backwards). The first barrier of the reaction in the intermediate environment is by 4.3 kcal/mol lower compared to its analogue in the reactant environment.

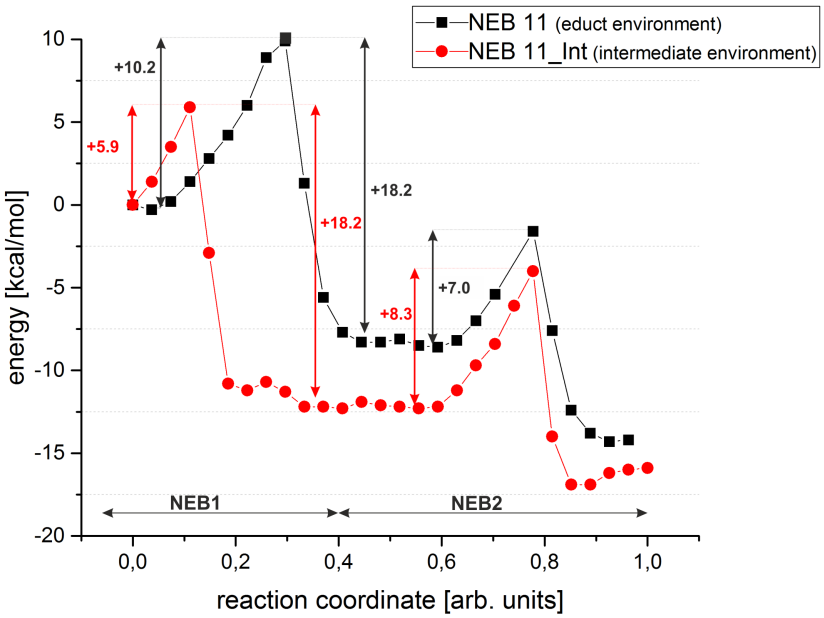


Abbildung 8.21: NEB path (QM/MM) of inhibitor **11** in two different environments adapted to the reactant structure (reactant environment) and to the intermediate structure (intermediate environment). Geometry optimizations were performed with B3LYP/cc-pVDZ as the QM method and the subsequent single point calculations utilized SCS-MP2/aug-cc-pVDZ as QM methods.

As the NEB calculations did not converge for the fluorinated system in the intermediate environment, relaxed scans of the second reaction step (protonation) were performed in order to quantize the influence of the histidine rotation on the reaction barrier. One calculation started with the histidine before the rotation (**FGA124'_Int_inside**; Figure 8.20) and one after the rotation (**FGA124'_Int_outside**; Figure 8.20).

The profile, shown in Figure 8.22, has a reaction barrier of 6.8 kcal/mol. This barrier is comparable to the one of the unfluorinated systems.

The reaction barrier for the system including a histidine which is twisted out of the active site, shown in Figure 8.23, is 32.1 kcal/mol high in energy. Due to the orientation of the ring, a protonation of the double bond is very unlikely, with regard to the sterical and energetic barriers. This result implies that it is necessary for the histidine to rotate back into its original orientation before the reaction can progress.

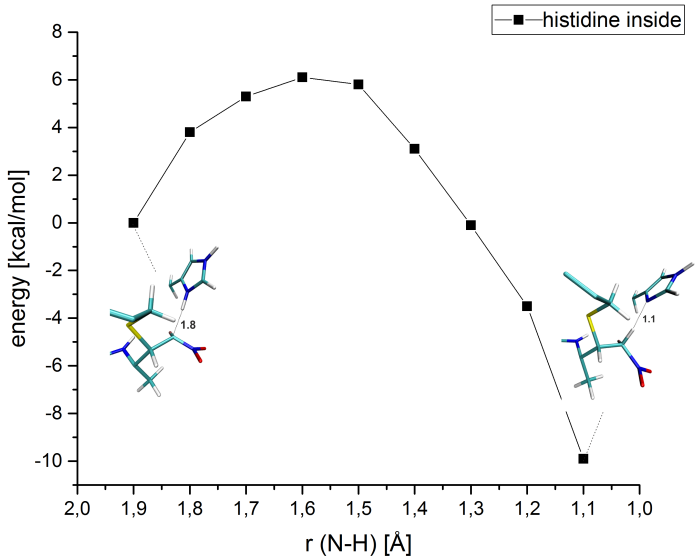


Abbildung 8.22: Reaction profile of the warhead protonation by histidine before its rotation out of the active side. Relaxed scan calculations (QM/MM) were performed with B3LYP/cc-pVDZ with subsequent SCS-MP2/aug-cc-pVDZ single point calcualtions as QM method. Distances are given in Å.

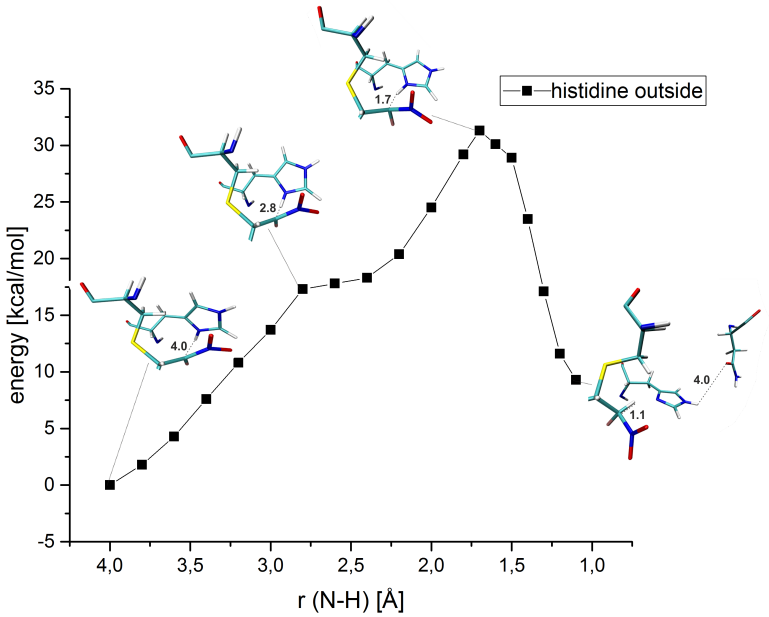


Abbildung 8.23: Reaction profile of the warhead protonation by histidine after its rotation out of the active side. Relaxed scan calculations (QM/MM) were performed with B3LYP/cc-pVDZ with subsequent SCS-MP2/aug-cc-pVDZ single point calculations as QM methods. Distances are given in Å.

Conclusion

The herein discussed results concerning fluorinated and unfluorinated nitroalkenes indicate a weaker interaction between His162 and the warhead for **FGA124'** in general, most likely caused by the reduced electron density in the halogenated warhead. This is especially the case for the intermediate structure **FGA124'_Int**, as the diffuse negative charge of Cys25 does not contribute to the stabilization of His162. This lower interaction can cause the five-membered histidine ring to rotate into a more distant orientation to the double bond and thus cause an extremely high energy barrier for the protonation. The slow binding behavior of fluorinated nitroalkenes can be explained by the necessity of the histidine ring to rotate back before the inhibition can proceed. This entails an extension of the residence time of the fluorinated intermediate structure. The comparable barrier heights of the fluorinated system, where the histidine is in a favourable position compared to the unfluorinated system, is able to explain the similar inhibition potentials, measured by the group of Prof. Schirmeister.[3]

8.2.7 Beta Fluorinated Nitroalkenes

As an additional aspect, the inhibition reaction of nitroalkenes in rhodesain was investigated with an inhibitor that was fluorinated at the β carbon of the nitroalkene double bond (see Figure 8.24).

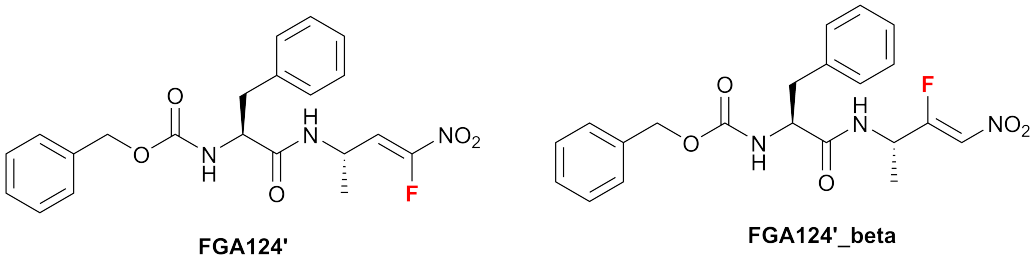


Abbildung 8.24: Structures of inhibitor **FGA124'** and **FGA124'_beta**.

The structure of the new inhibitor was obtained by the manual preparation of inhibitor **11** inside the active site of rhodesain. The system was optimized on QM/MM level using B3LYP/cc-pVDZ as QM method. The obtained relaxed structure was taken for the performance of relaxed scan calculations to create the intermediate and product structure of the system. Based on those structures, QM/MM NEB calculations were performed using B3LYP/cc-pVDZ as QM method, with subsequent SCS-MP2/aug-cc-pVDZ singe point calculations. Two NEB calculations were performed, one for the nucleophilic attack of Cys25 at the β carbon and one for the

protonation of the double bond at α position. Figure 8.25 shows the resulting NEB path, combining the two calculations in one normed reaction coordinate.

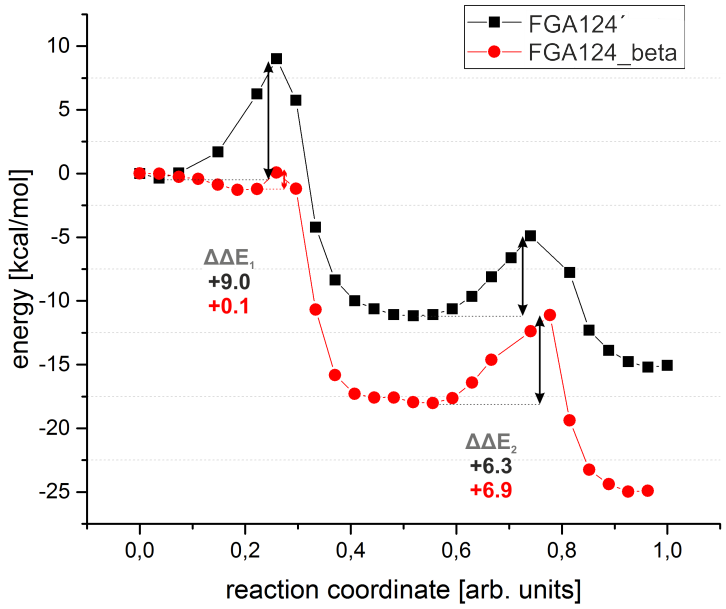


Abbildung 8.25: NEB path (QM/MM) of the α and β fluorinated inhibitors **FGA124'** and **FGA124'_beta** in rhodesain. The calculations were performed with B3LYP/cc-pVDZ and subsequent single point calculations with SCS-MP2/aug-cc-pVDZ as QM methods.

The first reaction barrier of the β fluorinated compound is nearly 0.0 kcal/mol. This result is expected, as fluor decreases the electron density at the β carbon, making it a stronger electrophile. The second barrier (8.4 kcal/mol) is in the same energetic range as the respective barriers of the unfluorinated and α fluorinated systems. In collaboration with the group of Prof. Schirmeister, the compounds are planned to be synthesized in the future.

8.3 Theoretical Studies on Inhibitor 11 and 15

8.3.1 Summary

As mentioned in the beginning of this chapter, a summary of all relevant results is given, followed by a detailed discussion of the respective topics.
The objective of the following considerations is to determine the origin of the different behaviour of the two inhibitors **11** and **15**, shown in Figure 8.26. With the help of a model system containing only the warhead and a cysteine representative (methanethiol), thermodynamic calculations were performed using DFT approaches (B3LYP/aug-cc-pVDZ). The computed reaction showed a slight energetic favour for the inhibition of compound **11**.

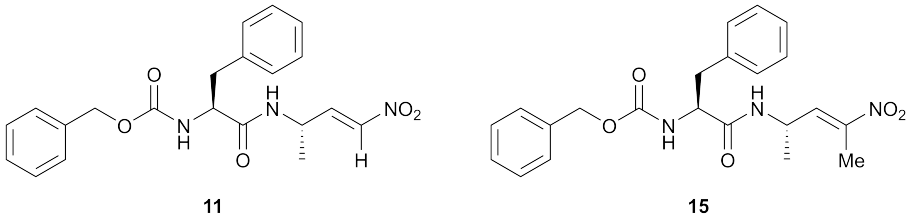


Abbildung 8.26: Structures of the dipeptidyl nitroalkene inhibitors **11** and **15**.

In a second approach, potential energy surfaces were calculated with the above mentioned model systems at the QM level of theory (B3LYP/aug-cc-pVDZ). The possibility of a histidine-independent mechanism was investigated. The nucleophilic attack and protonation were thus both assisted by the methanethiol in this mechanism. Both reaction barriers exceed 40 kcal/mol. The reactions would thus not take place under the given conditions. As the comparison of the electronic properties of the warheads did not yield sufficient answers to the question of the different inhibition behavior of the components, the orientation of the inhibitors inside the active site of rhodesain were compared in a second step. The group of Prof. Schirmeister from Mainz performed covalent and non-covalent docking for both inhibitors in rhodesain (crystal structure 2P7U[28]). All investigated docking poses had the same orientation of the inhibitors, with the warhead being placed near the active site's dyade. In most cases, the warhead is stabilized by hydrogen bonding with the two adjacent amino acids Trp184 and Gln19. The peptidyl backbones of all inhibitors lie in the active site cleft, also stabilized *via* hydrogen bonds through amino acids Asp161 and Gly66. Although slight differences appeared within the studies and inhibitor **11** is again favoured in the orientation and interaction contributions, no significant connection could be drawn to

the methyl group at the α position of inhibitor **15**. In a third attempt, MD simulations were performed for both systems in order to compare the behaviour and stabilization of the inhibitors over a certain period of time. Ten simulations are performed for each system, with a simulation time of 8 ns each and time steps of 1 fs. These trajectories showed significant differences between the inhibitors. The distance between the β carbon of the warhead and the sulfur of the active site's cysteine was compared for all MD runs and both systems. While for compound **11** this parameter stays in a range smaller than 5Å for seven of the ten simulations, compound **15** exceeds this range in seven of ten runs. The origin of this behaviour of compound **15** is assumed to result from the steric arrangement of the methyl group at the α carbon relative to the second methyl residue of the inhibitor. In the beginning of the simulations, the two residues face in the same direction (similar to a *"syn"* orientation). This orientation is sterically not favoured. A comparison of the dihedral angle between the residues and the distances between Cys25 and the double bond clearly showed that if the methyl groups rotate towards an "*anti*"position, the warhead is twisted out of the active site. As this steric aspect does not hold for component **11**, this explains the difference between compound **11** and **15**.

8.3.2 Thermodynamics

In a first attempt, the electronic properties of the warheads from both inhibitors were compared within a model system. The system consists of the respective warhead and double bond (see Figure 8.27), as well as methanethiol as a model representing cysteine. Thermodynamic QM calculations with B3LYP/aug-cc-pVDZ were performed in order to quantify the differences between the attack of the sulfur atom at both carbon atoms of the double bond and the following protonation. The continuous solvation model COSMO (ϵ=78.39) was additionally included in order to include possible polarization effects of the environment. The usage of a diffuse basis set is needed to describe the involved sulfur atom satisfactorily. The results of these calculations are depicted in Figure 8.28.

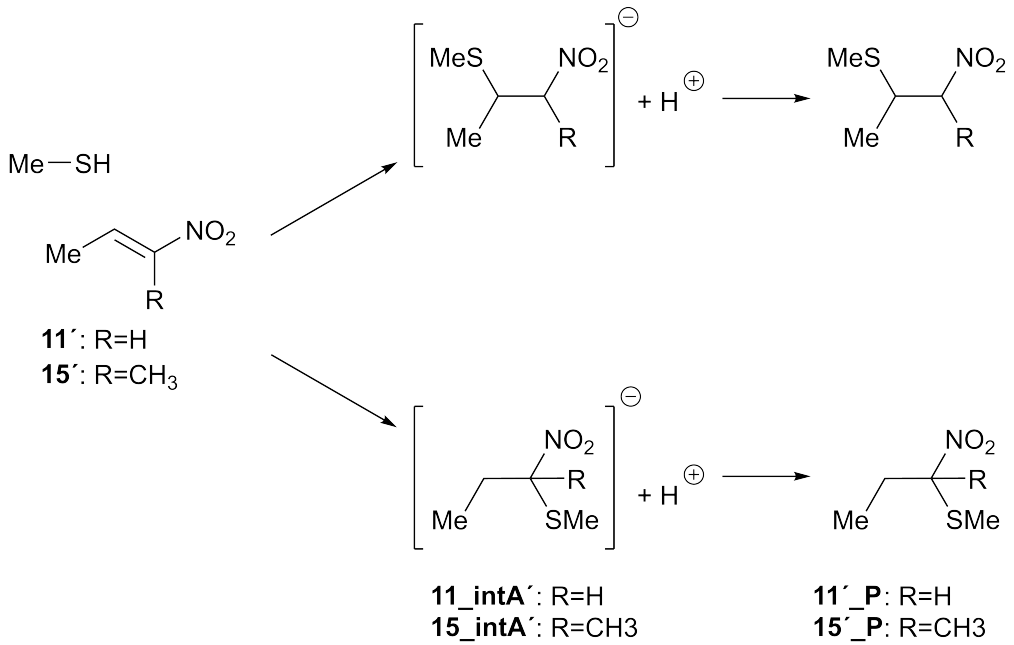


Abbildung 8.27: Inhibition reaction of the model system representing the reaction of two nitroalkene inhibitors with the active site's cysteine of rhodesain.

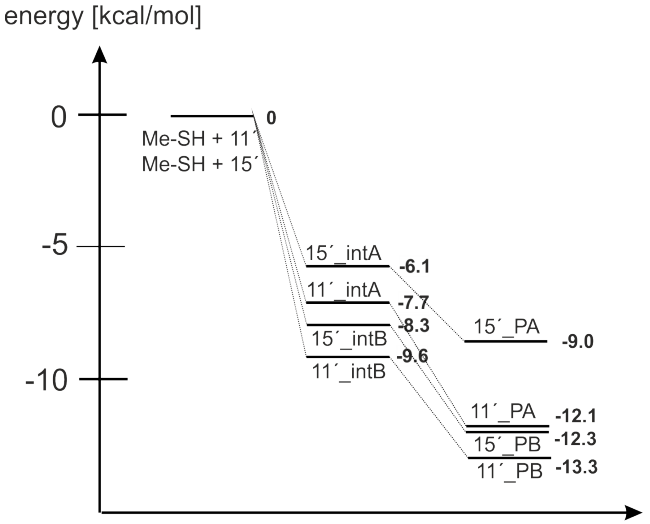


Abbildung 8.28: Results for the thermodynamic calculations of the reaction between the warheads **11'** and **15'** and the cysteine representative methanethiol. The calculations were performed with B3LYP/aug-cc-pVDZ and COSMO (ϵ = 78.39). Energies are given in kcal/mol.

The nucleophilic attack at the β carbon atom is more favourable for both compounds compared to the corresponding reaction at the α carbon (2.1 kcal/mol for **11'** and 2.2 kcal/mol for **15'**). Additionally, the reaction enthalpies for compound 11 are energetically favoured. However, the electronic properties of the warheads do not seem

to be the origin of the different performance of the inhibitors, at least when they are computed in an isolated, non-enzymatic environment.

8.3.3 Potential Energy Surfaces of an Alternative Pathway

The thermodynamic results of the model systems did not reveal significant differences between the two inhibitors. In order to get insight into the reaction barriers, potential energy surfaces (PES) of the model systems were performed using DFT (B3LYP/aug-cc-pVDZ) with an implicit solvation model (COSMO; $\epsilon = 78.39$). The idea of a histidine independent mechanism was investigated by including only a methanethiol as representative for the protonated cysteine residue and the inhibitor warheads (structure **11'** and **15'** in Figure 8.27). In this mechanism, the nucleophilic attack and protonation are both performed by the methanethiol.
The barriers for this mechanism are 31.6 kcal/mol for **11'** and 30.6 kcal/mol for **15'**. This result clearly shows that a *syn*-addition of a protonated cysteine residue will not take place. Additionally, the two systems show a very similar behaviour. As the pKa value of the thiol group can be influenced by the enzymatic environment, the barriers could decrease, if the whole system is included in the calculations. However, the cysteine side chain is a weak acid and the energies of the barriers are in an energetic area, implying the enlargement of the system to not be promising. The participation of histidine as a proton donor seems to be a crucial part of the mechanism. For both systems, an intermediate (**Int 11** and **Int 15** in Figure 8.29 and 8.30) was formed during the reaction through the protonation of the NO_2 group. In the enzymatic environment, this electron withdrawing group is stabilized *via* hydrogen bonding and this intermediate is unlikely to play a crucial role in the inhibition mechanism. Regarding the small influence of the methyl group on the electronic structure of inhibitor **15**, however, a more detailed investigation of the energetic barriers are not in the scope of this work. Instead, steric effects and the interaction of the inhibitors with the enzymatic environment are compared for both inhibitors.

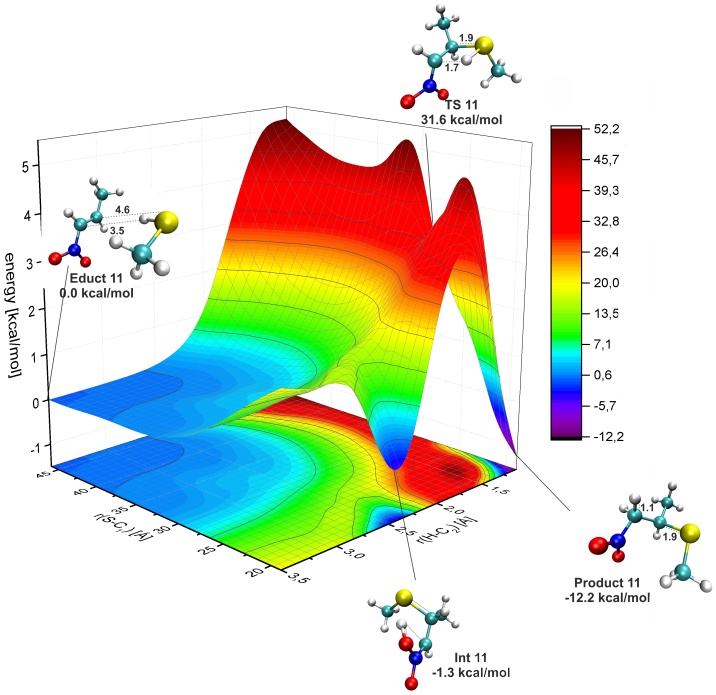


Abbildung 8.29: Potential energy surface of the model system with methanethiol as nucleophile and proton donor for inhibitor **11'**.

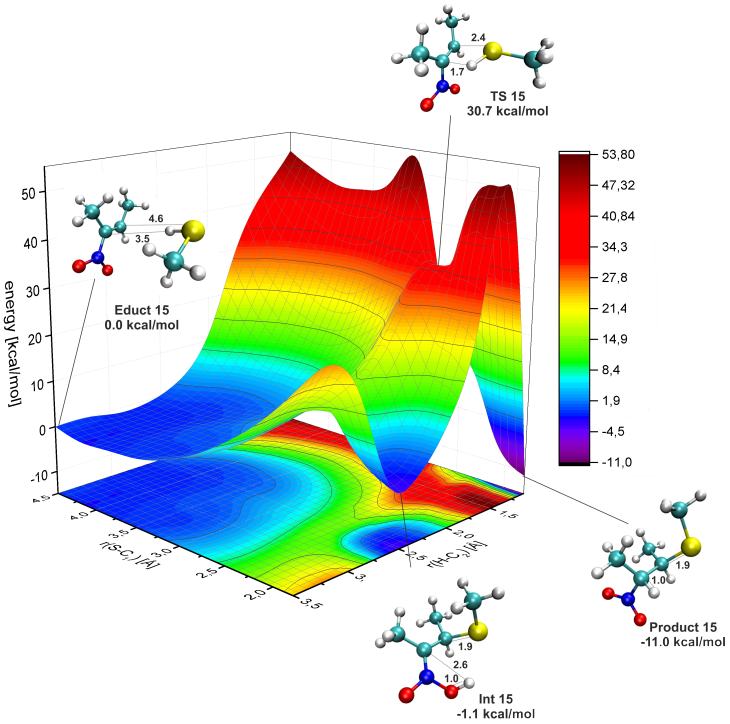


Abbildung 8.30: Potential energy surface of the model system with methanethiol as nucleophile and proton donor for inhibitor **15'**.

8.3.4 Docking Studies

The similar electronic properties of the warheads are not unusual, keeping their small structural difference in mind. The different behaviour of the compounds could instead result from a different orientation inside the active site of the enzyme. Therefore, docking studies, performed by the group of Prof. Schirmeister in Mainz, need to be evaluated and compared.[3] Compound **11** and **15** were docked covalently and non-covalently against rhodesain (crystal structure 2P7U[28]). The three structures with the highest ratings of each docking study were analyzed. Figure 8.31 shows a superimposition of the best non-covalent docking structures of **11** and **15** in the active site of rhodesain.

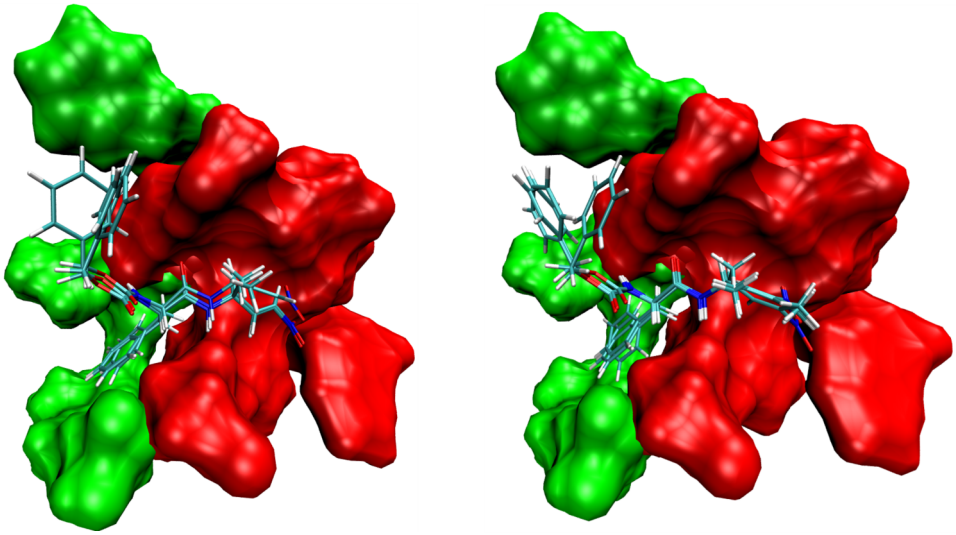


Abbildung 8.31: Results from the non-covalent docking of inhibitor **11** (left) and **15** (right) in rhodesain (crystal structure 2P7U[28]).

The enzymatic environment around the inhibitor (within a radius of 5 Å) is plotted as a surface. The red color represents residues, which form polar non-covalent interactions with the inhibitor, the green shaped surface are non-polar residues. Both inhibitors have the same orientation in all docking poses. The warhead and the recognition unit of the inhibitors lie inside the active site cleft and are stabilized by polar interactions. In Figure 8.32, the respective interacting amino acids are explicitly shown. The aromatic ring from the benzyl carbamate at the back end of the inhibitors is not positioned rigidly in the hydrophilic pocket. Its orientation differs within the different docking structures. However, this does not have an impact on the settled orientation of the inhibitor backbone and the position of the warhead. In Table 8.4, the polar non-covalent interactions of both inhibitors of the non-covalent docking study

are summarized and compared to those of K11777. The contributions are separated into polar interactions between enzyme and warhead (1 - 3) and polar interactions with the inhibitor backbone (4 - 6). All interactions are identical within the investigated inhibitors, with the exception of a hydrogen bond between the amino group of the backbone of Cys25 and the NO_2 group of inhibitor **11** and **15**, that only occurs in three of the docked structures.

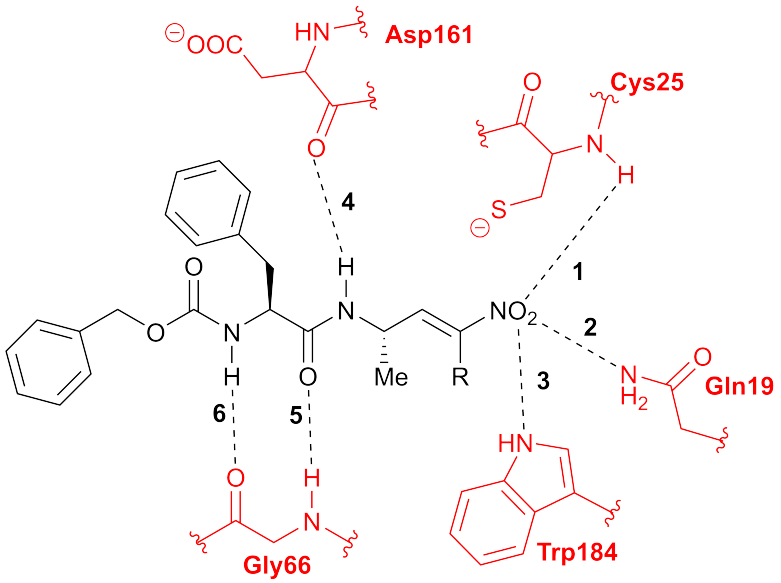


Abbildung 8.32: Scheme of possible non-covalent polar interactions between the inhibitors (**11**: R = H; **15**: R = Me) and the active site of rhodesain.

These structures, namely **11_1**, **15_1** and **15_2** show slight rotations of the warhead towards the cysteine backbone (see Figure 8.32). Through the rotation, the double bond of the warhead is pushed away from the active site Cys25. This orientation is unfavourable for the attack of the sulfur at the double bond. However, the resulting difference in the sulfur-β carbon distance of the two orientations is only about 1 Å. Furthermore, there are docking structures for both inhibitors, showing more favourable orientations. Thus, this issue could have only a small impact on the inhibition.

Tabelle 8.3: Contributing non-covalent polar interactions for non-covalent docking poses. Assignment of the interaction is referring to Figure 8.32.

	1	2	3	4	5	6
K11777		x	x	x	x	x
11_1	x	x	x	x	x	x
11_2		x	x	x	x	x
11_3		x	x	x	x	x
15_1	x	x	x	x	x	x
15_2	x	x	x	x	x	x
15_3		x	x	x	x	x

Another interesting aspect to be investigated is the orientation of both inhibitors, after the inhibition has taken place. For this purpose, covalent docking structures of compound **11** and **15** in rhodesain are analyzed. Through the covalent bond between Cys25 and the double bond, the inhibitors were closer to the active site, and steric effects gain impact on the orientation. Figure 8.33 shows the overlaid covalent docking structures of both inhibitors.

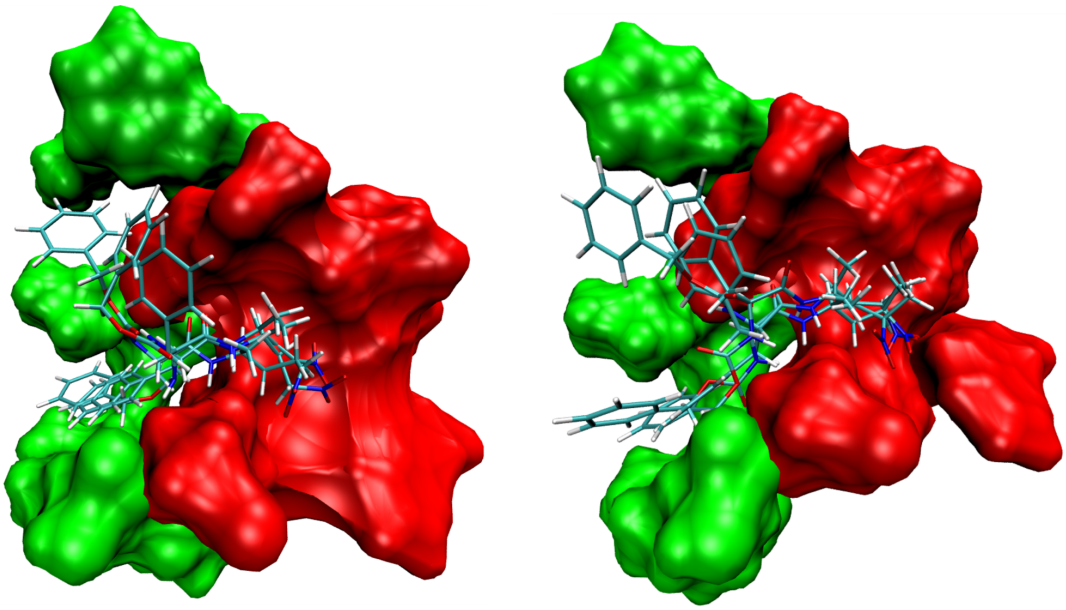


Abbildung 8.33: Results from the covalent docking of inhibitor **11** (right) and **15** (left) in rhodesain (crystal structure 2P7U[28]).

The general orientation of all inhibitors was similar. The recognition unit lies inside a network of polar interactions with the enzyme and the warhead is faced in the same direction. Still, in the covalently docked structures, more differences between

11 and **15** are found than within the non-covalently docked structures. The back end of the compounds is twisted in some cases, resulting in interchanging positions of the aromatic rings. The concerned structures lack interaction 6 (see Table 8.4), which is the interaction of the hydrogen bond between the amino group of the inhibitor backbone with Gly66 (see Figure 8.33). This orientation interchange does, however, take place for both inhibitors and seems to be independent of the methyl group at the α position of the double bond. On one of the structures of inhibitor **15** (**15_3**), the whole backbone of the molecule is rotated in a way that renders any polar interaction with the enzyme unfeasible. Regarding the warhead, three different orientations emerge. The first equals the orientation of the warhead from K11777, forming hydrogen bonds with the amino acids Gln19 and Trp184. This orientation is close to the non-covalently docked results and holds for the structures **11_2**, **15_2** and **15_3** (see Table 8.4). The second orientation includes an additional hydrogen bond between the warhead and the backbone of the active site Cys25. This orientation occurs in structure **11_1** and is known from the non-covalent docking poses. In this case, the double bond of the warhead is slightly inclined away from the active site. The third orientation is the most unfavourable one, as the warhead and double bond are twisted out of the active site, and thus no stabilization can take place. Again, this orientation occurs for both inhibitors (**11_3** and **15_1**) and thus the methyl group of compound **15** at the α position has no influence on it.

Tabelle 8.4: Contributing covalent polar interactions for non-covalent docking poses. Assignment of the interaction is referring to Figure 8.32.

	1	2	3	4	5	6
K11777		x	x	x	x	x
11_1	x	x	x	x	x	x
11_2		x	x	x	x	
11_3				x	x	
15_1				x	x	x
15_2		x	x	x	x	
15_3		x	x			

Although inhibitor **11** shows statistically slightly better results in all aspects discussed above, no significant differences can be observed within the docking structures of the compounds. The orientation of the inhibitors inside the active site does hence not seem to have an influence on the different inhibition potentials. The focus was therefore set

on another aspect, namely the residence time of the inhibitors in the active site. This behaviour was investigated *via* MD simulations.

8.3.5 Molecular Dynamic Simulations

For each inhibitor, ten simulations were computed, each running over a time of 8 ns with time steps of 1 fs. The starting structure of the simulations was taken from the docking study of the non-covalently docked inhibitors. The systems were prepared as described in chapter 6. As a first descriptor, the distance between the active site cysteine sulfur and the β-carbon C_2 of the double bond was compared for all MD runs. The simulations (see Figure 8.34) revealed a differing behaviour of the inhibitors. For compound **11**, the distance between S and C_2 increases to several Å (>> 5Å) in three of the ten simulations. Therefore, the warhead is placed outside of the reaction range of the active site amino acid. For compound **15**, this behaviour is found in seven out of ten times. To get a more detailed insight into this phenomenon, further descriptors have been investigated. The polar interactions between the inhibitor and the active site amino acids (see Figure 8.32), were compared for all simulations.

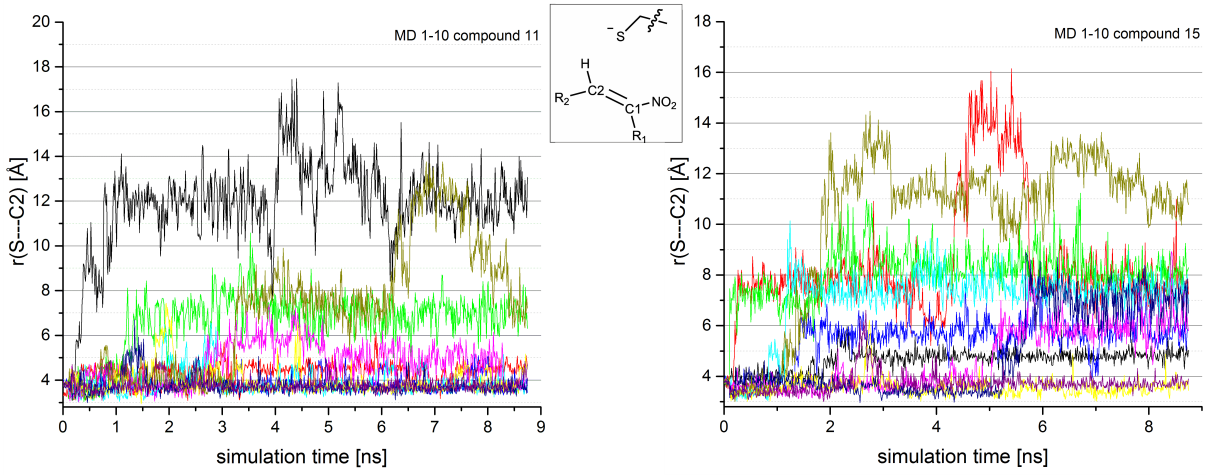


Abbildung 8.34: Distance between the active site sulfur S and the β carbon C_2 of the warhead throughout all MD simulations for both inhibitors.

The results do not reveal the origin of the different behaviour of the inhibitors, but show that they never completely leave the active site cleft. The rear part of the inhibitor's backbone stays inside the active site, while in some cases the frontal part, including the warhead, rotates out. A closer investigation of the warhead orienta-

tions leads to the observation of a peculiar behaviour of inhibitor **15** throughout all simulations.

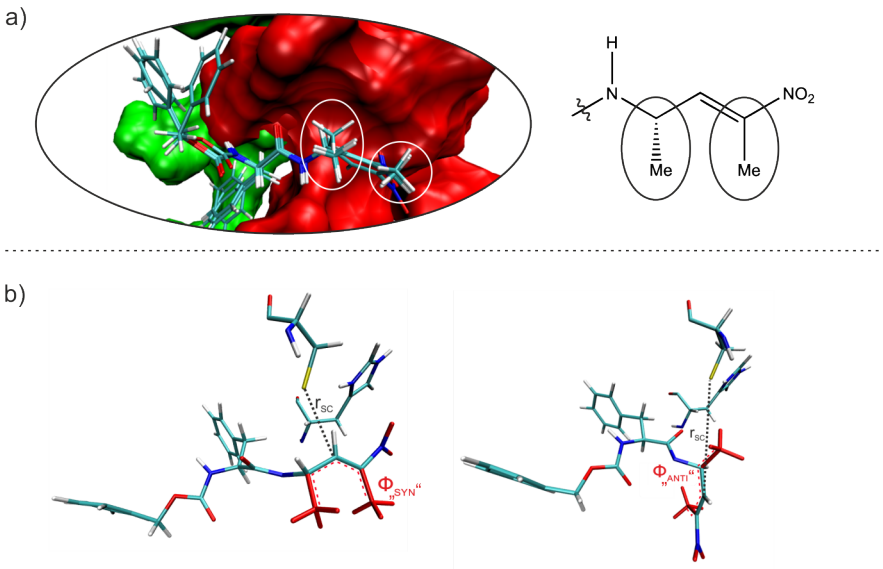


Abbildung 8.35: a) Orientation of the methyl groups of inhibitor **15** for the three best non-covalent docking poses; b) Two different orientations of the methyl groups of inhibitors **15**, occurring throughout the MD simulations.

In the non-covalent docking poses of **15**, the orientation of the methyl group of the inhibitor backbone relative to the methyl group at the α position matched for all poses (see Figure 8.35a). The two residues point in the same direction. This orientation will be called *syn*-orientation and is sterically not favoured. In each of the simulations, in which the warhead leaves the active side cleft (r(S-C_2) > 5Å), the two methyl groups rotate apart from each other. This results in a sterically more favourable "*anti*orientation, which entails a twist of the warhead out of the active site. In order to validate this finding, two variables were examined for all MD runs. The first one is the distance between the β carbon of the double bond and the sulfur of the active site Cys (r(S-C_2)). The second descriptor is the dihedral angle Φ between the two methyl groups. The results show a clear dependence of the descriptors for component **15**. Figure 8.36 illustrates the values for two MD simulations, with both variables of inhibitor **15** staying below 5 Å . The corresponding dihedral angles stay in a steady range around 60°. Figure 8.37, on the other hand, contains all data for the simulations, in which the inhibitor rotates out the active site. For the first ns, the inhibitor stays very close the active site Cys25 (r(S-C_2) < 4.0 Å). During this period, the angle Φ is again in a range around 60°. As soon as the distance increases, the

two methyl groups rotate away from each other. In some simulations, the dihedral settles at a value around -70°, in other cases it takes values around 180°. The effect of the rotation always entails a drastic increase of the intermolecular distance between the warhead and Cys25. The question whether the rotation of the methyl residues initiates the increase of the distance or *vice versa*, is difficult to answer. The change in both variables happens nearly at the same time. A plausible assumption could be that the small fluctuations of the intermolecular distance at some point give enough space to initiate the rotation, which then increases the distance, once again.

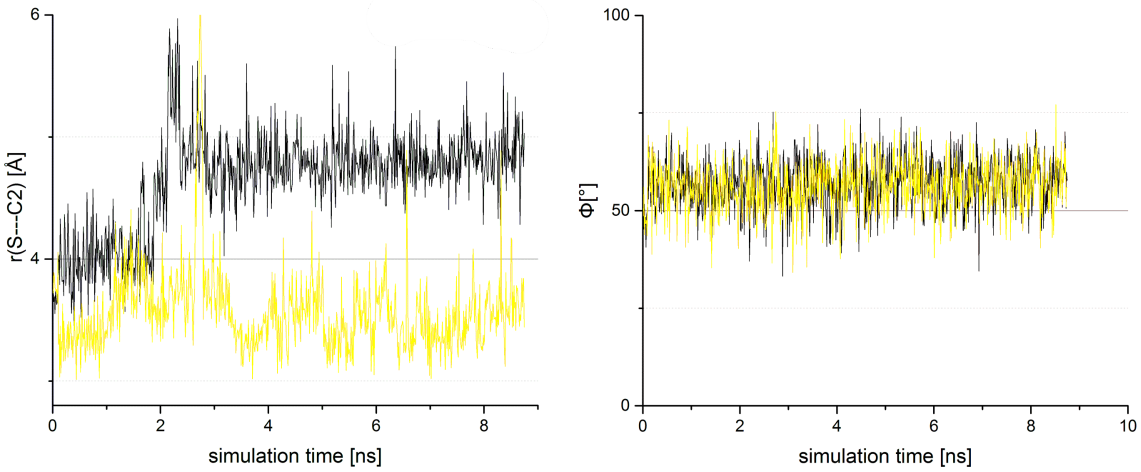


Abbildung 8.36: The two variables r(S-C_2) and Φ (defined in Figure 8.35) during several MD simulations, in which the warhead of compound **15** moved out of the active site of rhodesain.

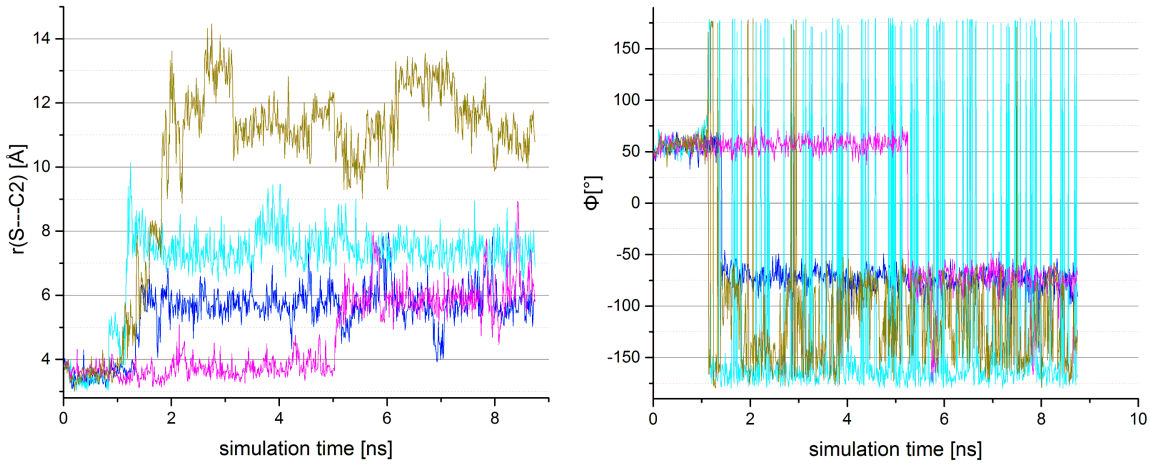


Abbildung 8.37: The two variables r(S-C_2) and Φ (defined in Figure 8.35) during several MD simulations, in which the warhead of compound **15** stayed inside the active site of rhodesain.

For inhibitor **11**, no steric hindrance is provided as its hydrogen atom at the β position is significantly smaller than a methyl group. Hence, the above observed dependency between the distance r(S-C_2) and Φ does not occur during the respective MD simulations. Figure 8.38 shows exemplary simulations of inhibitor **11**, in which the warhead rotates out the active site. The dihedral angel Φ is clearly independent to the intermolecular distance between sulfur and the double bond.

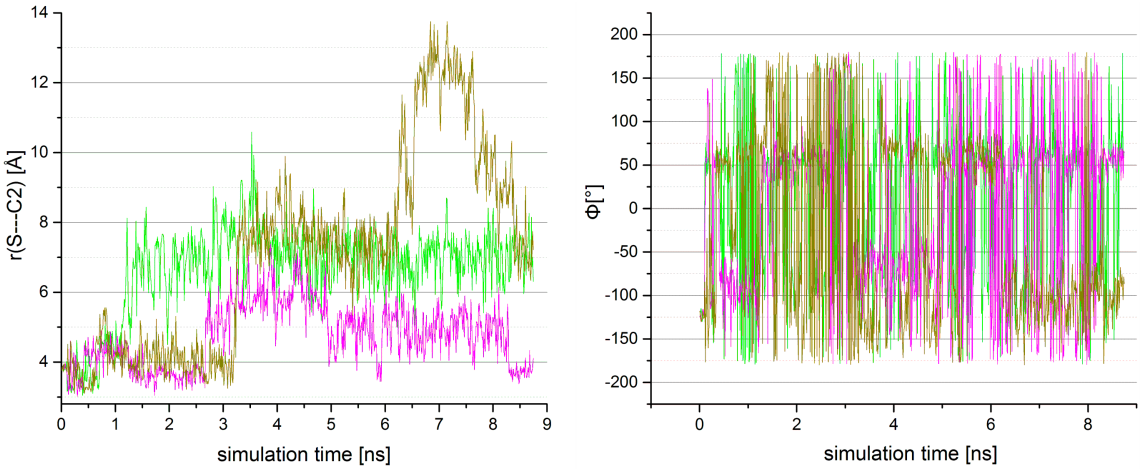


Abbildung 8.38: The two variables r(S-C_2) and Φ (defined in Figure 8.35) during all performed MD simulations of compound **11** in rhodesain.

Conclusion

The inhibition potential of the two nitroalkenes **11** and **15** in rhodesain were investigated *via* different theoretical approaches. MD simulations of the two inhibitors in rhodesain showed that the warhead of inhibitor **15** is twisted out of the active site due to sterical hinderance of two adjacent methyl groups of the inhibitor. The warhead of inhibitor **11**, in contrast, stays in a stable position inside the active site during the simulation. For this inhibitor, no sterical hinderance was found, as it differs from inhibitor **15** by a hydrogen atom (instead of a methyl group). The analysis of the MD simulations provided an explanation for the experimental results, which showed a higher inhibition potential for inhibitor **11**, compared to inhibitor **15**.[3]

9 Appendix

Experimental Data of the Group of Prof. Guetschow[1]

Tabelle 9.1: Experimental data concerning the inhibition efficiency of the four carbamate-based inhibitors in Cathepsin B shown in Figure 9.1 (from the working group of Prof. Guetschow). Measured data is given for an inhibitor concentration of [I] = 20 μM. The term k_{inac}/K_i is a measure for the overall inhibition efficiency, based on the scheme in Figure 9.2.

[H] Inhibitor	enzyme activity	k_{inac}/K_i $[M^{-1}s^{-1}]$
1	2%	745
2	4%	711
3	85%	-
4	98%	-

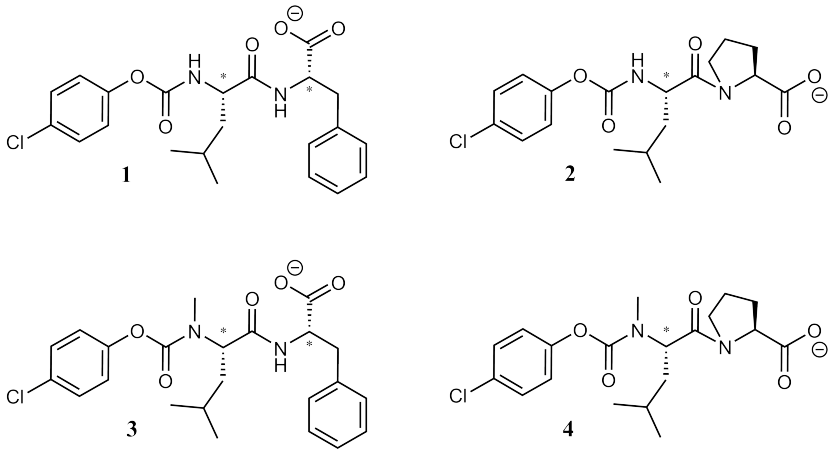


Abbildung 9.1: Investigated compounds from the working group of Prof. Guetschow.

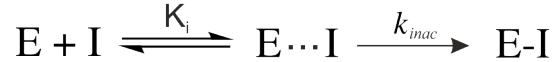


Abbildung 9.2: Scheme of a irreversible inhibition reaction with E: enzyme, I: inhibitor, K_i = inhibitor concentration of half maximal rate of covalent bond formation.

Further Results of chapter 7.3.4 - Water-Mediated Proton Transfer of Carbamate-Based Inhibitors in Cathepsin B

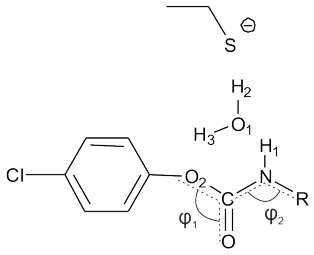


Abbildung 9.3: Definition of the geometric parameters of the system.

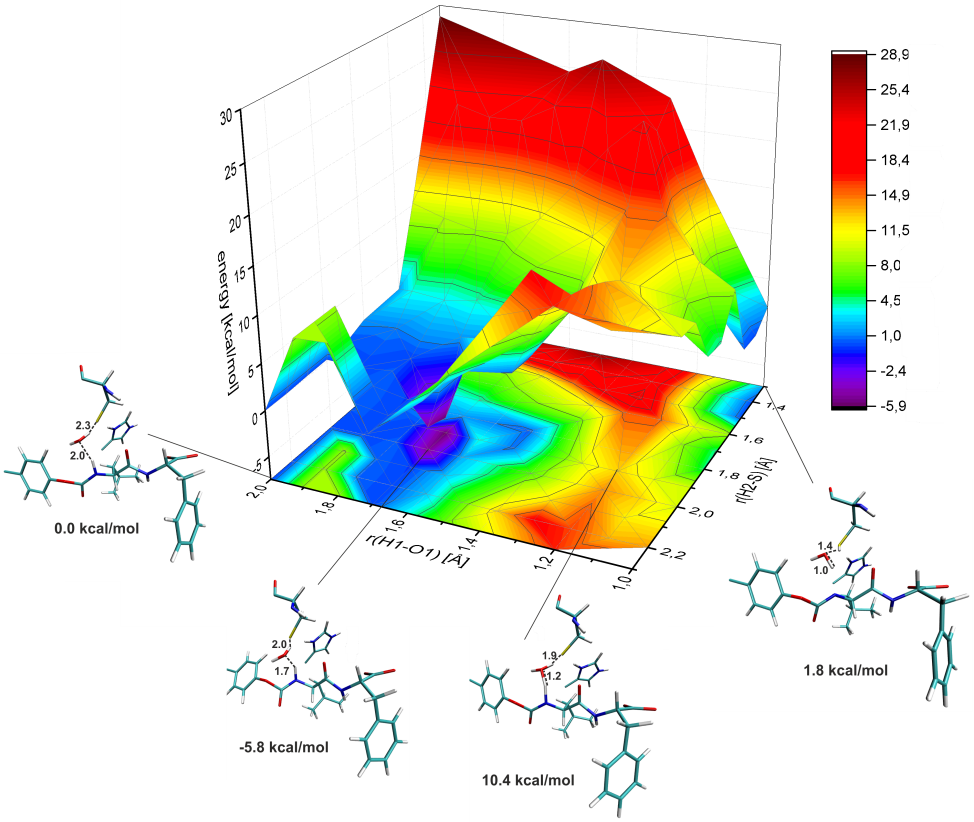


Abbildung 9.4: PES (QM/MM) of the water-bridged proton transfer from the inhibitor to the Cys29 (QM method PM3) with an alternative set of coordinates compared to the one described in chapter 7.3.4.

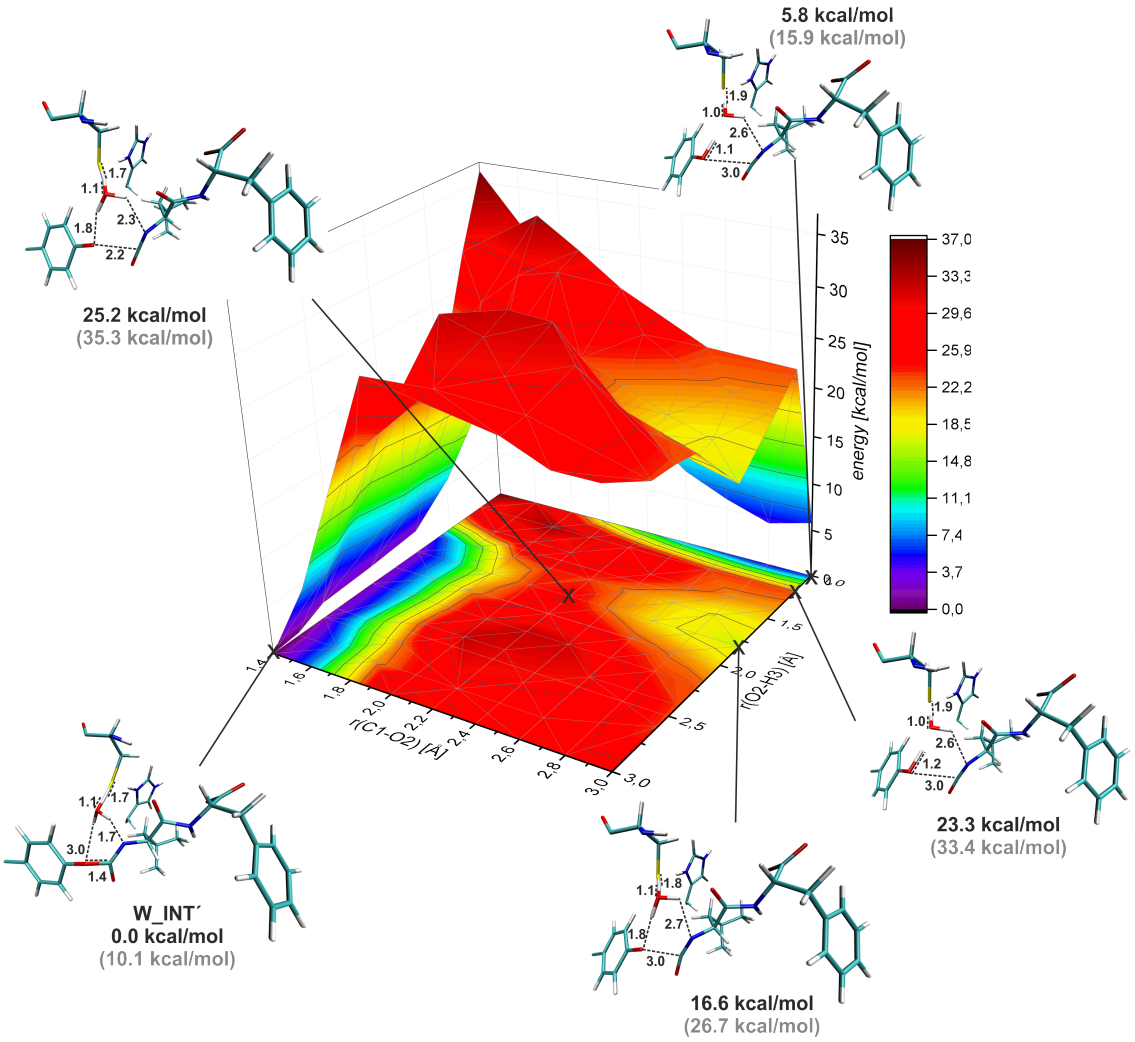


Abbildung 9.5: PES of the water-bridged dissociation of the inhibitor with the simultaneous protonation of the phenolate (PM3 as QM method) with starting structure W_INT′ from Figure 7.26 in chapter 7.3.4. Gray values give the energies relative to the reactant structure of the mechanism.

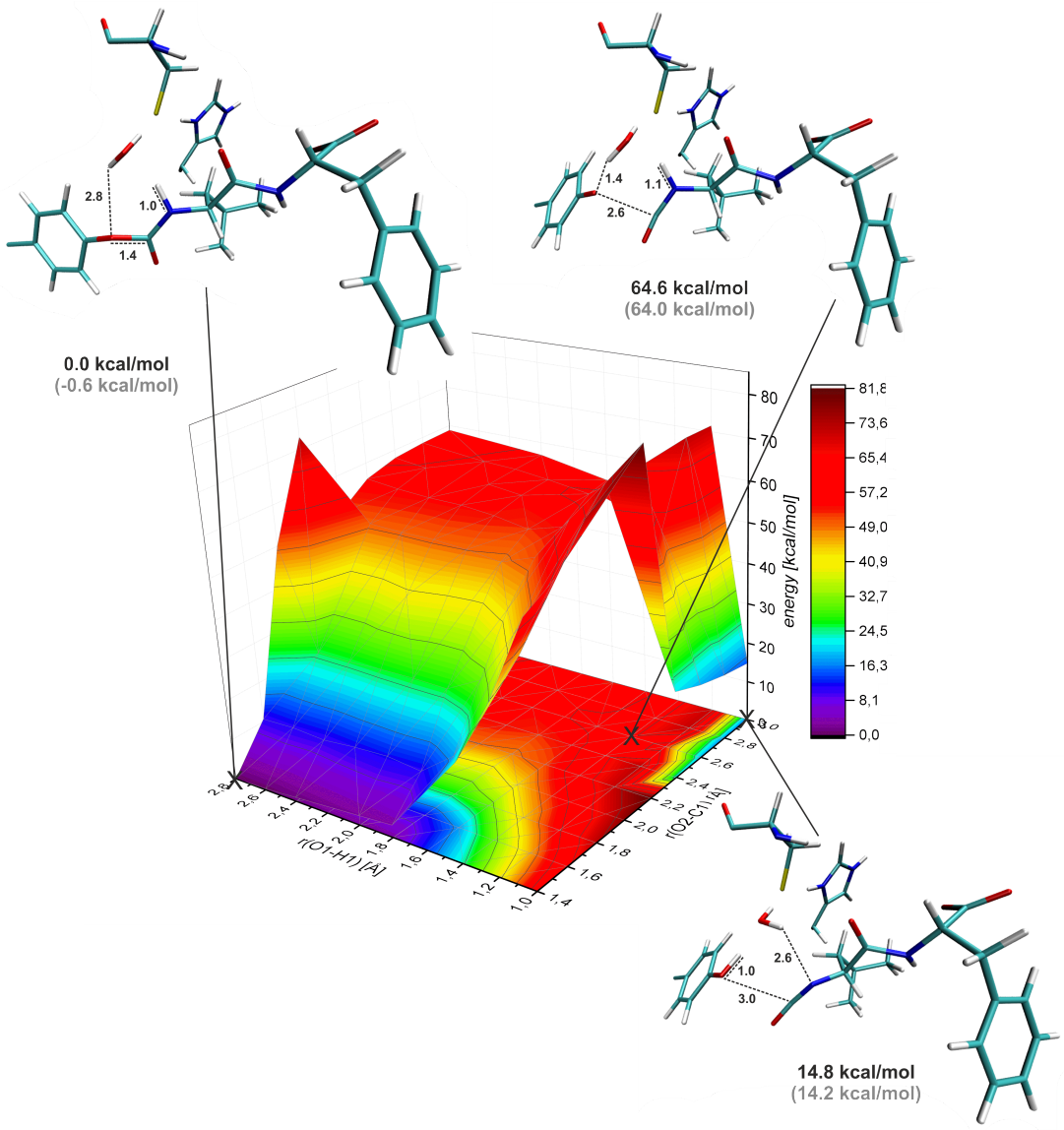


Abbildung 9.6: PES of the water bridged dissotiation of the inhibitor with the simultanious protonation of phenolat (PM3 as QM method) without the participation of Cys29 with an alternative set of coordinates compared to the one described in chapter 7.3.4. Gray values give the energies relative to the reactant structure of the mechanism.

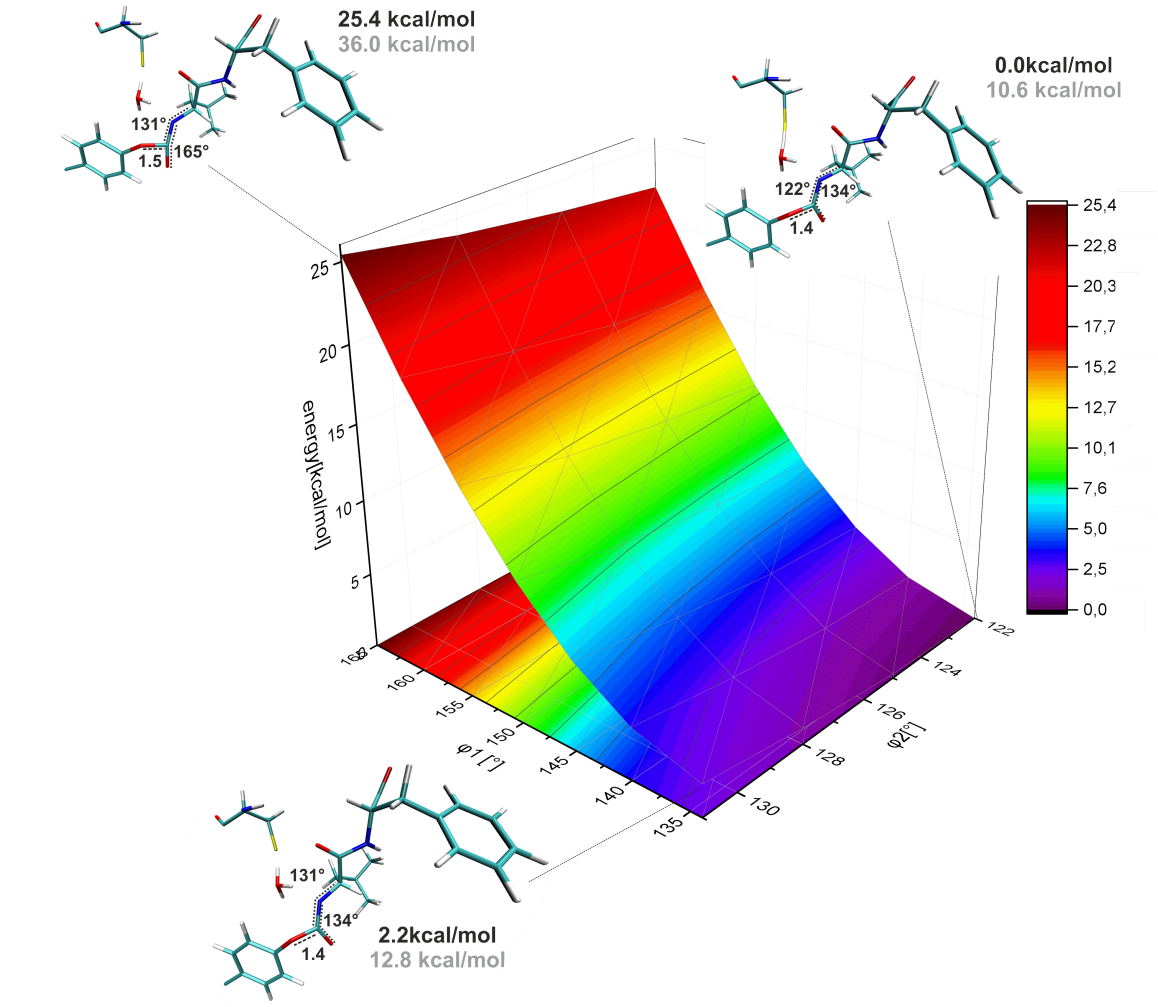


Abbildung 9.7: PES (PM3 as QM method) of the reaction from starting structure W_TS′ from Figure 7.45 in chapter 7.3.4 for the modification of the two angles ϕ_1 and ϕ_2. Gray values give the energies relative to the reactant structure of the mechanism.

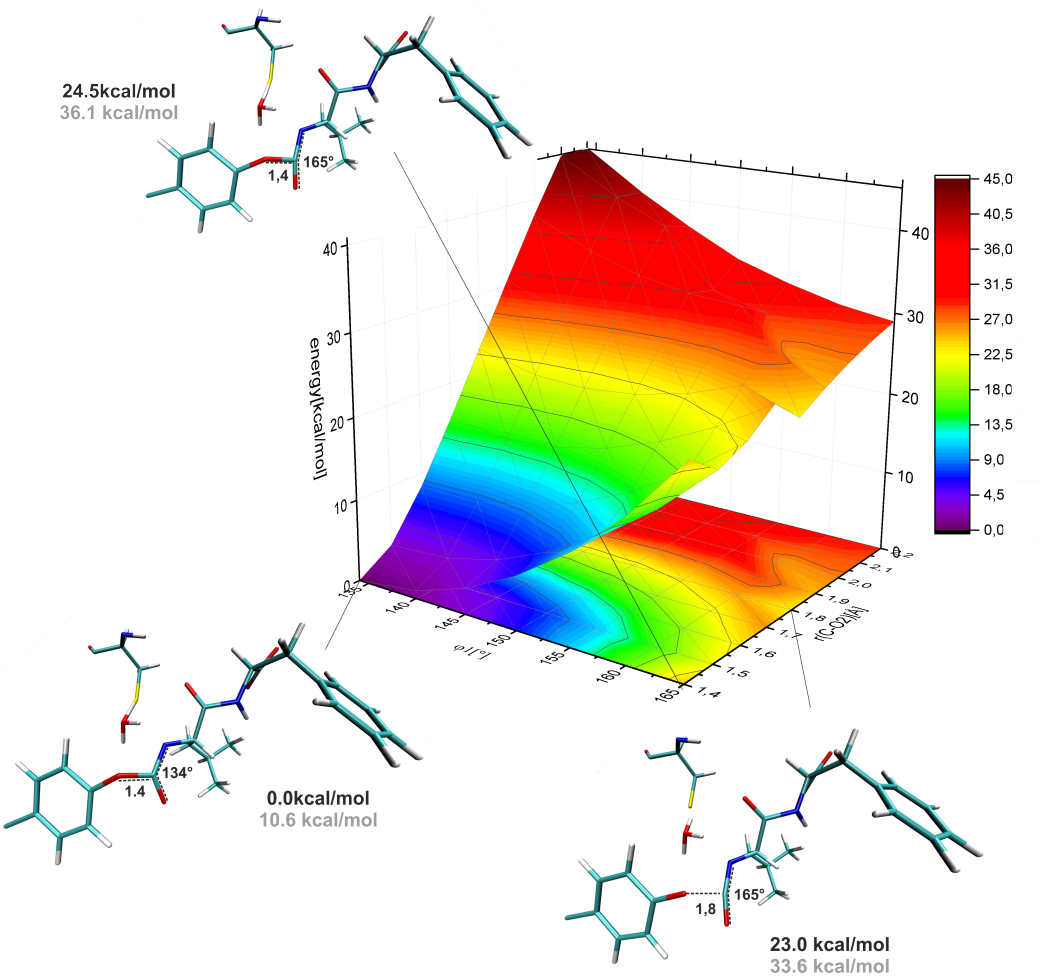


Abbildung 9.8: PES (PM3 as QM method) of the reaction from starting structure W_TS′ from Figure 7.45 in chapter 7.3.4 for the modification the angle ϕ_1 and the C-O_2 bond. Gray values give the energies relative to the reactant structure of the mechanism.

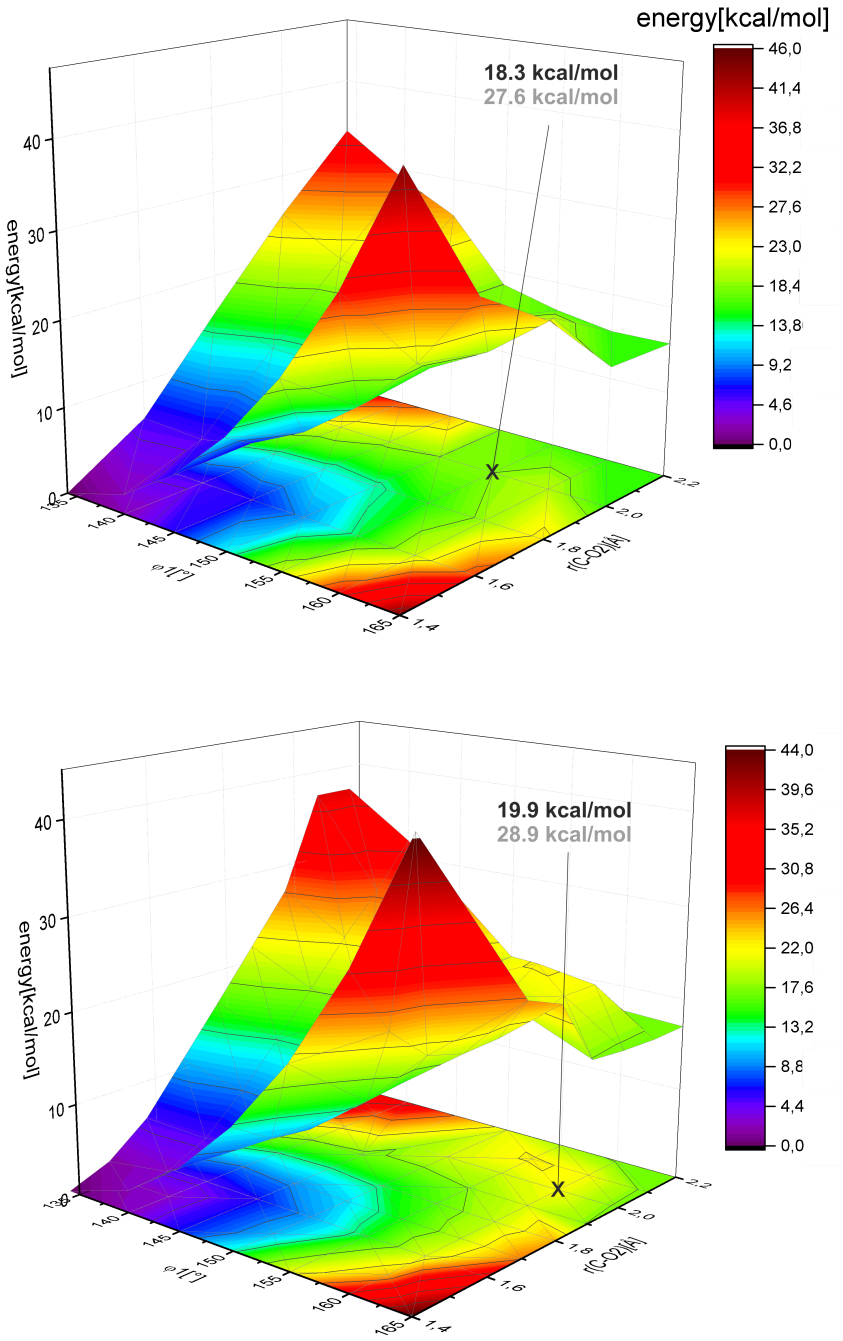


Abbildung 9.9: QM/MM PES (Single Point calculations with B3LYP/aug-cc-pVDZ (upper graphic) and SCS-MP2/aug-cc-pVDZ (lower graphic) as QM methods on the structures, optimized with PM3 es QM method) of the reaction from starting structure W_TS' from Figure 7.45 in chapter 7.3.4 for the modification of the angle ϕ_1 and the C-O_2 bond. Gray values give the energies relative to the reactant structure of the mechanism.

10 Abbreviations

ACh Acetylcholine

AChE Acetylcholine esterase

Ala Alanine

Arg Arginine

Asn Asparagine

Asp Aspartic acid

Cys Cysteine

DFT Density functional theory

FF Force field

GAFF Generalized AMBER force field

GGA Generalized gradient approximation

Gln Glutamine

Glu Glutamic acid

Gly Glycine

HAT Human African Trypanosomiasis

HF Hartree Fock

His Histidine

Ile Isoleucine

LDA Local density approximation

Leu Leucine

Lys Lysine

MD Molecular dynamics

Met Methionine

MM Molecular mechanics

MP2 Second order Møller-Plesset Perturbation Theory

NEB Nugded Elastic Band

Phe Phenylalanine

Pro Proline

QM Quantum Mechanics

RMSD Root mean square deviation

SCS Spin component scaling

Ser Serine

T.b. Trypanosoma brucei

Thr Threonine

Trp Tryptophan

Tyr Tyrosine

Val Valine

WHO World Health Organization

11 Literaturverzeichnis

[1] Working group of Prof. Dr. M. Guetschow, University of Bonn, Pharma-Zentrum; Unpublished data.

[2] Latorre, A.; Schirmeister, T.; Kesselring, J.; Jung, S.; Johe, P.; Hellmich, U. A.; Heilos, A.; Engels, B.; Krauth-Siegel, R. L.; Dirdjaja, N.; Bou-Iserte, L.; Rodriguez, S.; Gonzalez, F. V. *Acs Medicinal Chemistry Letters* **2016**, *7*, 1073–1076.

[3] Working group of Prof. Dr. T. Schirmeister, University of Mainz, Institute of Pharmacy and Biochemistry; Unpublished data.

[4] Puente, X. S.; Sanchez, L. M.; Overall, C. M.; López-Otín, C. *Nature Reviews Genetics* **2003**, *4*, 544.

[5] Rawlings, N. D.; Barrett, A. J.; Thomas, P. D.; Huang, X.; Bateman, A.; Finn, R. D. *Nucleic Acids Research* **2018**, *46*, D624–D632.

[6] Barrett, A. J.; Rawlings, N. D.; Woessner, J. F. *Handbook of proteolytic enzymes*; Academic Press: San Diego, **2004**.

[7] Davie, E. W.; Neurath, H. *Journal of Biological Chemistry* **1955**, *212*, 515–529.

[8] Davie, E. W.; Ratnoff, O. D. *Science* **1964**, *145*, 1310.

[9] Turk, B. *Nature Reviews Drug Discovery* **2006**, *5*, 785–799.

[10] Bernstein, H. G.; Kirschke, H.; Wiederanders, B.; Pollak, K. H.; Zipress, A.; Rinne, A. *Molecular and Chemical Neuropathology* **1996**, *27*, 225–247.

[11] Lecaille, F.; Kaleta, J.; Brömme, D. *Chemical Reviews* **2002**, *102*, 4459–4488.

[12] Yan, S. Q.; Sameni, M.; Sloane, B. F. *Biological Chemistry* **1998**, *379*, 113–123.

[13] DeClerck, Y. A.; Imren, S.; Montgomery, A. M. P.; Mueller, B. M.; Reisfeld, R. A.; Laug, W. E. *Chemistry and Biology of Serpins* **1997**, *425*, 89–97.

[14] Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chemical Reviews* **2002**, *102*, 4639–4750.

[15] Caffrey, C. R.; Hansell, E.; Lucas, K. D.; Brinen, L. S.; Hernandez, A. A.; Cheng, J. N.; Gwaltney, S. L.; Roush, W. R.; Stierhof, Y. D.; Bogyo, M.; Steverding, D.; McKerrow, J. H. *Molecular and Biochemical Parasitology* **2001**, *118*, 61–73.

[16] Cazzulo, J. J.; Stoka, V.; Turk, V. *Biological Chemistry* **1997**, *378*, 1–10.

[17] World malaria report; **2017**; Geneva: World Health Organisation; Licence: CC BY-NC-SA 3.0 IGO.

[18] Schirmeister, T.; Welker, A. *Pharmazie in unserer Zeit* **2009**, *38*, 564–574.

[19] Seife, C. *Science* **1997**, *277*, 1602.

[20] Rawlings, N. D.; Barrett, A. J. *Nucleic Acids Research* **1999**, *27*, 325–331.

[21] Ghosh, A. K.; Gemma, A. *Structure-Based Design of Drugs and Other Bioactive Molecules*; John Wiley and Sons, **2014**.

[22] Sajid, M.; McKerrow, J. H. *Molecular and Biochemical Parasitology* **2002**, *120*, 1–21.

[23] Turk, S.; Turk, B.; Turk, V. *Proteases and the Regulation of Biological Processes* **2003**, *70*, 15–30.

[24] Turk, D.; Guncar, G.; Podobnik, M.; Turk, B. *Biological Chemistry* **1998**, *379*, 137–147.

[25] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235–242.

[26] Musil, D.; Zucic, D.; Turk, D.; Engh, R. A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N.; Bode, W. *Embo Journal* **1991**, *10*, 2321–2330.

[27] Pickersgill, R. W.; Harris, G. W.; Garman, E. *Acta Crystallographica Section B-Structural Science* **1992**, *48*, 59–67.

[28] Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K. C.; Caffrey, C. R.; Legac, J.; Hansell, E.; McKerrow, J. H.; Craik, C. S.; Rosenthal, P. J.; Brinen, L. S. *Journal of Biological Chemistry* **2009**, *284*, 25697–25703.

[29] Schechter, I.; Berger, A. *Biochemical and Biophysical Research Communications* **1967**, *27*, 157–162.

[30] Paasche, A.; Schirmeister, T.; Engels, B. *Journal of Chemical Theory and Computation* **2013**, *9*, 1765–1777.

[31] Paasche, A.; Zipper, A.; Schafer, S.; Ziebuhr, J.; Schirmeister, T.; Engels, B. *Biochemistry* **2014**, *53*, 5930–5946.

[32] Michael, S.; Netaly, K.; Albeck, A. *Proteins: Structure, Function, and Bioinformatics* **2009**, *77*, 916–926.

[33] Shokhen, M.; Khazanov, N.; Albeck, A. *Proteins: Structure, Function, and Bioinformatics* **2010**, *79*, 975–985.

[34] Schmidt, T. C.; Welker, A.; Rieger, M.; Sahu, P. K.; Sotriffer, C. A.; Schirmeister, T.; Engels, B. *Chemical Physics Physical Chemistry* **2014**, *15*, 3226–3235.

[35] Brodie, B. B.; Reid, W. D.; Cho, A. K.; Sipes, G.; Krishna, G.; Gillette, J. R. *Proceedings of the National Academy of Sciences of the United States of America* **1971**, *68*, 160–164.

[36] Jollow, D. J.; Mitchell, J. R.; Potter, W. Z.; Davis, D. C.; Gillette, J. R.; Brodie, B. B. *Journal of Pharmacology and Experimental Therapeutics* **1973**, *187*, 195–202.

[37] Smith, A. J. T.; Zhang, X. Y.; Leach, A. G.; Houk, K. N. *Journal of Medicinal Chemistry* **2009**, *52*, 225–233.

[38] Schirmeister, T.; Kesselring, J.; Jung, S.; Schneider, T. H.; Weickert, A.; Becker, J.; Lee, W.; Bamberger, D.; Wich, P. R.; Distler, U.; Tenzer, S.; Johe, P.; Hellmich, U. A.; Engels, B. *Journal of the American Chemical Society* **2016**, *138*, 8332–8335.

[39] Vicik, R.; Helten, H.; Schirmeister, T.; Engels, B. *ChemMedChem* **2006**, *1*, 1021–1028.

[40] Schneider, T. H.; Rieger, M.; Ansorg, K.; Sobolev, A. N.; Schirmeister, T.; Engels, B.; Grabowsky, S. *New Journal of Chemistry* **2015**, *39*, 5841–5853.

[41] Ettari, R.; Tamborini, L.; Angelo, I. C.; Micale, N.; Pinto, A.; De Micheli, C.; Conti, P. *Journal of Medicinal Chemistry* **2013**, *56*, 5637–5658.

[42] Ludewig, S.; Kossner, M.; Schiller, M.; Baumann, K.; Schirmeister, T. *Current Topics in Medicinal Chemistry* **2010**, *10*, 368–382.

[43] Frlan, R.; Gobec, S. *Current Medicinal Chemistry* **2006**, *13*, 2309–2327.

[44] Stern, I.; Schaschke, N.; Moroder, L.; Turk, D. *Biochemical Journal* **2004**, *381*, 511–517.

[45] Buth, H.; Wolters, B.; Hartwig, B.; Meier-Bornheim, R.; Veith, H.; Hansen, M.; Sommerhoff, C. P.; Schaschke, N.; Machleidt, N.; Fusenig, N. E.; Boukamp, P.; Brix, K. *European Journal of Cell Biology* **2004**, *83*, 781–795.

[46] Turk, B.; Turk, D.; Salvesen, G. S. *Current Pharmaceutical Design* **2002**, *8*, 1623–1637.

[47] Buck, M. R.; Karustis, D. G.; Day, N. A.; Honn, K. V.; Sloane, B. F. *Biochemical Journal* **1992**, *282*, 273–278.

[48] Lutgens, S. P. M.; Cleutjens, K.; Daemen, M.; Heeneman, S. *Faseb Journal* **2007**, *21*, 3029–3041.

[49] Yasuda, Y.; Li, Z. Q.; Greenbaum, D.; Bogyo, M.; Weber, E.; Bromme, D. *Journal of Biological Chemistry* **2004**, *279*, 36761–36770.

[50] Montcourrier, P.; Mangeat, P. H.; Valembois, C.; Salazar, G.; Sahuquet, A.; Duperray, C.; Rochefort, H. *Journal of Cell Science* **1994**, *107*, 2381–2391.

[51] Ruan, J.; Zheng, H. Y.; Rong, X. D.; Rong, X. M.; Zhang, J. Y.; Fang, W. J.; Zhao, P.; Luo, R. C. *Molecular Cancer* **2016**, *15*.

[52] Wu, Z.; Ni, J. J.; Liu, Y. C.; Teeling, J. L.; Takayama, F.; Collcutt, A.; Ibbett, P.; Nakanishi, H. *Brain Behavior and Immunity* **2017**, *65*, 350–361.

[53] Kempson, G. E.; Tuke, M. A.; Dingle, J. T.; Barrett, A. J.; Horsfield, P. H. *Biochimica Et Biophysica Acta* **1976**, *428*, 741–760.

[54] Hummel, K. M.; Petrow, P. K.; Franz, J. K.; Muller-Ladner, U.; Aicher, W. K.; Gay, R. E.; Bromme, D.; Gay, S. *Journal of Rheumatology* **1998**, *25*, 1887–1894.

[55] Trabandt, A.; Aicher, W. K.; Gay, R. E.; Sukhatme, V. P.; Nilsonhamilton, M.; Hamilton, R. T.; McGhee, J. R.; Fassbender, H. G.; Gay, S. *Matrix* **1990**, *10*, 349–361.

[56] Trabandt, A.; Gay, R. E.; Fassbender, H. G.; Gay, S. *Arthritis and Rheumatism* **1991**, *34*, 1444–1451.

[57] Hou, W. S.; Li, Z. Q.; Gordon, R. E.; Chan, K.; Klein, M. J.; Levy, R.; Keysser, M.; Keyszer, G.; Bromme, D. *American Journal of Pathology* **2001**, *159*, 2167–2177.

[58] Esser, R. E.; Watts, L. M.; Angelo, R. A.; Thornburg, L. P.; Prior, J. J.; Palmer, J. T. *Journal of Rheumatology* **1993**, *20*, 1176–1183.

[59] Esser, R. E.; Angelo, R. A.; Murphey, M. D.; Watts, L. M.; Thornburg, L. P.; Palmer, J. T.; Talhouk, J. W.; Smith, R. E. *Arthritis and Rheumatism* **1994**, *37*, 236–247.

[60] Biroc, S. L.; Gay, S.; Hummel, K.; Magill, C.; Palmer, J. T.; Spencer, D. R.; Sa, S.; Klaus, J. L.; Michel, B. A.; Rasnick, D.; Gay, R. E. *Arthritis and Rheumatism* **2001**, *44*, 703–711.

[61] Visscher, D. W.; Sloane, B. F.; Sameni, M.; Babiarz, J. W.; Jacobson, J.; Crissman, J. D. *Modern Pathology* **1994**, *7*, 76–81.

[62] Emmertbuck, M. R.; Roth, M. J.; Zhuang, Z. P.; Campo, E.; Rozhin, J. R.; Sloane, B. F.; Liotta, L. A.; Stetlerstevenson, W. G. *American Journal of Pathology* **1994**, *145*, 1285–1290.

[63] Sinha, A. A.; Gleason, D. F.; Staley, N. A.; Wilson, M. J.; Sameni, M.; Sloane, B. F. *Anatomical Record* **1995**, *241*, 353–362.

[64] Sinha, A. A.; Gleason, D. F.; Deleon, O. F.; Wilson, M. J.; Sloane, B. F. *Anatomical Record* **1993**, *235*, 233–240.

[65] Mueller-Steiner, S.; Zhou, Y.; Arai, H.; Roberson, E. D.; Sun, B. G.; Chen, J.; Wang, X.; Yu, G. Q.; Esposito, L.; Mucke, L.; Gan, L. *Neuron* **2006**, *51*, 703–714.

[66] Palmer, J. T.; Rasnick, D.; Klaus, J. L. *Journal of the American Chemical Society* **1995**, *38*, 3193–3196.

[67] Steverding, D. *Parasites and Vectors* **2008**, *1:3*, 1–8.

[68] WHO Report on Human African Trypanosomiasis **2017**; Geneva: World Health Organisation.

[69] Barrett, M. P.; Burchmore, R. J. S.; Stich, A.; Lazzari, J. O.; Frasch, A. C.; Cazzulo, J. J.; Krishna, S. *Lancet* **2003**, *362*, 1469–1480.

[70] Cox, F. E. G. *Infectious Disease Clinics of North America* **2004**, *18*, 231.

[71] Njiokou, F.; Laveissere, C.; Simo, G.; Nkinin, S.; Grebaut, P.; Cuny, G.; Herder, S. *Infection Genetics and Evolution* **2006**, *6*, 147–153.

[72] Kennedy, P. G. E. *Journal of Clinical Investigation* **2004**, *113*, 496–504.

[73] Simo, G.; Asonganyi, T.; Nkinin, S. W.; Njiokou, F.; Herder, S. *Veterinary Parasitology* **2006**, *139*, 57–66.

[74] Kerr, I. D.; Wu, P.; Marion-Tsukamaki, R.; Mackey, Z. B.; Brinen, L. S. *Plos Neglected Tropical Diseases* **2010**, *4*, 1–9.

[75] McKerrow, J. H.; Engel, J. C.; Caffrey, C. R. *Bioorganic and Medicinal Chemistry* **1999**, *7*, 639–644.

[76] Pepin, J.; Milord, F. *Advances in Parasitology, Vol 33* **1994**, *33*, 1–47.

[77] Mackey, Z. B.; O'Brien, T. C.; Greenbaum, D. C.; Blank, R. B.; McKerrow, J. H. *Journal of Biological Chemistry* **2004**, *279*, 48426–48433.

[78] Lonsdale-Eccles, J. D.; Grab, D. J. *Trends in Parasitology* **2002**, *18*, 17–19.

[79] Michael, C.; Turner, R. *Journal of Cell Science* **1999**, *112*, 3187–3192.

[80] Barry, J. D.; McCulloch, R. *Advances in Parasitology, Vol 49* **2001**, *49*, 1–70.

[81] Warshel, A.; Levitt, M. *Journal of Molecular Biology* **1976**, *103*, 227–249.

[82] Senn, H. M.; Thiel, W. *Angewandte Chemie-International Edition* **2009**, *48*, 1198–1229.

[83] M. Sierka, J. S. *Handbook of Materials Modeling*; Springer, **2005**; Vol. A.

[84] To, J.; Sherwood, P.; Sokol, A. A.; Bush, I. J.; Catlow, C. R. A.; van Dam, H. J. J.; French, S. A.; Guest, M. F. *Journal of Materials Chemistry* **2006**, *16*, 1919–1926.

[85] Brown, J. M.; Hofmann, P. *Organometallic bonding and reactivity: fundamental studies*; Springer Verlag, **1999**.

[86] Yang, S.-Y.; Ziegler, T. *Organometallics* **2006**, *25*, 887–900.

[87] Rossbach, S.; Ochsenfeld, C. *Journal of Chemical Theory and Computation* **2017**, *13*, 1102–1107.

[88] Bakowies, D.; Thiel, W. *The Journal of Physical Chemistry* **1996**, *100*, 10580–10594.

[89] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *The Journal of Physical Chemistry* **1996**, *100*, 19357–19363.

[90] Patel, S.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 1–15.

[91] Patel, S.; Mackerell, A. D.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 1504–1514.

[92] Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D. *Journal of Physical Chemistry B* **2005**, *109*, 18988–18999.

[93] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *Journal of Physical Chemistry A* **2004**, *108*, 621–627.

[94] Tafipolsky, M.; Ansorg, K. *Journal of Chemical Theory and Computation* **2016**, *12*, 1267–1279.

[95] Lennard-Jones, J. E. *Proc. R. Soc. London, Ser. A* **1925**, *109*, 584–597.

[96] Freindorf, M.; Shao, Y.; Kong, J.; Furlani, T. R. *Journal of Inorganic Biochemistry* **2008**, *102*, 427–432.

[97] Heilos, A.; *Master Thesis*; **2014**; Insitute of physical and theoretical chemistry; university of Würzburg.

[98] Szabo, A.; Ostlund, S., Neil *Modern Quantum Chemistry*, first edition ed.; Dover Publications: New York, **1996**.

[99] Jensen, F. *Introduction to Computational Chemistry*, second edition ed.; John Wiley and Sons, Ltd, **2007**.

[100] Roothaan, C. C. J. *Reviews of Modern Physics* **1951**, *23*, 69–89.

[101] Hall, G. G. *Proceedings of the Royal Society of London Series A - Mathematical and Physical Sciences* **1951**, *205*, 541–552.

[102] Møller,C. and Plesset, M. S. *Am. Phys. Soc.* **1934**, *46*, 618–622.

[103] Grimme, S. *Journal of Chemical Physics* **2003**, *118*, 9095–9102.

[104] Pauli, W. *Z. Physik* **1927**, *43*, 601–623.

[105] Grimme, S.; Goerigk, L.; Fink, R. F. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012**, *2*, 886–906.

[106] Goumans, T. P. M.; Ehlers, A. W.; Lammertsma, K.; Wurthwein, E. U.; Grimme, S. *Chemistry - A European Journal* **2004**, *10*, 6468–6475.

[107] Gerenkamp, M.; Grimme, S. **2004**, *392*, 229–235.

[108] Antony, J.; Grimme, S. *Journal of Physical Chemistry A* **2007**, *111*, 4862–4868.

[109] Jung, Y. S.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *Journal of Chemical Physics* **2004**, *121*, 9793–9802.

[110] Lochan, R. C.; Jung, Y.; Head-Gordon, M. *The Journal of Physical Chemistry A* **2005**, *109*, 7598–7605.

[111] Koch, W.; Holthausen, M. *A Chemist´s Guide to Density Functional Theory,*, 2nd ed.; Wiley-VCH, **2001**.

[112] Engel, E.; Dreizler, R. *Density Functional Theory*; Springer, **2011**.

[113] Thomas, L. *Proc. Camb. Phil.Soc.* **1927**, *23*, 542–547.

[114] Schrödinger, E. *Annalen der Physik* **1926**, *385*, 437–490.

[115] Fermi, E. **1928**, *48*, 73–79.

[116] Dirac, P. *Roy. Soc.* **1930**, *126*, 360–365.

[117] Fock, V. *Z. Phys.* **1930**, *61*, 126–148.

[118] Hohenberg, P.; Kohn, W. *Physical Review B* **1964**, *136*.

[119] Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*.

[120] Perdew, J. P.; Schmidt, K. *AIP Conf. Proc.* **2001**, *577*, 1–20.

[121] Mardirossian, N.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2014**, *16*, 9904–9924.

[122] Lee, C. T.; Yang, W. T.; Parr, R. G. *Physical Review B* **1988**, *37*, 785–789.

[123] Zhao, Y.; Truhlar, D. G. *Journal of Chemical Physics* **2006**, *125*, 1–18.

[124] Peverati, R.; Truhlar, D. G. *Journal of Physical Chemistry Letters* **2011**, *2*, 2810–2817.

[125] Grimme, S. *Journal of Chemical Physics* **2006**, *124*, 1–16.

[126] Bremond, E.; Adamo, C. *Journal of Chemical Physics* **2011**, *135*, 1–6.

[127] Thiel, W. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2014**, *4*, 145–157.

[128] Thiel, W. *Semiempirical Methods*, **2000**; Vol. 3.

[129] Pariser, R.; Parr, R. G. *Journal of Chemical Physics* **1953**, *21*, 466–471.

[130] Pariser, R.; Parr, R. G. *Journal of Chemical Physics* **1953**, *21*, 767–776.

[131] Segal, G. *The Journal of Physical Chemistry* **1990**, *94*, 5436–5439.

[132] Dewar, M. J. S.; Thiel, W. *Journal of the American Chemical Society* **1977**, *99*, 4899–4907.

[133] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *Journal of the American Chemical Society* **1985**, *107*, 3902–3909.

[134] Stewart, J. J. P. *Journal of Computational Chemistry* **1989**, *10*, 209–220.

[135] Goldstein, H.; Poole Jr., H. C. P.; Safko, J. L. *Classical Mechanics*, 3rd ed.; Pearson, **2006**.

[136] Boyd, D. B.; Lipkowitz, K. B. *Journal of Chemical Education* **1982**, *59*, 269–274.

[137] Allinger, N. L. *Advances in Physical Organic Chemistry*; Academic Press, **1976**.

[138] Alder, B. J.; Wainwright, T. E. *Journal of Chemical Physics* **1957**, *27*, 1208–1209.

[139] Alder, B. J.; Wainwright, T. E. *Journal of Chemical Physics* **1959**, *31*, 459–466.

[140] Lamberti, V. E.; Fosdick, L. D.; Jessup, E. R.; Schauble, C. J. C. *Journal of Chemical Education* **2002**, *79*, 601–606.

[141] Verlet, L. *Physical Review* **1967**, *159*, 98–103.

[142] Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; *AMBER 11*; **2010**.

[143] Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A.; *AMBER 14*; **2014**.

[144] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 712–725.

[145] Maier, J. A.; Carmenza, M.; Koushik, K.; Lauren, W.; E., H. K.; Carlos, S. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

[146] Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

[147] Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2005**, *26*, 114–114.

[148] Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

[149] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926–935.

[150] Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *Journal of Molecular Structure-Theochem* **2003**, *632*, 1–28.

[151] *ChemShell, a Computational Chemistry Shell, see www.chemshell.org*.

[152] Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Journal of Physical Chemistry Letters* **1989**, *162*, 165–169.

[153] Treutler, O.; Ahlrichs, R. *Journal of Chemical Physics* **1995**, *102*, 346–354.

[154] Von Arnim, M.; Ahlrichs, R. *Journal of Computational Chemistry* **1999**, *19*, 1746–1757.

[155] Klamt, A.; Schuurmann, G. *Journal of the Chemical Society-Perkin Transactions 2* **1993**, 799–805.

[156] Thiel, W.; *MNDO2005 V7.0, a development of Max-Planck-Institut fuer Kohlenforschung*; **2005**.

[157] Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *Journal of Physical Chemistry A* **2009**, *113*, 11856–11865.

[158] *Origin (OriginLab, Northampton, MA).*

[159] Humphrey, W.; Dalke, A.; Schulten, K. **1996**, *14*, 33–38.

[160] Bajda, M.; Latka, K.; Hebda, M.; Jonczyk, J.; Malawska, B. *Bioorganic Chemistry* **2018**, *78*, 29–38.

[161] Mohammadi, T.; Ghayeb, Y. *Journal of Biomolecular Structure and Dynamics* **2018**, *36*, 126–138.

[162] Pizova, H.; Havelkova, M.; Stepankova, S.; Bak, A.; Kauerova, T.; Kozik, V.; Oravec, M.; Imramovsky, A.; Kollar, P.; Bobal, P.; Jampilek, J. *Molecules* **2017**, *22*.

[163] Taylor, P. *Neurology* **1998**, *51*, 30–35.

[164] Bergon, M.; Hamida, N. B.; Calmon, J. P. *Journal of Agricultural and Food Chemistry* **1985**, *33*, 577–583.

[165] Benson, W. H.; Dorough, H. W. *Pesticide Biochemistry and Physiology* **1984**, *21*, 199–206.

[166] Mata, G.; do Rosario, V. E.; Iley, J.; Constantino, L.; Moreira, R. *Bioorganic and Medicinal Chemistry* **2012**, *20*, 886–892.

[167] Cognetta, A. B.; Niphakis, M. J.; Lee, H. C.; Martini, M. L.; Hulce, J. J.; Cravatt, B. F. *Chemistry and Biology* **2015**, *22*, 928–937.

[168] Bachovchin, D. A.; Ji, T. Y.; Li, W. W.; Simon, G. M.; Blankman, J. L.; Adibekian, A.; Hoover, H.; Niessen, S.; Cravatt, B. F. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 20941–20946.

[169] Chang, J. W.; Cognetta, A. B.; Niphakis, M. J.; Cravatt, B. F. *Acs Chemical Biology* **2013**, *8*, 1590–1599.

[170] Clayden, J. *Organic chemistry*; Oxford University Press: Oxford; New York, **2007**.

[171] Ouertani, R.; El Atrache, L. L.; Ben Hamida, N. *International Journal of Chemical Kinetics* **2013**, *45*, 118–124.

[172] Sanz-Asensio, J.; Plaza-Medina, M.; Martinez-Soria, M. T. *Pesticide Science* **1997**, *50*, 187–194.

[173] Walz, I.; Schwack, W. *European Food Research and Technology* **2008**, *226*, 1135–1143.

[174] Scofield, R. E.; Werner, R. P.; Wold, F. *Biochemistry* **1977**, *16*, 2492–2496.

[175] Epple, R.; Urbina, H. D.; Russo, R.; Liu, H.; Mason, D.; Bursulaya, B.; Tumanut, C.; Li, J.; Harris, J. L. *Bioorganic and Medicinal Chemistry Letters* **2007**, *17*, 1254–1259.

[176] Dunning, T. H. *Journal of Chemical Physics* **1989**, *90*, 1007–1023.

[177] Engels, B.; Schmidt, T. C.; Gatti, C.; Schirmeister, T.; Fink, R. F. *Challenging Problems in Charge Density Determination: Polar Bonds and Influence of the Environment*; Structure and Bonding, Vol. 147; Springer: New York, **2012**.

[178] Schmidt, T. C.; Paasche, A.; Grebner, C.; Ansorg, K.; Becker, J.; Lee, W.; Engels, B. *QM/MM Investigations Of Organic Chemistry Oriented Questions*; Topics in Current Chemistry-Series, Vol. 351; Springer-Verlag Berlin: Berlin, 2014.

[179] Working group of Prof. Dr. F. V. Gonzales, Universitat Jaume, Spain; Departament de Quimica Inorganica i Organica; Unpublished data.

[180] Latorre, A.; Schirmeister, T.; Kesselring, J.; Jung, S.; Johe, P.; Hellmich, U. A.; Heilos, A.; Engels, B.; Krauth-Siegel, R. L.; Dirdjaja, N.; Bou-Iserte, L.; Rodriguez, S.; Gonzalez, F. V. *Acs Medicinal Chemistry Letters* **2016**, *7*, 1073–1076; reprinted (adapted) with permission. Copyright 2018 American Chemical Society.

[181] Geis, A.; *Bachelor Thesis*; **2017**; Insitute of physical and theoretical chemistry; university of Würzburg.