UiT
THE ARCTIC UNIVERSITY OF NORWAY

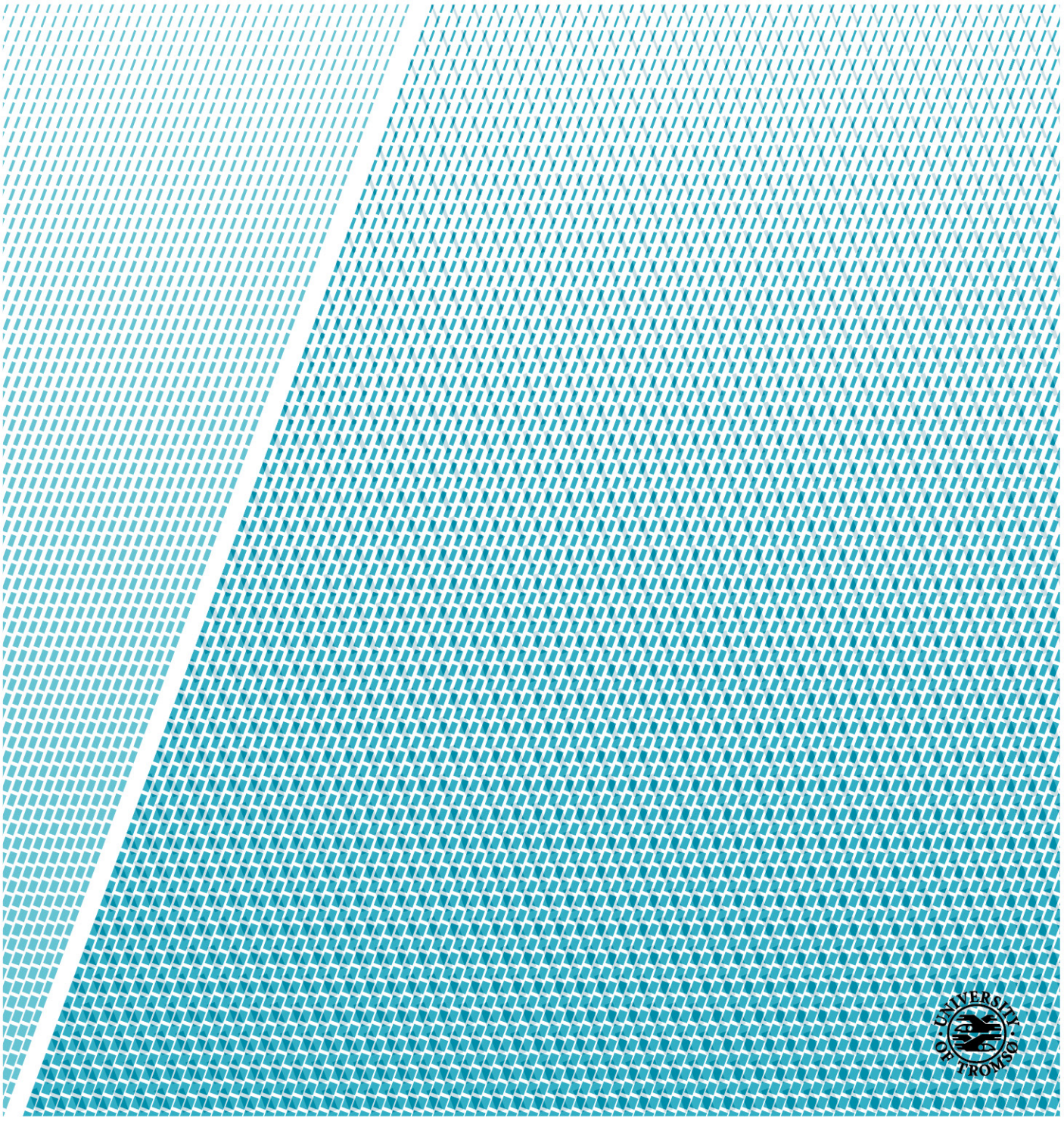
Faculty of Science and Technology
Department of Chemistry

Computer Simulations of Antimicrobial Tripeptides with Proteolytic Stability Towards Chymotrypsin

—

Bente Sirin Barge

KJE-3900 Master's thesis in Molecular Sciences, December 2019



Abstract

The emerging crisis of antimicrobial resistance calls for the development of new therapeutic drugs. Antimicrobial peptides (AMPs), found in a variety of species as part of the host-defence mechanisms, serves as promising candidates for clinical use. Studies of well-known AMPs like lactoferrin have lead to the preparation of short cationic AMPs (CAPs) with broad spectrum antimicrobial activity against bacteria, fungi, enveloped viruses, parasites and cancer cells. However, in order for the CAPs to have therapeutic effect it must face and overcome several *in vivo* challenges. One major hurdle is the proteolytic stability towards chymotrypsin, an enzyme present in the gastrointestinal tract. In this study, the interaction between chymotrypsin and active CAPs has been studied using molecular modelling and computer simulations. It was found that CAPs cleaved by the enzyme have stronger electrostatic interactions with the enzyme than the stable CAPs, both in the bound state and in the transition state. The stable CAPs were found to have structural features which can hinder optimal orientation of the main-chain in the transition state, and therefore lower the electrostatic stabilization. Furthermore, the enzyme was found to be preorganized with a polar environment that stabilizes the transition state more than the reactant state.

Acknowledgements

My first and biggest appreciation goes to Bjørn Olav Brandsdal— I cannot imagine a better or cooler supervisor than him. My co-supervisor Geir Villy Isaksen, an equally cool supervisor, I truly appreciate your effort to teach me and the guidance I have received. Thank you to the rest of the *Brandsdal group*, you have all made me feel welcome.

I am grateful for my time at UiT and for the people that I have had the pleasure of meeting during my five years as a student. Thank you to my fellow master students. I appreciate the time spent studying, but even more so the time spent not studying.

Lastly, I would like to thank my mother Mai for her support, and my father Bjørn, who has taught me many life lessons and has always challenged me do to my very best.

Contents

Abstract **i**

Acknowledgements **iii**

List of Abbreviations **vii**

1 Introduction **1**
1.1 Antimicrobial Resistance and Its Origin 1
1.2 Antimicrobial Peptides 2
1.3 The Pharmacophore of Cationic Antimicrobial Peptides 4
1.4 Serine Proteases and the Proteolytic Fate of CAPs 6
1.5 Computational Tools 8
1.5.1 Molecular Docking 8
1.5.2 Molecular Mechanics 10
1.5.3 Simulation Methods and Molecular Dynamics 15
1.5.4 The Linear Interaction Energy 19
1.5.5 Free Energy Perturbation 20
1.5.6 Empirical Valence Bond Theory 22
1.6 Aim of Study 25

2 Methods **27**
2.1 Protein Preparation and Docking 27
2.2 Molecular Dynamics 29
2.3 Binding Affinity of CAPs with LIE 32
2.3.1 Preparing Topology 32
2.3.2 MD Simulation Details 32
2.4 Free Energies with EVB 33
2.4.1 Preparing Topology 33
2.4.2 Preparation of Dynamic Control Input and FEP File 34
2.4.3 MD Simulation Details 35
2.5 Transition State Stabilization Using Modified LIE 35

2.5.1 MD Simulation Details 36

3 Results and Discussion **37**
3.1 Binding of CAPs to Chymotrypsin 37
3.1.1 Docking Results 38
3.1.2 Binding Free Energies with LIE 39
3.2 Activation and Reaction Free Energies 44
3.3 Binding Affinity of the TS Relative to RS 47

4 Concluding Remarks **53**

Bibliography **57**

Appendix A LIE Energy Fluctuation Graphs **65**

Appendix B Modified-LIE Energy Fluctuation Graphs **67**

List of Abbreviations

Agp	L-2-amino-(3-guanidino)propanoic acid
AMPs	Antimicrobial peptides
AMR	Antimicrobial resistance
App	L-2-amino-3-(4-aminophenyl)propanoic acid
Bip	L-2-amino-biphenylalanine
BPTI	Bovine pancreatic trypsin inhibitor
CAPs	Cationic antimicrobial peptides
EVB	Empirical valence bond theory
FEP	Free energy perturbation
Har	L-2-amino-(6-guanidino)hexanoic acid
ITC	Isothermal titration calorimetry
LIE	Linear interaction energy
MD	Molecular dynamics
MIC	Minimal inhibitory concentration
MM	Molecular mechanics
MRSA	Methicillin-resistant *Staphylococcus aureus*

NHBn	NH-benzyl
PDB	Protein Data Bank
RS	Reactant state
Tbt	L-2-amino-2,5,7-tri*tert*-butyl tryptophan
TS	Transition state
VB	Valence bond
XP	Extra precision (Glide docking)

1 Introduction

1.1 Antimicrobial Resistance and Its Origin

Modern medicine has revolutionized the treatment of infectious diseases. Diseases such as tuberculosis and syphilis are rarely a threat in developed countries and, by its population, usually only encountered in history textbooks.

The antibiotic era started in the 1940s with the distribution of penicillin.[1] The pharmaceutical industry flourished due to the succeeding new discoveries and mass distribution of novel antibiotic classes. Antibiotics and other antimicrobial therapies drastically improved the human health while creating huge profits for the pharmaceutical industry. In the 1970s, after decades of innovation, the situation changed. With few new discoveries of antimicrobics and the decline of infectious diseases, pharmaceutical companies shifted their focus to the more profitable research of chronic diseases (cardiovascular diseases, cancer etc.).[2]

Now, infectious diseases once again pose a serious threat for human health due to antimicrobial resistance (AMR). The use of antimicrobials and vaccines have drastically lowered the mortality rate related to infectious diseases in high-income countries. However, only one human infectious disease, smallpox, has successfully been eradicated by human effort.[3] Infectious diseases are still very much present world-wide and are among the main causes of death in low-income countries as of 2016.[4] Bacterial pathogens such as *Staphylococcus aureus* (causes septicemia among other things) and *Mycobacterium tuberculosis* (causes tuberculosis) have both evolved into multi-drug-resistant forms. It has been estimated that AMR causes at least 700,000 deaths a year, a figure which can increase to 10 million deaths per year by year 2050 if no preventing actions are taken.[5]

The ability of bacteria to resist antibiotics is obtained by their natural evolutionary process. Bacteria have a short generation turnover which ensure exponential growth and the occasional gene mutation. The rate of mutation accelerates under environmental stress, and the stress rising from the presence of antimicrobics is of no exception to the bacteria.[6][7] Antibiotic resistance can also arise from *horizontal gene transfer*, which is the exchange of genetic information between cells.[8] Horizontal gene transfer occurs between cells of both the same and between different species, and is therefore a rapid way for bacteria to attain resistance.

Antibiotics are not only used for therapeutic means but also widely used in aquaculture and agriculture to prevent infections and as growth promoters (though this practice was banned in the EU in 2006 and in the US in 2017).[9][10] As a result of human activity, the aquatic environment has been contaminated by man-made antibiotics, whether through sewage infrastructure, agricultural runoff or other inadequate disposal of unused antibiotics. Surface water has been suggested to have a key role in AMR development.[11] When antibiotics finds its way to surface water, it enforces antibiotic pressure on bacteria native to the aquatic environment. The pathogenic bacteria present in surface water may gain relevant resistance mechanisms through genetic exchange with environmental species.[7][11] As a result of antimicrobial overuse and inadequate handling, the environment, like coastal beaches, has become a plausible site for catching AMR infections.[12][13]

1.2 Antimicrobial Peptides

Antimicrobial peptides (AMPs) are a class of peptides widely distributed in nature as part of an organisms innate immune system. The class is structurally diverse and displays therapeutic activity as antibacterials, antifungals, antivirals and anticancers.[14]. Most AMPs share some common structural features; their chain is less than 60 amino acids long, they possess an overall positive charge and they have the ability to adapt into an amphipathic design, which means that they possess both a polar and non-polar region.[15]

The bacterial membrane is partially populated by negatively charged phospholipids that gives the membrane an overall negative charge. The amphipathic 3D structure of AMPs is key for their antimicrobial activity as it allows for interactions

with the bacterial membrane through both electrostatic and hydrophobic forces. Eukaryotic membrane surfaces are overall neutrally charged, which means that positively charged amphipathic structures interact with the eukaryotic membrane only through hydrophobic interactions. This makes the bacterial membrane a target for AMPs, even in the presence of eukaryotic cells. It is believed that the AMPs take an amphipathic form on contact with the cytoplasmic membrane thereby allowing the charged domain to interact with the negatively charged phospholipids while the lipophilic part buries itself deep into the membrane. The killing mechanism is still unclear, and it might vary between AMPs or even be a multiple action mechanism. Various membrane disruption models like "torosidal-pore", "barrel-stave" and "carpet model" have been proposed, all of which leads to membrane disruption and cell death by lysis.[14] AMPs are also capable of translocation over the membrane, making intracellular killing mechanisms just as plausible.[16]

It is clear that AMPs are promising candidates for broad-spectrum therapeutic use in clinics. They are potentially less likely to provoke resistance in target bacteria due to the mode of action; the membrane is a fundamental structure thought to be less prone to mutation. If AMPs have multiple killing mechanisms, this could also decrease the risk of inducing bacterial resistance. However, the discovery of plasmid-mediated resistance against the membrane active peptide colistin (an antibiotic) is a distressing finding.[17]

AMPs also possess attributes that are challenging for clinical application. Peptides face several obstacles *in vivo* on their path towards their bacterial target. Orally administrated peptides are subjected to low absorption in the gastrointestinal tract and first-pass metabolism, which is a phenomenon associated with drug administered orally, and will reduce the amount of drug that reaches the systemic circulation.[18] Parenteral administered drugs avoid the problem of absorption and lower the first-pass effect but still fall victim to renal clearance and/or enzymatic degradation. Plasma half-life, which is the time required for the concentration of a substance to be reduced by one half, has to be sufficiently long for a drug to display pharmacologic activity. Strategies for prolonging plasma half-life include substituting amino acids with corresponding D-amino acid, cyclization of the peptide, increase molecular mass to avoid renal secretion, co-administration with enzyme inhibitors or utilize liposomes or other delivery systems.[19]

Which strategy to follow depends on the peptide and the 'attacking' enzyme(s). Success will come from surveying the problems that lie ahead for the promising peptide and obtaining structural information on relevant enzymes. Molecular modelling can be of great help in this capacity as a useful tool for investigating cleavage specificity and the interaction between peptides and proteolytic enzymes.

1.3 The Pharmacophore of Cationic Antimicrobial Peptides

In 2003, Strøm *et al.* investigated the minimum requirement of charge and lipophilic properties necessary for antimicrobial activity in cationic antimicrobial peptides (CAPs).[20] The minimal motif required was surprisingly small. The anti-staphylococcal pharmacophore was found to consist of only two cationic and two bulky lipophilic moieties. The charged moiety can be arginine or lysine while the lipophilic can be tryptophan or phenylalanine. This suggest the CAPs as small as tri- and dipeptides with a positive N-terminal and a hydropohobic elements as C-terminal capping are conceivable as future antimicrobial agents. Further investigations were performed on the possibility of increasing potency by incorporating unnatural hydrophobic residues.[21] It was found that bulkier non-standard residues yielded lower minimal inhibitory concentration (MIC) against *Escherichia coli* and *Staphylococcus aureus* with biphenylalanine (Bip) and 2,5,7-tri*tert*-butyl tryptophan (Tbt) showing the best improvement among the unnatural residue candidates (Figure 1.1). The effect of modifying the charged moiety was investigated by incorporating unnatural residues, some of which are shown Figure 1.2.[22] Both peptides incorporating the short L-2-amino-(3-guanidino)propanoic acid (Agp) and the long L-2-amino-(6-guanidino)hexanoic acid (Har) showed somewhat lower MIC values against *S. aureus* and methicillin-resistant *S. aureus* (MRSA) when comparing to the corresponding peptide incorporating arginine. This indicates some effect on potency by adjusting length of the side-chain. The effect of changing the charged moiety to a primary amine using lysine and the effect of removing the positive charge by using L-2-amino-3-(4-aminophenyl)propanoic acid (App) (which have a $pK_a \sim 5$) was also investigated in the same study. The use of a primary amine over

a guanidine residue generally lowered the potency and the removal of the positive charge also reduced potency.

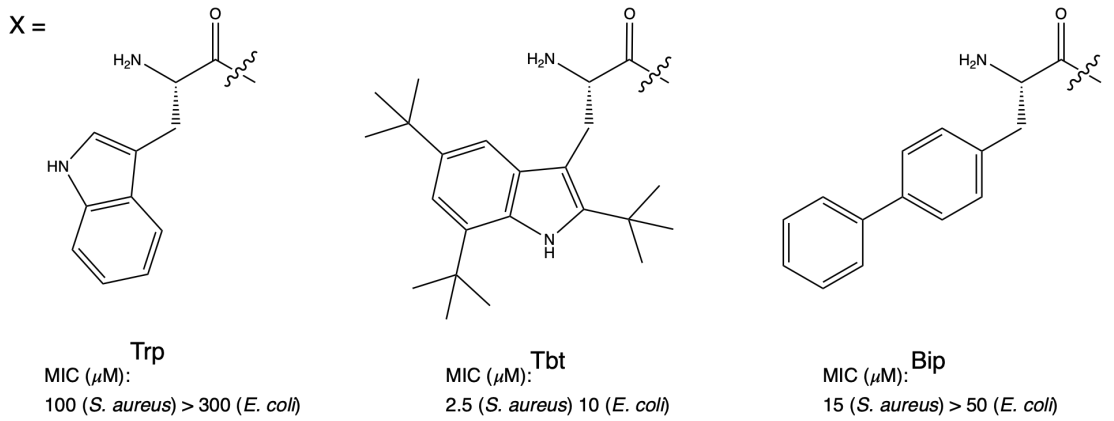


Figure 1.1: Structure of hydrophobic residues Trp, Tbt and Bip investigated in the study by Haug *et al.*[21] The MIC values were tested for dipeptides with a X-R-NHBn motif, where X is is the bulky residues of investigation, R is the residue arginine and NHBn is NH-benzyl, the hydrophobic C-terminal capping.

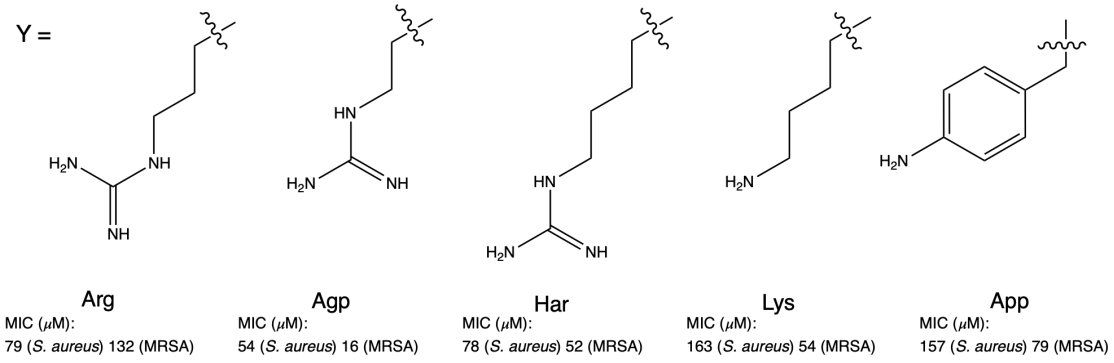


Figure 1.2: Structures of residues Arg, Agp, Har, Lys and App investigated in the study by Karstad *et al.*[22] The MIC values were tested for tripeptides with a R-W-Y-NHBn motif, where Y is is the charged residue of investigation, R is the residue arginine, W is the residue tryptophan and NHBn is NH-benzyl, the hydrophobic C-terminal capping.

1.4 Serine Proteases and the Proteolytic Fate of CAPs

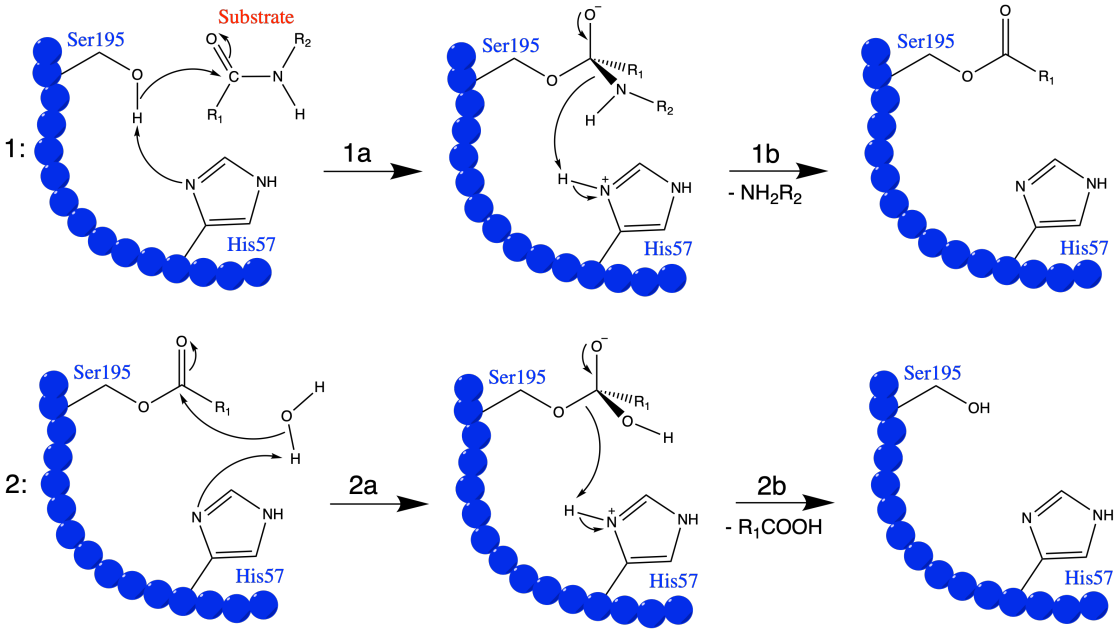


Figure 1.3: Reaction mechanism of serine proteases. Step 1 (on top) leads to the release of the first peptide and the formation of an acyl-enzyme intermediate. Step 2 (bottom) leads to the release of the second peptide.

Serine proteases is a group of enzymes that hydrolyze peptide bonds within a polypeptide chain. Serine proteases are widely distributed in nature and found in all three domains of life. The accepted reaction mechanism of serine proteases, as presented in Figure 1.3, proceeds in two steps. In the first step, Ser195 is deprotonated by His57, yielding a nucleophilic Ser195 residue that attacks the carbon of the substrates scissile bond. Ser195, His57 and Asp102 is known as the *catalytic triad*. The role of Asp102 is to stabilize His57 through hydrogen bonding interactions. The result of the nucleophilic attack is the formation of an intermediate state where the 'attacked' carbon has acquired a tetrahedral configuration. The intermediate tetrahedral state is stabilized by the *oxyanion hole*, which is a small pocket formed by the main-chain NH groups of Ser195 and Gly193. The NH groups are able to form hydrogen bond to the negatively charged oxygen atom bonded to the sp^3 carbon. The first reaction step is concluded when the sp^3 hy-

bridizated carbon returns to the sp^2 configuration and the adjacent nitrogen is protonated by the positive His57 residue, thus leading to cleavage of the scissile bond. The second step involves another tetrahedral transition state formed by a nearby water molecule attacking the electrophilic carbonyl carbon. The His57 residue is simultaneously protonated by the water molecule. Finally the second peptide is released from the enzyme by the Ser195-substrate bond breaking.

The active site of serine proteases is comprised of several subsite binding pockets that governs the specificity of the enzyme.[23] The naming convention for the binding pockets is presented in Figure 1.4. The most prominent binding site is named the S1 pocket, also called the *specificity pocket*. The specificity pocket lies in the active site and is oriented such that the residue adjacent to the peptide bond to be cleaved (adjacent to the carbon) is directed towards the pocket. The composition of the specificity pocket varies between enzymes and it determines whether the adjacent residue fits into the pocket (which promote cleavage) or not. For trypsin, there is a Asp residue present in the bottom of the pocket which makes the pocket complementary to positive side-chains. The pocket of chymotrypsin is designed for bulky aromatic side-chains because of the Ser residue located at the bottom and small residues like glycine lining the sides of the pocket.

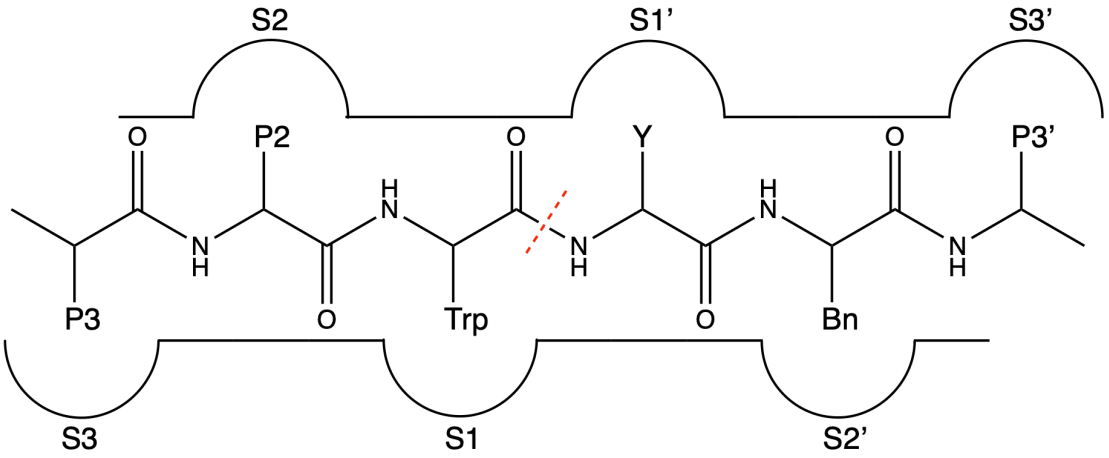


Figure 1.4: Subsites of the active site of serine proteases. Nomenclature on the binding pockets (S1-S3 and S1'-S3') and the substrate residues (P1-P3 and P1'-P3') is adopted from Schechter *et al.*[23] Red wiggly line indicate the site of cleavage (scissile bond).

CAPs are good substrates for both trypsin and chymotrypsin since their cationic and hydrophobic nature coincides with the specificity of the two enzymes. Interactions between CAPs and the binding pockets of trypsin and chymotrypsin have been thoroughly investigated by researchers at the Department of Chemistry at UiT in hope of increasing the proteolytic stability of CAPs. The various CAPs investigated are all tripeptides modelled after the tripeptide Arg-Trp-Arg-NHBn, where NHBn is the abbreviation of the hydrophobic capping NH-benzyl. The tripeptide Arg-Trp-Arg-NHBn show antimicrobial activity but it lacks tryptic and chymotryptic stability.

In 2010 Karstad *et al.* studied *in vitro* the chymotryptic stability of CAPs by substituting C-terminal Arg residue.[22] A library of CAPs was prepared where each CAP was of the design Arg-Trp-Y-NHBn, where Y had been substituted and the rest of the structure was kept constant. The prepared CAPs were subjected to chymotryptic degradation studies by testing the CAPs half-lives. Using this strategy, the researches were able to investigate the specificity of the S1' pocket which is the pocket residue Y is expected to bind to. The Y element was substituted by six different residues, including Har, Lys, Agp and App which are found in Figure 1.2. Har, the longer Arg analog, had a shorter half-life than the model CAP which means that the stability was decreased. In contrast, the short Arg analog Agp was shown to be stable against degradation. The conclusion was thus that stability decreased with increasingly long guanidine residue in the P1' position. The same trend was seen for the incorporation of a shorter primary amine. The CAP substituted with Lys was shown to be unstable while the incorporation of a short Lys analog increased the stability. Finally the incorporation of a naturally charged residue, App, was shown to yield a stable CAP.

1.5 Computational Tools

1.5.1 Molecular Docking

Molecular docking is a computational method that attempts to give accurate ligand-receptor complex predictions. Docking can be separated into two problems, 1) generating structures and 2) evaluating the structures to find the most

plausible result.

Structures are generated by using sampling algorithms. Various approximations in the sampling algorithms is commonly used as sampling all possible confirmations is too big of a computational task. DOCK is an example of an early program that ignores all degrees of freedom except for the six degrees of translational and rotational freedom of the ligand relative to the receptor.[24] Other algorithms like Monte Carlo[25] include the conformational degrees of freedom of the ligand to achieve better sampling. Molecular dynamics can also be used when flexibility of both ligand and protein is required, but the method is best for local minima search only as high-energy barriers could prevent proper conformation search.[26]

Scoring functions can be divided into three classes: (1) force field-based functions, which are based on existing force fields and parameter sets, (2) empirical-based where the binding energy is composed of individual uncorrelated terms fitted to reproduce experimental data, and (3) knowledge-based where the function is designed to reproduce experimental structures.[26] The scoring function has to rank all generated poses, which makes efficiency important. The scoring function should also refrain from generating false positives, i.e. poses which falsely have been attributed a good docking score.

Glide is a molecular docking program that aims for a complete conformational search by performing multiple docking steps.[27][28] Glide is grid-based, meaning a grid with evenly spaced site points is used to represent the shape and properties of the receptor. The use of a grid simplifies interaction energy calculations by using precomputed site point-receptor surface distances. The first step of Glide is an exhaustive systematic conformational search to limit the number of poses to be investigated further. The second step is a minimization step of the ligand of the kept poses using a molecular mechanics force field. In the third and final step, the poses of lowest energy are subjected to Monte Carlo sampling. Final scoring of poses are performed using GlideScore, an empirical scoring function. The Extra Precision (XP) Glide protocol uses the following scoring function[29]:

$$XPGlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \tag{1.1}$$

where E_{coul} and E_{vdW} are coulumb and van der Waals interaction energy respectively, E_{bind} is the sum of other favorable binding energy terms and $E_{penalty}$ is the sum of energy terms that obstruct binding. E_{bind} and $E_{penalty}$ is made up of the following energy terms:

$$E_{bind} = E_{phobic\ pair} + E_{hyd\ enclos} + E_{hbnn\ motif} + E_{hbcc\ motif} + E_{PI} + E_{hb\ pair} \quad (1.2)$$

$$E_{penalty} = E_{desolv} + E_{lig\ str} \quad (1.3)$$

$E_{phobic\ pair}$ is the pair-wise interaction between lipophilic ligand and protein atoms and $E_{hyd\ enclos}$ is an energy contribution arising from a hydrophobic enclose on lipophilic ligands, a situation which is believed to contribute to the free energy beyond that of $E_{phobic\ pair}$. $E_{hb\ pair}$ rewards hydrogen bonding pairs and $E_{hbnn\ motif}$ and $E_{hbcc\ motif}$ are for special H-bond motifs which give additional contribution to binding free energy. E_{PI} represent energy contribution from pi stacking and pi-cation interaction. Penalizing terms E_{desolv} and $E_{lig\ str}$ represent desolvation penalty (using a crude explicit water model) and energy strain that arises when a ligand has to adjust to a rigid cavity respectively.

1.5.2 Molecular Mechanics

Molecular modelling describes all methods of investigating molecular structure and function using modelling techniques. Computational resources are undoubtedly a powerful tool in molecular modelling, both as an aid for visualization and to accelerate computational needs. In molecular mechanics (MM), the molecular system follows the laws of classical mechanics. The atom is considered the smallest particle of the system and is modelled as a sphere with a given radius and point charge. This is not an exact description but it allows for larger biomolecular systems to be studied within reasonable computation time. The MM approach rely on the Born-Oppenheimer approximation which states that the nuclear and electronic motion can be separated due to the vast difference in mass between the nuclei and the electron. Because the atom is considered the smallest particle, the electronic motion is ignored and the energy of the system is calculated as a function

of the atomic positions only. Any change in the energy of the system corresponds to movement on a potential energy surface associated with the electronic ground state.

The potential energy is calculated using an empirical energy function referred to as a force field. The following theory on force fields is retrieved from *Molecular modelling: principles and applications* by A.R. Leach.[30] The force field defines the functional form and the parameters used to calculate the potential energy of a system. There are several force fields to choose from. Generally, all force fields consists of five energy terms derived from simple models of interaction. A typical functional form for a force field is:

$$\begin{aligned} U(\mathbf{r}^N) = &\sum_{Bond} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angle} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{tors.} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) \\ &+ \sum_{im.tors.} \frac{k_i}{2}(\zeta_i - \zeta_{i,0})^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \end{aligned} \tag{1.4}$$

The first term is bond stretching energy, the second is angle bending energy and the third is torsion energy. The fourth term is the improper torsion energy, a term introduced to assure correct conformation and the last term is energy arising from non-bonded interactions.

As stated previously, force fields are empirical functions, meaning no force field has the correct functional form nor the correct parameters to represent the true energy. Force fields are often designed for an intended purpose and its complexity reflects the accuracy needed and the acceptable computational time. For example, the MM4 force field is designed for alkanes and cycloalkanes and aims for accurate description of molecular properties including vibrational frequencies.[31] MM4 employs a complicated functional form including cross-terms that represent coupling between internal coordinates. OPLS on the other hand, is a force field designed for macromolecules in liquid simulation.[32] It uses a simple functional form (with no cross terms) similar to Equation 1.4 to lessen the computational efforts required for the large number of atoms.

Except for very specific purpose force fields, a force field has to be *transferable*.

Transferability means that the parameters used give reliable results for a series of related molecules. Without it the use of the force field gets limited and a new set of parameters is required for every new molecule in question.

Bond Stretch Energy

The bond stretching energy defines the energy arising from stretching the bond between two atoms. In the simplest model it is written as a harmonic potential as seen in Equation 1.5.

$$E_{bond}(l) = \frac{k}{2}(l - l_0)^2 \tag{1.5}$$

where k is the force constant that determines the stiffness of the harmonic oscillator, l is the bond length and l_0 is the reference bond length. Any deviation of l from the value of l_0 will give a quadratic increase of the potential. As a consequence of the simple model used, the potential moves towards infinity as the bond length increases. Certainly, the function should converge towards the dissociation energy to model the the correct behavior. The Morse potential is a useful model for describing correct function limit behavior. As seen in the Morse potential described in Equation 1.6, the dissociation energy D is part of the equation:

$$E_{Morse}(l) = D\big(1 - e^{-\alpha(l-l_0)}\big)^2, \quad \alpha = \sqrt{\frac{k}{2D}} \tag{1.6}$$

The Morse potential is a more accurate model than the harmonic potential, but solving the Morse equation requires more computational time. The harmonic potential is frequently used by force fields, OPLS for example, because bonds are not expected to deviate significantly from the equilibrium bond length. As long as the displacement from the minimum is small, the harmonic potential is a reasonable approach.

Angle Bending Energy

Angle bending like bond stretching is often approximated by the harmonic potential:

$$E_{angle}(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \tag{1.7}$$

The harmonic potential is in fact a Taylor expansion terminated at the second order, meaning it can also be written as $E_{angle}(\theta) = E(0)+\frac{dE}{d\theta}(\theta-\theta_0)+\frac{1}{2}\frac{d^2E}{d\theta^2}(\theta-\theta_0)^2$. The first term of the Taylor expansion is ignored since its only the zero point on the energy scale. The second term is also ignored as the derivatives are evaluated at the equilibrium and the first derivative of a extremum is always zero. The approximation can be improved by including higher order terms, but this naturally comes with computational costs and the problem of evaluating higher order force constants.

Torsion Energy

Torsion energy is the energy associated with the rotation around the central bond in a four atom sequence as measured by the angle formed by the two outermost bonds. The energy function has to describe the periodicity of the bond; a rotation of 360° of any torsional bond will render the same conformation. Some torsional angles may possess a periodicity of 120°, like all torsional angles found in ethane. A Fourier series as seen in Equation 1.8 can be used to model the torsion energy associated with torsion angle ω.

$$E_{tors}(\omega) = \sum_{n=1} V_n cos(n\omega) \tag{1.8}$$

V_n is a constant that effect the energy barrier of rotation and n is the multiplicity of the function and gives the number of energy minima as the angle increases from 0° to 360°.

Improper Torsion

Improper torsions are introduced to achieve correct planar arrangement in cases where a force field would treat a structure incorrectly. One approach is to intro-

duce an improper torsion term between four atoms which are not connected by a 'normal' sequence (1-2-3-4) but by an improper one (1-2(-3)-4, meaning atoms 1, 3 and 4 are all bonded to 2). The functional form of this energy term is a torsional potential that maintains the torsion angle at 0° or 180° (Equation 1.9). Improper torsions are used in planar systems like aromatic rings or peptide bonds which possess a double bond character.

$$E_{im.tors}(\omega) = k(1 - \cos 2\omega) \tag{1.9}$$

Non-Bonded Interactions

The non-bonded interactions is usually separated into two components, electrostatic and van der Waals contributions. The electrostatic interaction is modelled by point charge electrostatics between pairs of non-bonded particles within the system using Coulomb's law:

$$E_{elec}(r_{ij}) = \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.10}$$

where q_i and q_j are the assigned partial charges of particle i and j, respectively, r_{ij} is the distance between particle i and j and ϵ_0 is the electric constant of vacuum. The particles within a molecule are assigned partial charges to reproduce the unequal distribution of charge found in a molecule. This approach is the lowest of approximations for modelling the charge distributions. The representation of electronic potential can be improved by including higher-order electric moments (dipole, multipole, etc.).

The van der Waals energy describes the dispersive and repulsive interactions between non-bonded atoms. The Lennard-Jones 6-12 potential is a simple, empirical model for describing the van der Waals energy between two atoms i and j with distance r_{ij}:

$$E_{vdW}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{1.11}$$

The collision diameter σ is the distance where the potential between the atoms is zero and ϵ is the well depth of the energy minimum of the potential. The repulsive

part is the first term and the dispersive part is the last term. At large distances the dispersive part will dominate due to the r^{-6} dependency. As the distance decreases the repulsive part, with a r^{-12} dependency, will become increasingly dominant.

1.5.3 Simulation Methods and Molecular Dynamics

Statistical Mechanics

Macroscopic variables refers to experimentally measurable quantities, such as energy, that can be obtained in the lab. The computational effort needed to model a macroscopically sized system, from which to predict macroscopic variables, is too large for the task to be doable. Simulation methods are, however, able to predict macroscopic variables by considering the microstate of a smaller, microscopic model of the system in question and applying statistical mechanics.

The microstate of a system specifies the position $\mathbf{r}$ and momenta $\mathbf{p}$ for every particle present. If there are N particles, there are 6N variables in total ($\mathbf{r}$ and $\mathbf{p}$ are both three-variable vectors) needed to define the microstate. The term 'phase space' is used to describe an abstract space in the 6N-dimension in which every point refers to a single microstate. Simulation techniques are used to generate representative (microstate) configurations of a system, a task also referred to as 'sampling the phase space'. A collection of configurations in which all configurations are representative of a given macrostate is called an ensemble. In statistical mechanics, an average experimental observable is defined as the ensemble average. In other words, the observed average value of a property A seen in the lab is equal to the value obtained by calculating the average of property A over all configurations belonging to the ensemble. The ensemble average of a property A is defined as a 6N-integral equation, denoted as a double integral for convenience:

$$\langle A \rangle_{ensemble} = \iint A(\mathbf{p}^N, \mathbf{r}^N)\rho(\mathbf{p}^N\mathbf{r}^N)\ d\mathbf{p}^N d\mathbf{r}^N \tag{1.12}$$

where A is the observable property and ρ is the probability of the given configuration.

One important postulate in statistical mechanics is the Ergodic hypothesis which states that the ensemble average defined in Equation 1.12 is equal to the

time average, which is the average measurement of property A in a single system that evolves in time indefinitely:

$$\langle A \rangle_{time} = \lim_{\tau -> \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\big(\mathbf{p}^N(t), \mathbf{r}^N(t)\big)\ dt \tag{1.13}$$

This postulate is of great importance for the simulation technique called molecular dynamics (MD) as it enables MD to predict thermodynamic properties. MD generates configurations by mimicking the dynamic evolution of a single system to generate a *trajectory*, a set of configuration of the same system successive in time. An estimate of the time average (Equation 1.13) of a thermodynamic property A can be obtained from a MD simulation with sufficient sampling size (that is, an appropriate large number of configurations, M):

$$\langle A \rangle_{time} \approx \frac{1}{M} \sum_{t=1}^{M} A(\mathbf{p}^N, \mathbf{r}^N) \tag{1.14}$$

Molecular Dynamics

MD treats the system with MM and generates the trajectory of a system evolving in time. The trajectory defines how the velocity and position of each particle vary with time, and this information is obtained by applying Newton's second law of motion on each of the particles to obtain the acceleration vector $\mathbf{a}_i$ on each particle i:

$$\mathbf{F}_i = m_i \mathbf{a}_i \tag{1.15}$$

m_i is the particle mass and $\mathbf{F}_i$ is the force acting on particle and it is related to potential energy (which is obtained by a force field) by $-\nabla U(x) = F(x)$. Since the force acting on each particle is defined by the force field used, we have to look at the force field to understand what forces influence each particle. Using a force field defined by Equation 1.4, a particle will experience both intramolecular forces and non-bonded interactions. The force acting on a particle will change whenever either an interacting particle or the particle itself moves. This means the motion of all particles are coupled together and that the acceleration on each particle

is not constant. We use a finite difference method like the Verlet algorithm or variations of it to solve the equations of motion. The Verlet algorithm solves for the new atomic positions by breaking the dynamics into small steps with a fixed time duration δt. During each time step the acceleration on each particle is kept constant and the new position after each time step is determined by the current acceleration, time step duration and previous position, as seen in Equation 1.16.

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2 \mathbf{a}(t) \tag{1.16}$$

Equation 1.16 is obtained by summation of two Taylor series approximations, one for the position after the time step ($+\delta t$) and one for the position one time step earlier ($-\delta t$), both are seen in Equation 1.17. The first derivative of $\mathbf{r}$ is equal to the velocity $\mathbf{v}$ and the second derivative is equal to the acceleration $\mathbf{a}$.

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \tfrac{1}{2}\delta t \mathbf{a}(t) \tag{1.17a}$$
$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta t \mathbf{v}(t) + \tfrac{1}{2}\delta t \mathbf{a}(t) \tag{1.17b}$$

The size of the time step is crucial for the result of the MD simulation. The smaller the time step the better the approximation of a continuous trajectory. Decreasing the time step will however increase the computational effort of propagating the system forwards in time. In contrast, the continuous potential will not be accurately described if the time step is too big, and the infrequent time step update can lead to instabilities in the system arising from unnatural energy fluctuations. The velocity is not explicitly present in Equation 1.16. One of the Verlet variations is the leap-frog algorithm which is the algorithm used by the MD program package Q[33]. The leap-frog uses the following equations to solve for position and velocity:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t+\tfrac{1}{2}\delta t) \tag{1.18a}$$
$$\mathbf{v}(t+\tfrac{1}{2}\delta t) = \mathbf{v}(t-\tfrac{1}{2}\delta t) + \delta t \mathbf{a}(t) \tag{1.18b}$$

The algorithm goes as follows: (1) for every particle i; retrieve force acting on particle using the force field, (2) update velocity of particle i using Equation

1.18b, and last step (3) update position of particle i using Equation 1.18a. As seen in Equation 1.18b, the method has to be assigned starting velocities for every particle. This is usually done by random assignment of velocities from a Maxwell-Boltzmann distribution of velocities at a given temperature T. Unlike the classic Verlet algorithm, this approach explicitly defines the velocity. The total energy of a system includes both potential and kinetic energy, the latter of which is calculated from the velocities of all particles. Scaling the velocity is also a way of temperature control of the system, as the temperature is proportional to the kinetic energy.

Solvent Treatment and Boundary Conditions

Biomolecules in our body exists in a solvated state. To study the biomolecule of interest in vacuum is therefore not beneficial if the aim is to gather relevant biological results.

Simulations of biomolecules emerged in liquid water are more likely to replicate the correct biological behavior, but this entails that a large number of water molecules have to be included in the simulation. Different water models, at different computational costs, are used to include water in simulations. In simple water models, water molecules are kept rigid (meaning no bond stretching or angle bending) and interactions with water molecules are limited to pairwise Coulombic and Lennard-Jones expressions. TIP3P is an example of a simple water model.[34] In TIP3P all three atoms interact with electrostatic interactions while only one atom, the oxygen atom, partake in van der Waal interactions. The use of rigid water molecules and neglect of van der Waals interactions involving the hydrogens will lower the accuracy of the simulation but speed up the computation time. More advanced water models may be applied if needed, like water models that allow flexibility in the molecule and/or incorporates polarization effects.

Another challenge for simulation methods is the correct treatment of boundaries. As mentioned earlier, macroscopically sized systems are not feasible to simulate— the number of particles included in the simulation is limited by the computational resources available. To correctly model bulk properties, the particles should not sense the presence of a boundary at all. One approach is using a liquid droplet, which adds a spherical layer of explicit water molecules around the

biomolecule of interest and treats the space beyond the boundary as a continuum with a dielectric constant ε. The closer a given particle is to the center of the sphere, the less it is influenced by the boundary. The atoms inside the boundary are included in the simulation and are affected by a potential meant to mimic the potential of an infinite system. If the system posses a net charge then the effective potential should reflect the systems polarization on the continuum medium, which in turn affect the system. In such a case the Born equation can be used:

$$\Delta G_{elec}(q) = -\left(1 - \frac{1}{\varepsilon}\right)\frac{q^2}{2R_c} \tag{1.19}$$

where q is the net charge of the system, ε is the dielectric of the continuum and R_c is the radius of the water droplet.

1.5.4 The Linear Interaction Energy

The linear interaction energy (LIE) is a method developed in 1994 by Åqvist *et al.* for estimating the binding affinity of ligands in ligand-protein complexes.[35] The LIE model assumes that the binding free energy can be expressed as the sum of polar and non-polar components of the binding free energy only:

$$\Delta G_{bind} = \Delta G_{bind}^{non-polar} + \Delta G_{bind}^{polar} \tag{1.20}$$

The binding free energy of a ligand bound to protein is estimated as the difference in solvation free energies between two systems, the ligand and the ligand-protein complex. This means that there are only two states that have to be sampled by simulation methods, the ligand in solvent (free state) and ligand bound to protein (bound state). The LIE equation is defined as:

$$\Delta G_{bind} = \alpha\Delta\langle U_{l-s}^{vdW}\rangle + \beta\Delta\langle U_{l-s}^{EL}\rangle + \gamma \tag{1.21}$$

$\Delta\langle U_{l-s}^{vdW}\rangle$ is the difference in ligand-surrounding van der Waals interactions between the two states ($\Delta\langle U_{l-s}^{vdW}\rangle = \langle U_{l-s}^{vdW}\rangle_{bound} - \langle U_{l-s}^{vdW}\rangle_{free}$) and $\Delta\langle U_{l-s}^{EL}\rangle$ is the difference in ligand-surrounding electrostatic interactions between the two states ($\Delta\langle U_{l-s}^{vdW}\rangle = \langle U_{l-s}^{EL}\rangle_{bound} - \langle U_{l-s}^{EL}\rangle_{free}$). α and β are coefficients that weights the non-polar and polar contribution, respectively. γ is a constant used to calculate

the absolute binding energy in a given system, and can be neglected for relative binding free energy.

According to Equation 1.21, the non-polar component of the binding free energy is estimated by van der Waal interaction energies and the polar component is estimated by electrostatic interaction energies. For the non-polar component, this approximation is justified by the fact that both solvation free energy of non-polar compounds and van der Waals ligand-surrounding interactions in MD simulations scale linearly with ligand size. Thus, the non-polar component can be defined as a linear function of U^{vdW}.

The polar component is estimated by the linear response theory, which gives the electrostatic free energy contribution to the solvation free energy as:

$$\Delta G^{el}_{sol} = \tfrac{1}{2}\big(\langle U^{el}_{l-s}\rangle_{on} + \langle U^{el}_{l-s}\rangle_{off}\big) \tag{1.22}$$

The subscripts off and on denotes whether the electrostatic interaction between ligand and surrounding environments is turned off or on. Neglecting the $U_{l-s,\ off}$ and accounting for possible deviations from the scaling factor using in linear response theory ($\frac{1}{2}$), we get polar component approximation seen in Equation 1.21

1.5.5 Free Energy Perturbation

The Gibbs free energy (ΔG) is an experimentally determinable quantity that reflects the spontaneity of a process at a given temperature and, in the contexts of enzyme-ligand binding reactions, the strength of the interaction. ΔG is determined by the interplay between the change of enthalpy (ΔH) and the change in entropy (ΔS) as seen in Equation 1.23. For a process to proceed spontaneously, and therefore be favorable, ΔG has to be negative. This is the case if there is both a release of heat and an increase in disorder of the system (negative ΔH and positive ΔS) or if either of the two quantities are significantly high enough to render ΔG negative. The entropy contribution is determined by the temperature (T), meaning a process may be favorable at a given temperature, but not favorable at any temperature.

$$\Delta G = \Delta H - T\Delta S \tag{1.23}$$

MD simulations will predominantly sample low energy regions of the phase space. Quantities like the internal energy can be estimated by MD since the quantity is predominantly influenced by low energy conformations. Free energy is a quantity that is significantly influenced by high energy contributions. This means that high energy regions of phase space has to be sampled for accurate free energy estimations. Sufficient sampling of high energy conformations is not possible with a finite computer simulation method. Fortunately it is possible to calculate the free energy difference between states using the free energy perturbation (FEP) method. FEP uses sampling methods such as MD in combination with Zwanzig formula to calculate the free energy difference between states. The Zwanzig formula, which gives the free energy difference between systems 1 and 2, is shown in Equation 1.24.

$$\Delta G_{1->2} = -\beta \; ln \; \langle \exp[-(U_2 - U_1)\beta^{-1}] \rangle_1 \tag{1.24}$$

where T is the temperature, U_1 and U_2 is the potential energy of system 1 and 2, respectively, $\beta = k_B T$ where k_B is the Boltzmann constant and $\langle .. \rangle_1$ denotes that the ensemble average is collected from state 1 conformations. The latter means that only phase space of state 1 is sampled and that potential energies for state 2 are estimated using conformations of state 1 in combination with state 2 potential energy parameters.

The sampling of state 2 will be acceptable if state 1 and state 2 have sufficient overlap in phase space. To ensure overlap, one is allowed to include a number of alchemical, intermediate states between states 1 and 2. The total free energy change of moving from state 1 to state 2 will be the sum of intermediate energy terms, as seen in Equation 1.25. Every intermediate U_m is the potential energy of a non-existing state m, a state defined as a linear combination of state 1 and 2. U_m is given by: $U_m = \lambda U_1 + (1 - \lambda)U_2$ where λ is a parameter in the range 0 to 1. The use of intermediate, alchemical states is valid since the free energy is a state function, which means that the free energy only depends on the initial and the final state.

$$\Delta G_{1->2} = -\beta \sum_{m=1}^{n-1} ln \left\langle \exp[-(U_{m+1} - U_m)\beta^{-1}] \right\rangle_m \tag{1.25}$$

1.5.6 Empirical Valence Bond Theory

The trajectory produced by MD contains dynamic information about a given system but since electron motion is ignored, no breaking or forming of bonds can occur. To study the process of chemical reactions one has to use other methods like the empirical valence bond theory (EVB) which model reacting bonds using a Morse potential. EVB is a semi-empirical QM/MM approach developed by Warshel in the 1980s as a mean to study the potential energy surface of enzymatic reactions.[36] EVB evaluates the potential energy surface within a valence bond (VB) framework. However, the Hamiltonian matrix elements are approximated by analytical potential energy functions instead of using standard quantum mechanical treatment. The matrix elements of the enzymatic reaction is obtained by first modelling an analogous reaction in solvent. The solution reaction is refered to as a *reference reaction* and is used to calibrate the enzymatic potential surface. This approximation allows for reliable sampling of the potential energy surface of complex enzymatic processes, which would otherwise be unattainable by QM treatment alone. Furthermore, using the same parameters for different systems enables us to study the direct effect of changing the chemical environment while eliminating systematic errors.

To study a chemical reaction, EVB utilises a set of VB configurations where each configuration corresponds to a physically meaningful state of the system, like the product state or reactant state. The wavefunction of the system will be a linear combination of the VB states. For a two-state EVB model, where ϕ_1 and ϕ_2 correspond to the reactant and product states, respectively, the wavefunction will be defined as shown in Equation 1.26:

$$\psi = c_1\phi_1 + c_2\phi \tag{1.26}$$

where c_1 and c_2 are coefficients balancing the linear combination. The Hamiltonian matrix for this wavefunction is a 2×2 matrix where the diagonal elements H_{11} and H_{22} are the energies corresponding to the two states as given by $H_{ii} = \langle\phi_i|\mathbf{H}|\phi_i\rangle$ and the off-diagonal elements H_{12} and H_{21} are quantum couplings of the two states. The lowest electronic eigenvalue E_g, which defines the potential energy surface, is obtained by solving the 2×2 secular equation shown in Equation 1.27:

$$\begin{vmatrix} H_{11} - E_g & H_{12} - E_g S_{12} \\ H_{21} - E_g S_{21} & H_{22} - E_g \end{vmatrix} = 0 \tag{1.27}$$

where S_{ij} is the overlap integral given by $S_{ij} = \langle \phi_i | \phi_j \rangle$. What makes EVB a semi-empirical method is the description of the matrix elements by empirical functions, as supposed to using standard quantum mechanical methods. H_{11}, H_{22} and H_{12} are approximated by the following empirical potential functions, where the energy associated with binding U^i_{bind} is described using a Morse potential (Equation 1.28):

$$H_{ii} = \epsilon_i \ = U^i_{bind} + U^i_{ang} + U^i_{tors} + U^i_{imp.t.} + U^i_{nb,rr} + U^i_{nb,rs} + U^i_{ss} + \alpha^{i \neq 1} \tag{1.28a}$$

$$H_{ij} = H_{ji} = A_{ij} \exp(-\mu_{ij} r_{kl}) \tag{1.28b}$$

U^i_{ang}, U^i_{tors}, $U^i_{imp.t.}$ and U^i_{nb} are the angle, torsion, improper torsion and non-bonding potential energies respectively. r and s subscripts stands for reacting and surrounding fragments and U^i_{ss} encompass all interaction in between surrounding fragments. α is the gas-phase energy, which is the energy of the state when all fragments of the configuration are separated at infinite separation. The gas-phase term used is a constant that gives the gas-phase energy relative to state $i = 1$. A_{ij} and μ_{ij} are fitting parameters of the quantum coupling between the two states and r_{kl} is the distance between two atoms relevant for the reaction, usually two atoms involved in bond breaking or bond formation.

The nature of $\alpha^{i \neq 1}$ and H_{ij} is an important feature of the EVB approach. Once they have been calibrated by a reference reaction they can be used to describe the same reaction in a different chemical environment. That is, we can view the enzyme as an another solvent for the reaction, and if we have calibrated $\alpha^{i \neq 1}$ and H_{ij} for a suitable reaction in solvent, we can use the same $\alpha^{i \neq 1}$ and H_{ij} values to evaluate the enzymatic potential surface. The transferability of H_{ij} is an assumption made in the EVB approach, and the validation of the assumption is supported by constrained density functional theory studies of a S_N2 reaction where H_{ij} remained approximately the same upon moving from gas-phase and to solution.[37]

Neglecting the overlap integral S_{ij} by assuming the that the empirical de-

scription of H_{12} will account for the effect of the overlap integral, the solution to Equation 1.27 can be written as shown in Equation 1.29.

$$E_g = \frac{1}{2}\left[(\epsilon_1 + \epsilon_2) - \sqrt{(\epsilon_1 - \epsilon_2)^2 + 4H_{12}^2}\right] \tag{1.29}$$

The FEP approach described in Section 1.5.5 gives only the free energy difference between two states. EVB is able to calculate the free energy ΔG of moving on the actual ground-state potential E_g along a reaction coordinate and from which we can obtain the activation free energy $\Delta G^{\ddagger}$. Using the FEP/umbrella sampling expression defined by Equation 1.30 we can define ΔG. The reaction coordinate $\Delta\epsilon$ is defined as the energy gap between state 1 and 2 ($\Delta\epsilon = \epsilon_1 - \epsilon_2$) which is a highly useful definition since we avoid the problem of pre-definding the reaction path and we only have to deal with a one-dimensional coordinate.

$$\Delta G(\Delta\epsilon_n) = \frac{\sum\limits_{m \supset \Delta\epsilon_n} w_m \Big(\Delta G_m(\lambda_m) - \beta \ln\big\langle \exp[-\beta^{-1}(E_g(\Delta\epsilon_n) - \epsilon_m(\Delta\epsilon_n))] \big\rangle_m \Big)}{\sum\limits_{m \supset \Delta\epsilon_n} w_m} \tag{1.30}$$

where $\Delta\epsilon_n$ is the discretized reaction coordinate, ϵ_m is the mapping potential, $\Delta G_m(\lambda_m)$ is the free energy obtained by using Equation 1.25 on the mapping potential and the $\sum w_m / \sum w_m$ is for weighting the individual contributing vectors to the reaction coordinate interval such that they are weighted proportionally to the total contribution of the interval.

EVB is a sensible choice when extensive sampling of larger systems is crucial. Once the reference reaction is calibrated the same parameters can be used to study the potential surface of several chemical environments, as for example to study the effect of mutating an enzyme or the effect of changing the temperature. EVB has enabled researchers at UiT to shed light on the origin of catalytic efficiency of cold-adapted trypsin.[38] Isaksen and co-workers were able to produce high precision computational Arrhenius plots for the activation free energy versus temperature for trypsin, which can only be done by obtaining a sizable amount of free energy profiles at different temperatures.

1.6 Aim of Study

CAPs have the potential as new pharmaceuticals for oral-administration, but the proteolysis of CAPs in the gastrointestinal (GI) tract poses a challenge. The goal of this project is to model the interaction between CAPs and chymotrypsin using molecular modelling and computational tools. Chymotrypsin is a serine protease that presents a threat for CAPs in the GI tract due to its substrate affinity.

Four CAPs from the *in vitro* study of Karstad *et al.*[22] on the chymotryptic degradation of CAPs were selected for further investigations. The four CAPs are presented in in Figure 1.5.

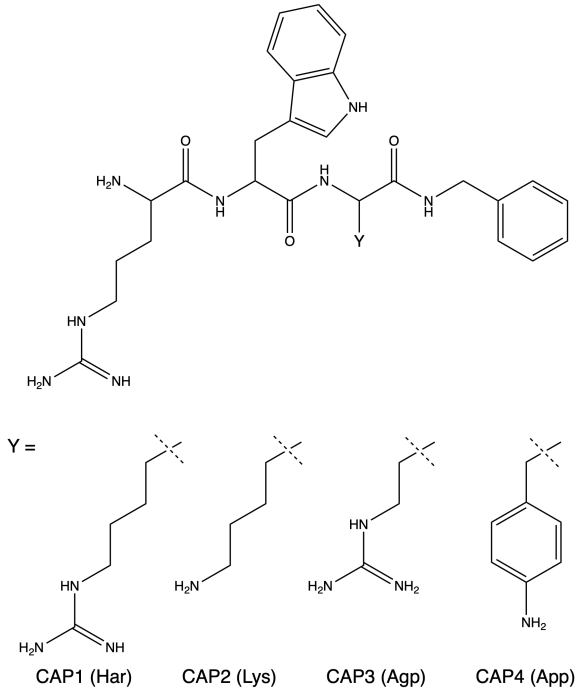


Figure 1.5: Structures of the four CAPs investigated in this project.

All CAPs share the same Arg-Trp-Y-NHBn structure, meaning only the Y residue differs between CAPs. The App residue possess no charge in physiological conditions, rendering CAP 4 with a total charge of +2. In physiological conditions CAP 1, 2 and 3 have an additional +1 charge in the Y residue, giving a total charge of +3. CAP 1 and 2 were found by the degradation study to have half-lifes of 1 and 4 hours, respectively. CAP 3 and 4 both had half-lifes of 24 hours or

more and were therefore considered stable against degradation. Using a set of stable and unstable CAPs may shed light on to why only some CAPs are found stable.

Specifically, the project intend to accomplish the following tasks:

1. Applying molecular docking in the preparation of CAP-chymotrypsin models.

2. Run MD simulation on CAP structures in both bound and unbound state and predict the binding free energy using the LIE method.

3. Parameterize a reference reaction analogous to the initial reaction step where a tetrahedral chymotrypsin-CAP intermediate is formed (reaction step 1a in Figure 1.3).

4. Perform EVB on the initial reaction step of the catalytic mechanism to predict the reaction free energy surface.

5. Employ a modified LIE method to investigate the stabilization of the transition state relative to the reactant state.

2 Methods

2.1 Protein Preparation and Docking

A crystal structure of the P1 Trp mutant bovine pancreatic trypsin inhibitor (BPTI) chymotrypsin complex (PDB ID: 1T8O) was used to obtain coordinates for chymotrypsin.[39] The structure was prepared using the Protein Preparation Wizard in Maestro.[40] In the preprocess stage, hydrogens were added, bond orders were assigned and water molecules beyond 5 Å from heavy atoms were deleted. In the refine menu the Interactive Optimizer option was selected, and Analyse Network was used with PROPKA[41] pK_a prediction (pH 7.0) and with current orientations selected. All residues were then optimized. His57 were locked in HID state and the orientation of Ser195 hydroxyl group was locked into a position facing His57 (Figure 2.1).

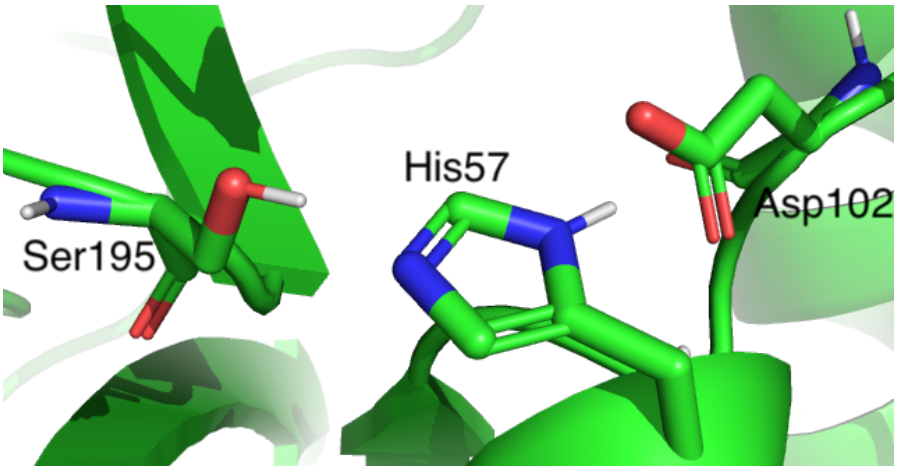


Figure 2.1: Orientation of Ser195, His57 and Asp102 in the active site of chymotrypsin after protein preparation.

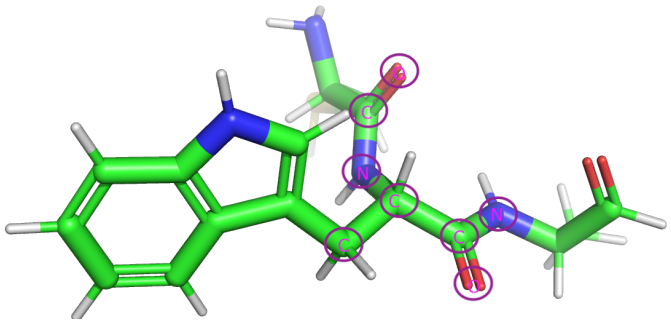


Figure 2.2: Structure of BPTI residues 14-16 (sequence Cys-Trp-Ala). Atoms marked with red circles were used to define the core pattern used in constrained docking.

CAP 2 was then build using residues 14-16 of BPTI as a starting point. Residue 15 (Trp), was left unaltered, residue 14 was modified into an Arg residue, residue 16 was modified into Lys and a NHBn capping was added to the C-terminal.

Glide[27][28] docking of CAP 2 was performed by first generating the receptor grid for chymotrypsin using the truncated (residue 14-16) BPTI as ligand molecule. A grid was generated with van der Waals radius scaling of 1.0 and a partial charge cutoff of 0.25. Flexible ligand docking was performed using XP precision with a van der Waals radii scaling factor of 0.80 and a partial charge cutoff of 0.15. Docking was constrained by defining an 8 atom core pattern (Figure 2.2) of the ligand molecule with a 0.5 Å tolerance. The docking was set to sample ring conformations. Non-planar conformations of amides were penalized and Epik[42] state penalties were added to the docking score. The initial phase was set to screen 10 000 poses. The scoring window was set to 100 kcal/mol and the number of poses to keep after the initial phase was set to 1000. Energy minimization was then performed with OPLS3e[43], dielectric constant of 2.0 and a maximum number of steps of 100. The best 100 poses were kept for post-docking minimization and the threshold for rejection was set to 0.50 kcal/mol.

After docking of CAP 2, CAP 1, 3 and 4 were prepared using the docked CAP 2 structure as a starting point. The Lys residue of CAP 2 was manually modified into Har, Agp and App for CAP 1, 3 and 4, respectively. The manually prepared CAP 1, 3 and 4 structures were scored in-place using Glide XP[29] and the previously generated grid. Using the in-place only option means no ligand sampling or minimization steps were performed.

2.2 Molecular Dynamics

All MD simulations performed in this project were carried out using Q, a molecular dynamics package consisting of four programs, Qprep, Qdyn, Qfep and Qcalc.[33] The program package is principally designed as a tool for LIE calculations, FEP simulations and EVB calculations. An overview of the workflow of a free energy calculation using Q is shown i Figure 2.3.

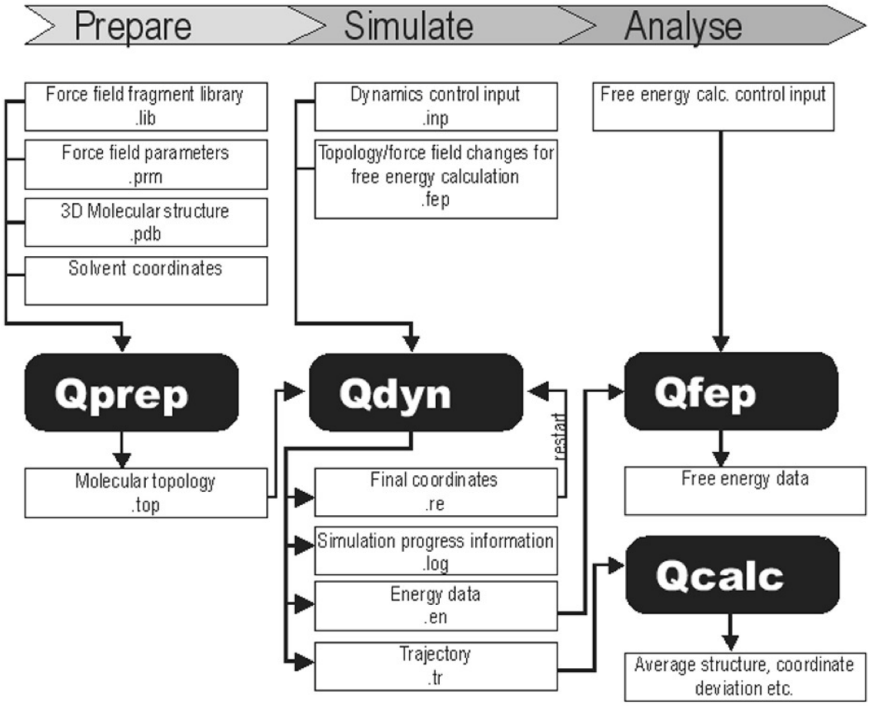


Figure 2.3: General workflow for free energy calculations with Q. The four Q programs are represented by black boxes. White boxes represent files. The arrows show how the files are related to the programs (either as input or output files). Figure is adapted from *Manual for the Molecular Dynamics Package Q v 5.06*[44].

The workflow can be separated into three steps, prepare, simulate and analyse. The first step involves the preparation of a molecular topology by using Qprep. A molecular topology is a file containing all information on the molecular system needed to run MD simulations with Q. The next step is the main part of Q where Qdyn carries out the MD simulation on the prepared system. The last step is Qfep, a program that uses the data produced by Qdyn to calculate free energies.

Qcalc is an optional step for both FEP/EVB and LIE and can be used as a tool to calculate average structures, average distances between atoms, etc.

Qgui is a graphical user interface developed to make the use of Q more accessible for novice users as well as to simplify the workflow and to make the use of Q more efficient.[45] The main window of Qgui is illustrated in Figure 2.4.

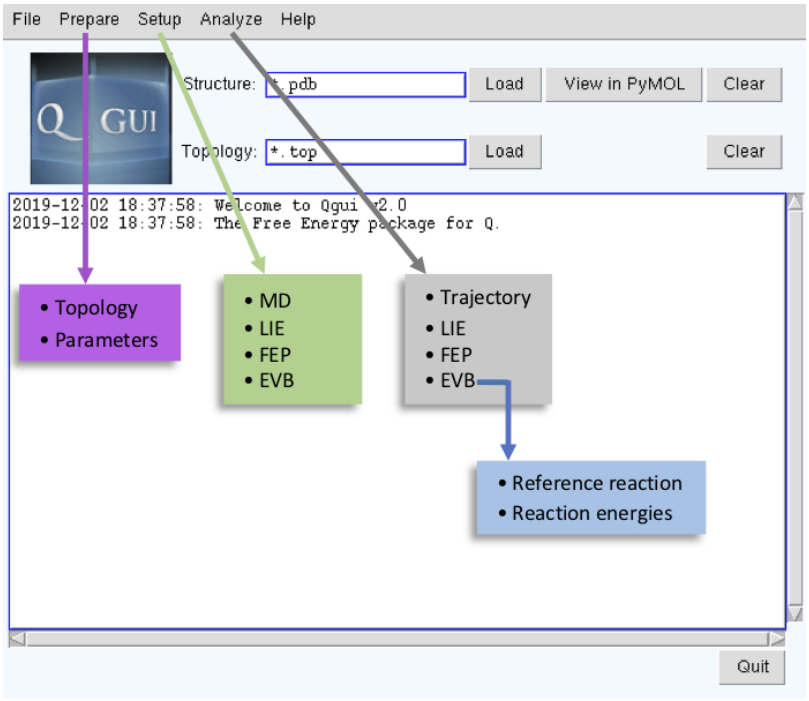


Figure 2.4: Main window of Qgui. Some of the build-in tools available are shown in the figure.

As seen in Figure 2.3, each step of the Q workflow requires additional input files which has to be prepared by the user. Qgui provides the user with easy preparation of all input data necessary for the use of Q. The remainder of this subsection will be used to highlight some of Qgui's features.

Prepare

The preparation of a topology can be done using Qgui. If the user wishes to add disulfide bonds, Qgui will automatically identify pairs of sulfate atoms which form bonds. Other features are the toggling of residue charges, defining simula-

tion sphere radius and selection of simulation sphere center. Qgui is paired with PyMOL[46] which can be used for visualizing the settings made.

Qgui not only ease the use of Qprep, but also provides an automated scheme to generate missing force field parameters and library files using the OPLS2005 force field[47]. The parametrization utilizes the ffld server utility which belongs to Maestro.[40]

Setup

As illustrated in Figure 2.4, the setup dropdown menu includes support for MD, LIE, FEP and EVB. There are two challenges that emerge for the novice user when preparing the FEP file. The first is how to define the state(s) used in the transformation and, the second is how to assign missing force field parameters to the newly defines state(s). Qgui is paired with PyMOL[46] which allows for visualization of the FEP transformation as well as simplified modelling of the states using a PyMOL window. The program then offers automatic assignments of force field parameters to the newly defined state(s).

Analyse

Crucial to the EVB method is the parametrization of a reference reaction by manually adjusting $\alpha^{i\neq1}$ and H_{ij} until the target free energies are reproduced. Qgui automatizes the calibration of $\alpha^{i\neq1}$ and H_{ij} for a two-state EVB. The user is only required to provide the target $\Delta G^{\ddagger}$ and ΔG_0 values of the reference reaction.

EVB is a method that in many cases requires extensive sampling. A way of efficiently processing the results will make the EVB method more applicable for extended studies. Qgui provides a high-throughput analysis of reaction free energies. The Qgui interface is designed to analyse any number of sets of EVB simulations simultaneously. This means that one can easily analyse and compare different simulations at once, whether it is the same system at different temperatures or different systems all together.

2.3 Binding Affinity of CAPs with LIE

Binding free energies were investigated using the LIE method. Eight systems were subjected to MD simulations in total, one for the free state (CAP free in water) and one for the bound state (bound to solvated protein) for each of the four CAPs. Consequently, eight topologies were prepared. For each CAP, the two topologies had to be prepared such that the MD simulation for bound and free state proceeded under identical boundary conditions. Besides having identical sphere center and boundary radius, the net charge of the two spheres has to be equal. The net charge will affect the Born term applied to the system (Equation 1.19), and the term will be unequal if the net charge is different. The net charge of the bound state may differ from the free state due to charges present in the protein. Chymotrypsin has a net charge of +3 which was neutralized by deprotonating the side-chains of Lys82, Lys202 and Lys203, which were found to be the three most distant positively charged residues in respect to the simulation center. The α-carbon of Lys82, Lys202 and Lys203 has a distance of 25.8, 24.4 and 26.6 Å from the simulation center.

2.3.1 Preparing Topology

Force field parameters and library files for the unnatural residues Har, Agp and App, as well as the hydrophobic capping NHBn, were prepared using Qgui[45] and the Maestro ffld-server[40]. The coordinates for the bound state and the free state were retrieved from the docking results. Qprep[33] and OPLS2005[47] were used to prepare the topologies. Both free and bound state were emerged in a water droplet where the carbonyl carbon was set as the center of a sphere with radius of 33 Å. Water molecules were added to the simulation sphere using the TIP3P[34] model.

2.3.2 MD Simulation Details

The preparation of the parameters for controlling the MD simulations and the FEP file was prepared using the *setup LIE* tool in Qgui[45]. Prior to the production phase, the system was heated stepwise from 1 K to 300 K using short simulation of 51 ps and followed by 1 ns of equilibrium at 300 K. The system was coupled to

a temperature bath with a relaxation time of 10 fs. 2 500 000 steps with stepsize 1.0 fs were performed for each simulation, and with four parallels (with random starting seed) the total simulation time for each of the eight systems were 10 ns. Bonds and angles of the solvent were constrained using SHAKE[48]. Non-bonded interactions were given a cut off distance of 10 Å between solvent-solvent, solute-solvent and solute-solute atoms. No cut off for Q-atoms. Energies were printed out every 5 steps.

2.4 Free Energies with EVB

The catalytic mechanism of chymotrypsin proceeds, as described in Section 1.4, in two steps. The initial nucleophilic attack by Ser195 (step 1a in Figure 1.3) was investigated by a two-state EVB. The structures from the docking results will be used as EVB state 1 and the tetrahedral intermediate resulting after the first catalytic step will be EVB state 2.

2.4.1 Preparing Topology

A new topology was prepared for EVB. Coordinates for each chymotrypsin-CAP complex were retrieved from the docking results. Qprep[33] and OPLS2005[47] were used to prepare four topologies, one for each complex. All complexes were emerged in a water droplet where the carbonyl carbon of Trp was selected as simulation center and the spherical boundary radius was set to 33 Å (whole protein soluted). Additional water was added to the simulation sphere using the TIP3P[34] model.

It was decided to use only one reference reaction for all four complexes. That is, the free energy profile obtained by EVB for all four complexes will be based on the same $\alpha^{i\neq 1}$ and H_{ij} values. To prepare the reference system, coordinates of His57, Ser195 and CAP 1 were retrieved from the CAP 1-chymotrypsin complex. His57 and Ser195 were chopped, His57 such that imidazol and β-carbon were left, and Ser195 such that the hydroxyl, β-carbon and α-carbon were left. CAP 1 was truncated to a scaffold molecule common to all the four CAPs. Hydrogens were added to the β-carbon of His57, the α-carbon of Ser195 and to the four terminal carbon atoms of the truncated CAP (8 hydrogens in total for all three fragments).

The reference system prior to the addition of hydrogens was identical to the EVB state 1 system in Figure 2.5. Force field parameters and library file for the three fragments of the reference system was prepared using Qgui[45]. A topology was prepared with Qprep[33] with OPLS2005[47] as force field, using a 22 Å radius simulation sphere and the carbonyl carbon of Trp as simulation center. Water was added using TIP3P[34].

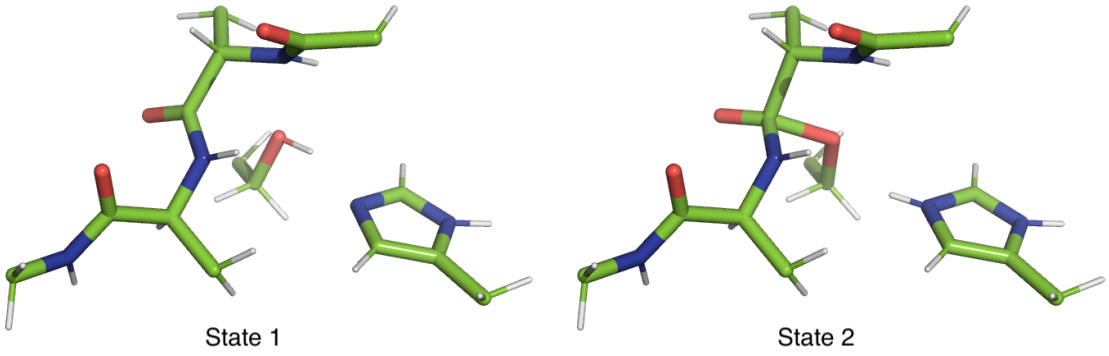


Figure 2.5: The Q-atoms selected and how their configuration change by the FEP transformation. Left figure is the Q-atoms in state 1. Right figure is the Q-atoms in state 2.

2.4.2 Preparation of Dynamic Control Input and FEP File

The two-state EVB was prepared using the *Setup EVB* tool in Qgui[45] and the exact same setup was performed for all four enzyme-CAP complexes and for the reference system. The atoms to be transformed are called Q-atoms and a total of 45 atoms were chosen, all of which are present in all the systems (Figure 2.5). Three bonds were added to the *form/break bond* window. Each state has to be assigned a value of 1 or 0 to mark whether the bond is present (1 = true, 0 = false). The three assigned bonds are; the O-H bond in Ser195 ($state_1$=1, $state_2$=0), the O-C bond between oxygen in Ser195 and the carbonyl carbon of Trp ($state_1$=0, $state_2$=1) and the N-H bond between the pyridine-like nitrogen of His57 and the hydrogen of the hydroxyl group of Ser195 ($state_1$=0, $state_2$=1). The *Sync pyMOL* option was selected, which displays both states in a pyMOL[46] window and allows the user to define the states visually by using edit mode in pyMOL. In state 2 the protonated imidazole ring of His57 had to be corrected to display a planar

conformation and the carbon of Trp involved in bond formation had to be corrected to display a tetrahedral conformation. The auto-assign option was then used, and after verifying the correct state 2 system, auto-assign successfully assigned all state dependent force field parameters missing except for the off-diagonal parameters, which was manually assigned to be the oxygen of Ser195 and the carbonyl carbon of Trp. The Q-atoms selected and the two EVB states are displayed in Figure 2.5.

2.4.3 MD Simulation Details

Prior to the production phase, the system was heated stepwise from 1 K to 300 K in a matter of 51 ps and followed by 1 ns of equilibrium at 300 K. A temperature bath coupling with relaxation time of 100 fs was used. The simulation ran in reversed order, meaning the intermediate FEP states drove the system from state 2 to state 1. 51 intermediate FEP states were used and each proceeded for 10 000 steps using step size 1.0 fs. 10 parallels (with random starting seed) were executed, giving a total simulation time of 5.1 ns for each system. Bonds and angles of the solvent was constrained using SHAKE[48]. The distance of the three bonds which forms/breaks were restrained with a harmonic potential force which is equal to zero in the given distance range and equal to 10.0 kcal*mol^{-1}*Å^2 beyond. In all five systems the O-C bond was restrained to 1.0 - 2.7 Å (both states), the H-N bond to 0.9 - 2.7 Å (state 1 only) and the H-O bond to 0.9 - 1.9 Å (state 2 only). Non-bonded interactions were given a cut off distance of 10 Å between solvent-solvent, solute-solvent and solute-solute atoms. No cut off for Q-atoms. Energies were printed out every 5 steps.

2.5 Transition State Stabilization Using Modified LIE

It it well established that enzymes accelerate chemical reactions by lowering the activation energy. The importance of enzymes as biocatalysts can not be underestimated. Without the help of enzymes, the rate of metabolic processes would be too slow to sustain life. Simulation methods have successfully shed light on the origin of catalytic power, which is found to be electrostatic preorganization of the enzymes dipoles to stabilize the transition state.[49][50][51] The stabilization of the

transition state (TS) is a difficult quantity to measure. Here, a modified LIE approach is used to estimate the stabilization of the TS relative to the stabilization of the reactant state (RS). In LIE, the binding free energy of a ligand is found as the difference in solvation free energy of the bound state and the free state. Applying this to TS and RS, the binding free energy of the TS can be estimated as the free energy of transfer between the RS and TS. The LIE equation (Equation 1.21) where the RS is considered the reference state and TS the bound state, binding free energy of TS relative to RS becomes:

$$\Delta G_{bind}^{TS-RS} = \alpha\big(\langle U_{l-s}^{vdW}\rangle_{TS} - \langle U_{l-s}^{vdW}\rangle_{RS}\big) + \beta\big(\langle U_{l-s}^{el}\rangle_{TS} - \langle U_{l-s}^{el}\rangle_{RS}\big) + \gamma \qquad (2.1)$$

Equation 2.1 can be solved by sufficient sampling of the RS and TS. Sampling of RS and TS will be estimated by sampling a state assumed close to the actual RS or TS state. In EVB, the reaction free energy profile is found by solving Equation 1.30. As seen in the equation, multiple intermediate mapping potentials ($\epsilon_m = \lambda_m\epsilon_1 + (1-\lambda_m)\epsilon_2$) contribute to the same reaction coordinate ($\Delta\epsilon_n$). By examining the output file generated by Qfep[33], one can identify the mapping potential that contributes the most to a given reaction coordinate.

2.5.1 MD Simulation Details

Topology, FEP file and PDB file was retrieved from the EVB input files. This means that the Q-atoms defined in EVB, will be Q-atoms here as well. The RS system, defined by using $\lambda = 0.84$, was heated stepwise from 1 K to 300 K in a matter of 51 ps followed by 1 ns of equilibrium at 300 K. The system was coupled to a temperature bath with a relaxation time of 10 fs. The production phase proceeded for 1 250 000 steps, using step size 1.0 fs. 8 parallels (with random starting seed) were performed, yielding 10 ns of total simulation time for each CAP-enzyme system. Solvent was constrained with SHAKE[48] and non-bonded interactions were given a cut off distance of 10 Å between solvent-solvent, solute-solvent and solute-solute atoms. Energies were printed out every 5 steps. Identical setup for the TS, except $\lambda = 0.40$ in both equilibrium and production phase.

3 Results and Discussion

Before a simulation of a system can be conducted, the configuration of the system must be properly defined. The first task was therefore to obtain initial structure of each of the four CAPs bound to chymotrypsin. Structural information on biomolecules are typically obtained by experimental methods such as X-ray crystallography or NMR. The initial structure of chymotrypsin was retrieved from such a model found in the Protein Data Bank (PDB), a data bank for biomolecular structures.[52] As described in Methods, a docking experiment was first performed to build a model of the CAPs bound to chymotrypsin. It should be noted that these complexes were built under the assumption that all four CAPs are substrates for chymotrypsin. This chapter will start with an evaluation of the binding of CAPs to chymotrypsin as obtained from docking and the binding free energies calculated using the LIE method. Next, the reaction mechanism of CAPs bound to chymotrypsin is investigated using reaction free energy surfaces obtained with EVB. Finally, the results of the investigation of the binding free energy of the TS relative to RS are presented.

3.1 Binding of CAPs to Chymotrypsin

The CAPs were modelled using the BPTI as a starting point. BPTI is a TS inhibitor for chymotrypsin and other serine proteases, which means that the structure of the inhibitor in the active site is assumed to resemble the TS of an actual substrate. The BPTI used here has a Trp as the P1 residue. The CAPs are expected to bind to chymotrypsin by the binding mode proposed by Karstad *et al.*, as shown in Figure 3.1.

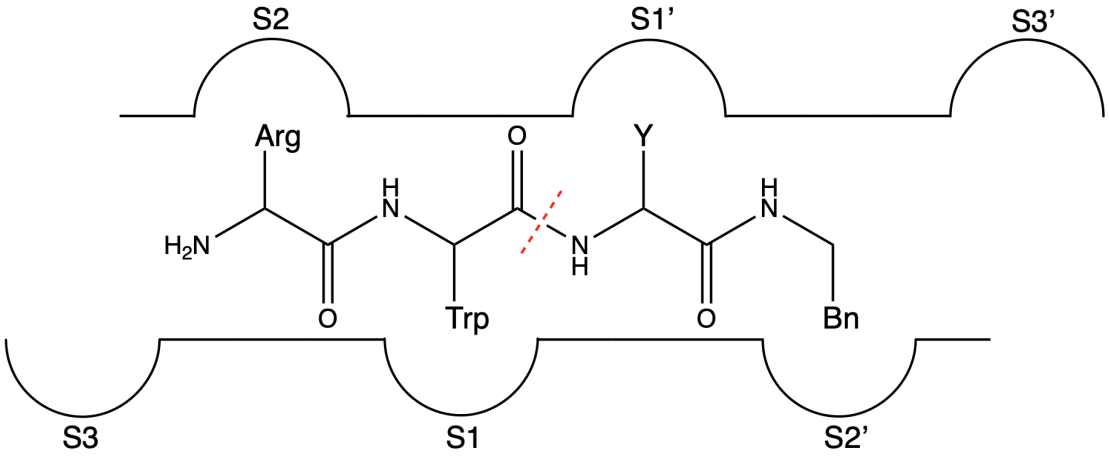


Figure 3.1: Binding site of chymotrypsin and the proposed mode of binding for CAPs included in this study.[22] The hydrophobic Trp is expected to bind to S1 (specificity pocktet), the charged Arg in the S2 pocket and the C-terminal bulk in S2'. Y residue is expected to enter S1'. The red line is the site of cleavage (scissile bond).

Y in Figure 3.1 represents the residues that varies between the CAPs, Har (CAP 1), Lys (CAP 2), Agp (CAP 3) and App (CAP 4). The S1' pocket of chymotrypsin is made up of surface loops 40 (34-41) and 60 (58-69).[53] The pocket contains the negatively charged residues Asp35 and Asp64, which means that the positively charged residues Har, Lys and Agp may interact with the pocket.

3.1.1 Docking Results

CAP 2 was docked using constraints designed to match the geometry of the main-chain of BPTI in complex with chymotrypsin. The Glide XP docking score was -8.1 kcal/mol. Under the assumption that all CAPs bind with the same binding mode, the remaining CAPs were manually prepared using CAP 2 as a starting point. In-place scoring of the prepared CAPs gave XP docking scores of -7.7, -7.9 and -7.4 kcal/mol for CAP 1, 3 and 4, respectively. The docked CAP 2 and manually prepared CAP 1, 3 and 4 are presented in Figure 3.2.

The fact that all CAPs are predicted to bind means that all CAPs are expected to be either substrates or inhibitors to the reaction mechanism of chymotrypsin. It should be noted that the proteolytic stability of the CAPs does not necessarily correlate with their binding affinity. CAP 3 and 4 can act as inhibitors instead of

substrates, which means that CAP 3 and 4 may bind as strongly as CAP 1 and 2, but still be stable against degradation.

Glide has shown great accuracy when compared with other docking programs.[54] However, Glide lacks conformational sampling of the protein structure. This make Glide unfit for binding free energy predictions, but well suited for Glide's primary aim, which is fast and accurate pose predictions.

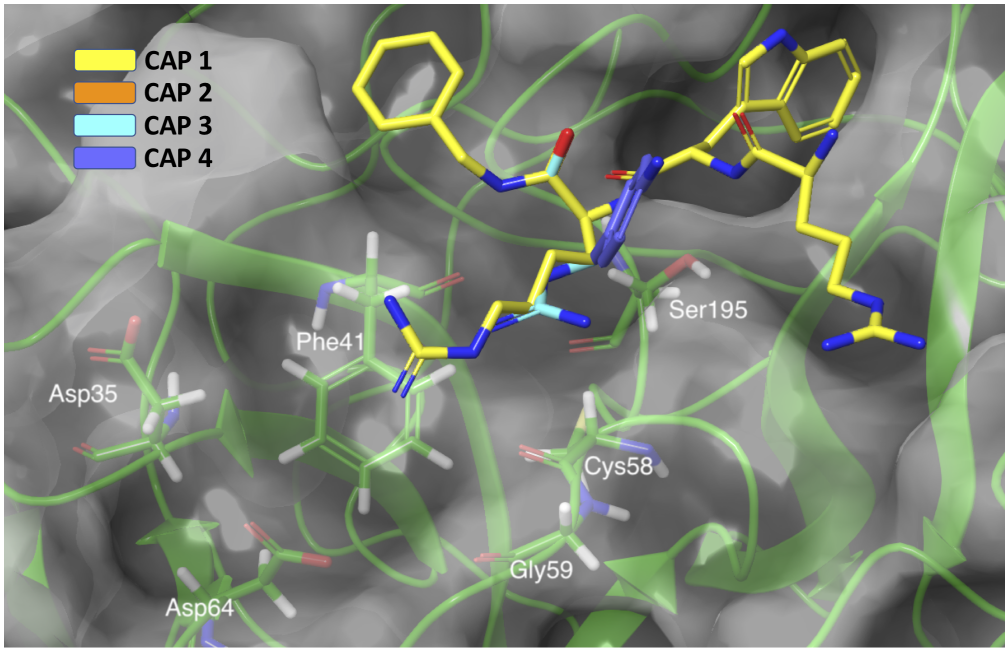


Figure 3.2: Models structures of CAPs bound to chymotrypsin, based on the docking result of CAP 2. The CAPs have same coordinates for all atoms except those belonging to the P1' residue. Yellow is used for CAP 1, cyan is used for CAP 3 and purples is used for CAP 4. CAP 2 is indeed included in the figure, but the superimposition leaves CAP 2 hidden under CAP 1.

3.1.2 Binding Free Energies with LIE

In order to include conformational sampling of protein structures, MD simulations was used. LIE has been shown to be a reasonable method for computing binding free energies of ligands bound to protein. Studies using LIE have reported a reproduction of experimental data within 1 kcal/mol.[55][56]

The energy fluctuations of the MD simulations were investigated by plotting the non-bonded energies against the simulation time. For all CAPs, both van der

Waals and electrostatic energies were overall stable in both states. The energies of CAP 1 are shown as an example in Figure 3.3. The energy fluctuations of CAP 2, 3 and 4 are given in Appendix A. From Figure 3.3, it seems that the van der Waal CAP-enzyme interactions oscillate between two energy levels. It is therefore likely that CAP 1 has two orientations (both stable) which it alternates between. The fact that the energies are stable over the whole simulation time means that the systems energies have converged, and that the MD simulations are stable.

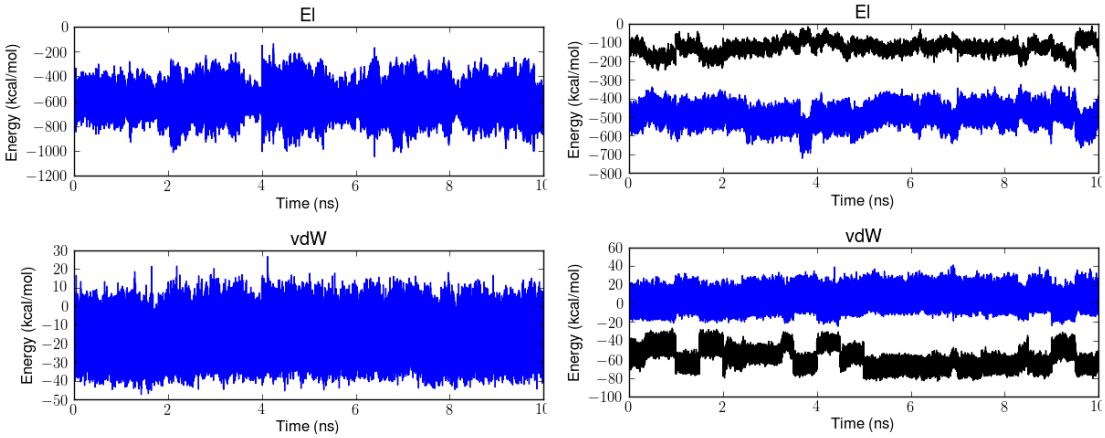


Figure 3.3: Energy fluctuations for CAP 1 as a function of simulation time (left: free state, right: bound state). Blue graphs represent ligand-solvent and black graphs represent ligand-enzyme.

Representative snapshots of the MD simulations for all CAPs in bound state are shown in Figures 3.4 - 3.7. The binding free energies, listed in Table 3.1, were analysed using the *Analyse LIE* tool in Qgui[45].

Table 3.1: Binding Free Energies (kcal/mol) for CAPs Calculated Using LIE (kcal/mol) and Corresponding Experimental Half-Life Results.

	ΔG_{bind}^{LIE}[a]	std.err.	$\Delta\Delta G$[b]	$\Delta\langle U_{l-s}^{vdW}\rangle$	$\Delta\langle U_{l-s}^{EL}\rangle$	$\tau_{1/2}$ (h)[c]
CAP1	-17.0	4.5	0.0	-39.0	-20.5	1
CAP2	-18.0	2.5	-1.0	-35.7	-23.2	4
CAP3	-6.0	4.0	11.0	-31.9	-0.5	stable
CAP4	-5.9	2.0	11.1	-41.3	3.0	stable

[a] Calculated using LIE parameters $\alpha = 0.18$, $\beta = 0.50$ and $\gamma = 0.00$.
[c] ΔG_{bind}^{LIE} energies relative to CAP 1.
[c] Experimental data from Karstad *et al.*[22]

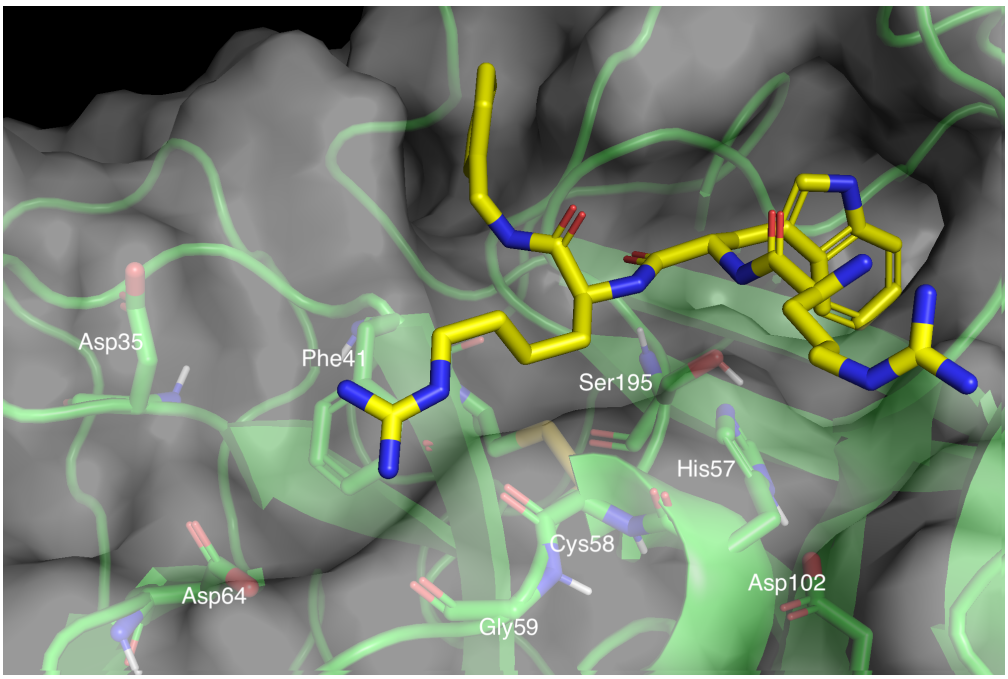


Figure 3.4: Representative snapshot as observed in the MD simulation of CAP 1 bound to chymotrypsin.

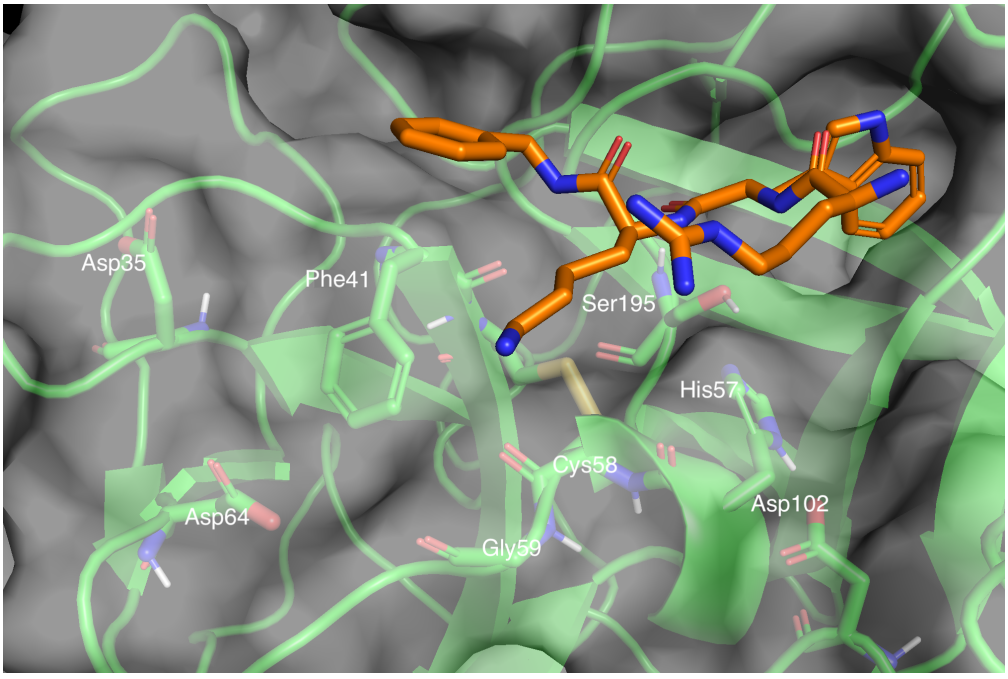


Figure 3.5: Representative snapshot as observed in the MD simulation of CAP 2 bound to chymotrypsin.

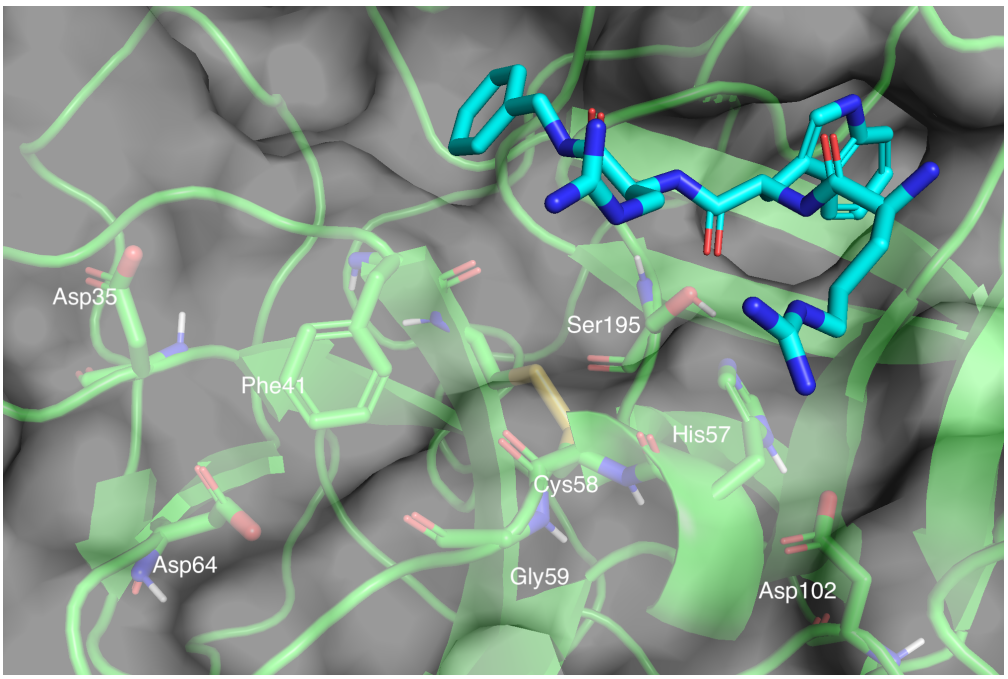


Figure 3.6: Representative snapshot as observed in the MD simulation of CAP 3 bound to chymotrypsin.

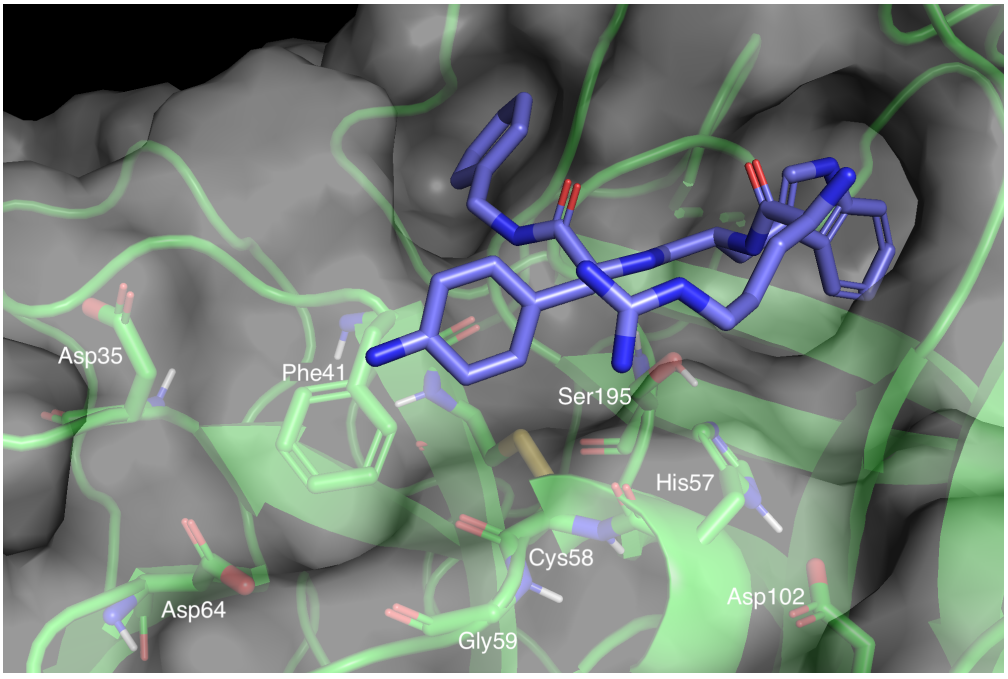


Figure 3.7: Representative snapshot as observed in the MD simulation of CAP 4 bound to chymotrypsin.

The MD trajectories reveals that the Y residue of CAP 1, 2 and 4 enters the S1' pocket during simulation. The Y residue of CAP 1, Har, is found to enter the pocket deeper than the Y residues of CAP 2 and 4 due to its extended hydrocarbon chain. Har has electrostatic interactions with main-chain oxygen of Cys58 and Gly59, as well as the side-chain of Asp64. The Y residue of CAP 2, Lys, have electrostatic stabilization from the same protein residues as Har. The Y residue of CAP 4 (App) is neutral and expected to be less stabilized by the protein residues Cys58, Gly59 and Asp64 than the positively charged Y residues CAP 1 and CAP 2. Agp, the Y residue of CAP 3 is not found to enter the S1' pocket. This may be due to repulsive forces between the bulky guanidine and the protein, or that the distance from Agp to Cys58, Gly59 and Asp64 are too long for the stabilizing interactions to 'compete' against the residue-solvent interactions.

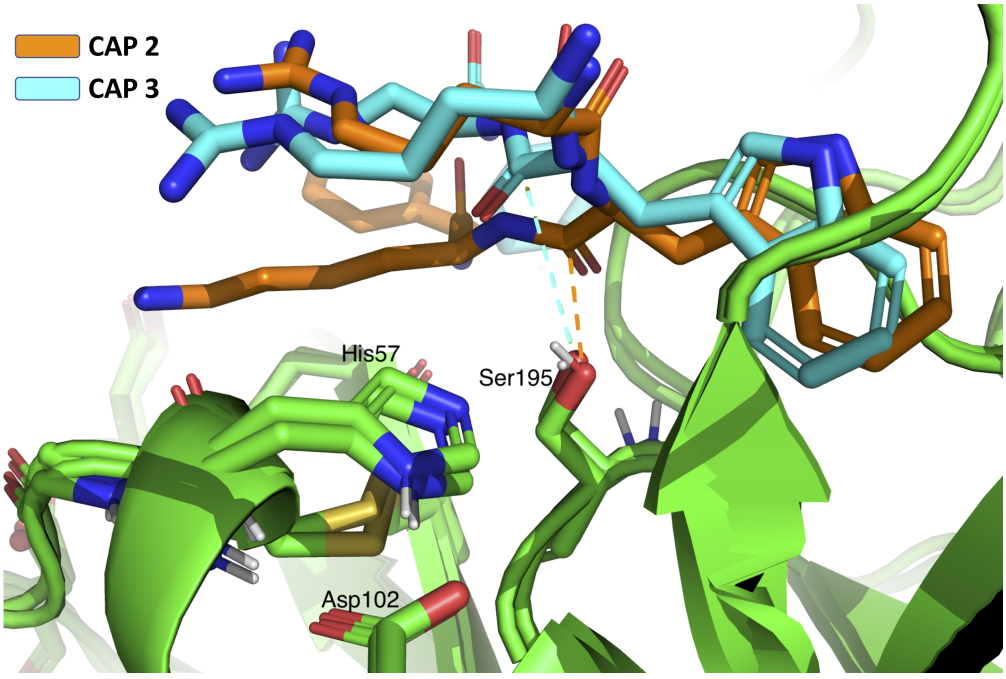


Figure 3.8: Representative snapshot as observed in the MD simulation of CAP 2 and 3 bound to chymotrypsin. The two snapshots are superimposed to illustrate the different structure of CAP 2 and 3. CAP 2 interacts with the Ser195 superimposed on top.

The main-chain structures of CAP 1, 2 and 4 are found to remain close to the initial configuration. CAP 3 is however oriented differently and does not lie as close to the protein as the other CAPs. The average distance between Ser195

(side-chain oxygen) and P1 (carbonyl carbon) over all simulations was found to be 3.53, 3.63, 4.46 and 3.48 Å for CAP 1, 2, 3 and 4, respectively. As can be seen in Figure 3.8, the P1 residue of CAP 3 does not enter the S1 pocket as deep as for the other CAPs and the main-chain has a different orientation. The main-chain oxygen of P1 is facing away from the oxyanion hole, which is the orientation found for the other CAPs. This may indicate that the CAP 3 is not stabilized for the catalytic reaction.

The experimental binding free energy is only available for CAP 3. Previous isothermal titration calorimetry (ITC) studies on CAP 3 reported an experimental binding free energy of -5.7 kcal/mol for binding of CAP 3 to chymotrypsin.[22] The LIE estimate for CAP 3, -6.0 kcal/mol, is in excellent agreement with experimental data. The binding free energies relative to CAP 1 show a clear separation between the experimentally stable and unstable CAPs— CAP 3 and 4 have approximately 11-12 kcal/mol higher binding free energy than CAP 1 and 2. The average van der Waals interactions in the bound state relative to the free state is similar for all CAPs and vary from -35.7 to -41.3 kcal/mol. The clear distinction in binding free energy between stable and unstable CAPs originate from the difference in electrostatic interactions between the two states. For CAP 1 and 2 the bound state provides stabilizing electrostatic interactions of -20.5 and -23.2 kcal/mol relative to the free state. CAP 3 only experience a stabilizing electrostatic effect of -0.5 kcal/mol and CAP 4 has instead a destabilizing electrostatic effect of 3.0 kcal/mol. These results indicate that the lack of electrostatic stabilization in the bound state may be the reason for why CAP 3 and 4 are stable towards chymotryptic degradation.

3.2 Activation and Reaction Free Energies

The reaction mechanism of chymotrypsin is described in detail in Section 1.4. The reaction mechanism is initiated by a proton transfer from Ser195 to His57, followed by the nucleophilic attack by Ser195 on the main-chain carbonyl carbon of the P1 residue, which leads to the formation of a tetrahedral intermediate. This reaction step is shown to be the rate-limiting step of the catalytic mechanism.[57] In the two-state EVB, the CAP bound to the active site (docking result) is state 1 and

the tetrahedral intermediate is state 2. The free energy results therefore refers to the free energy associated with the rate-limiting step, and not the free energy of the complete reaction.

$\alpha^{i\neq1}$ and H_{ij} values were retrieved by calibrating the reference reaction using the *EVB reference reaction calibration* tool in *Qgui*[45]. The target free energies used in the calibration are from Štrajbl *et al.* on the reference solution reaction for studies of serine proteases.[58] The activation and reaction free energies and the reorganization energy calculated with EVB method are listed in Table 3.2. The reaction free energy surfaces for all CAPs and the reference reaction in solvent are shown in Figure 3.9.

Table 3.2: Activation and Reaction Free Energies and the Reorganization energy (λ) (kcal/mol) for the Formation of the First Tetrahedral Intermediate.

	$\langle\Delta G^{\ddagger}\rangle$	$\langle\Delta G_0\rangle$	$\langle\lambda\rangle$
CAP1	17.6 ± 0.4	14.7 ± 0.7	336.6 ± 1.2
CAP2	18.1 ± 0.3	15.4 ± 0.3	336.0 ± 1.8
CAP3	15.7 ± 0.2	11.1 ± 0.3	339.5 ± 0.9
CAP4	18.5 ± 0.4	15.9 ± 0.5	336.7 ± 1.4

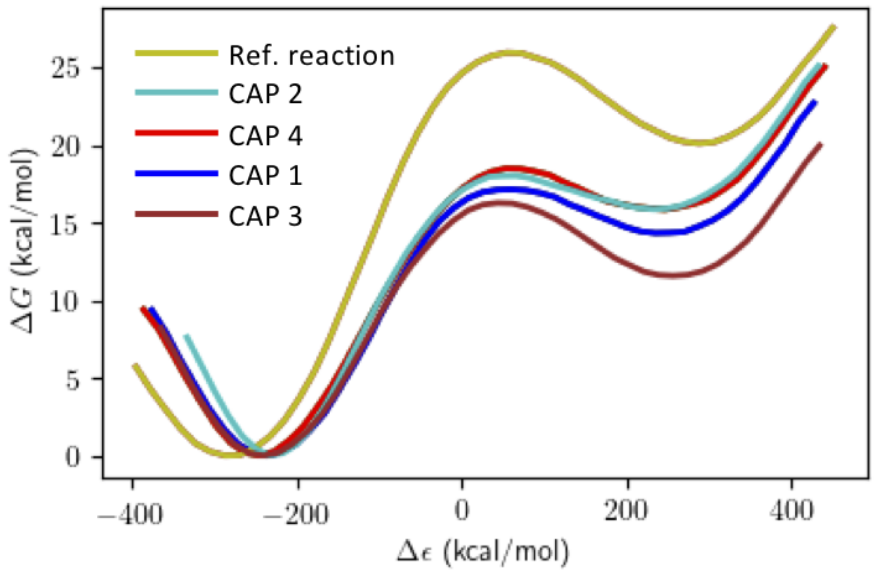


Figure 3.9: Reaction free energy surfaces of the catalytic reaction of all CAPs and the reference reaction in solvent.

The free energy surface is two-dimensional as a result of defining the reaction coordinate as the one-dimensional energy gap between state 1 and 2 ($\Delta\epsilon = \epsilon_1 - \epsilon_2$). The activation energies for all four CAPs are lower than the reference activation energy of 26 kcal/mol, which shows that the reaction is catalyzed by chymotrypsin with a catalytic effect of 8-10 kcal/mol. The EVB result show that chymotrypsin is able to catalyze the formation of the tetrahedral intermediate for all CAPs, but this does not necessarily mean that the CAPs are cleaved. The starting structure of the system forces the CAPs into an orientation optimized for reaction and the two-state FEP furthermore forces the system into the tetrahedral intermediate state. Therefore, the results can be a consequence of the starting structure.

The activation and reaction free energies are lower for CAP 3 than for the other CAPs which could indicate that the TS and the tetrahedral intermediate state for CAP 3 is more stabilized than for the other CAPs. However, the energies could also be lower due to ground-state destabilization, meaning the reactant state of CAP 3 is higher in energy than the reactant state of the other CAPs.

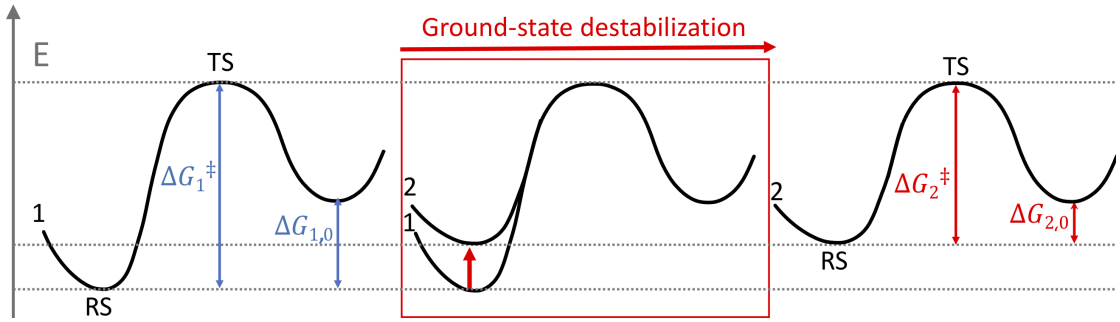


Figure 3.10: Ground-state destabilization, the process of moving from the scenario to the left and to the right (from high to low activation and reaction free energies).

The scenario illustrated in Figure 3.10 seems more plausible for CAP 3 considering the MD simulation findings of Section 3.1.2. The binding mode of CAP 3 (Figure 3.6 and Figure 3.8) does not have optimal stabilization from the oxyanion hole, and this indicate that the main-chain oxygen of P1 may not be stabilized for the catalytic reaction.

As mentioned in Section 2.5, enzymes are preorganized with respect to their dipoles to stabilize the TS. This means that enzymes have optimized a preorganized polar environment which requires less reorganization to stabilize the reaction coordinate than the reorganization of solvent dipoles in an uncatalysed reaction. The *reorganization energy* (λ) in protein is therefore lower than the reorganization energy in solvent. λ can be estimated by examining the diabatic EVB free energy profiles, which was done using Qgui[45]. λ values calculated from the EVB is 336.6, 336.0, 339.5 and 336.7 kcal/mol for CAP 1, 2, 3 and 4, respectively. CAP 3 has a somewhat higher λ than the other three CAPs, which indicate that the enzyme is less preorganized to accommodate the reaction coordinate of CAP 3, compared to the reaction coordinate of the other CAPs. λ of the reference reaction was found to be 371.8 kcal/mol ($\pm$ 4.4). These results show that the protein is preorganized, in respect to the solvent reference reaction, to stabilize the reaction coordinate of all CAPs.

3.3 Binding Affinity of the TS Relative to RS

In order to examine the stabilization of the TS relative to the RS, a modified LIE method was employed. Both TS and RS was identified by examining the reaction free energy profiles predicted in Section 3.2. ΔG_{bind}^{TS-RS} is a quantity that reflects the binding interactions of the CAP in TS. The lower the energy of ΔG_{bind}^{TS-RS}, the more stabilized the CAP is in the TS relative to the RS. ΔG_{bind}^{TS-RS} is however not a physical quantity that can be easily be measured experimentally, but can nonetheless provide valuable insights.

Energy fluctuation of the MD simulations were investigated in the same manner as previously shown for LIE (Section 3.1.2). The non-bonded interactions of all CAPs in both RS and TS were all stable, meaning that the systems energies had converged and the simulations were stable. The energy fluctuations for CAP 1 is shown as an example in Figure 3.11. The energy fluctuation of the CAP 2, 3 and 4 are given in Appendix B. The energies found by employing Equation 2.1 on the RS and TS estimated by the EVB results are listed in Table 3.3.

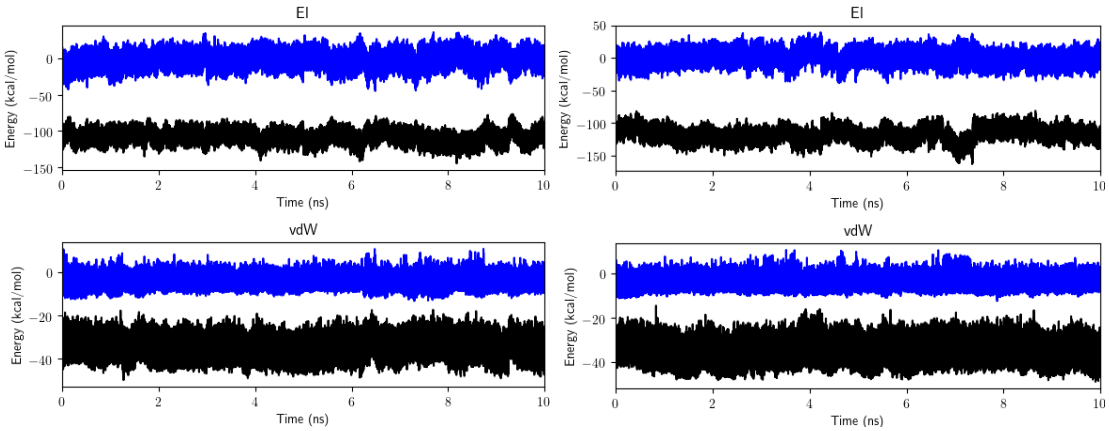


Figure 3.11: Energy fluctuations for CAP 1 as a function of simulation time (left: RS, right: TS). Blue graphs represent ligand-solvent and black graphs represent ligand-enzyme.

Table 3.3: Binding Free Energies of the TS relative to the RS as Defined by ΔG_{bind}^{TS-RS}(kcal/mol).

	ΔG_{bind}^{TS-RS}[a]	std.err.	$\Delta(\Delta G)$[b]	$\Delta\langle U_{l-s}^{vdW}\rangle$	$\Delta\langle U_{l-s}^{EL}\rangle$
CAP1	-4.1	1.0	0.0	0.5	-8.3
CAP2	-4.4	1.3	-0.2	0.1	-8.8
CAP3	-2.6	1.1	1.5	1.2	-5.5
CAP4	-3.3	1.7	0.7	1.1	-7.0

[a] Calculated using LIE parameters $\alpha = 0.18$, $\beta = 0.50$ and $\gamma = 0.00$.
[b] $\Delta(\Delta G_{bind}^{TS-RS})$ energies relative to CAP 1.

The results in Table 3.3 show that all CAPs have lower interaction energies in the TS, meaning they are more stabilized in that state. CAP 1, 2, 3 and 4 experience a stabilizing effect of -4.1, -4.6, -2.6 and -3.3 kcal/mol, respectively. It is interesting that the unstable CAPs, CAP 1 and 2, have a higher stabilization of the TS relative to the stable CAPs, CAP 3 and 4. ΔG_{bind}^{TS-RS} of CAP 4 is also somewhat lower in energy than CAP 3, making CAP 3 the least stabilized CAP in regard to the TS relative to RS. The stabilizing effect originate from the difference in average electrostatic interactions between the two states. The difference in average van der Waals interactions between the two states, ranging from -0.2 and 1.5 kcal/mol, is low compared to to the difference in electrostatic interactions, which range from -5.5 to -8.8 kcal/mol. The average non-bonded interactions which make up ΔU_{l-s}^{EL}

and ΔU_{l-s}^{vdW} in Table 3.3 are listed in Table 3.4.

Table 3.4: Average Non-Bonded Interactions Between Q-Atoms and Surrounding in TS relative to TS (kcal/mol).

	$\Delta\langle U^{vdW}\rangle_{Q-wat}$	$\Delta\langle U^{vdW}\rangle_{Q-prot}$	$\Delta\langle U^{EL}\rangle_{Q-wat}$	$\Delta\langle U^{EL}\rangle_{Q-prot}$
CAP 1	0.5	-0.1	1.3	-9.6
CAP 2	0.6	-0.6	2.1	-10.9
CAP 3	0.6	0.6	-2.7	- 2.8
CAP 4	0.5	0.6	-6.1	-0.9

The average van der Waals interactions for all CAPs are approximately equal in the two states, both regarding Q-water and Q-protein interactions. CAP 1 and 2 have a destabilizing effect in the Q-water electrostatic interactions, but the effect is small (1.3 kcal/mol for CAP 1 and 2.1 kcal/mol for CAP 2). The Q-protein electrostatic interactions between states differ significant1y for CAP 1 and 2. The TS of CAP 1 is stabilized with -9.6 kcal/mol while the CAP 2 TS is stabilized with -10.9 kcal/mol. Most interestingly, CAP 3 and 4 does not share the same features as the two stable CAPs. Unlike CAP 1 and 2, CAP 3 and 4 have a stabilization effect from the electrostatic Q-water interactions; -2.7 and -6.1 kcal/mol for CAP 3 and 4, respectively. Furthermore, the electrostatic stabilizing effect of the Q-protein interactions is minimal when compared to the stabilization experience by the unstable CAPs. CAP 3 has a -2.8 kcal/mol stabilization and CAP 4 has a even lower stabilization of -0.9 kcal/mol. This means that the increased stabilization of CAP 4 relative to CAP 3 originate from favorable Q-water interactions and not Q-protein interactions, which are actually found to be lower in CAP 4 than CAP 3.

The MD simulations were investigated to find electrostatic interactions which can explain the weak Q-protein interactions of CAP 3 and 4, compared to CAP 1 and 2. Of particular interest was the distance between the P1 main-chain oxygen and the two stabilizing main-chain NH groups of Ser195 and Gly193 in the oxyanion hole. The force field in the FEP transformation gives the oxygen a more negative charge in the TS relative to RS, and thus a stronger electrostatic potential. However, a decrease in bond length will also contribute to more negative energies. Also of interest is the main-chain NH group of NHBn (C-terminal cap-

ping) and its interactions with the main-chain oxygen of Phe41. The distance between the main-chain NH group of P1 and the main-chain of Ser208 was also investigated, but the result was found to be similar for all CAPs. Representative snapshots from the MD simulations of both states are shown in Figure 3.12 and the four distances of interest are summarized in Table 3.5.

Table 3.5: Atom-Atom Distances (Å) Between CAP Atoms (P1 Main-Chain O and NHBn Capping Main-Chain NH) and Stabilizing Atoms Inside and Close to the Oxyanion Hole.

		P1 (=O)		NHBn (-NH)	P1 (-NH)
	state	Ser195 (-NH)	Gly193 (-NH)	Phe41 (=O)	Ser208 (=O)
CAP1	RS	1.78	3.08	2.77	3.42
	TS	1.71	2.84	3.22	2.99
	Diff.	*-0.07*	*-0.24*	*0.45*	*-0.43*
CAP2	RS	1.76	2.92	2.67	3.49
	TS	1.73	2.65	3.19	2.83
	Diff.	*-0.04*	*-0.27*	*0.52*	*-0.66*
CAP3	RS	1.78	3.02	3.79	3.35
	TS	1.78	3.09	4.90	2.78
	Diff.	*0.00*	*0.07*	*1.11*	*-0.57*
CAP4	RS	1.76	3.04	2.21	3.44
	TS	1.73	2.81	3.43	2.98
	Diff.	*-0.03*	*-0.23*	*1.22*	*-0.46*

The MD trajectories shows that the main-chain oxygen of the P1 residue has better orientation towards the oxyanion hole in the TS of CAP 1, 2 and 4. This is verified by the average distances listed in Table 3.5 where we see that the atom-atom distances to both Ser195 and Ser193 has decreased. Surprisingly, the main-chain P1 oxygen of CAP 3 is not found closer to the oxyanion hole in TS relative to RS. As seen for CAP 3 in Figure 3.12, the main-chain NH group of the NHBn capping is longer facing Phe41, but the main-chain P1 oxygen instead. This interaction will stabilize the CAP but decrease the ligand-enzyme interactions, as the NH group is no longer oriented towards Phe41. This electrostatic CAP-CAP interaction is verified by the average distance which was found to be 2.41 Å in the TS of CAP 3 (and 3.99, 3.20 and 3.05 Å in CAP 1, 2 and 3, respectively). Furthermore it was found that the P1' residue of CAP 3 lie on average 2.87 Å from the main-chain

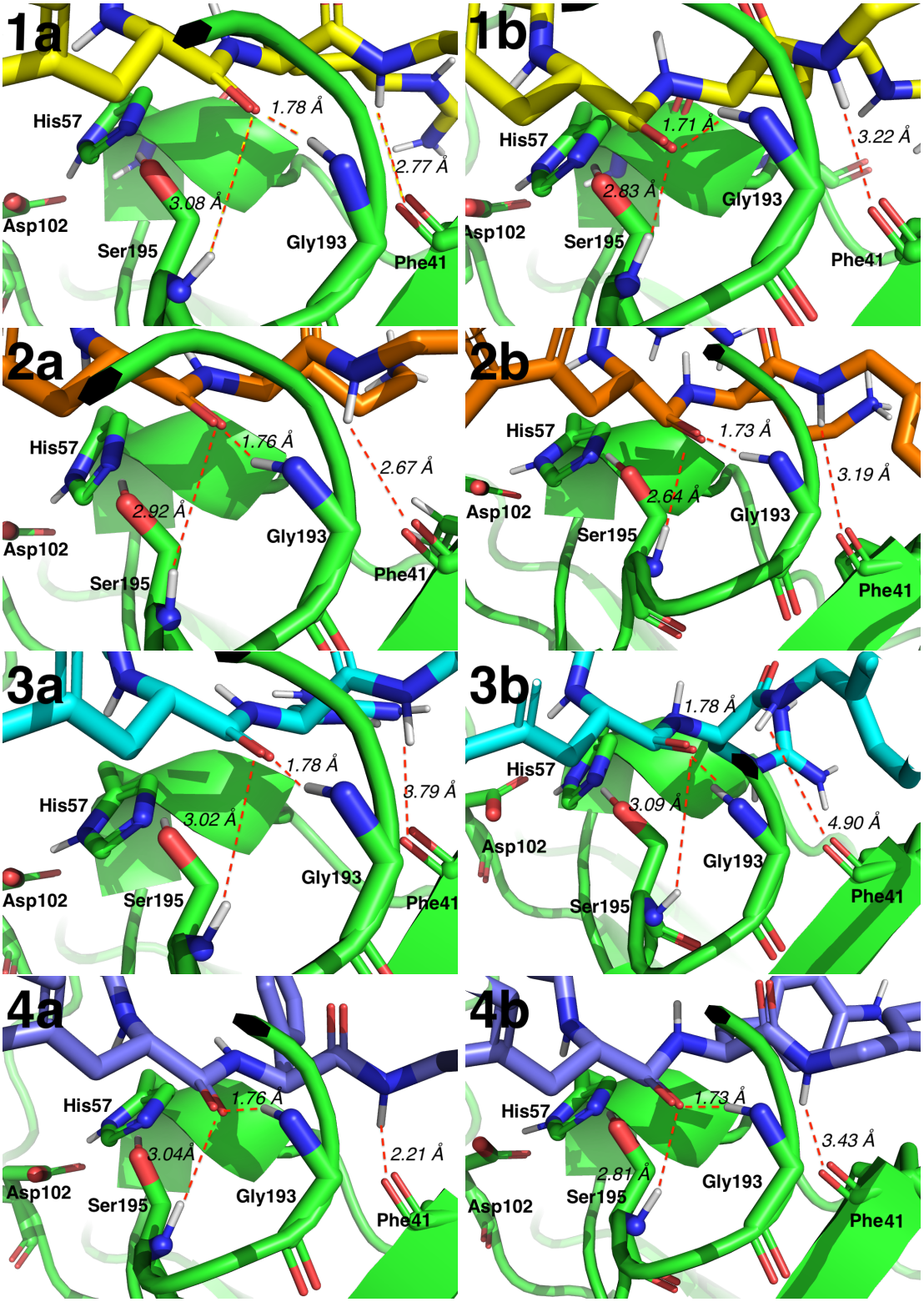


Figure 3.12: Representative snapshot of the oxyanion hole as observed in the MD simulation in RS (left) and in TS (right) for CAP 1 (1a-b), CAP 2 (2a-b), CAP 3 (3a-b) and CAP 4 (4a-b).

oxygen of Phe41, which could have a repulsive effect on the NH group of NHBn. The distance between Phe41 and the NHBn NH group in TS increases notably more for CAP 3 and 4 than for CAP 1 and 2, which means that CAP 1 and 2 experience greater stabilization by Phe41 than CAP 3 and 4.

To summarize, CAP 1 and 2 are more stabilized in the TS than RS due to favorable electrostatic Q-protein interactions. CAP 3 and 4 experience stabilizing Q-water interactions in TS, but low to minimal Q-protein stabilization. The MD simulations suggest the increased TS stabilization for CAP 1 and 2 stems from stronger stabilization of the oxyanion hole and overall favorable main-chain interactions. CAP 4 have better oxyanion hole stabilization in the TS than RS, but the main-chain NH group of NHBn is notably far from the stabilizing main-chain of Phe41 in TS relative to RS. This will destabilize CAP 4, and the increased length could be due to the TS orientation of the bulky P1 residue (App). The P1 main-chain oxygen of CAP 3 does not move closer to the oxyanion hole in the TS, which is the trend observed for the other CAPs. It also has a distinct orientation of the main-chain NH group of NHBn towards the P1 main-chain oxygen. The orientation of Phe41 may explain why CAP 3 was found to have a somewhat higher reorganization energy. The short P1' residue is shown to lie close to Phe41, which could induce Phe41 to rotate away from the NH group of NHBn.

4 Concluding Remarks

In this project, the use of MD simulations has shown to give valuable insight into the chymotryptic degradation of CAPs. The binding free energies were investigated using the LIE method. It was found that the two unstable CAPs have remarkably low binding free energies, which were around 12 kcal/mol lower than the binding free energies of the two stable CAPs. The strong binding of the unstable CAPs was found to be due to favorable electrostatic interactions in the bound state relative to the free state. The experimental binding free energy was only available for one CAP, but the calculated binding free energy for this compound was in excellent agreement with experimental data.

Reaction free energy surfaces of the rate-limiting step were obtained using the EVB method. The results showed that all CAPs are able to form the tetrahedral intermediate. However, this does not mean that all CAPs are cleaved. The formation of a tetrahedral intermediate may only be a result of the starting structure used, which was optimized for reaction. The reorganization energy of all enzyme catalyzed reactions were lower than the reorganization energy of the solvent reaction used as reference. This means that the enzyme is preorganized for the reaction, compared to the solvent reaction.

The stabilization of the TS relative to the RS was studied using a modified LIE method. The two unstable CAPs were found to be more stabilized by the TS relative to RS, than the two stable CAPs. The increased TS stabilization of the unstable CAPs originate from stronger electrostatic interactions with the protein. The stable CAPs have minimal difference in electrostatic interactions with protein in TS relative to RS, and instead experience stronger electrostatic interactions with water in the TS relative to RS. The MD simulations show that the main-chain

structure of the unstable CAPs lies closer to the protein in TS relative to RS, than what was found for the stable CAPs. A tighter binding of the main-chain in the TS could explain why the unstable CAPs are found to have lower TS binding free energy than the stable CAPs. It was also shown that the main-chain P1 oxygen lies closer to the oxyanion hole in the TS than the RS for all CAPs except CAP 3. Interestingly, the activation and reaction free energies of CAP 3 were found to be lower than for the other CAPs, and the binding mode of CAP 3 was found to lack optimized stabilization from the oxyanion hole and differ from the structure of the RS. The low energies can indicate ground-state destabilization of CAP 3, and the binding mode suggest that CAP 3 is not stabilized for the catalytic reaction. This may indicate that the RS structure, even with stabilization from the oxyanion hole, is unfavorable.

This computational study of four CAPs have been successful at separating the stable from the unstable CAPs. It is apparent from these results that electrostatic interactions plays a significant role for the stability of CAPs against chymotryptic degradation. The unstable CAPs have stronger electrostatic interactions with the protein than the stable CAPs, both in the bound state and in the TS relative to the RS. The stable CAPs possess structural features which obstruct stabilizing interactions with protein in the TS.

This project has not only contributed to the understating of chymotryptic degradation of CAPs, but also identified some interesting features of the TS. The cleaved CAPs have stronger electrostatic stabilization in the TS than the RS. This implies that the protein is preorganized with a polar environment that stabilize the TS more than the RS.

Future Work

The overall goal of this project was to investigate the chymotryptic degradation of CAPs. While interesting findings were found that separate the stable from the unstable CAPs, a larger set of CAPs should be studied to clearly establish a trend. The same methods, LIE, EVB and modified LIE, should be employed, but more extensively used. The EVB method was only used to investigate the reaction free energy surface of the assumed rate-limiting step. An EVB study on other reaction

steps of the catalytic mechanism may reveal other trends that separate the stable from unstable CAPs. Furthermore, it is possible that other catalytic steps may be the rate-limiting steps for these CAPs. The individual free energy surfaces of CAP 1, 3 and 4 obtained in this study are somewhat scattered. The systems should be investigated as to why the free energy surfaces are not more uniform, and focus should be on producing a closer set of free energy surfaces. The findings of this project indicate that the RS is an unfavorable structure for CAP 3. It could be interesting to further investigate the stability of the RS of CAP 3 relative to the RS of the other CAPs. This could for example be done by performing a FEP transformation between the RS of different CAPs to establish if there is a difference in free energy, which would verify if the RS of CAP 3 is higher in energy than for the other CAPs.

Bibliography

[1] F. Malerba and L. Orsenigo. "The evolution of the pharmaceutical industry". In: *Business History* 57.5 (2015), pp. 664–687.

[2] Cole, S.T. "Who will develop new antibacterial agents?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1645 (2014), p. 20130430.

[3] M. Roser, S. Ochmann, H. Behrens, and H. Ritchie. "Eradication of Diseases". In: *Our World in Data* (2019). https://ourworldindata.org/eradication-of-diseases.

[4] World Health Organization. *The top 10 causes of death*. May 24, 2018. URL: `https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death` (visited on 04/07/2019).

[5] UN Interagency Coordination Group (IACG). *No Time to Wait: Securing the future from drug-resistant infections*. 2019. URL: `https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf?ua=1` (visited on 07/07/2019).

[6] H. Long, S.F. Miller, C. Strauss, C. Zhao, L. Cheng, Z. Ye, K. Griffin, R. Te, H. Lee, C. Chen, and M. Lynch. "Antibiotic treatment enhances the genome-wide mutation rate of target cells". In: *Proceedings of the National Academy of Sciences* 113.18 (2016), E2498–E2505.

[7] J. Davies and D. Davies. "Origins and Evolution of Antibiotic Resistance". In: *Microbiology and Molecular Biology Reviews : MMBR* 74.3 (2010), pp. 417–433.

[8] M. Madigan, J. Martinko, K. Bender, D. Buckley, and D. Stahl. "Brock Biology of Microorganisms". In: Pearson Education Limited, Essex, 2015, pp. 323–331.

[9] J.I.R. Castanon. "History of the use of antibiotic as growth promoters in European poultry feeds". In: *Poultry Science* 86.11 (2007), pp. 2466–2471.

[10] U.S. Food and Drug Administration. *FDA Releases Annual Summary Report on Antimicrobials Sold or Distributed in 2017 for Use in Food-Producing Animals Showing Declines for Past Two Years*. Dec. 18, 2018. URL: https://www.fda.gov/animal-veterinary/cvm-updates/fda-releases-annual-summary-report-antimicrobials-sold-or-distributed-2017-use-food-producing (visited on 07/07/2019).

[11] Isabella Sanseverino, Anna Navarro Cuenca, Robert Loos, Dimitar Marinov, and Teresa Lettieri. "State of the Art on the Contribution of Water to Antimicrobial Resistance". In: (2018).

[12] A.F.C. Leonard, L. Zhang, A.J. Balfour, R. Garside, P.M. Hawkey, A.K. Murray, O.C. Ukoumunne, and W.H. Gaze. "Exposure to and colonisation by antibiotic-resistant E. coli in UK coastal water users: Environmental surveillance, exposure assessment, and epidemiological study (Beach Bum Survey)". In: *Environment International* 114 (2018), pp. 326–333.

[13] P. Fresia, V. Antelo, C. Salazar, M. Giménez, B. D'Alessandro, E. Afshinnekoo, C. Mason, G.H. Gonnet, and G. Iraola. "Urban metagenomics uncover antibiotic resistance reservoirs in coastal beach and sewage waters". In: *Microbiome* 7 (2019).

[14] M. Mahlapuu, J. Håkansson, L. Ringstad, and C. Björn. "Antimicrobial Peptides: An Emerging Category of Therapeutic Agents". In: *Frontiers in Cellular and Infection Microbiology* 6 (2016).

[15] G. Diamond, N. Beckloff, A. Weinberg, and K.O. Kisich. "The Roles of Antimicrobial Peptides in Innate Host Defense". In: *Current pharmaceutical design* 15.21 (2009), pp. 2377–2392.

[16] M. Wu, E. Maier, R. Benz, and R.E.W. Hancock. "Mechanism of Interaction of Different Classes of Cationic Antimicrobial Peptides with Planar Bilayers and with the Cytoplasmic Membrane of *Escherichia coli* †". In: *Biochemistry* 38.22 (1999), pp. 7235–7242.

[17] Y. Liu, Y. Wang, T.R. Walsh, L. Yi, R. Zhang, J. Spencer, Y. Doi, G. Tian, B. Dong, X. Huang, L. Yu, D. Gu, H. Ren, X. Chen, L. Lv, D. He, H. Zhou, Z. Liang, J. Liu, and J. Shen. "Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study". In: *The Lancet Infectious Diseases* 16.2 (2016), pp. 161–168.

[18] B.J Bruno, G.D. Miller, and C.S. Lim. "Basics and recent advances in peptide and protein drug delivery". In: *Therapeutic delivery* 4.11 (2013), pp. 1443–1467.

[19] M. Werle and A. Bernkop-Schnürch. "Strategies to improve plasma half life time of peptide and protein drugs". In: *Amino Acids* 30.4 (2006).

[20] M.B. Strøm, B.E. Haug, M.L. Skar, W. Stensen, T. Stiberg, and J.S. Svendsen. "The pharmacophore of short cationic antibacterial peptides". In: *Journal of Medicinal Chemistry* 46.9 (2003), pp. 1567–1570.

[21] B.E. Haug, W. Stensen, T. Stiberg, and J.S. Svendsen. "Bulky nonproteinogenic amino acids permit the design of very small and effective cationic antibacterial peptides". In: *Journal of Medicinal Chemistry* 47.17 (2004), pp. 4159–4162.

[22] R. Karstad, G. Isaksen, B.O. Brandsdal, J.S. Svendsen, and J. Svenson. "Unnatural Amino Acid Side Chains as S1, S1, and S2 Probes Yield Cationic Antimicrobial Peptides with Stability toward Chymotryptic Degradation". In: *Journal of Medicinal Chemistry* 53.15 (2010), pp. 5558–5566.

[23] I. Schechter and A. Berger. "On the size of the active site in proteases. I. Papain". In: *Biochemical and Biophysical Research Communications* 27.2 (1967), pp. 157–162.

[24] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. "A geometric approach to macromolecule-ligand interactions". In: *Journal of Molecular Biology* 161.2 (1982), pp. 269–288.

[25] T. N. Hart and R. J. Read. "A multiple-start Monte Carlo docking method". In: *Proteins* 13.3 (1992), pp. 206–222.

[26] D.B. Kitchen, H. Decornez, J.R. Furr, and J. Bajorath. "Docking and scoring in virtual screening for drug discovery: methods and applications". In: *Nature Reviews Drug Discovery* 3.11 (2004), pp. 935–949.

[27] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin. "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy". In: *Journal of Medicinal Chemistry* 47.7 (2004), pp. 1739–1749.

[28] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, and J.L. Banks. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening". In: *Journal of Medicinal Chemistry* 47.7 (2004), pp. 1750–1759.

[29] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, and D.T. Mainz. "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for ProteinLigand Complexes". In: *Journal of Medicinal Chemistry* 49.21 (2006), pp. 6177–6196.

[30] A.R. Leach. *Molecular Modeling: Principles and Applications*. Vol. 5. Essex: Prentice Hall., Jan. 2001.

[31] N.L. Allinger, K. Chen, and J. Lii. "An improved force field (MM4) for saturated hydrocarbons". In: *Journal of Computational Chemistry* 17.5 (1996), pp. 642–668.

[32] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives. "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids". In: *Journal of the American Chemical Society* 118.45 (1996), pp. 11225–11236.

[33] J. Marelius, K. Kolmodin, I. Feierberg, and J. Åqvist. "Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems." In: *Journal of Molecular Graphics and Modelling* 16.4 (1998), pp. 213–225.

[34] W. L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. "Comparison of simple potential functions for simulating liquid water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935.

[35] J. Aqvist, C. Medina, and J. E. Samuelsson. "A new method for predicting binding affinity in computer-aided drug design". In: *Protein Engineering* 7.3 (1994), pp. 385–391.

[36] A. Warshel and R.M. Weiss. "An empirical valence bond approach for comparing reactions in solutions and in enzymes". In: *Journal of the American Chemical Society* 102.20 (1980), pp. 6218–6226.

[37] G. Hong, E. Rosta, and A. Warshel. "Using the Constrained DFT Approach in Generating Diabatic Surfaces and Off Diagonal Empirical Valence Bond Terms for Modeling Reactions in Condensed Phases". In: *The Journal of Physical Chemistry B* 110.39 (2006), pp. 19570–19574.

[38] G.V. Isaksen, J. Åqvist, and B.O. Brandsdal. "Protein surface softness is the origin of enzyme cold-adaptation of trypsin". In: *PLoS computational biology* 10.8 (2014), e1003813.

[39] H. Czapinska, R. Helland, A.O. Smalås, and J. Otlewski. "Crystal Structures of Five Bovine Chymotrypsin Complexes with P1 BPTI Variants". In: *Journal of Molecular Biology* 344.4 (2004), pp. 1005–1020.

[40] *Maestro*. Version 11.8.012. Schrödinger, LLC; New York, NY, USA: 2018-4.

[41] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, and J.H. Jensen. "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions". In: *Journal of Chemical Theory and Computation* 7.2 (2011), pp. 525–537.

[42] J.C. Shelley, A. Cholleti, L.L. Frye, J.R. Greenwood, M.R. Timlin, and M. Uchimaya. "Epik: a software program for pKaprediction and protonation state generation for drug-like molecules". In: *Journal of Computer-Aided Molecular Design* 21.12 (2007), pp. 681–691.

[43] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J.Y. Xiang, L. Wang, D. Lupyan, M.K. Dahlgren, J.L. Knight, J.W. Kaus, D.S. Cerutti, G. Krilov, W.L. Jorgensen, R. Abel, and R.A. Friesner. "OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins". In: *Journal of Chemical Theory and Computation* 12.1 (2016), pp. 281–296.

[44] *Manual for the Molecular Dynamics Package Q v 5.06*. Jan. 1, 2004. URL: `http://xray.bmc.uu.se/~aqwww/q/documents/qman5.pdf`.

[45] G.V. Isaksen, T.A.H. Andberg, J. Åqvist, and B.O. Brandsdal. "Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems". In: *Journal of Molecular Graphics and Modelling* 60 (2015), pp. 15–23.

[46] Schrödinger, LLC. "The PyMOL Molecular Graphics System, Version 2.3.2". 2019.

[47] G. A. Kaminski, Richard A. Friesner, Julian Tirado-Rives, and William L. Jorgensen. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides". In: *The Journal of Physical Chemistry B* 105.28 (2001), pp. 6474–6487.

[48] J. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341.

[49] A. Warshel. "Energetics of enzyme catalysis". In: *Proceedings of the National Academy of Sciences* 75.11 (1978), pp. 5250–5254.

[50] M. Fuxreiter and A. Warshel. "Origin of the Catalytic Power of Acetylcholinesterase: Computer Simulation Studies". In: *Journal of the American Chemical Society* 120.1 (1998), pp. 183–194.

[51] A.J. Adamczyk, J. Cao, S.C.L. Kamerlin, and A. Warshel. "Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions". In: *Proceedings of the National Academy of Sciences* 108.34 (2011), p. 14115.

[52] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. "The Protein Data Bank". In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242.

[53] T. Kurth, D. Ullmann, H.D. Jakubke, and L. Hedstrom. "Converting Trypsin to Chymotrypsin: Structural Determinants of S1' Specificity". In: *Biochemistry* 36.33 (1997), pp. 10098–10104.

[54] J.B. Cross, D.C. Thompson, B.K. Rai, J.C. Baber, K.Y. Fan, Y. Hu, and C. Humblet. "Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy". In: *Journal of Chemical Information and Modeling* 49.6 (2009).

[55] H.S. Leiros, B.O. Brandsdal, and S.M. McSweeney. "Biophysical characterization and mutational analysis of the antibiotic resistance protein NimA from Deinococcus radiodurans". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1804.4 (2010), pp. 967–976.

[56] S. Bjelic, M. Nervall, H. Gutiérrez-de-Terán, K. Ersmark, A. Hallberg, and J. Åqvist. "Computational inhibitor design against malaria plasmepsins". In: *Cellular and Molecular Life Sciences* 64.17 (2007), pp. 2285–2305.

[57] M. W. Hunkapiller, M. D. Forgac, and J. H. Richards. "Mechanism of action of serine proteases: tetrahedral intermediate and concerted proton transfer". In: *Biochemistry* 15.25 (1976), pp. 5581–5588.

[58] M. Štrajbl, J. Florián, and A. Warshel. "Ab Initio Evaluation of the Potential Surface for General Base- Catalyzed Methanolysis of Formamide: A Reference Solution Reaction for Studies of Serine Proteases". In: *Journal of the American Chemical Society* 122.22 (2000), pp. 5354–5366.

Appendix A LIE Energy Fluctuation Graphs

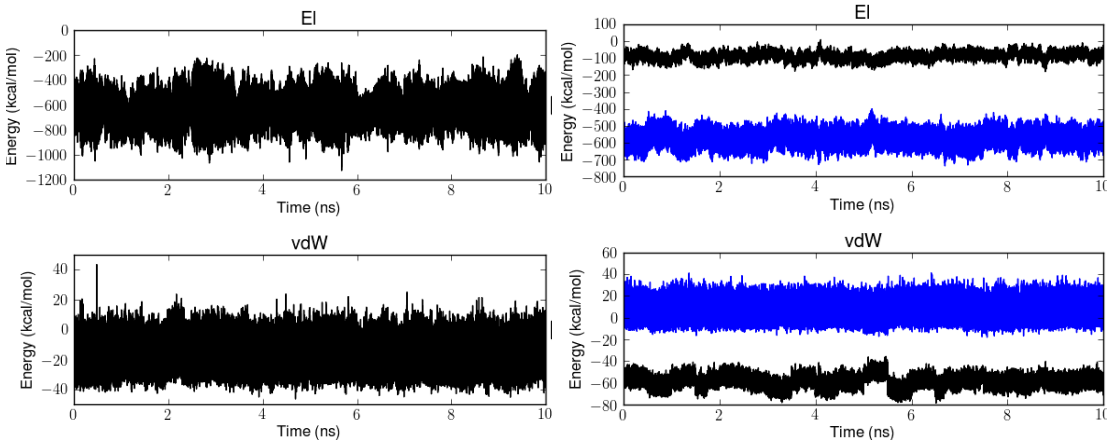


Figure 1: Energy fluctuations for CAP 2 as a function of simulation time. Left: free state where black graph represent Q-water interactions. Right: bound state where blue graph represent Q-water interactions and black graph represent Q-protein.

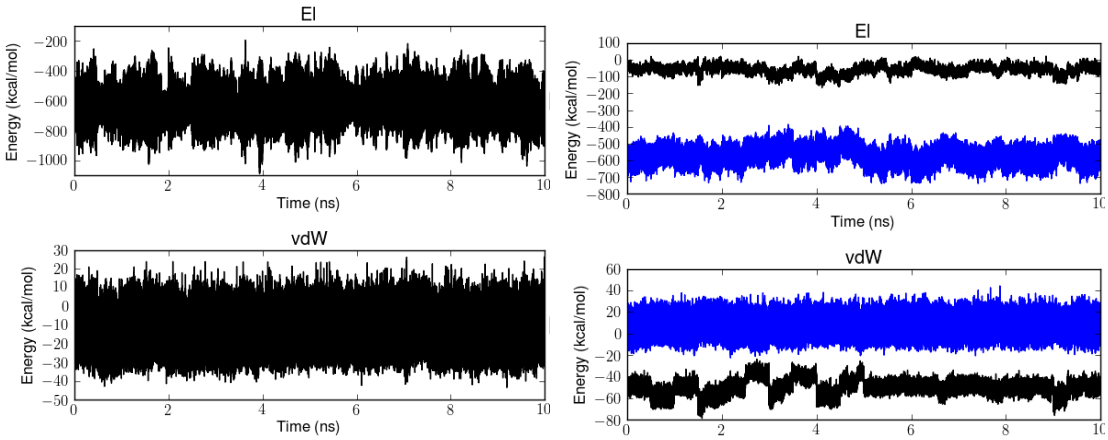


Figure 2: Energy fluctuations for CAP 3 as a function of simulation time. Left: free state where black graph represent Q-water interactions. Right: bound state where blue graph represent Q-water interactions and black graph represent Q-protein.

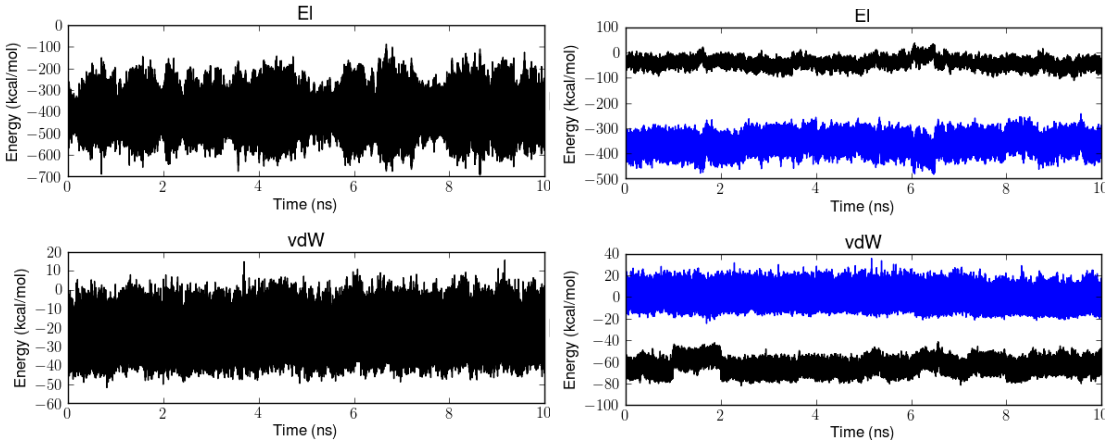


Figure 3: Energy fluctuations for CAP 4 as a function of simulation time. Left: free state where black graph represent Q-water interactions. Right: bound state where blue graph represent Q-water interactions and black graph represent Q-protein.

Appendix B Modified-LIE Energy Fluctuation Graphs

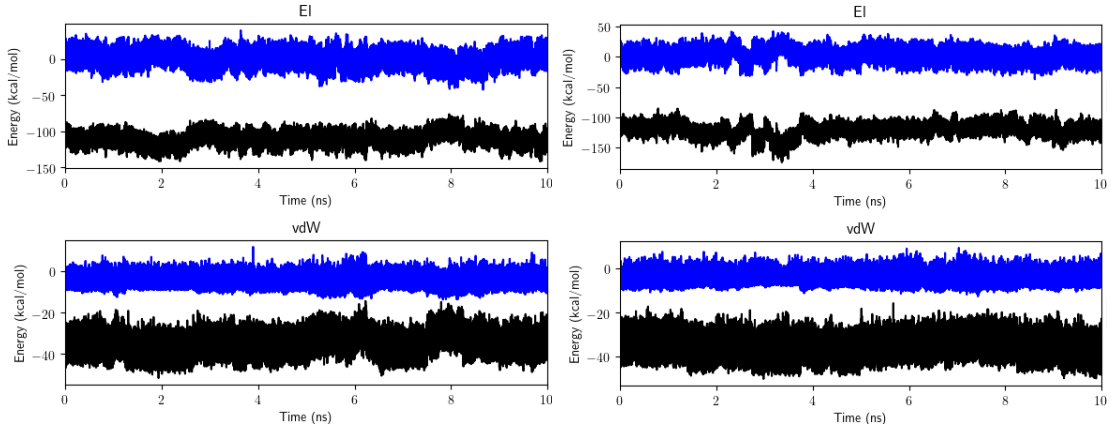


Figure 4: Energy fluctuations for CAP 2 as a function of simulation time. (Left: RS, right: TS). Blue graph represent Q-water interactions and black graph represent Q-protein

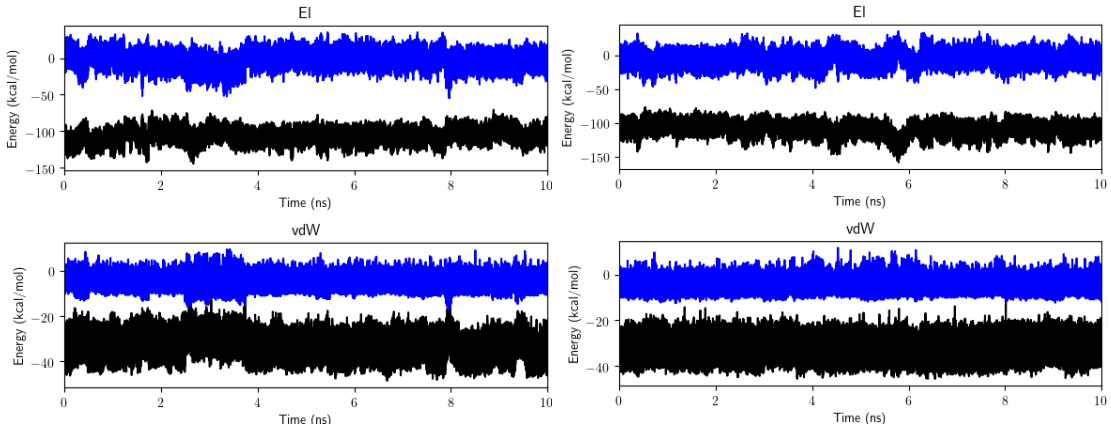


Figure 5: Energy fluctuations for CAP 3 as a function of simulation time. (Left: RS, right: TS). Blue graph represent Q-water interactions and black graph represent Q-protein

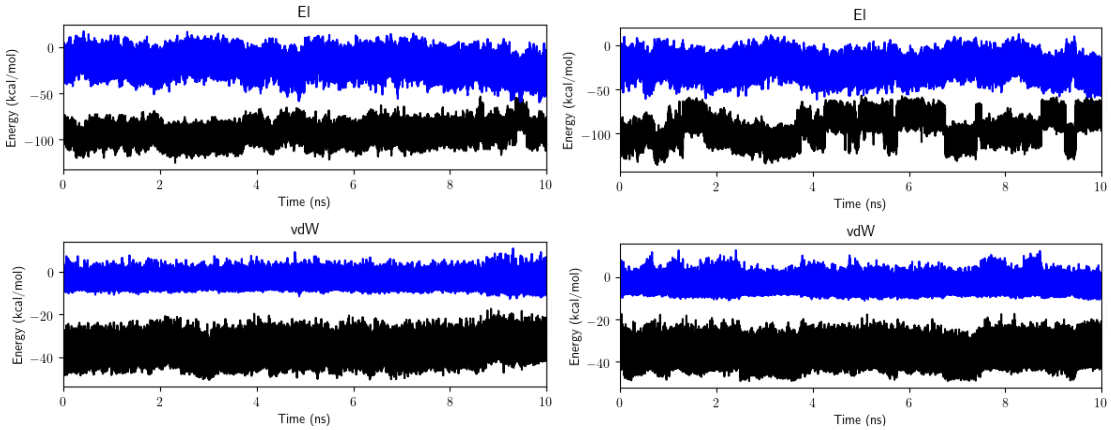


Figure 6: Energy fluctuations for CAP 4 as a function of simulation time. (Left: RS, right: TS). Blue graph represent Q-water interactions and black graph represent Q-protein