

**Julius-Maximilians-Universität
Würzburg**

Fakultät für Chemie und Pharmazie

Theoretical Investigation into the Inhibition of Cysteine Proteases

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrads
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Milena Mladenovic

aus Niš

Würzburg 2007

Eingereicht am: _____
bei der Fakultät für Chemie und Pharmazie.

1. Gutachter: _____
2. Gutachter: _____
der Dissertation.

1. Prüfer: _____
2. Prüfer: _____
3. Prüfer: _____
des öffentlichen Promotionskolloquiums.

Tag des öffentlichen Promotionskolloquium: _____

Doktorurkunde ausgehändigt am: _____

Die vorliegende Arbeit wurde unter Anleitung von Prof. Dr. Bernd Engels von Juli 2003 bis November 2007 am Institut für Organische Chemie der Julius-Maximilians-Universität Würzburg angefertigt.

Teilergebnisse dieser Arbeit waren Gegenstand von Publikationen sowie von Postern und Kurzvorträgen.

Hiermit erkläre ich an Eides statt, dass ich die Dissertation

“Theoretical Investigations into the Inhibition of Cystein Proteases”

selbständig angefertigt habe und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ich erkläre weiterhin, dass diese Dissertation weder in gleicher oder andere Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ich habe früher außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Würzburg, 14.11.2007

Milena Mladenovic

Table of Contents

Chapter 1	Introduction	1
Chapter 2	Background	5
2.1	Cysteine Proteases and Their Inhibitors.....	5
2.1.1	Occurrence and structure of cysteine proteases	6
2.1.2	Mechanism of proteolysis	9
2.1.3	Inhibitors of cysteine proteases.....	10
2.1.4	Inhibition mechanism.....	13
2.2	Background on Theoretical Methods Used in This Work.....	15
2.2.1	Molecular modeling (MM) methods.....	15
2.2.1.1	Energy terms	16
2.2.2	Quantum mechanical (QM) methods.....	21
2.2.2.1	The basics.....	21
2.2.2.2	Semi-empirical methods.....	26
2.2.2.3	Density functional theory (DFT).....	28
2.2.3	Combined quantum mechanics / molecular modeling (QM/MM) methods.....	30
2.2.3.1	General principles	30
2.2.3.2	Treating covalent bonds between the QM and MM regions	33
2.2.3.3	Coupling of the QM and MM regions.....	35
2.2.4	Molecular Dynamics (MD).....	38
Chapter 3	Results and discussion	43
3.1	QM Calculations.....	47
3.1.1	The method	47
3.1.2	The role of His199 in protonation of the inhibitor.....	49
3.1.3	Substitution at the N-atom of aziridine	53
3.2	QM/MM and MD Calculations.....	65
3.2.1	The method	65
3.2.2	QM/MM treatment of inhibitor protonation by the His199 residue.....	69
3.2.3	QM/MM calculations clarifying the inhibition potency of epoxide- and aziridine-based inhibitors.....	80

3.2.4	QM/MM calculations explaining the regioselectivity of the reaction with epoxide- and aziridine-based inhibitors.....	98
3.2.5	Investigations into the stereospecificity of the cysteine protease inhibition reaction with epoxysuccinyl derivatives.....	115
3.2.6	Using MD Simulations to correct the existing X-ray data.....	131
3.3	QM/MM MD Calculations.....	137
3.3.1	The method.....	137
3.3.2	Explaining the stability of the active site HisH ⁺ /CysS ⁻ ion-pair in cysteine proteases	138
Chapter 4	Summary	157
Chapter 5	Zusammenfassung	165
Chapter 6	References.....	173

Chapter 1 Introduction

The most challenging field in computational chemistry of today, and probably its most promising field of the future, is the calculation of large molecular entities. Proteins are by far the most interesting macromolecules, due to their astonishing range of structure and function, and particularly their vital role in nature.

Proteolytic activity is essential for the normal functioning of an organism and failures in its control mechanisms can result in a vast number of diseases. For many illnesses resulting from excess proteolysis no medications exist, making research into the physiological roles of proteases and development of substances to inhibit them of highest importance to science and the pharmaceutical industry.

Infectious diseases - worldwide casue of death No. 1.

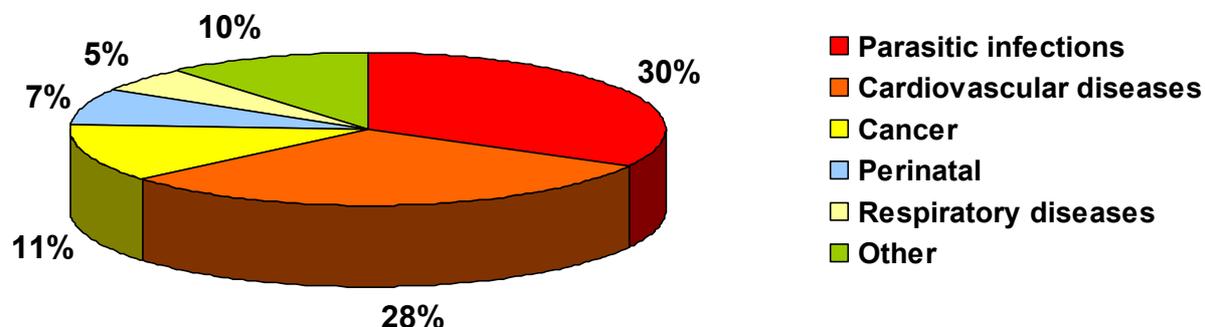


Figure 1.1.1. Worldwide death causing diseases in percent (Source: WHO). According to the WHO in 1997 50 million people worldwide died of infectious diseases.

Papain-like lysosomal cysteine proteases are processive and digestive enzymes found in a wide variety of living organisms, from bacteria to humans.¹ For many years these enzymes were believed to be responsible for redundant “house-keeping” activities and as such were considered to have little, if any, value as drug targets. However, recent progress in identifying new members of this class, as well as a better understanding of their physiological roles, has altered this perception. It has become clear that lysosomal cysteine proteases fulfill specific functions in extracellular matrix turnover, antigen presentation, and processing events. As such they are involved in many diseases and represent promising drug targets for osteoporosis,² arthritis,³ cancer,^[4] Alzheimer’s disease² and a wide variety of parasitic infections such as malaria⁵ and African trypanosomiasis (“sleeping sickness”)². According to the investigations of the world health organization (WHO), each year 50 million people die worldwide from parasitic infectious diseases alone (Figure 1.1.1). A large part of these deaths is caused by so called tropical diseases, especially malaria, sleeping sickness, leishmaniasis,² amoebiasis, etc., for which no vaccines currently exist. All of these diseases involve cysteine proteases.

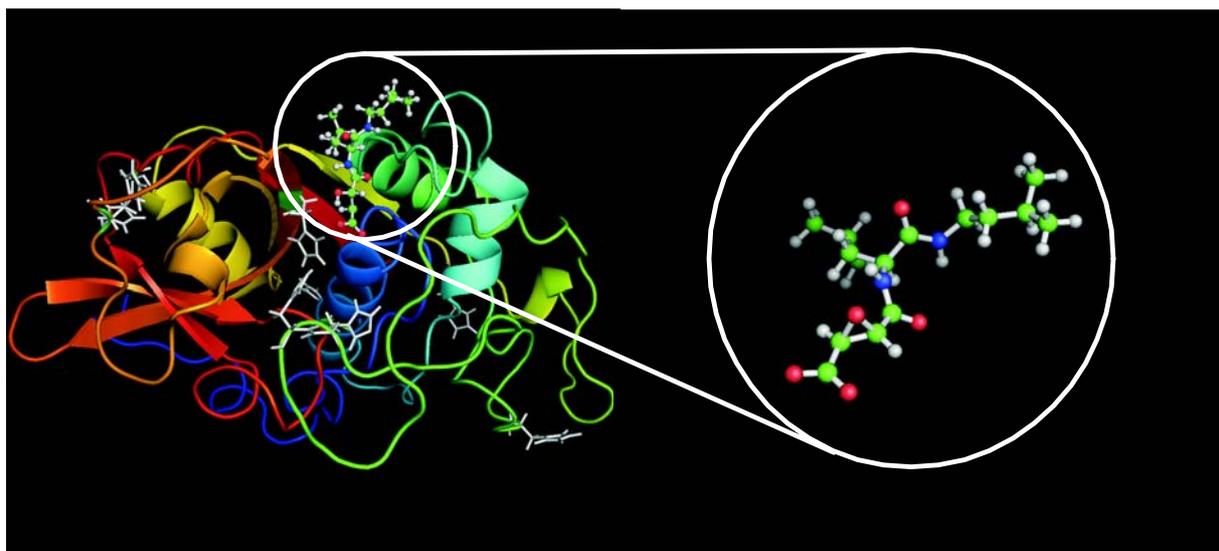


Figure 1.1.2. Left: cathepsin B–E64c enzyme-inhibitor complex. Right: E64c cysteine protease inhibitor, a derivative of the natural E64 inhibitor isolated from *Aspergillus japonicus* in the late 70’s.

A typical irreversible cysteine protease inhibitor consists of an electrophilic building-block attached to a peptidic or peptidomimetic sequence.⁶ An important class of low molecular weight cysteine protease inhibitors contains small heterocycles as electrophilic warheads. Epoxide- and aziridine-based cysteine protease inhibitors are the most important examples of this class. They have been developed as analogues to the naturally occurring E64, an epoxysuccinyl peptide discovered in the late 1970’s.⁷ Indeed, epoxysuccinic acid derivatives and analogous aziridine compounds have turned out to be promising drug candidates in many diseases.^{8,9,10,11,12,13,14} Some epoxysuccinic acid derivatives have even reached clinical development as anti-cancer drugs.¹⁵ Like epoxysuccinic acid derivatives aziridine-2,3-dicarboxylic acid derivatives have been proven to be selective, non-toxic cysteine protease inhibitors suitable as antiparasitic¹⁶ and affinity labelling¹⁷ agents. The appearance of over 120 publications and patents featuring these compounds in the last 3 years alone underlines the importance of these inhibitor classes in various fields.

Although known about and investigated since the late 1970’s, our picture of the basic principles governing inhibitor strengths and the structure-activity relationships of the cysteine protease inhibition mechanism is still very incomplete. Computational approaches can be a very useful tool for investigating such questions, as they allow the inspection of single, specific effects in isolation from all others, in a manner very difficult to achieve

experimentally. Furthermore, theoretical modelling can be used to inform the development of new, improved inhibitor species.

The *ab initio* treatments of such large systems like proteins are still not feasible. However, there is a vast number of computational approaches capable of dealing with protein structures with reasonable accuracy. This work presents a summary of theoretical investigations into cysteine protease cathepsin B, an enzyme consisting of 253 amino acids (over 3000 atoms), using a range of methods. We have concentrated on the investigation of cysteine protease inhibition by epoxide- and aziridine-based inhibitors in order to obtain better insight into these important topics. Various model systems are simulated by means of pure quantum mechanical methods and by hybrid methods (QM/MM) that combine quantum mechanical (QM) approaches for the description of bond forming and breaking processes, and molecular mechanics (MM) for the treatment of the influence of the protein environment and the surrounding solvent. Both approaches provide a static picture. Dynamical effects are then accounted for by additional molecular dynamics (MD) simulations, using both classical and QM/MM MD approaches.

This thesis is organized as follows: at the beginning a short overview of the cysteine proteases and their inhibitors, as well as of the theoretical background of the methods used in the investigations, is given. This is followed by a chapter presenting the results obtained from our calculations, organized by subchapters according to the method used and the problems treated. These subchapters also include experimental and theoretical background from the literature as well as discussion of the respective results. At the end a short summary will be given.

Chapter 2 Background

2.1 Cysteine Proteases and Their Inhibitors

Proteases (also known as peptidases) are peptide bond-cleaving enzymes (group EC 3.4). This large family of molecules can be further categorized into groups according to the position in the polypeptide chain at which the cleavage takes place (exopeptidases (EC 3.4.11-19) and endopeptidases (EC 3.4.21-99)), which terminus of the substrate is involved (aminopeptidases and carboxypeptidases), or according to which residues of the active site are involved in the cleavage process (serine (EC 3.4.21), cysteine (EC 3.4.22), aspartic (EC 3.4.23), metallo (EC

3.4.24), and threonine (EC 3.4.25) endopeptidases). Cysteine proteases (also termed thiol proteases) are endo- or exopeptidases (EC 3.4.18) in which the cysteine residue of the active site is involved in the catalysis.

2.1.1 Occurrence and structure of cysteine proteases

Cysteine proteases occur in a wide variety of living organisms, from viruses (e.g. picornain of hepatitis A and rhinovirus) and bacteria (e.g. staphylo- and streptococcal cysteine proteases) to humans (e.g. cathepsins and calpains). Most proteinases isolated from parasitic protozoa also belong to this group. They are directly involved in parasitic nutrition and invasion of the host cells, and hence play an important role in many diseases (malaria [falcipain 1-3], American trypanosomiasis [cruzipain], African trypanosomiasis (“sleeping sickness”) [rhodesain], leishmaniasis [cathepsin B and C like proteases], trichomoniasis,...) making them promising targets for the treatment of parasitic infections.²

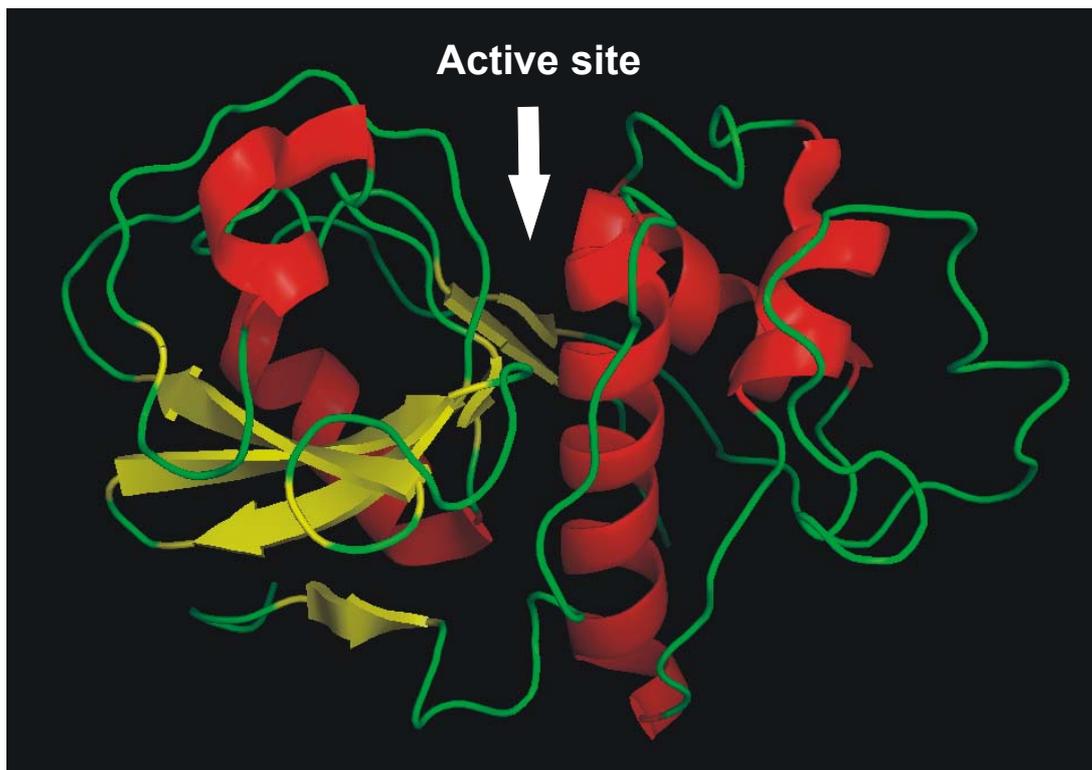


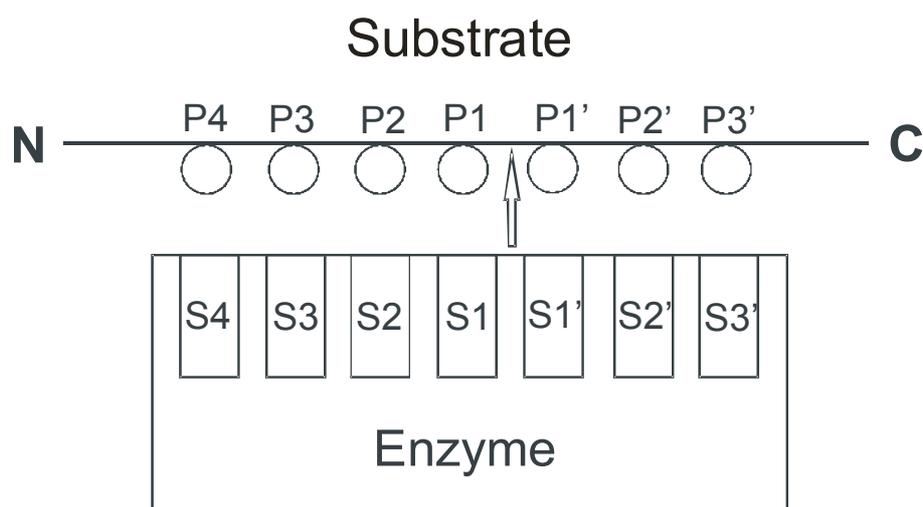
Figure 2.1.1. X-ray structure of papain with the active site labeled. (2PAD.pdb from the MEROPS protease databank.)

Chapter 2 Background

The most extensively investigated plant cysteine protease is papain, found in the latex of the melon tree, after which the largest subgroup (the papain superfamily, clan CA, class C1) in the cysteine protease class was named. The most important mammalian cysteine proteases can be divided into two groups: lysosomal cathepsins and cytoplasmic calpains.

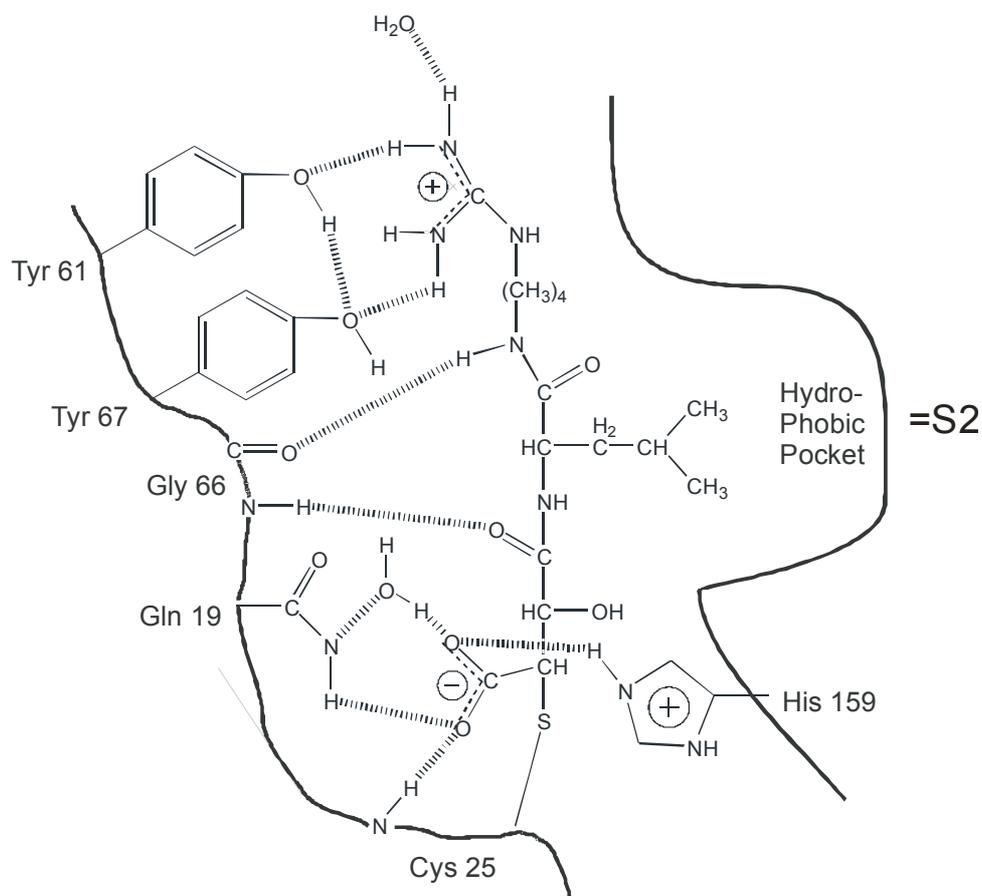
The largest and the most important family of cysteine proteases is the papain family. The members of this superfamily do not all have similar sequences but their molecular architecture is remarkably similar. This typical structure will be explained using papain as an example.

Papain is a monomeric polypeptide consisting of 212 amino acid residues. The X-ray structure of active papain (2PAD.pdb) is depicted, with the active site marked, in Figure 2.1.1. In addition to the sequence and the 3D structure of the lone protein, the X-ray structures of complexes of papain with a number of different inhibitors are also known (E64C, BPACK, CLIK148,...). Schechter and Berger¹⁸ have proposed that the active site of papain consists of seven subsites: S_1 to S_4 bind the four peptidyl substrate residues on the N-terminal side of the scissile bond, while S_1' to S_3' bind the three residues on the C-terminal side. In Scheme 2.1.1 a schematic diagram of the interaction between papain and a substrate is depicted. Scheme 2.1.2 shows the papain – E64c interaction within the active site in more detail.



Scheme 2.1.1. Terminology of papain's active site specificity according to Schechter and Berger.¹⁸

The Cys25 and the His159 (papain numbering) residues together form the most important part of the binding site, as they are directly involved in the cleavage of the covalent bond of the substrate. These two amino acid residues are charged and form a thiolate/imidazolium ion pair, which explains the unusually high reactivity towards electrophilic reagents. The existence of this ion pair is no longer disputed in the literature, but the reason for its occurrence is not yet determined. Another important constituent of the binding pocket is the side chain of the Gln19 residue. This, together with the backbone NH group of Cys25, forms the oxyanion hole in which the oxyanion of the reacting substrate is trapped.



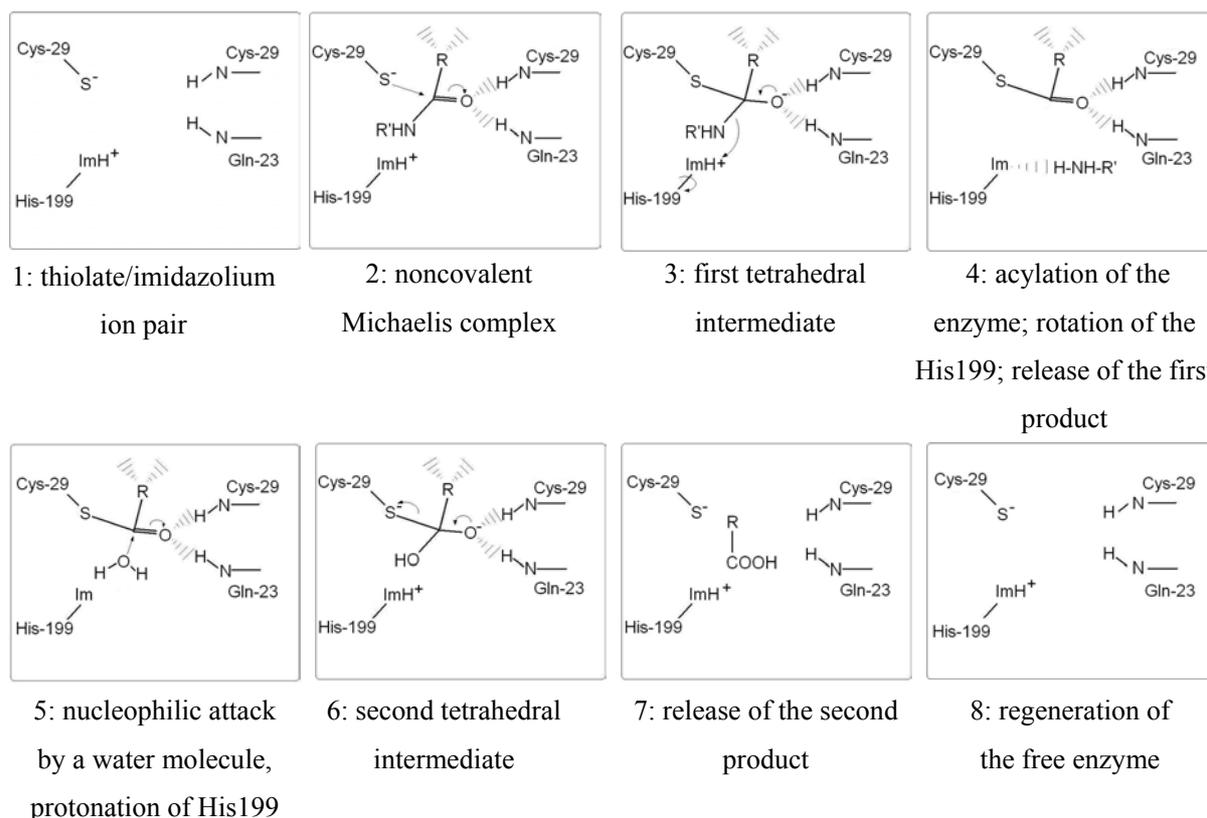
Scheme 2.1.2. Diagram of the active site of the papain–E64c enzyme–inhibitor complex, modified after ref. 19.

The protein structure and, more importantly, the structure of the binding pocket are very similar for all members of the papain family, and hence the proteolysis and inhibition mechanisms also proceed in a similar manner.

In this work the behavior of cysteine proteases is studied by using the cathepsin B structure, and hence all further residue numbering will be given according to this enzyme.

2.1.2 Mechanism of proteolysis

Scheme 2.1.3 depicts the proposed mechanism of proteolysis by cysteine. In the first picture the thiolate/imidazolium ion pair and the oxyanion hole of the active site are shown together. The first step of the proteolysis reaction involves the formation of the non-covalent enzyme–substrate complex. Once this complex is formed and fixed with hydrogen bonds, the negatively charged S atom of the cysteine residue can easily attack the carbon of the reactive peptide bond, producing the first tetrahedral intermediate in the reaction. This intermediate is stabilized by hydrogen bonding between the substrate oxyanion and the oxyanion hole (a proposition actually supported by X-ray structures of the intermediate).²⁰

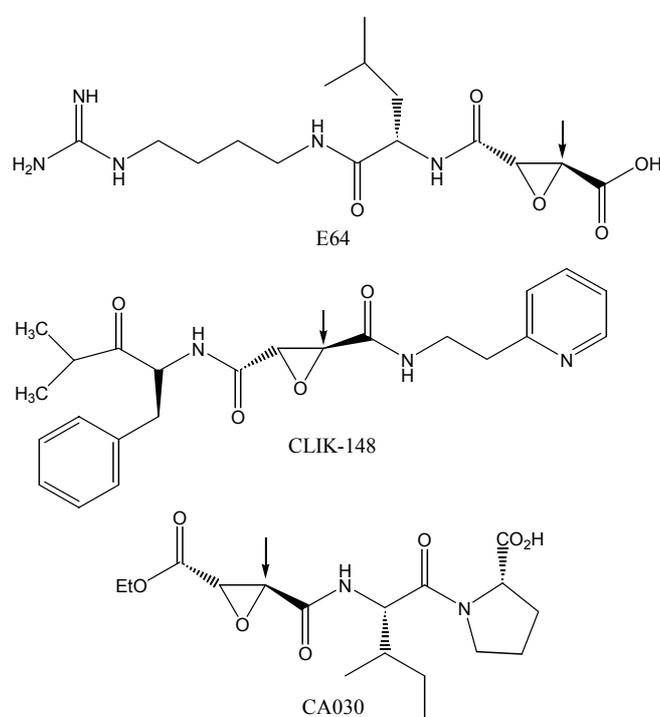


Scheme 2.1.3. Mechanism of proteolysis by cysteine protease (cathepsin B numbering).

Esterification of the thiol makes the protonated His199 ring sufficiently acidic ($pK_a = 4$) to donate its proton to the negatively charged N atom of the leaving group. Rotation of the histidine residue is necessary for this step to occur. In this way the enzyme is acylated. Deacylation can then occur via a typical base-catalyzed mechanism, wherein the imidazole nitrogen polarizes a water molecule which then attacks the acyl-enzyme at the carbonyl carbon forming a second tetrahedral intermediate. This then dissociates into the other half of the cleaved substrate and the regenerated enzyme.

2.1.3 Inhibitors of cysteine proteases

All inhibitors can be classified into two major groups: those which interact with the active site (active site directed) and those which attack another site on the enzyme (allosteric effectors). According to the type of interaction, the active site directed inhibitors can be further divided into covalent/non-covalent and irreversible/reversible inhibitors. Reversible inhibition usually involves non-covalent interaction between the enzyme and the inhibitor, although this is not always the case (e.g. peptidyl aldehydes²¹ or nitriles²²); on the other hand, irreversible inhibition always involves covalent binding of the inhibitor to the enzyme.



Scheme 2.1.4. Low molecular weight cysteine protease inhibitors containing an epoxide ring as an electrophilic warhead.

Inhibitors of cysteine proteases can also be divided into high molecular weight (HMW) and low molecular weight (LMW) inhibitors. Members of the first group have protein structure and are usually naturally occurring cysteine protease inhibitors. The second group is comprised of molecules with proteinic, polypeptidic, or peptidomimetic structure. Most of these are synthetic, and they typically display certain common structural features. These include a peptide segment for recognition by the enzyme (at least two amino acids to achieve a good affinity) and a substitutable electrophilic group to irreversibly react with the cysteine residue of the active site. This latter functionality has to mimic the partially positively charged carbonyl carbon of the enzyme's normal substrate. Examples of such building blocks are three-membered heterocycles like epoxides and aziridines.

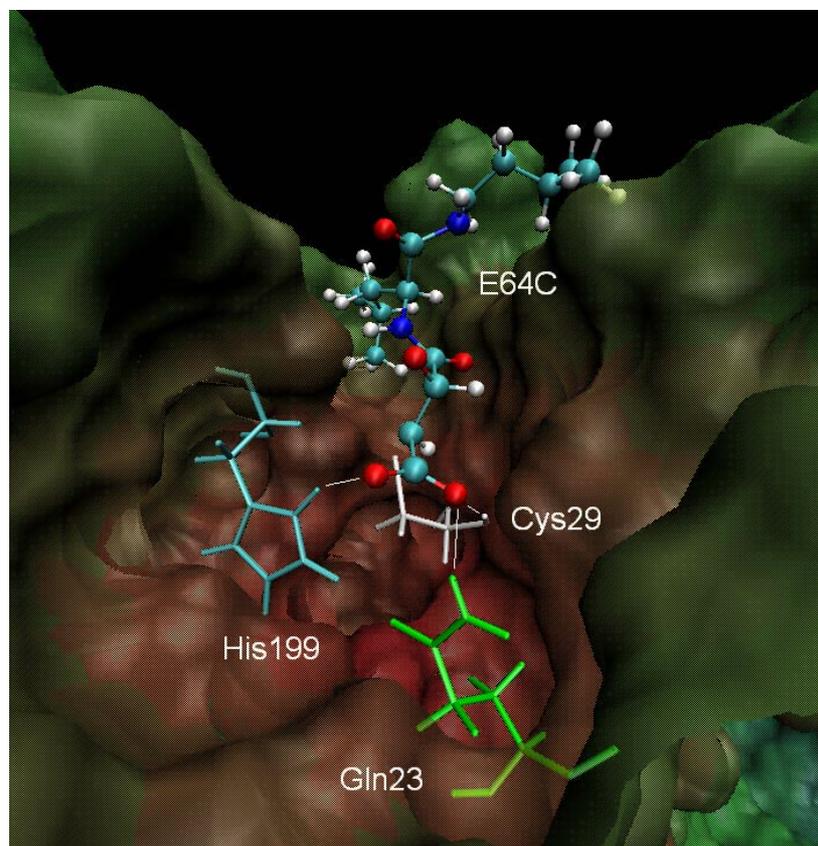
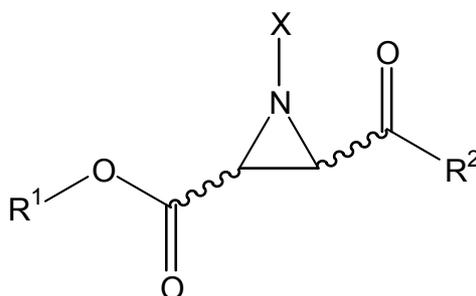


Figure 2.1.2. The E64c inhibitor bound to the active site of cathepsin B.

One of the first cysteine protease inhibitors to exploit the electrophilicity of small heterocyclic rings was the epoxysuccinyl peptide E-64 (Scheme 2.1.4), isolated from *Aspergillus japonicus* in 1978.⁷ Since that time a vast number of structurally-related inhibitors with similar

properties have been discovered (cathestatins,²³ estatins,²⁴ WF14861,²⁵ ...). Although all these compounds are selective inhibitors of the papain superfamily enzymes, they show little or no selectivity towards other cysteine proteases. Overcoming this limitation was the reason for the development of synthetic epoxysuccinyl peptides. The first success came after elucidating the 3D enzyme structures and binding modes of epoxysuccinyl peptides to several cysteine proteases via X-ray crystallography. (X-ray structural data of the papain-E64 enzyme-inhibitor complex is available; a detail is shown in Figure 2.1.2) CA030 (Scheme 2.1.4) and CA074 were the first cathepsin B selective inhibitors.

The inhibition mechanism of epoxysuccinyl peptides with the epoxide ring acting as an electrophilic trap led to development of peptides containing an aziridine ring (Scheme 2.1.5). Due to their lower chemical reactivity these types of inhibitors should be weaker than epoxide-containing ones. However, reduction of the pH of the reaction medium leads to an increase in deactivation activity down to pH 4, whereas the contrary applies for epoxides (most active at pH 6-7). N-unsubstituted aziridines also show an inverted stereoselectivity and are more active with an (*R,R*)-configured aziridine ring.



Scheme 2.1.5. N-acylated aziridines as cysteine protease inhibitors

The differences in the inhibition behavior of aziridines and epoxides can be explained by the difference in properties, such as aziridine's decreased ring strain, enhanced basicity, and potential H-bond donation (when X=H in Scheme 2.1.5). An advantage that the replacement of oxygen with nitrogen brings is the possibility of attaching additional functional groups, including extra peptide chains, to the aziridine nitrogen.

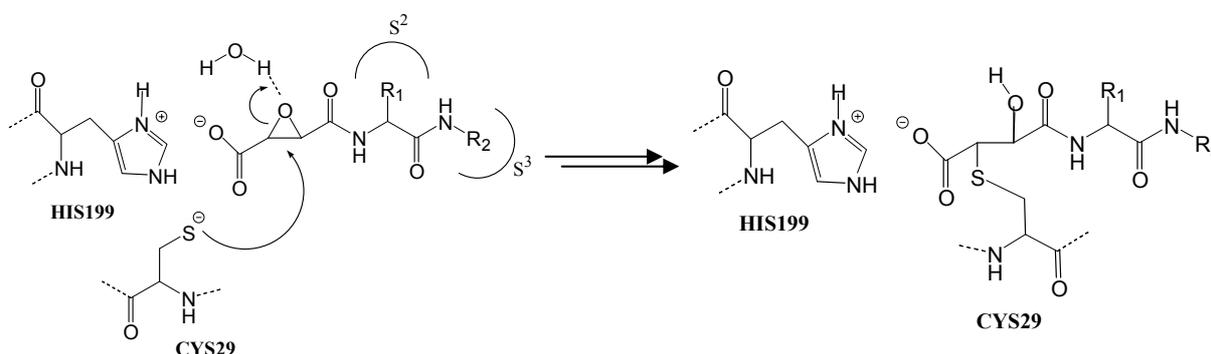
2.1.4 Inhibition mechanism

The reaction of irreversible inhibition of enzymes by such inhibitors can be represented by the following minimal two-step mechanism.



Scheme 2.1.6. Minimal two-step mechanism for irreversible enzyme inhibition.

In Scheme 2.1.7 the inhibition mechanism is depicted in more detail, using cathepsin B as an example.



Scheme 2.1.7. Mechanism of inhibition of cysteine proteases. Left: Non-covalent complex between the enzyme and the inhibitor. Right: Covalently bound inhibitor in the active site of the protease. Cathepsin B numbering is used.

The first step is the reversible formation of an enzyme inhibitor complex (EI). This preliminary step is characterized by the dissociation constant K_i . The efficiency of this step is mainly determined by the peptidic or peptidomimetic sequence. Following this, the negatively charged active site's thiolate attacks the electrophilic three-membered ring. It is generally believed that this is the rate-determining step of the overall inhibition process. A ring-opening

reaction ensues, leading to alkylation of the active site cysteine residue (E-I). As a consequence, the enzyme is irreversibly deactivated. The energy barrier and reaction energy of this ring-opening reaction determine the kinetics (described by the first-order rate constant of inhibition k_i) and thermodynamics of the second step. These quantities are believed to be primarily determined by the chemical nature of the electrophilic warhead, *i.e.* the epoxide or the aziridine itself. The overall inhibition potency of irreversible inhibitors is quantified by the second-order rate constant of inhibition (defined as $k_{2nd}=k_i/K_i$).

2.2 Background on Theoretical Methods Used in This Work

One of the most common problems in computational chemistry is calculating the electronic energy for a given molecular structure. There are two directions in which one can go in trying to solve this problem. On the one side there are *Quantum Mechanical* (QM) methods which, with a certain level of approximation, deal with the explicit electronic configuration of the molecule and energy expressions involving nuclear, electronic, and spin coordinates. On the other side there are so called *Molecular Modeling* (MM) methods which neglect both the electronic structure of the molecule, by writing its electronic energy as a parametric function of nuclear coordinates, and the quantum aspects of nuclear motion, by treating it at the classical level (using Newton's second law). Both methods have their strengths and weaknesses. QM methods deliver more reliable results and can deal with systems in which new bonds between atoms are created or old ones cleaved, but require a lot of computational time and are hence useless for treating big systems like proteins. MM methods are faster, but cannot be used to calculate changes in the molecular structure involving creation and cleavage of bonds, and are less reliable since the result depends on parameters fitted only for certain cases.

2.2.1 Molecular modeling (MM) methods

The basis of MM methods (also called force field methods) is the observation that molecules tend to be composed of units which have similar structure and behavior independently of their molecular environment. This corresponds to the concept of functional groups in organic chemistry. For example, all C=O bonds are approximately 1.21 Å long and have a vibrational frequency of around 1700 cm⁻¹. Another convenient feature of big organic molecules (for the treatment of which force fields were designed in the first place) is that they include in their structure only a handful of chemical elements. Different atom types can be identified for the same element according to the structural unit the atom is featured in, and different parameters can be assigned to each type. For example, a carbon atom could be classified as an sp³-, sp²-, sp-, carbonyl-, cyclopropane carbon, and so forth. The more different atom types a force field includes, the more flexible, and potentially more accurate, it is.

Another assumption that is fundamental for the function of the force field methods is the Born-Oppenheimer approximation, which assumes that the energy can be written as a function of nuclear spatial coordinates. The molecules are modeled as atoms held together by bonds. This ansatz is also known as the “ball and spring” model, since atoms are seen as balls, to which different properties are assigned according to atom type, and bonds are considered as springs, whose stiffness is determined by the atom types involved.

Two things characterize every force field: the form of the expression for the total energy of the system, and the parameters for the different atom types needed to evaluate this expression.

2.2.1.1 Energy terms

The force field energy is written as a sum of at least five terms, each describing the change in one of the five key energy contributions (Figure 2.2.1) arising from a change in atomic coordinates:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} \quad \text{Eq 2.1}$$

The first three terms on the right hand side of equation 2.1 describe interactions between atoms that are connected by bonds, whereas the last two terms account for the so called non-bonded interactions. E_{str} and E_{bend} correspond to stretching and bending vibrations in a molecule, while E_{tors} describes the torsional energy for rotation around a bond. The last two terms describe van der Waals and electrostatic interactions between non-bonded atoms.

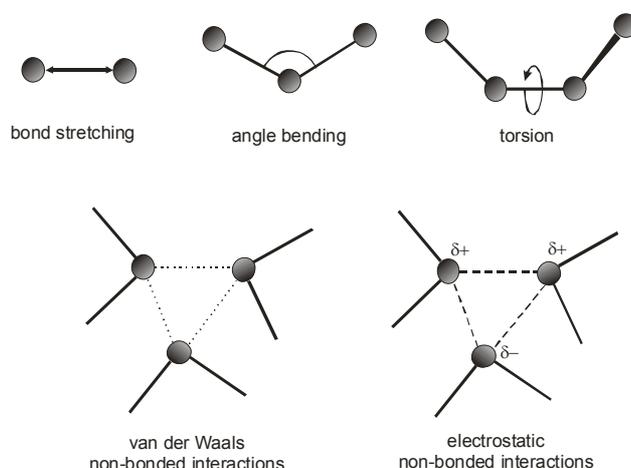


Figure 2.2.1. Schematic representation of the five key contributions to the force field energy.

Using such an energy function, relative energies can be calculated from the positions of atoms in the system.

The energy expression used in molecular mechanics calculations is not necessarily limited to these five terms, however. Contributions to the energy arising from the coupling between stretching, bending, and torsional motions are taken into account by energy cross terms (E_{cross}).

Bond-stretching energy

This part of the total energy covers the energy changes due to bond vibration. At its simplest, this term takes the form of a second order Taylor expansion around an “equilibrium” bond length R_0 (which is not the actual equilibrium distance between the bonded atoms A and B, but a fitted parameter producing results closest to the experimental values.):

$$E_{str}(R^{AB} - R_0^{AB}) = E(R^{AB} = R_0^{AB}) + \frac{dE(R^{AB} = R_0^{AB})}{dR}(R^{AB} - R_0^{AB}) + \frac{1}{2} \frac{d^2E(R^{AB} = R_0^{AB})}{dR^2}(R^{AB} - R_0^{AB})^2 \quad \text{Eq 2.2}$$

The first term is the zero point of the energy scale, so can be set to zero, and the second term is also zero, due to the energy having a minimum at R_0 . Therefore, Eq 2.2 maybe simplified to:

$$E_{str}(R^{AB} - R_0^{AB}) = \frac{1}{2} \frac{d^2E(R^{AB} = R_0^{AB})}{dR^2}(R^{AB} - R_0^{AB})^2 = k^{AB}(\Delta R^{AB})^2 \quad \text{Eq 2.3}$$

where k^{AB} is the force constant for the A-B bond and ΔR^{AB} the displacement from the “equilibrium” bond length. k^{AB} and R_0 are parameters that are specific for every force field and atom type, and that need to be parameterized by using experimental results or calculations at a much higher level of theory.

This simple harmonic form of the stretch energy is sufficient for determining the equilibrium geometry in most cases. However, problems arise when ΔR^{AB} becomes too large. One solution is to include more terms in the Taylor expansion. However, this still has problems, since higher-order polynomial expressions of the stretch energy would still not have the correct

limiting behavior. Furthermore, this would also call for the determination of more parameters. Correct limiting behavior in the case of bond length elongation would be convergence toward dissociation energy. A simple function which satisfies this criterion is the Morse potential, but this has proven problematic in practical application. Luckily, for minimization purposes and simulations at room temperature (300 K) it is sufficient that the potential is reasonably accurate up to ~ 10 kcal/mol above the energy minimum. For this reason, most force fields use either the harmonic approximation, or add more terms to the polynomial expansion.

The bending and out-of-plane bending energy

The bending energy term takes into account energy contributions induced by bending motions of three bonded atoms forming an angle θ . Similarly to the stretching term, bending energy is usually represented by a Taylor expansion around an “equilibrium” angle value up to the second order:

$$E_{bend}(\theta^{ABC} - \theta_0^{ABC}) = k^{ABC} (\theta^{ABC} - \theta_0^{ABC})^2 \quad \text{Eq 2.4}$$

This harmonic approximation works quite well for most applications. Nevertheless, higher accuracy can be achieved by including Taylor series terms of higher order. The simple harmonic approximation is seen to be accurate up to about $\pm 30^\circ$ from the equilibrium geometry, while the cubic one performs well up to $\pm 70^\circ$.

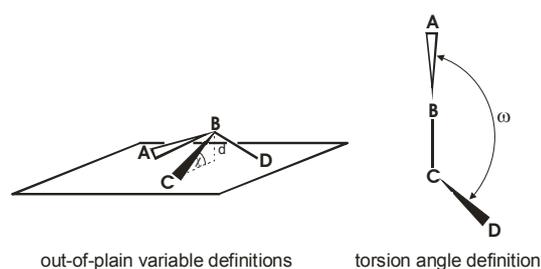


Figure 2.2.2. Definition of variables describing out-of-plane bending and torsion

Out-of-plane bending is the deviation from planarity of four atoms, of which one atom is central and bonded to the other three. This is shown on the left of Figure 2.2.2, with the atom B central and connected to the three atoms A, C, and D. This deviation from planarity can be expressed as the distance d between B and the plane formed by A, C and D, or the angles χ

between the plane and a B-X bond. The equilibrium value of d is zero for sp^2 -hybridized centers, which prefer planarity, and nonzero for, e.g., NH_3 . If d deviates from its equilibrium value, this raises the energy of the system. This is partially accounted for by the change in bending energy caused by the change in the angles $\angle A-B-C$, $\angle C-B-D$, and $\angle B-D-A$. However, this does not completely cover the energy change, and an extra energy term has to be added, the out-of-plane bending energy. E_{oop} is typically taken as having quadratic dependence on d or χ (in which case there is a contribution to E_{oop} for each of the three atoms A, C, and D):

$$E_{oop}(\Delta d^B) = k^B (\Delta d^B)^2 \quad \text{Eq 2.5}$$

Here Δd^B stands for the deviation between d^B and its equilibrium value.

The torsional energy

The torsional energy term describes the change in overall energy arising from the torsional angle ω , which describes the relative rotation of atoms A and D around bond B-C, with the four atoms connected in the sequence A-B-C-D (Figure 2.2.2, right). There are two major differences between E_{tors} and E_{stre} or E_{bend} . First, the change in energy due to torsion has to be periodic, since in many cases ω has to be free to vary from 0° to 360° (or -180° to 180°), and second, large deviations from the equilibrium geometry do not cause large energy variations if the B-C bond is a single bond. For these two reasons the use of Taylor series is not an efficient solution, and one must turn to Fourier series:

$$E_{tors}(\omega) = \sum_{n=1} V_n \cos(n\omega) \quad \text{Eq 2.6}$$

where n describes the period of the rotation, usually set to be 2π or π , and V_n determines the height of the barrier to rotation around the B-C bond.

The van der Waals energy

The van der Waals energy term describes the interaction between two atoms A and B that are not directly bonded. One can picture it as an interaction between the electron clouds of the atoms, but strictly speaking, they are transient, induced dipole-dipole, quadrupole-dipole, quadrupole-quadrupole, etc. interactions. At large distances the interaction is equal to zero. As

the atoms move closer they start to attract each other and the potential energy becomes negative, till the equilibrium distance between the two atoms is reached. If the atoms move still closer towards each other they start to repel, and the potential energy rises steeply and asymptotically with decreasing distance. As dispersion interactions decay as R^{-6} at long distances, one can express E_{vdw} in the following form:

$$E_{vdw}(R^{AB}) = E_{repulsive}(R^{AB}) - \frac{C^{AB}}{(R^{AB})^6} \quad \text{Eq 2.7}$$

The expression for the repulsive part cannot be derived theoretically, but it is known that the energy increases much more steeply on the repulsive side of the potential than on the attractive side. One of the most commonly used potentials which fits all these requirements is the Lennard-Jones (L.-J.) potential:

$$E_{LJ}(R^{AB}) = \varepsilon^{AB} \left[\left(\frac{R_0^{AB}}{R^{AB}} \right)^{12} - 2 \left(\frac{R_0^{AB}}{R^{AB}} \right)^6 \right] \quad \text{Eq 2.8}$$

where R_0^{AB} is the equilibrium distance (van der Waals distance) and ε^{AB} is the “softness” of the atoms, both of which depend on the atom types involved. In order to simplify the parameterization, the van der Waals distance between atoms A and B is taken as the sum of van der Waals radii for atom A and atom B, and the mutual “softness” is approximated by a square root of the product of the individual, atomic “softnesses”.

The electrostatic energy

The other part of the non-bonded interactions results from coulombic interaction between charged parts of the molecule. The charge distribution of the whole molecule can be described by assigning partial charge to each atom; for instance, in a carbonyl group oxygen would be slightly negatively charged, while the carbon is positively charged. Alternatively, the whole bond can be treated as a dipole, and a dipole moment assigned to it. These charges and dipole moments are determined by calculations at a much higher level of theory. The interaction between the point charges is described by a Coulomb potential:

$$E_{el}(R^{AB}) = \frac{Q^A Q^B}{\varepsilon R^{AB}} \quad \text{Eq 2.9}$$

where ε is a dielectric constant. If the dipole moments of the bond are used instead of atomic charges, the following energy expression is used:

$$E_{el}(R^{AB}) = \frac{\mu^A \mu^B}{\varepsilon (R^{AB})^3} (\cos \chi - 3 \cos \alpha_A \cos \alpha_B) \quad \text{Eq 2.10}$$

where the angles are as depicted in Figure 2.2.3.

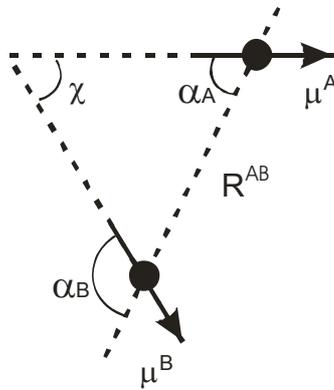


Figure 2.2.3. Definition of important variables for dipole-dipole interaction

2.2.2 Quantum mechanical (QM) methods

2.2.2.1 The basics

The basis of all QM methods is the attempt to find the solutions to the Schrödinger equation, which in its time-independent non-relativistic form is:

$$\hat{H}\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, s_1, s_2, \dots, s_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) = E\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, s_1, s_2, \dots, s_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) \quad \text{Eq 2.11}$$

where \hat{H} is the Hamilton operator (Hamiltonian) for the system in the absence of electrostatic or magnetic fields and R_M , r_N , and s_N are nuclear, electron, and spin coordinates, respectively. Ψ is the wavefunction containing all the information about the observed system, including its total energy, E . The Hamiltonian is the differential operator for the total energy:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \frac{1}{M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad \text{Eq 2.12}$$

where A and B denote the M nuclei (with the mass M) in the system and i and j the N electrons, Z is the nuclear charge, while R and r represent, respectively, the internuclear and the electron-electron or nucleus-electron distances. The Hamiltonian has been formulated here using atomic units. The Laplacian operator involved in the expressions for the nuclear and the electronic kinetic energy (the first two terms) is given (in Cartesian coordinates) as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad \text{Eq 2.13}$$

The other three terms of Eq. 2.12 represent electrostatic interactions between nuclei and electrons and constitute the potential energy part of the Hamiltonian.

The total energy of the system is computed as the expectation value of the Hamilton operator:

$$\langle E \rangle = \int \dots \int \Psi^* \hat{H} \Psi dV \equiv \langle \Psi | \hat{H} | \Psi \rangle \quad \text{Eq 2.14}$$

In Eq 2.14 dV represents integration over all coordinates. The right-hand part of the equation shows the Dirac notation, which is a briefer way of writing wavefunction-related integrals.

Nuclei are much heavier than electrons, and therefore move much slower; for this reason, we can make the approximation that the nuclear positions are fixed and the electrons are moving in the potential energy well of the nuclei. In this case, the nuclear kinetic energy is zero and the repulsion between the nuclei merely a constant. This is the well-known Born-Oppenheimer approximation, which leads to the reduction of the complete eigenvalue problem (Eq 2.12) to the electronic eigenvalue problem:

$$\hat{H}_{elec} \psi_{elec} = E_{elec} \psi_{elec} \quad \text{Eq 2.15}$$

where ψ_{elec} is the electronic wavefunction which depends only parametrically on the nuclear coordinates. The total energy of the system is then the sum of the electronic energy E_{elec} and the constant nuclear repulsion term.

The goal of every QM calculation is to find a wavefunction of the system, typically the ground state wavefunction, that is an eigenfunction of the Hamiltonian. Once this wavefunction is obtained, all other properties of the system can be extracted from it by applying the appropriate operators. However, the exact solution of the eigenvalue problem 2.14 is possible only for one-electron systems, which have very little practical relevance. Further approximations are necessary to treat multielectron systems.

The wavefunction of the system in the ground state can be systematically approached starting from any acceptable (continuous and quadratic-integrable) trial wavefunction by using the famous variational principle, which states that the energy expectation value of any approximate wavefunction Ψ_{trial} will be an upper bound to the true energy of the ground state:

$$\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle = E_{trial} \geq E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \quad \text{Eq 2.16}$$

where the equality $E_{trial} = E_0$ holds if and only if $\Psi_{trial} = \Psi_0$ (to within a unitary transformation). The result of this principle is that the ground-state may be found by minimizing the energy E_{trial} by searching through acceptable N -electron wavefunctions until the lowest possible energy is reached, at which point the exact E_0 and Ψ_0 have been found. For practical reasons it is only possible to search within a subset of possible wavefunctions, and hence one can only obtain the best approximation to the exact wavefunction possible within that subset. In the Hartree-Fock approximation, the cornerstone of almost all wavefunction-based quantum chemical methods, this subset consists of single-determinantal wavefunctions (called Slater determinants, Eq 2.17).

$$\Psi_0 \approx \Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\bar{x}_1) & \chi_2(\bar{x}_1) & \cdots & \chi_N(\bar{x}_1) \\ \chi_1(\bar{x}_2) & \chi_2(\bar{x}_2) & \cdots & \chi_N(\bar{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\bar{x}_N) & \chi_2(\bar{x}_N) & \cdots & \chi_N(\bar{x}_N) \end{vmatrix} \quad \text{Eq 2.17}$$

where

$$\chi(\bar{x}) = \phi(\vec{r})\sigma(s) \quad \sigma = \alpha, \beta \quad \text{Eq 2.18}$$

represents a spin-orbital, which is constructed out of a spatial $\phi(\vec{r})$ and a spin $\sigma(s)$ (having either α or β spin) part.

In the Hartree-Fock method the Hamiltonian 2.12 is applied to a wavefunction of the type shown in Eq 2.17. After applying the Born-Oppenheimer approximation, this yields the following energy expression:

$$E = \sum_i^N \langle \phi_i | \mathbf{h}_i | \phi_i \rangle + \frac{1}{2} \sum_{ij}^N \left(\langle \phi_j | \mathbf{J}_i | \phi_j \rangle - \langle \phi_j | \mathbf{K}_i | \phi_j \rangle \right) + V_{nn} \quad \text{Eq 2.19}$$

in which:

\mathbf{h} is a one-electron Hamilton operator,

\mathbf{J} is a Coulomb operator, which corresponds to the classical electrostatic repulsion between two electrons, and

\mathbf{K} is an exchange operator, which has no classical analogy.

\mathbf{J} and \mathbf{K} are two-electron operators which act on orbitals as follows:

$$\begin{aligned} \mathbf{J}_i | \phi_j(2) \rangle &= \langle \phi_i(1) | \mathbf{g}_{12} | \phi_i(1) \rangle | \phi_j(2) \rangle \\ \mathbf{K}_i | \phi_j(2) \rangle &= \langle \phi_i(1) | \mathbf{g}_{12} | \phi_j(1) \rangle | \phi_i(2) \rangle \end{aligned} \quad \text{Eq 2.20}$$

The next step is to determine a set of MOs which would provide a minimal (or at least stationary) energy for the system. This is usually done in a constrained optimization by using Lagrange multipliers (λ_j) and imposing the condition that the Lagrange function (L) must remain stationary with respect to changes in the orbitals:

$$\delta L = \delta E - \sum_{ij}^N \lambda_{ij} \left(\langle \delta \phi_i | \phi_j \rangle + \langle \phi_i | \delta \phi_j \rangle \right) = 0 \quad \text{Eq 2.21}$$

The energy variation, after introducing the Fock operator \mathbf{F}_i , has the form:

$$\delta E = \sum_i^N \left(\langle \delta \phi_i | \mathbf{F}_i | \phi_i \rangle + \langle \phi_i | \mathbf{F}_i | \delta \phi_i \rangle \right) \quad \text{Eq 2.22}$$

with

$$F_i = h_i + \sum_j^N (J_j - K_j) \quad \text{Eq 2.23}$$

The Fock operator is an one-electron operator. It describes the kinetic energy of an electron and its electrostatic interactions with the nuclei and other electrons .

After combining Eq 2.21 and 2.22 and diagonalization of the matrix of Lagrange multipliers, one obtains a set of pseudo-eigenvalue equations known as the Hartree-Fock equations:

$$F_i \phi_i' = \varepsilon_i \phi_i' \quad \text{Eq 2.24}$$

where ϕ_i' are called canonical MOs and ε_i could be interpreted as the corresponding molecular orbital energies. Since, in order to determine one of the Fock orbitals, one needs to know the form of all others, the application of an iterative method is necessary. Orbitals which are a solution to equation 2.24 are called Self-Consistent Field (SCF) orbitals.

In practice one varies MOs by expanding them in terms of atom-centered basis functions, i.e. by representing them as a linear combination of atomic functions $\phi_i = \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha}$ (where M is the number of basis functions). After introducing such a basis set and multiplying from the left by a specific basis function and integrating, Eq 2.24 becomes:

$$FC = SC\varepsilon \quad \text{Eq 2.25}$$

which is the Roothaan-Hall equation. The **F** matrix contains the Fock matrix elements and the **S** matrix the overlap integrals between the basis functions. The **C** matrix contains the coefficients of the expansion of the molecular orbital into the atomic basis. These coefficients are varied in an iterative procedure in order to obtain the minimum energy. The procedure begins by making an initial estimate for the C matrix; the **F** matrix is formed and diagonalized, yielding a new set of (improved) coefficients, which are then used to construct a new Fock matrix, which is diagonalized again, and so on. The procedure is continued until the change in the coefficients from one iteration to the next drops below a specified convergence criterion.

The importance of the Hartree-Fock method is that it serves as a baseline, which may be modified either by introducing further approximations to decrease computational cost (as in semi-empirical methods), or by adding more determinants to improve the accuracy of the calculation (as in MP2, CI).

2.2.2.2 Semi-empirical methods

As the size of the molecule increases, leading to an expansion of the basis set, the cost of HF calculations rises steeply (varying as $\sim N^4$, where N is the number of basis functions), due to the calculation of two-electron integrals. In semi-empirical methods the number of explicitly calculated two-electron integrals is reduced, and with it, the computational costs.

Semi-empirical methods can be classified into several groups according to the approximations they make. The method used in this work is Austin Model 1 (AM1),²⁶ and in this chapter we will describe only this model and other models of the same type (modified (parameterized) neglect of diatomic differential overlap approximation, or NDDO, models).

All semi-empirical models are based on two fundamental approximations:

1. Only the valence electrons are explicitly treated, by using a minimal basis set, while the core electrons are combined with the nucleus to form a single electrostatic potential. This approximation demands either reduction of the nuclear charge or the introduction of spatial functions to describe the combined electrostatic effect of the nucleus and core electrons.
2. All products of basis functions depending on the same electron coordinates, but on different atoms, are neglected (the zero differential overlap approximation), i.e. $\mu_A(i) \cdot \nu_B(i) = 0$, where μ and ν are basis functions, i the electronic coordinates, and A and B are atoms. The consequences of this approximation are that the overlap matrix becomes a unitary matrix, one-electron integrals involving three centers (one of the operator and two of the basis) are set to zero, and all three- and four-center two-electron integrals are neglected.

In order to compensate for these approximations the remaining integrals are made into parameters, which are fitted to either calculations on a higher level of theory or experimental results.

These two approximations by themselves comprise the neglect of diatomic differential overlap (NDDO) model. The most popular semi-empirical methods, MNDO,²⁷ AM1,²⁶ and PM3,²⁸ are parameterizations of NDDO. The parameters involved are atomic variables describing the properties of single atoms. These three methods differ in the way the core-core repulsion is treated and how the parameters are assigned.

In the MNDO (modified neglect of diatomic overlap) method the core-core repulsion has the form:

$$V_{nn}^{MNDO}(A, B) = Z'_A Z'_B \langle s_A s_B | s_A s_B \rangle (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad \text{Eq 2.26}$$

where α are the fitted parameters. Z'_A and Z'_B are the effective nuclear charges (as screened by the core orbitals) of atoms A and B, and s are the s -type basis functions.

The main drawbacks of MNDO are: overestimation of the nuclear repulsion between atoms for interatomic distances of 2-3 Å. As a consequence, this method is not able to predict H-bonds and overestimates the activation energies in reactions involving cleavage and formation of bonds.

Correction of the MNDO core-core repulsion term by adding Gaussian functions and reparameterization of the whole model yields the Austin Model 1 (AM1) method. In this model the core-core repulsion has the form:

$$V_{nn}^{AM1}(A, B) = V_{nn}^{MNDO}(A, B) + \frac{Z'_A Z'_B}{R_{AB}} \times \left(\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2} \right) \quad \text{Eq 2.27}$$

Where a_k , b_k , and c_k are constants fitted to molecular data.

AM1 delivers more accurate activation energies than MNDO and can predict approximately correct H-bond strengths, but it often delivers the wrong minimum energy geometry.

2.2.2.3 Density functional theory (DFT)

DFT is an alternative method to wavefunction-based approaches. The basic concept of the method is provided by the theorem of Kohn and Hohenberg which states that the electrons in a system feel an external potential that is uniquely determined by the electron density, *i.e.* that the potential is the functional of the density and not of the wavefunction. A functional is defined as a prescription to obtain a number from a function of certain variables. It is to be distinguished from a function, which is a prescription to obtain a number for certain values of a set of variables. In order to differentiate these two, the functional is denoted as $F[f]$, where f is the function acted upon.

If the exact expression for this functional were possible to determine, it could be used to calculate electronic structure without having to use the operator-based formalism and would deliver the exact solution without having to use, for example, full CI. The exact universal functional is not known. However, several approximations for it exist.

The earliest method, which used the electron density as the sole variable and did not refer to the wavefunction at all, was invented by Thomas and Fermi in the late 1920s.^{29, 30} The Thomas-Fermi theory had a big drawback in that it used drastic approximations to construct the energy functional. One of the consequences was that molecules had higher energies than the separate atoms they were composed of.

In the 1960s, Kohn and Sham came up with an alternative approach that explicitly constructs the electron density from a single-determinantal wavefunction. The orbitals of the Kohn-Sham wavefunction are optimized within the external potential provided by a total electron density, without direct reference to each other. The KS wavefunction is not the real wavefunction, but a model intended to give the same electron density as the real system, and therefore the same energy under application of the universal functional.

The energy of the system can be divided into kinetic, $T[\rho]$, and potential, $U[\rho]$ parts, both of which are functionals of the electron density:

$$E_{DFT}[\rho] = T[\rho] + U[\rho] \quad \text{Eq 2.28}$$

In the Kohn-Sham theory kinetic energy is split into two further terms:

$$E_{DFT}[\rho] = T_S[\rho] + \Delta T[\rho] + U[\rho] \quad \text{Eq 2.29}$$

$T_S[\rho]$ describes the kinetic energy if the electrons do not interact, while $\Delta T[\rho]$ is a small correction for the fact that the electrons do interact.

The total kinetic energy of the system cannot be exactly determined from the density. However, $T_S[\rho]$ can be easily calculated by applying the quantum mechanical kinetic energy operator to a single-determinantal wavefunction corresponding to the exact density. The ease of calculation of this term is one of the main advantages of the KS method.

The potential energy term can be represented as a sum of three terms: the electrostatic attraction energy between the nuclei and the electrons V_{ne} , the repulsion between electrons V_{ee} , and a correction term V_{XC} called the exchange-correlation energy. The latter cannot be calculated exactly.

$$E_{DFT}[\rho] = T_S[\rho] + \Delta T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] + V_{XC}[\rho] \quad \text{Eq 2.30}$$

The terms in the Eq 2.30 that cannot be determined exactly ($\Delta T[\rho]$ and $V_{XC}[\rho]$) are together referred to as the exchange-correlation functional, and their principal function is as a correction to the electrostatic electron-electron energy. This correction covers the unphysical self-interaction of the electrons arising from the fact that V_{ee} describes the repulsion between an electron and the whole diffuse electron density cloud (corresponding to all N electrons), whereas in the reality an electron interacts only with the other $N-1$ electrons. This term also takes into account two correlation effects which cause the instantaneous positions of the electrons to correlate to avoid each other, thus decreasing the electron-electron repulsion energy below that calculated for electrons as delocalized waveforms. Fermi correlation (or exchange interaction) causes electrons of the same spin to avoid each other, and is a purely quantum mechanical effect arising from the fact that electrons are Fermions (i.e. have half-integer spin). Coulomb correlation occurs between all electrons, due to Coulomb repulsion causing electrons to avoid each other.

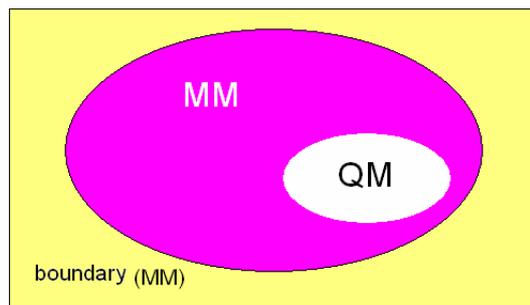
2.2.3 Combined quantum mechanics / molecular modeling (QM/MM) methods

Molecular systems such as organic molecules in solution or biological macromolecules have been the subjects of many theoretical studies. Computational studies have become especially useful in studying protein behaviour since X-ray diffractometry enabled the elucidation of many protein structures on an atomic level. Due to the size of such systems, which usually include not only the macromolecule, but also explicitly-simulated solvent molecules, the first choice of method would be molecular modeling. However, although this method allows reasonably fast calculations on large systems it has a major drawback in being useless for studies on reactivity, since it does not support the simulation of bond formation or bond breaking. A reaction between an enzyme and a substrate at an active site involving covalent interaction therefore has to be treated with a QM method. However, the rest of the system cannot be neglected, since the structure and molecular properties of the surroundings play a very important role in the reaction.

In 1976 Warshel and Levitt introduced the idea of hybrid quantum mechanical/molecular mechanical (QM/MM) method.³¹ They treated a small portion of a system quantum mechanically while the rest, which did not need such detailed description, was represented by an empirical force field.

2.2.3.1 General principles

The fundamental idea behind this method is that the system is partitioned into several parts which are modeled at different levels of theory (Scheme 2.2.4). The atoms of the residues or molecules that are participating in the reaction are treated quantum mechanically. The rest of the system, which usually includes protein environment and solvent, is treated with standard force field methods. Finally a boundary region is defined to account for the finite size of the system.



Scheme 2.2.4. The fundamental idea behind the QM/MM approach: the system is divided into several parts treated at different levels of theory.

The energy of the system is obtained by solving the time-independent Schrödinger equation of the form:

$$\hat{H}_{eff} \Psi(r_i, R_\alpha, R_M) = E(R_\alpha, R_M) \Psi(r_i, R_\alpha, R_M) \quad \text{Eq 2.31}$$

where the following notation is used:

H_{eff} – effective Hamiltonian,

R_α and r_i – nuclear and electron coordinates in the QM region, respectively,

R_M – coordinates of the atoms in the MM region.

For the case when the system is partitioned into three regions (QM, MM, and boundary) the effective Hamiltonian can be broken down in four terms:

$$\hat{H}_{eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} + \hat{H}_{boundary} \quad \text{Eq 2.32}$$

The form of the first term depends on the quantum mechanical method used in the calculations, and for the case of a simple *ab initio* method it would be (in atomic units):

$$\hat{H}_{QM} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{ij} \frac{1}{r_{ij}} - \sum_{i\alpha} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} \quad \text{Eq 2.33}$$

where indexes i, j refer to electrons and α, β to nuclei, Z is the nuclear charge, and r is the distance between particles.

The MM Hamiltonian includes only the interactions between the atoms of the MM region, and can therefore be replaced by the standard expression for the energy of the force field used:

$$\hat{H}_{MM} = E_{MM} = E_{str} + E_{bend} + E_{tors} + E_{vdW} + E_{el} \quad \text{Eq 2.34}$$

The form of the QM/MM Hamiltonian will depend on the forms of the separate QM and MM Hamiltonians. For the case where there are only electrostatic and Lennard-Jones interactions between the QM and MM regions, a possible Hamiltonian would be:

$$\hat{H}_{QM/MM} = -\sum_{iM} \frac{q_M}{r_{iM}} + \sum_{\alpha M} \frac{Z_{\alpha M}}{r_{\alpha M}} + \sum_{\alpha M} \left\{ \frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^6} \right\} \quad \text{Eq 2.35}$$

The first two terms on the right hand side of the equation 2.23 describe the electrostatic (Coulombic) interaction between the electrons and nuclei of the QM region and atoms of the MM region. The third term is the Lennard-Jones term which takes into account the close-range exchange-repulsion and long-range dispersion interactions which cannot be treated by a simple Coulomb interaction expression.

$H_{boundary}$ in Eq 2.20 is the boundary Hamiltonian and can take any of the forms usually used in pure MM calculations, e.g. periodic boundary conditions. This term typically consists of two parts, the QM and MM contributions.

The total energy of the system is obtained by calculating the expectation value of the wavefunction and the effective Hamiltonian, and for the Hamiltonian 2.20 this is:

$$E_{tot} = \frac{\langle \Psi | \hat{H}_{QM} + \hat{H}_{QM/MM} + \hat{H}_{boundary(QM)} | \Psi \rangle}{\langle \Psi | \Psi \rangle} + E_{MM} + E_{boundary(MM)} \quad \text{Eq 2.36}$$

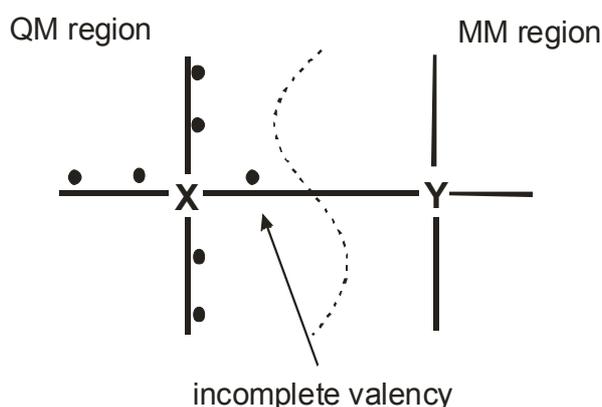
Once the total energy is known, the forces on the atoms are easily derived.

$$F_{\alpha} = \frac{\partial E_{tot}}{\partial R_{\alpha}} \quad F_M = \frac{\partial E_{tot}}{\partial R_M} \quad \text{Eq 2.37}$$

The total energy and the derived forces can then be employed for geometry optimization, molecular dynamics, or Monte Carlo simulations in exactly the same way as for purely QM or MM computations.

2.2.3.2 Treating covalent bonds between the QM and MM regions

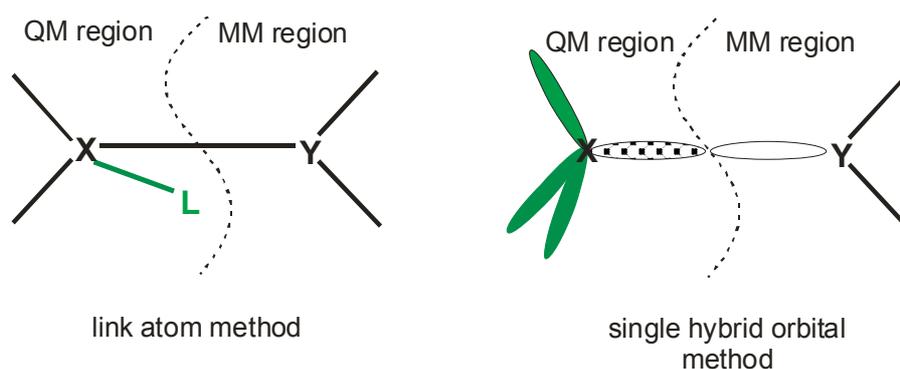
One of the biggest problems in QM/MM calculations is treating the boundary region between the QM and MM parts. If a small molecule in solution is treated then the straightforward separation of the system into QM-molecule and MM-solvent part is possible and no covalent bonds have to be “broken”. But if the studied system is an enzyme in solution, and the reaction of interest takes place at the active site of the enzyme, then only one small part of the enzyme can be treated quantum mechanically and some covalent bonds have to be cut by the boundary between two regions. Let us consider a frontier bond between a QM atom X and an MM atom Y. A problem occurs because the electron of X involved in the covalent bond with Y does not have its pair, since in the MM region the electronic structure is not explicitly described (Scheme 2.2.5). The various approaches have been suggested in the literature, which can be divided into two groups (Scheme 2.2.6): those which involve addition of an atom or a pseudoatom, and those that deal specifically with the frontier bond orbital.



Scheme 2.2.5. Schematic representation of the problem occurring when covalent bonds cross the QM/MM boundary region.

Link and connection atom method

The first and the simplest method to be implemented was the link atom method.³² This involves adding a monovalent atom (a link or dummy atom) along the frontier bond to fill the valency hole of X (left hand side of Scheme 2.2.6). Usually a hydrogen-like atom is used, but there are also programs which use halogen-like atoms (chlorine and fluorine). The main advantage of the link atom method is its easy implementation in the current quantum chemistry codes and its reliability in providing accurate results, as long as the frontier bonds are far enough from the atoms involved in the reaction. However, its main disadvantages are: *i.*) that a C-H bond will have a slightly different effect on the rest of the quantum subsystem than the bond it replaces, and *ii.*) additional degrees of freedom that occur with its implementation.



Scheme 2.2.6. Two alternative methods to treat the junction between the QM and MM region.

To avoid these disadvantages, some authors have suggested a so-called connection atom method, in which the classical atom Y is replaced in the quantum calculation by a monovalent pseudoatom that is parameterized to reproduce the behavior of the X-Y bond. AM1 and PM3³³ parameters, as well as a density functional theory (DFT)³⁴ pseudopotential, have been developed for a monovalent atom capable of representing covalent frontier bond properties. The main advantage of this method in comparison to the link atom approach is that no additional atoms have to be added to the system, as the Y atom exists as both a quantum and a classical atom. On the other hand, this method requires parameterization to be done for each type of covalent frontier bond.

LSCF and generalized hybrid orbital method

Building on the pioneering work of Warshel and Levitt, Rivail *et al.*³⁵ developed the so-called LSCF (local self-consistent field) method, in which the two electrons of the frontier bond inhabit a strictly localized bond orbital (SLBO). As in the case of link or connection atoms, such a bond orbital has to be far enough from the reactive atoms that its deficiencies (compared to a real QM bond) do not affect the reaction. The SLBO is included among the molecular orbitals of the quantum mechanical subsystem and kept frozen. Molecular orbitals for the rest of the quantum subsystem are kept orthogonal to the SLBO and optimized using the local self-consistent procedure. This approach avoids the use of dummy atoms and properly describes the chemical properties of the frontier bond, but is more difficult to implement at the *ab initio* level.

Gao *et al.*³⁶ developed an extension to the LSCF method called the generalized hybrid orbital method (GHO). In this method the classical frontier atom is described by two sets of orbitals: auxiliary and active orbitals. The latter set is included in the SCF calculations, while the former generates an effective core potential for the frontier atom. The advantages and disadvantages are similar to those of the LSCF method, with the exception that the GHO method introduces new semiempirical parameters necessary to describe both the auxiliary and active orbitals.

2.2.3.3 Coupling of the QM and MM regions

The basic concept of the QM/MM approach is the division of the system into two parts: an electronically important reactive core treated quantum mechanically, and everything outside, which is described by molecular mechanics methods. The influence of the MM region on the reactive core may be mainly mechanical (the classical part forcing the reactive part into a particular geometry), but it may also include electrostatic and polarization effects. Three different coupling modes between the QM and the MM region have accordingly been suggested by Bakowies and Thiel³⁷: mechanical embedding, electrostatic embedding, and polarized embedding. In the following text these three embedding schemes will be discussed for a case in which the valency hole is filled with the link atom. As in Scheme 2.2.5, X denotes the atoms in the QM part, whereas Y stands for the MM atoms; the link atom is denoted by L.

Mechanical embedding

The total energy of the QM/MM system, which implicitly includes a molecular mechanical correction for the energetic contributions due to the link atom, can be represented as follows:

$$E_{tot}(X-Y) = E_{MM}(X-Y) - E_{MM}(X-L) + E_{QM}(X-L), \quad \text{Eq 2.38}$$

where:

$$E_{MM}(X-Y) = E_{MM}(X) + E_{MM}(X,Y) + E_{MM}(Y), \quad \text{Eq 2.39}$$

$$E_{MM}(X-L) = E_{MM}(L) - E_{MM}(X,L) + E_{MM}(X). \quad \text{Eq 2.40}$$

$E_{MM}(X,Y)$ represents the intermolecular interaction energy between X and Y atoms calculated using the MM approach. (X-L) represents the interaction of the QM atoms with the link atom. By combining Eq 2.38-2.40 one obtains a simplified formula which avoids the double evaluation of the energy terms for X:

$$E_{tot}(X-Y) = E_{MM}(Y) + E_{MM}(X,Y) + E_{QM}(X-L) + E_{link} \quad \text{Eq 2.41}$$

where the link atom correction is:

$$E_{link} = -E_{MM}(L) + E_{MM}(X,L). \quad \text{Eq 2.42}$$

The preceding derivation is only formally valid, since the term $E_{MM}(X-Y)$ is not always well-defined. If the QM subsystem X undergoes a chemical reaction, it cannot be described classically, and calculation of $E_{MM}(X)$ breaks down. However, this term, due to cancellation, does not appear in Eq 2.29.

In the mechanical embedding model all interactions between the QM and MM regions are included in $E_{MM}(X,Y)$ and are thus calculated molecular mechanically.

Electrostatic embedding

In this approach the electrostatic interactions between the QM and MM part are represented as interaction between point charges and treated quantum mechanically. Atomic point charges can explicitly be included in the Hamiltonian for X-L interactions:

$$\hat{H}_{el}(X-L;Y) = \hat{H}_{el}(X-L) - \sum_i^X \sum_J^Y \frac{q_J}{r_{iJ}}. \quad \text{Eq 2.43}$$

$H_{el}(X-L;Y)$ and $H_{el}(X-L)$ are the electronic Hamiltonians for X-L with and without the external field, respectively, which is generated by the MM point charges q_J . The former includes no interactions between link and MM atoms.

The inclusion of point charges changes the form of the core Hamiltonian $h_{\mu\nu}$ in the LCAO approximation:

$$\tilde{h}_{\mu\nu} = h_{\mu\nu} - \sum_J^Y q_J V_{\mu\nu}^J. \quad \text{Eq 2.44}$$

The resulting electrostatic (Coulomb) energy is given by:

$$E^{coul}(X,Y) = \sum_{\mu}^X \sum_{\nu}^X \sum_J^Y P_{\mu\nu} q_J V_{\mu\nu}^J + \sum_A^X \sum_J^Y Z_A q_J V^{AJ} \quad \text{Eq 2.45}$$

where $P_{\mu\nu}$ and Z_A are density matrix elements and nuclear charges, respectively, and $V_{\mu\nu}^J$ and V^{AJ} denote nuclear attraction integrals and Coulomb repulsion terms which describe the interaction between a unit charge at MM atom J and an electron or core A in the QM region, respectively. Performing an SCF run with such a perturbed-core Hamiltonian yields a perturbed density matrix and consequently a different SCF energy $\tilde{E}_{QM}(X-L)$. This process may be identified as polarization or induction in the QM part and yields an energy gain of

$$E^{ind}(X) = \tilde{E}_{QM}(X-L) - E_{QM}(X-L) - E^{coul}(X,Y) \quad \text{Eq 2.46}$$

The potential and the total energy are given by:

$$E_{pot}(X-Y) = E_{MM}(Y) - E^{coul}(X,Y) + E^{ind}(X) + E_{QM}(X-L) + E_{MM}^{bonded}(X,Y) + E_{MM}^{vdW}(X,Y) \quad \text{Eq 2.47}$$

$$E_{tot}(X) = E_{pot}(X-Y) + E_{link} \quad \text{Eq 2.48}$$

The difference between this and the previous model is that in the electrostatic embedding scheme, the electrostatic part of the QM/MM interactions from the mechanical embedding method is replaced by Coulomb and induction energies (Eq 2.33 and Eq 2.34, respectively).

Van der Waals and bonded interactions are kept unchanged and are treated molecular mechanically.

Polarized embedding

The disadvantage of the electrostatic embedding approach is the unsatisfactory description of the non-bonded QM/MM interactions. The polarization of the QM part due to the point charges in the MM part is calculated, while the MM part stays “unpolarized”, i.e. a term is missing which would describe the polarization of the MM part due to the presence of the QM electric field. Classically this term could be represented in the following form:

$$E^{ind}(Y) = -\frac{1}{2} \sum_J^Y \mu_\alpha^J \langle F_\alpha^J \rangle \quad \text{Eq 2.49}$$

where

$$\langle F_\alpha^J \rangle = \sum_\mu^X \sum_\nu^X \tilde{P}_{\mu\nu} \nabla_\alpha V_{\mu\nu}^J + \sum_A^X Z_A \nabla_\alpha V^{AJ} \quad \text{Eq 2.50}$$

is the expectation value of the unit electric field operator, μ the induced dipole moment, and \tilde{P} the perturbed density matrix which corresponds to the perturbed Hamiltonian \tilde{h} .

The total energy is then the total energy of the system treated with electrostatic embedding plus a correction term for induction energy, E^{ind} .

2.2.4 Molecular Dynamics (MD)

A certain combination of spatial coordinates of the atoms (nuclei) in a molecule can be represented as a point in a phase space. Changes in molecular conformation correspond to moving through the phase space. Both the force field energy and the energy obtained by solving the Schrödinger equation are functions of nuclear coordinates. One of the most common problems in computational chemistry is determination of the molecular structure that corresponds to a stationary point in a phase space, such as local and global energy minima and transition state structures. This can be achieved by optimizing the nuclear coordinates with respect to the energy.

One of the methods of investigating the phase space corresponding to a particular molecular system is the molecular dynamics method. The idea behind this approach is to generate a trajectory consisting of successive configurations of the system by solving Newton's equations of motion.

$$\frac{d^2 \vec{r}_i}{dt^2} = \frac{\vec{F}_i}{m_i} \quad \text{Eq 2.51}$$

In principle, every MD run is an iterative procedure in which the following two steps are repeated many thousands of times:

1. Computation of the forces on all particles.
2. Integration of Newton's equations of motion to update particle positions and velocities.

Before the simulation begins, the initial configuration and velocities of the system must be assigned. The starting structure of the molecules can be obtained from experimental data, or from a theoretical model, or from a combination of these two. The initial velocities are usually obtained by a random selection from a Maxwell-Boltzmann distribution at the specified temperature:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right] \quad \text{Eq 2.52}$$

This describes the probability p that an atom i of mass m_i has a velocity v_{ix} in direction x at a temperature T . The velocities thus generated are usually further adjusted so that the total momentum of the system equals zero.

The force calculation is often the most time-consuming part of an MD simulation. The force acting on each atom is determined by differentiating the potential function (Eq.2.33) and may include contributions from a wide range of bonded and long- and short-range nonbonded terms, depending on the force field used. Luckily, by applying Newton's third law, which says that the force between two particles is equal in magnitude but opposite in direction (along the line connecting the centers of the particles), the force between two atoms has to be computed only once.

$$F_x(\vec{r}) = -\frac{\partial E_{FF}(\vec{r})}{\partial x} \quad \text{Eq 2.53}$$

Once the forces between the particles are computed, Newton's equations of motion can be solved. However, since the force acting on each atom has a complex dependence on the change in positions of all the other atoms, the task of integrating Newton's equations cannot normally be solved analytically. The usual solution in such cases is to break up the integration into many small steps. This numerical solution of the integration is called the finite difference method.

Finite difference methods

The idea behind finite difference techniques is to break down the integration of Newton's equations of motion into small stages corresponding to a small, fixed time step δt . The force acting on each atom in a molecular configuration at time t is then computed as a vector sum of its interaction with surrounding atoms. From the force one can then determine the accelerations ($\vec{a}(t)$), velocities ($\vec{v}(t)$), and positions ($\vec{r}(t)$) of particles at a time t and use them to calculate the positions and velocities at a time $t + \delta t$. The whole cycle is then repeated for the new configuration. The force is considered a constant during a time step. There are many algorithms based on finite difference methods used in molecular dynamics. However, they are all based on the assumption that the positions and dynamic properties of atoms in a system can be approximated by Taylor series expansions:

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}(t) + \frac{1}{2} \delta t^2 \vec{a}(t) + \frac{1}{6} \delta t^3 \vec{b}(t) + \frac{1}{24} \delta t^4 \vec{c}(t) + \dots \quad \text{Eq 2.54}$$

$$\vec{v}(t + \delta t) = \vec{v}(t) + \delta t \vec{a}(t) + \frac{1}{2} \delta t^2 \vec{b}(t) + \frac{1}{6} \delta t^3 \vec{c}(t) + \dots \quad \text{Eq 2.55}$$

$$\vec{a}(t + \delta t) = \vec{a}(t) + \delta t \vec{b}(t) + \frac{1}{2} \delta t^2 \vec{c}(t) + \dots \quad \text{Eq 2.56}$$

The most widely used method for integration of the equations of motion is probably the Verlet algorithm. It uses the positions from the current and previous time step and accelerations from the current step to calculate the new positions for the next step in the

simulation. The relationship between the positions in three subsequent steps can be represented as follows:

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}(t) + \frac{1}{2} \delta t^2 \vec{a}(t) + \dots \quad \text{Eq 2.57}$$

$$\vec{r}(t - \delta t) = \vec{r}(t) - \delta t \vec{v}(t) + \frac{1}{2} \delta t^2 \vec{a}(t) - \dots \quad \text{Eq 2.58}$$

By combining Eqs 2.37 and 2.38 one obtains:

$$\vec{r}(t + \delta t) = 2\vec{r}(t) + \vec{r}(t - \delta t) + \delta t^2 \vec{a}(t) \quad \text{Eq 2.59}$$

Eq 2.39 shows that in the Verlet algorithm the velocities do not appear explicitly. However, they can be calculated by, for instance, dividing the difference in positions at times $t - \delta t$ and $t + \delta t$ by a time corresponding to two time steps, as shown in Eq 2.40, or by estimating the velocity at the half-step, as in Eq 2.41.

$$\vec{v}(t) = \frac{[\vec{r}(t + \delta t) - \vec{r}(t - \delta t)]}{2\delta t} \quad \text{Eq 2.60}$$

$$\vec{v}\left(t + \frac{1}{2}\delta t\right) = \frac{[\vec{r}(t + \delta t) - \vec{r}(t)]}{\delta t} \quad \text{Eq 2.61}$$

Although this method of integration is popular, it has a number of drawbacks:

- a. It is not a “self-starting” algorithm, since the positions at time $-\delta t$ are necessary to start the integration. Usually they are obtained by truncating the Taylor expansion in Eq 2.34 after the first term: $\vec{r}(-\delta t) = \vec{r}(0) - \delta t \vec{v}(0) - \frac{1}{2} \delta t^2 \vec{a}(0)$.
- b. A loss of precision may occur since in Eq 2.39 the first two terms on the left hand side are large numbers, whereas the term $\delta t^2 \vec{a}(t)$ is far smaller.
- c. The lack of an explicit velocity term in the equations makes calculation of the velocities difficult, and furthermore, the velocities at a time t cannot be calculated without the nuclear positions at the subsequent step.

There are several variations on the Verlet algorithm, one of which is the so-called leap-frog algorithm, which uses the following relationships:

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}\left(t + \frac{1}{2} \delta t\right) \quad \text{Eq 2.62}$$

$$\vec{v}\left(t + \frac{1}{2} \delta t\right) = \vec{v}\left(t - \frac{1}{2} \delta t\right) + \delta t \vec{a}(t) \quad \text{Eq 2.63}$$

The velocities and positions are not calculated at the same time step; moreover, the velocities “leap-frog” over the positions calculated at time t to yield their new values for the time $t + \frac{1}{2} \delta t$.

The velocities at time t can be calculated as:

$$\vec{v}(t) = \frac{\left[\vec{v}\left(t + \frac{1}{2} \delta t\right) + \vec{v}\left(t - \frac{1}{2} \delta t\right) \right]}{2} \quad \text{Eq 2.64}$$

The advantages of this method over the standard Verlet algorithm are that the velocities are explicitly computed and that it does not include calculation of differences of large numbers. However, the obvious disadvantage is that the positions and velocities are not synchronized, so the kinetic (velocity dependent) and potential (position dependent) energy contributions cannot be calculated at the same time.

Another variation on the Verlet algorithm is made in the velocity Verlet algorithm. In this case, positions, velocities, and accelerations are computed at the same time:

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}(t) + \frac{1}{2} \delta t^2 \vec{a}(t) \quad \text{Eq 2.65}$$

$$\vec{v}(t + \delta t) = \vec{v}(t) + \frac{1}{2} \delta t [\vec{a}(t) + \vec{a}(t + \delta t)] \quad \text{Eq 2.66}$$

Chapter 3 Results and discussion

In this chapter, the results will be presented of the calculations on cysteine protease inhibition mechanism performed using pure quantum mechanical (QM), combined quantum mechanics/molecular modeling (QM/MM), molecular dynamics (MD), and QM/MM molecular dynamics (QM/MM MD) methods. This chapter is divided into three parts, one for the calculations performed with each simulation method (QM/MM and MD calculations are described in the same subchapter, as they were performed on the same system and support each other). Each subchapter is further subdivided into a section containing theoretical details about the method involved and a part presenting and discussing the results obtained using that method.

All calculations presented here (except the QM/MM MD simulations) deal with the second step of the inhibition, namely the irreversible deactivation of the enzyme by the inhibitor. By treating this interaction with a variety of different methods we managed to isolate and observe individual effects governing the reaction mechanism.

The quantum mechanical approach was used to study very small model systems consisting only of the electrophilic warhead of the inhibitor (both substituted and not) and molecular moieties simulating a very simplified protein active site (methylthiolate instead of Cys29 and methylimidazolium instead of His199 residue) and solvent surroundings (two waters or two ammonium ions, in combination with a continuum solvent model). Although simple, such a system provides a good description of the most important interactions involved in the inhibition reaction. It also allows investigation of the influence of the properties of the electrophilic warhead on the reaction rate. Substitution at the heteroatom of the aziridine ring was examined, yielding results which could prove useful in rational drug design. Possible protonation of the ring heteroatom by the His199 residue was studied to inform the configuration of subsequent hybrid QM/MM computations.

Beside the properties of the electrophilic warhead, the protein and solvent environment is also an important factor in the irreversible deactivation of the enzyme active site by the inhibitor. The non-covalent interactions of the inhibitor with the oxyanion hole and other subsites of the enzyme need to be explicitly taken into account in the calculations, because of their possible impact on the reaction profile. Since the active site of cysteine proteases is situated at the surface of the enzyme, it is completely accessible to the solvent. As a consequence, solvent water molecules also have to be explicitly simulated. As molecular modeling methods allow the treatment of such large systems, but lack the possibility of describing covalent interactions, our method of choice was the combined quantum mechanics/molecular modeling approach. By splitting the system into a smaller part that undergoes the bond cleavage/formation process and must be treated quantum mechanically, and a larger part, comprised of the rest of the protein, which could be treated using force fields, we managed to simulate the system at the desired precision. Our investigations concentrated on the role of His199 in the inhibition mechanism as well as on the structure-reactivity relationships between cysteine protease and various inhibitors, yielding new insight into the kinetics, regio- and stereospecificity of the inhibition. In particular, our calculations provide the following insights: *i.*) an explanation for the regioselectivity of the reaction, and original insight into

which interactions affect the stereoselectivity; *ii.*) a clear model which explains the known structure-activity relationships and connects these effects with the pH-dependency of the inhibition; *iii.*) our computations question the generally accepted two-step model by showing that substituent effects accelerate the irreversible step to such an extent that the achievement of an equilibrium in the first step is doubtful; *iv.*) by way of theoretical characterizations of aziridine models, the reasons for similarities and differences in the mode of action of epoxide- and aziridine-based inhibitors are elucidated; and finally, *v.*) combining our results with experimental knowledge will allow rational design of new inhibitors.

To account for dynamical effects as well, molecular dynamics (MD) computations were also performed. In these calculations the potential energy was computed at the force field level. The results of these calculations not only supported and clarified the QM/MM results, but comparison with previous X-ray structures helped correct existing errors in the available geometrical models and resolved inconsistencies in the weighting of various factors governing the inhibition.

In the final part of this chapter the first QM/MM MD calculations on the active site of the cysteine proteases will be presented. In contrast to the MD simulations discussed above, these calculations used potential energies computed at the QM/MM-level. With the help of these computations we sought to address strongly disputed questions about the reasons for the existence of the active site ion pair and its role in the high activity of the enzyme.

3.1 QM Calculations

3.1.1 The method

In order to determine the relevance of the different molecular moieties which take part in the inhibition of cysteine proteases and to decide which of them should be treated quantum mechanically in the later QM/MM investigations, pure QM calculations were done on small model systems. In these the inhibitor was represented only by its electrophilic warhead moiety (the oxirane/aziridine ring) and no protein surroundings were taken into account. Two sets of calculations have been performed. In one set, the enzyme environment was simulated only by the methylthiolate (representing the Cys29 residue of the active site). The influence of substituents at the N center of the aziridine ring was treated using this model. In the other set of calculations, the presence of a protonated His199 residue was also taken into account and its role in protonation of non-N-substituted aziridines and oxiranes was investigated. All the calculations were done at the same level of theory.

The TURBOMOLE³⁸ program package was used in the computations. To mimic the electrostatic effects of the solvent, the continuum solvent model (COSMO³⁹) was employed using a dielectric constant of $\epsilon = 78.39$. During geometry optimizations and scans of potential energy surfaces (PESs) or potential energy paths (PEPs) the calculations were done at the RI⁴⁰DFT/B-LYP⁴¹/TZVP⁴² level of theory. This approach was found to be sufficiently accurate for the determination of the geometrical structure of the stationary points, while underestimating reaction barriers and energies. Therefore single point computations at the DFT/B3⁴³-LYP/TZVP level were performed to determine the activation and reaction energies of various processes more precisely. For the gas phase this combined approach was found to yield similar results to the CCSD(T) method.⁴⁴

The computed PESs and PEPs for the reaction must capture the energy changes involved in the inhibitor-ring-opening and the S-C₁ bond creation steps (for notation see Figure 3.1.1). In the first set of calculations (those using only methylthiolate to represent the protein) only the reaction path is calculated. The reaction coordinate contains the angle XC₂C₁ (X=O or NH) of the inhibitor ring and the distance between S and C₁. Although the reaction paths obtained proved sufficient for locating the transition state structures, for the second set of calculations

two-dimensional PESs were also computed by varying $\angle XC_2C_1$ and R_{C_1-S} independently. The PESs were computed to ensure that both coordinates determining the reaction coordinate are properly taken into account. The results of these calculations (see Figures 3.1.7 – 3.1.10) show that the S-C₁ distance is the most important coordinate at the beginning, while the NC₂C₁ angle becomes more and more significant as the reaction proceeds. For each point of the two-dimensional PES and the one-dimensional reaction path all other internal degrees of freedom were optimized. Only the reaction coordinates were kept fixed at certain values.

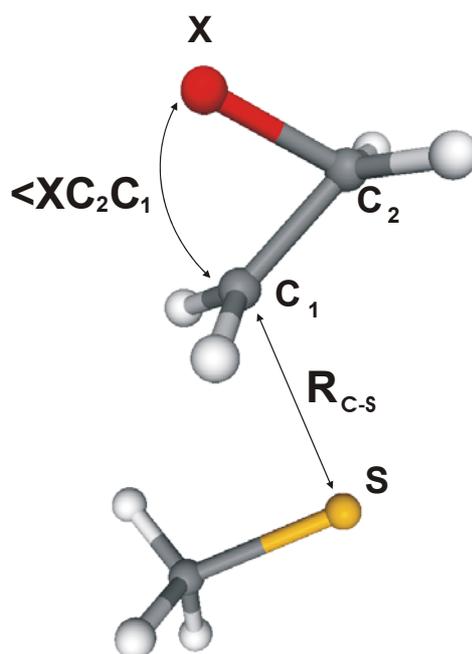


Figure 3.1.1. Notation of important ring-opening reaction coordinates examined in the QM calculations. X=O, NH, or N-R.

Points close to the transition state on the PES and PEP then served as input for a numerical frequency analysis and subsequently a transition state (TS) search. The resulting TS structures were then relaxed by a careful downhill optimization towards reactants and products to ensure that these stationary points were connected by contiguous paths. The geometries obtained in this process were used for the determination of barriers and reaction energies. Again, these optimizations were done at the DFT/B-LYP/TZVP level of theory, while single-point calculations at the DFT/B3-LYP/TZVP level were performed to achieve more accurate relative energies.

3.1.2 The role of His199 in protonation of the inhibitor

Two different model systems were set up to investigate whether the electrophilic ring of the inhibitor is protonated by the His199 residue of the protease during the inhibition reaction. The first consisted of the oxirane ring, methylthiolate (representing the Cys29 residue and henceforth referred to as Cys29), methylimidazolium (representing the His199 residue and henceforth referred to as His199) and a water molecule oriented as a bridge [oxirane-system (Figure 3.1.3)]. In the second model system the oxirane ring was replaced by an aziridine ring [aziridine-system (Figure 3.1.5)].

The starting geometries were chosen in such a manner as to represent the real positions of Cys29, His199 and inhibitor ring in the X-ray structure of Cathepsin B - E64c enzyme-inhibitor complex. During the calculations the relative positions of the methyl groups of methylthiolate and methylimidazolium with respect to each other were kept fixed to mimic the rigidity of the backbone of the enzyme. By careful inspection of the X-ray structure of the enzyme-inhibitor complexes, one can see that the distance between the hydrogen of the protonated His199 residue and the inhibitor ring is too large for a protonation to take place, as was reported by Varughese *et al.*^{19,45} However, since the active site lies on the surface of the enzyme, the large distance between the histidine residue and the inhibitor could possibly be bridged by a water molecule which could then serve as a proton transmitter. In order to investigate this possibility the computations were performed both with and without such a bridging water molecule.

In the PEP calculations only the reaction path corresponding to the simultaneous opening of the ring ($<NC_2C_1$ or $<OC_2C_1$) and shortening of the S-C₁ distance was calculated. Nevertheless, for this simple system this proved to be sufficient for the determination of transition state structures.

The reaction path and the corresponding stationary point structures for the oxirane-system are shown in Figures 3.1.2 and 3.1.3, respectively. In the figures and discussion the reactants will be denoted as R, transition state structures as TS, and products as P. The corresponding protonated entities will be denoted as RH, TSH and PH.

In the presence of an explicit water molecule the oxirane ring undergoes an exothermic inhibition reaction in which the TS is not protonated. The protonation takes place at the end of

the reaction, when the structure is near P. As a result, the kinetics of the inhibition reaction are not influenced by the protonation, but the thermodynamics are. The potential energy barrier of the reaction is 9 kcal/mol. The corresponding geometrical parameters are $\angle OC_2C_1 = 74^\circ$ and $R_{Cl-S} \approx 2.8 \text{ \AA}$ (Table 3.1.1)

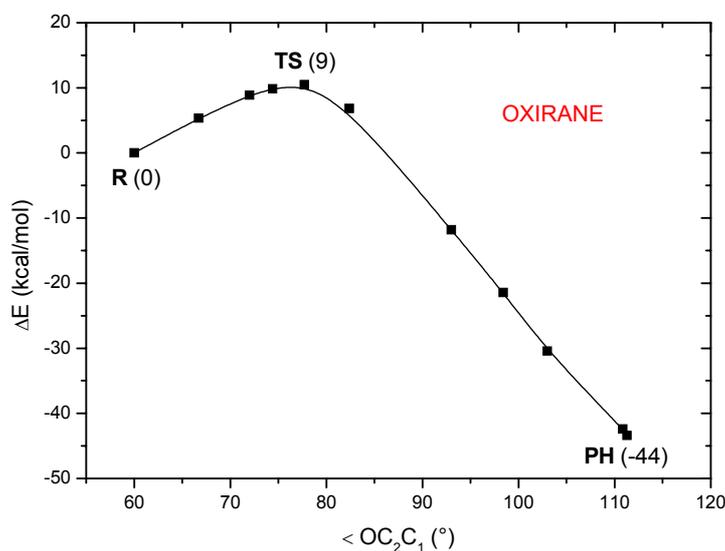


Figure 3.1.2. Reaction path of the ring-opening reaction obtained in calculations on the system consisting of the oxirane ring, Cys29, His199 and a water molecule. Stationary point energies in parentheses in kcal/mol.

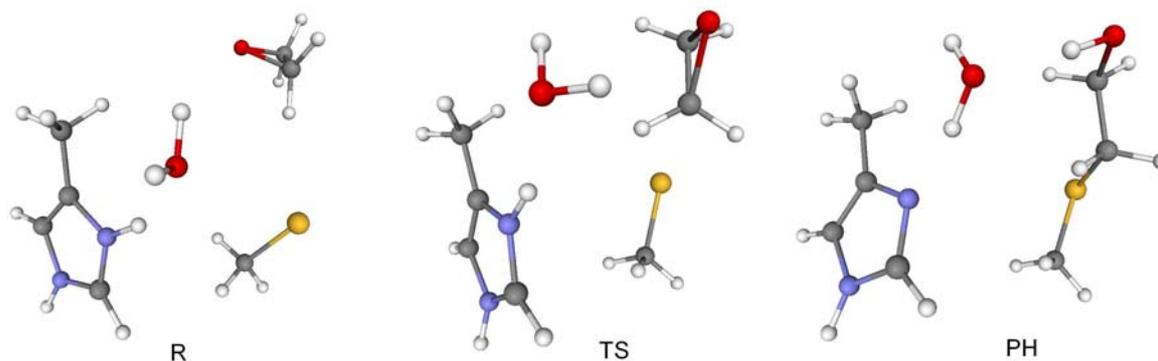


Figure 3.1.3. Geometrical structures of stationary points on the reaction path obtained for the oxirane system.

Computations on the aziridine-system were done in the same manner as for oxirane, except that in this case two different reaction paths are possible. One is a reaction in which protonation of the aziridine ring by the imidazolium ring of His199 via a water molecule occurs long before the TS, at the very beginning of the reaction. Along the other path the protonation of the ring occurs after the transition state, and takes place near the product structure. The reaction paths and the stationary point structures are depicted in Figures 3.1.4 and 3.1.5, respectively.

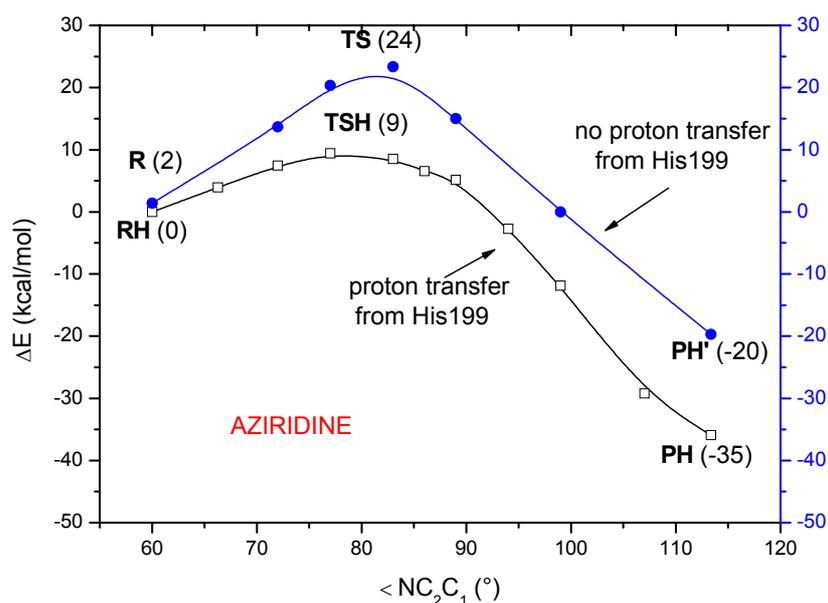


Figure 3.1.4. Reaction path for the ring-opening reaction obtained for the system consisting of the aziridine ring, Cys29, His199 and a water molecule. The blue line represents the reaction in which the product is only protonated after the transition state. The black one describes the reaction in which the protonation takes place before the TS. Stationary point energies in parentheses in kcal/mol.

To investigate the non-protonated TS the protonation had to be suppressed. This was obtained by freezing the NH bond of the labile proton at its equilibrium distance in the R state. The protonation of the aziridine ring, which takes place at the end of the reaction, is due to the presence of the water molecule, which is reduced to a hydroxyl anion. This explains the higher energy. Since in this case His199 is not the proton source, this product structure is

denoted with PH' to differentiate it from the protonated product structure due to the proton transfer from His199 (PH).

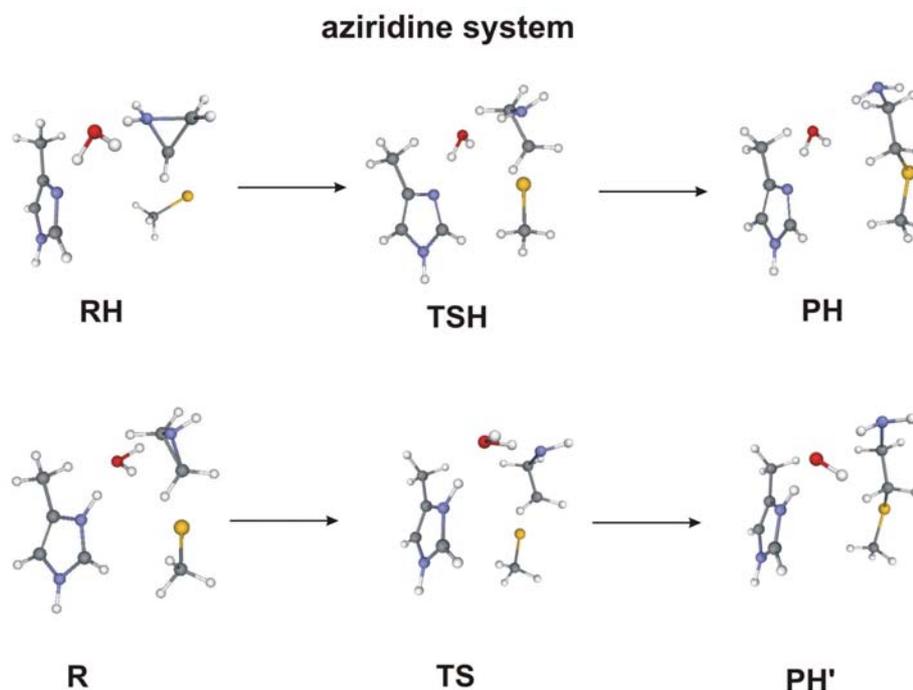


Figure 3.1.5. Structures corresponding to the stationary points on the both reaction paths obtained for the aziridine-system. Upper part: water-mediated proton transfer from His199 takes place. Lower part: Protonation of the ring by His199 does not take place.

Table 3.1.1. The reaction coordinate values and relative energies for the stationary point structures found along both aziridine reaction paths and the oxirane reaction path.

	aziridine			Oxirane		
	$\angle \text{NC}_2\text{C}_1$ (°)	$R_{\text{C1-S}}$ (Å)	ΔE (kcal/mol)	$\angle \text{OC}_2\text{C}_1$ (°)	$R_{\text{C1-S}}$ (Å)	ΔE (kcal/mol)
RH	60	3.631	0	-	-	-
R	60	3.906	2	60	3.784	0
TS	83	2.642	24	74	2.799	9
TSH	73	2.794	8	-	-	-
PH	109	1.868	-35	111	1.874	-44
PH'	113	1.875	-20	-	-	-

The parameters characterizing the stationary points, the value of the reaction coordinate, and the energy relative to the reactant along the ring-opening reaction paths for both the aziridine and in oxirane systems are given in Table 3.1.1.

Obviously, for the aziridine system, migration of the proton from His199 to aziridine is favorable, even before the reaction begins. Consequently, proton donation strongly influences both the kinetics and the thermodynamics of the reaction. If protonation by His199 is not allowed the TS is reached at the $\angle\text{NC}_2\text{C}_1$ angle of 83° and the $R_{\text{C1-S}}$ distance of 2.64 Å. A reaction barrier of about 24 kcal/mol is computed. Compared to the oxirane inhibitor such a reaction would be rather slow. However, if protonation is enabled for the aziridine system, the protonated TS occurs earlier, at $\angle\text{NC}_2\text{C}_1 = 73^\circ$ and $R_{\text{C1-S}} = 2.79$ Å, and a much lower barrier of about 8 kcal/mol is predicted.

Our QM calculations showed that His199 can indeed protonate the inhibitor ring if the proton is transmitted via a water molecule. This indicates that His199 should be incorporated into the QM part of the QM/MM calculations. If it is left in the MM part the proton transfer could not be described.

If one compares the geometries and energetics of the oxirane reaction with the aziridine reaction in which the protonation by His199 is enabled, one sees that both the stationary point geometries and the reaction barriers are similar (Table 3.1.1). On the other hand, the aziridine reaction in which the protonation does not occur until the end of the reaction path is characterized by a 15 kcal/mol higher reaction barrier. This shows that in systems in which the surroundings are acidic enough to allow the protonation of the aziridine TS, the inhibition by aziridines should happen as fast as with oxirane-based inhibitors; whereas if this is not the case and the TS is not protonated, the aziridine-based inhibitors should be much slower. This observation correlates nicely with the results of previous calculations,^{44,46} as well as with experimental results showing a similar change in the kinetics of the aziridine-based inhibition reaction on changing the pH of the system from basic to acidic.

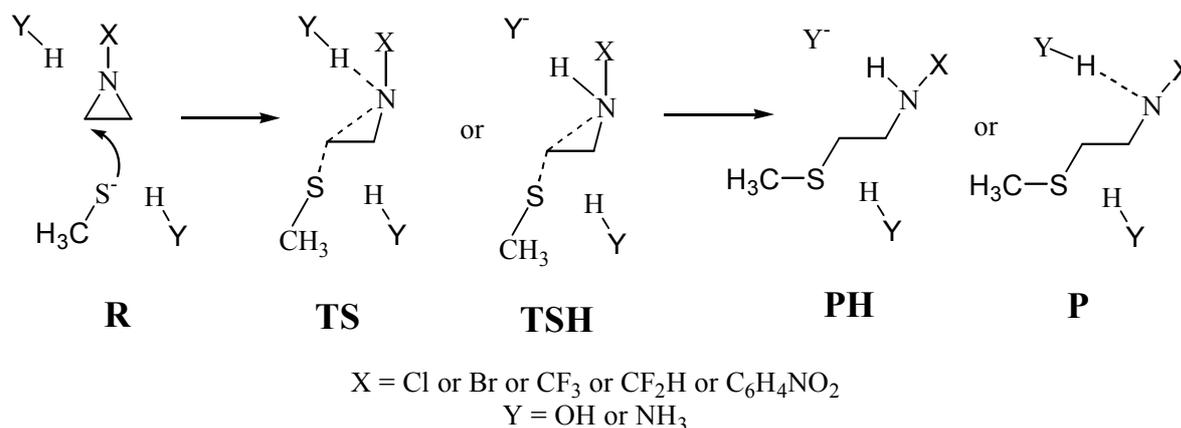
3.1.3 Substitution at the N-atom of aziridine

Epoxide- and aziridine-based inhibitors of cysteine proteases^{6,47,48} are typically developed using the naturally occurring epoxysuccinyl peptide, E64, as a starting point. They represent a

very potent class of inhibitors. Although both epoxide and aziridine heterocycles share a common principal inhibition mechanism, some differences exist. The most important, and those which we will focus on this work, are the influence of the medium pH, alkylation rates, and substituent effects.

As indicated in the results already presented and the work of Helten *et al.*,^{44,46} the aziridine ring opening is influenced by acidic media to a greater extent than the epoxide ring opening. This is in line with the experimental data relating to aziridines as inhibitors of cysteine proteases.⁴⁴ The acidic environment enables both inter- (via His199 residue) and intra- (via carboxyl group substituted at ring-carbon atom) molecular water-mediated protonation of the aziridine ring to take place, and directly influences the kinetics of the inhibition reaction.^{44,46}

In contrast to oxirane, the aziridine ring allows additional substitution at the ring N atom. This has two favorable consequences. 1.) Substituents could be introduced which enhance the affinity to certain enzyme subsites, leading to an enhancement of the specificity of the inhibitor to certain protease. 2.) Substituted species could modify the properties of the aziridine ring relating to the nucleophilic ring-opening reaction, by influencing the kinetics and thermodynamics of the inhibition reaction.



Scheme 3.1.1. Model reaction of the alkylation of methyl thiolate by N-substituted aziridine ring. R stands for reactants, TS, for transition state, P for unprotonated product and PH for protonated product structure.

Comparison of aziridine-based inhibitors showed that the N-acylated derivatives without a free acid group at the aziridine ring carbon atoms^{49,50} are, in general, more potent than the corresponding N-unsubstituted or N-alkylated derivatives. It is believed that this is due to the electron-withdrawing property of the acyl group, the presence of which activates the ring towards nucleophilic attack. Calculations⁵¹ support this assumption. Attachment of a CHO group at the N-position of the aziridine ring lowers the reaction barrier for nucleophilic attack by methylthiolate by 14 kcal/mol in aqueous surroundings compared to the unsubstituted aziridine ring. To further investigate the influence of electron-withdrawing substituents on the character of the ring-opening reaction, and to provide a more detailed background for the rational design of new aziridine-based inhibitors, calculations were performed on various new model systems. Ten different model systems were created: each of the N-aziridine substituents Cl, Br, CF₃, CF₂H, and C₆H₄NO₂ were tested, and for each substituent the influence of the polar surroundings was probed by adding either two water molecules (pK_s≈15) to simulate a low proton-donating ability of the environment or two ammonium ions (pK_s≈9.3) to simulate a more acidic environment. The reaction under examination is the nucleophilic attack by the methylthiolate (representing the Cys29 residue) at the substituted aziridine ring, which is shown in Figure 3.1.6 together with the notation for reaction coordinates and geometry parameters in the following discussion.

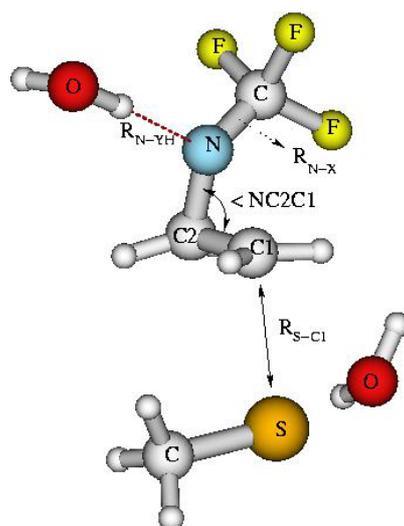


Figure 3.1.6. The notation for reaction coordinates and other geometrical parameters used in Table 3.1.3., shown using the example of the system with a CF₃ moiety as a ring substituent (X = CF₃) and 2 water molecules as explicit solvent (Y = OH).

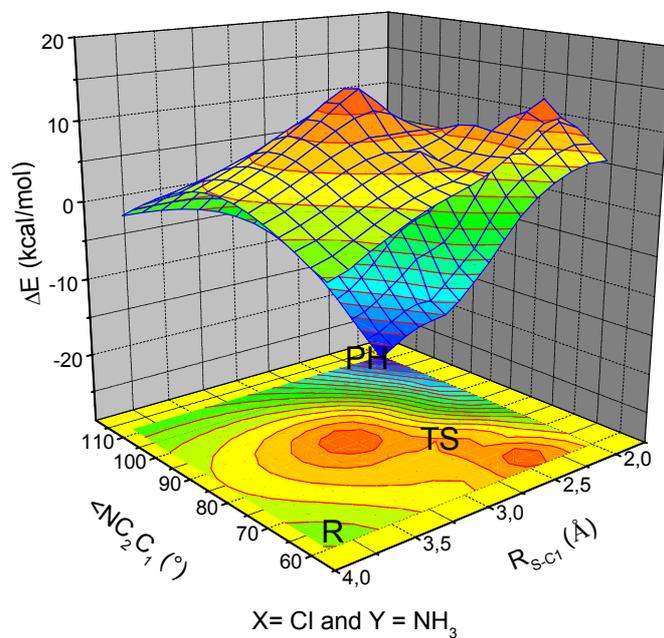
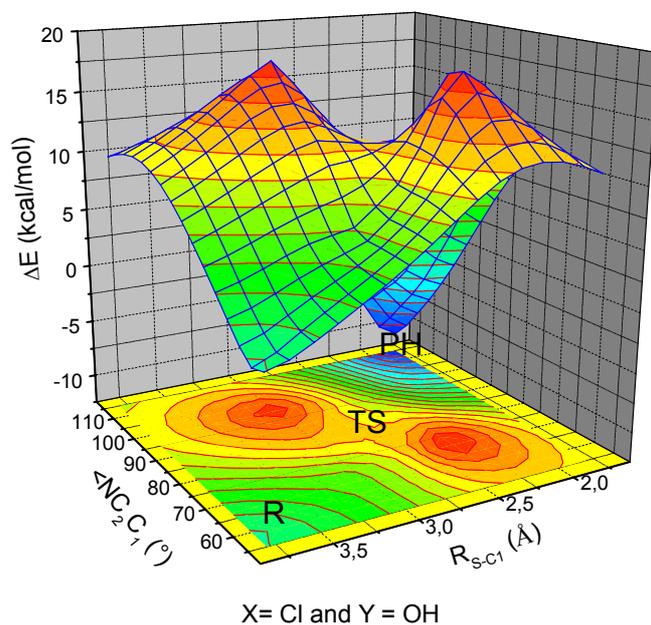


Figure 3.1.7. Potential energy surfaces (PES) computed for the systems containing Cl as an N-substituent, with two explicit water molecules (top) or two explicit ammonium ions (bottom) as solvent.

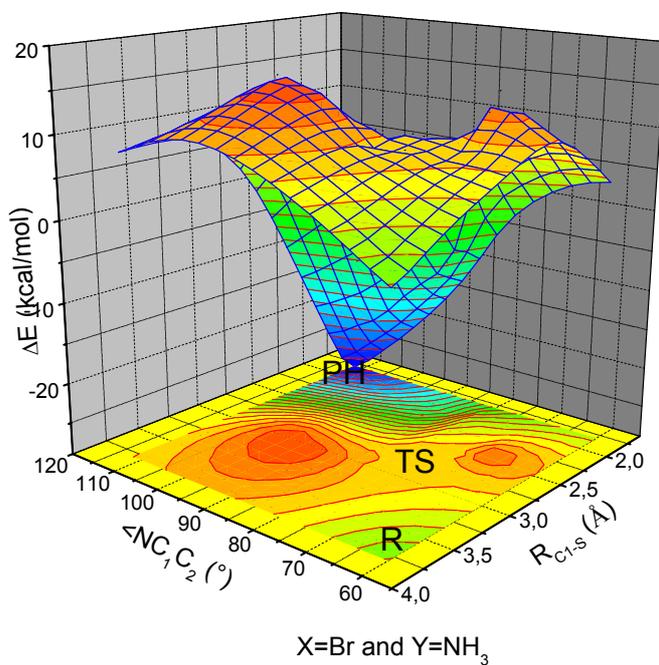
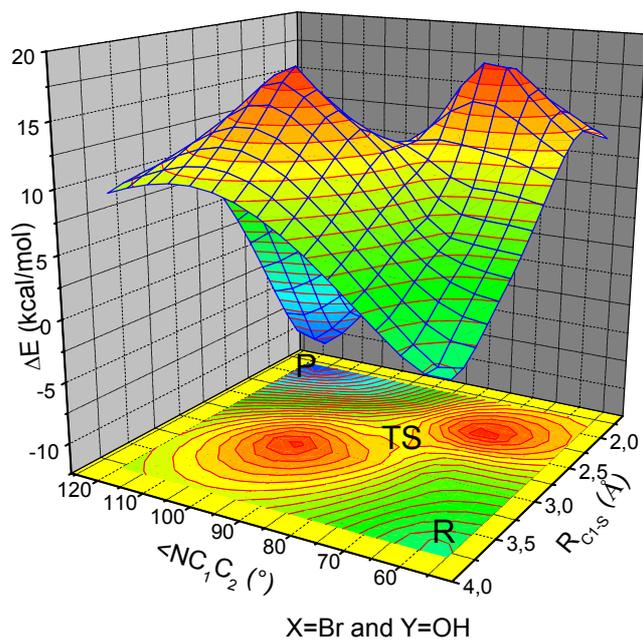


Figure 3.1.8. Potential energy surfaces (PES) computed for the systems containing Br as an N-substituent, with two explicit water molecules (top) or two explicit ammonium ions (bottom) as solvent.

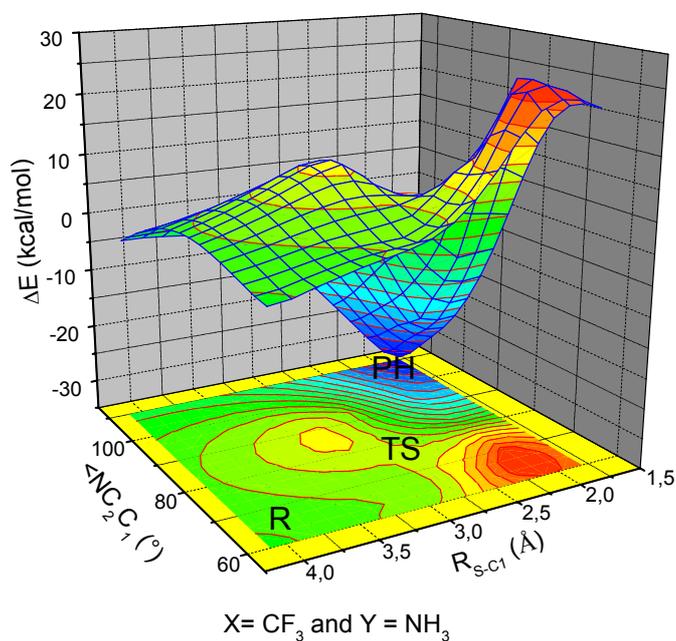
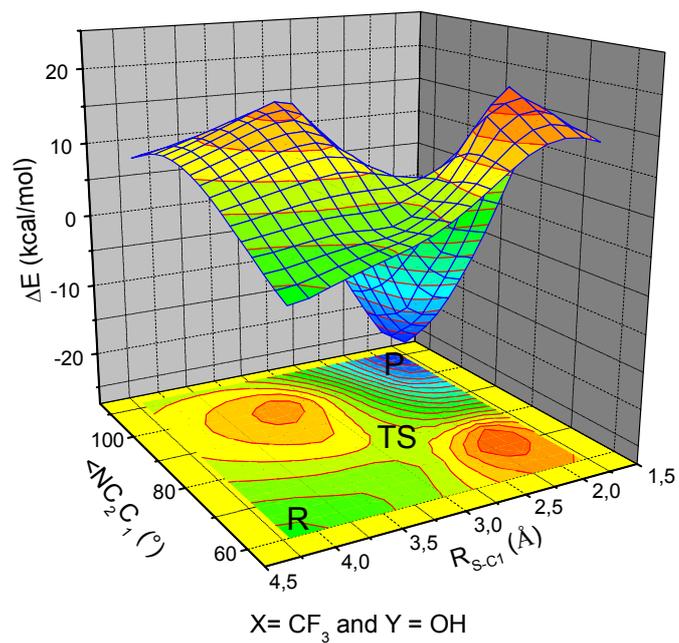


Figure 3.1.9. Potential energy surfaces (PES) computed for the systems containing CF_3 as an N-substituent, with two explicit water molecules (top) or two explicit ammonium ions (bottom) as solvent.

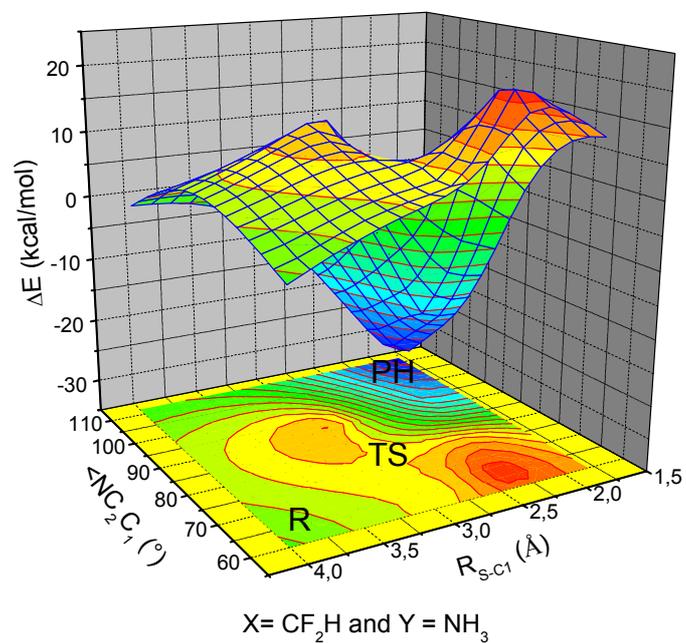
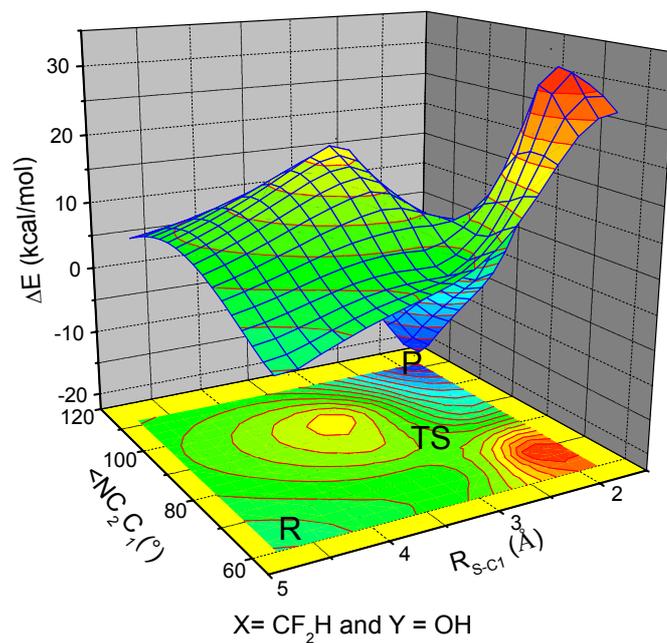


Figure 3.1.10. Potential energy surfaces (PES) computed for the systems containing CF₂H as an N-substituent, with two explicit water molecules (top) or two explicit ammonium ions (bottom) as solvent.

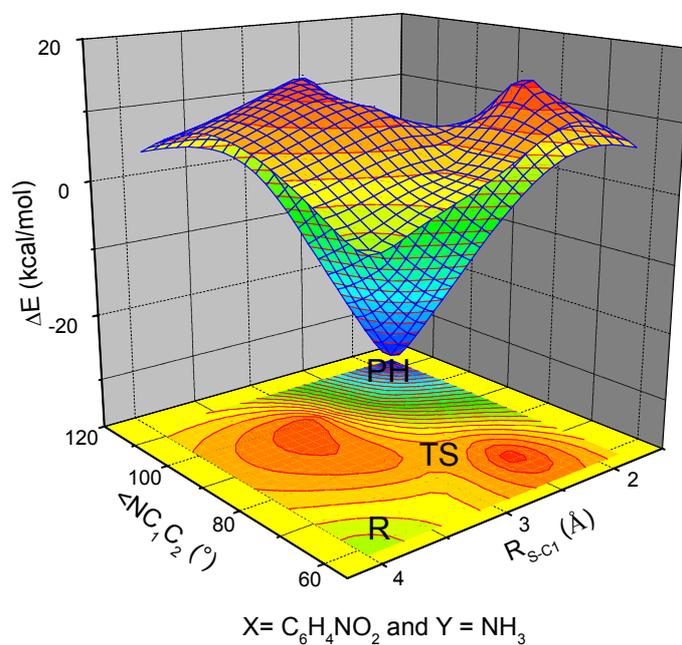
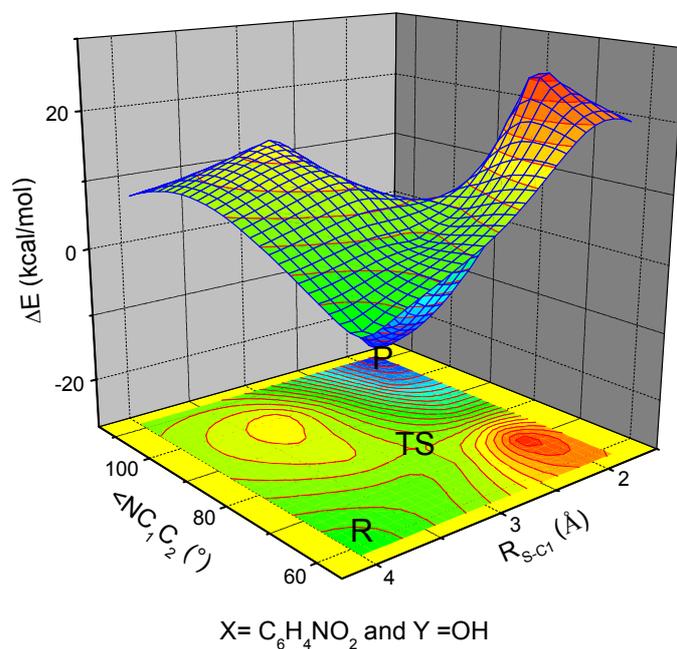


Figure 3.1.11. Potential energy surfaces (PES) computed for the systems containing C₆H₄NO₂ as an N-substituent, with two explicit water molecules (top) or two explicit ammonium ions (bottom) as solvent.

Chapter 3 Results and discussion

Table 3.1.2. Relative energies (in kcal/mol) describing the influence of the substituents and explicit solvent molecules on the reaction profiles of the ring-opening reaction.

entry	X	Y = OH				Y = NH ₃			
		TS	TSH	P	PH	TS	TSH	P	PH
1	H ^a	28	-	-	6	-	14	-	-24
2	CHO ^a	14	-	-25	-23	13	-	-	-41
3	Cl	16	-	-12	-10	15	17	-	-31
4	Br	16	-	-12	-10	15	17	-	-31
5	CF ₃	13	-	-23	-	12	-	-	-36
6	CF ₂ H	16	-	-17	-	15	-	-	-34
7	C ₆ H ₄ NO ₂	14	-	-25	-	14	12	-	-38

^a The results are taken from Ref.51

Table 3.1.3. Geometry parameters for systems containing different N-substituents (X) and explicit solvent molecules (Y) for the stationary points along the reaction path.

system X / Y		<NC ₂ C ₁ (°)	R _{S-C1} (Å)	R _{C1-C2} (Å)	R _{N-C1} (Å)	R _{N-C2} (Å)	R _{N-YH} (Å)	R _{N-X} (Å)
Cl/OH	R	59.9	3.767	1.492	1.493	1.498	1.900	1.868
	TS	80.4	2.666	1.469	1.906	1.484	1.800	1.862
	P	115.0	1.874	1.548	2.530	1.451	1.611	1.929
Cl/NH ₃	R	60.1	3.847	1.493	1.497	1.497	1.789	1.845
	TS	79.6	2.660	1.471	1.890	1.483	1.659	1.848
	PH	111.8	1.863	1.534	2.485	1.478	1.052	1.858
Br/OH	R	59.7	3.701	1.492	1.489	1.498	1.891	2.036
	TS	80.3	2.642	1.469	1.904	1.483	1.793	2.006
	P	115.1	1.874	1.548	2.532	1.452	1.625	2.064
Br/NH ₃	R	59.9	3.771	1.492	1.492	1.496	1.773	1.999
	TS	79.4	2.645	1.471	1.887	1.482	1.636	1.992
	PH	111.8	1.864	1.534	2.496	1.480	1.052	2.001
CF ₃ /OH	R	60.2	4.146	1.490	1.499	1.497	1.967	1.410
	TS	75.4	2.773	1.469	1.808	1.487	1.871	1.374
	P	110.2	1.869	1.543	2.475	1.473	1.764	1.296
CF ₃ /NH ₃	R	60.4	4.028	1.488	1.504	1.502	1.881	1.417
	TS	75.5	2.780	1.468	1.811	1.490	1.740	1.382
	PH	109.4	1.864	1.539	2.469	1.486	1.051	1.372
CF ₂ H/OH	R	59.9	4.685	1.495	1.491	1.492	1.927	1.420
	TS	76.8	2.741	1.471	1.835	1.484	1.814	1.383
	P	111.5	1.870	1.545	2.495	1.472	1.724	1.300
CF ₂ H/NH ₃	R	60	3.994	1.493	1.495	1.498	1.824	1.428
	TS	77.4	2.738	1.472	1.849	1.486	1.671	1.392
	PH	110.2	1.865	1.541	2.476	1.478	1.042	1.368
C ₆ H ₄ NO ₂ /OH	R	59.5	3.928	1.504	1.481	1.479	1.935	1.406
	TS	74.9	2.766	1.478	1.791	1.469	1.835	1.385
	P	110.8	1.867	1.548	2.482	1.467	1.757	1.332
C ₆ H ₄ NO ₂ / NH ₃	R	59.9	3.911	1.500	1.490	1.485	1.802	1.418
	TS	75.1	2.775	1.475	1.798	1.474	1.674	1.395
	PH	112.1	1.864	1.548	2.500	1.466	1.042	1.359

In these calculations the same reaction coordinates were used as described in the previous section (Figure 3.1.1 and Figure 3.1.6). This time, however, two-dimensional PESs were calculated instead of only the one-dimensional reaction paths. The reason for this is the independence of the two internal coordinates on each other. The computed potential energy surfaces are shown in Figures 3.1.7-11. The points on the PESs near the saddle point served as input structures for numerical frequency analysis and, later on, for transition state optimization. The results are given in Table 3.1.2. All energies are given relative to the corresponding reactant structure energies. Both in Figures 3.1.7-11 and in Tables 3.1.2 and 3.1.3, PH and TSH represent the protonated product and transition state species respectively, while R and TS stand for the corresponding non-protonated stationary points. In Table 3.1.3 selected geometrical parameters of the stationary points of all ten PESs are given.

Calculations by Vicik *et al.*⁵¹ predict a very high reaction barrier for the ring-opening reaction of the unsubstituted aziridine ring in the presence of two explicit water molecules, representing a low-acidity environment (entry 1 of Table 3.1.2, 28 kcal/mol). The high barrier results from the protonation of the ring occurring after the TS. In the more acidic environment (with two ammonium ions) the reaction barrier drops to 14 kcal/mol due to protonation of the TS structure (TSH, Table 3.1.2). However, attaching a CHO substituent to the ring nitrogen changes the picture. Both in the presence of water and ammonium ions the transition state stays unprotonated, but the barrier heights are nonetheless quite low at 14 kcal/mol and 13 kcal/mol, respectively (entry 2 of Table 3.1.2). The substituents treated in the present work have the same impact on the inhibition reaction energy profile.

Characteristic of all PESs in Figure 3.1.7-11 is that none of the transition state structures are protonated, whether with water or ammonium ions in the vicinity (see also Table 3.1.2). The proton donation takes place later, to the product structure, and only in case of ammonium ions being employed as solvent molecules. Nevertheless, if the height of the reaction barriers is compared to the results obtained in our previous calculations on aziridine systems (Table 3.1.1) and the calculations performed by Vicik *et al.* (Table 3.1.2., entry 1), we see a drop of about 10 – 16 kcal/mol, depending on the N-substituent. With chlorine and bromine atoms as substituents the energy barrier in water is 16 kcal/mol and in ammonium 15 kcal/mol. The same activation energies are seen with the CF₂H substituent in both solvents. The C₆H₄NO₂ moiety leads to an energy barrier of 14 kcal/mol in both solvation environments, and the CF₃ substituent shows the greatest influence of all on the reaction kinetics with reaction barriers of

13 and 12 kcal/mol for water and ammonium respectively. The reason for this drop in activation energy, and hence acceleration of the inhibition reaction, is the electron-withdrawing behaviour of the substituents, which lowers the basicity of the aziridine N atom. Due to this the charge accumulated on the N atom along the reaction path is delocalized, making the nitrogen moiety a much better leaving group in the S_N2 reaction. The stabilization of the negative charge on the opening ring is reflected in the non-protonated TS structures of N-substituted rings even in cases where NH_4^+ ions serve as explicit solvent molecules. For comparison, the non-substituted ring experiences protonation under such circumstances; but even in cases where a protonated TS (denoted TSH) could be found for an N-substituted system (bromine and chlorine substituents), this state is actually 2 kcal/mol higher in energy than the non-protonated transition state, TS. This underlines the degree of charge stabilization at the ring, and also explains why a change in the proton-donor affinity of the surroundings (water vs. ammonium ions) does not influence the kinetics of the reaction much. The increase of the reaction barrier due to the presence of water as solvent is 1 kcal/mol, at most (Table 3.1.2). A difference between the systems solvated in water and those solvated in ammonium appears only at the end of the reaction path, in the product structures. The reactions with NH_4^+ are thermodynamically preferred, since protonation of the product decreases the reaction energy by 13-20 kcal/mol. This is consistent with the basicity of the non-protonated product and the higher acidity of ammonium in comparison to water.

The drop in the basicity of the aziridine nitrogen atom is also reflected in the existence of the non-protonated product structures when the solvent is represented by two water molecules. Such products, when they appear (as in the cases of the CHO, Br, and Cl substituents), are only 2 kcal/mol more stable than their protonated counterparts. Calculations with an unsubstituted aziridine ring and water molecules as explicit solvent yield only protonated PH structures (Table 3.1.1 and 3.1.2).

To summarize, one big advantage of inhibitors containing an aziridine ring in comparison to oxirane-based inhibitors is the possibility of N-substitution. The calculations presented here show that acceleration of the inhibition reaction of aziridine-based inhibitors could be achieved by the addition of electron-withdrawing N-substituents to the ring. In this way, the inhibitors with inhibition rates comparable to those of oxirane-based inhibitors could be generated.

Another important aspect is the low pH dependence of the kinetics of the inhibition reaction of aziridines with an electron-withdrawing substitute attached to the nitrogen. The activation energy in acidic surroundings (simulated by the inclusion of two explicit ammonium ions and a continuum solvent model) is at most only 1 kcal/mol lower than the corresponding calculations in the neutral environment. This property is extremely important if an inhibitor is to be used for medical purposes. A broad pH range of activity implies a broad range of possibilities for implementation.

Previously-studied systems with CHO attached to the N-position of the aziridine ring showed hydrolytic instability in DMSO and, furthermore, a possible reaction involving nucleophilic attack of the methylthiolate moiety at the carbonyl carbon of the substituent.⁵¹ Substituents described in this work have been experimentally found to show no such behaviour.⁵² A model reaction with 4-methoxythiophenolate in D₂O at pH 7.6 was experimentally performed and the ring opening reaction was followed by NMR spectroscopy. The characteristic NMR signals of the substituents at N remained constant, proving the stability of the inhibitor towards hydrolysis in DMSO. Furthermore, enzyme assays with FP2 showed an increased value for k_i , proving that these inhibitors do behave as originally expected. No competing reactions to nucleophilic attack at the C atoms of the ring occur.

An especially promising substituent is CF₂H, since it allows further substitution. The hydrogen of the CF₂H group could be substituted in such a way that it improves the stabilization of the enzyme-inhibitor non-covalent complex structure by interacting with one of the enzyme's subsites.

3.2 QM/MM and MD Calculations

3.2.1 The method

I: QM/MM method

The second step in the inhibition reaction of cysteine proteases (Section 2.1.4) is found to be irreversible. It takes place at the active site of the enzyme and comprises the creation of a covalent bond between the carbon atom of the Cys29 residue and the carbon atom of the inhibitor ring, as well as the breaking of a C-X (X = N or O) bond during the ring opening. If possible, a protonation of the X moiety also occurs. Due to the presence of bond-breaking and -forming processes, classical force field approaches are not the proper choice to study such systems, and QM methods have to be used. However, the enzymatic environment plays an important role in the inhibition. The His199 residue can act as proton source, and the nearby oxyanion-hole (the side chain of Gln23 and backbone NH-group of Cys29) are expected to stabilize the enzyme-inhibitor complex by non-covalent interactions. Also, the solvent has to be taken into account. Since such large systems cannot be treated quantum mechanically, the combined quantum mechanics/molecular modeling (QM/MM) approach becomes the method of choice.

QM/MM methods⁵³ divide the total system (enzyme, solvent and inhibitor) into the active centre and the rest. The active centre comprises all parts which are directly involved in the bond-breaking/formation processes and which for that reason have to be described by quantum chemical approaches. The rest of the system contains all those parts which indirectly influence the studied bond breaking/formation processes only via non-covalent intermolecular interactions. The QM and MM regions interact by electrostatic and van der Waals terms. Electrostatic interactions cause the electronic structure of the QM system to adapt to the electrostatic field exerted by the protein environment. In our work the electrostatic embedding scheme³⁷ was employed to cover these interactions. In this scheme the MM charges are incorporated into the one-electron Hamiltonian of the QM procedure. Dangling bonds at the QM/MM boundary are capped with hydrogen link atoms⁵⁴ in the framework of the charge-shift method.

The whole system used to model the irreversible step in the cysteine protease inhibition mechanism comprises the cysteine protease, the inhibitor, and additional water molecules. The X-ray structure of cathepsin B complexed with E64c as inhibitor was used as a starting point for the protease geometry (pdb file 1ITO⁵⁵). Water molecules are included to describe solvent effects. They need to be considered since the active sites of most cysteine proteases are solvent-accessible. The various systems calculated with the QM/MM approach differ in the composition of the QM region, while the classical part remains the same.

The QM part generally contains the electrophilic warhead of the inhibitor (with or without substituents) and the active site of the cysteine protease enzyme, consisting of the side chains of the His199 and Cys29 residues (cathepsin B numbering).⁵⁶ On the basis of a large number of experimental results¹⁰ the zwitterionic form (Cys29-S⁻ - His199-H⁺) was adopted. The protein residues which build up the oxyanion hole were not included in the QM part. Their interaction with the active site is mainly of electrostatic nature and is thus taken into account through the electrostatic embedding scheme. Nevertheless, since the role of the oxyanion hole in the inhibition reaction is important, additional computations in which the oxyanion hole was included in the QM part were performed. Reaction energies obtained for this larger QM part differed less than 0.3 kcal/mol (0.7 %) from the results obtained with our standard QM/MM division, proving its accuracy. The MM part also comprises the rest of the protein plus a water sphere to account for additional solvent effects.

The starting structure for the systems used in the QM/MM calculations was obtained by the following procedure. In the X-ray structure of the cathepsin B - E64c complex the E64c was replaced by the respective inhibitor. Protein and inhibitor were solvated in a water sphere ($r = 50$ Å) and optimized using the CHARMM program package.⁵⁷ In this optimization the water molecules of the outer layer (5 Å), as well as the atoms of the inhibitor, were kept fixed. For subsequent QM/MM calculations, only an inner water layer of $r = 10$ Å around the active site was included. All QM/MM computations employed the ChemShell software⁵⁸ using the TURBOMOLE program suite³⁸ for QM, and the DIPoly⁵⁹ code for MM calculations. Geometry optimizations were carried out with the hybrid delocalized internal coordinate (HDLC) optimizer.⁶⁰

QM calculations were done at the RI⁴⁰DFT/B-LYP⁴¹/TZVP⁴² level of theory for all geometry optimizations and PES scans. In most cases this approach was found to be sufficiently

accurate for the determination of the geometrical structure of the stationary points, although reaction barriers and energies were underestimated. Therefore, additional single point computations employing the DFT/B3⁴³-LYP/TZVP level were performed on stationary point structures to determine more precisely the activation and reaction energies of various processes. Points on the PES close to the transition state served as input for a TS search with the HDLC optimizer. The TS structures thus found were then relaxed to obtain the reactants and products and thus ensure that the stationary points were connected by a contiguous path. The geometries of the stationary points were used for the subsequent determination of barriers and reaction energies.

In some cases the B-LYP-calculated PESs were so flat in the vicinity of saddle point that we experienced difficulties in locating the transition state structures. Due to the poor curvature we could not find a good structure to optimize a TS from, and consequently the transition state structure calculation failed to converge. Therefore, in such cases we performed B3-LYP single point calculations for RIDFT/B-LYP-obtained structures surrounding the probable location of the TS, and in this way calculated an energy profile from which a better starting point for the TS calculation could be obtained. We then performed a TS search at the DFT/B3-LYP/TZVP level of theory, which proved to be sufficient for locating the saddle point. Subsequent optimizations to the product and reactant structures were also performed on the DFT/B3-LYP/TZVP level of theory.

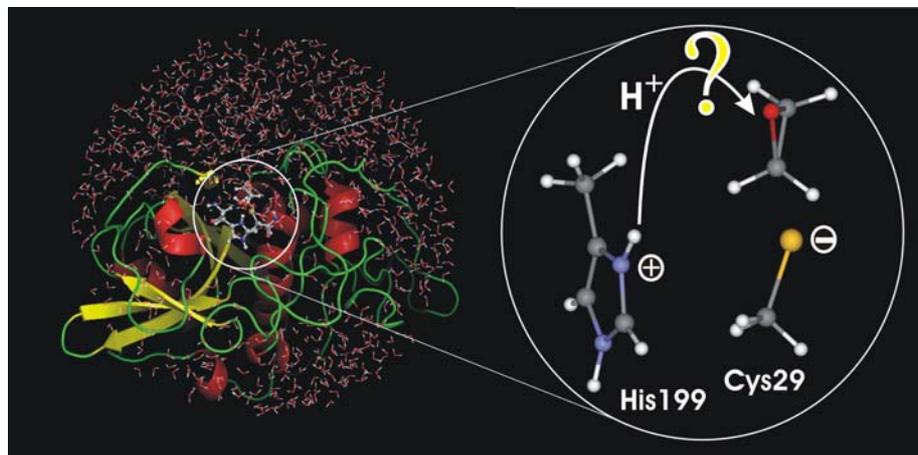
DFT is well known to offer excellent "value for money", in terms of quality of results against computational cost, for the description of many properties.^{61,62,63} However, this is often due to error compensation, which sometimes results in failure to get reasonable results,^{64,65,66,67} and in many cases multi-reference approaches are necessary to obtain reliable reaction paths.^{68,69,70,71} In the present case, however, the B3LYP/TZVP//BLYP/TZVP method has proven to be reasonably accurate.⁴⁴ Problems in treating protonated epoxides⁷² are not relevant to the present case since the protein environment is not sufficiently acidic for protonation of the oxirane ring. The protonation takes place after the TS on the reaction path, so that the kinetics are not influenced. The products are substituted alcohols, for which similar problems are not known.

II: Molecular Dynamics (MD) Runs

In addition to static QM/MM calculations, some systems were also submitted to exhaustive classical MD simulations in order to study their flexibility. All MD calculations were done by using the CHARMM program and force field. The starting structure was the QM/MM optimized (reactant and product) structure from which all water molecules were deleted. The structure was then solvated in a water sphere with a radius of 50Å. Since the overall structure had a charge of -7, seven counterions had to be added. We used seven sodium atoms and placed them at the outer surface of the solvent shell. The whole system was then submitted to a series of consecutive constrained optimizations, MD runs (heating from 100 to 300 K and equilibration at 300 K; time step 1 fs) and addition of further water molecules, until the MD runs no longer resulted in areas of low solvent density forming. During the whole procedure the enzyme and the inhibitor in the active site were allowed to “move” only during the optimization, and were held fixed during the MD runs. The outer 5 Å layer of water was kept fixed the whole time. This first step in our MD simulations was necessary to make sure that the molecules of the water sphere were well distributed along the protein surface as well as properly equilibrated within the solvation sphere, and that no bad contacts (strongly repelling groups abnormally close to each other) existed. After the solvation was done we started the “long” MD run, consisting of initial constrained optimization, heating to 300 K (15 ps), equilibration (200 ps), and sampling in a NVT system (200 ps) using the Berendsen thermostat.⁷³ In all MD runs the Verlet leap-frog algorithm⁷⁴ was used with a time step of 1 fs.

In MD simulations performed on the reactant structure of E64c the S atom of Cys29 was given a charge of -0.07. If Cys29 sulphur were charged it would interfere with the non-covalent interactions between E64c and the enzyme, since its covalent binding to the ring is not possible within the force field approximation. In other words, a charged S atom denied the possibility of reaction would force the inhibitor to adopt conformations that could not lead to the irreversible step of the inhibition reaction.

3.2.2 QM/MM treatment of inhibitor protonation by the His199 residue



Can the His199 residue donate its proton to the opening inhibitor ring within the enzyme environment?

Both the reversible and the irreversible step in the cysteine protease inhibition mechanism are influenced by the protein and general physiological environment. In the work of Helten *et al.*^{44,46} the influence of various surroundings on the reaction profile of the second step was investigated at a quantum-chemical level. Gas-phase calculations do not yield reasonable energy profiles since the shapes of the resulting potential curves are falsified by the deep minimum of the anion-molecule complex. In a polar medium this complex is strongly destabilized so that it no longer represents a local minimum. To model the effect of the surroundings, continuum solvation models were used to mimic the bulk effects of the solvent. Explicit water ($pK_s \approx 15$), ammonium ($pK_s \approx 9.3$) or formic acid ($pK_s \approx 3.8$) molecules were employed to assess the specific effects of neighbouring molecules. Using these models it was possible to explain the influence of the pH value of the surroundings and of the substituents on the inhibition potency of aziridine- and epoxide-based inhibitors. The computations were even able to contribute to the design of improved cysteine-protease inhibitors.⁵¹

Despite their successes, these models cannot account for the influence of the actual molecular structure of the protein environment which may also be important for the inhibition mechanism. The best example of such an influence is the controversial role of the His residue of the active site of cysteine proteases (His159 residue in papain numbering). All experimental findings agree that, in the active site of the cysteine proteases of the papain

family (Merops classification: Clan A, family C1), the Cys25 and His159 units (papain numbering) exist as an ion pair, Cys-S⁻/His-H⁺.¹ Due to its acidity the protonated His159 can act as a proton donor. Indeed, based on the bell-shaped dependence of the second-order rate constant of inhibition k_{2nd} (defined as $k_{2nd}=k_i/K_i$, see Section 2.1.4) on the pH value, it has been postulated for epoxide-based inhibitors that during the inactivation of papain the epoxide moiety is protonated by His159.⁷⁵ This suggestion was rejected by Varughese *et al.*¹⁹ Based on the crystal structure of papain alkylated by the epoxide-based inhibitor E64 they argued that the epoxide was most likely to have been protonated by water because of the distance of the resulting alcoholate from His159. Their finding was supported by Meara and Rich⁷⁶ who investigated the pH-dependence of the inhibition of papain by the carboxamide analog of EP-475. They did not find an alkaline pH dependence of the first-order rate constant k_i (see Section 2.1.4) for this process. The deviation from the bell shaped behaviour of k_{2nd} in earlier findings was explained by the influence of dissociation constant K_i on k_{2nd} . The insensitivity of k_i to increasing pH values was taken as evidence that His159 does not take part in the inhibition mechanism, since its participation should lead to a decrease of k_i at pH values high enough for His159 to be deprotonated. Our previous computations also indicated that the k_i value connected with the epoxide ring opening should not be influenced by the pH value of the surroundings.⁴⁴ They predicted that neither the ammonium ion ($pK_s \approx 9.3$) nor formic acid ($pK_s \approx 3.8$) protonates the epoxide before the transition state of the ring-opening reaction is reached. As a consequence, the barrier height, and hence k_i , change only slightly when the acidity of the environment increases. While experiment and theory agree on the influence of the possible protonation on k_i they are not necessarily in agreement about the role of His159. The computations clearly indicate that the protonation of the epoxide takes place after the transition state (TS) of the ring-opening reaction. This means that His159 could take part in the overall mechanism without influencing the reaction velocity.

Meara and Rich⁷⁶ also showed that at $pH \approx 3$ the k_i value ($\lg k_i \approx -1.5$) was 1-2 orders of magnitude lower than at $pH \geq 5$. This pH dependence was attributed to protonation of the cysteine moiety, which possesses a pK_a value of around 4. Theoretical computations have drawn the same conclusions.⁴⁴

For both *N*-unsubstituted and *N*-alkylated aziridine-based inhibitors the influence of the pH value on k_{2nd} is much larger than for epoxide-based inhibitors.^{49,77} This finding, too, is nicely reflected by our previous computations, which indicate that an ammonium ion already

protonates the aziridine group before the TS of the ring-opening reaction is reached. As a result of protonation the energy barrier drops considerably, so protonation influences both kinetics and thermodynamics. These findings lead to the hypothesis that the protonated His159 influences the inhibition mechanisms of epoxides and aziridines in different ways: it affects both the kinetics and thermodynamics of the aziridine reaction, but only the thermodynamics of the epoxide reaction.

Investigations at the purely quantum mechanical level predicted that the His199 residue of the active site can serve as proton donor in the inhibition reaction (Section 3.1.2). To examine how the reaction is affected by inclusion of the protein and solvent environment we turned to the QM/MM approach. As in the previous QM treatment, the inhibitor was represented only by the electrophilic warhead, i.e. either an aziridine or an oxirane ring. The side chains of the inhibitor were omitted since their role in the inhibition process is limited to reversible, noncovalent interactions between the protease and the inhibitor which influence the formation of the enzyme-inhibitor complex but not the ring-opening reaction itself.

To test whether water can still mediate the proton transfer step when the surrounding solvent and proteins are treated explicitly, rather than by COSMO, we performed computations both with and without the mediating water molecule in the QM part. This leads to four different model systems (AZR, AZR-H, OX, and OX-H), which are depicted in Figure 3.2.1. The abbreviation AZR indicates the model system with the aziridine warhead, and AZR-H refers to the same system but with the mediating water molecule included in the QM part of the system. OX and OX-H are the corresponding abbreviations for the oxirane system.

The reaction coordinates used for the construction of the PES are, like in the previous QM calculations, the distance between the attacking nucleophile and the attacked C atom of the inhibitor ring and the angle corresponding to ring opening (Figure 3.2.1). The computed PESs for the model systems AZR, AZR-H, OX, and OX-H are shown in Figures 3.2.2 and 3.2.3. The energies are given relative to the energies of the corresponding reactants. The structures of the stationary points are depicted in Figures 3.2.4 and 3.2.5. The relative energies of the various stationary points at the B3LYP/TZVP level are collected in Table 3.2.1.

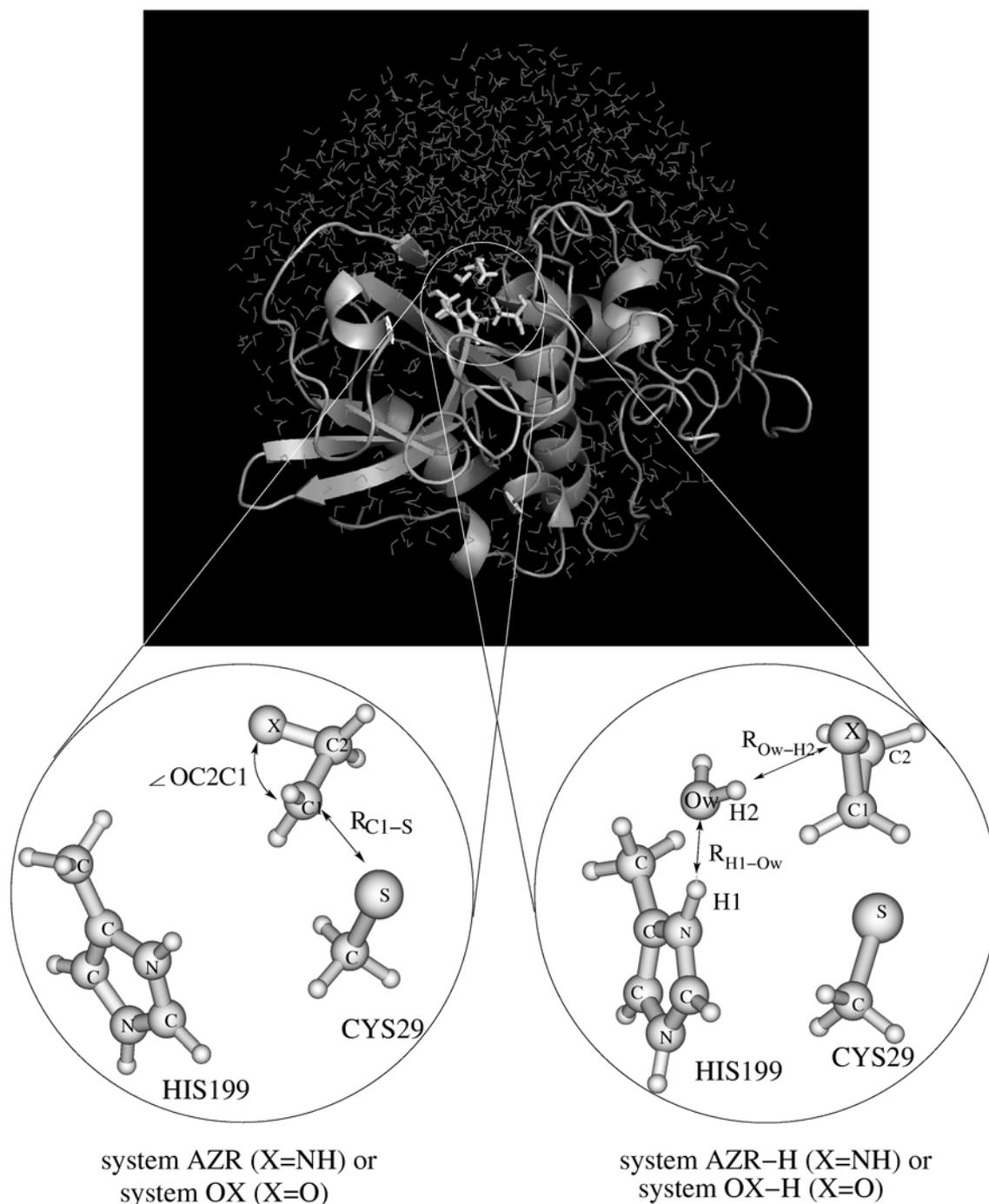


Figure 3.2.1. The overall structure of the model system used in the QM/MM computations (top). The part treated quantum mechanically is shown at the bottom. X is oxygen for oxirane systems or an NH moiety for aziridine systems. Bottom left: the reaction coordinates used to examine the reaction. Bottom right: parameters describing the proton-transfer process. Below these is listed the notation used to refer to the QM-treated systems (AZR, AZR-H, OX, and OX-H), in which those including a water molecule are denoted with -H.

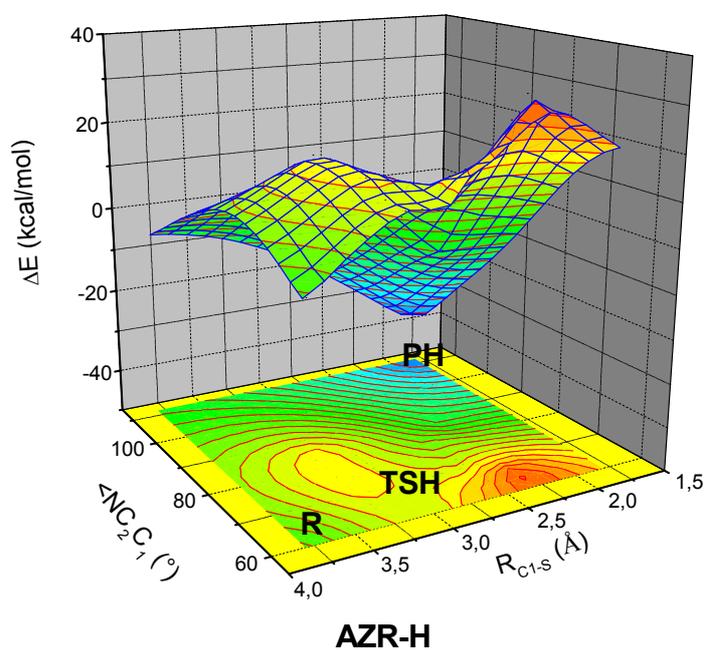
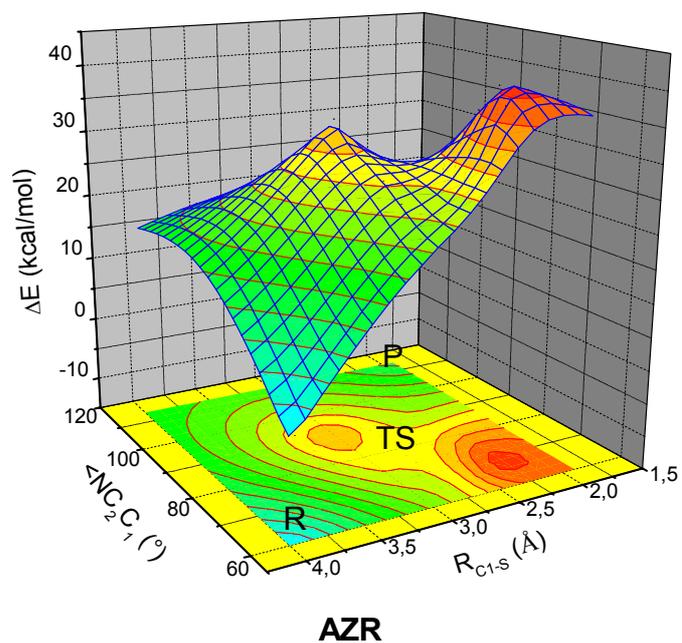


Figure 3.2.2. PES obtained in QM/MM calculations for the AZR system (top) and the AZR-H system (bottom).

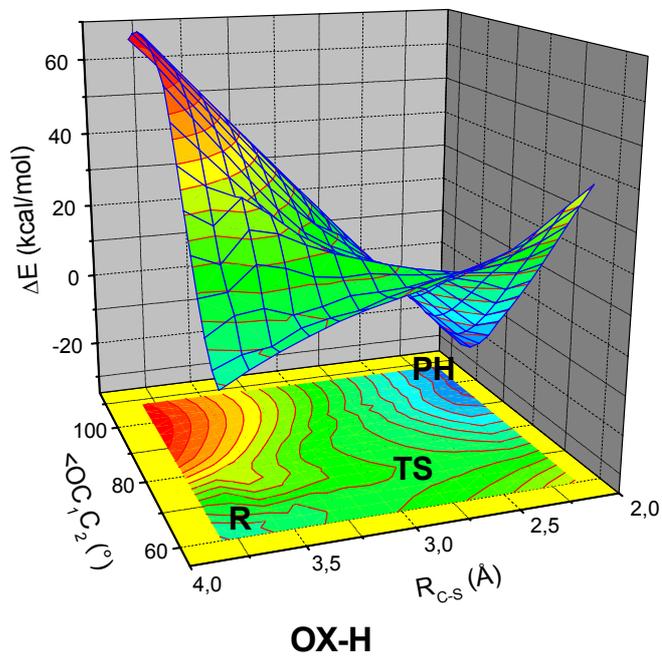
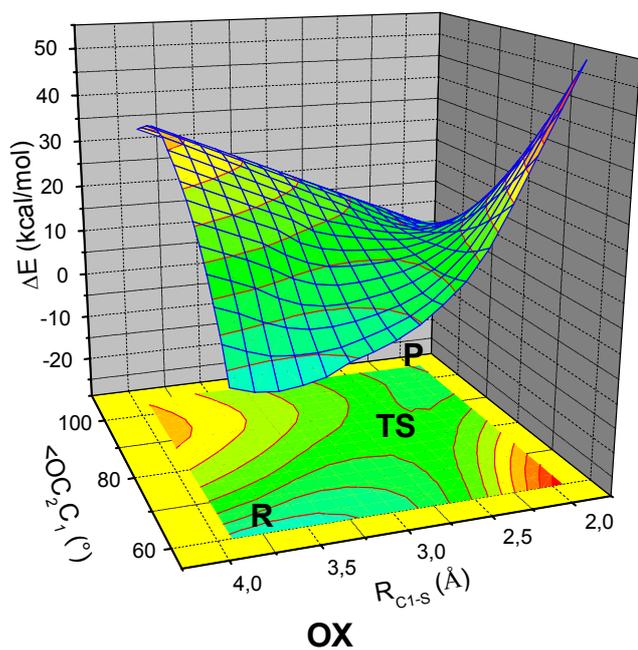


Figure 3.2.3. PES obtained in QM/MM calculations for the OX system (top) and the OX-H system (bottom).

The upper part of Figure 3.2.2 shows the PES for the AZR system, which features aziridine as the inhibitor but contains no additional water molecule between His199 and the inhibitor. The profile of the PES (BLYP/TZVP level) indicates a high reaction barrier and an endothermic reaction. At the B3LYP/TZVP level a barrier of 30 kcal mol⁻¹ and a reaction energy of 12 kcal mol⁻¹ are predicted (Table 3.2.1). In Figure 3.2.4 some selected geometrical parameters for the reactants (R), transition states (TS), and products (P) are shown. For the transition state of the AZR system we find $\angle \text{NC}_2\text{C}_1 = 79^\circ$ and $R_{\text{Cl-S}} = 2.71 \text{ \AA}$, indicating that the transition state is slightly more reactant-like ($\angle \text{NC}_2\text{C}_1 = 60^\circ$) than product-like ($\angle \text{NC}_2\text{C}_1 = 113^\circ$). This is unexpected considering the endothermicity of the reaction. From the energy contour plot of the PES (Figure 3.2.2, upper part) it is obvious that both $R_{\text{Cl-S}}$ and $\angle \text{NC}_2\text{C}_1$ contribute to the reaction coordinate. At the beginning of the reaction the change is mostly in $R_{\text{Cl-S}}$, while $\angle \text{NC}_2\text{C}_1$ changes very little. Just before the TS $\angle \text{NC}_2\text{C}_1$ starts to vary as well. The proton remains at His199 during the whole reaction, i.e. the inhibitor is not protonated. Protonation cannot take place since the distance between the proton of the histidine moiety and the N centre of the inhibitor is too large. It decreases during the reaction, but even for the products the distance is 3.58 Å.

The AZR-H system includes a water molecule in the QM region, bridging the distance between the histidine moiety and the N atom of the inhibitor. The PES for this system is given in the lower part of Figure 3.2.2. The corresponding geometrical changes along the reaction path are shown at the bottom of Figure 3.2.4. The large difference in the energy profiles of the PESs of AZR and AZR-H is obvious. Going from AZR to AZR-H, the barrier height significantly decreases and the reaction becomes considerably exothermic. The reaction energy calculated using the B3LYP/TZVP method drops from 12 to -36 kcal mol⁻¹, while the barrier decreases from 30 to only 16 kcal mol⁻¹ (Table 3.2.1). This is caused by the protonation of the emerging azanide group (Figure 3.2.4, lower part). The proton from HisH⁺199 - mediated by the water molecule - is transferred to the nitrogen of the inhibitor ring. Due to the strong basicity of the emerging azanide group, the protonation occurs almost at the very beginning of the inhibition reaction. The computations predict a stable protonated reactant structure which lies only 1.3 kcal mol⁻¹ above the non-protonated reactant (RH in Table 3.2.1 and Figure 3.2.4). Protonation also changes the geometrical arrangement of the TS considerably (Figure 3.2.4). For the TS we find an $\angle \text{NC}_2\text{C}_1$ angle of 66° and an $R_{\text{Cl-S}}$ distance of 3 Å. Compared to the values predicted for the AZR system (without protonation), this represents a more reactant-like TS, which is as expected from the Hammond postulate,

since the reaction is now quite exothermic. The strong shift of the TS towards the reactants is also obvious from the contour plots of the PES (Figure 3.2.2). The contour plot for the AZR-H system also shows that the R_{Cl-S} distance dominates the reaction coordinate at the beginning of the reaction, as in the case of the AZR system. The $\langle NC_2C_1$ angle changes even less than for AZR before the TS, whereas after the TS it becomes the dominant contributor to the reaction coordinate. The products of the AZR and of the AZR-H systems possess quite similar values for $\langle NC_2C_1$ and R_{Cl-S} (Figure 3.2.4).

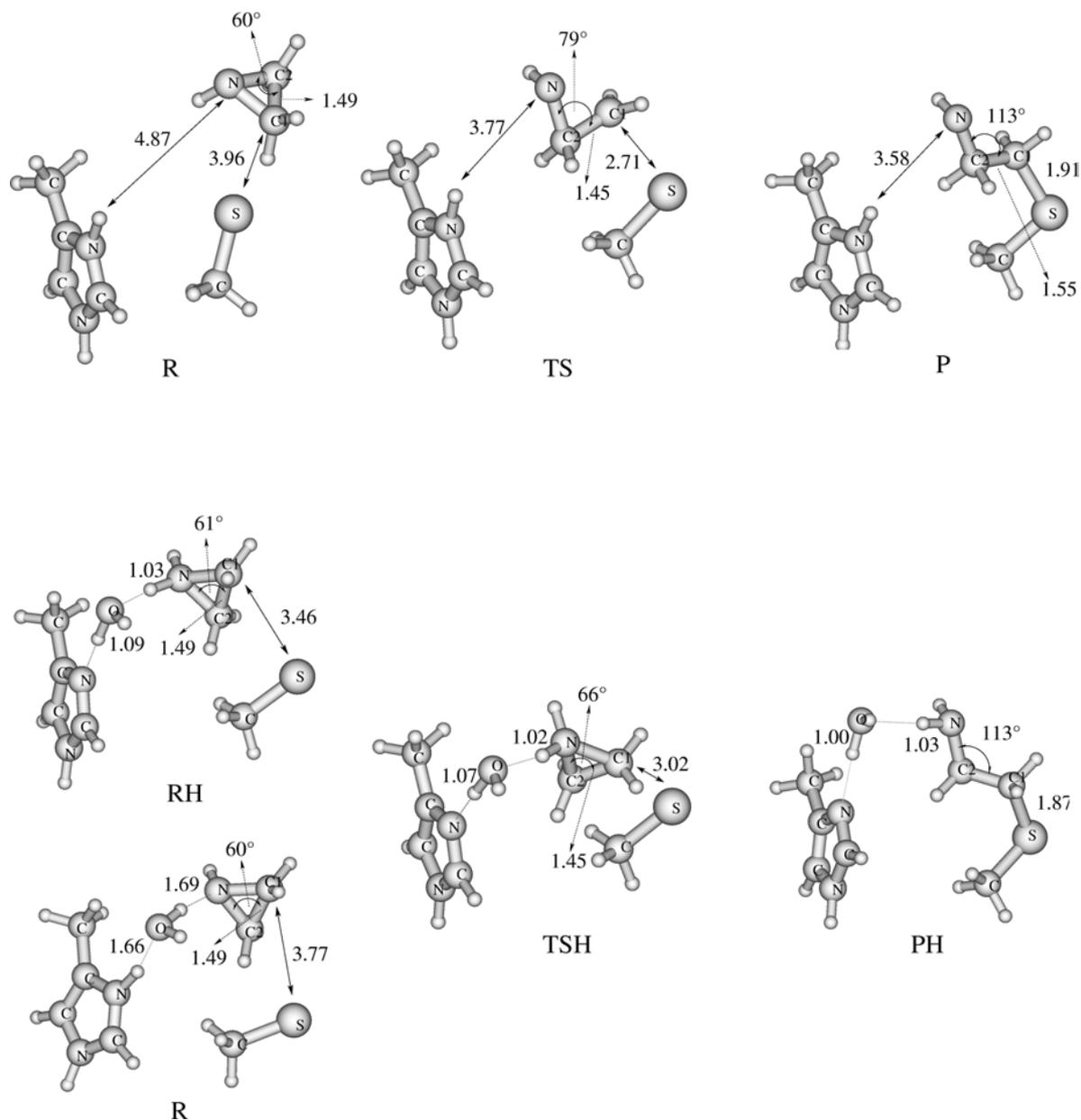


Figure 3.2.4. Selected geometrical parameters for the stationary point of the reaction profile of the AZR (top) and AZR-H system (bottom). All distances are given in Å and angles in degrees ($^\circ$).

It is important to note that the geometry of the OX product was derived from the available X-ray data, albeit with oxirane substituents replaced by hydrogen. The initial guess for the AZR product structure was obtained by replacing the oxygen atom of the OX product with NH.

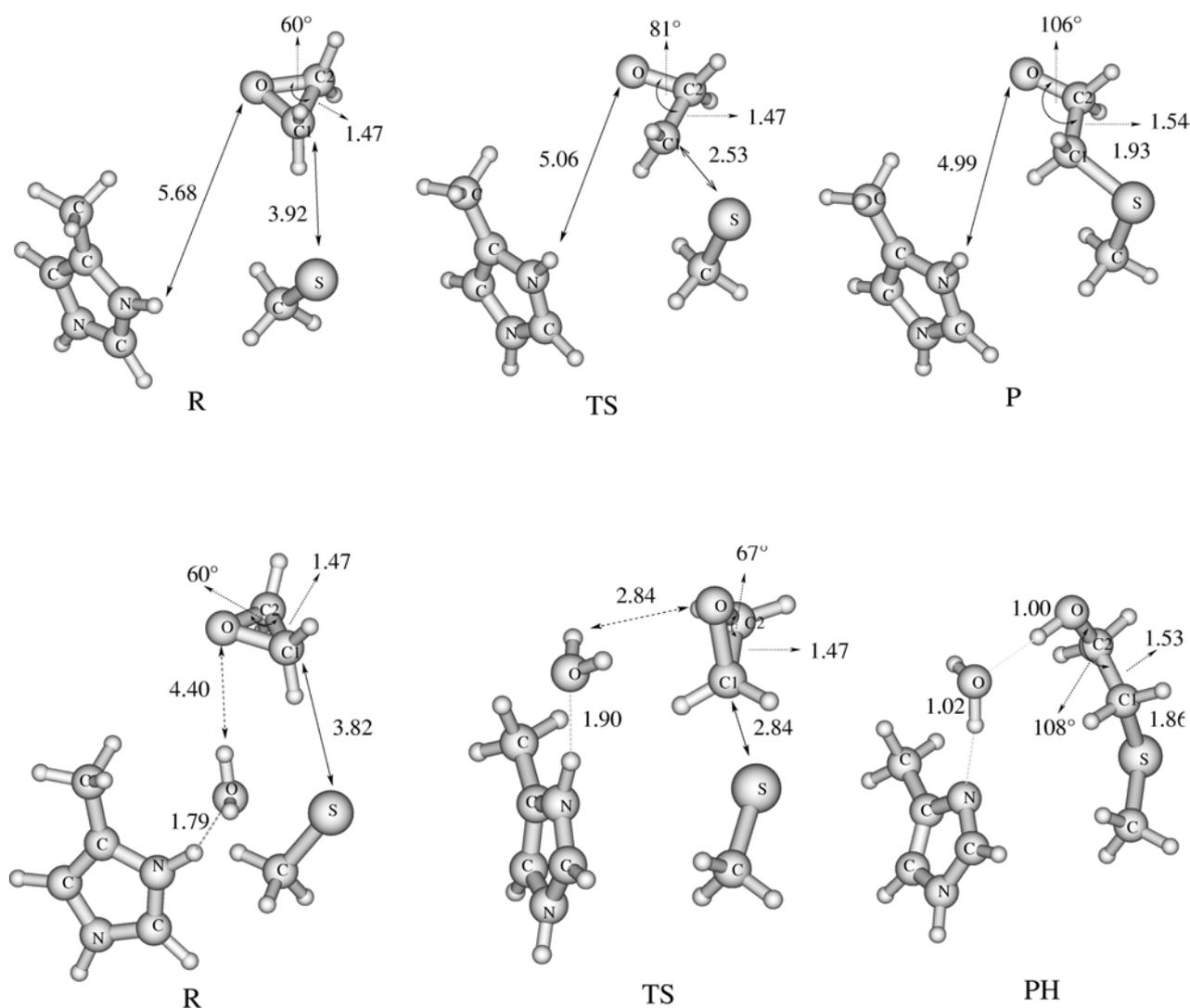


Figure 3.2.5. Selected geometrical parameters for the stationary point of the reaction profile of the OX (top) and OX-H systems (bottom). All distances are given in Å and angles in degrees (°).

In Figure 3.2.3 the potential energy surfaces for the systems containing oxirane as inhibitor are shown. The chart at the top shows the PES for the OX system, which contains no additional water molecule as possible proton mediator. The relative energies of the stationary points at the B3LYP/TZVP level are given in Table 3.2.1. The data indicate an endothermic reaction for the OX system, with an activation energy of 18 kcal mol⁻¹ and a reaction energy of about 9 kcal mol⁻¹. Some selected geometrical parameters for the OX system are given in Figure 3.2.5 (upper part). We obtain $\angle\text{NC}_2\text{C}_1 = 81^\circ$ and $R_{\text{Cl-S}} = 2.5 \text{ \AA}$ for the TS, and $\angle\text{NC}_2\text{C}_1 = 106^\circ$ and $R_{\text{Cl-S}} = 1.93 \text{ \AA}$ for the product. The geometrical parameters (Figures 3.2.4, 3.2.5) and the contour plots of the PESs (Figures 3.2.2, 3.2.3) show that the TS of the OX system is more product-like than its AZR counterpart. As for the AZR system no protonation of the inhibitor takes place along the reaction path. The reason is again the large distance between His199 and the inhibitor.

Table 3.2.1. Relative energies of the stationary points (R = reactant, TS = transition state, P = product) of the various PESs calculated at the B3LYP/TZVP level. All energies are given relative to the reactants (kcal mol⁻¹). An added H (e.g. TSH) indicates the relative energies of the structures in which the proton moved from the His199 moiety to the inhibitor.

	ΔE (kcal/mol)				
	RH	TS	TSH	P	PH
AZR	-	30	-	12	-
AZR-H	1.3	-	16	-	-36
OX	-	18	-	9	-
OX-H	-	17	-	-	-36

The computed PES of the OX-H system containing a water molecule to bridge the gap between the His199 moiety and the inhibitor is depicted in the lower part of Figure 3.2.3. Comparing the two PESs in Figure 3.2.3 (see also Table 3.2.1), it is obvious that the main change is from an endothermic to an exothermic reaction (9 vs. -36 kcal mol⁻¹ reaction energy) while the barrier height remains nearly constant (18 vs. 17 kcal mol⁻¹). A closer look at the geometrical arrangements of the TS and products (Figure 3.2.5) shows the reason. The basicity of the emerging alcoholate is less than that of the emerging azanide group of the AZR-H system. The result is that only the product gets protonated in the OX-H system, while at the TS the proton remains at the His199 residue. The proton shift is again mediated by a water molecule. It takes place at about $\angle\text{NC}_2\text{C}_1 = 85^\circ$ and $R_{\text{Cl-S}} = 2.6 \text{ \AA}$. It is interesting to note

that for $\langle \text{NC}_2\text{C}_1 \rangle > 90^\circ$ the proton shift occurs at all $R_{\text{Cl-S}}$ distances. One would expect the PES to differ from that of the OX system only for those reaction stages in which the proton shift has already happened in the OX-H system. However, this is not the case, as can be seen from Figure 3.2.3. It shows that introduction of the water molecule to form OX-H moves the TS towards the reactants, as would be expected from the Hammond postulate. This may also be seen by examining the structures shown in Figure 3.2.5.

Our computations clearly indicate that protonation of epoxide- or aziridine-based inhibitors by the protonated histidine residue can only take place if the gap between HisH^+199 and the inhibitor is bridged by a water molecule. This is in line with the analysis of Varughese *et al.*¹⁹ who ruled out direct proton transfer from the His199 to the inhibitor on the basis of X-ray investigations revealing a large distance between these moieties. However, the computations also show that in both cases a single water molecule can establish a very efficient relay system that allows the proton to move quite easily from the His199 to the heteroatom of the inhibitor. HisH^+199 can thus act as the proton source in the inhibition mechanism, provided that water molecules can penetrate deeply enough into the active site of the protein to establish the required relay system. This is supported by X-ray experiments showing the presence of such water molecules in the binding pocket of the cathepsin B–E64c complex. HisH^+199 is a more efficient proton source than water, as can be seen from the pK_a values of both compounds (6.8 vs. 15). Thus, water serves only as a bridge, and not as a proton source itself.

The importance of inhibitor protonation can be seen from the computed energies. These indicate that without protonation, the aziridine-based inhibitors would not be active at all, since the reaction barrier is much too high and the reaction is predicted to be endothermic. The kinetics of the epoxide reaction are less influenced by protonation, but the computations show that for such inhibitors the overall reaction energy is only sufficiently negative to make the reaction effectively irreversible if protonation can take place. Both findings are consistent with our previous calculations.⁴⁴ For both inhibitor types the $\text{HisH}^+199\text{-H}_2\text{O-inhibitor}$ relay system must be considered important for a wide range of pH values due to the proximity of the proton source (HisH^+) and its high efficiency in the proton transfer.

Our results are supported by the solvent accessibility of the active site and by the results of X-ray investigations. The latter show that single water molecules are present in the active site,

and hence, that the predicted relay system can indeed be formed. Given its importance for the activity of aziridine-based inhibitors, substituents which block the formation of the water bridge would strongly diminish the activity of the inhibitors. On the other hand, substituents which stabilize the emerging azanide through, for instance, charge delocalization would greatly enhance the activity of aziridine-based inhibitors at higher pH-values where the histidine residue is only partially protonated. For inhibitors which would need to operate only at lower pH values such substituents would be less important, as proton transfer would boost their activity. The reason for the strong pH dependence is that the azanide formed during the reaction of aziridine inhibitors is so basic that the proton switches position by the time the TS is reached. The enhancement of inhibitor activity by N-substituents has in fact been found experimentally using formyl as substituent.

3.2.3 QM/MM calculations clarifying the inhibition potency of epoxide- and aziridine- based inhibitors

Epoxides and aziridines are important compounds in polymer,⁷⁸ supramolecular,⁷⁹ and organic chemistry.^{80,81} In medicinal chemistry epoxysuccinic acid derivatives and analogous aziridines are important as building blocks for inhibitors of cysteine proteases,⁶ which present promising drug targets for many diseases.⁸⁻¹⁴ Some epoxysuccinic acid derivatives have even reached clinical development as anti-cancer drugs.¹⁵ Like epoxysuccinic acid derivatives aziridine-2,3-dicarboxylic acid derivatives have been proven to be selective, non-toxic cysteine protease inhibitors suitable as antiparasitic¹⁶ and affinity labelling agents.¹⁷

The activity of the overall inhibition process is determined by a complicated interplay between substituents and environmental effects. The results of the *experimental* findings may be shortly summarized as follows:

1. The known difference between epoxides and aziridines in susceptibility to nucleophilic ring opening is reflected in their inhibition potencies: for pH values around 7 the inhibition data reveal lower second-order rate constant k_{2nd} values (for definition see Section 2.1.4) for aziridines, but for pH values around 4, *N*-unsubstituted and *N*-alkylated aziridines are as potent as epoxides. The lower inhibition potency at pH 6 – 7 is mainly due to lower k_i values.

2. For both the epoxide- and the aziridine-based inhibitors a free acid attached to the three-membered ring enhances inhibition potency compared to esters.^{49,83,84,85} However, this is only the case for compounds which bind solely to the non-primed binding sites of the proteases (E-64 analogs).
3. An experimental differentiation between the reversible (described by the dissociation constant K_i) and irreversible (described by the first-order rate constant k_i) step of the inhibition has been achieved by Meara and Rich with E-64 analogues.^{76,83} The following order of inhibition activity was found for ring substituents:⁷⁶ $\text{CO}_2\text{H} > \text{CONH}_2 > \text{COCH}_3 > \text{CO}_2\text{Et} \approx \text{CO}_2\text{Me} > \text{CH}_2\text{OH} \approx \text{H}$. The replacement of the acid (CO_2H) by hydroxamic acid (CONHOH), amide (CONH_2), or ketone (COCH_3) mainly influences K_i while k_i remains nearly unaffected.
4. These results obtained for the ring-opening reaction of epoxide-based inhibitors by the Cys residue of the proteases are not in line with the ring-opening reactions of epoxycarbonyl compounds with methylthiolate in solution.⁸⁶ For these reactions Bihovsky found the following trend for the rate constant k : $\text{COCH}_3 > \text{CO}_2\text{CH}_3 > \text{CONH}_2 > \text{CO}_2^-$.

This list demonstrates the overwhelming amount of experimental data available on the subject; however, the picture about the basic principles which govern the structure-activity relationships (SARs) is still very incomplete. Quantum mechanical calculations^{44,46} were able to provide a complete picture of the ring-opening reaction with methylthiolate in a polar solvent. However, corresponding information on the reactions in enzyme environments are missing. In this work various model systems have been simulated by means of QM/MM methods that combine quantum mechanical methods (QM) for the description of bond-forming and -breaking processes and molecular mechanical (MM) approaches to account for the influence of the protein environment and the surrounding solvent. These provided a static picture of the reaction by optimizing the geometry to a saddle point or minimum. The influence of dynamical effects was also investigated by performing additional Molecular Dynamic (MD) simulations. By comparing the various model systems the computed data allow a detailed analysis of the kinetics and thermodynamics of the mode of action of epoxide- and aziridine-based inhibitors and elucidation of the main interactions governing these processes.

In the calculations the system shown in Figure 3.2.6 was used. Different inhibitor structures were obtained by varying the three-membered ring ($X = O, NH$) or by addition of a peptidomimetic chain in the R position. By varying the reaction coordinates ($\langle XC3C2$ and R_{C2-S}) potential energy surfaces for different reactions corresponding to the irreversible part of the inhibition mechanism were constructed.

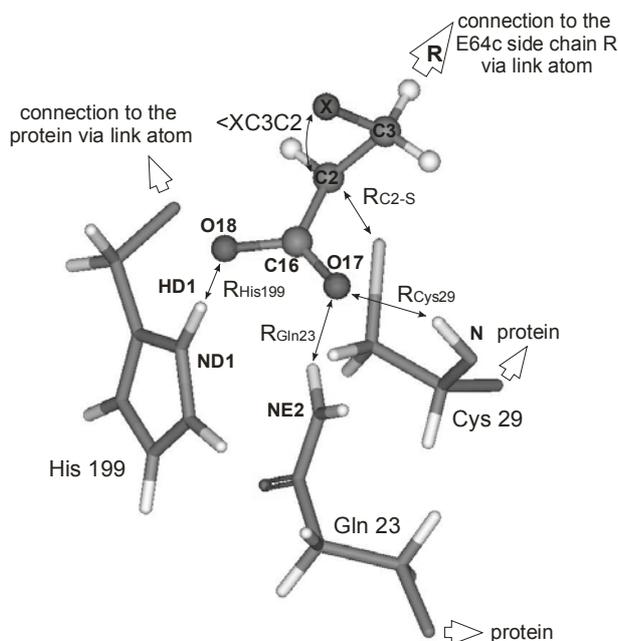


Figure 3.2.6. Composition of the QM part, description of reaction coordinates ($\langle XC3C2$ and R_{C2-S}), and residues forming the oxanion hole (Gln23 and Cys29), as well as the notation for the hydrogen bonds in the active site. Atom numbering corresponds to the 1ITO pdb structure. C2 atom has *S* absolute configuration. R denotes the side chain of the E64c inhibitor. X is O, NH or NH_2 .

Oxirane-based inhibitors

The kinetics and thermodynamics of the irreversible step of the inhibition process will be discussed in terms of computed reaction barriers and energies listed in Table 3.2.2, which gives the relative energies (difference from reactant state R in kcal mol^{-1}) of the transition states TS (i.e. the reaction barrier) and the protonated (PH) or deprotonated (P) products (i.e. the reaction energies). PH and P differ in the protonation of the alcoholate group which results from the ring opening of the oxirane. As already noted before (see previous section) the transition state invariably remains unprotonated for oxirane-based inhibitors even if e.g. NH_4^+

Chapter 3 Results and discussion

is used as proton source. The reactants R and products PH or P are obtained by optimizations along the intrinsic reaction path starting from the transition state (TS) structure. To distinguish between the various effects the inhibitor was modeled by an oxirane ring with a carboxylate substituent (OX-COO⁻; columns 4 and 6 in Table 3.2.2, X=O in Figure 3.2.6) and by the complete E64c inhibitor (E64c; columns 5 and 7 in Table 3.2.2, X=O and R=Leu-NH-(CH₂)₂-CH(CH₃)₂ in Figure 3.2.6). In all cases the computations simulate the attack of the thiolate group of the Cys29 residue on the C2 centre, which is the only reaction found by X-ray investigations (although attack on the C3 centre is also conceivable; we address this issue in 3.2.4.). To elucidate the influence of the enzyme environment computations were performed modeling the reaction in the presence of both a protonated (abbreviated as protonated His199) and a deprotonated His199 residue (abbreviated as deprotonated His199; as for normal His199 but with hydrogen HD1 removed (see Figure 3.2.6)). PESs for all of the systems are given in Figures 3.2.7 and 3.2.8 For comparison, Table 3.2.2 also contains the energies obtained for the reaction of an unsubstituted oxirane ring with the enzyme (OX (previous Section); Table 3.2.2 column 3) and for the reaction of a thiolate with unsubstituted and substituted oxiranes in a polar solvent, represented by COSMO³⁹ (solvent model OX, OX-COO⁻; Table 3.2.2 column 8, 9).⁴⁶ A characterization of the stationary points by selected geometrical parameters is given in Table 3.2.3. In addition to the length of hydrogen-bond bridges it also contains important distances between heavy atoms for comparison with X-ray data. The definitions for the various parameters are also illustrated in Figure 3.2.6. The computed geometrical arrangement of the reaction products for the E64c inhibitor obtained for an enzyme with a protonated and deprotonated His199 residue is shown in Figure 3.2.9.

Table 3.2.2. Relative energies of the stationary points of the reaction profiles computed for the reactions of oxirane-based inhibitors relative to the reactants R. All energies are given in kcal/mol. For definition of the different systems see text.

entry		protonated His199			deprotonated His199		solvent model ⁴⁶	
		OX	OX-COO ⁻	E64c	OX-COO ⁻	E64c	OX	OX-COO ⁻
1	R2	--	--	--	-7	--	--	--
2	R	0	0	0	0	0	0	0
3	TS	17	2	1	2	1	14	17
4	PH	-36	-47	^a n.p.	-56	^a n.p.	-26	-38
5	P	--	^a n.p.	-22	^a n.p.	-33	^a n.p.	^a n.p.

^a not performed

For protonated products (entry 4 of Table 3.2.2) the exothermicity of the ring-opening reaction is greater than 40 kcal mol⁻¹, and even for non-protonated products values of more than 20 kcal mol⁻¹ are predicted. This indicates that the ring-opening reaction is irreversible even if no protonation can take place (e.g. at pH values greater than 9). Since no new information about the irreversibility would be obtained we refrain from computing protonated products for all model systems.

Table 3.2.3. Selected geometrical parameters for the stationary points of the reaction profiles computed for the reactions of oxirane-based inhibitors. All distances are in Å and angles in degrees (°). For definition of the various distances see Figure 3.2.6 and Figure 3.2.9.

entry		R _{C2-S}	∠OC3C2	R _{O18-ND1}	R _{O17-NE2}	R _{O17-NCys29}	R _{O5-Gly74}	R _{N6-Gly198}
QM/MM:OX-COO⁻ / protonated His199								
1	R	3.07	64.7	1.66/2.72	1.94/2.91	2.31/3.25	--	--
2	TS	2.81	71.1	1.62/2.62	1.92/2.88	2.20/3.14	--	--
3	PH	1.92	112.2	1.58/2.65	1.88/2.84	1.99/2.94	--	--
QM/MM:E64c / protonated His199								
4	R	3.07	64.0	1.65/2.68	2.21/3.15	2.97/3.85	2.00/2.94	3.88/4.02
5	TS	2.80	72.4	1.64/2.65	2.15/3.10	2.67/3.59	1.99/2.93	3.69/4.03
6	P	1.91	111.0	1.67/2.64	1.91/2.88	1.96/2.91	2.01/2.95	3.75/4.05
MM/MD: E64c / protonated His199								
7	P	1.91	111.0	1.69/2.66	1.87/2.82	2.05/3.00	2.20/3.11	2.26/3.10
QM/MM:OX-COO⁻ / deprotonated His199								
8	R2	3.48	61.0	--/4.18	2.16/3.04	3.40/4.19	--	--
9	R	3.38	61.5	--/4.02	2.02/2.97	3.23/4.08	--	--
10	TS	2.96	67.0	--/3.95	1.99/2.94	2.82/3.74	--	--
11	PH	1.87	113.7	--/3.66	1.81/2.80	1.95/2.92	--	--
QM/MM:E64c / deprotonated His199								
12	R	3.41	63.5	--/4.38	2.26/3.14	2.13/2.93	1.91/2.85	2.91/3.83
13	TS	3.14	70.7	--/4.30	2.28/3.15	2.08/2.91	1.91/2.85	3.18/4.00
14	P	1.92	111.1	--/4.20	2.17/3.05	1.80/2.73	2.01/2.93	3.73/4.28
MM/MD: E64c / deprotonated His199								
15	P	1.92	111.1	--/3.69	2.11/2.87	1.91/2.93	2.33/3.22	2.57/3.38

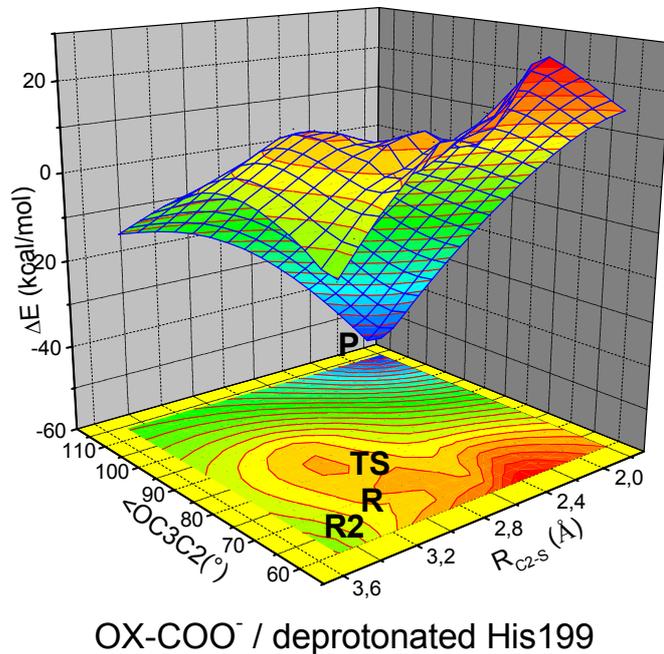
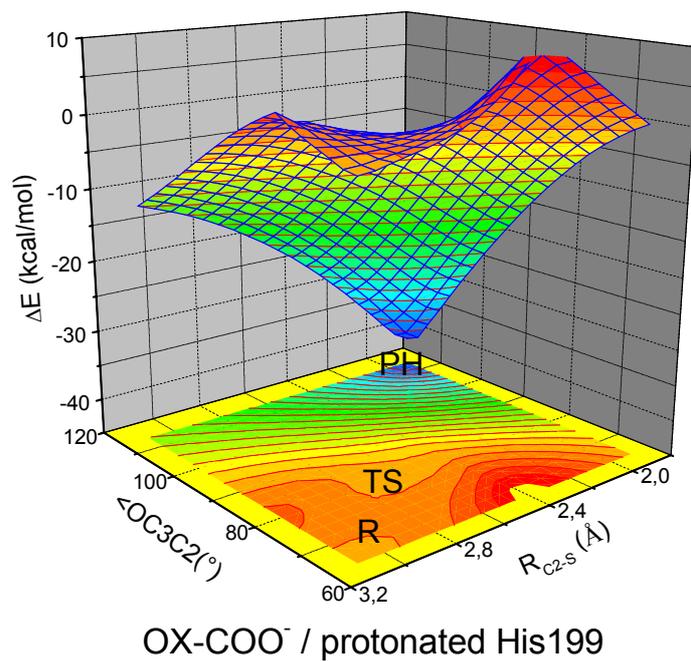
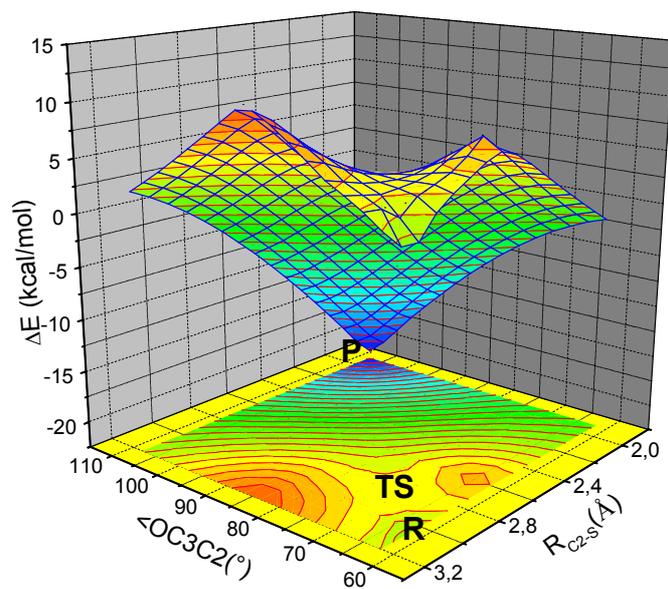
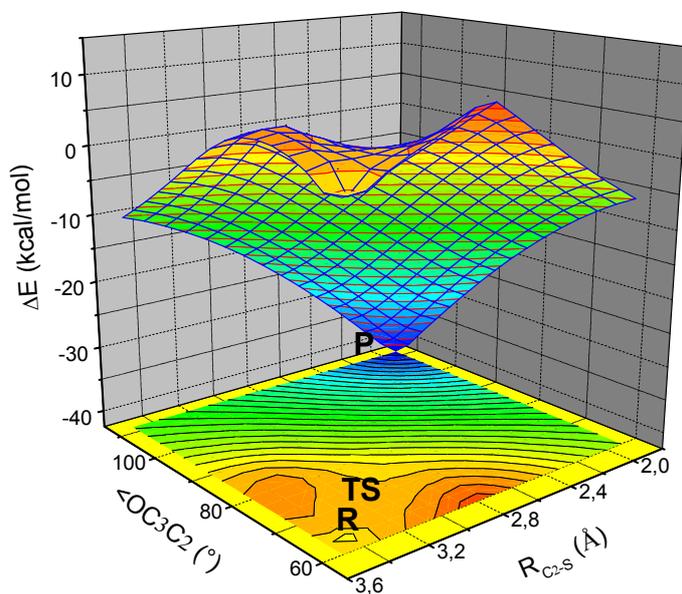


Figure 3.2.7. PES obtained in QM/MM calculations for the OX-COO⁻ system in the presence of protonated (top) and deprotonated (bottom) His199 residue.



E64c / protonated His199



E64c / deprotonated His199

Figure 3.2.8. PES obtained in QM/MM calculations with the E64c inhibitor in the presence of the protonated (top) and deprotonated (bottom) His199 residue of the enzyme.

All model inhibitors studied lead to irreversible inhibition, but variations in the computed reaction barriers for the different model systems indicate qualitative differences in the inhibition mechanism. Let us first concentrate on situations in which the His199 residue is protonated (columns 2-4 in Table 3.2.2, entries 1-7 in Table 3.2.3). For an oxirane ring without any substituents a reaction barrier of 17 kcal mol⁻¹ is predicted (Table 3.2.2, TS, column 3). For such barrier heights the generally-accepted two-step model is valid, since the corresponding ring-opening reactions are so slow that the formation of the reversible enzyme-inhibitor complex can indeed reach equilibration. If, however, a carboxylate substituent is attached, the barrier drops by about 15 kcal mol⁻¹ to 2 kcal mol⁻¹ (Table 3.2.2, TS, column 4). Including the whole E64c molecule in the calculation causes no significant change in the barrier (Table 3.2.2, TS, column 5). For such small energy barriers the reaction can be expected to be so fast that the formation of the reversible enzyme-inhibitor complex cannot reach equilibration anymore. As this would essentially merge the two steps of the inhibition into one single step, discriminating between influences on the first (represented by the equilibrium constant *K*) and the second step (represented by the reaction constant *k_i*) should be done with caution.

In contrast to the large decrease it causes in an enzyme environment, in solution addition of a carboxyl group slightly increases the activation barrier (Table 3.2.2: last two columns), indicating the strong influence of the enzyme environment on the irreversible step. The interactions involved can be examined from the geometrical data in Table 3.2.3. Computations in which the inhibitor is solely mimicked by the epoxide warhead with a carboxylate substituent indicate a very strong hydrogen bond between O18 and HD1 (Table 3.2.2: $R_{O18-ND1} = 1.6-1.7 \text{ \AA}$). Additional hydrogen bonds are found to the oxyanion hole, i.e. between O17 and the NH₂ group of the Gln23 residue ($R_{O17-NE2} \approx 1.9 \text{ \AA}$) and between O17 and the NH group of the Cys29 residue ($R_{O17-NCys29} = 2.0 - 2.3 \text{ \AA}$). The short distances show these to be strong hydrogen bonds and thus potentially important factors affecting the inhibition process. The influence becomes clear if the geometrical parameters of the reactants are compared. For the reaction of an unsubstituted oxirane with the active site of cathepsin B previous studies⁸⁷ predicted $R_{C2-S} \approx 4 \text{ \AA}$ and $\angle OC3C2 = 60^\circ$ for the reactants. For the transition state $R_{C2-S} = 2.8 \text{ \AA}$ and $\angle OC3C2 = 67^\circ$ were computed. For the corresponding reaction of a carboxylated oxirane the present study predicts $R_{C2-S} \approx 3 \text{ \AA}$ and $\angle OC3C2 \approx 65^\circ$ for the reactants and $R_{C2-S} \approx 2.8 \text{ \AA}$ and $\angle OC3C2 = 71^\circ$ for the transition state (Table 3.2.3,

entries 1 and 2), i.e. the interactions strongly influence the reactants, even though the transition state structures are quite similar. This indicates that the strong interactions between the carboxylate moiety and enzyme pull the inhibitor into the active site, towards the attacking thiolate group of the Cys29 residue. This makes the reactants more similar to the transition state, allowing the reaction to start from further along the reaction path.

If the carboxylated oxirane is replaced by the whole E64c inhibitor the barrier height decreases only slightly (column 5 of Table 3.2.2). The small barrier is reflected by small changes between the geometrical parameters of R and TS (entries 4 and 5 of Table 3.2.3). This shows that the ring-opening reaction is mainly influenced by substituents directly attached to the warhead, while the peptidomimetic side chains serve only to stabilize the reversible enzyme/inhibitor complex EI. The upper graph in Figure 3.2.10 shows the correlation between the ring-opening angle in the TS structure (relative to the corresponding R structure) and the reaction barrier height for all calculated systems with a carboxylated ring (oxirane and aziridine). The linear dependence is in accordance with our assumptions. The same correlation is found between the relative change in R_{C2-S} distance and the TS energy (Figure 3.2.10, bottom).

While the ring-opening reaction is less influenced by the peptidomimetic side chain its importance for the formation of the reversible enzyme-inhibitor complex becomes obvious from Figure 3.2.9 (upper part) and Table 3.2.3 (entries 4-6). Beside the interactions already discussed there is an additional strong hydrogen bond between O5 and the Gly74 residue ($R_{O5-Gly74} = 2.94 \text{ \AA}$). The hydrogen bond between N6H and the Gly198 residue is weaker but still worth mentioning (its importance for the stereospecificity will be discussed later). The heavy atom distances between the inhibitor and the hydrophobic pockets' residues are much larger (about 5 \AA), which indicates weaker interactions than expected. These latter interactions are responsible for the considerable increase in $R_{O17-NCys29}$ if OX-COO⁻ (Table 3.2.3 entries 1-3) is compared to E64c (entries 4-6). This shows that these interactions reduce the flexibility of the inhibitor, leading to improved preorganisation for the attack of the thiolate group of the Cys29 residue.

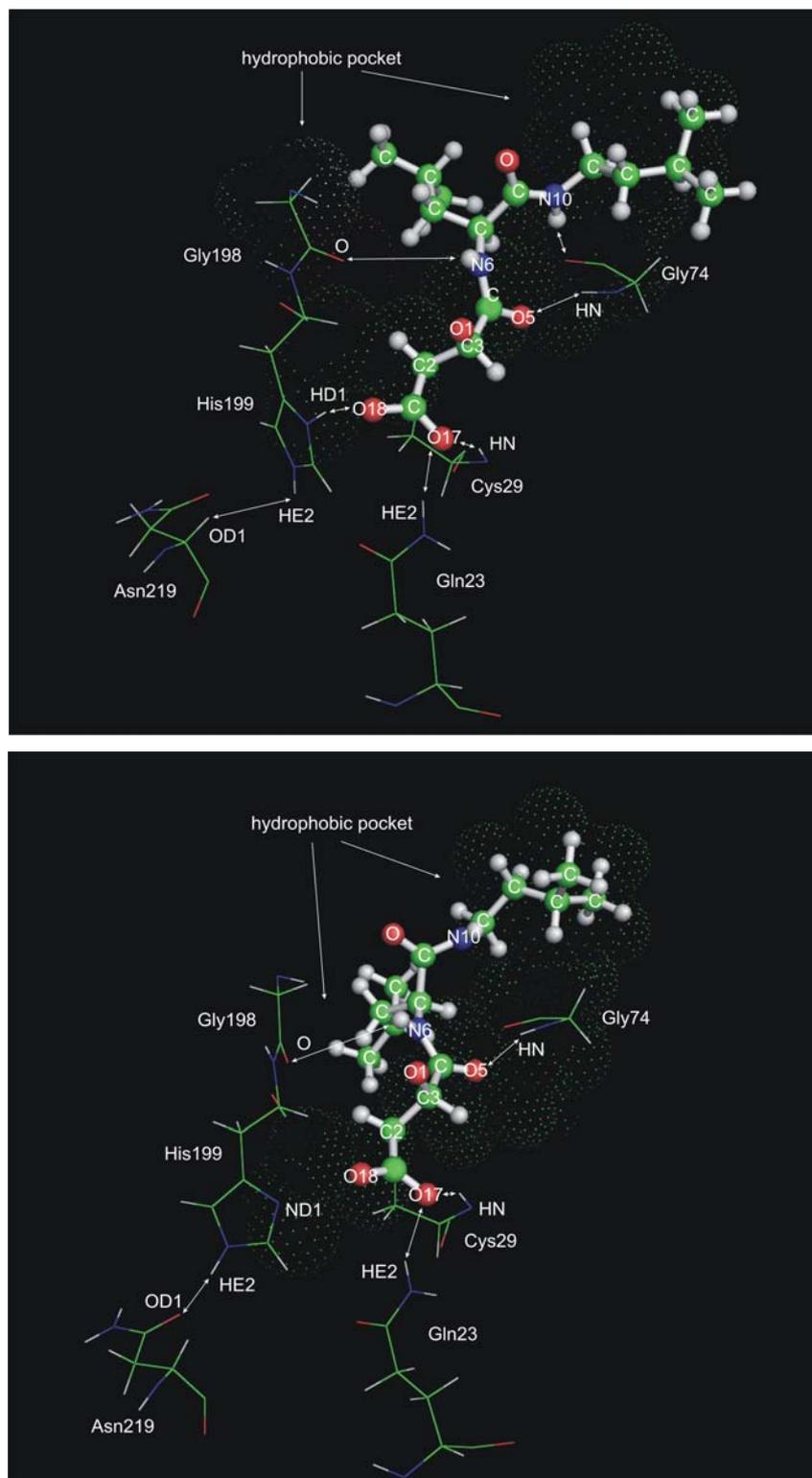


Figure 3.2.9. Reaction product for the E64c inhibitor in the presence of a protonated (top) and a non-protonated (bottom) His199 residue in the active site. The residues of the active site and the most important interactions are shown, and the locations of the hydrophobic pockets are labeled by dots. Atom notation corresponds to that of the 1ITO.pdb structure.

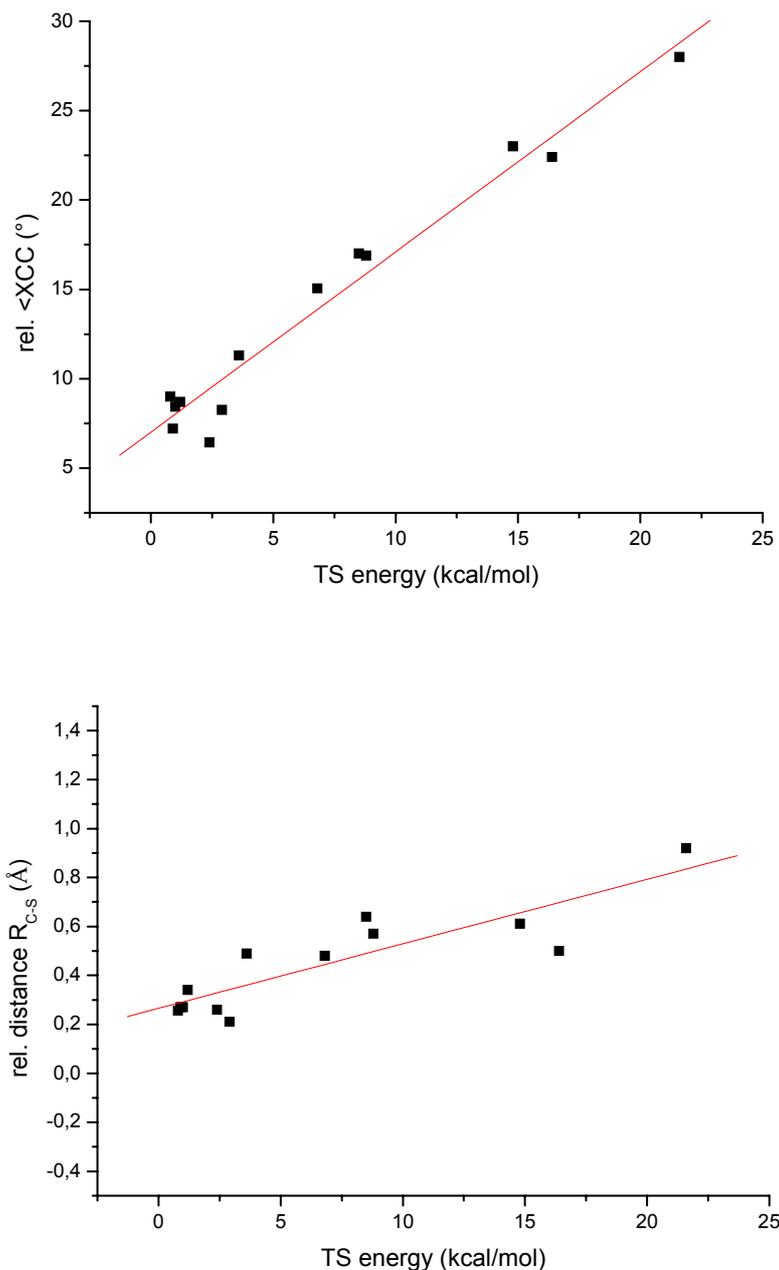


Figure 3.2.10. Graphs showing correlation between: (top) increase in ring-opening angle, $\angle XC3C2$ ($X=O, N,$ or NH) of the transition state structure relative to the reactant and the reaction barrier, and (bottom) increase in R_{C-S} distance for the TS, relative to R, and the TS energy. The data covers 13 different computed model systems (all presented in this work) with a carboxylate moiety substituted at the ring in the C2 position.

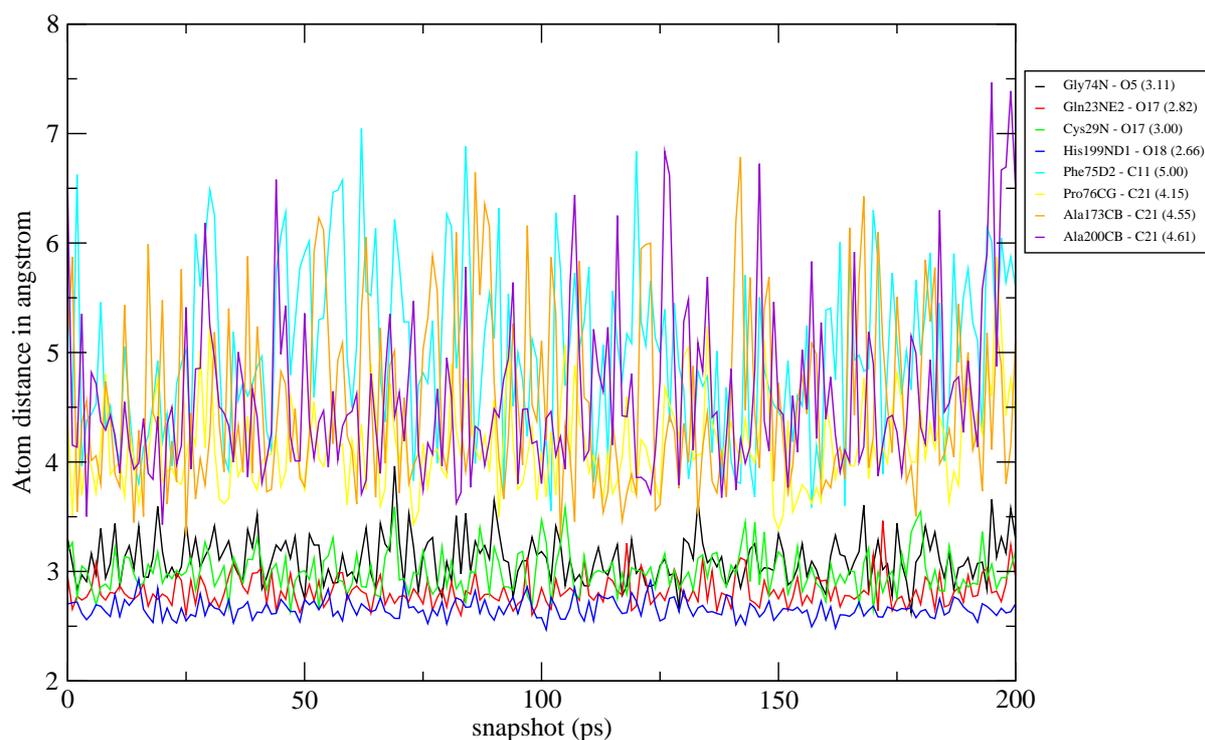


Figure 3.2.11. Time dependence of selected distances between heavy atoms of E64c and cathepsin B during the course of the molecular dynamics simulation. Larger (4-7 Å) distances: interactions with Phe75, Pro76, and Ala173 across the hydrophobic pocket. Smaller distances (2.7-3.5 Å): interaction by hydrogen bonds with Gly74, Gln23, Cys29, and His199. In the top-right a legend is given for colour-coding of the different distances plotted. The nomenclature of the atoms is the one used in the 1ITO pdb file. The numbers in parenthesis are average distances in Å.

The distance between two centres is not an unambiguous proof of the strength of their interaction. However, there are other indications of bond strength that can be obtained from MD simulations. Strong bonds mostly show small oscillations around the average distance, while distances between weakly-interacting atoms vary considerably. The variations of selected distances along the MD trajectory of the product system (PH) are plotted in Figure 3.2.11. This chart clearly supports the picture obtained from the QM/MM computations. The interactions between the inhibitor and hydrophobic pockets built out of Phe75, Pro76 and Ala200 residues (the S2 and S3 binding subsites) are characterized by distance variations of up to 3 Å. On the other hand, distances between atoms bridged by the stronger hydrogen bonds (to His199, the oxyanion hole and Gly74 in particular) vary by only about 0.5 Å, confirming the strength of the interactions.

Rich and Meara⁷⁶ measured the rate constants for papain inhibition by E64c-based inhibitors with different substituents at the C2 position (Figure 3.2.6). Their results, as well as the overall rate constant for the inhibition of papain by HO-*tEps-Ile-Pro-OH* and HO-*tEps-Ile-NEt₂*⁸⁸, are given in Table 3.2.4. The results show that replacement of the carboxylate group by an amide or ketone group does not change the kinetics of the irreversible step. Rather, the reversibly-formed complex becomes less stable, as indicated by the considerably lower k_{2nd} values but similar k_i values (Table 3.2.4). One would expect the carbonyl group of the amide or ketone group to interact with the oxyanion hole rather than the His199 residue due to steric hindrance. Therefore, replacement of the carboxylate group by substituents interacting solely with the oxyanion hole can be mimicked by simulations of the carboxylated systems in interaction with the His199 is switched off, e.g. by deprotonation of the His199 moiety. Such simulations have the advantage that besides being an appropriate model for such substituents they additionally mimic situations where His199 really is deprotonated, i.e. at high pH values.

Table 3.2.4. Kinetic constants determined for the inhibition of papain by various E64c derivatives.

Entry	Inhibitor	k_{2nd} ($M^{-1}s^{-1}$)	k_i (s^{-1})
1	HOOC- <i>tEps-Ile-Pro-OH</i> ⁸⁸	78 000	-
2	HOOC- <i>tEps-Ile-NEt₂</i> ⁸⁸	103 000	-
3	HOOC- <i>tEps-Leu-NH-(CH₂)₂-CH(CH₃)₂</i> ⁷⁶	290 000	0.71
4	HONHCO- <i>tEps-Leu-NH-(CH₂)₂-CH(CH₃)₂</i> ⁷⁶	100 000	0.67
5	NH ₂ CO- <i>tEps-Leu-NH-(CH₂)₂-CH(CH₃)₂</i> ⁷⁶	6 800	0.35
6	CH ₃ CO- <i>tEps-Leu-NH-(CH₂)₂-CH(CH₃)₂</i> ⁷⁶	2 300	0.47
7	HOCH ₂ - <i>tEps-Leu-NH-(CH₂)₂-CH(CH₃)₂</i> ⁷⁶	4.1	0.019

The corresponding results for the system with deprotonated His199 are also summarized in Tables 3.2.2 (columns 6-7) and 3.2.3 (entries 8-15). Let us concentrate on the results obtained for the whole E64c inhibitor (Table 3.2.2 column 7, Table 3.2.3 entries 11-13). All geometrical parameters indicate that deactivation of the interaction between His199 and the carboxylate is partially compensated by the interaction between the carboxylate and the backbone NH group of the Cys29 residue. For reactants and transition states the deprotonation of the His199 residue leads to a decrease of about 0.8 Å in $R_{O17-NCys29}$ (Table 3.2.3, entries 1-2 and 12-13). The differences in the product structures are much smaller. The other hydrogen bonds change little on deprotonation of His199. While a small increase is seen in $R_{O17-NE2}$, $R_{O5-Gly74}$ and $R_{N6-Gly198}$ decrease slightly. The TS becomes somewhat more reactant-like

(compare entries 2 and 13), in line with the increased exothermicity (-22 vs. -33 kcal/mol), although the geometrical differences are small. Overall, while the stability of the reversible enzyme-inhibitor complex could be expected to decrease considerably due to the switching off of a strong binding interaction, the effects of His199 deprotonation on the irreversible step seem to be largely compensated for by other interactions becoming stronger.

For the OX-COO⁻ system (Table 3.2.3, entries 8-11) the situation is somewhat more complicated, since the OX-COO⁻ inhibitor is much more flexible than the complete E64c inhibitor. Consequently, it can more easily reorient itself to avoid repulsion between the deprotonated His199 residue and O18. This leads to an additional possible reactant structure (R2, entry 8) which lies 7 kcal mol⁻¹ lower than R. Since R is obtained by optimization along the intrinsic reaction path it represents the starting point of the irreversible step, and is thus more suitable for comparison to the R structures obtained for the other systems. R2 and R are separated only by a small barrier. The geometrical arrangements of R and R2 differ in the orientation of the carboxylate group with respect to His199 and the oxyanion hole. The repulsive interaction pushes carboxylate and oxirane away from each other, reflected in the larger distance between O18 and ND1. As a result, $R_{O17-NE2}$ and $R_{O17-NCys29}$ also increase (Table 3.2.3, entries 1 and 8). This shows that for the OX-COO⁻ system at least, both distances are significantly influenced by the His199-carboxylate interaction.

However, for reaction with thiolate to take place, the OX-COO⁻ system has to move into a similar orientation to all the other reactant systems discussed. For the OX-COO⁻ system (Table 3.2.3: entries 8 and 9) $R_{O17-NCys29}$ decreases to 3.23 Å, which is still considerably larger than the value obtained for the protonated His199 (entry 1). $R_{O17-NE2}$ approaches the values found for the protonated His199 (table 3.2.3: entries 9 and 1). Interestingly, the OX-COO⁻ (entries 9-11) and the E64c systems (entries 12-13) differ considerably in the $R_{O17-NCys29}$ values while the $R_{O17-NE2}$ distances resemble each other. This indicates that the stationary points found for the OX-COO⁻ system represent local minima which perhaps do not reflect the minimum energy path of the reaction. However, due to the limited significance of this model system and limited computational power we refrain from searching for a minimum that more closely matches the E64c reactant state. It is interesting to note that the product structures of both systems are very similar.

The strong influence of the carboxylate group for the protonated His199 system results from making the reactants more transition state-like. When the His199 residue is deprotonated, the remaining interactions are sufficiently strong to induce nearly identical effects (Table 3.2.3: entries 9 and 12). Nevertheless, the overall stability of the reversible enzyme-inhibitor complex is expected to decrease on deprotonation since the deactivated His199-carboxylate interaction represented the strongest hydrogen bond to the inhibitor. This explains the combination of nearly constant k_i values and strongly decreasing k_{2nd} values for the entries 3-6 of Table 3.2.4. The inhibitors listed in entries 3 and 4 can form interactions to both His199 and the oxyanion hole, while for 5 and 6 only the latter interaction is possible. Only if the interaction with the oxyanion hole is also no longer possible is a strong decrease in k_i seen, as indicated by comparison of the OX and the E64c systems (Table 3.2.2, columns 3 and 5). This explains the sharp decrease in k_i going from entry 5 or 6 to 7 of Table 3.2.4, as for the alcohol substituent of entry 7, interaction with the oxyanion hole is denied. However, for the case of E64c, the computed barriers are so small that the measured k_i values can also be expected to subsume effects normally attributed to the formation of the reversible complex.

The results of the simulations also explain the influence of the environmental pH value on the inhibition potency of oxirane ring-based inhibitors. Studies show that inhibitors with carboxylated oxiranes as warheads decrease strongly in activity if the pH value increases above 7.⁷⁶ If the carboxylate group is replaced by a carbonyl group the activity remains constant, but is considerably smaller overall. Our computations show this loss in activity to be due to the His199-carboxylate interaction being switched off when the His199 residue is deprotonated. Additionally, they clearly indicate that the effect must be connected with the stability of the reversible enzyme-inhibitor complex, since the kinetics of the irreversible step remain nearly unaltered. For inhibitors which can only interact with the oxyanion hole deprotonation will have no impact, since no significant interaction between the His199 residue and the inhibitor exists.

In summary, our computations explain why two carbonyl groups at the α position to the carbon centers of the oxirane ring are essential for the potency of oxirane-based inhibitors. Additionally, they make clear why the replacement of one carbonyl group by a carboxylate moiety favors the formation of the reversible enzyme-inhibitor complex (constant K_i) but only slightly influences the irreversible step characterized by the reaction constant k_i . However, the computed reaction barriers are so low that the two-step model seems to be problematic.

Finally, our computations predict weak interactions between the E64c side chains and the S2 and S3 subsites. This explains why two polymorphous forms of the papain-E64c complex exist⁶ which differ only in whether the leucyl and isoamylamide groups interact with, respectively, the S2 and S3 subsites, or *vice versa*.

Aziridine-based inhibitors

The mechanisms and binding modes of aziridine-based inhibitors have been subject to considerably less investigation than their oxirane counterparts, among the consequences of which is a lack of available X-ray structures. Computations predict their inhibition mechanism to be similar to that of oxiranes, but differences arise due to the stronger basicity. As a consequence, weak proton donors protonate the warhead even before the transition state is reached, so that both kinetics and thermodynamics are influenced by the pH value of the environment. To study whether different substituent effects arise for these inhibitors as well, we performed computations for a model system in which the inhibitor is modeled by an aziridine ring with an attached carboxylate moiety (AZR-COO⁻). To investigate the strong influence of the pH-value we also modeled the system with the nitrogen centre of the aziridine ring protonated (AZRH-COO⁻). The computed relative energies are summarized in Table 3.2.5. To characterize the stationary points we present selected geometrical parameters in Table 3.2.6. We desist, however, from computing the aziridine analogues of E64c, since comparison of the OX-COO⁻ and E64c computations did not indicate any cooperative effects between the carboxylate moiety and the peptidomimetic side chain.

Table 3.2.5. Relative energies (compared to the reactants R) of the stationary points of the reaction profiles computed for the reactions of aziridine-based inhibitors. All energies are given in kcal/mol. For definition of the various simulations see text.

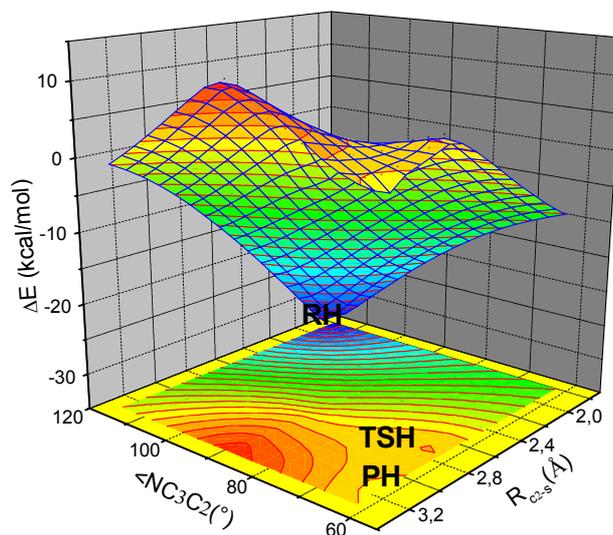
	protonated His199			deprotonated His199		solvent model ⁴⁶	
	AZR	AZR-COO ⁻	AZRH-COO ⁻	AZR-COO ⁻	AZRH-COO ⁻	AZR	AZR-COO ⁻
R2	-	-	-	0 (-7)	0(-7)	-	-
R	0	0	0	(0)	-	0	0
TS	30	15	-	21 (14)	-	28	-
P	12	-5	-	-15 (-8)	-	-	-
RH2	-	-	-	-	1 (-6)	-	-
RH	1	-	1	-	(0)	-	-
TSH	18	-	2	-	11 (5)	-	14
PH	-36	-	-33	-	-50 (-43)	-6	-26

Table 3.2.6. Selected geometrical parameters for the stationary points of the reaction profiles computed for the reactions of aziridine-based inhibitors. All distances are in Å and angles in degrees (°). For definition of the various distances see Figure 3.2.6.

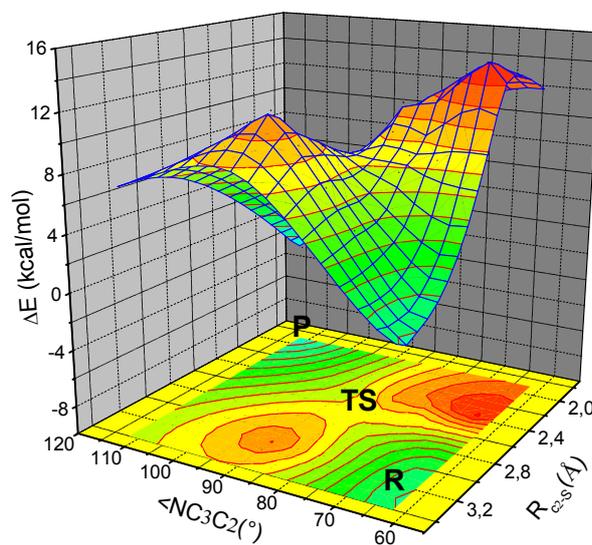
Entry		R _{C2-S}	<NC3C2	R _{O18-ND1}	R _{O17-NE2}	R _{O17-NCvs29}
QM/MM: AZR-COO⁻ /protonated His199						
1	R	3.31	62.0	1.70/2.74	2.80/3.59	4.19/5.16
2	TS	2.70	85.0	1.60/2.66	2.11/2.98	2.93/3.91
3	PH	1.93	110.2	1.63/2.65	2.12/2.95	2.18/3.16
QM/MM: AZRH-COO⁻ /protonated His199						
4	RH	3.14	63.1	1.83/2.85	1.89/2.86	2.95/3.87
5	TSH	2.80	71.8	1.79/2.81	1.86/2.84	2.74/3.66
6	PH	1.89	115.1	1.77/2.78	1.78/2.76	2.02/2.97
QM/MM: AZR-COO⁻ /deprotonated His199						
7	R2	3.33	60.3	-	1.76	2.29
8	R			not found		
9	TS	2.90	82.1	-	1.96	2.03
10	P	1.88	111.1	-	1.88	1.75
QM/MM: AZRH-COO⁻ /deprotonated His199						
11	R2			not found		
12	RH2	3.36	60.7	-	1.76	2.30
13	RH			not found		
14	TSH	2.95	68.3	-	1.83	2.05
15	PH	1.88	116.5	-	1.75	1.77

As with the corresponding oxirane system, the carboxylate substituent lowers the reaction barrier by about 15 kcal mol⁻¹ for both the deprotonated (AZR-COO⁻) and the protonated (AZRH-COO⁻) models. This leads to barriers of 15 and 1 kcal mol⁻¹ for the AZR-COO⁻ and the AZRH-COO⁻ system, respectively. The calculated potential energy surfaces for both systems are given in Figure 3.2.12. The drop in reaction barrier of 14 kcal mol⁻¹ upon protonation (AZR-COO⁻ vs. AZRH-COO⁻), along with comparison to the reaction barrier predicted for the OX-COO⁻ model (2 kcal mol⁻¹), explains why aziridine-based inhibitors are considerably less potent than oxirane ones for higher pH-values but become similar active for low pH values. The similarity between aziridine and oxirane systems is also reflected in the geometrical parameters (compare Table 3.2.6 to Table 3.2.3, entries 1-3). Protonation of the nitrogen centre of the aziridine ring leads to somewhat larger H-bond distances overall. This can be explained by a lowering of the electron density of the carboxylate moiety due to a transfer of density onto the positively charged ring. If the product is not protonated, the reaction is only slightly exothermic. However, due to the strong basicity of the emerging

negatively charged nitrogen center, protonation can be expected even at pH values around 12. For the protonated product the reaction is still less exothermic than the corresponding oxirane system (-33 vs. -47 kcal mol⁻¹), but all values are in line with the experimentally measured irreversibility of the inhibition.



AZRH-COO⁻ / protonated His199



AZR-COO⁻ / protonated His199

Figure 3.2.12. PES obtained in QM/MM calculations for the AZRH-COO⁻ system (top) and the AZR-COO⁻ system (bottom) in the presence of protonated His199 residue.

3.2.4 QM/MM calculations explaining the regioselectivity of the reaction with epoxide- and aziridine-based inhibitors

Many experimental studies have investigated the complex regioselectivity of the inhibition process of cysteine proteases by substituted oxiranes, but the effects involved are only partly understood. The regioselectivity of the alkylation step was discovered by X-ray investigations: for compounds bearing an acid at C2 (CO₂H or CO₂⁻), attack takes place only at the α -carbon (C2; see Figures 3.2.6 and 3.2.13).^{55,89,90,91} For the corresponding ester and amide derivatives, however, attack at both C2 and C3 is observed, depending on the overall binding mode (binding to S, S' or both).⁹² For cathepsin B a general rule can be drawn from the experimental findings: the nucleophilic attack of the Cys sulphur atom occurs at the oxirane carbon atom that possesses the P'-substituent. Reactions with the carbon atom that possesses the P-substituent are not observed.⁹²

However, these results for the ring-opening reaction by the Cys residue of the proteases do not match the ring-opening reactions of α,β -epoxycarbonyl compounds with methyl thiolate in solution, which represent a minimal model reaction for the alkylation step of the inhibition.⁸⁶ For these reactions Bihovsky found that the rate constant for epoxides with an ester attached to the ring carbon (CO₂CH₃, $k = 0.44$) is much higher than the corresponding value for epoxides containing a negatively charged carboxylate (CO₂⁻, $k = 0.01$). Overall, the trend COCH₃ > CO₂CH₃ > CONH₂ > CO₂⁻ was determined for the rate constant k . These investigations also revealed a different regioselectivity. For ester substituents the rate constants for α - and β -attack are quite similar ($k_{\alpha} = 0.25$ vs. $k_{\beta} = 0.19$), whereas for CO₂⁻ they differ by a factor of 2 ($k_{\alpha} = 0.0036$ vs. $k_{\beta} = 0.0064$), with the β -attack preferred.

The quantum chemical computations of Helten *et al.*^{44,46} nicely explain the experimental solvent effects on reaction rates and regioselectivity of ring-opening reactions of substituted epoxides by methyl thiolate. Their analysis of the various effects underlines the greater importance of electronic effects in comparison to steric effects. Assuming that the alkylation step of the inhibition reaction is determined by the same factors as the model ring-opening reaction, the differences in the regioselectivities of the reaction with cysteine protease compared to the reaction with methyl thiolate in solution must be attributed to the formation of the reversible enzyme inhibitor complex EI. Indeed, the overall trends in the inactivation of

cysteine proteases by oxirane carboxylic acids are generally explained by a favourable ionic interaction between the negatively charged carboxylate and the positively charged imidazolium ion of His159 (papain numbering) of the active site.⁹³ Available X-ray data support this assumption.^{55,89-91}

This explanation is also in accordance with the experimental data described above. However, the X-ray data only reflect the situation at the end of the inhibition process, i.e. they only provide information about the effects governing the reaction if the orientation and geometries of reactants and products are very similar. Direct information on the effects influencing the inhibition process of cysteine proteases by epoxide- or aziridine-containing inhibitors is only obtainable from time-resolved measurements or by way of computations which account for the protein environment. According to quantum mechanical calculations,^{44,46} aziridines and epoxides do not differ in their behaviour concerning the regioselectivity of nucleophilic attack.

To shed some light on this complicated issue the inhibition process was simulated on a QM/MM level of theory for inhibitors with various substitution patterns. The computations are done by using the same systems as in the previous sub-chapter. The notation used in the following text is given in Figure 3.2.13. Figure 3.2.13 is similar to Figure 3.2.6 but also shows how α - and β -attack of the substituted inhibitor ring are defined. The negatively charged sulphur atom of the Cys29 residue could attack either at the C_α (calculations on which situation will be specified as E64c- α , OX-COO⁻- α , AZR-COO⁻- α or AZRH-COO⁻- α) or at the C_β (calculations specified as E64c- β , OX-COO⁻- β , e.g. AZR-COO⁻- β or AZRH-COO⁻- β) atom.

Table 3.2.7. In addition to the reaction coordinates, Table 3.2.7 also contains (in brackets) the distances and angles for the other carbon centre. So for OX-COO⁻- α the table lists $R_{C_{\alpha-S}}$ ($R_{C_{\beta-S}}$) and $\angle OC_{\alpha}C_{\beta}$ ($\angle OC_{\beta}C_{\alpha}$) and *vice versa*.

Table 3.2.7. Selected geometrical parameters for the stationary points of the reaction profile of the alkylation of cysteine protease by an inhibitor, represented by an oxirane ring with a carboxylate substituent, and with an ammonium ion acting as a proton donor. Two possible geometrical attacks of the Cys29 S atom are included. Values in parentheses are connected with the other C-centre, e.g. the C _{β} centre in case of α -attack and *vice versa*. All distances are given in Å, angles in degrees (°), and energies in kcal/mol.

	rel. E	$R_{C_{\alpha/\beta-S}}$	$\angle OC_{\beta/\alpha}C_{\alpha/\beta}$	R_{His199}	R_{Gln23}	R_{Cys29}
OX-COO⁻-α						
R	0	3.07 (3.44)	64.7 (57.9)	1.66	1.94	2.31
TS	2	2.81	71.1	1.62	1.92	2.20
PH	-47	1.92	112.2	1.58	1.88	1.99
OX-COO⁻-β						
R	0	3.33 (3.28)	59.2 (62.5)	1.67	1.95	2.51
TS	9	2.76	76.1	1.66	1.90	2.52
PH	-48	1.88	108.3	1.59	1.87	2.21

Table 3.2.7 shows that the α -attack is kinetically preferred. In the case of the OX-COO⁻- α mechanism the reaction barrier is 7 kcal/mol lower than for OX-COO⁻- β . If the reaction paths for α - and β -attack are compared (Figure 3.2.14), it is obvious that for the α -attack both of the reaction coordinates change in a concerted fashion over the course of the reaction path, going from $R_{C_{\alpha-S}} = 3.07$ Å and $\angle OC_{\beta}C_{\alpha} = 64.7^\circ$ for R, to $R_{C_{\alpha-S}} = 2.81$ Å and $\angle OC_{\beta}C_{\alpha} = 71.1^\circ$ for TS, to $R_{C_{\alpha-S}} = 1.92$ Å and $\angle OC_{\beta}C_{\alpha} = 112.2^\circ$ for P). At the beginning of the β -attack reaction, going from R to TS, the $R_{C_{\beta-S}}$ distance plays a more important role and shows a more significant change (3.33 Å to 2.76 Å) than the angle (59.2° to 76.1°). Once the TS is reached the angle starts to change more drastically, so that both reaction coordinates change simultaneously. For the reactant R of the α -attack a value of $\angle OC_{\beta}C_{\alpha} = 64.7^\circ$ indicates that the ring has already started to open. For reactions in gas phase or solution an angle of 60° has been computed for the reactants.⁴⁴ The reactant of the OX-COO⁻- β reaction path has $\angle OC_{\beta}C_{\alpha} = 59.2^\circ$ and $\angle OC_{\alpha}C_{\beta} = 62.5^\circ$, while $R_{C_{\beta-S}} = 3.33$ Å and $R_{C_{\alpha-S}} = 3.28$ Å. Although this structure was obtained by optimization along the intrinsic reaction path from the OX-COO⁻- β transition state, its geometrical parameters still favor the α -attack mechanism.

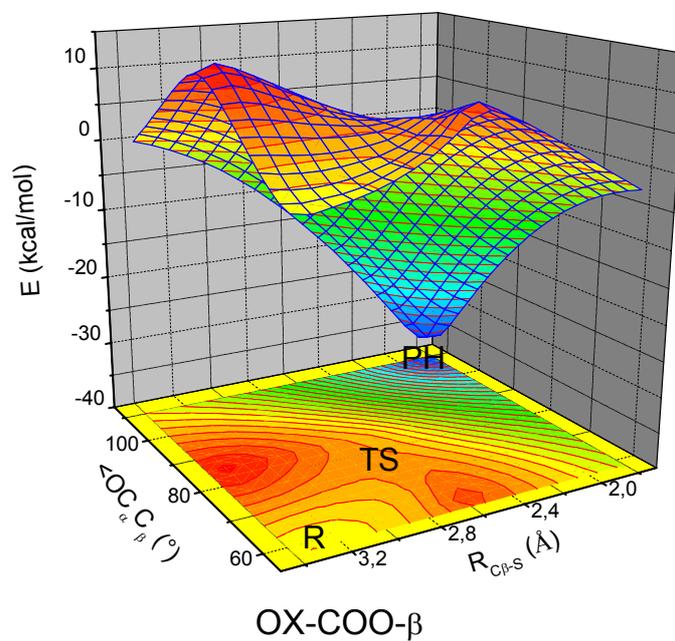
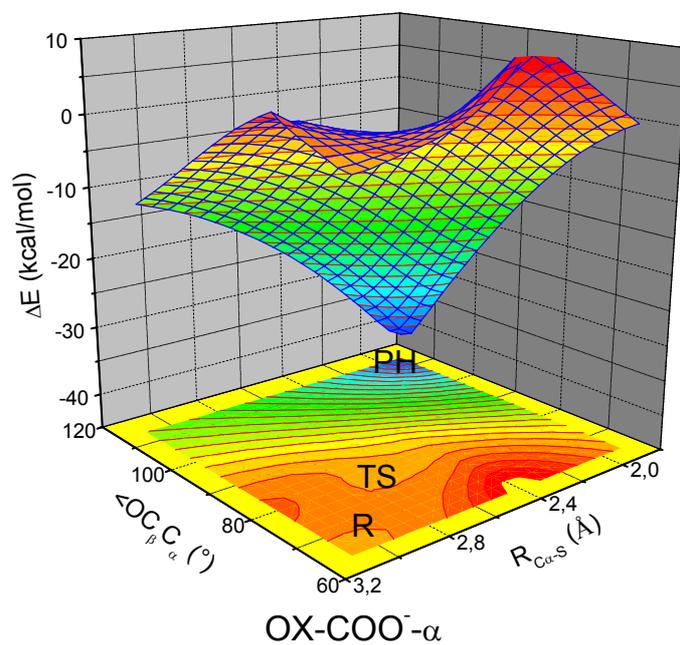


Figure 3.2.14. PES obtained in QM/MM calculations for the OX-COO⁻- α system (top) and the OX-COO⁻- β system (bottom) for the substitution in which the C _{α} atom of the reactant has (*S*)-configuration.

The preference of the α -attack results from the interaction of the carboxylate substituent at the warhead with the oxyanion hole and the His199 residue. This means that it is dictated by the structure of the non-covalent enzyme-inhibitor complex which is formed in the first step of the inhibition reaction. In order to perform the β -attack, the geometrical arrangement within the oxyanion hole has to be unfavorably changed. This variation is for example reflected in the change of the R_{Cys29} distance (Figure 3.2.13), which represents the bond length of the H-bond between the carboxylate and the backbone amide hydrogen (Table 3.2.7). In the case of α -attack the distances are 0.2 Å (in the TS, 0.3 Å) shorter than for the OX-COO⁻- β reaction. If one compares how this parameter changes over the reaction path, one notices that for the OX-COO⁻- β reaction R_{Cys29} hardly changes on going from R to TS and then decreases by 0.32 Å on going to P. For OX-COO⁻- α , however, the R_{Cys29} distance decreases steadily along the reaction path. The R_{Gln23} and R_{His199} distances (Figure 3.2.13), on the other hand, behave quite similarly for the α - and the β -reaction, so the interactions these represent cannot be contributing to the difference in activation energy. The insensitivity of these interactions may result from these two residues interacting with the carboxylate moiety via their side chains, which are flexible and can easily follow the movement of the inhibitor. With Cys29, however, the H-bond (R_{Cys29}) is to the backbone, which is stiff and does not allow such an adjustment. The position of the NH unit of the Cys29 residue is also determined by the position of its sulphur atom, which needs an optimal arrangement for the nucleophilic attack.

Calculations on the small model system yield a reaction barrier difference of 7 kcal/mol in favour of the α -attack, which mainly results from to the interaction of the carboxylate substituent with the residues of the oxyanion hole. However, although our small model reproduces the regioselectivity, it is too simple to include all the effects that would contribute (in whichever direction) to the regiospecificity of the complete inhibitor. Therefore, in order to study the remaining effects not covered by the simple model, we additionally performed calculations at the same level of theory but using the whole E64c inhibitor. The QM region in these calculations is identical to the previous computations (Figure 3.2.13). The C $_{\beta}$ substituent of the full E64c molecule, a long peptidomimetic chain, could be treated within the MM part since it is not directly involved in bond-breaking or -formation processes. The ammonium ion which served as a proton donor was omitted since protonation takes place after the TS, and hence does not influence the kinetics. As a consequence the obtained product structures are not protonated, resulting in a less exothermic character of the reactions. Since irreversible

bonding is also predicted for unprotonated products, all the effects which influence the regioselectivity should be nicely captured.

Product structures of α - and β -attacks at the E64c inhibitor are given in Figure 3.2.15. The corresponding potential energy surfaces are shown Figure 3.2.16. The relative energies of stationary points are tabulated, together with the geometrical parameters in Table 3.2.8. In the reactants both C atoms of the inhibitor ring are (*S*)-configured. To distinguish between both model systems (warhead-only vs. the complete inhibitor) the two inhibition pathways involving the complete E64c inhibitor will be denoted E64c- α and E64c- β .

With the complete inhibitor in the calculation, the difference between the α and β activation energies is 16 kcal/mol. This shows an increased regioselectivity with respect to the model system discussed before which only took the warhead into account (OX-COO⁻- α vs. OX-COO⁻- β), for which a reaction barrier difference of only 7 kcal/mol was computed. This increase results from the long peptidic side chain of the E64c interacting with the enzyme on both sides of the epoxide ring via hydrogen bonds and hydrophobic interactions. The Gly74 residue especially forms a strong H-bond with the carbonyl group of the inhibitor (Figure 3.2.15). This additional interaction makes the β -attack geometrically even more unfavorable. The interactions between the carboxylate substituent and His199 and the oxyanion hole remain nearly unchanged (as happened in the smaller model). Moving from the R to the TS structure along the E64c- β reaction path reduces $R_{C\beta-S}$ by 0.5 Å, and $\angle OC_\alpha C_\beta$ varies by over 22°, from 57.2° to 79.6°. For the E64c- α path, the corresponding changes are only around 0.1 Å and 8°. Furthermore, just like in the smaller model system OX-COO⁻- β , in E64c- β the geometrical parameters of the reactant favor α -alkylation (see values in parenthesis in Table 3.2.10). Both results for OX-COO⁻ and for E64c accord nicely with the general linear dependence of the reaction barrier height on the change in the internal coordinates (Figure 3.2.10).

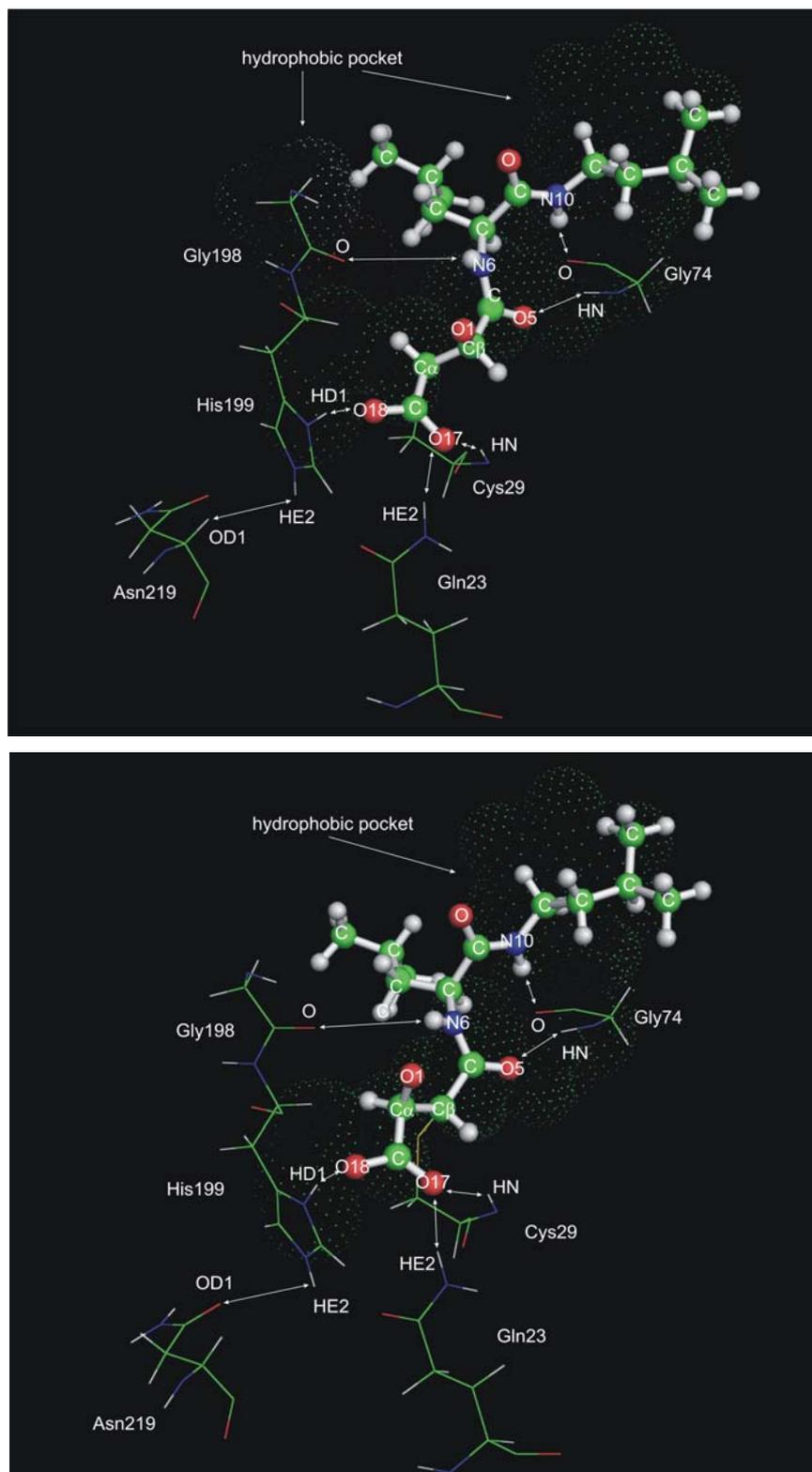


Figure 3.2.15. α - (top) and β - (bottom) product structures of the E64c inhibitor. The residues of the active site and the most important interactions are shown. Dots represent the residues of the hydrophobic pocket.

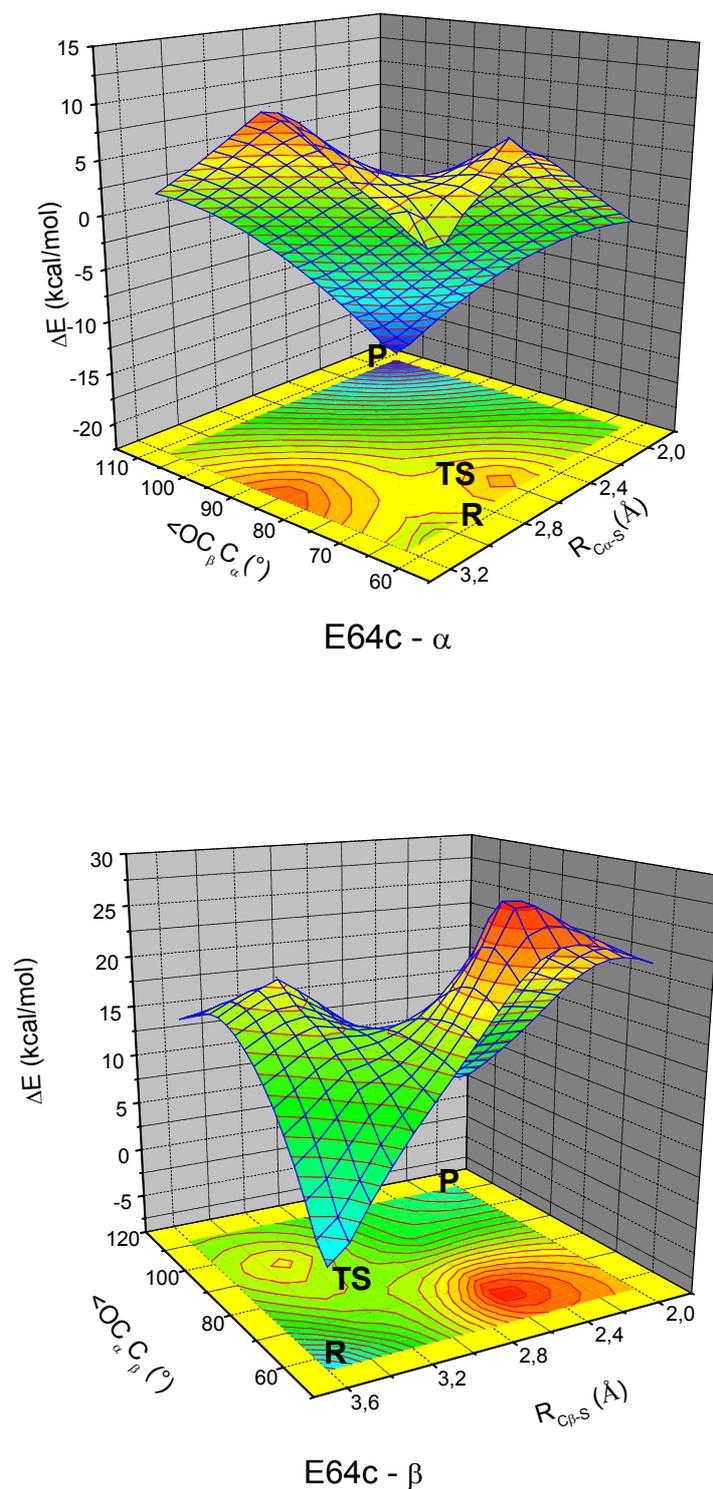


Figure 3.2.16. PES obtained in QM/MM calculations with the E64c inhibitor. Top: α -attack; bottom: β -attack.

Table 3.2.8. Selected geometrical parameters for the stationary points of the reaction profiles for the alkylation of the cysteine protease by E64c. Values in parentheses relate to the non-reactive C-centre, i.e. the C_β centre in case of α attack and *vice versa*. All distances are given in Å, angles in degrees (°), and energies in kcal/mol.

	rel. E	R _{Cα/β-S}	<OC _{β/α} C _{α/β}	R _{His199}	R _{Gln23}	R _{Cys29}	R _{Gly74}
E64c - α							
R	0	3.07 (3.49)	64.0 (58.2)	1.65	2.21	2.97	2.00
TS	1	2.80	72.4	1.64	2.15	2.67	1.99
P	-22	1.91	111.0	1.67	1.91	1.96	2.01
E64c - β							
R	0	3.56 (3.14)	57.2 (66.0)	1.66	2.38	3.40	2.10
TS	16	3.06	79.6	1.66	2.28	3.30	2.08
P	-2	2.01	107.4	1.62	1.96	2.64	2.00

It is interesting to note that the additional interactions disfavor the β-attack by a further 8 kcal/mol, while the reaction barrier of the E64c-α reaction is practically identical to that of the smaller OX-COO⁻-α model (2 vs. 1 kcal/mol, respectively. Tables 3.2.7 and 3.2.8). Such a small barrier calls into question the generally accepted two step model for the inhibition. The computed barriers are so low that the irreversible part of the inhibition cannot be looked upon as being separated from the reversible one. Instead our computations indicate that the reversible binding step does not reach equilibrium since the subsequent irreversible step is too fast. This is also shown by the angles of the reactant structures (64-65°, Tables 3.2.7 and 3.2.8), indicating that the ring opening has already started.

The MD simulations performed on the reactant structure of E64c provide further confirmation that α-attack is preferred as a result of the geometrical arrangement of the inhibitor within the active site. In Figure 3.2.17 the change of the R_{Cα-S} and R_{Cβ-S} with time is plotted. It is obvious that the C_α atom of the reactant lies closer to the attacking sulphur (average distance = 3.62 Å) than the C_β atom (average distance = 4.04 Å). The average R_{Cα-S} and R_{Cβ-S} distances obtained in the MD runs are around 0.5 Å larger than the corresponding ones obtained in QM/MM calculations (3.07 Å and 3.56 Å in Table 3.2.10). The reason for this is that in the MD simulations the S atom of Cys29 was given a charge of -0.07. Such a near-neutral S atom cannot show the correct electrostatic attraction towards the ring carbons. If the Cys29 sulphur were charged, however, this it would interfere with the non-covalent interactions between E64c and the enzyme, since its covalent binding to the ring is not possible within the force field approximation. Despite this overestimate, the difference between R_{Cα-S} and R_{Cβ-S} stays

the same, around 0.4 Å, proving that the MD simulations describe the interactions in the non-covalent E64c/enzyme complex properly.

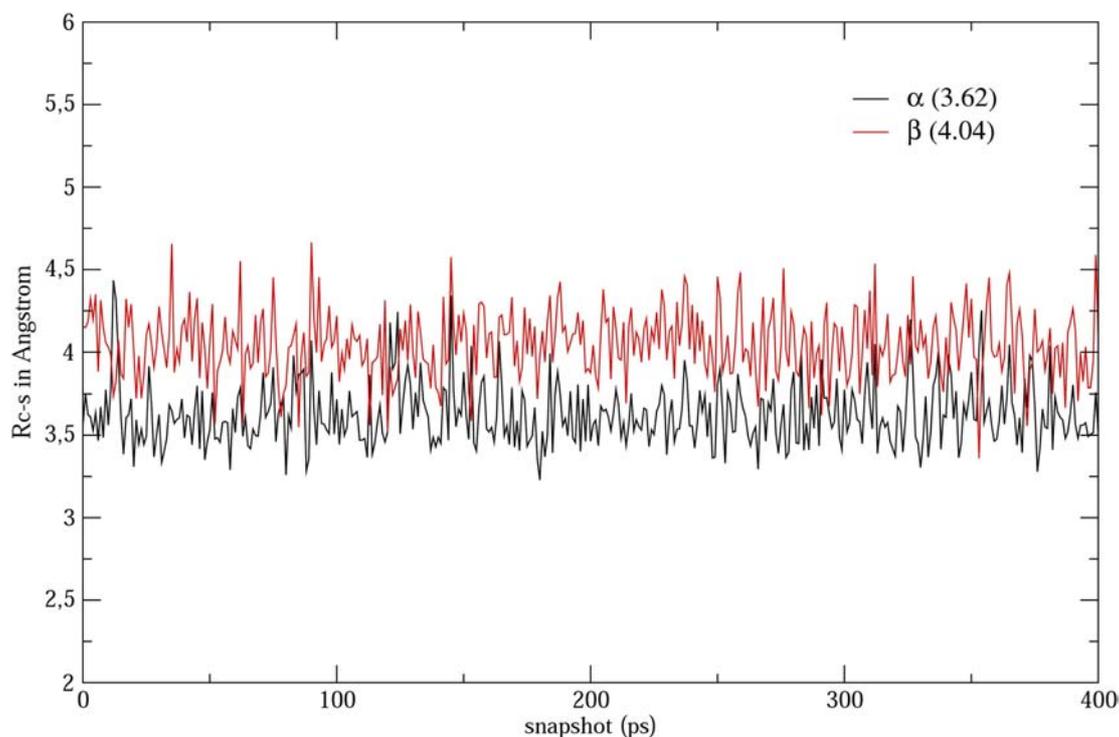


Figure 3.2.17. Time dependence of R_{C-S} distances between C_{α} and C_{β} atoms of E64c reactant structure and sulphur of cathepsin B obtained in molecular dynamics calculations. In parentheses: average distance.

To gain more insight into the strength of the various non-covalent interactions, MD runs were also performed for the product structure of E64c. For the weaker and consequently less important interactions, the MD should show large variations in the bond length, while only small variations are expected for stronger bonding situations. In Figure 3.2.11 (previous subchapter), the time dependence of distances between the heavy atoms of the E64c and the enzyme corresponding to the non-covalent interactions of the S1 and S2 binding pockets are shown. One can clearly differentiate two groups of interactions from the differences in bond distances. The distances in the upper part of the diagram are somewhat longer (4-7 Å), and fluctuate significantly during the simulation. The other group is has distances of around 3 Å and considerably smaller variations (2.7-3.5 Å). The first, longer set of enzyme-inhibitor contacts represent interactions between the inhibitor side chain and the hydrophobic pocket

formed by the side chains of residues Phe75, Pro76, Ala173, Gly198, and Ala200. This part of the inhibitor side chain is relatively far away from the Cys29 residue where the irreversible bonding takes place. Additionally, the bonds are very flexible and should consequently be less important for the regioselectivity of the alkylation. However, these interactions are crucial for the recognition of the active site by the inhibitor, and therefore very important also for the development of enzyme-specific agents.

The second type of interactions is hydrogen bonds between the E64c moiety and the backbone of Cys29, Gly74, and Gly198, as well as side chains of Gln23 and His199 (see Figure 3.2.15). The resulting H-bond network is situated close to both sides of the active site. It brings the inhibitor into position for the attack by the nucleophile and consequently determines the regioselectivity of the inhibition (see also Figure 3.2.17). The computed reaction paths show that the geometrical arrangement organized by this network favors an attack at the C atom of the inhibitor in the position nearest to the oxyanion hole (α position). A very important part of the hydrogen bond network is formed by the backbone NH units of the Cys29 and Gly74 residues. The RMSD (root mean square displacement) values of the N atoms of these two residues involved in the H-bonds with the inhibitor are 0.373 Å and 0.435 Å, respectively. The RMSDs of the corresponding acceptor atoms, the oxygen atoms O17 and O5 (Figure 3.2.12), are also small (0.438 Å, and 0.557 Å, respectively). Comparison of these values with the RMSD of the rest of the inhibitor (1.707 Å) emphasizes the strength of the corresponding H-bonds. As the O17 and O5 centers of the inhibitor are situated directly to the left and right of the inhibitor warhead, these two H-bonds are obviously very important for fixing the warhead in place (overall warhead RMSD = 0.696 Å).

Another important hydrogen bond interaction is the one between the side chain His199 ND1 centre and the O18 atom of E64c (for notation see Figure 3.2.12). The RMSD values of these atoms are 0.464 Å for ND1 and 0.444 Å for O18. The imidazolium ring of the His199 (average RMSD = 0.506 Å) has a rather stable position within the active site due to a hydrogen bond to the Asn219 residue (RMSD = 0.572 Å). This shows that the ND1-O18 hydrogen bond represents another important anchor for the carboxylate group of the inhibitor. However, in spite of what one might have expected, the presence of this H-bond is not a precondition of the regioselectivity. This had already been indicated by experiments which found the same regioselectivity for inhibitors which cannot interact directly with the His199 residue as for those which can.^{94,95,96} Also, the results obtained in MD simulations of the

cathepsin B/E64c enzyme-inhibitor complex in which the His199 moiety was not protonated at the ND1 position (Table 3.2.3) show no significant changes in the average value of the R_{Gln23} , R_{Cys29} , and R_{Gly74} distances (the details are described in the previous subchapter). This proves that these interactions are strong enough to compensate for the missing His199-O18 hydrogen bond.

Regiospecificity of carboxylated aziridines

In the previous chapter it has been shown that the inhibition reaction of the aziridine-based inhibitors is similarly affected by the substitution in the C2 position of the ring as the reaction of the oxirane-based ones. To see if the similarity also holds for the regiospecificity, further calculations were done on carboxylated aziridines. Again, both deprotonated (AZR-COO⁻) and protonated (AZRH-COO⁻) inhibitor moieties were computed, in order to take into account the influence of the pH value of the surroundings. For both systems potential energy surfaces were calculated for both α - and β -attack pathways. The computed relative energies and selected geometrical parameters are collected in Table 3.2.9. The calculated potential energy surfaces for all systems are shown in Figures 3.2.18-19.

As for the corresponding oxirane system, in the case of deprotonated (AZR-COO⁻) inhibitor the reaction barrier for β -attack lies 7 kcal mol⁻¹ higher than the one for α -attack. For the AZRH-COO⁻ inhibitor, however, this difference is only 3 kcal/mol. The higher barrier for the β -attack for AZR-COO⁻ inhibitor could result from the endothermicity of the non-protonated product (10 kcal/mol), according to the Hammond postulate. This would also explain why the protonation of the ring leads to an energy drop of 14 kcal/mol for α -attack and 18 kcal/mol for β -attack. The endothermicity of the non-protonated product of β -attack could stem from less stabilization of the negatively charged NH moiety of the opened ring. The NH moiety of the α -attack product is stabilized by four H-bonds to surrounding water molecules and an additional H-bond to the backbone carbonyl group of Gly198 residue. However, due to its position within the active site, the product of the β -attack is only stabilized by three hydrogen bonds to nearby water molecules. This effect is thus unrealistic and results from the lack of a proton source in the QM region – in reality the NH group would be protonated by water straight away, rather than forming hydrogen bonds.

Table 3.2.9. Selected geometrical parameters for the stationary points of the reaction profile for the alkylation of the cysteine protease by an inhibitor represented by a protonated aziridine ring with a carboxylate substituent. Results from α - and β -attack pathways are compared. Values in parentheses refer to the nonreacting C-centre, i.e. the C_β centre in case of α attack and *vice versa*. All distances are given in Å, angles in degrees ($^\circ$), and energies in kcal/mol.

	rel. E	$R_{C\alpha/\beta-S}$	$\angle NC_{\beta/\alpha}C_{\alpha/\beta}$	R_{His199}	R_{Gln23}	R_{Cys29}
AZRH-COO⁻ -α						
RH	0	3.14 (3.34)	63.1 (60.0)	1.83	1.89	2.95
TSH	1	2.80	71.8	1.79	1.86	2.74
PH	-34	1.89	115.1	1.77	1.78	2.02
AZRH-COO⁻ -β						
RH	0	3.21 (3.34)	60.7 (62.0)	1.85	1.90	3.30
TSH	4	2.72	72.0	1.88	1.86	3.19
PH	-34	1.85	109.0	1.82	1.79	2.52
AZR-COO⁻ -α						
R	0	3.31 (3.57)	62.0 (60.2)	1.70	2.80	4.19
TS	15	2.70	85.0	1.60	2.11	2.93
P	-5	1.93	110.2	1.63	2.12	2.18
AZR-COO⁻ -β						
R	0	3.47 (3.33)	60.2 (62.0)	1.70	3.01	4.41
TS	22	2.55	88.2	1.64	2.54	3.91
P	10	1.93	109.4	1.57	2.11	3.29

For both the AZRH-COO⁻ and AZR-COO⁻ inhibitors, the reactant geometry for the α -attack is more favorable. The $\angle NC3C2$ angle in this case is larger (63.1° for AZRH-COO⁻ and 62.0° for AZR-COO⁻) than the corresponding β -attack values (60.7° and 60.2° , respectively). This data fits the angle–transition state energy correlation depicted in Figure 3.2.10, showing the same dependence as the corresponding oxirane systems. The same applies for the R_{C-S} distances.

Let us concentrate on the protonated aziridine inhibitor. If the results in Table 3.2.9 are compared with the corresponding ones for the OX-COO⁻ system one sees that while the numbers connected to α -attack are similar, β -attack seems more favorable in the case of aziridine-based inhibitors. (Please note that the α -systems for AZR-COO⁻, AZRH-COO⁻, and OX-COO⁻ are the same as AZR-COO⁻, AZRH-COO⁻, and OX-COO⁻ in the previous chapter.) Geometry parameters for both α - and β -attack at AZRH-COO⁻ show weaker interaction between the carboxylate moiety and the Cys29 and His199 residue. As mentioned in the previous chapter, this is probably due to less negative charge being present on the carboxyl

group because of protonation making the aziridine ring N more electronegative to compensate for its positive charge. The electron-withdrawing effect of the protonated nitrogen also increases the electrophilicity of the C2 and C3 atoms of the ring (C_α and C_β respectively, in Figure 3.2.12). The reactant structure of the β -attack makes clear that the β -attack pathway is more favorable than for oxirane: for the protonated aziridine-based inhibitor the $R_{C\beta-S}$ distance is 3.21 Å and $R_{C\alpha-S} = 3.34$ Å, whereas for the oxirane-based inhibitor the same distances are 3.33 Å and 3.28 Å, respectively. The corresponding angles are 60.7° and 62.0° for AZRH-COO $^-$ - β , and 59.2° and 62.5° for OX-COO $^-$ - β . The geometrical arrangements show that the reactant for β -attack in the AZRH-COO $^-$ may not be favorable as the one for the α -attack, but the difference is much less if compared to the OX-COO $^-$ inhibitor. For the same reason $\angle NC_\alpha C_\beta$ changes by only around 11° on going from R to TS, while $\angle OC_\alpha C_\beta$ undergoes a change of around 17° . As a result the reaction barrier for the β -attack of the AZRH-COO $^-$ inhibitor is 5 kcal/mol lower than the corresponding value for the OX-COO $^-$ inhibitor.

When the aziridine ring is not protonated, the picture changes. The behavior of this inhibitor is more similar to the OX-COO $^-$ systems – for both the α - and β -attack pathways, the C_α atom in the reactant looks like the preferred reaction site. However, the electron-donor effect of the non-protonated nitrogen decreases the electrophilicity of the ring carbons. As a result, the reactant structures in both α - and β -attack have large $R_{C\alpha-S}$ and $R_{C\beta-S}$ distances (3.31 Å and 3.47 Å, respectively). Due to these electron-donation effects, the carboxylate substituent interacts more strongly with the His199 residue ($R_{His199} = 1.70$ Å), but the interaction between the substituent and the residues of the oxyanion hole ($R_{Gln23} = 2.80$ Å and $R_{Cys29} = 4.19$ Å for α -attack, and $R_{Gln23} = 3.01$ Å and $R_{Cys29} = 4.41$ Å for β -attack) gets weaker.

Since X-ray structures of aziridine-based inhibitor/enzyme complexes do not exist, and one cannot be sure about the regioselectivity of such inhibition reactions, we desist from making further conclusions due to the possible lack of accuracy in our model. For now, measurements have only shown that the aziridine-based analogues of E64c with (*R,R*)-configuration of the warhead are better inhibitors than the (*S,S*)-configured ones, in contrast to the behavior of E64c. According to our calculations the reason for this could be much weaker interaction of the aziridine ring carboxylate substituent with the oxyanion hole in comparison to oxirane-based inhibitors. This weaker interaction could allow slightly different positioning of the

inhibitor within the active site and consequently a different stereospecificity of the inhibition reaction.

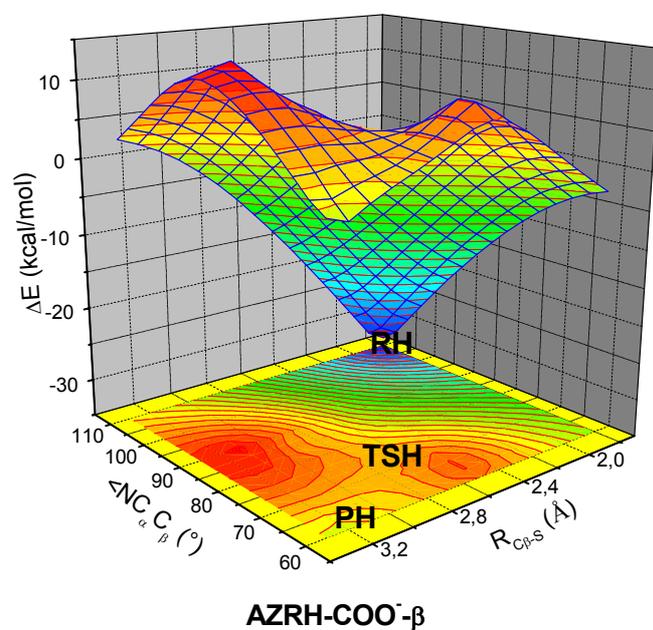
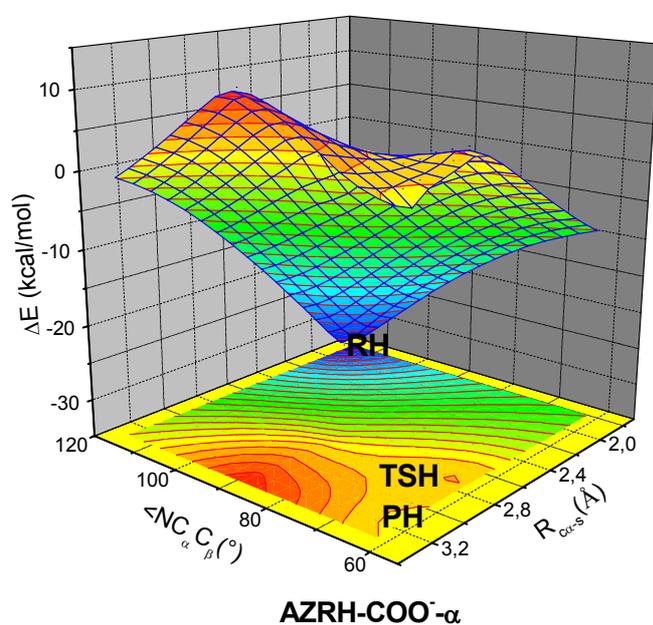


Figure 3.2.18. PES obtained in QM/MM calculations for the AZRH-COO⁻- α system (top) and the AZRH-COO⁻- β system (bottom).

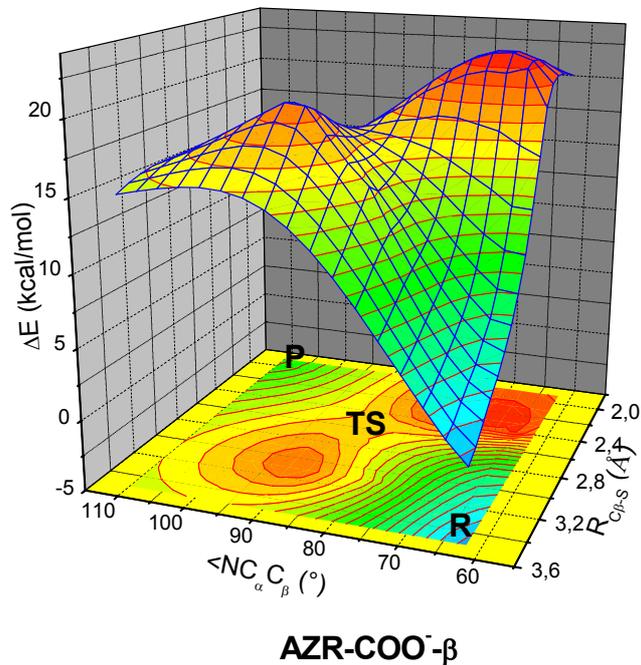
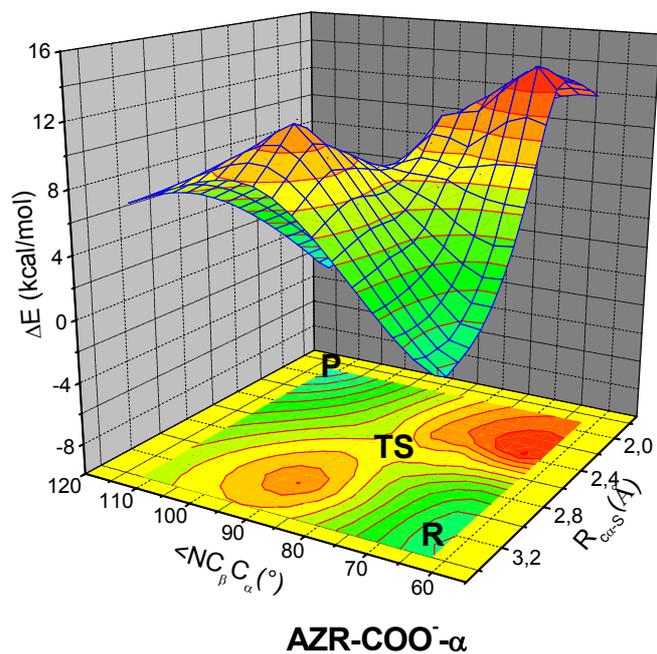


Figure 3.2.19. PES obtained in QM/MM calculations for the AZR-COO⁻- α system (top) and the AZR-COO⁻- β system (bottom).

3.2.5 Investigations into the stereospecificity of the cysteine protease inhibition reaction with epoxysuccinyl derivatives

It was shown by the earliest studies into epoxysuccinyl-based inhibitors that a *trans*-configured epoxide ring is essential for inhibition and that *cis*-configuration abolishes inhibition potency.⁹⁷ Thus, all further studies concentrated on the two *trans*-isomers with (*R,R*)- or (*S,S*)-configuration at the ring carbons. For derivatives with a free acid at C2 (Figure 3.2.9) the (*S,S*)-isomers were identified as eutomers (for summary see refs. 98 and 99). Ongoing studies^{98,99} with derivatives bearing esters or amides at both ring carbons showed that cathepsin-B selective inhibitors which address the primed substrate binding sites (e.g. CA030) are more active with an (*R,R*)-configured epoxide ring, whereas for inhibitors spanning both primed and non-primed binding sites the (*S,S*)-configuration leads to higher inhibition potency. Thus the binding modes, i.e. the interactions with the protein environment, strongly determine the stereoselectivity: inhibitor-enzyme interactions at the S-subsites are favoured by the (*S,S*) and reverse interactions with the S'-subsites generally by the (*R,R*) configurations of the epoxysuccinyl moiety. The reason for this behavior is not known.

In order to inspect the stereoselectivity of the inhibition of cysteine proteases by epoxide-based inhibitors the system depicted in Figure 3.2.6 was used. The QM part of the QM/MM calculations consisted of the carboxylated epoxide ring of the E64c inhibitor (X=O) and the active site residues. The rest of the enzyme and the side chain of the E64c moiety were treated by the force field and were connected to the QM part using link atoms.

The oxirane ring of the E64c has two stereocenters and therefore four possible diastereomers: two *trans*-configurations, corresponding to (*S*) or (*R*) absolute configuration on both C2 and C3 atoms (these systems are denoted E64c-(*S,S*) and E64c-(*R,R*), respectively, see Figure 3.2.6), and two *cis*-configurations, corresponding to (*2S,3R*) and (*2R,3S*) absolute configurations on stereo centers, where 2 and 3 stand for the C2 and C3 atoms, respectively. Since the *cis*-conformers are very inactive and there are almost no experimental results concerning them in the literature, the calculations were performed only for the (*2R,3S*) absolute configuration for comparison to (*R,R*) and (*S,S*). Calculations on this system are denoted by E64c-(*R,S*). All three computed diastereomers of the E64c inhibitor are shown in Figure 3.2.20.

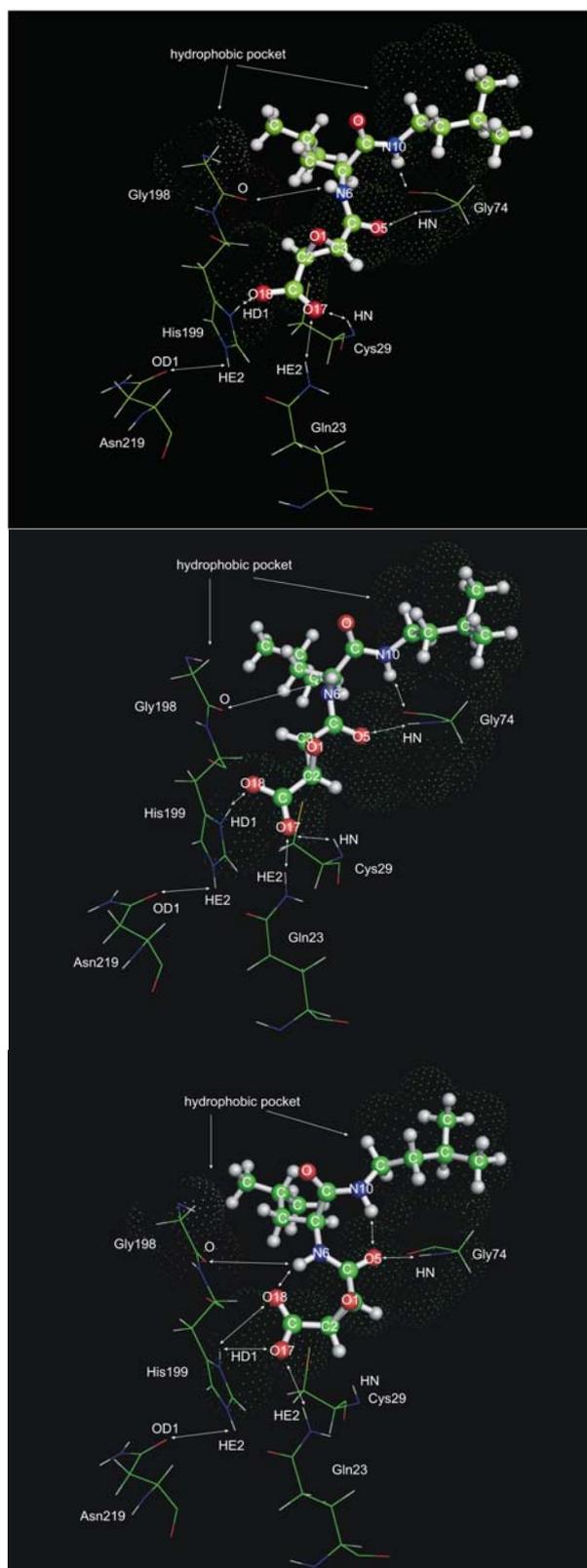


Figure 3.2.20. Reactant structure of configurations (*S,S*) (top), (*R,R*) (middle) and (*R,S*) (bottom) of the E64c inhibitor. Dots indicate residues of the hydrophobic pocket. Atom notation is also given.

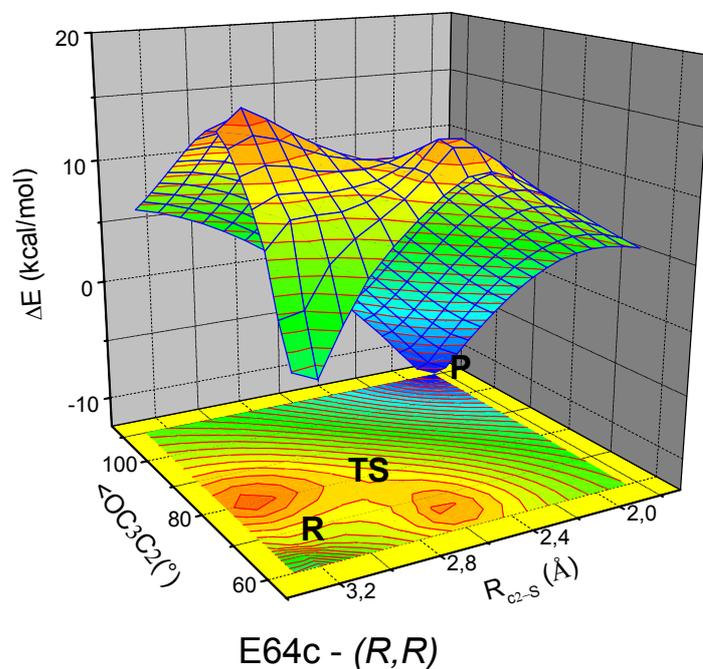
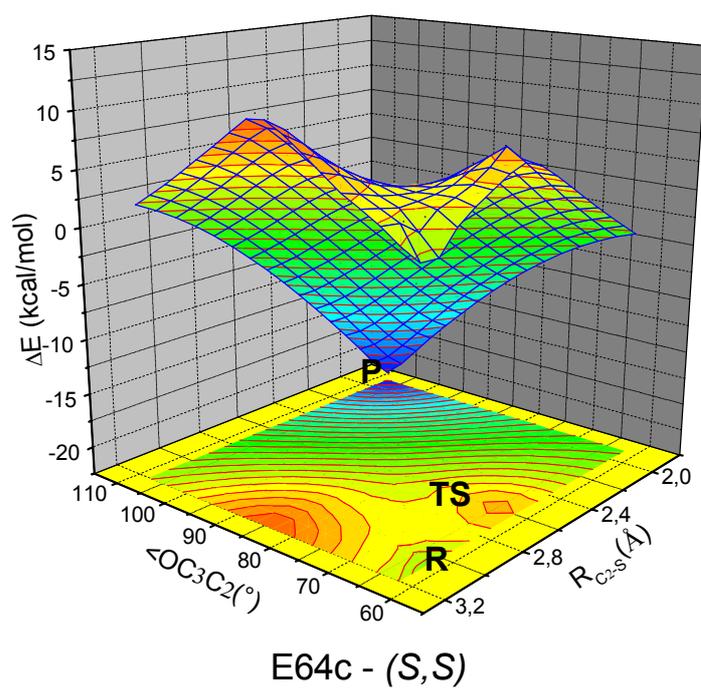


Figure 3.2.21. PES obtained in QM/MM calculations for the E64c-(*S,S*) system (top) and E64c-(*R,R*) system (bottom).

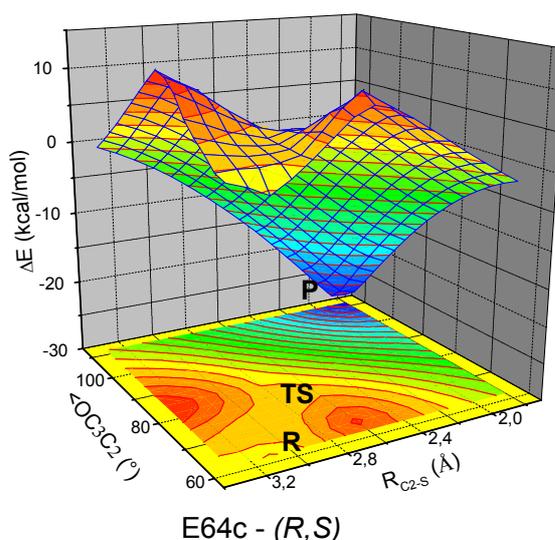


Figure 3.2.22. PES obtained in QM/MM calculations for the E64c-(R,S) system.

Table 3.2.10. Selected geometrical parameters for the stationary points of the reaction profile for the acylation of the cysteine protease by E64c and OX inhibitors. The influence of the absolute configuration on C2 and C3 atoms is compared. All distances are given in Å, angles in degrees (°), and energies in kcal/mol. For notation see Figure 3.2.20.

	rel. E	R_{C2-S}	$\angle\text{OC3C2}$	R_{His199}	R_{Gln23}	R_{Cys29}	R_{Gly74}	R_{Gly198}
E64c-(S,S)								
R	0	3.07	64.0	1.65	2.21	2.97	2.00	3.68
TS	1	2.80	72.4	1.64	2.15	2.67	1.99	3.69
P	-22	1.91	107.4	1.67	1.91	1.96	2.01	3.75
E64c-(R,R)								
R	0	3.21	61.9	1.70	1.86	2.71	2.10	4.75
TS	7	2.73	76.9	1.57	1.89	2.57	2.08	4.80
P	-12	1.90	109.0	1.51	1.83	2.35	2.06	4.52
E64c-(R,S)								
R	0	3.09	64.0	2.75	2.42	4.19	2.06	3.78
TS	1	2.83	73.0	2.70	2.31	4.09	2.05	3.81
P	-29	1.88	110.9	2.87	1.91	3.86	2.04	3.79
OX-(S)								
R	0	3.07	64.7	1.66	1.94	2.31	-	-
TS	2	2.81	71.1	1.62	1.92	2.20	-	-
PH	-47	1.92	112.2	1.58	1.88	1.99	-	-
OX-(R)								
R	0	3.02	64.1	1.65	1.93	2.05	-	-
TS	3	2.81	72.4	1.64	1.90	2.02	-	-
PH	-38	1.91	110.0	1.55	1.87	1.93	-	-

The potential energy surfaces (PESs) for the irreversible step of the inhibition reaction for different systems were obtained by varying the angle $\angle\text{OC3C2}$ corresponding to the ring opening and the distance between the epoxide C2 atom and the sulphur of the attacking Cys29 residue, $R_{\text{C2-S}}$. These parameters are shown in Figures 3.2.21 and 3.2.22 and their values at the stationary points along the reaction profile are given in Table 3.2.10. The alkylation took place at the C2 position in all cases, since due to the regioselectivity of the reaction in the protein environment this is the only possible position at which the nucleophilic attack can take place (see previous subchapter). The starting point for construction of PESs was the relevant product structure. Since there is no X-ray data for the E64c-(*R,R*) and E64c-(*R,S*) product structures they were obtained by manipulation of the E64c-(*S,S*) structure. The change of (*S,S*)- to (*R,S*)-configuration was achieved by switching only the positions of the hydrogen atom and the carboxyl group at the C2 atom of the ring, whereas to obtain the E64c-(*R,R*) product structure we had to interchange the position of the substituents on both sides of the ring.

If one compares the energy profiles of the E64c-(*S,S*) and E64c-(*R,R*) (Figure 3.2.21 and Table 3.2.10), one notices a difference in the barrier height of 6 kcal/mol. Another difference is the 10 kcal/mol more stable product structure of the (*S,S*)-configured inhibitor. This is in good accordance with the experimental results, which show that the (*S,S*)-configured E64-like inhibitors are much more efficient than the (*R,R*)-configured ones.

Table 3.2.10 shows that the reason for the activation energy difference between (*S,S*) and (*R,R*)-configurations may be the difference in geometrical arrangement during the nucleophilic attack. For the (*S,S*)-configuration the reaction starts at $\angle\text{OC3C2}$ value of 64° and $R_{\text{C2-S}}$ distance of 3.07 Å and the transition state (TS) structure is already reached after a change of angle of about 8° and distance of approximately 0.3 Å. In the case of (*R,R*)-configuration, however, in the reactant (R) structure the varied internal coordinates are 61.9° and 3.21 Å, and in order to reach the TS the angle has to increase for 15° and the distance has to decrease for about 0.5 Å. The dependence of the reaction barrier height on the $\angle\text{OC3C2}$ and $R_{\text{C2-S}}$ distance in the TS, relative to the R structure, agrees with the correlation represented in Figure 3.2.10. Giving the epoxide ring (*R,R*)-configuration leads to the position of the inhibitor within the active site being less susceptible to nucleophilic attack. To determine which part of the E64c inhibitor plays the most important role in the stereoselectivity,

additional QM/MM calculations were made on a system containing only a carboxylated warhead as an inhibitor.

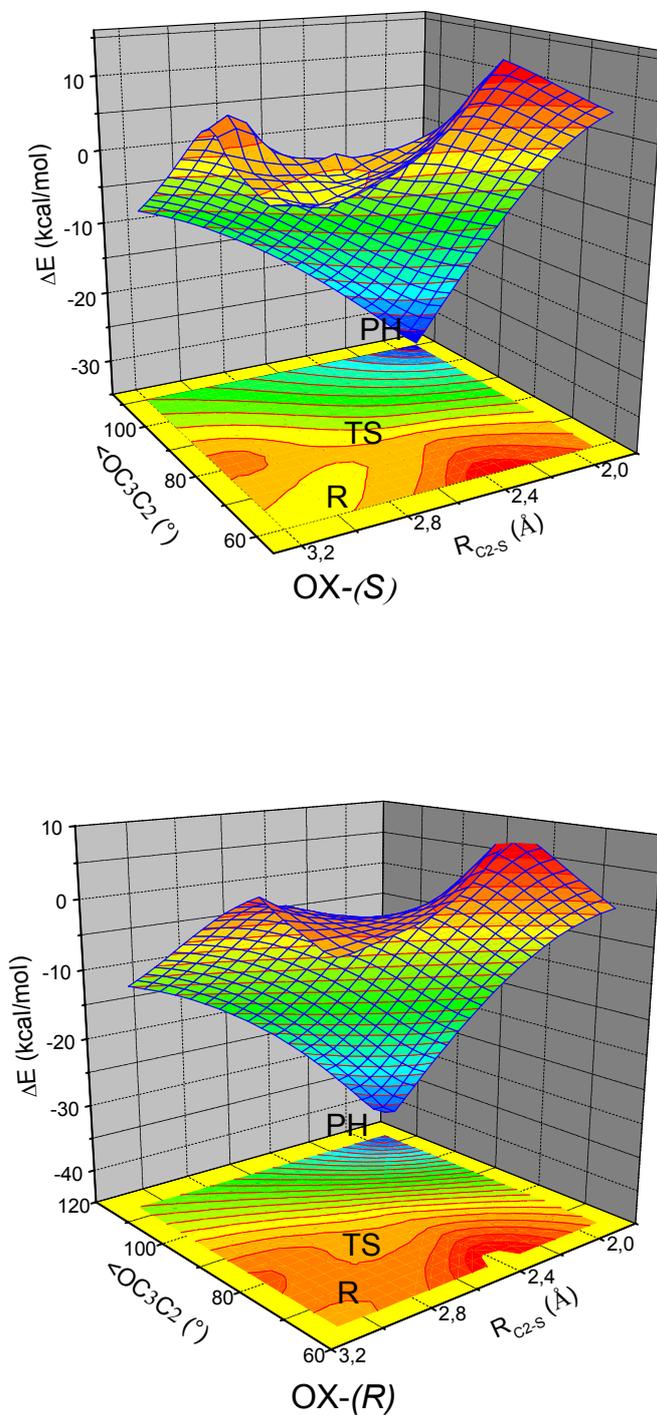


Figure 3.2.23. PES obtained in QM/MM calculations for the OX-(S) system (top) and the OX-(R) system (bottom).

In Table 3.2.10 and Figure 3.2.23 the geometrical parameters and PESs respectively of the inhibition reaction with the inhibitor modelled only by a carboxylated oxirane ring are presented. Such an inhibitor has only one stereo-centre (C2 atom) and for that reason only two stereoisomers exist. We denote them OX-(*S*) and OX-(*R*). Consequently, two PESs were calculated (Figure 3.2.23). The internal coordinates and the structure of the QM part are the same as those in the computations with the whole E64c inhibitor (Figure 3.2.6). The only difference is that this time an ammonium ion was included (treated within the QM part) to allow the protonation of the product (P) structure (the same technique was used in our previous calculations concerning the regioselectivity of the inhibition mechanism). Since the TS structure is not protonated the inclusion of NH₄⁺ influences neither the barrier height nor the reaction rate, and the only consequence is the ~25 kcal/mol lower energy of the P structures due to stabilization of the system by protonation.

Comparison of systems E64c and OX in Table 3.2.10 reveals the same trend for both in the change of reaction energies between (*S*)- and (*R*)-conformers (once the difference of 25 kcal/mol due to protonation is disregarded). The (*S*)-conformer of the P structure is about 10 kcal/mol more stable. While the trend is preserved in the reaction energy, the trend in the kinetics is different. For OX no significant difference in the relative TS energy between OX-(*S*) and OX-(*R*) is found. The barrier heights for the two cases are only 1 and 2 kcal/mol higher, respectively, than for E64c-(*S,S*). If the values of the reaction coordinates are compared one notices that both OX-(*S*) and OX-(*R*) have R structures of similar geometry to that of E64c-(*S,S*) (angle of about 64° and distance of about 3 Å). Furthermore, they undergo a similar change in these internal coordinates on the way to the TS, and again fit nicely the TS energy – TS <C2C3/R_{C2-S} correlation in Figure 3.2.10. These results clearly show that the carboxylate substituent alone cannot be responsible for the stereoselectivity. Such a small inhibitor, although anchored within the oxyanion hole and bound via hydrogen bonds to His199 (see Figure 3.2.6), is still flexible enough to place its C2 atom in the optimal position for the nucleophilic attack regardless of stereochemistry.

The stereoselectivity must be due to the C3-substituted peptidomimetic side chain of E64c, which interacts with the substrate (S) binding site of the enzyme in several ways. Three of its atoms build hydrogen bonds with the residues of cathepsin B (Figure 3.2.20): O5 with backbone N of Gly74, N6 with backbone O of Gly198, and N10 with backbone O of Gly74. In addition the aliphatic residues of the inhibitor side chain fit the hydrophobic pockets

formed by the side chain of Phe75, constituting the S3 subsite, and the Pro76, Ala 173, Gly198, and Ala200 residues which constitute the S2 subsite. To determine the importance of these interactions MD computations at the pure molecular modelling level were performed on the reactant and product structures for all three E64c systems.

The MD calculations are also important for the following reasons. Since no X-ray structures are available, the reaction products for the other two stereoisomers were derived from the E64c-(*S,S*) crystal structure by reorienting the corresponding substituents and minimizing the resulting system. These structures were taken as starting points for calculation of the minimum energy paths of the respective reactions, leading to the corresponding reactants. While such a profile correctly reflects the reactions necessary to obtain a given product, it does not provide information whether the starting point, the reactants, can feasibly be reached – which they might not be, if the non-covalent interactions between the inhibitor and the enzyme lead to geometrical arrangements which are considerably lower in energy than the reactant found by using QM/MM to back-extrapolate from the product. To gain information about whether such arrangements exist, MD simulations were performed on the reactant structures. Since in this step we focus solely on the reversible interactions a classical force field simulation is sufficiently accurate, and actually better suited since all interactions leading to the new S-C bond are not included. The goal of the MD simulation is to analyse how the different stereoisomers of the E64c inhibitor warhead influence its position in the active site. Of special interest are the energy differences of conformations in which the carbon atoms are oriented towards the S⁻ center of the attacking Cys29. However, other arrangements, in which the S⁻ forms hydrogen bonds with hydrogen bond donors of the inhibitor (N6-H or N10-H), are also likely to occur. These are of less interest for the present study since they cannot lead to the ring opening reaction; however, MD simulations with a negatively-charged sulphur center are actually biased towards such conformations: the corresponding hydrogen bonds are attractive while, within a classical approach, the sulphur-carbon interactions are strongly repulsive, which would not be the case in a quantum mechanical approach. Since QM/MM MD calculations are prohibitively expensive for the treatment of the questions at hand, some way of avoiding unwanted conformations had to be found. Quantum chemical approaches found only small barriers for the ring-opening reaction within the active site in all cases. Consequently, at the reactant structure, the S⁻ ... C interaction is only slightly repulsive, or not at all. This situation is obtained in a force field if the charge on the sulphur centre is switched off. In addition, this measure leads to a

destabilization of those inhibitor conformers possessing hydrogen bridges to the sulphur centre. This is a second advantage since those conformations are not of interest for the present question. Therefore, all MD simulations on reactant structures were performed with a neutral sulphur. Our assumption that once the active site sulphur is assigned negative charge in the force field calculations, conformers would appear which cannot bind to the active site, was investigated in further MD runs and proved to be true.

Figures 3.2.24 and 3.2.25 present, for reactant and product structures respectively, the time dependence of some of the heavy-atom distances between E64c and the cathepsine B binding site obtained in MD computations for all three configurations of the ring. These include distances corresponding to both hydrophobic and H-bond interactions. The distance variations given in these two figures can clearly be separated into two groups: those belonging to the upper part of the graph (4 to 6 Å), and those belonging to the lower part (around 3 Å). The longer distances in the upper part are longer and fluctuate more with the time. The interactions described by these distances are all from the S2 and S3 subsites and hydrophobic in nature. Our MD modelling describes these interactions as weak ones, and hence predicts – as expected – a high flexibility of this part of the inhibitor. This correlates nicely with the fact that X-ray structures are known for two forms of E64c inhibitor bound to papain (cystein protease with a similarly-constructed the binding site to cathepsine B) in which the leucyl side chain and the isoamylamide moiety (see Scheme 3.2.1 and Figure 3.2.20, and also next sub-chapter) are rotated in opposite direction (the 1PE6⁹¹ and 1PPP¹⁰⁰ pdb structures). In the 1PE6 structure the leucyl side chain is in the S2 pocket, while the terminal isoamylamide group occupies the S3 pocket. 1PPP, however, shows the reverse orientation.

In the lower part of the chart in Figures 3.2.24 and 3.2.25 distances corresponding to H-bonded interactions are found. These are, specifically: O17 – side chain amide of Gln23, O17 – backbone carbonyl of Cys29, O18 – ND1 of His199, O5 – backbone amide of Gly74 and N6 – backbone carbonyl group of Gly198 (Figure 3.2.20).

Additionally, in Figure 3.2.26 the time dependence of the distance between the C2 atom of the ring and the non-protonated sulphur of Cys29 in the reactant structure are plotted for all three configurations.

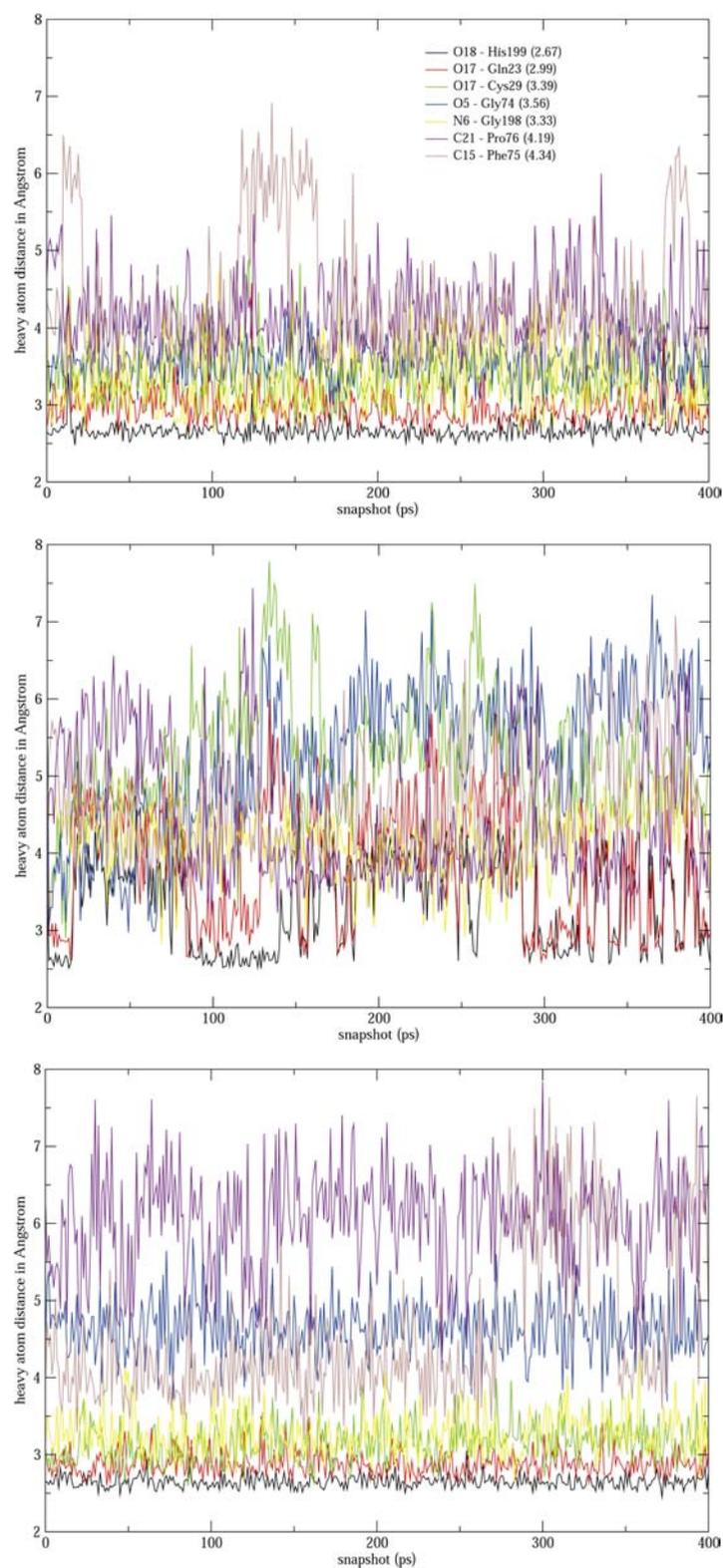


Figure 3.2.24. Time dependence of selected distances between E64c molecule and cathepsin B (heavy atoms) obtained in molecular dynamics calculations on reactant structures. Top: E64c-(*S,S*), middle: E64c-(*R,R*), and bottom: E64c-(*R,S*) system. For details see legend at the top.

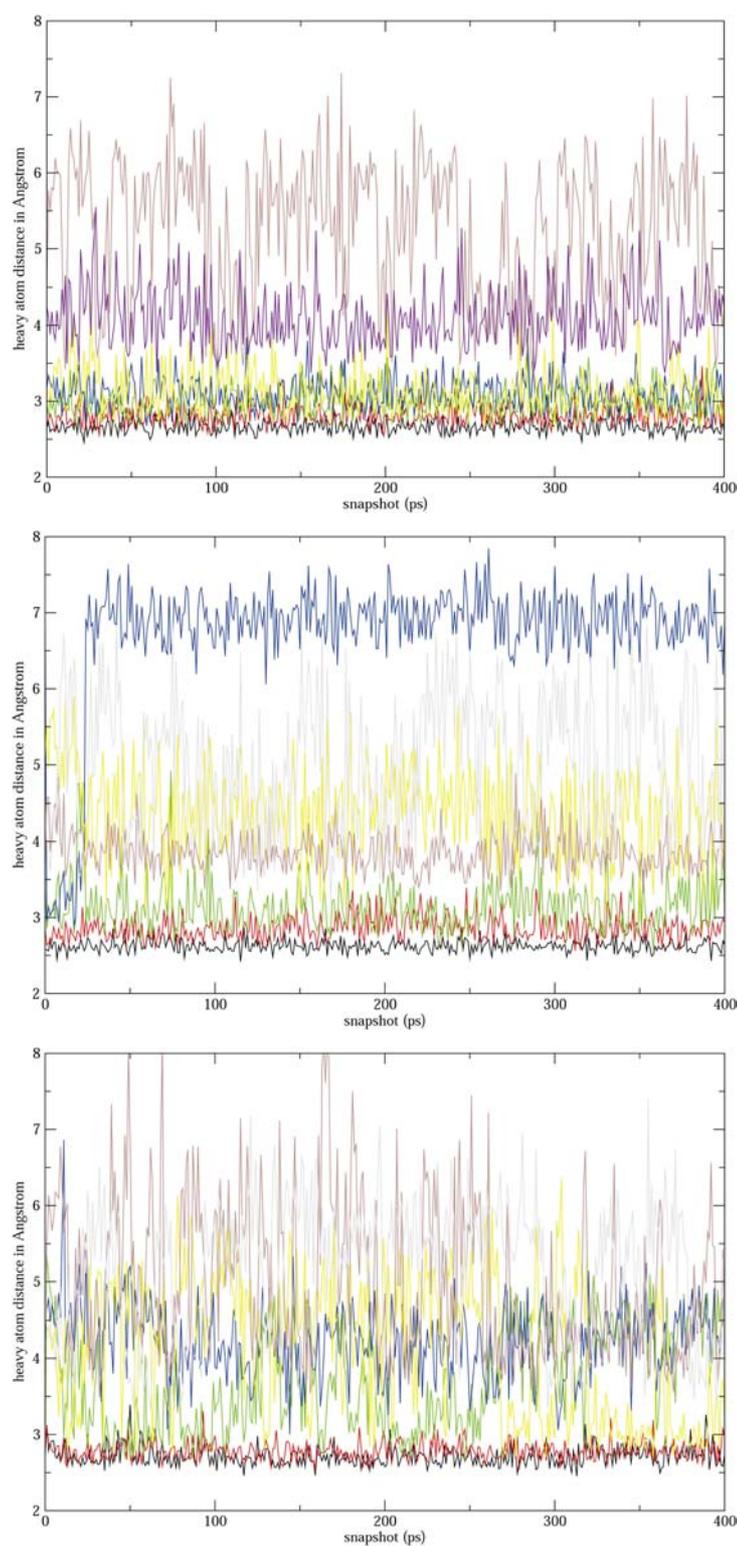


Figure 3.2.25. Time dependence of selected distances between E64c molecule and cathepsin B (heavy atoms) obtained in molecular dynamics calculations on product structures. Top: E64c-(*S,S*), middle: E64c-(*R,R*), and bottom: E64c-(*R,S*) system. For details see legend at the top of Figure 3.2.24.

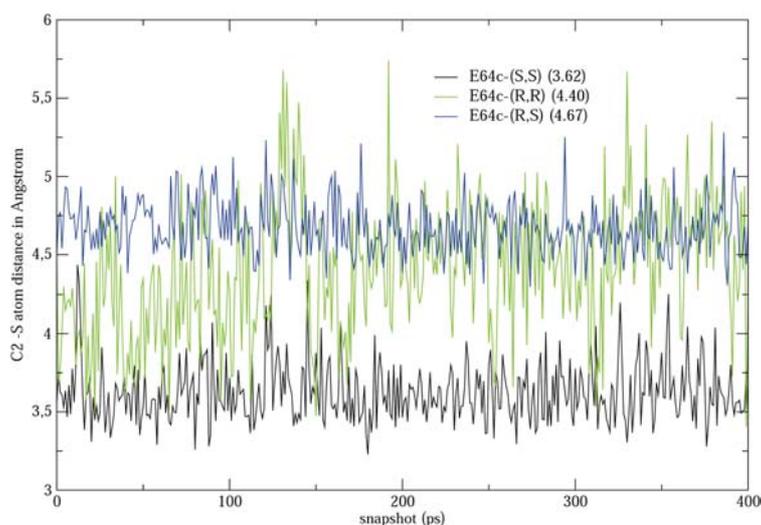


Figure 3.2.26. Time dependence of distance between C2 atom of the ring and sulphur of Cys29 residue for different configurations of the E64c molecule obtained in molecular dynamics calculations on reactant structures. Values in parentheses are average distances.

Looking at the MD trajectories for both the reactant and product structures of different systems it becomes obvious that the hydrophobic interaction is not configuration dependent, but stays unchanged in all three systems. For this reason it also cannot be directly responsible for the stereospecificity of the inhibition reaction. Let us, then, concentrate on H-bond interactions, starting with the E64c-(*S,S*) system. In both the reactant and product MD hydrogen bonds are well preserved, indicating a stable structure. The position of the C2 atom is dictated by the hydrogen bond network, and its distance from the attacking S⁻ of the thiolate group of the Cys29 moiety fluctuates between 3.2 and 4 Å (Figure 3.2.26), making this configuration a perfect starting point for the alkylation step. In the reactant structure of the E64c-(*R,R*) system, however, this H-bond pattern is disrupted. These hydrogen bond interactions become so weak that the carboxylate group of the inhibitor starts rotating within the binding site. In Figure 3.2.24, this is represented by the increase of the O18-His199, O17-Gln23 and O17-Cys29 distances to about 4Å. In such an arrangement the carboxylate is rotated and O17 interacts with His199, while O18 fits into the oxyanion hole. However, this new hydrogen bond network is much less stable. As a consequence the inhibitor is not fixed in place within the active site and the ring is not brought into the optimal position with respect to Cys29, which can also be seen from Figure 3.2.26. Since the S-C distance gets longer, the \angle OC3C2 angle of the reactant decreases to 60°, which is the value of the undisturbed ring. According to the correlation between the barrier height and the \angle OC3C2 angle of the reactant

(Figure 3.2.10), the actual reaction barrier will therefore increase by 1-2 kcal/mol. The influence of the newly formed S-C bond is seen in Figure 3.2.25. Here the distances O18-His199 and O17-Gln23 remain at values around 2.5-3 Å. They are constrained to do this by the newly-formed covalent bond. The O5-Gly74 H-bond, however, is a weaker interaction than in E64c-(*S,S*). After 30 ps of the MD simulation, this interaction completely dissociates (see the blue line in Figure 3.2.25). Although the (*R,R*)-configuration is not optimal for the inhibitor, the alkylation of this configuration can still take place. This can be ascertained from Figure 3.2.26: while the average R_{C2-S} distance in the reactant structure of E64c-(*R,R*) is indeed longer than in the E64c-(*S,S*), it does sometimes drop to values that would allow a successful attack by the sulphur centre. The reaction will not proceed as readily as for E64c-(*R,R*) than for E64c-(*S,S*), but it will take place.

For E64c-(*R,S*), however, this is not the case; the inhibitor does not bind at all. The (*R,S*)-configured system is the most intriguing one, since the QM/MM and the MD results for it do not match. QM/MM calculations predict that the E64c-(*S,S*) system should inhibit the enzyme as fast as E64c-(*R,S*) does. A reaction barrier of 1 kcal/mol is found for both systems (Table 3.2.10). The (*R,S*)-configuration even has the most favorable reaction energy, according to the QM/MM computations (-12 vs. -22 vs. -29 kcal/mol). However, geometrical parameters in Table 3.2.10 reveal that in E64c-(*R,S*) the H-bond between the O17 atom of the inhibitor and the Cys29 backbone (part of the oxyanion hole) is broken. O17 interacts instead with His199 via a hydrogen bond (see Figure 3.2.20, bottom). This interaction actually becomes the strongest H-bond between the inhibitor and the enzyme in this configuration. The O18 atom of the carboxylate still interacts with the His199 residue, but this previously strong interaction is now weakened due to the intramolecular H-bond between O18 and N6-H. The strong non-covalent interaction between the inhibitor and the oxyanion hole has to be abolished for the C2 atom of the ring to be brought into the vicinity of the attacking sulphur, which is not the case for the other configurations. The inhibitor adopts such a geometrical arrangement because the QM/MM calculation brings us from the TS to the nearest local minimum on the energy hypersurface, not taking into account the possibility of other, energetically more favorable, conformers nearby. Such minima can be found, however, by the MD simulations. At the bottom of Figure 3.2.24 the time dependence of the important non-covalent interactions between the enzyme and the E64c-(*R,S*) is shown. It is clear that the H-bond network, which was found to be disrupted in the QM/MM calculations of the reactant structure, in fact remains intact, since it strongly stabilizes the complex. However, to enable

the hydrogen bonding network (O18-His199, O17-Cys29, O17-Gln23 and N6-Gly198), the inhibitor ring has to be moved away from the position represented at the bottom of Figure 3.2.20. When this happens, the C2 atom (and C3 as well) is too far away to be attacked by the negatively charged sulphur. This can be clearly seen from the plot of R_{C2-S} along the trajectory length (the blue line in Figure 3.2.26).

Our computations indicate that the hydrogen bonds between the enzyme and the epoxysuccinic base of the inhibitor formed in the reversible step of the inhibition reaction are responsible for the stereospecificity of the reaction by determining the position of C2 in respect to S⁻. The (*S,S*)-configuration of the warhead provides the best geometrical arrangement for the formation of these H-bridges. However, some (*R,R*)-configured inhibitors interacting with the S' subsite show better reactivity than their (*S,S*)-counterparts. According to our calculations, the reason for this is the formation of an H-bond between the amide group of the peptidomimetic side chain of the inhibitor and the backbone carbonyl of the Gly198 residue.

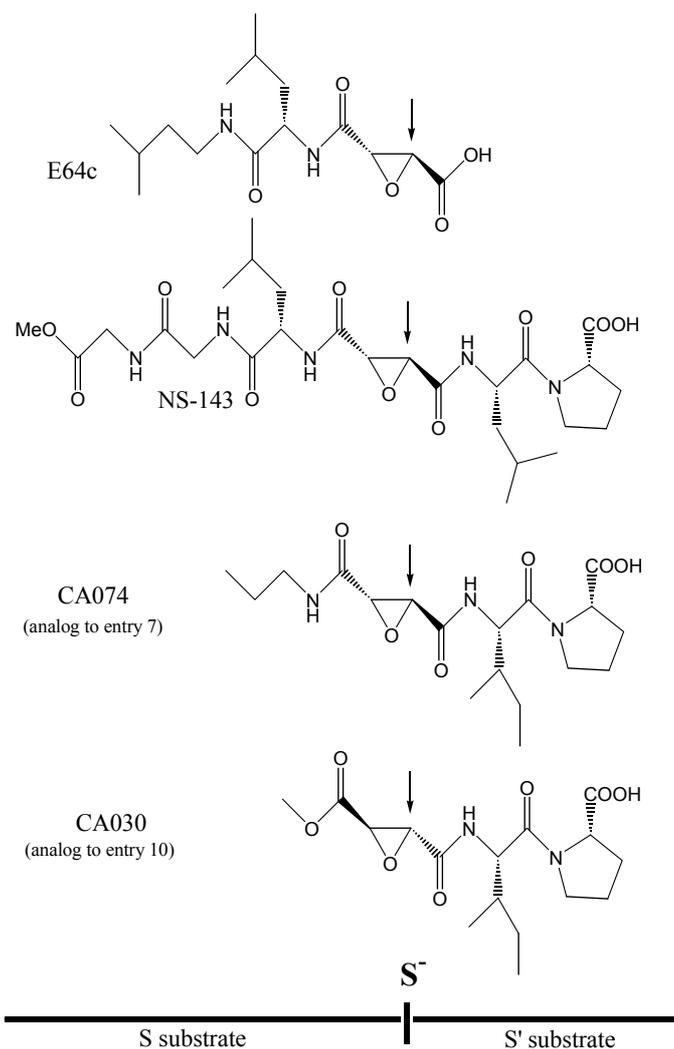
Table 3.2.11 gives information on all five inhibitor-enzyme H-bonds found in the QM/MM stationary point structures obtained for the various configurations. Comparing E64c-(*S,S*) to E64c-(*R,R*), the only significant difference appears in the interaction between the inhibitor and the backbone of the Gly198 residue (approximately 1 Å difference), which is also reflected in the pictures in Figure 3.2.20. Since the QM/MM structures represent the one minimum among many closest to the 1ITO pdb structure from whence they are derived, it is important to take into account the results of MD investigations, which scan all nearby minima. The average N6-Gly198 H-bond distance obtained via molecular dynamics is 2.09 Å in the product and 3.33 Å in the reactant structure of E64c-(*S,S*). For E64c-(*R,R*) the picture changes. The strong fluctuations in N6-Gly198 indicate that in the reactant structure this H-bond bridge almost ceases to exist. The reason for this difference is that the N6 atom is close to a very flexible and large aliphatic side chain of the inhibitor. Its distance to Gly198 backbone (the yellow line in Figures 3.2.24 and 3.2.25) varies more with time than the distances representing H-bond interactions closer to the active site, indicating that it represents a weaker interaction. The weakened H-bond network between the epoxysuccinyl moiety of the inhibitor and the active site found for the E64c-(*R,R*) configuration allows the aliphatic chain move deeper into the hydrophobic pocket. This pulls the N6H amide group away from the Gly198, breaking the H-bond.

Table 3.2.11. Second-order rate constants of cathepsin B inhibition by (*S,S*)- and (*R,R*)-epoxysuccinyl-based inhibitors.

Entry	Inhibitor	k_2/K_i ($M^{-1}s^{-1}$)
1	Agm-Leu-(<i>S,S</i>)-tEps-OH (E64)	81400 ± 2760^{98}
2	Agm-Leu- (<i>R,R</i>)-tEps-OH	1170 ± 40^8
3	(CH_3) ₂ HC-(CH_2) ₂ -HN-Leu-(<i>S,S</i>)-tEps-OH (E64c)	175000^6
4	(CH_3) ₂ HC-(CH_2) ₂ -HN-Leu-(<i>R,R</i>)-tEps-OH	-
5	MeO-Gly-Gly-Leu-(<i>S,S</i>)-tEps-Leu-Pro-OH (NS-143)	1520000 ± 88800^{99}
6	MeO-Gly-Gly-Leu-(<i>R,R</i>)-tEps-Leu-Pro-OH	214600 ± 11980^{10}
7	nPrNH-(<i>S,S</i>)-tEps-Leu-Pro-OH	153000 ± 7590^8
8	nPrNH-(<i>R,R</i>)-tEps-Leu-Pro-OH	29400 ± 1320^8
9	EtO-(<i>S,S</i>)-tEps-Leu-Pro-OH	44400 ± 1690^8
10	EtO-(<i>R,R</i>)-tEps-Leu-Pro-OH	567000 ± 21500^8
11	EtO-(<i>R,R</i>)-tEps-Leu-Agm	73 ± 5^8

The observations made on our model system allow a general explanation of the stereoselectivity of the cysteine protease inhibition by epoxysuccinyl-based inhibitors. In Scheme 3.2.1 the binding modes of some inhibitors are given. Table 3.2.11 shows the second-order rate constants for cathepsin B inhibition by different stereoisomers of various inhibitors. For all of the inhibitors in Table 3.2.11, the (*S,S*)-stereo-isomers show higher activity - with the exception of EtO-tEps-Leu-Pro-OH, for which the second order rate measured for the (*R,R*)-configuration (entry 10) is approximately 12 times larger than that of its (*S,S*)-counterpart (entry 9). It is important to note that this situation reverses on changing the side chain binding to the S subsite from EtO to nPrNH (compare entries 9 and 10 to entries 7 and 8, Table 3.2.11). These two inhibitors are identical to CA030 and CA074 apart from one side chain amino acid (Ile is replaced with Leu). If the inhibitor structures given in Scheme 3.2.1 are compared, one notices that only CA030 has an oxygen atom attached to the first carbonyl group on the left-hand side of the ring, whereas all others have an NH moiety in this position. This position corresponds to the N6 position in E64c and, if occupied by a proton donor group (like NH in CA074), could build a hydrogen bond with the backbone carbonyl of the Gly198 residue. As seen from our calculations, the (*S,S*)-configuration would favour this interaction. However, if there is a proton acceptor group (as in CA030) in this position a strong repulsion would occur and a larger distance between these two moieties would be preferred, which would occur with the (*R,R*)-stereoisomer. Indeed, the distances between heavy atoms that correspond to this interaction in the X-ray structures of the CA030–cathepsin B ((*R,R*)-configuration, 1CSB pdb structure⁹⁶) and CA074–cathepsin B ((*S,S*)-configuration, 1QDQ)

inhibitor–enzyme complexes are 3.77 Å and 3.40 Å, respectively. In other cysteine protease–(*S,S*)-inhibitor complexes this distance is also approximately 3 Å (1PE6, 1PPP, 1ATK⁹⁰, 1CVZ¹⁰¹).



Scheme 3.2.1. Binding modes of epoxysuccinyl-based inhibitors to cysteine proteases based on the X-ray structures. S⁻ is the negatively charged sulphur of the active site Cys29 residue. From top to bottom: **E64c**: HO-(2*S*,3*S*)-tEps-Leu-NH-(CH₂)₂-CH(CH₃)₂; **NS-143**: MeO-Gly-Gly-Leu-(2*S*,3*S*)-tEps-Leu-Pro-OH; **CA030**: EtO-(2*R*,3*R*)-tEps-Ile-Pro-OH; **CA074**: nPrNH-(2*S*,3*S*)-tEps-Ile-Pro-OH. The arrow indicates the acylation position.

It can also be seen from Table 3.2.11 that the side chain of the inhibitor plays an important role in its activity. Replacement of Pro-OH by Agm in EtO-(2*R*,3*R*)-tEps-Leu-Pro-OH (the last two rows of Table 3.2.11) leads to almost complete deactivation of the inhibitor. By

contrast, the side-chain structure and the additive effect of ring substitution in both the P and P' directions in NS-143 raise its k_2/K_i (second-order rate constant) value to an enormous $1520000 \text{ M}^{-1}\text{s}^{-1}$. The importance of the side-chain interaction with the enzyme binding pocket for the inhibitor activity is dramatically demonstrated by the calculations on the E64c-(R,S) system.

3.2.6 Using MD Simulations to correct the existing X-ray data

To check the accuracy of our approach we carefully compared the 1ITO pdb structure with the geometrical arrangement obtained after the QM/MM optimization procedure described above, both for the system containing only a carboxylated ring as model for the inhibitor and for the system with E64c (Table 3.2.12). An obvious difference is in the distance between the E64c O18 atom (Figure 3.2.12) and the nitrogen atom (ND1 in Figure 3.2.12) of the His199 residue. In the pdb structure a value of 3.18 \AA is given, while a value of around 2.65 \AA is obtained after the optimization procedure (Table 3.2.12). However, the simulated models and the starting 1ITO structure agree nicely on all other distances. For the present study the difference in $R_{O18-ND1}$ is an important issue, since the theoretically-determined distance of 2.65 \AA indicates a strong interaction is present, which the experimental distance of 3.18 \AA contradicts; and hence, theory predicts the CO_2^- -His199 interaction to be one of the most important effects governing the inhibition reaction rate, while the X-ray structure does not. As a consequence of the large $R_{O18-ND1}$ distance in the 1ITO X-ray structure neither Yamamoto *et al.*⁵⁵ nor Powers *et al.*⁶ consider the strong salt bridge between the carboxylate and the His199 as important factor for the inhibition potency which plays a dominant role in our picture. The 1PE6⁹¹ structure also indicates this interaction to be negligible, as the distance it gives is even longer ($R_{O18-ND1} = 3.8 \text{ \AA}$). By contrast, however, the structures 1ATK,⁹⁰ 1AEC⁴⁵ and 1PPP¹⁰⁰ give distances of 2.8 - 2.9 \AA , indicating the strong bond predicted by our computations.

The deviation may result from experimental uncertainties. Table 3.2.13 shows the experimental conditions under which the X-ray structures of various cysteine proteases were obtained. Although generally reliable, the resolutions of 2.10 \AA and 2.29 \AA at which 1PE6 and 1ITO structures respectively were taken seem insufficiently accurate to enable precise location of the inhibitor within the active site.¹⁰² Also, a high temperature factor (B-factor) is seen for both the E64c moiety and the His199 residue.

The distances obtained from X-ray experiments are averaged over the nuclear motion, so differences could also result from a shallow potential energy well allowing large amplitude vibrations or librations. We decided to examine these effects by performing MD runs.

Another discrepancy between our results and the experimental findings is found for the $R_{O17-NE2}$ distance. The 1PPP structure gives a value of 2.4 Å, which conflicts with the values given by our computations, and indeed all other X-ray structures (2.8 - 2.9 Å, Table 3.2.12). A distance $R_{O17-NE2} = 2.4$ Å would indicate a different ordering of the importance of the various factors in the inhibition reaction than our computations predict, calling our model into question.

Table 3.2.12. Comparison of the computed and X-ray structure distances between H-bond-linked heavy atoms in the active site of different cysteine proteases. For notation see Figures 3.2.12 and 3.2.6. All distances in Å. Notations in parentheses correspond to different residue numbering in the enzyme. Distances presented for the MD calculations are the time-averaged values.

Structure	O18-ND1	O17-NE2	O17-N	O5-N
QM/MM: OX- α	2.65	2.84	2.94	-
QM/MM: E64c / His199 protonated	2.64	2.88	2.91	2.95
QM/MM: E64c / His199 not protonated	4.20	3.07	2.73	2.93
MD: E64c / His199 protonated	2.66	2.82	3.00	3.11
MD: E64c / His199 not protonated	3.69	2.86	2.93	3.22
1ITO ⁵⁵	3.18	2.84	3.07	3.12
1ATK ⁹⁰	2.95	2.95	3.06	3.07 (Gly66)
1AEC ⁴⁵	2.80	2.91	2.94	2.92 (Gly68)
1PPP ¹⁰⁰	2.94	2.46	3.09	3.56 (Gly66)
1PE6	3.76	2.90	2.75	3.16 (Gly66)

The 1ITO X-ray structure was obtained at a pH value of 4.5. In this pH regime three different protonation states are imaginable (pK_a (His) \approx 6; pK_a (COOH) \approx 2-4). The histidine could be protonated while the carboxylate is deprotonated (HisH⁺/CO₂⁻), both could be protonated (HisH⁺/CO₂H), or neither could be protonated (His/CO₂⁻). (The His/CO₂H situation is highly unlikely since the CO₂H substituent is considerably more acidic than the His group.) To test the influence of the protonation state on the geometrical arrangement MD runs were performed for the HisH⁺/CO₂⁻ and the His/CO₂⁻ situation. While the distances connected with hydrophobic groups varied strongly along the MD trajectories, the ND1–O18 distance varied

by only 0.5 Å in the HisH⁺/CO₂⁻ simulation (Figure 3.2.11), and its average value was less than 0.1 Å greater than the QM/MM prediction (Table 3.2.12). Both facts underline the strength of the interaction involved. On the other hand, an average ND1–O18 distance of about 3.7 Å was found in the His/CO₂⁻ simulation, which is far too large, compared to experiment. It is possible that the experimentally measured ND1–O18 distance of about 3.2 Å arises as an average value for a system in which about 50% of the His199 residues are protonated, with $R_{O18-ND1} = 2.7$ Å, while the rest are deprotonated, with $R_{O18-ND1} = 3.7$ Å. This theory would be supported by the higher average B-factor of the inhibitor (B = 39) compared to the protein (B = 17) (Table 3.2.13). No simulation was performed on the HisH⁺/CO₂H system, as even larger distances would be expected in this case.

The existence of a partially-deprotonated His199 moiety might explain not only the $R_{O18-ND1}$ distance of 3.2 Å found in the 1ITO structure, but also the even longer $R_{O18-ND1}$ distance (3.8 Å) found in the 1PE6 structure. Our QM/MM simulations with a deprotonated His199 residue (Table 3.2.12) predict $R_{O18-ND1} = 4.2$ Å, which is longer than the measured value. However, if the His199 group is protonated in only a fraction of the enzymes, this might explain the experimental result. Additionally, due to the repulsive interaction between the deprotonated His199 residue and the carboxylate moiety strong anharmonic effects can be expected, which would cause thermal H-bond-lengthening. Such effects are indeed indicated by preliminary MD simulations. They predict distances around 3.7 - 3.9 Å with strong fluctuations along the MD trajectories (Figure 3.2.27).

Table 3.2.13. Comparison of resolution, experimental pH value, and average B-factor of active site histidine residue and inhibitor of X-ray experiments on different cysteine proteases.

Protease	inhibitor	MEROPS PDB name	resolution (Å)	X-ray-diffraction pH value	average B-factor of His (K)	average B-factor of E64(-c) ^{a)} (K)
actinidin	E64	1AEC	1.86	6.0	4	7
cathepsin K	E64	1ATK	2.20	6.2	9	7
cathepsin B	E64c	1ITO	2.29	4.5	17	39
papain	E64c	1PPP	1.90	9.2	8	24
papain	E64c	1PE6	2.10	7.0	15	23

a) Only the temperature B-factor of the atoms of the inhibitor ring and its carboxylate substituent are taken into account.

To clarify the situation a superimposition of the backbones of the 1ITO pdb structure and of the structure obtained from the optimization of the $\text{HisH}^+/\text{CO}_2^-$ system was performed. This showed that both structures match perfectly except in the position of the opened three-membered ring inhibitor. A close inspection of all corresponding structural parameters revealed a striking difference in the out-of-plane angle C3-C16-C2-S and in the angle $\angle\text{SC2C3}$ (Figure 3.2.28). The optimized structure possesses $\text{C3-C16-C2-S} = 122^\circ$ and $\angle\text{SC2C3} = 110^\circ$, reflecting the expected sp^3 environment of the of the C2 centre. The 1ITO structure gives 162° and 115° , respectively, indicating a more sp^2 -hybridized C2 centre. This indicates that the optimization procedure corrected an error within the pdb structure.¹⁰³

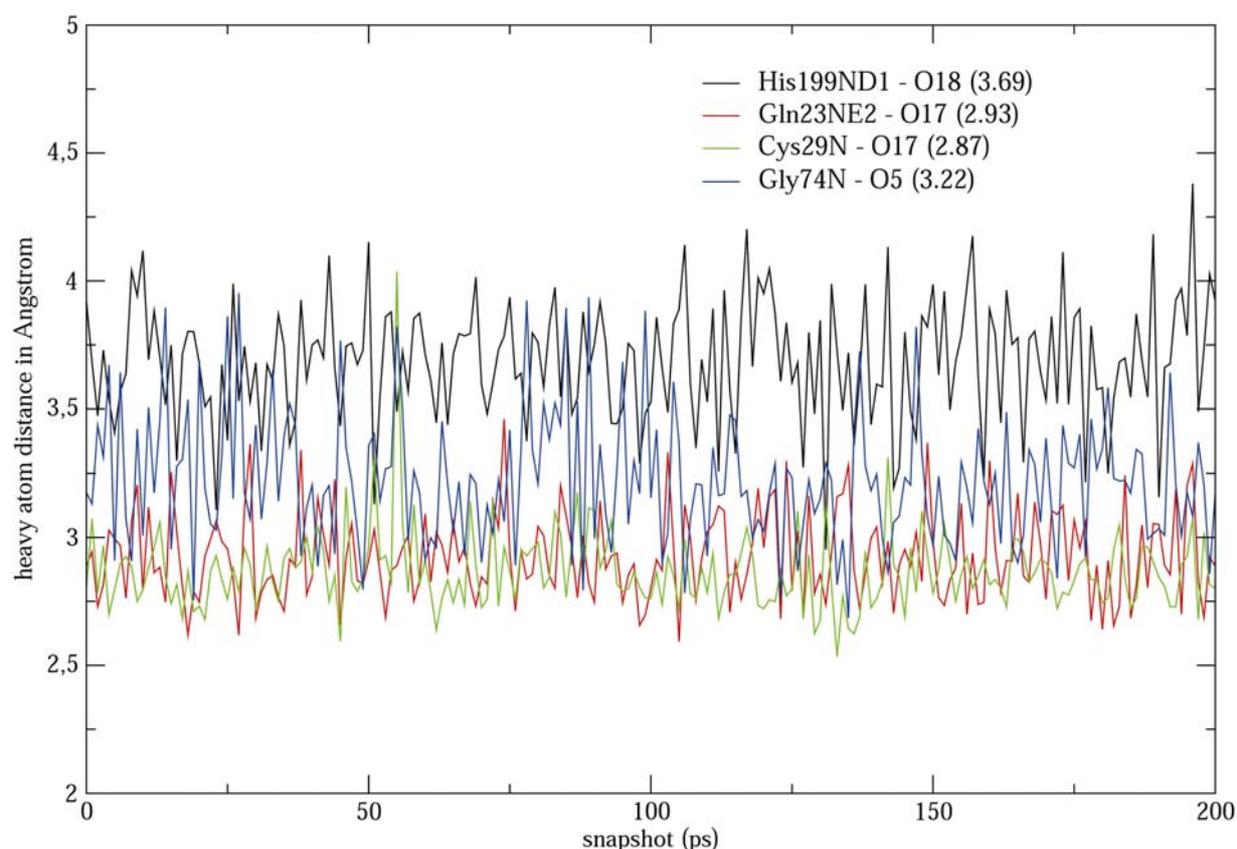


Figure 3.2.27. Time-dependence of selected distances between E64c and cathepsin B (heavy atoms) obtained from molecular dynamics calculations in which His199 residue was not protonated. Distances representing hydrogen-bond interactions with Gly74, Gln23, Cys29, and His199 are presented. In the top-right corner a legend for the colour-coding of the interactions is given. The nomenclature of the atoms is the same as used in the 1ITO pdb file. The numbers in parenthesis are simulation-average distances in Å.

In the 1PPP X-ray structure the isoamylamide group of E64c interacts with the S₂ subsite of the enzyme, while the leucyl group binds to the S₃ binding site. In the 1PE6 structure the situation is opposite, although the enzyme and the inhibitor are the same (Figure 3.2.29). This conformational difference, however, does not interfere with the binding motif of the succinylcarboxy moiety.¹⁰⁰

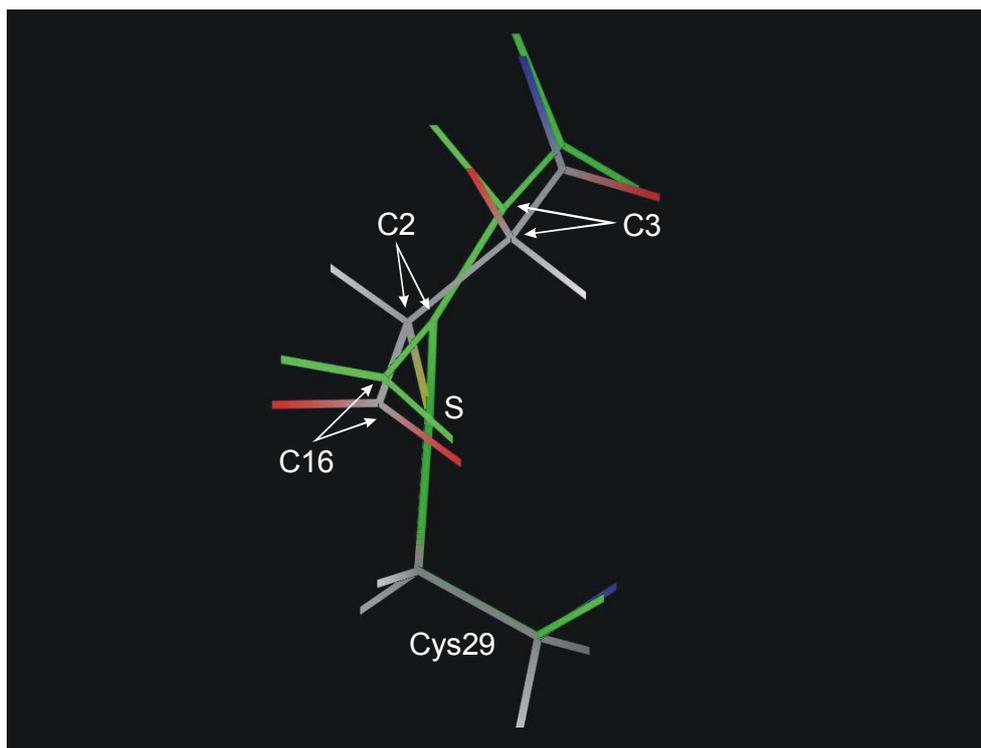


Figure 3.2.28. Difference in inhibitor positions from the 1ITO structure (green) and the structure obtained in calculations on the HisH⁺-CO₂⁻ system (colored) after superposition of the enzyme backbone. Atom numbering corresponds to the 1ITO pdb structure.

The backbone superimposition of the two structures (Figure 3.2.29) showed that in the 1PPP structure (green), the side-chain beta C atom (CB) of Gln23 residue takes a different conformation than in the 1PE6 structure (in red). Further investigation revealed that the other structures in Table 3.2.12 also have the so-called “a” alternate conformation of the CB atom of the corresponding glutamine residue, just like in the 1PE6 structure, in contrast to the 1PPP structure which possesses the less occupied “b” alternate conformation. The difference in the conformation of the CB atom also influences the position of the NE2 atom, shifting it closer

to the carboxyl moiety of the inhibitor and making the $R_{O17-NE2}$ distance shorter. Our calculations start from the 1ITO structure and retain the starting “a” alternate configuration of the CB atom of Gln23 residue.

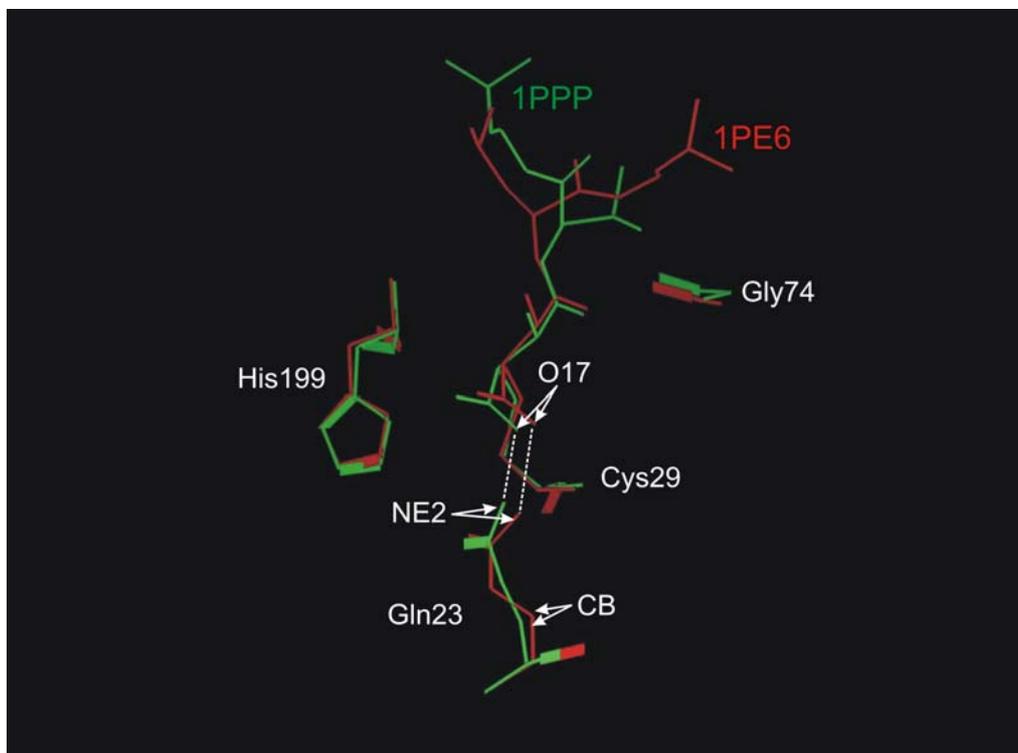


Figure 3.2.29. Difference in inhibitor position between 1PE6 structure (red) and 1PPP structure (green) after backbone superposition. Atom and residue numbering corresponds to the 1ITO pdb structure.

3.3 QM/MM MD Calculations

3.3.1 The method

The starting structure used to model the active cathepsin B protease was obtained by deleting the inhibitor moiety in the cathepsin B-E64c enzyme-inhibitor complex (1ITO⁵⁵ X-ray structure). Preparation of the system was done by means of fully classical MD simulations using the CHARMM program package⁵⁷ in the same way as already described in Section 3.2.1. The whole system was solvated in water, and then subjected to a series of harmonically restrained minimizations, MD runs, and addition of further solvent molecules until uniform solvent density was reached. Finally, a 200 ps long MD run (NVT, 300 K) was performed in which all constraints except those for the residue ion-pair of the active site (His199 and Cys29) were released. During this run snapshots were taken every 40ps. The five structures obtained as snapshots at 40, 80, 120, 160, and 200 ps were used as starting structures in QM/MM MD equilibration and production runs. We forbore from using other cysteine protease X-ray structures without a bound inhibitor (1HUC,¹⁰⁴ 1PPN,¹⁰⁵ 9PAP¹⁰⁶) since close inspection of the active sites revealed no significant differences in the geometrical arrangement of the ion-pair and surrounding residues compared to the 1ITO structure.

Due to the enormous computational effort needed for calculation of the QM part in the QM/MM MD approach the size of the system had to be reduced to make the computations feasible. This was achieved by restricting the water layer around the protein surface to a thickness of 12 Å. The counterions were kept and placed at the new surface of the water sphere. During the simulations the enzyme and a water layer of 3 Å around the enzyme were allowed to move. The 9 Å thick “frozen” water shell was used to create a boundary between the system and the surrounding vacuum, preventing the mobile water molecules from escaping. It thus provides a kind of effective boundary potential to the system. To simplify the simulation, all hydrogen atoms in the system were replaced by deuterium, i.e. had their mass set to $m_d=2.014$ a.u. The calculations were done using the ChemShell program⁵⁸ where the DIPoly force field⁵⁹ was used to describe the behavior of the MM part and the AM1²⁶ semi-empirical method was employed in the QM part. The active water molecules were only allowed to move freely in the rotational and translational degrees of freedom; internal

vibrational motion was frozen. The QM region contained the residues of the active site, His199, and Cys29. The His199 residue was terminated at the CA position (see Figure 3.3.1) by replacing this atom with the H link atom. The same was done for the Cys29 residue. At the QM/MM boundary the charge-shift scheme was applied and fixed MM point charges were included into the QM Hamiltonian to provide electrostatic embedding of the QM region. The reversible non-iterative leapfrog-type integrator¹⁰⁷ was used for integration of the equations of motion. The simulated temperature was 300 K (NVT conditions), controlled by a Berendsen thermostat⁷³ (coupling time 0.1 ps) during the 10 ps long equilibration and by a Nosé-Hoover chain thermostat¹⁰⁸ (chain length was 4, characteristic coupling time of the thermostat to the physical system was 0.02 ps) during the 20 ps-long production phase. The time step used was 1 fs.

To isolate various effects and to analyse the observations made in QM/MM MD runs, additional QM and QM/MM calculations were performed at the RIDFT⁴⁰/B-LYP⁴¹/TZVP⁴² level of theory. These calculations were carried out for various model systems focusing on different parts of the system. For all model systems the distance between the His199 ND1 atom and the HD1 hydrogen (for atom notation see Figures 3.3.1 and 3.3.5) was used as a reaction coordinate. Since finding the reaction energies was the goal in these calculations and the precise localization of the transition state structures was not necessary, we omitted further calculations on the DFT/B3⁴³-LYP/TZVP level. Such calculations are important for accurately determining the height of reaction barriers; however, they yield the same results for the energy difference between the reactants and the products.⁴⁴

3.3.2 Explaining the stability of the active site HisH⁺/CysS⁻ ion-pair in cysteine proteases

Cysteine proteases are a large group of enzymes whose proteolytic activity depends on a thiol group of cysteine residue situated in the active site. In an active enzyme an ion-pair is established between the cysteine residue and the nearby histidine. In Figure 3.3.1 the active site of cysteine protease cathepsin B in form of His199H⁺/Cys29S⁻ ion-pair is depicted. The existence of this ion-pair is no longer disputed. In the late 70's and early 80's its presence was experimentally proved by potentiometric difference titrations,¹⁰⁹ isotope effect measurements,¹¹⁰ and proton nuclear magnetic resonance spectroscopy.^{111,112} Kinetic studies

on different thiol proteases showed that there are two ionisable groups involved in the catalysis, one with a pK_a value of around 4 and another with pK_a around 8.5. The former is usually attributed to the deprotonation of the Cys29 residue (cathepsin B numbering), whereas the latter is believed to reflect the ionization of His199. However, the pK_a values of cysteine and histidine residues in proteins are usually about 8-9 and 6, respectively. These unusual pK_a values can only be explained by considering the local enzyme environment of both residues in more detail.

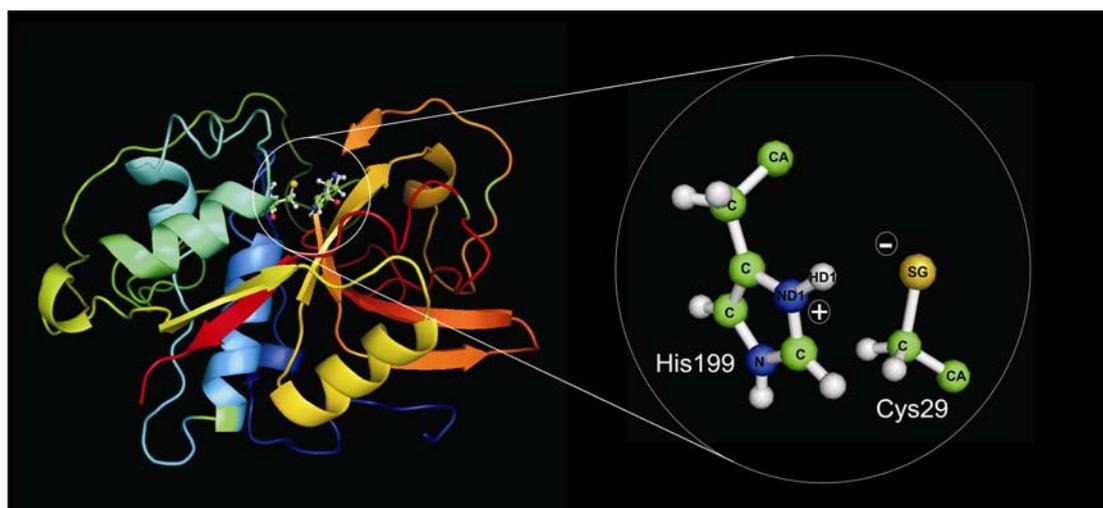


Figure 3.3.1. Active site ion-pair in cathepsin B (His199H⁺ and Cys29S⁻) (right) together with its position within the enzyme (left). Notation for the atoms of the active site is given.

Many molecular groups around the active site are considered to be of importance for the proper functioning of the cysteine proteases.¹¹³ Earlier quantum chemical studies¹¹⁴ on papain have proposed that the macrodipole of the helix near the active site (L1-helix, Figure 3.3.2) facilitates the proton transfer from cysteine to histidine residue and in this way stabilizes the zwitterionic form of the active site. The helix macrodipole is the sum of the dipole moments of the individual peptide units oriented along the helix axis. These microdipoles, one of which is contributed by each peptide bond, are the result of the partial double-bond character of the C-N bond and are parallel to the N-H and C=O bonds. While initial investigations into this phenomenon stated that the entire α -helix is involved in stabilization of the ion-pair via its macrodipole, recent results show that it is the first turn of the helix that accounts for about 80% of the overall charge stabilization effect, possibly via hydrogen bond interactions.¹¹⁵

Further indication of the importance of the terminal hydrogen bonds has been provided by both mutagenesis studies¹¹⁶ and Hartree-Fock calculations.¹¹⁷ In order to shed more light onto this subject, and to try to explain the high reactivity of the negatively charged Cys29 residue despite its stabilization, we performed QM/MM MD calculations on the protein. Figure 3.3.2 shows how the negatively charged sulphur center of the cysteine residue of the active site is situated on top of the L1 α -helix (in red) and is involved in a network of H-bond bridges to His199, Ala200, Gln23, and Trp30 residues as well as to an amide residue of its own backbone and a couple of solvent water molecules. In the calculations the active site residues' side chains (right hand side in Figure 3.3.1) were described by using a quantum mechanical approach while the rest (protein and water surroundings) was calculated at on the classical force field level. Five different systems were calculated, with their starting structures taken from the classical MD production run. A snapshot was taken from the classical MD trajectory every 40 ps and used as a starting point for the QM/MM MD calculations. All the QM/MM MD trajectories are labelled according to the time (40ps, 80ps, 120ps, 160ps or 200ps) at which their starting structure was extracted from the classical MD (e.g. the 40ps system had a starting structure taken at 40 ps, 80ps one taken at 80 ps, etc.).

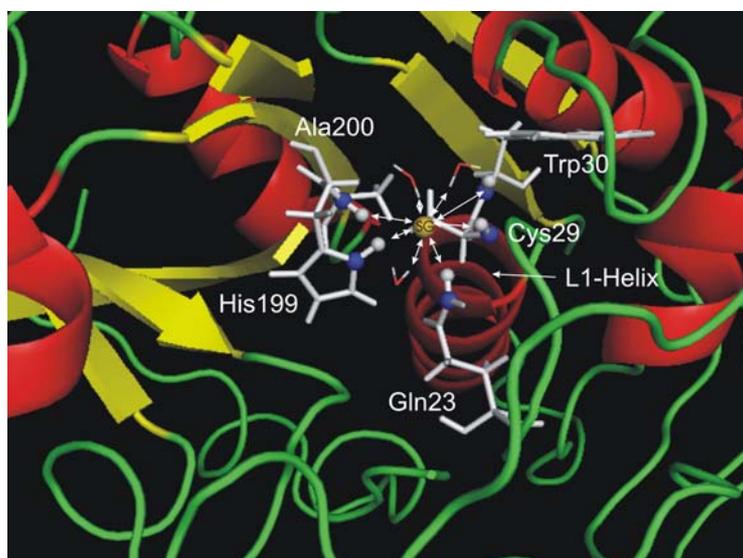


Figure 3.3.2. The active site ion-pair in cathepsin B (His199H^+ and Cys29S^-) together with the nearby water solvent molecules and protein surroundings. The hydrogen bond network between the sulphur of the Cys29 and the surrounding water and enzyme atoms is shown. The amide groups of the enzyme involved in H-bonds, as well as S^- of Cys29, are given in ball-and-stick representation.

In Figures 3.3.3 and 3.3.4 the time dependence is given of interatomic distances representing the H-bond network depicted in Figure 3.3.2. For all systems the first 200 snapshots represent equilibration with the Berendsen thermostat⁷³ while the rest is the production phase performed using the Nose-Hoover thermostat¹⁰⁸ (see the method section above). Snapshots were extracted from the QM/MM MD calculations every 0.05 ps.

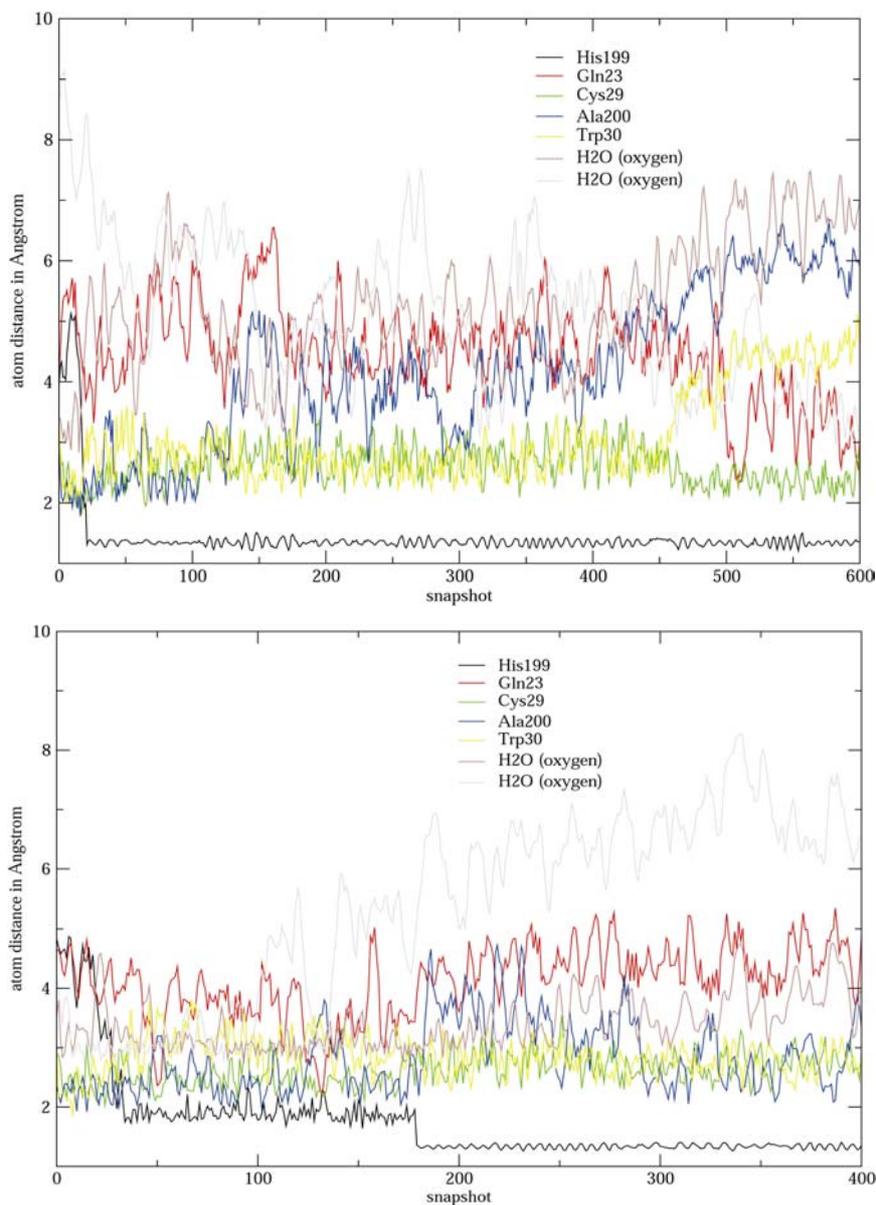


Figure 3.3.3. Time dependence of the hydrogen bond network between the sulphur of the Cys29 and the surrounding water and enzyme atoms (Figure 3.3.2) for the 40ps (top) and 160ps (bottom) trajectories. Distances to water molecules are presented as distances between the sulphur of Cys29 and the oxygen atom of water. Otherwise distances are between sulphur and hydrogen. Protonation of S⁻ of Cys29 occurs.

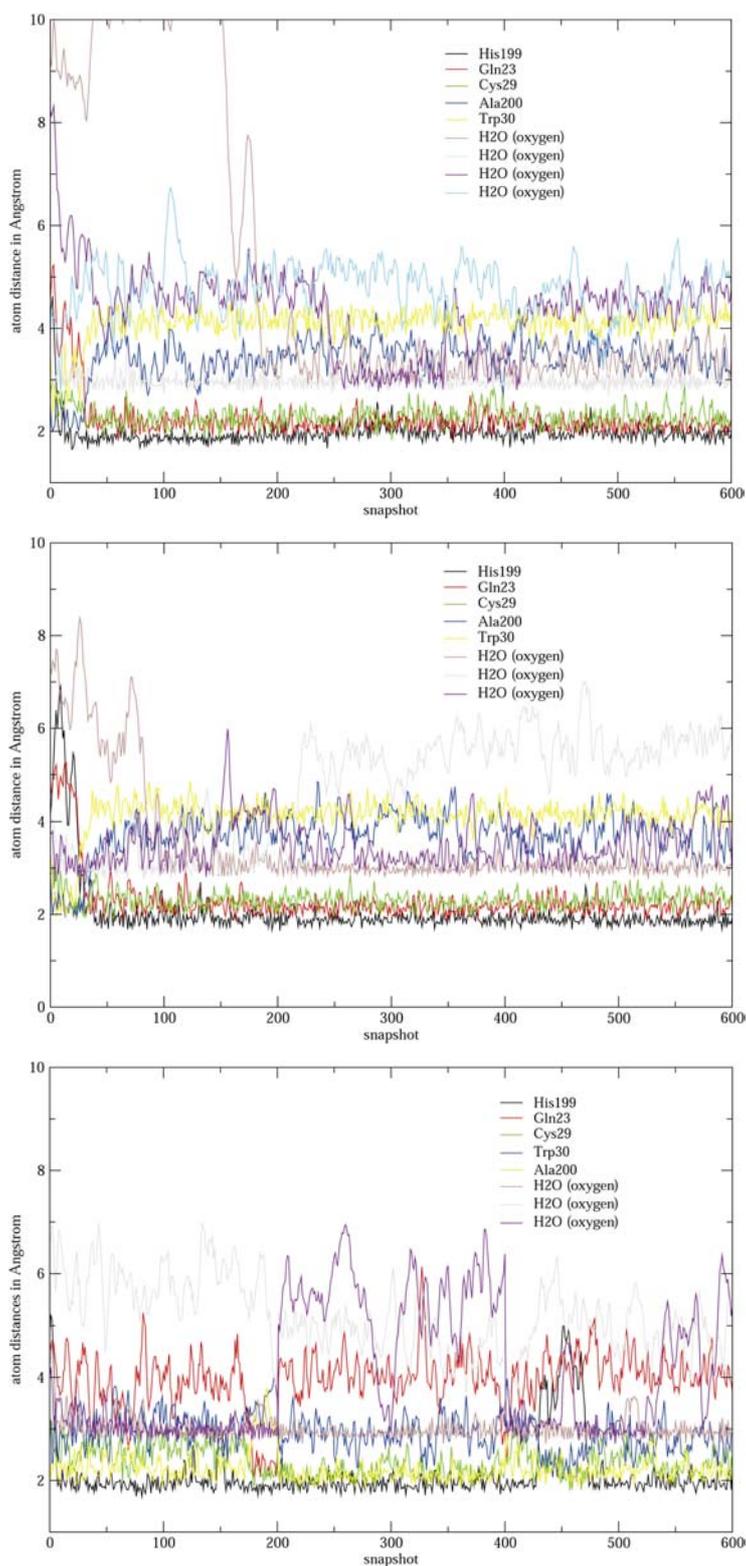


Figure 3.3.4. Time dependence of the H-bond network between the sulphur of the Cys29 and the surrounding water and enzyme atoms (Figure 3.3.2) for the 80ps (top), 120ps (middle), and 200ps (bottom) trajectories. Distances to water molecules are presented as distances between the sulphur of Cys29 and the oxygen atom of water. Other distances are between sulphur and hydrogen. No protonation of S^- of Cys29 occurs.

Since in the hybrid QM/MM MD calculations the ion-pair is within the QM part, the proton transfer between the His199 and Cys29 residues is allowed (in contrast to pure classical calculations which cannot describe bond-creation or bond-cleavage processes). Indeed, in two of the five trajectories (40ps and 160ps, Figure 3.3.3) protonation of the S⁻ of the Cys29 residue by the His199 residue does happen. Detailed analysis of Figures 3.3.3 and 3.3.4 reveals the reasons for this behavior.

First we concentrate on the systems in which no proton transfer appears, which are the systems 80ps, 120ps and 200ps. The time evolution of their H-bond networks (Figure 3.3.2) is given in Figure 3.3.4. One clearly sees that in all three systems there are 4 or 5 strong and stable hydrogen bonds, around 2 Å in length, stabilizing the negatively charged sulphur of the Cys29 residue. Approximately two of these are H-bridges to the solvent water molecules. (Please note that H-bonds to water molecules are described by the distance between the S⁻ and an oxygen atom of water, since the water molecules are mobile and can, by rotation, easily switch the hydrogen atom involved in the H-bond. The actual H-bond distance is typically about 1Å shorter than the S-O distance plotted in the figures.) The negatively charged sulphur center forms three other strong hydrogen bonds – either to the side chains of the His199 and the Gln23 residues and the amide group of Cys29 backbone (80ps and 120ps), or to the His199 side chain and the amide groups of the Cys29 and the Ala200 residue backbones (200ps) (see Figure 3.3.2). Beside these strong interactions, there are also somewhat weaker hydrogen bonds (at distances of 3–4 Å) providing additional stabilization for the deprotonated Cys29 residue. These are usually formed between S⁻ and Trp30, Gln23, Ala200 and additional water molecules. It is important to note that, since the potential hydrogen bond donors (His199, Cys29 backbone, Gln23, Ala200, Trp30, see Figure 3.3.2) are located in every direction around the Cys29 residue, the thiolate residue will be stabilized regardless of the direction in which it moves. If it moves downwards in Figure 3.3.2 it will be stabilized mostly by Gln23, His199 and Cys29 (80ps and 120ps). If it moves up the Ala200 and Trp30 residues take over (200ps).

In both the 40ps and 160ps trajectories (Figure 3.3.3), proton transfer is found to take place, in the form of simultaneous deprotonation of the His199 residue and protonation of Cys29. In both systems proton transfer occurs during the equilibration period (Berendsen thermostat⁷³) where larger potential energy fluctuations are possible. By looking at the time evolution of the hydrogen bond distances in these two systems (Figure 3.3.3) one notices that the pattern of 4–

5 strong, stable H-bonds is missing. In the 40ps system protonation of the S⁻ takes place at the very beginning. This QM/MM MD run starts with four H-bonds (Ala200, Trp30, Cys29 and H₂O), but the distance variations are very large. Once the water molecule starts to drift away (the brown line in Figure 3.3.3) there is no nearby hydrogen bond donor that could take over, and the H-bond net falls apart. The influence of the water molecule leaving is even better reflected in the system 160ps. The collapse of the hydrogen bond net in this case is obviously due to the water molecule which starts to drift away after 100 snapshots (5 ps) (gray line in Figure 3.3.3).

The calculations indicate that stabilization of the active site ion-pair does not solely result from the L1 α -helix total dipole. If this were the main stabilizing factor, deprotonation of His199 would not occur once the H-bond network collapses, since the dipole of the helix would not be changed. Although the calculations strongly indicate the importance of the hydrogen bonds between the Cys29 sulphur atom and the surrounding enzyme residues and water molecules, they do not deliver a quantitative picture. To inspect in detail the influence of each of the members of the H-bond network, and to determine their importance for the survival of the ion-pair, pure QM calculations were done on several simplified systems. Two of them are depicted in Figure 3.3.5.

In the QM calculations the influence of single residues was investigated by a so-called “knock-out” method, in which different interactions are switched off one by one. In our case that means that the residues were removed from the calculations one at a time, leading to six different systems. For each system the reaction path for the proton transfer from His199 to Cys29 was calculated using the His199ND1-HD1 atom distance (see Figure 3.3.5) as the reaction coordinate. To obtain the minimum energy paths (MEPs) for the proton transfer this distance was fixed at a certain value while all other geometrical parameters were relaxed. To maintain a geometrical arrangement similar to the one within the enzyme the distances between the two terminal C-atoms of His199 residue and the two carbons of each of the other residues (in Cys29 all 4 carbons, Figure 3.3.5) were kept fixed. The starting structure was the last snapshot in the 80ps calculation. To keep the QM calculations feasible, complete residues were represented by smaller model units (Figure 3.3.5) which could still capture the major interactions between the enzyme environment and the ion-pair. His199 was represented by an imidazolium ion and Gln23 and Asn219 residues by acetamide. The Cys29 residue was represented by its side chain and the amide group of the backbone. This amide group forms a

peptide bond together with the carboxyl group of the Ser28 residue (not included). The whole peptide bond (terminated by a methyl group on the other side) was included in the calculations. In all calculations the COSMO³⁹ continuum solvent model was used to mimic the polarizability of the surrounding water and the peptide environment.

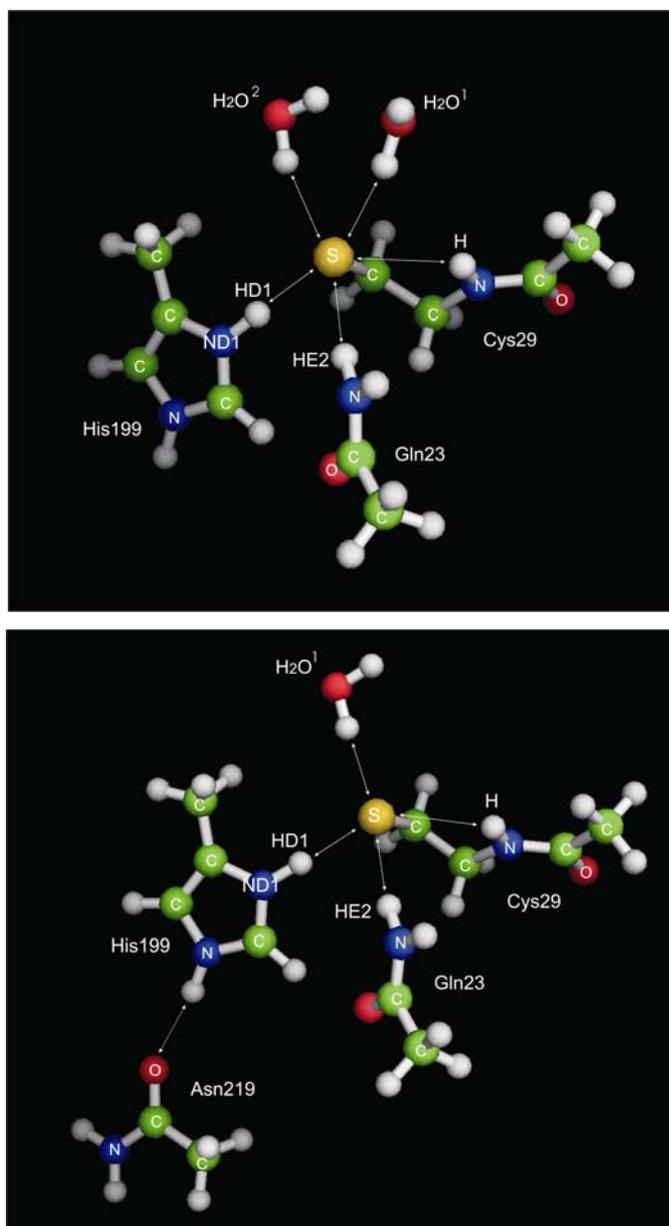


Figure 3.3.5. Top: model system calculated at the QM level of theory, comprising His199 (represented by methylimidazolium ion), Cys29 (with amide group of backbone), Gln23 (represented by acetamide) and two water molecules. Atom notation is given. Bottom: the same system only with one water molecule less and an additional acetamide molecule representing the Asn219 residue.

Table 3.3.1. Selected geometrical parameters for points along the reaction paths corresponding to the deprotonation of the His199 residue computed for different systems in pure QM calculations. All distances are given in Å. For definition of the various distances see Figure 3.3.5. Figures in bold correspond to optimized product or reactant structures.

entry	ΔE (kcal/mol)	$R_{\text{ND1-HD1}}$	$R_{\text{S-HD1}}$	$R_{\text{S-HE2}}$	$R_{\text{S-H(Cys29)}}$	$R_{\text{S-H}_2\text{O}}^1$	$R_{\text{S-H}_2\text{O}}^2$
System 1: ion-pair + Cys29 (b.b.), Gln23, 2 H₂O, COSMO							
1	0	1.063	2.112	2.403	2.742	2.319	2.338
2	3	1.2	1.874	2.384	2.822	2.323	2.391
3	5	1.3	1.719	2.388	2.962	2.323	2.391
4	7	1.4	1.607	2.400	3.154	2.323	2.391
5	7	1.5	1.531	2.421	3.243	2.323	2.391
6	7	1.6	1.478	2.460	3.282	2.323	2.391
7	8	1.7	1.443	2.501	3.305	2.323	2.391
8	8	1.8	1.426	2.535	3.288	2.323	2.391
System 2: ion-pair + Cys29 (b.b.), Gln23, 1 H₂O, COSMO							
9	0	1.074	2.079	2.389	2.689	2.246	...
10	2	1.2	1.862	2.370	2.753	2.271	...
11	4	1.3	1.71	2.378	2.847	2.294	...
12	5	1.4	1.601	2.392	2.937	2.318	...
13	5	1.5	1.527	2.405	3.061	2.342	...
14	5	1.6	1.476	2.445	3.16	2.357	...
15	4	1.7	1.441	2.490	3.171	2.369	...
16	4	1.798	1.417	2.535	3.215	2.379	...
System 3: ion-pair + Cys29 (b.b.), Gln23, 1 H₂O, Asn219, COSMO							
17	0	1.069	2.150	2.314	2.719	2.241	...
18	2	1.2	1.887	2.337	2.760	2.265	...
19	4	1.3	1.726	2.362	2.803	2.290	...
20	5	1.4	1.61	2.387	2.901	2.314	...
21	5	1.5	1.533	2.404	3.032	2.335	...
22	4	1.6	1.482	2.416	3.160	2.353	...
23	4	1.7	1.449	2.435	3.161	2.367	...
24	4	1.888	1.409	2.461	3.235	2.383	...
System 4: ion-pair + Cys29 (b.b.), Gln23, COSMO							
25	0	1.080	2.059	2.378	2.664
26	2	1.2	1.849	2.362	2.727
27	3	1.3	1.702	2.365	2.782
28	4	1.4	1.594	2.390	2.888
29	3	1.5	1.520	2.405	2.968
30	3	1.6	1.471	2.432	3.100
31	3	1.7	1.437	2.464	3.068
32	2	1.872	1.403	2.530	3.132
System 5: ion-pair + Cys29 (b.b.), COSMO							
33	0	1.099	1.991	...	2.661
34	1	1.2	1.812	...	2.710
35	2	1.3	1.673	...	2.762
36	2	1.4	1.578	...	2.849
37	1	1.5	1.510	...	2.918
38	0	1.6	1.464	...	3.027
39	0	1.7	1.434	...	3.029
40	-1	1.993	1.387	...	3.107

Continuation of **Table 3.3.1**.

System 6: ion-pair + COSMO							
41	0	1.133	1.962
42	0	1.2	1.811
43	1	1.3	1.668
44	1	1.4	1.574
45	0	1.5	1.509
46	-1	1.6	1.465
47	-2	1.7	1.434
48	-4	2.074	1.381

Table 3.3.1 gives relative energies as well as some geometrical parameters for points along the reaction path for all six calculated systems. Figure 3.3.6 summarizes all reaction paths. The energies are given relative to the structure in which ion pair is formed in the active site. Let us begin at the system represented in the top of Figure 3.3.5 (Table 3.3.1, entry 1-8, black line in Figure 3.3.6). Apart from the residues of the active site, it also contains the side chain of the Gln23 residue and two additional water molecules. In such a system, despite the lack of L1 helix, the ion-pair form of the active site is 8 kcal/mol more stable than the neutral form. It is not even possible to find a local minimum for the protonated sulphur state, and the structure for entry 8 of Table 3.3.1 had to be obtained in a single point calculation with a reaction coordinate set to the value that corresponds to those obtained for the other reaction paths. The energy difference between the ion-pair and the neutral form is slightly overestimated since one of the water molecules stabilizing the ion-pair had to be kept fixed to a certain distance to avoid its drifting away during the course of the reaction. If only one water molecule was included in the system, it showed no tendency towards wandering away and no constraints had to be set. In such a system (entry 9-16, red line in Figure 3.3.6) the ion-pair is still 4 kcal/mol more stable, and the proton transfer reaction has a barrier of around 5 kcal/mol. When the second water molecule is removed from the system (Table 3.3.1, entry 25 - 32, blue line in Figure 3.3.6), the neutral form is only 2 kcal/mol less stable than the ion-pair.

In systems 5 and 6 of Table 3.3.1 (light blue and pink lines in Figure 3.3.6) the Gln23 residue and both Gln23 and the Cys29 backbone were taken out of the calculations, respectively. Comparison of the relative energies of system 4 and 5 indicates a 3 kcal/mol stabilization of the ion-pair due to the presence of Gln23 residue. Once the Gln23 is not included in the system, the neutral form of the active site becomes more stable (albeit only by around 1 kcal/mol) than the zwitterionic form. According to the energy difference between systems 5

and 6, the hydrogen bond between the Cys29 backbone and the S^- stabilizes the ion-pair by only 3 kcal/mol. In the literature¹¹⁷ one finds that this stabilization energy is estimated to be around 7 kcal/mol. In our pure QM calculations the distance between the NH group of the Cys29 backbone and the Cys29 sulphur (column 6 of Table 3.3.1) is longer than the one within the enzyme. This is because, in the enzyme, the NH group cannot relax by moving away from the protonated sulphur due to the constraints of the enzyme structure. This destabilizes the neutral state, and hence increases the relative stabilization energy of the ion-pair. In the simplified QM model system, however, the NH group is free to move further away from the protonated sulphur atom, additionally stabilizing the neutral state. The QM/MM calculations describe the situation in the enzyme more accurately as they include these geometrical constraints.

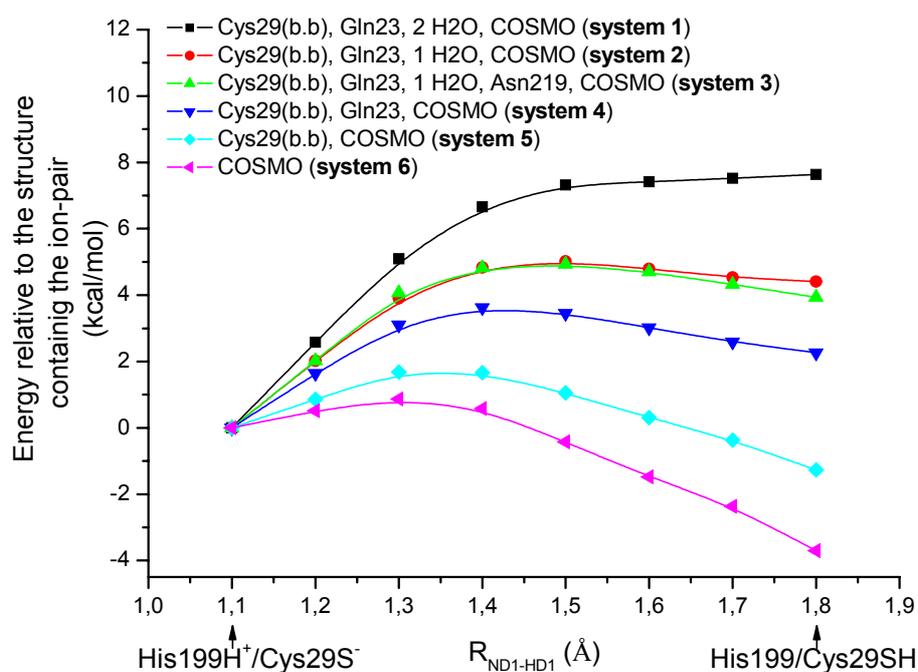


Figure 3.3.6. Reaction paths corresponding to the deprotonation of His199 residue in various systems calculated at the pure QM level. Energy is given relative to the ion-pair form of the active site ($His199H^+/Cys29S^-$).

Another source of stabilization of the ion-pair form of the active site discussed in the literature is the presence of the Asn219 residue.^{117,118} It interacts indirectly with the active site

by keeping the His199 residue in place using a hydrogen bond (Figure 3.3.5, bottom). The experimentally predicted stabilization energy is about 3.5 kcal/mol.¹¹⁸ In our QM calculations the influence of Asn219 is investigated by adding an acetamide, representing the side chain of Asn219, to system 2 of Table 3.3.1, yielding system 3 (Figure 3.3.5, bottom, green line in Figure 3.3.6). If systems 2 and 3 are compared (red and green line in Figure 3.3.6) one sees no difference in the reaction path and no additional stabilization of the zwitter-ion. Since our calculations take into account only stabilization due to the electronic interaction, the result indicates that the Asn219 plays an important role as an anchor point, directing the position of His199. In this way it participates in the creation of the H-bond network described above, but it plays no role in the direct electronic stabilization of the active site ion-pair.

In summary, by comparing the results obtained by QM computations the following conclusions can be drawn: 1) water seems to play an important role in the stabilization of the zwitterionic form of the active site, 2) this stabilization energy is estimated to be around 2-4 kcal/mol per water molecule, 3) the Gln23 residue contributes 2 kcal/mol to the ion-pair stabilization, and the H-bond from the backbone NH group of Cys29 at least 3 kcal/mol, and finally, 4) the Asn219 residue plays no direct role in the stabilization of the ion-pair, but serves as an anchor point for the His199, keeping it in the optimal position for interaction with the thiolate.

Pure QM calculations have thus shed light onto the importance of the various units comprising the H-bond network which stabilizes the ion-pair of the active site. They indicate that the existence of the L1 α -helix is not crucial for keeping the S-atom of Cys29 residue deprotonated – although the complete enzyme environment, as well as the explicit solvent shell, was not taken into account. To investigate the influence of these factors, and especially the role of the L1 helix, QM/MM calculations were performed containing the entire enzyme structure. These calculations started from the structure obtained in the 80ps QM/MM MD run. The size and the partition of the system remained the same as in QM/MM MD calculation, but the level of theory for the QM part was changed from a semiempirical method (AM1) to the more reliable RIDFT/B-LYP/TZVP level. As reaction coordinate we again adopted $R_{\text{ND1-HD1}}$, as in the QM calculations.

Table 3.3.2. Selected geometrical parameters and relative energies for points along the reaction paths for proton transfer between the His199 and Cys29 residues obtained in QM/MM calculations on various systems. All distances are given in Å and all relative energies in kcal/mol. For notation of the various distances see Figure 3.3.8. Figures in bold correspond to optimized structures.

entry	ΔE	$R_{\text{ND1-HD1}}$	$R_{\text{S-HD1}}$	$R_{\text{S-HE2}}$	$R_{\text{S-H(Cys29)}}$	$R_{\text{S-H(Ala200)}}$	$R_{\text{S-H}_2\text{O}}^1$	$R_{\text{S-H}_2\text{O}}^2$
System 7: 7 Å active atom range, 2 H₂O protonation of Cys29								
1	0	1.07	2.1	2.23	2.29	3.29	2.20	2.42
2	1	1.16	1.94	2.23	2.32	3.27	2.21	2.45
3	4	1.26	1.79	2.23	2.35	3.25	2.23	2.52
4	6	1.36	1.66	2.24	2.40	3.25	2.26	2.61
5	8	1.45	1.57	2.24	2.43	3.26	2.29	2.72
6	9	1.55	1.51	2.25	2.43	3.27	2.31	2.76
7	10	1.65	1.47	2.26	2.44	3.30	2.33	2.86
System 8: 12 Å active atom range, 2 H₂O protonation of Cys29								
8	0	1.07	2.07	2.21	2.34	3.27	2.17	2.81
9	1	1.16	1.93	2.22	2.36	3.26	2.19	2.88
10	3	1.26	1.76	2.23	2.40	3.25	2.20	3.00
11	6	1.36	1.64	2.24	2.43	3.26	2.22	3.04
12	7	1.45	1.57	2.25	2.45	3.27	2.24	3.12
13	7	1.55	1.51	2.25	2.47	3.29	2.26	3.38
14	8	1.65	1.47	2.23	2.48	3.33	2.27	3.30
System 9: 12 Å active atom range, 2 H₂O protonation of His199								
15	7	1.68	1.45	2.29	2.53	3.42	2.29	3.26
16	7	1.65	1.45	2.28	2.53	3.47	2.30	3.25
17	7	1.55	1.49	2.28	2.53	3.36	2.28	3.22
18	7	1.45	1.55	2.28	2.48	3.3	2.27	3.2
19	5	1.35	1.65	2.25	2.47	3.29	2.23	3.07
20	3	1.26	1.76	2.24	2.43	3.29	2.21	3.01
21	1	1.17	1.91	2.23	2.38	3.30	2.18	2.87
22	0	1.08	2.06	2.22	2.37	3.31	2.18	2.80
System 10: 12 Å active atom range, 1 H₂O, protonation of Cys29								
23	0	1.09	2.02	2.25	2.31	3.14	2.17	...
24	1	1.16	1.92	2.26	2.32	3.13	2.18	...
25	3	1.26	1.75	2.26	2.36	3.15	2.21	...
26	4	1.36	1.64	2.26	2.38	3.17	2.23	...
27	5	1.45	1.56	2.26	2.39	3.19	2.23	...
28	6	1.55	1.50	2.26	2.43	3.23	2.27	...
29	6	1.65	1.46	2.27	2.43	3.27	2.27	...
System 11: 12 Å active atom range, 1 H₂O protonation of His199								
30	4	1.73	1.43	2.26	2.54	3.37	3.07	...
31	4	1.65	1.45	2.25	2.55	3.34	3.09	...
32	4	1.55	1.50	2.27	2.49	3.26	2.60	...
33	3	1.45	1.56	2.27	2.45	3.20	2.50	...
34	3	1.35	1.64	2.26	2.43	3.17	2.43	...
35	2	1.26	1.75	2.25	2.4	3.15	2.41	...
36	1	1.17	1.87	2.26	2.37	3.13	2.35	...
37	0	1.10	1.99	2.25	2.33	3.1	2.29	...

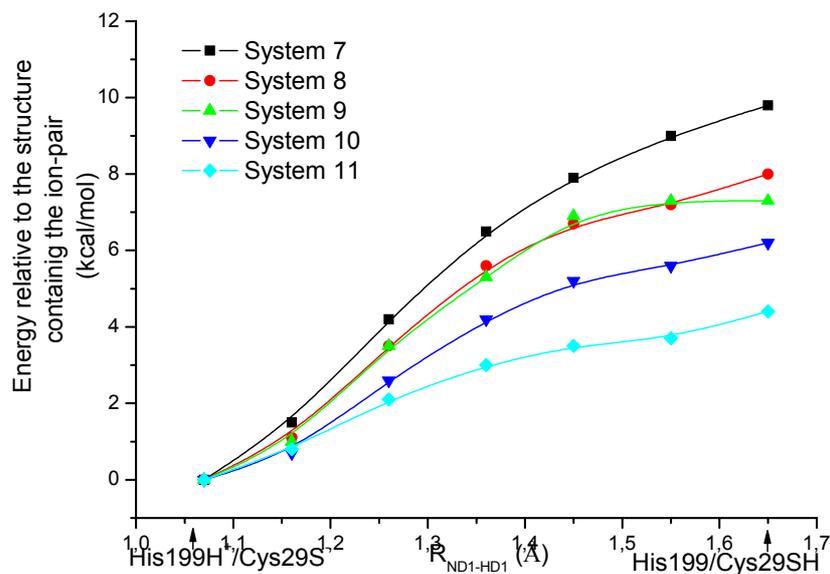


Figure 3.3.7. Reaction paths corresponding to the proton transfer from the His199 to the Cys29 residue in various systems calculated using the QM/MM method. Energy is given relative to the structure with the active site in an ion-pair form ($\text{His199H}^+/\text{Cys29S}^-$).

It has been discussed already for the pure QM calculations that the composition and the nuclear arrangement of the first solvent layer changes considerably if one goes from the zwitterionic structure ($\text{Cys29-S}^-\cdots\text{H}^+-\text{His199}$) to the neutral form ($\text{Cys29-SH}\cdots\text{His199}$). In the zwitterionic structure two water molecules form strong hydrogen bonds to the negatively charged sulphur center of the Cys29 residue. As expected, this hydrogen-bonding network is weakened on protonating the sulphur anion. As a consequence, one water molecule tends to break its hydrogen bond to the protonated sulphur. Only the H-bond between the second water and the Cys29 residue remains, albeit weaker than in the zwitterionic system. Since the water molecules are expected to behave in the same way at the QM/MM level, various model QM/MM calculations were performed to improve our understanding of the factors in play. The results are summarized in Table 3.3.2 and Figure 3.3.7.

In System 7 (Table 3.3.2) we restrained the second water molecule from moving away by restricting the range in which atoms were allowed to move to only 7 Å around the Cys29 residue. With this constraint in place, we found for System 7 that the zwitterionic form is the most stable. In subsequent calculations this restriction was lifted by extending the range in

which the atoms were allowed to move to 12 Å. In System 8 two water molecules (H_2O^1 and H_2O^2 in Figure 3.3.8) were incorporated in the hydrogen bonding network around the Cys29/His199 complex, while System 10 was manipulated in such a way that only one H-bridge is formed between water (H_2O^1 in Figure 3.3.8) and the sulphur of the active site. In this system the other water molecule was manually shifted away into the water bulk. For both System 8 and System 10 the proton migration from the His199 to the Cys29 residue was computed, i.e. the starting point for the calculation of the minimum energy path (MEP) was in both systems the zwitterionic form. Only the position of the water molecule forming the weaker bond to the sulphur center was modified as described above. These starting structures are depicted in Figure 3.3.8. The MEPs (Figure 3.3.7) were obtained by fixing the $R_{\text{ND1-HD1}}$ distance to certain values and optimizing all other geometrical parameters within the 12 Å range.

The stabilizing influence of water molecules on the zwitterionic form of the active site can be extrapolated from comparison of systems 7, 8 and 10 (Table 3.3.2, Figure 3.3.7, black, red, and blue line, respectively), since in all three systems all other stabilization factors remain almost unchanged. In System 7 two water molecules interact with the sulphur atom. One of them is bound via a strong H-bond (length around 2.3 Å) and the other via a somewhat weaker one (the length of which changes from 2.42 Å to 2.96 Å during the course of reaction). In this case no energy minimum could be found for the neutral active site form; the stabilization energy of the zwitterion is 10 kcal/mol, relative to the structure in which the ND1-HD1 distance (see Figure 3.3.8 for notation) is 1.65 Å (entry 7) (the distance corresponding to the optimal neutral structure). In System 8 (Figure 3.3.8, top) the first H-bond stays as strong as in System 7, but the second one becomes longer (changes from 2.81 Å to 3.26 Å) due to the relaxation of the neutral form of the active site. This decreases the energy by 2 kcal/mol (the energy difference between entry 7 and 14). In System 10 (Figure 3.3.8, bottom) the second H-bond disappears completely, while the first remains strong. The abolition of the H-bridge between the S^- and H_2O^2 destabilizes the ion-pair by an additional 2 kcal/mol (entry 29, Table 3.3.2).

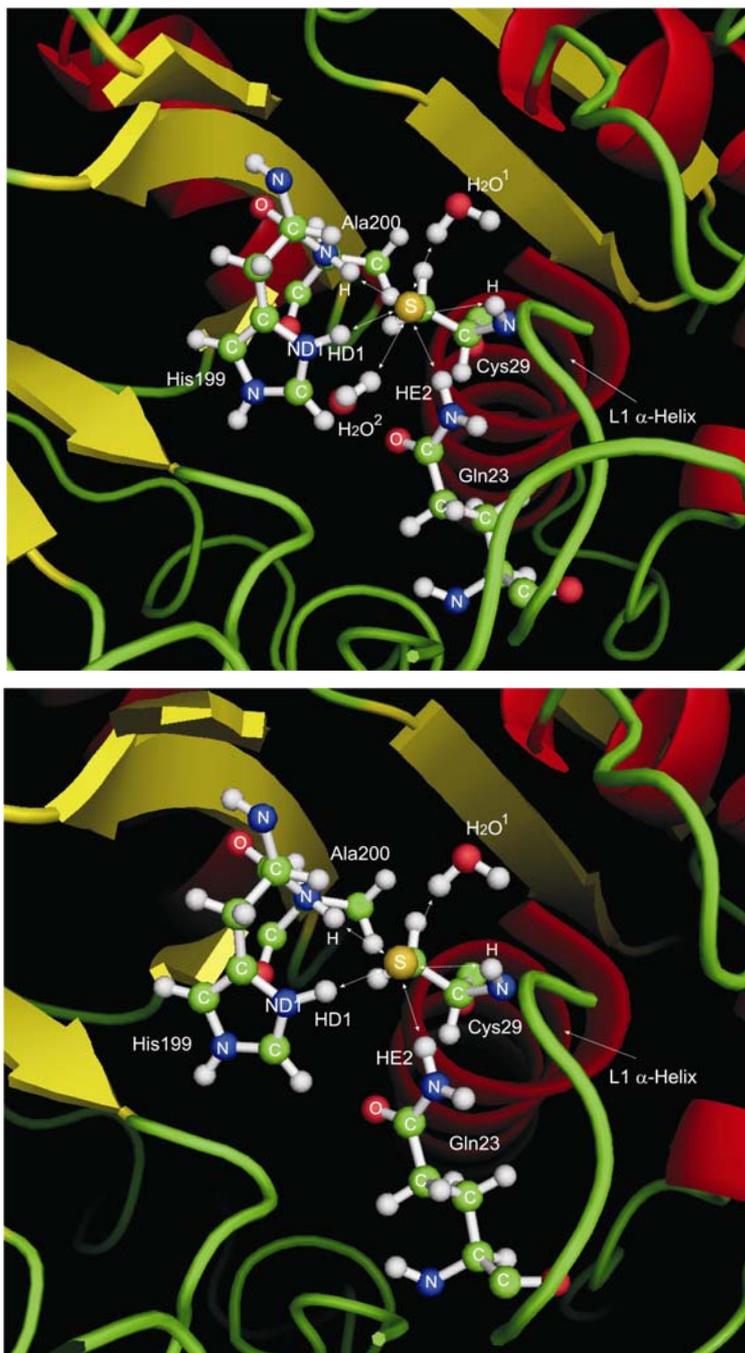


Figure 3.3.8. System 8 (top) and System 10 (bottom) calculated at the QM/MM level of theory. His199, Cys29, Gln23 and Ala200 residues are represented in ball-and-stick fashion. The additional water molecules H_2O^2 and H_2O^1 are also shown. In both top and bottom pictures the L1 α -helix is shown in red.

Since the water shell in Systems 8 and 10 was able to relax more than in 7 during the reaction course, we were able to obtain optimized neutral form structures (entries 15 and 30 in Table

3.3.2), by manually shifting the H_2O^1 molecule away from the protonated sulphur and letting the system relax. However, since these structures experienced large geometrical changes their energy should not be compared directly to the zwitterionic form. For this reason additional computations were made starting from these optimized neutral form structures and moving back along the reaction path to the ion-pair. In other words, the reaction path for protonation of His199 residue was performed in a stepwise manner, yielding System 9 (starting from the “product” of System 8, green line in Figure 3.3.7) and System 11 (starting from the “product” of System 10, light blue line in Figure 3.3.7) of Table 3.3.2. Corresponding manipulation of System 7 did not lead to the neutral form, but relaxed back to the zwitterionic structure.

For System 9 (two water molecules) the zwitterionic form is predicted to be 7 kcal/mol more stable than the neutral one. This is only 1 kcal/mol less stabilization than predicted for System 8. The difference in stabilization between Systems 11 and 10, however, is 3 kcal/mol. This is due to the larger freedom of movement of the H_2O^1 within the solvent shell in System 11 than in System 10, which is also reflected in the $R_{S-\text{H}_2\text{O}^1}$ values in Table 3.3.2. For System 11 the corresponding distance changes from 3.07 Å in the neutral form to 2.29 Å in the zwitterion state, whereas in System 10 this distance is reduced only by 0.1 Å. The change in reaction energy between Systems 8 and 9 is smaller while this distance discrepancy does not exist. In both cases both the distance between S^- and H_2O^1 and the distance between S^- and H_2O^2 change in the same manner. With this in mind System 9 (or System 8, as they differ little) seems to be the most accurate model, since the reorganization in the water shell is described properly, and, equally importantly, the second water molecule can move away from the sulphur center (Table 3.3.2, entry 15: $R_{S-\text{H}_2\text{O}^2} = 3.26$ Å) for the neutral state. Taking into account that System 7, which predicts a stabilization of 10 kcal/mol, is biased towards the zwitterionic form, while System 11, resulting in reaction energy of 4 kcal/mol, is more favorable to the neutral state, the correct value for the stabilization of the zwitterionic state would be around 7 kcal/mol, the one predicted by System 9.

The QM/MM calculations were performed to investigate the influence of the enzyme environment not directly involved in the hydrogen bond network, particularly the influence of the L1 α -helix. This can be achieved by comparing the results of the QM computation, which only took the stabilizing H-bond network into account, with the results obtained in the QM/MM calculations, which accounted for the whole enzyme. System 1 included the residues directly involved in the hydrogen bonding network (Cys29, His199, Gln23) and two

additional water molecules, while the rest of the enzyme was simulated by a polarizable medium. The whole enzyme was taken into account in System 9 which consequently also accounts for the macrodipole of the L1 helix. As in System 1, in System 9 two water molecules are involved in hydrogen bonding to the sulphur centre. For System 1 the zwitterionic form was computed to be 8 kcal/mol more favorable. The energy difference in System 9 is about 7 kcal/mol, indicating that the influence of the rest of the enzyme is about 1 kcal/mol. However, this is within the margin of error for the theoretical approaches. For Systems 2 and 11 (one water molecule) the computations predict reaction energies of 4 kcal/mol. This shows that in this case as well, the rest of the enzyme has virtually no effect on the calculations.

Combining the QM/MM MD results with the QM and QM/MM data leads to the following picture. The QM/MM MD simulations indicate that the zwitterionic form of the active site (Cys29-S⁻...H⁺-His199) is more stable than the neutral form (Cys29-SH...His199). Since the pK_a values of His199H⁺ (pK_a=6) and of Cys29-SH (pK_a=8-9) would lead one to expect a neutral state, the zwitterionic form must be stabilized by the interactions in the active site. The QM/MM MD computations show that the water molecules play a crucial role, since the proton migration from the His199 to the thiolate group of the Cys29 is always induced by a weak interaction between the S centre and the surrounding water network. In all simulations in which no proton transfer occurs a strong hydrogen bond network involving 2-4 water molecules exists along the whole trajectory.

The QM and QM/MM computations performed to quantify the influence of the various stabilizing effects also underline the importance of the water network. If both water molecules are taken into account the zwitterionic form is about 7-8 kcal/mol more stable than the neutral one. If only one water molecule is considered the energy difference shrinks to 4 kcal/mol. With no water molecules the difference lowers to 2 kcal/mol. The influence of the side chain of the Gln23 residue and the backbone NH group of the Cys29 residue was also evaluated. Both contribute 2-3 kcal/mol to the stabilization of the zwitterion. In addition, the sulphur center of Cys29 is also stabilized by hydrogen bonds to the backbone of the Ala200 and Trp30 residues. Together with the protonated His199 residue, these residues form a cage around the negatively charged S which can stabilize it regardless of the direction in which it moves. If this hydrogen bond network is replaced by a simple polar, and polarizable, environment, the zwitterionic form becomes about 4 kcal/mol less stable than the neutral one.

The picture of the cysteine protease active site that our calculations deliver can explain the high reactivity of the S⁻ centre towards the substrate despite its stabilization within the “empty” active site. Once the substrate enters the active site cleft it dislodges the water molecules that are situated inside. According to our model this would already destabilize the negatively charged sulphur (by 4-10 kcal/mol, corresponding to 2-4 water molecules). The S⁻ is also stabilized by hydrogen bond interactions to the Gln23 side chain and to the NH group of its own backbone, which form the so-called oxyanion hole. The oxyanion hole is an important anchor point for the substrate which forms strong H-bonds to it. Once the hydrogen bond interaction between the substrate and the enzyme is created (see previous chapters) the S⁻ is almost completely destabilized and ready to covalently bind to the substrate. The theory that the L1 helix macrodipole is responsible for the stabilization of the thiolate cannot explain the high reactivity of the sulphur towards the substrate warhead, since the presence of the substrate does not induce a big change in the macrodipole. X-ray data reveal that the geometrical arrangements of the active site with and without the substrate do not differ considerably.

Chapter 4 Summary

Cysteine proteases are widely distributed among living organisms, from viruses, bacteria, and parasitic protozoa to humans. They are directly involved in parasitic nutrition and invasion of host cells and hence play an important role in many diseases (malaria, cancer, “sleeping sickness”, leishmaniasis,...)² which makes them attractive targets for the development of new drugs.

One of the first cysteine protease inhibitors to exploit the electrophilicity of small heterocyclic rings was the epoxysuccinyl peptide E-64 (Scheme 2.1.4), isolated from *Aspergillus japonicus* in the late 70s.⁷ Since that time a vast number of natural structurally related inhibitors with similar inhibition properties have been discovered.²³⁻²⁵ Although all these compounds are

selective inhibitors of cysteine proteases of the papain superfamily, they show little or no selectivity towards other members of this large enzyme clan. This was the motivation for the development of synthetic epoxysuccinyl peptides. Indeed, syntactical epoxysuccinic acid derivatives have turned out to be promising drug candidates in the treatment of many diseases.⁸⁻¹⁴ Some have even reached clinical development as anti-cancer drugs.¹⁵ The inhibition mechanism of epoxysuccinyl peptides with the epoxide ring as an electrophilic trap led to development of peptides containing an aziridine ring (Scheme 2.1.5).

Although known about and investigated since the late 1970s, our picture of the basic principles governing the inhibition potencies and the structure-activity relationships of the cysteine protease inhibition mechanism is still very incomplete. In this work quantum mechanical, molecular dynamics and hybrid quantum mechanics/molecular modeling calculations were used to investigate the second step of the inhibition reaction, irreversible inactivation of the enzyme by the inhibitor. The power of computational approaches is the possibility to inspect the influence of a single, particular effect by selectively deactivating all others, something very difficult to achieve experimentally.

The quantum mechanical approach was used to study very small model systems consisting only of the electrophilic warhead of the inhibitor and molecules representing a very simplified protein active site. Solvent surroundings were two waters or two ammonium ions, in combination with continuum solvent model. Although simple, such a system provides a good description of the most important interactions affecting the inhibition reaction, and allows investigation of the influence of the properties of the electrophilic warhead on the reaction rate.

In chapter 3.1.2 substitution at the heteroatom of the aziridine ring was examined. Like epoxysuccinic acid derivatives, aziridine-2,3-dicarboxylic acid derivatives have been proven to be selective, non-toxic cysteine protease inhibitors suitable as antiparasitic¹⁶ and affinity labelling agents.¹⁷ Although both epoxide and aziridine heterocycles share a common principal inhibition mechanism, some differences exist. Aziridine ring opening is influenced by acidic media to a greater extent than epoxide ring opening, and unlike oxirane the aziridine ring allows additional substitution at the ring N atom. To investigate the influence of electron-withdrawing substituents on the character of the ring opening reaction and to provide a background for the rational design of new aziridine-based inhibitors, calculations were performed on various new model systems. Each of the potential N-aziridine substituents Cl,

Br, CF₃, CF₂H, and C₆H₄NO₂ was examined in either an acidic environment, simulated by including two ammonium ions (pK_s≈9.3) in the calculation, or in neutral solution, simulated by including two explicit water molecules (pK_s≈15), in addition to a polarizable solvent model.³⁹

The results showed that the inhibition reaction of aziridine-based inhibitors could be accelerated by the addition of electronegative N-substituents to the ring. In this way, inhibitors of the inhibition rate comparable to oxirane based inhibitors could be created. Another important advantage is the low pH dependence of the inhibition reaction kinetics and hydrolytic stability of these inhibitors. The activation energy in acidic surroundings is at most only 1 kcal/mol lower than in the corresponding neutral environment. Earlier studies on systems with CHO substituted at the N-position of the aziridine ring showed hydrolytic instability in DMSO, and furthermore, a possible side-reaction involving nucleophilic attack of the methylthiolate moiety at the carbonyl carbon of the substituent.⁵¹ It has been experimentally confirmed that the substituents treated in this work do not show this behaviour.⁵² An especially promising substituent is CF₂H, since it allows addition of further functionalities to the inhibitor. The hydrogen of the CF₂H group could be substituted in such a way as to further stabilize the non-covalent enzyme-inhibitor complex by interacting with one of the enzyme's subsites.

Chapter 3.1.1 deals with the possible protonation of the heteroatom of the oxirane and aziridine rings by the His199 residue at the quantum mechanical level of theory, as the basis for further QM/MM calculations.

The X-ray structure of the enzyme-inhibitor complexes reveals that the distance between the hydrogen of the protonated His199 residue and the inhibitor ring is too large for protonation to take place, as already reported by Varughese *et al.*^{19,45} However, since the active site lies on the surface of the enzyme the large distance between the histidine residue and the inhibitor could possibly be bridged by a water molecule which could then serve as a proton transmitter. Our QM calculations showed that the His199 can indeed protonate the inhibitor ring if the proton is transmitted via a water molecule. This indicates that His199 should be included in the QM part of QM/MM calculations; if left in the MM region, proton transfer could not be described.

Beside the properties of the electrophilic warhead, the protein and solvent environment are also very important for the irreversible deactivation of the enzyme active site by the inhibitor. The non-covalent interactions of the inhibitor with the oxyanion hole and other subsites of the enzyme as well as with the solvent molecules should be explicitly taken into account in the calculations, because of their possible impact on the reaction profile. As molecular modeling methods do allow the treatment of such large systems, but lack the possibility of describing changes in the bonding situation, our method of choice was the combined quantum mechanics/molecular modeling approach. By splitting the system into a smaller part in which bond cleavage and formation occurs (treated at the QM level), and a larger part including the rest of the protein (treated with MM), we managed to simulate the system with the desired precision.

Investigations into the role of His199 in the inhibition mechanism are presented in chapter 3.2.1. Pure quantum mechanical investigation had predicted that the His199 residue of the active site can serve as a proton donor in the inhibition reaction (Section 3.1.2). To examine the influence of the explicit protein and solvent environment on this process we turned to the QM/MM approach. These computations again make clear that the protonation of epoxide or aziridine based inhibitors by the protonated histidine residue can only take place if the gap between the HisH⁺199 and the inhibitor is bridged by a water molecule. They also show that in both cases it takes only one water molecule to establish a very efficient relay system in which the proton can easily move from the His199 to the heteroatom of the inhibitor. HisH⁺199 can thus act as the proton source in the inhibition mechanism, provided that there are any water molecules deep enough in the active site of the protein to establish the required relay system. That this is the case has been shown by X-ray experiments which detected the presence of such water molecules in the binding pocket of the cathepsin B–E64c complex. HisH⁺199 is a more efficient proton source than water, due to its lower pK_a value (6.8 vs. 15). For this reason, water serves only as a bridge, but not as the proton source itself.

The importance of inhibitor protonation can be estimated from the computed energies. These indicate that without protonation the aziridine-based inhibitors would not be active at all, since the reaction barrier is much too high and the reaction is predicted to be endothermic. However, in systems in which the surroundings are acidic enough to allow the protonation of the aziridine TS, inhibition by aziridines should happen as fast as with oxirane based inhibitors. The kinetics of the epoxide reaction are less influenced by protonation. Nevertheless, calculations show that for such inhibitors the protonation of the alcoholate

group of the opened ring is the main reason for the irreversibility of the inhibition reaction. For both inhibitor types the HisH⁺199-H₂O-inhibitor relay system must be considered important over a wide range of pH values due to the proximity of the proton source (HisH⁺) and its high efficiency in the proton transfer.

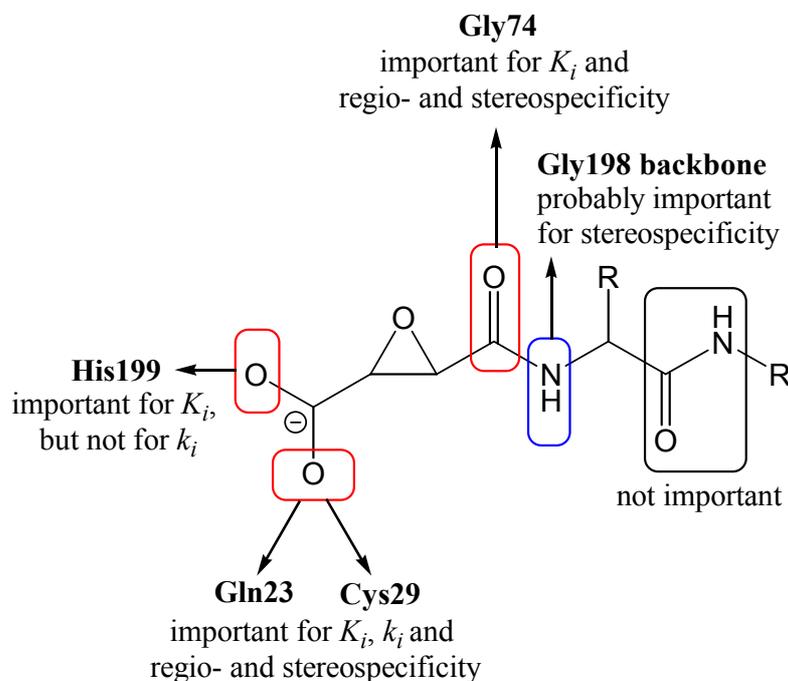


Figure 4.1.1. Diagram showing the various interactions determining the inhibition potency (K_i , k_i) as well as the regio- and stereospecificity of epoxide-based inhibitors.

In chapters 3.2.2-3.2.4 the structure-reactivity relationships between cysteine protease and various inhibitors are presented and the kinetics, regio- and stereospecificity of the inhibition are examined. The computations enabled careful analysis of the various interactions determining the mode of action of epoxide-based inhibitors. Examination of the behavior of various inhibitor-enzyme models even allowed differentiation between the interactions influencing the two different steps of the inhibition mechanism. Additional MD simulations indicate that the hydrogen bonds between the enzyme and the epoxysuccinic base of the inhibitor (in red in Figure 4.1.1) formed in the reversible step of the inhibition reaction are responsible for the regio- and stereospecificity of the reaction by determining the position of C2 in respect to S⁻ (Figures 3.2.17 and 3.2.26). It is shown that the change in C2-S distance on going from reactants to transition states correlates with the barrier height of the inhibition reaction (Figure 3.2.10). The computations allowed quantification of the interactions between

inhibitor and protease, and hence ranking of the importance of the various contacts. The resulting picture of the importance of the single interactions for the potency as well as the regio- and stereospecificity of such inhibitors is summarized in Figure 4.1.1. They indicate that even simple epoxide-2-carboxamide- and epoxide-2-carbonyl-based compounds should meet the minimal structural requirement for high inhibition potency of cathepsin-like cysteine proteases. Computations performed for corresponding aziridine-based inhibitors showed similar substituent effects and showed the differences in the pH-dependency of the inhibitor activities to be due to the chemical properties of the heterocycles. Cooperative effects between the substituents and the heterocycles, which are very important for the reactions in solution, were found to be small in the enzyme environment.

The results also indicate that the general assumption of a two step mechanism for the inhibition process does not hold in all cases. Instead, due to the strong hydrogen bond network within the active site, there is a smooth transition between both steps. As a consequence the first-order rate constant of inhibition, k_i , is affected not only by the irreversible ring-opening reaction but also by the reversible network interactions. In the generally accepted two-step model these interactions are subsumed into the dissociation constant of the non-covalent enzyme inhibitor complex, K_i . If K_i is increased by improving the hydrogen-bonding network, k_i increases as well. In other words, a clear differentiation between k_i and K_i seems not to be possible, and the second order-rate constant of inhibition, k_{2nd} , which includes the effects of both reversible and irreversible interactions seems to be more suitable for characterization of this class of irreversible inhibitors.

To account also for dynamical effects classical molecular dynamics (MD) computations were done. The results of these calculations not only supported and clarified the QM/MM results, the comparison with previous X-ray structures helped correct existing errors in the available geometrical models and resolved inconsistencies in the weighting of various factors governing the inhibition (chapter 3.2.5).

At the end in chapter 3.3.2 the first QM/MM MD calculations on the active site of the cysteine proteases were presented. In contrast to the MD simulations mentioned above, in these simulations the potential energies were computed at the QM/MM-level. In these investigations we tried to answer strongly disputed questions about the reasons for the existence of the active site ion pair and its role in the enzyme's high activity. Kinetic studies on different thiol proteases showed that there are two ionisable groups involved in the

catalysis, one with a pK_a value of around 4 and another with pK_a around 8.5. The former is usually assigned to deprotonation of the Cys29 residue (cathepsin B numbering), whereas the latter is believed to reflect the ionization of His199. However, the pK_a values of cysteine and histidine residues in proteins are generally about 8–9 and 6, respectively. These unusual pK_a values can only be explained by considering the local enzyme environment of both residues in detail. The QM/MM MD simulations indicate that the zwitterionic form of the active site (Cys29-S⁻...H⁺-His199) is more stable than the neutral form (Cys29-SH...His199) and that water molecules in the active site play a crucial role in its stabilization. Proton migration from the His199 to the thiolate group of the Cys29 seems to happen only when interactions between the S centre and the surrounding water network are weak (Figure 3.3.2). In all simulations in which no proton transfer occurs, a strong hydrogen bond network involving 2-4 water molecules could be observed.

The additional QM and QM/MM computations performed to quantify the influence of the various stabilizing effects on the ion pair also underline the importance of the water network. The ion-pair stabilization stemming from this hydrogen bond interaction is 2-4 kcal/mol per water molecule. In addition to the water molecules, the influence of other residues was also evaluated. The hydrogen bond with the Gln23 side chain contributes 2 kcal/mol to the stabilization of the zwitterion, while the interaction with the Cys29 backbone brings in at least 3 kcal/mol. In addition, the sulphur center of Cys29 is also stabilized by hydrogen bonds to the backbone of Ala200 and Trp30 residues. Together with the protonated His199 residue, these four residues form a cage around the negatively charged S and stabilize it independently of the direction in which it moves.

The picture built up by these computations can explain the high reactivity of the S⁻ centre towards the substrate despite its stabilization within the empty active site. When the substrate enters the active site cleft it dislodges the water molecules that are situated in there and binds to the His199 and Gln23 side chains and to the Cys29 backbone amide group via hydrogen bonds. This leads to the almost complete demolition of the hydrogen bond network that stabilizes the S⁻, leaving it significantly destabilized and ready to covalently bind to the substrate.

Chapter 5 Zusammenfassung

Cysteinproteasen sind in lebenden Organismen weit verbreitet. Dies reicht von Viren, Bakterien und parasitären Protozoen bis hin zum menschlichen Organismus. Sie sind direkt bei der parasitären Ernährung und beim Eindringen in die Wirtszelle beteiligt und spielen daher bei vielen Krankheiten (z.B. Malaria, Krebs, Schlafkrankheit, Leishmaniose etc.)² eine wichtige Rolle, was sie zu attraktiven Angriffspunkten in der Entwicklung neuer Medikamente macht.

Einer der ersten Cysteinprotease-Inhibitoren, welcher die Elektrophilie kleiner heterozyklischer Ringe ausnutzte, war das Epoxisuccinylpeptid E-64 (Schema 2.4.1).⁷ Es wurde erstmals in den späten 70er Jahren aus *Aspergillus japonicus* isoliert und seitdem

wurde eine große Anzahl von natürlichen, strukturell eng verwandten Inhibitoren mit ähnlichen Inhibierungseigenschaften entdeckt.²³⁻²⁵ Obwohl alle diese Verbindungen selektive Inhibitoren von Cysteinproteasen der Papain-Überfamilie darstellen, zeigen sie nur wenig bis keine Selektivität gegenüber einzelner Mitglieder dieses großen Enzymstammes. Aufgrund dessen wurde versucht, synthetische Epoxy Succinylpeptide zu entwerfen. Synthetische Epoxybernsteinsäurederivate scheinen tatsächlich vielversprechende Kandidaten für Medikamente bei vielen Krankheiten zu sein.⁸⁻¹⁴ Einige hiervon haben bereits die klinische Entwicklungsphase als Anti-Krebs-Mittel erreicht.¹⁵ Der Hemmmechanismus von Epoxy Succinylpeptiden mit einem Epoxidring als elektrophilen Angriffspunkt führte zu der Entwicklung von Peptidstrukturen, welchen einen Aziridinring beinhalten (Schema 2.1.5).

Obwohl bereits seit den späten 70ern bekannt und untersucht, ist das Bild über die grundsätzlichen Prinzipien, welche die Wirksamkeit der Inhibition und die Struktur-Aktivitäts-Beziehungen (SAB) der Cysteinprotease-Hemmmechanismen beeinflussen, immer noch sehr unvollständig. In dieser Arbeit wurden quantenmechanische (QM), molekulardynamische (MD) und gemischte quantenmechanische/molekularmechanische (QM/MM) Berechnungen durchgeführt, um den zweiten Schritt der Inhibierungsreaktion von Cysteinproteasen, d.h. die irreversible Inaktivierung des Enzyms durch den Inhibitor, eingehender zu untersuchen. Die Stärke von computergestützten Verfahren liegt in der Möglichkeit, den Einfluss einzelner spezifischer Effekte durch das Ausblenden von Umgebungseffekten zu untersuchen. Diese Herangehensweise ist nur sehr schwer im Experiment zu erreichen. Darüber hinaus konnten basierend auf theoretischen Modellen neue, verbesserte Arten von Inhibitoren entwickelt werden.

Die quantenmechanische Methode wurde benutzt, um sehr kleine Modellsysteme zu untersuchen, welche lediglich aus dem elektrophilen Kerngerüst des Inhibitors und einigen, das aktive Zentrum im Protein vereinfachend darstellenden, Molekülen bestanden. Die Solvensumgebung wurde durch zwei explizite Wassermoleküle und zwei Ammoniumionen in Kombination mit einem Kontinuum-Solvensmodell berücksichtigt. Ein solches vereinfachtes System liefert bereits eine gute Beschreibung der meisten wichtigen Wechselwirkungen während der Inhibierungsreaktion. Es erlaubt die Untersuchung des Einflusses der Eigenschaften des elektrophilen „Warheads“ auf den Reaktionsverlauf.

In Kapitel 3.1.2 wurde die Substitution am Heteroatom des Aziridinrings untersucht. Wie bei den Epoxybernsteinsäurederivaten konnte auch bei Aziridin-2,3-dicarbonsäurederivaten eine selektive, nicht-toxische Inhibition von Cysteinproteasen gezeigt werden, wodurch sie als antiparasitäre¹⁶ und affinitätsmarkierende¹⁷ Wirkstoffe Verwendung finden. Obgleich sowohl Epoxid- und Aziridinheterozyklen prinzipiell den gleichen Inhibitionsmechanismus aufweisen, können doch einige Unterschiede gefunden werden. Die Öffnung des Aziridinrings wird von einem saurem Medium im größeren Maße beeinflusst als die Ringöffnungsreaktion des Epoxids und im Gegensatz zu Oxiranen ist bei Aziridinen eine zusätzliche Substitution am Ring-Stickstoffatom möglich. Um den Einfluss von elektronenziehenden Substituenten auf den Charakter der Ringöffnungsreaktion zu untersuchen und um Hintergrundwissen für ein rationales Design neuer Aziridin-basierter Inhibitoren zu gewinnen, wurden Berechnungen für verschiedene neuartige Modellsysteme durchgeführt. Hierbei wurden als N-Aziridin-Substituenten Cl, Br, CF₃, CF₂H und C₆H₄NO₂ verwendet. Für jedes dieser Substituenten wurde der Einfluß von polaren Umgebungen durch die Zugabe von zwei expliziten Wassermolekülen (pK_s=15) oder zwei expliziten Ammoniumionen (pK_s=9.3) simuliert, welche entweder eine neutrale oder saure Umgebung berücksichtigen.

Diese Ergebnisse zeigten, dass die Beschleunigung der Inhibierungsreaktion von Aziridin-basierten Inhibitoren durch das Anfügen von elektronegativen N-Substituenten am Ring erreicht werden könnte. Es könnten so Inhibitoren mit vergleichbar guter Inhibierungsrate wie die der Oxiran-basierten Inhibitoren geschaffen werden. Weitere wichtige Aspekte sind eine geringe pH-Abhängigkeit der Kinetik der Inhibierungsreaktion solcher Inhibitoren und deren hydrolytische Stabilität. Die Aktivierungsenergie in basischer Umgebung ist meist 1 kcal/mol geringer als in neutralem Medium. Frühere Studien von N-acylierten Aziridinringen zeigten die Instabilität in DMSO gegenüber Hydrolyse und zusätzlich einen nukleophilen Angriff des Methylthiolatrests am Carbonylkohlenstoff des Substituenten.⁵¹ Die Substituenten, wie sie in dieser Arbeit beschrieben werden, zeigen dieses Verhalten nicht, wie tatsächlich experimentell bestätigt werden konnte. Ein besonders viel versprechender Substituent ist CF₂H, da es eine weitergehende Substitution des Ringes erlaubt.⁵² Das Wasserstoffatom der CF₂H Gruppe könnte derart ersetzt werden, dass die Stabilisierung des nicht-kovalent gebundenen Enzym-Inhibitor-Komplexes durch eine zusätzliche Wechselwirkung mit den Nebenbindungsstellen des Enzyms erreicht werden würde.

Das Kapitel 3.1.1 befasst sich mit der möglichen Protonierung des Heteroatoms des Oxiran- und Aziridinrings durch den His199 Rest auf einem quantenmechanischen Theorieniveau, um als Ausgangspunkt für weitere QM/MM Berechnungen zu dienen.

Wie bereits von Varughese *et al.*^{19,45} berichtet zeigt die Röntgenstruktur des Enzym-Inhibitor-Komplexes, dass der Abstand zwischen dem Wasserstoff des protonierten His199 und dem Ring des Inhibitors zu groß ist, damit eine Protonierung stattfinden kann. Da jedoch das aktive Zentrum auf der Oberfläche des Enzyms liegt, könnte der lange Abstand zwischen dem Histidinrest und dem Inhibitor durch ein Wassermolekül überbrückt werden, welches als Protonenüberträger fungieren könnte. Unsere QM Berechnungen zeigten, dass das His199 tatsächlich den Inhibitorring zu protonieren vermag, wenn das Proton über ein Wassermolekül übertragen wird. Dies zeigt, dass His199 in den QM Teil der QM/MM Berechnungen aufgenommen werden sollte. Verbliebe es im MM Teil, so könnte der Protonentransfer nicht beschrieben werden.

Neben den Eigenschaften des elektrophilen „Warheads“ ist sowohl die Protein- als auch die Solvensumgebung von großer Bedeutung für die irreversible Deaktivierung des aktiven Zentrums des Enzyms durch den Inhibitor. Aufgrund eines möglichen Einflusses auf das Reaktionsprofil müssen Solvensmoleküle sowie die nicht-kovalenten Wechselwirkungen des Inhibitors mit dem Oxyanionloch und anderen Nebenbindungsstellen des Enzyms explizit behandelt werden. Berücksichtigt man, dass nur molekularmechanische Methoden eine Behandlung von solch großen Systemen erlauben, im Gegenzug aber nicht in der Lage sind, kovalente Wechselwirkungen zu beschreiben, so wurde als Methode der Wahl ein kombinierter QM/MM Ansatz gewählt. Durch das Aufteilen des Gesamtsystems in einen kleinen Bereich, der die Bindungsspaltung- und Bindungsbildungsreaktionen beinhaltet (mit QM behandelt), und in einen großen Teil, welcher den Rest des Proteins umfasst (mit MM behandelt), waren wir in der Lage, das System mit gewünschter Genauigkeit zu simulieren.

Die Untersuchungen, welche sich mit der Rolle des His199 im Inhibierungsmechanismus beschäftigen, sind in Kapitel 3.2.1. angeführt. Reine quantenmechanische Untersuchungen sagen vorher, dass der His199 Rest des aktiven Zentrums die Rolle des Protonendonators in der Inhibierungsreaktion übernimmt (Sektion 3.1.2). Um den Einfluss der expliziten Behandlung der Protein- und Solvensumgebung auf dieses Phänomen zu untersuchen wurden QM/MM Verfahren eingesetzt. Diese Berechnungen unterlegen die bisherige Vermutung, dass die Protonierung von Epoxid- oder Aziridin-basierten Inhibitoren durch den protonierten

Histidinrest nur dann stattfinden kann, wenn der Abstand zwischen dem HisH⁺199 und dem Inhibitor durch ein Wassermolekül überbrückt wird. Es kann ebenfalls gezeigt werden, dass in beiden Fällen jeweils bereits ein Wassermolekül ausreicht, um eine sehr effiziente kaskadierte Protonenübertragungsreaktion vom His199 zum Heteroatom des Inhibitors zu bewirken. Das HisH⁺199 kann somit als Protonenquelle im Inhibitionsmechanismus dienen, jedoch nur unter der Voraussetzung, dass überhaupt Wassermoleküle tief genug in das aktive Zentrum eindringen können, um die erforderliche Protonenübertragungskaskade auslösen zu können. Diese Vermutung wird durch Röntgenbeugungsexperimente gestützt, welche die Anwesenheit solcher Wassermoleküle in der Bindungstasche des Cathepsin B-E64c Komplexes zeigen. Durch den Vergleich der pKa Werte lässt sich zeigen, dass HisH⁺199 tatsächlich eine effizientere Protonenquelle als Wasser darstellt (6.8 vs. 15). Somit fungiert Wasser nur als Protonenbrücke, jedoch nicht als Protonenquelle selbst.

Die Wichtigkeit der Protonierung des Inhibitors kann anhand der berechneten Energien abgeschätzt werden. Es kann gezeigt werden, dass ohne eine Protonierung die Aziridin-basierten Inhibitoren überhaupt nicht aktiv wären, da die Reaktionsbarriere deutlich zu hoch läge und eine mögliche Reaktion endotherm verlaufen würde. In Systemen mit hinreichend azider Umgebung, welche eine Protonierung des Aziridin-Übergangszustands erlauben würde, sollte die Inhibierung durch Aziridine ähnlich schnell ablaufen wie bei Oxiran-basierten Inhibitoren. Die Kinetik der Epoxidreaktion wird durch die Protonierung weniger beeinflusst. Nichtsdestotrotz zeigen die Berechnungen, dass für solche Inhibitoren die Protonierung der Alkoholatgruppe des geöffneten Rings der Hauptgrund für die Irreversibilität der Inhibierungsreaktion ist. Die HisH⁺199-H₂O-Inhibitor-Kaskade ist sehr wichtig für beide Inhibitortypen für einen großen pH-Bereich, was auf die räumliche Nähe der Protonenquelle (HisH⁺) und ihre hohe Effizienz beim Protonentransfer zurückzuführen ist.

In den Kapiteln 3.2.2 bis 3.2.4 werden die Struktur-Reaktivitäts-Beziehungen (SRB) zwischen der Cysteinprotease und verschiedenen Inhibitoren vorgestellt. Es wird gezeigt, dass die SRB eine entscheidende Rolle in Kinetik, Regio- und Stereoselektivität der Inhibierung spielen. Die Untersuchungen haben nicht nur eine genaue Analyse verschiedener Wechselwirkungen ermöglicht, die den Charakter der Wirkung der Epoxidhaltenden Inhibitoren bestimmen, sondern auch die Unterscheidung zwischen den Wechselwirkungen, die einzelne Schritte der Inhibierungsreaktion beeinflussen. Zusätzliche MD Simulationen deuten an, dass die

Wasserstoffbindungen zwischen dem Enzym und dem epoxisuccinischen Basis des Inhibitors (Abb. 5.1.1, im Rot), die in dem reversiblen Schritt der Inhibition entstehen, für die Regio- und Stereospezifität der Reaktion verantwortlich sind. Diese Wasserstoffbrücken bestimmen die Position des C2 Atoms im Bezug zu S⁻ und dadurch die erwähnte Eigenschaften der Inhibierung (Abb. 3.2.17 und 3.2.26). Es wird gezeigt, dass die Änderung des C2-S Abstands auf dem Weg von Edukten zu Übergangszuständen mit der Energiebarriere der Inhibitionsreaktion korreliert (Abb. 3.2.10). Die Rechnungen erlaubten die Quantitätsbestimmung der Wechselwirkungen zwischen dem Inhibitor und der Protease, und so die Bedeutung verschiedener Kontakte (Abb. 5.1.1). Das Abb. 5.1.1 deutet an, dass schon die epoxide-2-carboxamid- und epoxide-2-carbonylbasierte Moleküle die minimalen Strukturanforderungen für eine hohe Inhibierungspotenz in cathepsinähnlichen Cysteineproteasen erfüllen sollten. Die Untersuchungen entsprechender aziridinbasierten Inhibitoren zeigen ähnliche Substituenteneffekte wie für epoxydenthaltenden Inhibitoren, und geringere Unterschiede in der pH-Abhängigkeit der Wirkung von den chemischen Eigenschaften der zwei Heterozyklen. Kooperativeneffekte zwischen den Substituenten und den Heterozyklen, die die Reaktion in der Lösung bestimmen, zeigten sich als geringe.

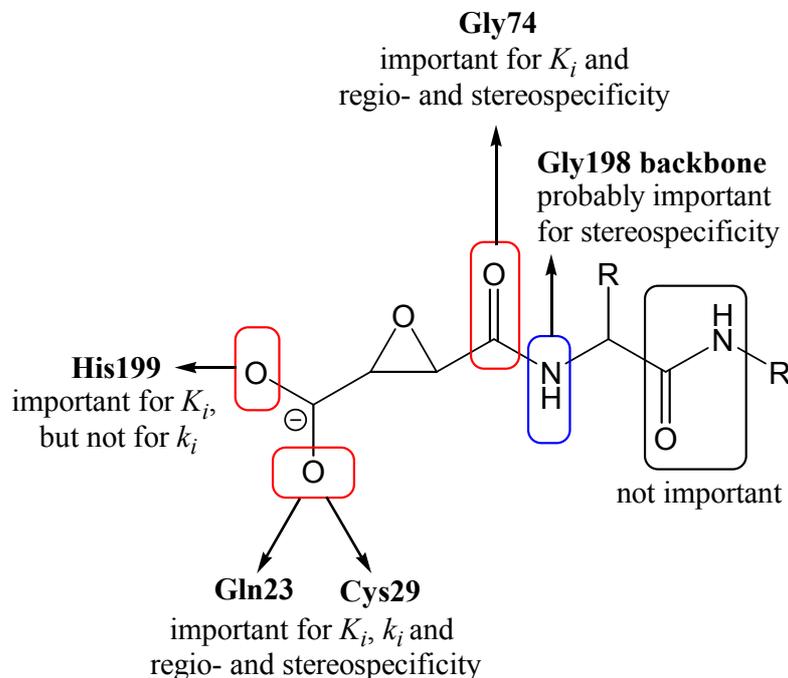


Abb. 5.1.1. Bedeutung verschiedener Wechselwirkungen für Inhibierungspotenz (K_i , k_i) und Regio- und Stereoselektivität der Inhibierungsreaktion der epoxydbasierten Inhibitoren.

Die Ergebnisse deuten auch an, dass die allgemeine Meinung, dass der Inhibitionsprozess ein zwei-stufiges Mechanismus ist, nicht in allen Fällen hält. Statt dessen zeigen sie, dass ein milder Übergang vom reversiblen zum irreversiblen Schritt statt findet als Ergebnis der starken Wasserstoffbindung-Netzwerks in der Bindungstasche. Folglich ist die Inhibitionsrate erster Ordnung, k_i nicht nur von der irreversiblen Ring-Öffnung Reaktion beeinflusst, sondern auch von den reversiblen Netzwerk-Wechselwirkungen. Im allgemein angenommenen Zwei-Schritt Model sind diese Wechselwirkungen unter der Dissoziationskonstante des nicht-kovalenten Enzym-Inhibitor Komplexes, K_i , zusammengefasst. Wenn K_i durch das Wasserstoffbindung-Netzwerk aufgebessert wird, nimmt k_i begleitend zu. Mit anderen Worten, eine absolute Unterscheidung zwischen K_i und k_i scheint nicht möglich zu sein und die Inhibitionsrate zweiter Ordnung, k_{2nd} , die sowohl die Effekte der reversible als auch irreversible Reaktion beinhaltet, scheint besser für die Beschreibung dieser Klasse irreversibler Inhibitoren geeignet zu sein.

Um auch die dynamischen Effekte im Anspruch zu nehmen, wurden klassische molekulardynamische (MD) Simulationen durchgeführt. Die Ergebnisse dieser Berechnungen haben nicht nur unterstützt und die QM/MM Ergebnisse näher erklärt, sondern auch durch der Vergleich mit bereits bestehenden Röntgenstrukturen verschiedener Cysteineprotease-Inhibitor-Komplexen geholfen, Fehler in vorhandenen Proteinkristallstrukturen zu korrigieren. Somit wurden Widersprüche in der Bewertung verschiedener Funktionen, die die Inhibierungsreaktion bestimmen, aufgelöst (Kapitel 3.2.5).

Im Kapitel 3.3.2 werden die ersten QM/MM MD Berechnungen vom aktiven Zentrum der Cysteineprotease vorgestellt. Im Kontrast zu den klassischen MD Simulationen oben, wird in diesen Untersuchungen die potentielle Energie auf dem QM/MM Theorieniveau berechnet. Hier haben wir versucht die stark umstrittene Frage über die Gründe für das Vorhandensein des aktiven Zentrums als His199⁺/Cys29⁻ Ion-paar und seine Rolle für die hohe Aktivität des Enzyms zu beantworten. Kinetische Studien von verschiedenen Thiol-Proteasen haben gezeigt, dass es zwei ionisierbare Gruppen gibt, die in die Katalyse verwickelt sind, eine mit dem pK_a Wert um 4 und die Andere mit pK_a von ungefähr 8,5. Die Erste wird üblicherweise der Deprotonierung des Cys29 zugeschrieben und die Zweite zur Ionisierung des His199. Dahingegen ist in der Proteinumgebung der pK_a Wert des Cysteins normalerweise 8-9 und der des Histidinrests bei 6. Diese ungewöhnlichen pK_a Werte können nur unter genauer

Berücksichtigung der lokale Enzymumgebung der zwei Reste erklärt werden. Die QM/MM MD Rechnungen deuten darauf hin, dass die Bindungstasche in Form eines Zwitterions (Cys29-S⁻···H⁺-His199) stabiler ist als in ihrer neutralen Form (Cys29-SH···His199), und, dass die Wassermoleküle eine entscheidende Rolle zu ihrer Stabilisierung spielen. Die Wanderung des Protons von His199 zum Thiolat des Cysteinrestes wird immer durch eine nachlassende Wechselwirkung zwischen dem Schwefel und den umgebenden Wassermolekülen verursacht (Abb. 3.3.2). In allen Fällen, in den keine Protonwanderung stattfindet, ist ein starkes Wasserstoffbindung-Netzwerk, zusammengesetzt von 2 zu 4 Wassermolekülen, vorhanden.

Die zusätzlichen QM und QM/MM Rechnungen die ausgeführt wurden um den Einfluss verschiedener Stabilisierungseffekte quantitativ zu bestimmen, unterstreichen ebenfalls die Wichtigkeit des Wasser-Netzwerks. Die Stabilisierung des Ion-Paares, die von der Wasserstoffbindungen stammt, beträgt 2-4 kcal/mol pro Wassermolekül. Außer dem Einfluss der Wassermoleküle, wurde auch der Einfluss anderer Reste ausgewertet. Die Wasserstoffbindung zur Gln23 Seitenkette trägt 2 kcal/mol zu der Stabilisierung des Zwitterions bei, während die Wechselwirkung mit dem Backbone des Cys29 mindestens 3 kcal/mol beisteuert. Der Schwefel des Cys29 Restes ist auch durch Wasserstoffbindungen zu den Backbones der Ala200 and Trp30 Reste zusätzlich stabilisiert. Die erwähnten vier Reste, zusammen mit dem protonierten His199, bilden einen Käfig um den negativgeladenen Schwefel und stabilisieren ihn unabhängig von der Richtung, in der er sich bewegt.

Das Model, das aus unseren Rechnungen entsteht, kann die hohe Reaktivität des S⁻ Zentrums gegenüber dem Substrat, trotz seiner Stabilisierung in der „leeren“ Bindungstasche, erklären. Wenn das Substrat in die Spalte der Bindungstasche eintritt, verdrängt es die sich dort befindenden Wassermoleküle und bindet durch Wasserstoffbindungen sowohl zu den His199 und Gln23 Seitenketten, als auch zu dem Cys29 Backbone. Auf diese Weise wird das, den negativen Schwefel stabilisierenden, Wasserstoffbindung-Netzwerk fast vollständig aufgelöst und der instabile Schwefelrest S⁻ zur kovalenten Substratbindung vorbereitet.

Chapter 6 References

¹ H.-H. Otto, T. Schirmeister, *Chem. Rev.*, 1997, **97**, 133.

² F. Lecaille, J. Kaleta, D. Brömme, *Chem. Rev.*, 2002, **102**, 4459.

³ G. Huet, R. M. Flip, C. Richet, C. Thiebet, D. Demeyer, M. Balduyck, B. Duquesnoy, P. Degand, *Clin. Chem.*, 1992, **38**, 1694.

⁴ B. F. Sloane, K. Moin, E. Krepela, J. Rozhin, *Cancer Metastasis Rev.*, 1990, **9**, 333.

⁵ P. J. Rosenthal, J. H. McKerrow, M. Aikawa, H. Nagasawa, J. J. Leech, A. Malarial, *J. Clin. Invest.*, 1988, **82**, 1560.

-
- ⁶ For some reviews see: J. C. Powers, J. L. Asgian, O. D. Ekici, K. E. James, *Chem. Rev.*, 2002, **102**, 4639-4750; T. Schirmeister, A. Klockow, *Mini Rev. Med. Chem.*, **2003**, *3*, *6*, 585-596.
- ⁷ K. Hanada, M. Tamai, S. Morimoto, T. Adachi, S. Ohmura, J. Sawada, I. Tanaka, *Agric. Biol. Chem.*, 1978, **42**, 537.
- ⁸ T. S. Kim, A. S. Tasker, *Curr. Top. Med. Chem.*, 2006, **6**, 355.
- ⁹ R. Roberts, *Drug News Perspect.*, 2005, **18**, 605.
- ¹⁰ V. Gocheva, J. A. Joyce, *Cell Cycle*, 2007, **6**, 60.
- ¹¹ V. Y. Ook, *BioDrugs*, 2006, **20(2)**, 105.
- ¹² P. J. Rosenthal, *Int. J. Parasitol.*, 2004, **34**, 1489.
- ¹³ C. R. Caffrey, S. Scory, D. Steverding, *Curr. Drug Targets*, 2000 **1**, 155.
- ¹⁴ J. Jose Cazzulo, V. Stoka, V. Turk, *Curr. Pharm. Des.*, 2001, **7**, 1143.
- ¹⁵ G. F. Eilon, J. Gu, L. M. Slater, K. Hara, J. W. Jacobs, *Cancer Chemother. Pharmacol.*, 2000, **45**, 183.
- ¹⁶ R. Vicik, V. Hoerr, M. Glaser, M. Schultheis, E. Hansell, J. H. McKerrow, U. Holzgrabe, C. R. Caffrey, A. Ponte-Sucre, H. Moll, A. Stich, T. Schirmeister, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 2753; A. Ponte-Sucre, R. Vicik, M. Schultheis, T. Schirmeister, H. Moll, *Antimicrob. Agents Chemother.*, 2006, **50**, 2439; F. Schulz, C. Gelhaus, B. Degel, R. Vicik, S. Heppner, A. Breuning, M. Leippe, J. Gut, P.J. Rosenthal, T. Schirmeister, *ChemMedChem*, 2007, **2**., 1214.
- ¹⁷ C. Gelhaus, R. Vicik, R. Hilgenfeld, C. L. Schmidt, M. Leippe, T. Schirmeister, *Biol. Chem.*, 2004, **385**, 435; C. Gelhaus, R. Vicik, T. Schirmeister, M. Leippe, *Biol. Chem.*, 2005, **386**, 499; R. Vicik, M. Busemann, C. Gelhaus, N. Stiefl, J. Scheiber, W. Schmitz, F. Schulz, M. Mladenovic, B. Engels, M. Leippe, K. Baumann, T. Schirmeister, *ChemMedChem*, 2006, **1**, 1126.
- ¹⁸ I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* 1967, **27**, 157; A. Berger, I. Schechter, *Phil. Trans. R. Soc. London*, 1970, **B257**, 1330.
- ¹⁹ K. I. Varughese, F. R. Ahmed, P. R. Carey, S. Hasnain, C. P. Huber, A. C. Storer, *Biochemistry*, 1989, **28**, 1330-1332; K. I. Varughese, Y. Su, D. Cromwell, S. Hasnain, N. H. Xuong, *Biochemistry*, 1992, **31**, 5172-5176;
- ²⁰ J. Drenth, K. Kalk, H. Swen, *Biochemistry*, 1976, **15**, 3731-3738.
- ²¹ T. Aoyagi, H. Umezawa, In *Proteases and Biological Control*; E. Reich, D. Rifkind, E. Shaw, Eds.; Cold Spring Harbor: New York, 1975; pp 429-454.
-

- ²² G. Lowe, Y. Yuthavong, *Biochem. J.*, 1971, **124**, 107.
- ²³ C. M. Yu, J. M. Curtis, J. A. Walter, J. L. Wright, S. W. Ayer, J. Kaleta, L. Querengesser, Z. R. Fathi-Afshar, *J. Antibiot (Tokyo)*, 1996, **49**, 395; J. T. Woo, H. Ono, T. Tsuji, *Biosci. Biotechnol. Biochem.*, 1995, **59**, 350.
- ²⁴ S. Yaginuma, A. Asahi, A. Morishita, M. Hayashi, M. Tsujino, M. Takada, *J. Antibiot. (Tokyo)*, 1989, **42**, 1362.
- ²⁵ T. Otsuka, Y. Muramatsu, T. Higaki, T. Nakanishi, S. Takase, H. Hatanaha, M. Okamoto, M. Hino, S. Hashimoto, *J. Antibiot. (Tokyo)*, 1999, **52**, 536; T. Otsuka, Y. Muramatsu, K. Niikura, M. Okamoto, M. Hino, S. Hashimoto, *J. Antibiot. (Tokyo)*, 1999, **52**, 542.
- ²⁶ M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.*, 1981, **103**, 1425.
- ²⁷ J. J. P. Stewart, *Rev. Compt. Chem.*, 1990, **1**, 45.
- ²⁸ J. J. P. Stewart, *J. Compt. Chem.*, 1989, **10**, 209, 221.
- ²⁹ T. Thomas, *PCPS*, 1927, **23**, 542
- ³⁰ E. Fermi, *Z. Phys.*, 1928, **48**, 73
- ³¹ A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
- ³² M. Field, P. Bash, M. Karplus, *J. Comput. Chem.*, 1990, **32**, 904
- ³³ I. Antes, W. Thiel, *J. Phys. Chem. A*, 1999, **103**, 9290.
- ³⁴ Y. K. Zhang, T. S. Lee, W. T. Yang, *J. Chem. Phys.*, 1999, **110**, 46.
- ³⁵ V. Thery, D. Rinaldi, J.-L. Rivail, B. Maigret, B. Frenzy, *J. Comput. Chem.*, 1994, **15**, 269; G. Monard, M. Loos, V. Thery, K. Baka, J.-L. Rivail, *Int. J. Quantum Chem.*, 1996, **58**, 153; X. Assfeld, J.-L. Rivail, *Chem. Phys. Lett.*, 1996, **263**, 100.
- ³⁶ J. Gao, P. Amara, C. Alhambra, M. J. Field, *J. Phys. Chem. A*, 1998, **102**, 4714.
- ³⁷ D. Bakowies, W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580
- ³⁸ TURBOMOLE, Versions 5.6 R. Ahlrichs, M. Bär, H.-P. Baron, R. Bauernschmitt, S. Böcker, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, H. Horn, C. Huber, U. Huniar, M. Kattaneck, C. Kölmel, M. Kollwitz, K. May, C. Ochsenfeld, H. Öhm, A. Schäfer, U. Schneider, O. Treutler, M. v. Arnim, F. Weigend, P. Weis, H. Weiss, University of Karlsruhe, Germany, **since 1988**.
- ³⁹ A. Klamt, G. J. Schürmann, *J. Chem. Soc. Parkin. Trans. 2*, 1993, **2**, 799; A. Schäfer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert, *Phys. Chem. Chem. Phys.*, 2000, **2**, 2187.
- ⁴⁰ O. Vahtras, J. Almlöf, M. W. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514; K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **242**, 652.

-
- ⁴¹ A. D. Becke, *Phys. Rev. B.*, 1988, **38**, 3098; C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B.*, 1988, **37**, 785.
- ⁴² A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* 1997, **100**, 5829
- ⁴³ A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
- ⁴⁴ H. Helten, T. Schirmeister, B. Engels, *J. Phys. Chem. B*, 2004, **108**, 7691.
- ⁴⁵ K. I. Varughese, Y. Su, D. Cromwell, S. Hasnain, N. H. Xuong, *Biochemistry* 1992, **31**, 5172.
- ⁴⁶ H. Helten, T. Schirmeister, B. Engels, *J. Org. Chem.*, 2005, **70**, 233.
- ⁴⁷ H. H. Otto, T. Schirmeister, *Chem. Rev.* 2002, **102**, 4459.
- ⁴⁸ T. Schirmeister, U. Käppler, *Mini Rev. Med. Chem.*, 2003, **3**, 361.
- ⁴⁹ V. Martichonok, C. Plouffe, A. C. Storer, R. Menard, J.B. Jones, *J. Med. Chem.* 1995, **38**, 3078.
- ⁵⁰ T. Schirmeister, M. Peric, *Bioorg. Med. Chem.* 2000, **8**, 1281.
- ⁵¹ R. Vicik, H. Helten, T. Schirmeister, B. Engels, *ChemMedChem*, 2006, **1**, 1021.
- ⁵² Private communications
- ⁵³ For some reviews see: J. Aqvist, A. Warshel, *Chem. Rev.* 1993, **93**, 2523; J. Gao, In *Reviews in Computational Chemistry*, Vol.7, K.B. Ipkowitz, D. B. Boyd, Eds.; VCH: Weinheim, 1995, p 119; T. Z. Mordasini, W. Thiel, *Chimia*, 1998, **52**, 288; G. Monard, K. M. Merz, Jr., *Acc. Chem. Res.*, 1999, **32**, 904; P. Sherwood, in *Modern Methods and Algorithms of Quantum Chemistry*, J. Grotendorst, Ed.; NIC Series 3 John von Neumann Institute for Computing, Jülich, 2000, p 285.; J. Gao, D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467; M. J. Field, *J. Comput. Chem.*, 2002, **23**, 48; G. Monard, X. Prat-Resina, A. Gonzalez-Lafont, J. M. Lluch, *Int. J. Quantum. Chem.*, 2003, **93**, 229; L., Ridder, A. Mulholland, *Curr. Top. Med. Chem.*, 2003, **3**, 1241.
- ⁵⁴ M. J. Field, P. A. Bash, M. J. Karplus, *J. Comput. Chem.*, 1990, **11**, 700; P. Amara, M. J. Field, *J. Theor. Chem. Acc.*, 2003, **109**, 43; N. Reuter, A. Dejaegere, B. Maigret, M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720; U. C. Singh, P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718, E. Derat, J. Bouquant, S. Humbel, *THEOCHEM*, 2003, **632**, 61; A. H. deVries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, G. J. Kramer, *J. Phys. Chem. B*, 1999, **103**, 6133.
- ⁵⁵ A. Yamamoto, T. Tomoo, K. Matsugi, T. Hara, Y. In, M. Murata, K. Kitamura, T. Ishida, *Biochim. Biophys. Acta*, 2002, **1597**, 244.
-

- ⁵⁶ Please note that the His residue in the active site is denoted as His159 for the papain numbering while His199 is used within the cathepsin B numbering. For the introduction the papain numbering was employed since more investigations were performed for this enzyme. For the present study cathepsin B was used since more appropriate X-ray structures exist. Therefore the cathepsin B numbering is used for the discussion.
- ⁵⁷ A. D. Jr. MacKarell, et al., *J. Phys. Chem. B*, 1998, **102**, 3586; B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus *J. Comp. Chem.*, 1983, **4**, 187; A. D. Jr. MacKerell, B. R. Brooks, C. L. III Brooks, L. Nilsson, B. Roux, Y. Won, M. Karplus, in *The Encyclopedia of Computational Chemistry*, Vol. 1, P. v. R. Schleyer et al., Eds.; John Wiley & Sons, Chichester, 1998, p. 271.
- ⁵⁸ P. Sherwood, et al., *THEOCHEM* 2003, **632**, 1.
- ⁵⁹ W. Smith, T. Forester, *J. Mol. Graph.*, 1996, **14**, 136.
- ⁶⁰ S. R. Billeter, A. J. Turner, W. Thiel, *Phys. Chem. Chem. Phys.*, 2000, **2**, 2177.
- ⁶¹ W. Adam, N. Bottke, B. Engels, O. Krebs, *J. Am. Chem. Soc.* 2001, **123**, 5542.
- ⁶² S. Schlund, M. Mladenovic, E. M. Basilio Janke, B. Engels, K. Weisz, *J. Am. Chem. Soc.* 2005, **127**, 16151.
- ⁶³ T. Hupp, C. Sturm, E. M. Basilio Janke, M. Perez Cabre, K. Weisz; B. Engels, *J. Phys. Chem.* 2005, **109**, 1703.
- ⁶⁴ B. Engels, J. C. Schoeneboom, A. F. Munster, S. Groetsch, M. Christl, *J. Am. Chem. Soc.* 2002, **124**, 287.
- ⁶⁵ H. U. Suter, V. Pleß, M. Ernzerhof, B. Engels, *Chem. Phys. Lett.* 1994, **230**, 398.
- ⁶⁶ H. U. Suter, B. Engels, *J. Chem. Phys.* 1994, **100**, 2936.
- ⁶⁷ B. Engels, *Theo. Chim. Acta* 1993, **86**, 429.
- ⁶⁸ M. Schmittel, J. P. Steffen, B. Engels, et al., *Angew Chem Int Ed*, 1998, **37**, 2371.
- ⁶⁹ B. Engels, S. D. Peyerimhoff, *J. Phys. Chem.*, 1989, **93**, 4462.
- ⁷⁰ B. Engels, M. Hanrath, Ch. Lennartz *Computer & Chemistry*, 2001, **25**, 15-38
- ⁷¹ P. W. Musch, B. Engels, *J. Am. Chem. Soc.*, 2001, **123**, 5557.
- ⁷² P. R. Carlier, N. Deora, T. D. Crawford, *J Org Chem*, 2006, **71**, 1592.
- ⁷³ H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.*, **81**, 1984, 3684.
- ⁷⁴ L. Verlet, *Phys. Rev.*, 1967, **159**, 98; R. W. Hockney, *Methods in Computational Physics*, 1970, **9**, 136.

-
- ⁷⁵ D. H. Rich, in *Proteinase Inhibitors*; A. J. Barrett, G. S. Salvesen, Eds.; Elsevier Science Publishers B.V., Amsterdam, 1986, p153.
- ⁷⁶ J. P. Meara, D. H. Rich, *J. Med. Chem.* 1996, **39**, 3357-3366.
- ⁷⁷ T. Schirmeister, M. Peric, *Bioorg. Med. Chem.* 2000, **8**, 1281-1291.
- ⁷⁸ Z. P. Chen, X. L. Zhu, G. D. Fu, E.T. Kang, K.G. Neoh *Macromolecules*, 2005, **38**, 7187.
- ⁷⁹ P. R. Carlier, *Angew Chem Int Ed*, 2004, **43**, 2602.
- ⁸⁰ A. Padva, S. Murohree, *ARKIVOC*, 2006, **3**, 6.
- ⁸¹ Aziridines and epoxides in organic synthesis, A. K. Yudin, Ed.; Wiley-VCH, Weinheim, Germany, 2006.
- ⁸³ J. P. Meara, D. H. Rich, *Bioorg. Med. Chem. Lett.*, 1995, **5**, 2277.
- ⁸⁴ R. Bihovsky, J. C. Powers, C. M. Kam, R. Walton, R. C. Loew, *J. Enzyme Inhib.*, 1993, **7**, 15.
- ⁸⁵ T. Schirmeister, *Arch. Pharm. Pharm. Med. Chem.*, 1995, **329**, 239.
- ⁸⁶ R. Bihovsky, *J. Org. Chem.*, 1992, **57**, 1029.
- ⁸⁷ M. Mladenovic, T. Schirmeister, S. Thiel, W. Thiel, B. Engels, *ChemMedChem*, 2007, **2**, 120.
- ⁸⁸ K. Matsumoto, M. Murata, S. Sumiya, K. Mizoue, K. Kitamura, T. Ishida, *Biochem. et Biophys. Acta*, 1998, **1383**, 93.
- ⁸⁹ A. Fujishima, Y. Imai, T. Nomura, Y. Fujisawa, Y. Yamamoto, T. Sugawara, *FEBS Lett.*, 1997, **407**, 47.
- ⁹⁰ B. Zhao, C. A. Janson, B. Y. Amegadzie, K. D'Alessio, C. Griffin, C. R. Hanning, C. Jones, J. Kurdyla, M. McQueney, X. Qiu, W. W. Smith, S. S. Abdel-Meguid, *Nat. Struct. Biol.*, 1997, **4**, 109.
- ⁹¹ D. Yamamoto, K. Matsumoto, H. Ohishi, T. Ishida, M. Inoue, K. Kitamura, H. Mizuno, *J. Biol. Chem.*, 1991, **266**, 14771.
- ⁹² D. Watanabe, A. Yamamoto, K. Tomoo, K. Matsumoto, M. Murata, K. Kitamura, T. Ishida, *J. Mol. Biol.*, 2006, **362**, 979.
- ⁹³ K. Varughese, F. Ahmed, P. Carey, S. Hasnain, C. Huber, A. Storer, *Biochemistry*, 1989, **28**, 1330.
- ⁹⁴ I. Stern, N. Schaschke, L. Moroder, D. Turk, *Biochem. J.*, 2004, **381**, 511.
- ⁹⁵ A. Yamamoto, K. Tomoo, T. Hara, M. Murata, K. Kitamura, T. Ishida, *J. Biochem. (Tokyo)*, 2000, **127**, 635.
-

- ⁹⁶ D. Turk, M. Podobnik, T. Popovic, N. Katunuma, W. Bode, R. Huber, V. Turk, *Biochemistry*, 1995, **34**, 4791.
- ⁹⁷ K. Hanada, M. Tamai, S. Morimoto, T. Adachi, S. Ohmura, J. Sawada, I. Tanaka, *Agric. Biol. Chem.*, 1978, **42**, 537.
- ⁹⁸ N. Schaschke, I. Assfalg-Machleidt, W. Machleidt, D. Turk, L. Moroder, *Bioorg. Med. Chem.*, 1997), **5**, 1789.
- ⁹⁹ N. Schaschke, I. Assfalg-Machleidt, W. Machleidt, L. Moroder, *FEBS Lett.*, 1998, **421**, 80.
- ¹⁰⁰ M.J. Kim, D. Yamamoto, K. Matsumoto, M. Inoue, T. Ishida, H. Mizuno, S. Sumiya, K. Kitamura, *Biochem.J.*, 1992, **287**, 797.
- ¹⁰¹ H. Tsuge, T. nishimura, Y. tada, T. Asao, D. Turk, V. Turk, N. Ktunuma, *Biochem. Biophys. Res. Commun.*, 1999, **266**, 411.
- ¹⁰² Private communications with one of the authors.
- ¹⁰³ The error may result from uncertainties in the description of C-S bonding.
- ¹⁰⁴ D. Musil, D. Zucic, D. Turk, R. A. Engh, I. Mayr, R. Huber, T. Popovic, V. Turk, T. Towatari, N. Katunuma, W. Bode, *EMBO J.*, 1991, **10**, 2321.
- ¹⁰⁵ R. W. Pickersgill, G. W. Harris, E. Garman, *Acta Crystallogr., Sect.B*, 1992, **48**, 59.
- ¹⁰⁶ I. G. Kamphuis, K. H. Kalk, M. B. Swarte, J. Drenth, *J.Mol.Biol.*, 1984, **179**, 233.
- ¹⁰⁷ S. Jang, G. A.Voth, *J. Chem.Phys.*, 1997, **107**, 9514.
- ¹⁰⁸ S. J. Nosé, *J.Chem.Phys.*, 1984, **81**, 511-519; S. J. Nosé, *Mol.Phys.*, 1984, **52**, 255; W. G. Hoover, *Phys.Rev. A*, 1985, **31**, 1695; G. J. Martyna, M. L. Klein, M. Tuckerman, *J.Chem.Phys.*, 1992, **97**, 2635.
- ¹⁰⁹ S. D. Lewis, F. A. Johnson, J. A.Shafer, *Biochemistry*, 1976, **82**, 5009.
- ¹¹⁰ D. J. Creighton, M. S. Gessouroun, J. M. Heapes, *FEBS Lett.*, 1980, **110**, 319.
- ¹¹¹ S. D. Lewis, F. A. Johnson, J. A. Shafer, *Biochemistry*, 1981, **20**, 48.
- ¹¹² F. A. Johnson, S. D. Lewis, J. A. Shafer, *Biochemistry*, 1981, **20**, 44.
- ¹¹³ A. C. Storer, R. Ménard, *Methods Enzymol.* 1994, **244**, 486.
- ¹¹⁴ P. T. van Duijnen, B. T. Thole, W. G. Hol, *Biophys. Chem.*, 1979, **9**, 273.
- ¹¹⁵ J. Aquist, H. Luecke, F. A. Quioco, A. Warshel, *Proc. Natl. Acad. Sci. U.S.A.*, 1991, **88**, 2026.
- ¹¹⁶ J. Sancho, L. Serrano, A. R. Fersht, *Biochemistry*, 1992, **31**, 22.
- ¹¹⁷ J. A. Rullmann, M. N. Bellido, P. T. van Duijnen, *J. Mol. Biol.*, 1989, **206**, 101.

¹¹⁸ T. Vernet, D. C. Tessier, J. Chatellier, C. Plouffe, T. S. Lee, D. Y. Thomas, A. C. Storer, R. Ménard, *J. Biol. Chem.*, 1995, **270**, 16645.

Acknowledgements

I wish to thank all the people who have helped me with advice and encouragement and made this thesis possible.

In the first place I wish to thank my boss, Prof. Dr. Bernd Engels, for giving me the opportunity to work on this interesting research. I thank him for the freedom he gave me in doing my work and for the trust he had in me.

I thank all past and present members of the working group of Prof. Dr. Bernd Engels, especially Christian Sturm, Mario Arnone, Thomas Hupp, and Alex Paasche for making work a better place. Special thanks to Sebastian Schlund, who was not just a good colleague, but also the best friend one could ever have.

Thanks to Priv.-Doz. Dr. Reinhold Fink, for the time he invested in helping me with all sorts of problems encountered while trying to understand the mysteries of quantum chemistry.

Thanks to Uschi Rüppl, for encouragement and for help in the struggle against the German bureaucracy.

Special thanks to James Asher for having the time and nerves to read and correct my English.

I wish to thank Prof. Dr. Tanja Schirmeister (University of Würzburg), our collaboration partner, for answering all my questions about pharmaceutical chemistry. Thanks to Verena Schulz, a member of Prof. Dr. Schirmeister's group, for inspiring conversations.

I wish to thank Dr. Stephan Thiel, a former group member of the working group of Prof. Dr. Walter Thiel (MPI, Mülheim), for helping me fight through the tcl scripts and get started with the QM/MM calculations, and Prof. Dr. Walter Thiel for the fruitful collaboration.

Finally, I thank my parents, Branislava and Dusan Mladenovic, for setting me on the right path and always giving me support and encouragement, and my brother, Borislav Mladenovic, for bringing light into even the gloomiest and hardest days. I wish to thank my boyfriend, Philipp Beisheim, for his support and encouragement and for the patience and understanding he had when the working days just never seemed to end. I thank all of my friends.

LIST OF PUBLICATIONS

1. Mladenovic, M.; Schirmeister, T.; Thiel, S.; Thiel, W.; Engels, B. The importance of the active site histidine for the activity of epoxide- or aziridine-based inhibitors of cysteine proteases. *ChemMedChem* (2007), 2(1), 120-128.
2. Vicik, R.; Busemann, M.; Gelhaus, C.; Stiefl, N.; Scheiber, J.; Schmitz, W.; Schulz, F.; Mladenovic, M.; Engels, B.; Leippe, M.; Baumann, K.; Schirmeister, T. Aziridine-based inhibitors of cathepsin L: synthesis, inhibition activity, and docking studies. *ChemMedChem* (2006), 1(10), 1126-1141.
3. Schlund, S.; Mladenovic, M.; Janke, E. M. B.; Engels, B.; Weisz, K.. Geometry and Cooperativity Effects in Adenosine-Carboxylic Acid Complexes. *Journal of the American Chemical Society* (2005), 127(46), 16151-16158.
4. Mladenovic, M.; Peric, M.; Engels, B. An ab initio study of the vibronic, spin-orbit, and magnetic hyperfine structure in the X²Π electronic state of NCO. *Journal of Chemical Physics* (2005), 122(14), 144306/1-144306/9.
5. Mladenovic, M.; Peric, M.; Rankovic, R.; Engels, B. An ab initio study of the hyperfine structure in the X²Π electronic state of HCCS-calculation of vibronically averaged components of the anisotropic hyperfine tensor. *Molecular Physics* (2005), 103(4), 587-598.
6. Schirmeister, T.; Breuning, A.; Murso, Al.; Stalke, D.; Mladenovic, M.; Engels, B.; Szeghalmi, A.; Schmitt, M.; Kiefer, W.; Popp, J.. Conformation and Hydrogen Bonding Properties of an Aziridiny Peptide: X-ray Structure Analysis, Raman Spectroscopy and Theoretical Investigations. *Journal of Physical Chemistry A* (2004), 108(51), 11398-11408.
7. Mladenovic, M.; Peric, M.; Jerosimic, S.; Engels, B. Ab initio study of the hyperfine structure of the X²Π electronic state of HCCS. *Molecular Physics* (2004), 102(23-24), 2623-2634.
8. Mladenovic, M.; Peric, M.; Engels, B. An ab initio calculation of the anisotropic hyperfine coupling constants in the low-lying vibronic levels of the X²Π electronic state of CCCH. *Journal of Chemical Physics* (2004), 121(24), 12361-12370.
9. Peric, M.; Mladenovic, M.; Engels, B. An ab initio study of the hyperfine structure in the X²Π electronic state of CCCH. *Journal of Chemical Physics* (2004), 121(6), 2636-2645.
10. Peric, M.; Mladenovic, M.; Engels, B. An ab initio study of the vibronic, spin-orbit and hyperfine coupling in the X²Π electronic state of the CCCD radical. *Chemical Physics Letters* (2004), 393(4-6), 552-557.
11. Peric, M.; Mladenovic, M.; Tomic, K.; Marian, C. M. Ab initio study of the vibronic and spin-orbit structure in the X²Π electronic state of CCCH. *Journal of Chemical Physics* (2003), 118(10), 4444-4451.