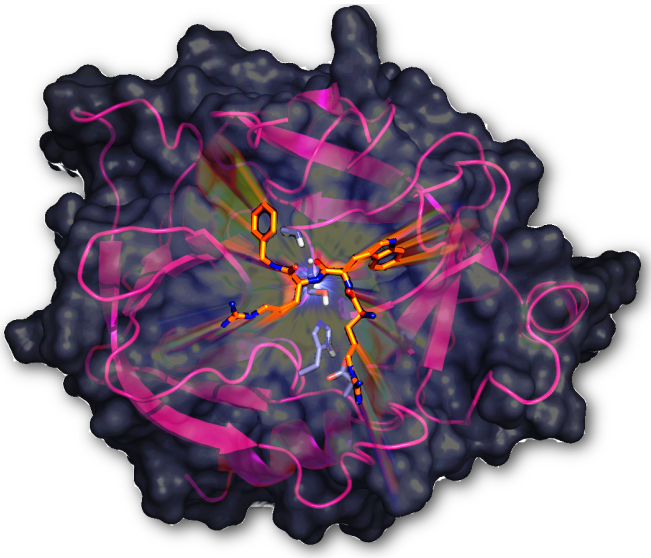
KJE-3900

Master's Thesis in Chemistry

Flexible membrane active antimicrobial tripeptides with stability towards chymotryptic degradation

Geir Isaksen

May, 2010



Faculty of Science
Department of Chemistry

University of Tromsø

KJE-3900

Master's Thesis in Chemistry

Flexible membrane active antimicrobial tripeptides with stability towards chymotryptic degradation

Geir Isaksen

May, 2010

Faculty of Science
Department of Chemistry

University of Tromsø

Flexible membrane active antimicrobial tripeptides with stability towards chymotryptic degradation

Geir Isaksen

May, 2010

Keywords: Antimicrobial peptides; tripeptides; geometry optimization; molecular dynamics; phase space; face flipping; membrane interaction; chymotrypsin; docking; isothermal titration calorimetry; molecular modelling; crystallization; crystal structure.

Abstract: In recent years antimicrobial peptides have gained a lot of attention due to their potential as a new generation of antibiotics combating the growing problem of antibiotic resistance. It is believed that the amphipathic structure of cationic peptides is a key feature for antimicrobial activity, and that this enables them to interact with the bacterial cell membrane. The conformational space of a range of cationic tripeptides have in this project been studied in solvent using density functional theory and molecular dynamics simulations. The results indicate that the cationic tripeptides are able to change between different, but equally stable, conformations that are both amphipathic and non-amphipathic, a property referred to as face flipping. Based on this, face flipping is proposed to be a key feature for the membrane interaction mechanism. The tripeptides mode of interaction was therefore studied with cellular model systems in more detail using MD simulations. The results show that the peptides first interact with the negatively charged head groups of the membrane with their cationic charges and then flip the hydrophobic groups into the membrane bilayer. The results thus provide strong support to the face flipping hypothesis.

A problem with antimicrobial peptides is that oral administration is difficult due to the degradation by digestive enzymes. The stability towards chymotryptic degradation has therefore been investigated by probing the S1, S1' and S2' binding pockets with unnatural amino acid side chains. The effect of different side chain substitutions were examined by combining isothermal titration calorimetry, crystallization experiments and extensive molecular modelling. Through these studies it was possible to investigate the preferential binding of several relevant unnatural amino acid side chains to the subsites of chymotrypsin. Important structural and mechanistic features were revealed in a fashion not feasible through the use of native peptide substrates. It was also found that proteolytic stability can be controlled not only by probing the S1 pocket, but also the notably less studied S1' pocket.

Acknowledgments

The presented work was done at the University of Tromsø in the period between August 2008 and May 2010. I would like to greatly acknowledge my supervisor Bjørn Olav Brandsdal for giving me the opportunity to be a part of a very interesting and challenging project, and for always supporting and guiding my ideas in the right direction. Thank you for all the necessary help and feedback through the writing process, and for giving me the freedom to grow, not only as a chemist, but also as a scientist. I would like to bring a special thanks to Hanna-Kirsti Leiros who has been of great help during the crystallization part of the project; I appreciate the time you spent explaining and sharing from your experience as a crystallographer making the world of structure determination a lot easier to maneuver in. I would also like to thank Ronny Helland for all the nice discussions we have had during my master, for all the advises concerning crystallization conditions, and for donating chymotrypsin. Johan Svenson and Rasmus Karstad for contributing with Half-life assays and MIC values. Johan for always bringing me more peptides for my never ending crystallization experiments. Kathrin Helen Hopman for all her computational help and personal effort during my attempt to geometry optimize the petides in solvent with density functional theory. Annfrid Sivertsen for always offering a helping hand and for borrowing me peptides when I was all out (and Johan could not bring me more fast enough). Thank you for all the time you spent teaching me with the ITC and introducing me to the world in the "wet lab". I would also like to thank Bjarte A. Lund for learning me how to use the crystallization robot (several times). Antoine Durdek for his mathematical expertise and for making our office a nice and funny place to be in. Arnfinn Hykkerud Steindal for providing a helping hand whenever I was struggling with linux or LaTeX.

As a member of both NorStruct and CTCC I have met incredibly many friendly and helpfull people. I would therefore like to thank all the present and former members of NorStruct and the CTCC that have contributed to my work and working environment. All my friends for skiing, serious squash playing, concerts (playing and listening), cold beers at flyt, and in general for reminding me that there is a life outside the office during the hard writing period of this thesis. I would also like to thank my parents for all their unconditional support.

My special heartfelt gratitude goes to my fiancee, Silje Jørgensen, for all her love and support, and for having a huge tolerance for my ever workaholic tendencies. Thank you for making all of this possible ♡.

Geir Isaksen

Abbreviations & symbols

∇^2	Laplace operator; $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$
χ	Atomic orbital
ϕ	Molecular orbital
Φ	Slater determinant
Ψ	Many electron wave function
AO	Atomic orbital
Agp	L-2-amino-(3-guanidino)propanoic acid
Amphiphile	Chemical compound possessing both hydrophilic and lipophilic properties. (such compunds are called amphiphilic or amphipathic)
AMPs	Antimicrobial peptides
App	L-2-amino-3-(4-aminophenyl)propanoic acid
ATTC	American type culture collection
Bip	L-biphenyl alanine
Bn	Benzyl
BPTI	Bovine Pancreatic Trypsin Inhibitor
CAP	Cationic Antimicrobial Peptide
CC	Coupled cluster
CHT	Chymotrypsin
CI	Configurational interaction
CM	Cytoplasmic membrane
DFT	Density functional theory
DIIS	Direct Inversion of the Iterative Subspace (convergence algorithm)
DI	Differential algorithm
Dip	L-β-phenyl-phenylalanine
DPG	Diphosphatidylglycerol
DPI	Cruickshank's dispersion precision index for coordinate error: $DPI = \sqrt{(N_{atom}/(N_{refl} - N_{param}))} R_{fact} D_{max} compl^{-\frac{1}{3}}$
EC	Energy correlation
Et	Ethyl
FISA	Hydrophilic solvent accessible surface area
Force field	A sum of different molecular mechanics terms expressing the energy of a system
FOSA	Hydrophobic solvent accessible surface are
GA	Genetic algorithm

GGA	Generalized gradient approximaton
Gpp	L-2-amino-3-(4-guanidinophenyl)propanoic acid
GRP	Generalized random phase
Har	L-2-amino-(4-guanidino)hexanoic acid
HF	Hartree-Fock
ITC	Isothermal titration calorimetry
LSDA	Local spin density approximation
MBPT	Many-body perturbation theory
MC	Monte carlo
MD	Molecular dynamics
MIC	Minimal inhibitory concentration
MR	Molecular replacement
MRSA	Methicillin-resistant *staphylococcus aurerus*
MO	Molecular orbital
OM	Outer membrane
Orn	L-ornithine
PCM	Polarizable continuum model
PE	Phosphatidylethanolamine
Ph	Phenyl
Phatogen	Organism, usually microorganism, that causes disease
PISA	π component of the solvent accessible surface area
POPE	1-palmitoyl-2-oleoyl-glycero-3-phosphoethanol-amine
PG	Phosphatidylgylcerol
PRCG	Polak-Ribiere-type conjugate gradient
QM	Quantum mechanics
RBD	Rigid-body docking
RMSD	Root mean square distance between pairs of atoms
	$\text{RMSD}= \sqrt{\frac{\sum_{i=1}^{N_{atoms}} d_i^2}{N_{atoms}}}; \qquad d_i = \lvert x_i - x_i' \rvert$
RP-HPLC	Reversed Phase High Performance Liquid Chromatography
SA	Simulated annealing
SASA	Solvent accessible surface area
SCF	Self-Consistent Field
SCRF	Self-consistent reaction field
Tbt	β-(2,5,7-tri-tert-butylindol-3-yl)alanine
UFF	Universal Force Field
VRE	vancomycin-resistant *Enterococcus faecium*
XP	Extra precision (in Glide docking)

Contents

List of Tables

List of Figures

"It would be possible to describe everything scientifically, but it would make no sense; it would be without meaning, as if you described a Beethoven symphony as a variation of wave pressure."

-Albert Einstein

1. Introduction

1.1 Antimicrobial resistance

In 1929 a British scientist named Alexander Flemming characterized an antibacterial product of the fungus *Penicillium chrysogenum* as the β-lactam antibiotic *penicilllin* G. He had discovered what became the first clinically effective antibiotic. In 1945 penicillin became available for general use, and pharmaceutical companies soon began to look for other antibiotics, which in turn led to the discovery of drugs that revolutionized treatment of infectious diseases [1].

Throughout the years many effective drugs and vaccines have been discovered or synthesized. However, in spite of the availability of effective drugs and vaccines, the battle against infectious diseases is far from being over. The emergence and spread of antimicrobial resistance is now threatening to undermine the ability to treat infections and save lives [2]. The drug resistance itself is an outcome of natural selection and should perhaps be viewed as an expected phenomenon of the Darwinian biological principle of "survival of the fittest". The need for new and better drugs against microorganisms like bacterias, viruses and fungous is increasing rapidly and the ongoing race between new drugs and antimicrobial resistance seems like a neverending process. Even though resistance was recognized early it is only just beginning to be considered as a societal issue and, in economic terms, as a negative externality in the health care context. It has recently been described as a threat to global stability and national security [3]. Some pathogens have now developed resistance to all known antimicrobial agents. Among these are several isolates of methicillin-resistant *Staphulococcus aurerus* (MRSA), vancomycin-resistant *Enterococcus faecium* (VRE) and isolates of *Mycobacterium tuberculosis* [1]. However, the biggest concern today is that the pipeline of new drugs is running dry and the incentives to develop new antimicrobials to address the global problems of drug resistance are weak [3].

One of the main reasons for resistance is overuse of drugs and antimicrobial resistant microorganisms typically emerge from hospitals. Due to these serious problems the world health organization (WHO) started an implementation workshop on the global strategy for containment of antimicrobial resistance in November 2002. Guidelines are being provided to slow the emergence and reduce the spread of antimicrobial resistant microorganisms. Some of the key points made by the WHO framework are to slow the emergence and reduce the spread of resistance through [3]:

- Reducing the disease burden and the spread of infection.
- Improving access to appropriate antimicrobials.

- Improving use of antimicrobials to slow the emergence and reduce the spread of antimicrobialresistant microorganisms.
- Strengthening health systems and their surveillance capabilities.
- Enforcing regulations and legislation.
- Encouraging the development of appropriate new drugs and vaccines.

The development of new and appropriate drugs are clearly necessary, but as the pipeline of new drugs is running dry, this is today a challenging task.

1.2 Antimicrobial peptides

Antimicrobial peptides (AMPs), which sometimes goes under the name host defense peptides, are evolutionarily conserved components of the innate immune system which are found among a wide variety of organisms as their first line of defense [4]. Antimicrobial peptides are a unique and diverse group of molecules typically ranging from 12 to 100 amino acids having a net positive charge along with an amphiphillic structure. These AMPs have been isolated from single-celled microorganisms, plants, amphibians, birds, fish, mammals (including humans), insects and other invertebrates [5, 6]. In the Antimicrobial Peptide Database there are currently (August, 2009) reported 1474 such peptides with an average length of 29.87 amino acids and an average charge of 3.81 [6, 7]. Among these AMPs 5.83% are antiviral, 29.3% antifungual, 6.37% anticancer (antitumor) and 77.81% are antibacterial. Even though it is found that such peptides may have a direct effect on the microbe, by damaging or destabilizing the bacterial, fungal, or viral membrane or acting on other targets, they appear to be broadly involved in the orchestration of the innate immune and inflammatory response [8].

Despite the AMPs similar general physical properties it appears that there are very limited sequence homologies and a wide variety of secondary structures with at least four major themes. To date the most prominent structures reported are amphiphillic peptides with two to four β-strands, amphipathic α-helices, loop structures, and extended structures [9, 10]. Of special interest is the fact that cationic AMPs (CAPs) have been found to show an unusual broad antimicrobial activity spectrum, and therefore constitute an attractive class of compounds for further development against bacterial strains resistant to the antibiotics of today. Their amphiphillic and positively charged structure have been proposed to interact with the negatively charged components of the cytoplasmic membrane of bacteria and therebye increasing the permeability to ions and solutes, which in turn may lead to cell death [11]. It is believed that this takes place via a disruption of the membrane, either via an ordered pore formation, which is described by the barrel-stave model and the toroidal model [12], or in a disordered manner, where the peptides interact with the surface and destabilize the membrane, resulting in collapse, which often is referred to as the carpet model [13]. A combination of the ordered and disordered mechanism has also been suggested as a possible mechanism [15–18]. The models are illustrated in Figure 1.1. All these models require, either explicitly or implicitly, that a certain threshold concentration, given as lipid:peptide (L:P) ratio, in the membrane must be crossed for disruption to occur. In some cases the

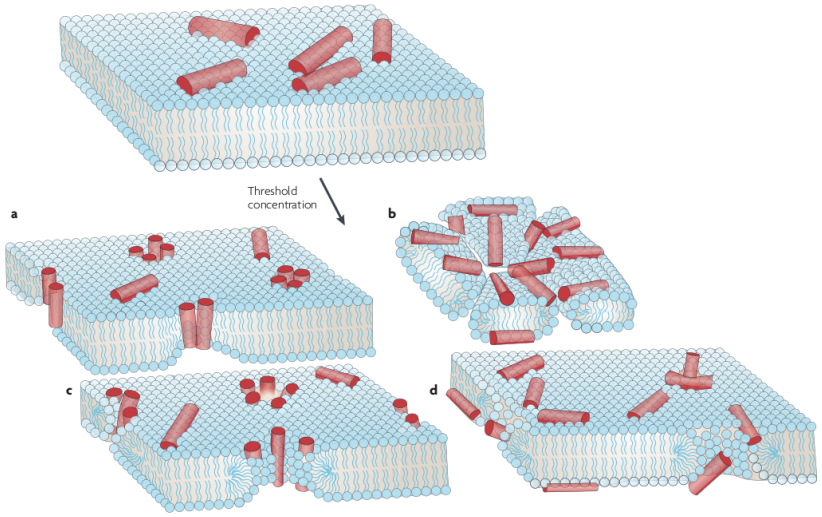


Figure 1.1: Illustration of proposed mechanisms of antimicrobial peptidemediated membrane disruption (adapted from Melo et al [14]): a) The barrel-stave pore, b) the carpet mechanism, c) the toroidal pore and d) the disordered toroidal pore.

threshold concentration can be as high as one peptide molecule per six phospholipid molecules, which is close to full membrane saturation [19]. The minimum inhibitory concentration (MIC) for AMPs are typically in the low micromolar range, and scepticism has arisen regarding the relevance of these thresholds and their importance *in vivo* [20, 21]. The macroscopic effects of crossing the MIC are easily observed by for example cell lysis, loss of viability or bacteriostasis (a phase in which microbial organisms are prevented from undergoing further cell growth, without being killed) [22]. It is however not clear if the final killing event of CAPs is due to damaged cytoplasmic membrane or a combination of membrane permeabilization and peptide interactions with intracellular targets [23, 24]. Thus, the exact mechanism of interaction is still not known in detail, and the models proposed so far might be too rigid to fully account for the many interactions that these molecules can establish in a complex membrane environment. This has previously been demonstrated with molecular dynamics simulations of AMPs interacting with phospholipid bilayers [25–27]. The observations from these studies indicated multiple co-existent structures, non-specific peptide-peptide interactions and membrane perturbation dictated by stochastic events. This indefinite behavior has the advantage that the bacteria seem to find it hard to circumvent the AMP action and the evolution of resistance is consequently more difficult [28].

Even though the mechanisms of AMPs action are not always precisely defined, some main factors emerge that lead to high levels of binding and selectivity towards bacteria (Table 1.1) [29]. Peptide cationic charges are considered as one of the key features for stronger interaction with the anionic bacterial membrane. Further enrichment of the peptides with hydrophobic amino acids together with cationic residues have been found to correlate with stronger partitioning to membranes, but also with higher levels of hemolytic activity. Another key feature is the amphipathicity of the peptides. During membrane interaction, AMPs adopt

Table 1.1: Antimicrobial peptides and bacterial characteristics [29].

Characteristic	Observed effect
Peptide:	
Cationic charge	Stronger interaction with anionic bacterial membrane.
Hydrophobicity	Stronger portioning to membranes. Higher levels of hemolytic activity.
Amphipathicity	Higher levels of peptide internalization and membrane perturbation.
Bacteria:	
Anionic charge	More fluid bilayers which are more susceptible to binding and probable disruption by cationic AMPs.
Transmembrane potential	Promotes AMP interaction while acting as a potential driving force for for peptide insertion and translocation.
Bilayer asymmetry	Differences in inner and outer leaflet composition provides a basis for further discrimination.

preferred conformations that frequently involve segregation of polar and apolar residues to opposite regions of the structure. Amphipathic structures therefore leads to higher levels of peptide internalization and membrane perturbation.

In contrast to mammalian membranes, which mainly consist of neutral phospholipids enriched with sterols, the bacterial membranes are composed of a large proportion of anionic phospholipds. This feature makes cationic peptides more target specific towards bacterial membranes, where they also can enter the fluid bilayers more easily compared with mammalian membranes[29]. A schematic presentation of the bacterial cell membranes for Gram-positive and Gram-negative bacteria is illustrated in Figure 1.2. The main difference between

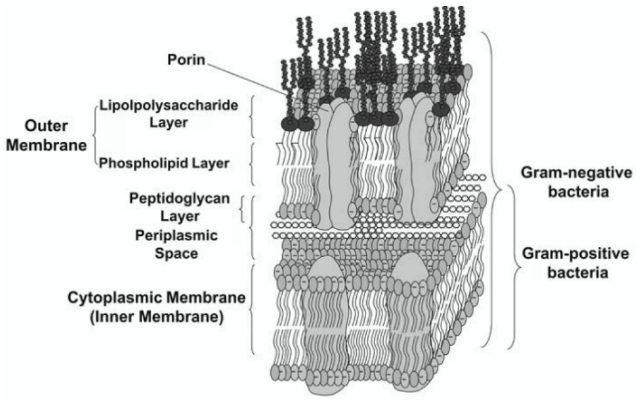


Figure 1.2: Schematic presentation of molecular organization of bacterial cell membranes. (Adapted from Lohner *et al.* [30])

these two is that Gram-negative bacteria consist of an outer membrane with an asymmetric distribution of lipopolysaccharides (LPS) and phospholipids, predominantly phosphatidylethanolamines (PE), and a cytoplasmic membrane.

Gram-positive bacteria on the other hand only have a cytoplasmic membrane, protected by a peptidoglycan layer, which is also found in the periplasmic space of Gram-negative bacteria. Common for both bacteria is that phosphatidylgylcerol (PG) is the most abundant negatively charged phospholipid species [31]. The phospholipid composition in Gram-positive bacteria has further been found to vary more from one species to another compared with Gram-negative bacteria [32]. Some examples of the membrane composition of selected Gram-positive and Gram-negative bacteria are shown in Table 1.2. Detailed knowledge of the

Table 1.2: Phospholipid composition of selected Gram-negative and Gram-positive bacteria[α,β].

Bacteria species		PG[f]	DPG[g]	L-lysl PG[h]	PE[i]	Others
Gram-negative						
Escherichia coli	OM[a]	3	6	0	91	0
	CM[b]	6	12	0	82	0
Pseudomonas cepacia	OM	13	0	0	87	0
	CM	18	0	0	82	0
Salmonella typhimurium	OM	17	2	0	81	0
	CM	33	7	0	60	0
Erwinia carotovora	OIM[c]	14	8	0	78	0
Gram-positive						
Staphylococcus aureus	CM	57	5	38	0	Trace
Micrococcus luteus	CM	26	67	0	0	7[d]
Bacillus megaterium	CM	40	5	15	40	0
Bacillus subtilus	CM	29	47	7	10	6[e]

[α] Data from refs. 33 and 34
[β] Values are in percentage of total phospholipid composition.
[a] OM, outer membrane.
[b] CM, cytoplasmic membrane.
[c] Phospholipid composition not different between outer and inner membrane.
[d] Almost exclusively phosphatidylinositol
[e] Including phosphatidic acid and glycolipids
[f] PG, Phosphatidylgylcerol
[g] DPG, Diphosphatidylglycerol
[h] L-lysl PG, Aminoacyl derivative of PG
[i] PE, Phosphatidylethanolamine

specific membrane composition can thus serve as a potential guide in the AMP target specificity, such as differing Gram-positive from Gram-negative bacteria.

A lot of research has been done throughout the years on CAPs since they first were discovered in the hemolymph of insects in the late 1970s [35]. The amphiphilic structure of cationic antibacterial peptides is today considered to be a key requisite for the peptides to interact with the bacterial cell membrane [11]. In contrast to antibiotics in regular clinical use, which interact with intracellular target, the membrane-active peptides show low propensity for development of resistance, and CAPs have therefore been heralded as new antibiotic drug candidates targeting multiresistant pathogens. Despite all promising features of CAPs, they have seen limited success so far. Some of the explanation for this might be that they are quite large (∼30 residues) compared with traditionally antibiotics and they have unresolved issues regarding toxicity and stability. The large size of CAPs makes the production cost as well as a potential lack of sys-

tematic applicability a big concern. It is therefore of great interest to develop smaller CAPs in order to alleviate these concerns.

1.3 Short CAPs targeting multiresistant pathogens

Our research group has for more than a decade focused on simplifying the complex AMP molecules in an attempt to bring antimicrobial peptides closer to clinical use [36]. One important step in this direction, discovered by using derivatives of lactoferricin, was the identification of a minimal set of structural motifs, i.e., a pharmacophore (an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger, or block, its biological response [37]) for antibiotic activity, consisting of two cationic charges and two hydrophobic bulk groups [38]. These small peptides (2 to 6 residues) showed substantially less hemolytic activity and are thus more selective towards prokaryotic cells [39]. However, while it is possible to prepare a wide array of pharmacologically active peptides, their applicability *in vivo* are limited by their unfavorable properties, with regard to administration, narrow therapeutic index, toxicity, and stability. In order to reach an infected target, the peptide drugs can be placed either directly on the skin, injected into the blood or taken orally. The two last options both face the problems of digestive enzymes. Oral administration of these small CAPs is of great interest, but is difficult because of digestive enzymes in the gastrointestinal tract where peptides usually are metabolized by enzymatic cleavage of the amide bonds joining the amino acids. Stability towards proteolytic degradation is thus important for therapeutic peptides as a means of increasing the plasma half life time [40].

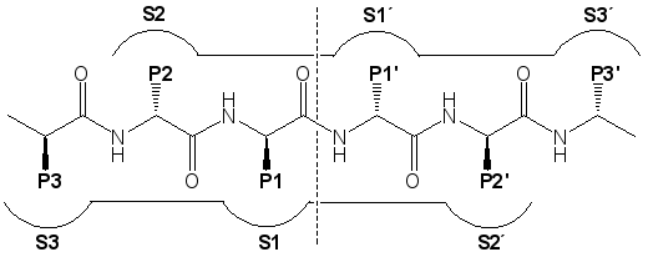


Figure 1.3: Schematic representation of the nomenclature used for serine proteases. The binding pockets are named S1-Sn and S1'-Sn' where S1 is the binding site for the peptide residue (P1) on the acyl side of the scissile bond and S1' is the binding site for the peptide residue (P1') on the leaving group side of the scissile bond [41].

The proteolytic enzymes responsible for peptide degradation fall into two general classes, the endopeptidases, which cleave the peptide sequence at more or less specific sites, and the exopeptidases, which cleave the terminal amino acids [42]. Proteinases, also called proteases or peptidases, describes a group of enzymes that catalyze the hydrolysis of covalent peptidic bonds. Proteinases form the four functional families serine, cysteine, aspartic and metallo proteinases [43]. Chymotrypsin and trypsin, both members of the serine proteinase family characterized by a nuclephillic serine residue in the active site, are key enzymes involved in protein and peptide degradation in the gastrointestinal (GI)

tract. These two endopeptidases have similar three-dimensional structure, but different substrate specificities [44, 45] which is governed by the substrate recognition sites. These include the polypeptide binding site and the binding pockets for the peptide side chains according to the nomenclature illustrated in Figure 1.3 [41].

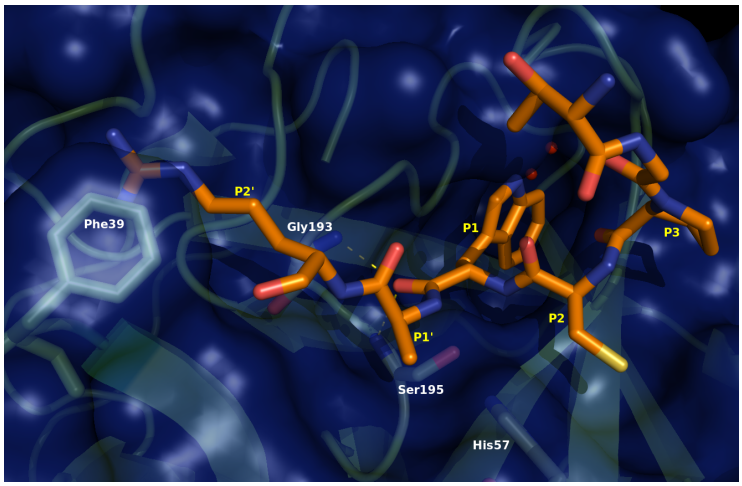


Figure 1.4: Orientation of the P1 Thr BPTI-bovine chymotrypsin complex around the catalytic triad from crystal structure with PDB entry code 1t8o [46].

Chymotrypsin-like proteases are two domain structures arranged into two six-stranded β-barrels with the active site located between these two barrels [47]. The catalytic triad spans the active site with Ser195 on one side and His57 and Asp102 on the other side. The S1 pocket of chymotrypsin is formed by the residues 189-192, 214-216, and 224-228 where Ser189, Gly216 and Gly226 are considered to be the main residues that defines the substrate specificity [48, 49]. These S1 residues create a deep and hydrophobic pocket and chymotrypsin typically cleaves C-terminally to large lipophillic amino acids, such as Trp and Phe, where Phe as P1 is preferred over for example Ala by a factor of 5000 [50, 51]. In contrast, residue 189 of trypsin in the S1 pocket is Asp, and trypsin consequently preferres to cleave C-terminally to cationic residues, such as Arg and Lys [52]. In Figure 1.4 the crystal structure of the P1 Thr BPTI-bovine chymotrypsin complex is shown illustrating the orientation of the inhibitor in the active site[46].

When a peptide-substrate is oriented correctly in the active site it will be cleaved between the P1 and P1' as illustrated in Figure 1.5 where the general serine proteinase reaction mechanism in shown. Ser195 attacks the carbonyl of the peptide substrate assisted by His57 acting as a general base, to yield a tetrahedral intermediate. The protonated His57 is then stabilized by a hydrogen bond to Asp102, whereas the oxyanion of the tetrahedral intermediate is stabilized by interaction with the main chain NHs from Ser195 and Gly193, defined as the oxyanion hole (Figure 1.4). The peptide bond is then cleaved as the tetrahedral intermediate collapses with expulsion of the leaving group, assisted by the protonated His57, now acting as a general acid, to yield the acyl enzyme intermediate. In the deacylation half of the reaction water attacks the acyl en-

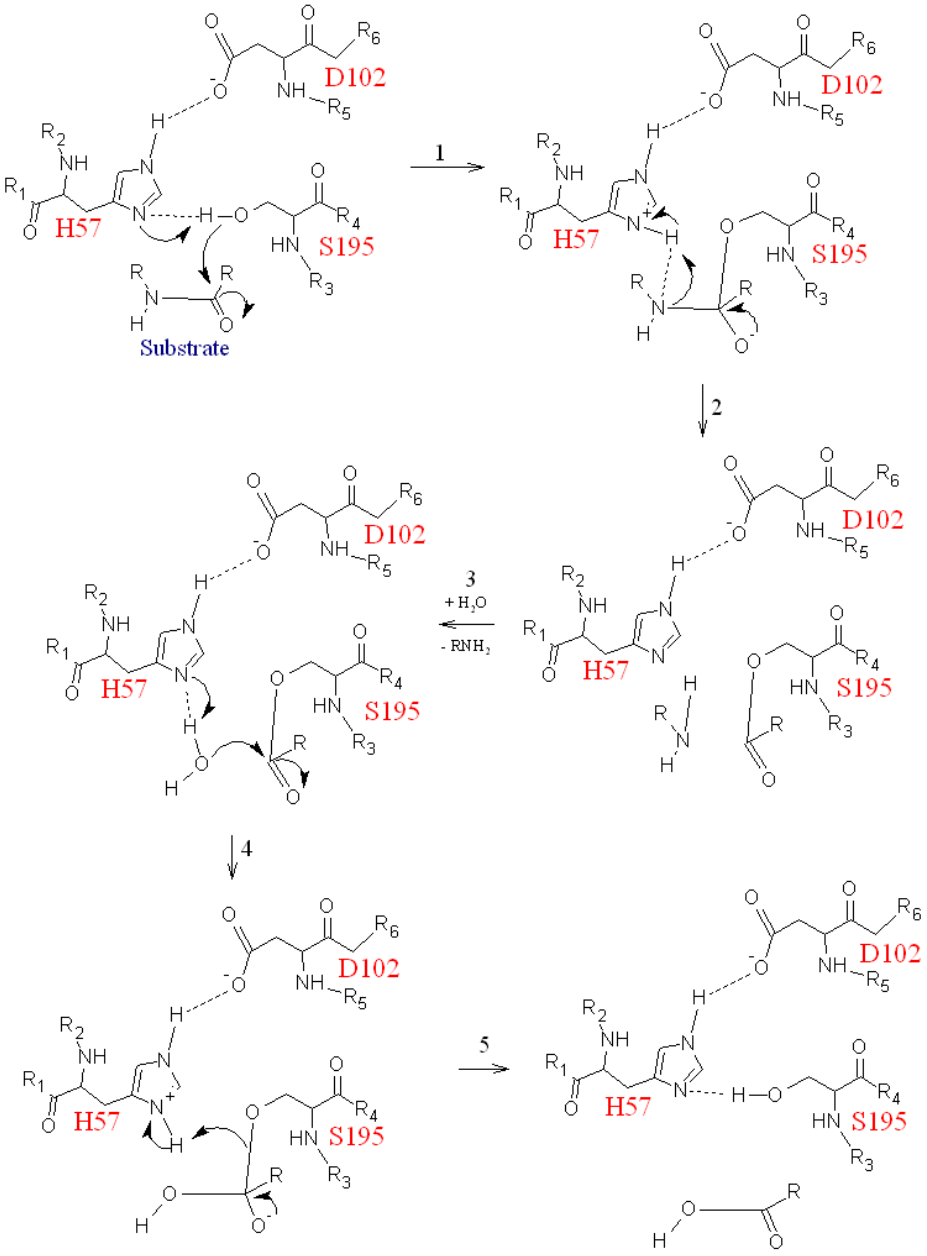


Figure 1.5: Serine proteinase reaction mechanism.

zyme, assisted by His57, yielding a new tetrahedral intermediate. When this second intermediate collapses a carboxylic acid product is expelled and Ser195 returns to the native state being hydrogen bonded to His57. It must however be kept in mind that this is a simplification of the real reaction which involves a network of hydrogen bonding interactions that links the substrate binding sites to the catalytic triad. This in turn means that as the reaction proceeds,

changes in bonding and charges at the scissile bond will propagate to more remote enzyme-substrate interactions, and vice versa. Supporting evidence for this mechanism has been taken from references [53–57].

In order to avoid proteolytic degradation of CAPs it is necessary to design them in a way that can circumvent the reaction mechanism of these enzymes. In recent years several strategies have been developed to improve the peptides pharmaceutical properties. In 1999 Brinckerhoff *et al.* [58] reported that the action of exopeptidases can be prevented by so-called end capping, which involves chemical modification of the N- and C-terminal amino or carboxylic acid functionalities, respectively. Inhibition of degradation by endopeptidases has been obtained through methylation of the amide nitrogen atom or inversion of the stereochemistry of the amino acid responsible for substrate specificity in the enzyme [59, 60]. It has also been shown that constrained cyclic analogues of linear peptides and incorporation of unnatural amino acids at selected positions increase the stability towards degradation and the duration of action [61–63]. Other strategies that have been successfully employed to generate peptides with both increased stability and activity involve the replacement of the scissile amide bond bond with peptidomimetic elements or co-administration of specific enzyme inhibitors [64, 65].

Our group has for the last ten years been developing short CAPs based on truncated analogs of bovine lactoferricin in an effort to generate novel potent antibiotic drugs with a low susceptibility for bacterial resistance [66]. The antimicrobial properties of these small cationic peptides have been increased by including unnatural amino acids in the structures [39]. Substitution of amino acids within a peptide sequence with unnatural analogues is thus a well established strategy to change the peptide properties. The reported effects of such substitutions involve reduced hemolytic activity, increased stability towards proteolytic degradation, and increased potency [64, 67, 68]. To date some of the promising CAPs are tripeptides based on arginine. In a recent report the stability towards tryptic degradation of some promising tripeptides was studied with RP-HPLC and isothermal titration calorimetry. It was found that trypsin was able to cleave some of the small cationic antimicrobial peptides surprisingly well and binds them with dissociation constants ranging from 1 to 20 μM [36]. Important insight in the detailed binding mode of short peptides to trypsin was gained in this study, and it was found that the peptides should be kept as small as possible in order to limit interactions with as many binding sites in trypsin as possible. It was also suggested that the amino acid occupying the P2 site should have a side chain sterically preventing it from binding to the S2 unit of trypsin, and that the length and the stereochemistry of the C-terminal end capping had a major influence on the stability. Peptides with for example a phenyl group connected via an ethyl chain to the amide nitrogen were found to be significantly more stable than those connected via a methyl or a propyl link [36]. The previous report presented several stable peptides towards tryptic degradation with good antimicrobial efficacy against *S. aurerus*. However, while tryptic stability is important for a potential peptidic drug, so is the stability towards degradation by chymotrypsin, which is topic of the work presented here.

1.4 Structure, Binding & Catalysis

In order to design new drug candidates effectively it is of utmost importance to know the target (protein, DNA, cell membrane etc.) and the drug-target interactions in detail. Today there are many methods available for structure determination, and the choice of method is strongly dependent on factors such as the molecule size, solubility and stability. The number of protein structures in the protein data bank (PDB) [69] has been increasing fast for the last 10-15 years (Figure 1.6) with 63438 experimentally determined structures currently deposited (February 22^{nd} 2010). X-ray crystallography and NMR are to date

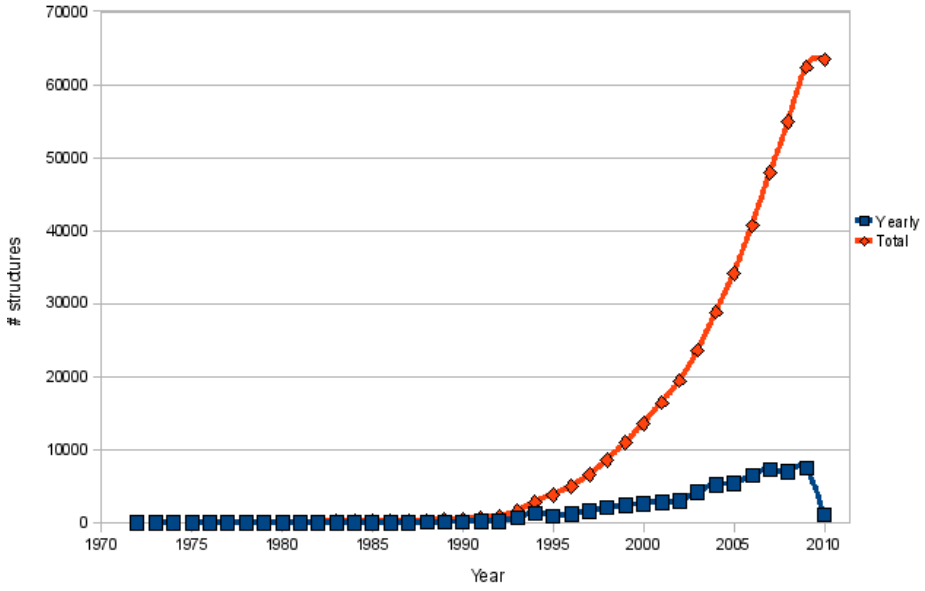


Figure 1.6: Yearly growth of total structures in the protein data bank.

the most dominant methods for structure determination, both with limitations and advantages. NMR spectroscopy has the advantage that it can capture motion and is thus particularly well suited for analyzing proteins that contain extensive flexible regions or do not harbour well-defined structures. The main disadvantage with NMR spectroscopy is the molecule size limitation (∼80 kDa). X-ray crystallography on the other hand has no size limitations and often provides better resolution compared with NMR. The main disadvantage with X-ray crystallography is that the molecule must crystallize, which in many cases can be very difficult.

Binding constants and reaction rates can be experimentally determined with techniques like equilibrium dialysis, radio-ligand binding assays, ultracentrifugation, or isothermal titration calorimetry (ITC). It is also possible to determine binding constants more indirect with spectroscopic methods. However, with increasing computer power, computational chemistry is becoming an increasingly important and reliable method for both structure prediction and interaction studies. Computational modeling of molecules can give important insight in molecular properties such as geometric and electronic structures, frequencies

and relative energies. One of the main strengths with computational methods is the opportunity to study species, processes or conditions that are difficult or even impossible to obtain experimentally.

1.4.1 X-ray Crystallography

Solving a protein structure with X-ray crystallography requires crystals. Proteins, like many molecules, can be prompted to form crystals when placed in the appropriate conditions. Crystals are typically grown by slow, controlled precipitation from aqueous solution under conditions that do not denaturate the protein. As a result, individual protein molecules align themselves in a repeating series of unit cells by adopting a consistent orientation. This serves as the basis for X-ray crystallography where the crystallized protein is used to determine the protein's three-dimensional structure via X-ray diffraction. There are many substances that can potentially cause proteins to precipitate. Typically this involves the use of salts which precipitate proteins by a process called "salting out". However, obtaining a crystal can in many cases be extremely difficult because there are so many variables involved such as pH, salt, organic solvents, additives, concentration, volume, temperature and so on. Crystallization experiments therefore involves a great deal of trial and error, and there are no guarantees for obtaining good crystals, or crystals at all.

Crystals are formed in the two stages nucleation and growth (Figure 1.7). In theory precipitation should occur when the combination of protein and precipitant concentrations exceeds a certain threshold. Ideally one would start with conditions corresponding to the purple region of the phase diagram in Figure 1.7. Once the nuclei has formed, one would like to move into the green region, where growth, but not additional nucleation can form [70].

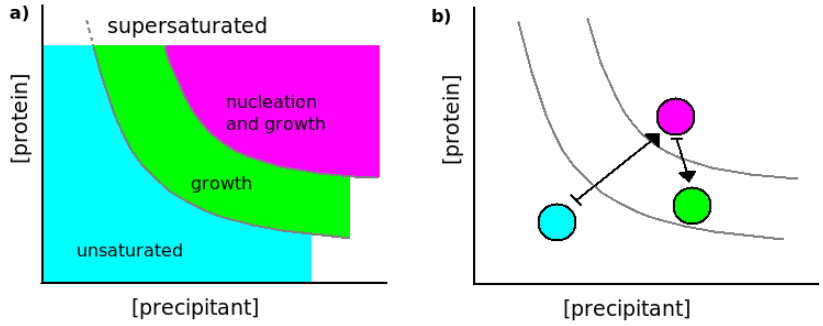


Figure 1.7: a) Illustration of a phase diagram for crystallization mediated by a precipitant. The blue region represents concentrations of protein and precipitant at which the solution is not saturated with protein, so neither nucleation nor growth occurs. The green and purple regions represent unstable solutions that are supersaturated with protein. In the purple region both nucleation and growth are supported, whereas in the green region only growth is supported. b) Illustration of an ideal strategy for growing large crystals. First nucleation is allowed to occur under the conditions in the purple region and secondly the condition is moved to the green region until crystal growth ceases.

There are several techniques for crystallization, but the most popular is the

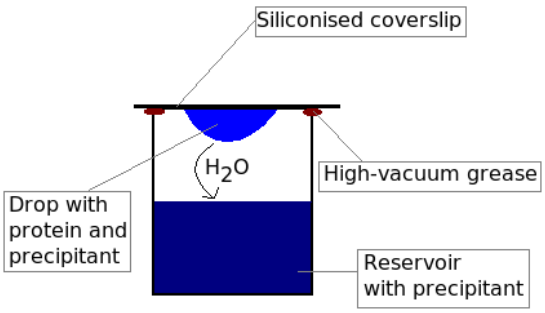


Figure 1.8: Illustration of the hanging drop vapor diffusion method.

vapor diffusion method. Hanging drop (Figure 1.8) and sitting drop are both vapor diffusion methods based on the same principle [70]. The main idea is that protein, buffer and precipitant is mixed in a drop which is placed on a glass coverslip. The reservoir solution normally contains the same as the drop, except from the protein. Normally the precipitant concentration in the drop is half of that in the reservoir. When the coverslip is placed on top of the well and the system becomes closed, water from the drop will diffuse through the vapor space and into the reservoir due to the differences in precipitant concentration (Figure 1.8). As water leaves the drop, the protein and precipitant concentration increase, bringing the system into the nucleation and growth region in the phase diagram (Figure 1.7). If the conditions are ideal the system will fall back to the growth region after being in the nucleation phase. Bad conditions either give many tiny crystals or amorphous precipitation.

The atoms in a crystal are arranged in a regular lattice, and the spacing between planes of atoms (referred to as Bragg planes) is comparable to the wavelength of X-rays. Crystals can therefore diffract X-rays. In 1913 W. L. Bragg interpreted a diffraction pattern obtained when X-rays were directed at crystal through the well known Bragg's law;

$$2d_{hkl} \sin\theta = n\lambda \tag{1.1}$$

where d is the spacing between the planes, hkl defines the plane, θ the angle between the incident wave and the plane, λ the wavelength, and n is an integer. In order for waves to interfere constructively, the difference in travel path must be equal to integer multipliers of the wavelength. When constructive interference occurs, a diffracted beam of X-rays will leave the crystal at an angle equal to that of the incident beam and intense reflected X-rays are produced (Figure 1.9). There are many planes (defined as miller planes with the indices h, k, l) within a crystal, and the crystal will consequently diffract the source beam into many discrete beams, each which produces a distinct reflection on a detector (usually image plate detectors or charge-coupled devices, CCD). The intensity of the reflection from a plane hkl is dependent on the electron density $\rho(x, y, z)$ on that plane. This in turn is dependent on the atom types present, and thus gives the information for potentially solving the structure. However, the detector only measures the amplitude, which is related to the intensity, of the diffracted wave. The phase of the wave is consequently lost in the measure-

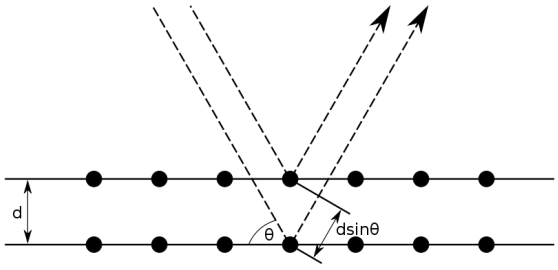


Figure 1.9: Bragg diffraction. When Bragg's law is fulfilled, the diffracted X-rays from the different planes are interfering constructively.

ment, a problem referred to as the "phase problem". If all the phases are known, the electron density could have been directly obtained by a Fourier transform operation [70]. Several techniques exist for obtaining the phases, such as multiwavelength anomalous diffraction (MAD), multiple isomorphous replacement (MIR), and single wavelength anomalous dispersion (SAD). However, with an increasing number of solved protein structures available in the protein data bank, a process called molecular replacement (MR) is often the choice of method for obtaining the phases when doing for example ligand studies. MR makes use of similar molecules by simulating the molecule's packing in the crystal and therebye obtaining theoretical phases which are to be used on the collected data set. The method is however less desirable for solving new structures since it can severely bias the solution of the structure. It must however be kept in mind that crystal structures only give information about the atom coordinates, and not the underlying complex formation.

1.4.2 Isothermal titration calorimetry

When a ligand binds to a protein, heat is either generated or absorbed. Isothermal titration calorimetry (ITC) is a unique thermodynamic technique for directly measuring such biomolecular interactions through the heats absorbed or released. With an ideal adiabatic calorimeter there is no heat exchange between the calorimeter and the surroundings, and the heat quantity Q evolved during the experiment is thus directly proportional to the observed temperature change ΔT, and the heat capacity ϵ of the reaction vessel and its content [71]:

$$Q = \epsilon \cdot \Delta T \tag{1.2}$$

By measuring the interaction heats, the association constant (K_a), reaction stoichiometry (n), and thermodynamic parameters including free energy of binding (ΔG), enthalpy (ΔH) and entropy (ΔS) can be determined using the relationship:

$$\Delta G^0 = -RT \ln K_a = \Delta H^0 - T\Delta S^0 \tag{1.3}$$

where T is the temperature in Kelvin and R the universal gas constant. K_a is the reciprocal of the equilibrium dissociation constant K_d ($K_a = 1/K_d$). It is also possible, by doing experiments at different temperatures, to determine the change in heat capacity at constant pressure:

$$\Delta C_p = \frac{d(\Delta H)}{dT} \tag{1.4}$$

Measurement of ΔC_p in addition to ΔG^0, ΔH^0 and ΔS^0 allows a full thermodynamic characterization of the binding interaction, as it is this parameter that controls the temperature variation of the other three parameters [72]. The main advantage with ITC is thus that the enthalpy change for a biomolecular binding interaction at a constant temperature is measured directly and without the need of a predominated model [73].

A modern ITC instrument typically consists of two identical cells composed of a highly efficient thermal conducting material surrounded by an adiabatic jacket [74] (Figure 1.10). Very sensitive circuits detect the temperature dif-

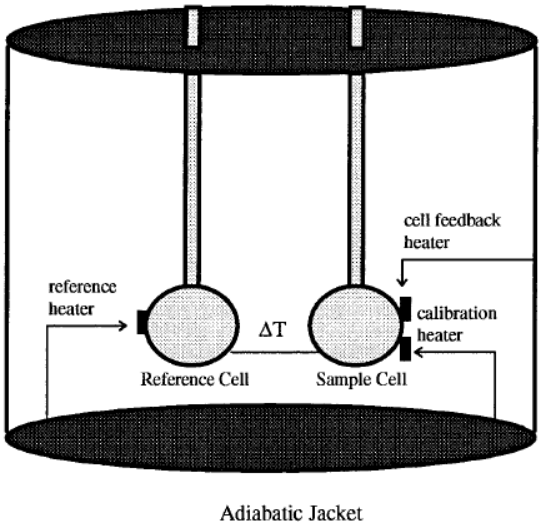


Figure 1.10: Illustration of an ITC instrument. Two cells are contained within an adiabatic jacket. A small continuous power is applied by the heater on the reference cell. Thermopile/thermocouple detectors sense temperature differences between the reference and sample cells. On interaction of ligand and macromolecule, heat is either absorbed or generated. Depending on the nature of the association, the feedback circuit will either increase or decrease power to the sample cell to maintain equal temperature with the reference cell. The heat per unit time supplied to the sample cell is the observable signal in an ITC experiment and a direct measure of the heat evolved on binding of a ligand to a macromolecule. (Adapted from reference [74])

ferences between the reference cell and the sample cell as the ligand is being stepwise injected into the calorimetric cell containing the second reagent (protein, DNA, etc.). The most sensitive instruments are able to measure interaction heat effects of reactant concentrations as low as 1-10 nmol [71]. The heats of interaction during a calorimetric titration is proportional to the fraction of bound ligand. It is therefore extremely important to determine accurately the initial concentrations of both the macromolecule and the ligand. It is also important to chose appropriate concentrations of the interacting components in order to determine K_a, ΔH, and n, which are determined by fitting the ITC isotherm to obtain the change in free ligand concentration with respect to the total ligand concentration. The shape of the binding isotherm is dependent on both K_a and the concentration of the interacting components in the calorimetric cell (the total binding site concentrations). Choosing the right concentrations is thus

important for obtaining an isotherm that provides maximum data points for the fitting process. This can be obtained through the relationship [75]:

$$C = nK_a[M_T] \tag{1.5}$$

where C is a unitless parameter and M_T is the concentration of the macromolecule being titrated. A rule of thumb says that the C value should be between 10 and 100 in to obtain maximum data points [73].

A simple reversible association between a macromolecule M and a ligand L,

$$\mathrm{M} + \mathrm{L} \underset{K_d}{\overset{K_a}{\rightleftharpoons}} \mathrm{ML} \tag{1.6}$$

is characterized by the binding constant K_a:

$$\mathrm{K_a} = \frac{[ML]}{[M][L]} \tag{1.7}$$

The heat released or absorbed during injection can be expressed by

$$q = V_0 \Delta H \Delta[\mathrm{ML}] \tag{1.8}$$

where q is the heat associated with the change in complex concentration Δ[ML] and ΔH is the molar enthalpy of binding. V_0 is the reaction volume of the sample cell (Figure 1.10). As the concentration of unoccupied binding sites begins to decrease, the heat change decrease correspondingly as more ligand is added. The total cumulative heat afther the ith addition, Q, can be expressed as

$$Q = V_0 \Delta H \sum \Delta[ML]_i = V_0 \Delta H [ML]_i \tag{1.9}$$

where $[ML]_i$ is the total concentration of complex after the ith injection. However, evaluation of microcalorimetric data requires the consideration of the observable response in terms of the total ligand added or the total ligand concentration [71]. The binding equations must therefore be expressed as a function of total ligand and macromolecule concentration:

$$[M_T] = [ML] + [M] \tag{1.10}$$

$$[L_T] = [ML] + [L] \tag{1.11}$$

$[M_T]$ and $[L_T]$ are now denoting the total macromolecule and ligand concentrations, whereas $[M]$ and $[L]$ are the free concentrations of macromolecule and ligand, respectively. $[ML]$ is the concentration of the formed complex.

The model used for fitting of the obtained isotherm from an ITC experiment depend on how many binding sites the macromolecule has. In the simplest case each macromolecule consists of only one type of binding sites within a finite number of identical non-interacting binding sites, all exhibiting the same intrinsic affinity for the ligand. These models are referred to as "single set of independent sites model". Within this system, the binding constant K_a is given by

$$K_a = \frac{\Theta}{(1-\Theta)[L]} \tag{1.12}$$

where Θ is the fractional saturation and $[L]$ is the concentration of the unbound ligand. By mass conservation this is related to the total ligand $[L_T]$ and macromolecule concentration $[M_T]$:

$$[L] = [L_T] - n\Theta[M_T] \tag{1.13}$$

Inserting this expression for $[L]$ into Eq.1.12 gives the quadratic equation

$$\Theta^2 - \Theta\left(1 + \frac{1}{nK_a[M_T]} + \frac{[L_T]}{n[M_T]}\right) + \frac{[L_T]}{n[M_T]} = 0 \tag{1.14}$$

whose only meaningful root is

$$\Theta = \frac{1}{2}\left(1 + \frac{1}{nK_a[M_T]} + \frac{[L_T]}{n[M_T]} - \sqrt{\left(1 + \frac{1}{nK_a[M_T]} + \frac{[L_T]}{n[M_T]}\right)^2 - \frac{4[L_T]}{n[M_T]}}\right) \tag{1.15}$$

The integral heat of reaction Q after the ith injection can now be represented as

$$Q = n[M_T]V_0\Delta H\Theta_i = \frac{V_0\Delta H_n[M_T]K_a[L]}{1 + K_a[L]} \tag{1.16}$$

where V_0 is the volume of the cell, ΔH is the enthalpy of binding per mole of ligand. Similarly the differential heat of the ith injection can be represented as

$$q_i = n[M_T]V_0\Delta H\left(\Theta_i - \Theta_{i-1}\right) \tag{1.17}$$

In order to obtain the parameters K_a, ΔH and n from a single experiment one must do a nonlinear fit based on Eq.1.16 to the hyperbolic saturation curve in the integral mode (Q vs. $[L_T]$). The same parameters can also be obtained based on Eq.1.17 by fitting the titration data to the sigmoid saturation curve in the differential heat mode (q_i vs. $[L_t]$, or q_i vs. $[L_T]/[M_T]$). However, an ITC experiment only gives the macroscopic information about the system under investigation and is thus not appropriate for obtaining microscopic information.

1.4.3 Quantum chemistry

The Schrödinger equation

Quantum mechanics tries to describe the molecular properties based on the electron distribution. The state of a system is described by a wavefunction which gives all possible information about the system [76]. In order to gain knowledge about possible future states of a quantum mechanical system from its present state, it is desirable to have an equation that describes how the wave function changes with time and space. The Austrian physicist Erwin Schrödinger postulated the well known Schrödinger equation, which in shorthand operator form is given in Eq.1.18.

$$\hat{H}\Psi = E\Psi \tag{1.18}$$

This is the time-independent equation and is consequently known as the time-independent Schrödinger equation. Ψ denotes the molecular wave function, $\hat{H}$ the Hamiltonian operator, and E the total energy of the state. The Hamiltonian operator consists of the kinetic energy operator $\hat{T_p}$ for each particle and a potential energy operator $\hat{V_{pq}}$ for each distinct pair of particles. The kinetic- and

potential energy operators are given in atomic units by the Laplace operator and the Coulomb potential (ignoring interactions due to particle spin) as:

$$\hat{H} = \sum_p \hat{T}_p + \sum_{p>q} \hat{V}_{pq} \tag{1.19}$$

$$\hat{T}_p = -\frac{1}{2m_p}\nabla_p^2 = -\frac{1}{2m_p}\left(\frac{\partial^2}{\partial x_p^2} + \frac{\partial^2}{\partial y_p^2} + \frac{\partial^2}{\partial z_p^2}\right) \tag{1.20}$$

$$\hat{V}_{pq} = \frac{q_p q_q}{r_{pq}} = \frac{q_p q_q}{\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2}} \tag{1.21}$$

Here m_p are the particles' masses and q_p the charges. The dynamics of a many-electron system is very complex, and the Schrödinger equation is only possible to solve exactly for one- and two-particle systems. A variety of methods for obtaining approximate solutions have therefore been developed, ranging from semiempirical methods with adjusted parameters to highly advanced analytical methods based on different mathematical formalisms (many-body perturbation theory, coupled cluster, configuration interaction etc.).

Born-Oppenheimer approximation

A huge simplification to the problem of many-electron systems can be made by realizing that the nuclei are the heaviest particles in a molecule. The lightest nucleus is the proton, ^{1}H, which is ≈ 1836 times as heavy as an electron, while the most abundant carbon nucleus, ^{12}C, is ≈ 21863 times as heavy. This means that the electrons and the nuclei will act differently according to molecular motion. When the nuclei change their configuration slightly the electrons will immediately adjust. If one assume that the Hamiltonian is separable into two or more terms, then the total eigenfunctions are products of the individual eigenfunctions of the separated Hamiltonian terms, and the total eigenvalues are sums of individual eigenvalues of the separated Hamiltonian terms. For example, if $\psi(q_1, q_2)$ can be written as $\psi(q_1)\psi(q_2)$ and the Hamiltonian separable into two terms $\hat{H} = \hat{H}_1(q_1) + \hat{H}_2(q_2)$. Assuming that the wavefunction can be written in this way and that $\psi(q_1)$ and $\psi(q_2)$ are eigenfunctions of $\hat{H}_1$ and $\hat{H}_2$ then

$$\begin{aligned}
\hat{H}\psi(q_1, q_2) &= (\hat{H}_1 + \hat{H}_2)\psi(q_1)\psi(q_2) \\
&= \hat{H}_1\psi(q_1)\psi(q_2) + \hat{H}_2\psi(q_1)\psi(q_2) \\
&= E_1\psi(q_1)\psi(q_2) + E_2\psi(q_1)\psi(q_2) \\
&= (E_1 + E_2)\psi(q_1)\psi(q_2) \\
&= E\psi(q_1, q_2)
\end{aligned} \tag{1.22}$$

Thus the eigenfunctions of $\hat{H}$ are products of the eigenfunctions of $\hat{H}_1$ and $\hat{H}_2$, and the eigenvalues are the sums of eigenvalues of $\hat{H}_1$ and $\hat{H}_2$. This allows for an approximation where the nuclei and electronic wavefunctions are separated, a result known as the Born-Oppenheimer approximation [77]. First the nuclear kinetic energy operators $\hat{T}_n$ are separated from the electronic Hamiltonian $\hat{H}^{el}$, which then consists of zero-electron, 1-electron and 2-electron parts (n,m

denoting nuclei and e,f denoting electrons):

$$\hat{H}^{tot} = \sum_n \hat{T}_n + \hat{H}^{el} \tag{1.23}$$

$$\hat{H}^{el} = \sum_{n>m} \hat{V}_{nm} = h_{nuc}, \tag{1.24}$$

$$+ \sum_e \left(\hat{T}_e + \sum_n \hat{V}_{en} \right) = \hat{h}, \tag{1.25}$$

$$+ \sum_{e>f} \hat{V}_{ef} = \hat{g} \tag{1.26}$$

The electronic Schrödinger equation is then solved for all electronic states (k) where the nuclear coordinates r_n enter h_{nuc} and $\hat{h}$ as parameters:

$$\hat{H}^{el}\psi_k^{el}(r_e; r_n) = E_k^{el}\psi_k^{el}(r_e; r_n), \qquad k = 0, 1, 2...\infty \tag{1.27}$$

The solutions from the electronic Schrödinger equation, E_k^{el}, are called the potential energy surface (PES). The equilibrium geometry is defined as the configuration of r_n that gives the lowest energy on the groundstate PES (Figure 1.11). In the second step, the complete Schrödinger equation is solved with an

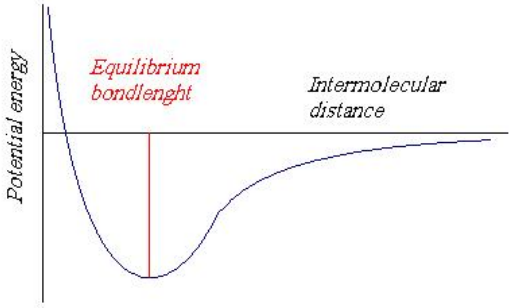


Figure 1.11: Illustration of the potential energy surface from the electronic Schrödinger equation in one dimension.

expansion over the electronic solutions ψ_k^{el} as

$$\psi^{tot}(r_n, r_e) = \sum_k \psi_k^{nuc}(r_n)\psi_k^{el}(r_e; r_n), \qquad \hat{H}^{tot}\psi^{tot} = E^{tot}\psi^{tot} \tag{1.28}$$

where the ψ_k^{nuc} are the coefficients and the ψ_k^{el} the basis of the expansion. This finally leads to an infinite set of coupled Schrödinger equations given as:

$$\left(\sum_n \hat{T}_n + E_k^{el}(r_n) \right) \psi_k^{nuc}(r_n) = E^{tot}\psi_k^{nuc}(r_n) \tag{1.29}$$

In practice, the Born-Oppenheimer approximation implies that the electronic wavefunction can be solved in a stationary nuclear framework. Equation 1.29 is however no less complicated than the original Schrödinger equation, but it can be truncated to a good approximation, both in the number of PESs included, and in the range and precision of each PES.

Wavefunction-based methods

The eigenfunctions of the electronic Schrödinger equation can for a single-electron system properly be called molecular orbitals. If the system only has one nucleus, the equation can thus be solved exactly, and the eigenfunctions would be hydrogen-like atomic orbitals. It is thus possible to use these orbitals as a starting point for constructing more complicated molecular orbitals through an approach called the "linear combination of atomic orbitals" (LCAO) [78]. First one would construct a starting guess wave function as a linear combination of atomic wavefunctions ϕ;

$$\psi = \sum_{i=1}^{N} a_i \phi_i \tag{1.30}$$

where the set of N functions ϕ_i comprise the basis set and each ϕ_i is associated with some coefficient a_i. The variational principle states that an approximate wavefunction always has an energy which is above or equal to the exact energy, E_0 [79]:

$$\frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0 \tag{1.31}$$

The "bra-ket" $\langle ...|...\rangle$ is a short hand notation for integrals over all coordinates where the "bra" $\langle ...|$ is complex conjugated and the denominator ensures that the wavefunction is normalized. The energy is only equal if the wavefunction is exact, providing a powerful tool for solving the wavefunction. As long as the energy keeps dropping, one is on the right track.

In the self-consistent field (SCF) approximation the problem of solving the N-electron Schrödinger equation is tackled by writing the wavefunction as a slater determinant. The slater determinant is an anti-symmetrized product of N orthonormal orbitals $\phi_1, \phi_2 ... \phi_N$ which are 1-electron wavefunctions:

$$\Phi(r_1, r_2 ... r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(r_1) & \phi_2(r_1) & \cdots & \phi_N(r_1) \\ \phi_1(r_2) & \phi_2(r_2) & \cdots & \phi_N(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(r_N) & \phi_2(r_N) & \cdots & \phi_N(r_N) \end{vmatrix} \tag{1.32}$$

The slater determinant ensures that the wavefunction switches sign when two electrons are interchanged, as required by the Pauli principle. The factor $1/\sqrt{N}$ ensures that ψ is normalized. By applying the variational principle to the slater determinant, one obtains the Hartree-Fock model [79]. The energy may then be

written in a symmetrical form as

$$E = \sum_{i=1}^{N_{elec}} h_i + \frac{1}{2}\sum_{i=1}^{N_{elec}}\sum_{j=1}^{N_{elec}} (J_{ij} - K_{ij}) + V_{nn} \tag{1.33}$$

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{a}^{N_{nuclei}} \frac{q_a}{|R_a - r_i|} \tag{1.34}$$

$$J_{ij} = \left\langle \phi_i(i)\phi_j(j) \left| \frac{1}{|r_i - r_j|} \right| \phi_i(i)\phi_j(j) \right\rangle \tag{1.35}$$

$$K_{ij} = \left\langle \phi_i(i)\phi_j(j) \left| \frac{1}{|r_i - r_j|} \right| \phi_j(i)\phi_i(j) \right\rangle \tag{1.36}$$

$$V_{nn} = \sum_{a}^{N_{nuclei}} \sum_{b>a}^{N_{nuclei}} \frac{q_a q_b}{|R_a - R_b|} \tag{1.37}$$

where the factor of 1/2 allows the double sum to run over all electrons. h_i describes the motion of electron i in the field of all the nuclei. J_{ij} is the Coulomb integral, K_{ij} is called the exchange integral, and V_{nn} is the nuclear potential energy. The SCF class of models have in common that they attempt to solve an N-electron Schrödinger equation by solving coupled 1-electron Schrödinger equations. The term "self-consistent" is derived from the coupling between the 1-electron Hamiltonian, called the Fock operator and the solutions (orbitals). The Fock operator in these terms become:

$$F_i = h_i + \sum_{j}^{N_{elec}} (J_j - K_j) \tag{1.38}$$

where the J_j and K_j operators are defined as

$$J_i|\phi_j(2)\rangle = \langle\phi_i(1)|g_{12}|\phi_i(1)|\phi_j(2)\rangle \tag{1.39}$$

$$K_i|\phi_j(2)\rangle = \langle\phi_i(1)|g_{12}|\phi_j(1)|\phi_i(2)\rangle \tag{1.40}$$

and g_{ij} is a two-electron operator giving the electron-electron repulsion

$$g_{ij} = \frac{1}{|r_i - r_j|} \tag{1.41}$$

The Fock operator is thus an effective one-electron operator which describes the kinetic energy of an electron and the attraction to all the nuclei h_i, as well as the repulsion to all the other electrons. The variation of the energy is then written in terms of the Fock operator F_i as

$$\delta E = \sum_{i}^{N_{elec}} (\langle\delta\phi_i|F_i|\phi_i\rangle + \langle\phi_i|F_i|\delta\phi_j\rangle) \tag{1.42}$$

where the Fock operator is associated with the variation of the total energy and not the energy itself. In terms of the Lagrange function, the variation becomes

$$\delta L = \sum_{i}^{N_{elec}} (\langle\delta\phi_i|F_i|\phi_i\rangle + \langle\phi_i|F_i|\delta\phi_i\rangle) - \sum_{ij}^{N_{elec}} \lambda_{ij} (\langle\delta\phi_i|\phi_j\rangle + \langle\phi_i|\phi_j\rangle) \tag{1.43}$$

The desired orbitals are those that make $\delta L = 0$. The Lagrange multipliers [80] are however elements of a Hermitian matrix ($\lambda_{ij} = \lambda_{ij}^*$). Therefore the final set of Hartree-Fock equations may be written as

$$F_i\phi_i = \sum_j^{N_{elec}} \lambda_{ij}\phi_j \tag{1.44}$$

The equation can again be simplified by choosing a unitary transformation (the total wave function is unchanged by a unitary transformation of the occupied MOs among themselves, because changing rows and columns in a determinant is not affecting the determinant itself) that makes the matrix of Lagrange multipliers diagonal (i.e $\lambda_{ij} = 0$ and $\lambda_{ii} = \epsilon_i$). This special set of molecular orbitals (ϕ') is called *canonical* MOs, and transforms eq.1.44 into a set of pseudo-eigenvalue equations

$$F_i\phi_i' = \epsilon_i\phi_i' \tag{1.45}$$

The Lagrange multipliers are the expectation value of the Fock operator in the MO basis

$$\langle\phi_i'|F_i|\phi_i'\rangle = \epsilon\langle\phi_i'|\phi_i'\rangle = \epsilon_i \tag{1.46}$$

The Hartree-Fock equations form a set of pseudo-eigenvalue equations as the Fock operator depends on all the occupied MOs. A specific Fock orbital can only be determined if all the other occupied orbitals are known. Iterative methods must therefore be employed for solving the problem. A set of functions that is a solution to $F_i\phi_i' = \epsilon_i\phi_i'$ are called self-consistent field orbitals. The total energy can then be written in terms of MO energies as

$$E = \sum_i^{N_{elec}} \epsilon_i - \frac{1}{2}\sum_{ij}^{N_{elec}} (J_{ij} - K_{ij}) + V_{nn} \tag{1.47}$$

where

$$\epsilon_i = \langle\phi_i|F_i|\phi_i\rangle = h_i + \sum_j^{N_{elec}} J_{ij} + K_{ij} \tag{1.48}$$

Since the Fock operator contains terms describing the repulsion to all other electrons (J and K), and the sum over MO energies it therefore counts the electron-electron repulsion twice, which must be corrected for. Moreover, the energy can not be exact as it only accounts for the repulsion in an average fashion. This is due to the approximation of a single Slater determinant of N spin orbitals as a trial wave function and the HF method is therefore often referred to as the mean-field approximation.

The SCF method (HF) is the simplest wavefunction-based method, but since each particle is assumed to experience a mean field created by the other particles the total energy can not be 100% correct. If one use a complete set of orbitals, the solution of the Schrödinger equation would yield the exact single determinant representation, and representing the wavefunction as an expansion in a complete basis of determinants would yield the exact solution of the wavefunction within the Born-Oppenheimer approximation. This is a powerful concept, because increasing the number of basis functions generally would improve the accuracy of the models, and this generally holds well for molecular modeling. In a sufficient large basis set the HF wave function Φ_{HF} is thus able to account for ∼99% of

the total energy. The remaining 1% on the other hand can often be of great importance for describing chemical phenomena. The difference between the HF energy and the lowest possible energy in a given basis is called the electron correlation (EC) energy, which correspond to the motion of the electrons being correlated. In order to improve on the HF results the trial wave function must contain more than one Slater determinant. Typically this is done by using the HF wave function as a starting point for improvements (because the HF method determines the energetically best one-determinant trial wave function within a given basis set), creating a generic multi-determinant trial wave function as

$$\Psi = a_0\Phi_{HF} + \sum_{i=1} a_i\Phi_i \tag{1.49}$$

where a_0 is usually close to one. The main methods for calculating the electron correlation by using multi-determinant trial wave functions (Eq.1.49) are many-body perturbation theory (MBPT), configurational interaction (CI) and coupled cluster (CC).

Kohn-Sham Density Functional Theory

The original idea behind density functional theory was to have an orbital-free expression for the energy of the electron in terms of the electronic probability density ρ. The main flaw with this concept was the poor representation of the kinetic energy and thus not a significant improvement compared with wave mechanics. The success of modern DFT was the introduction of orbitals by Kohn and Sham. They made DFT available for computational chemistry by introducing the concept of a non-interacting reference system built on one-electron functions [81]. Even though the electrons are interacting, the difference between the exact kinetic energy and that calculated assuming non-interacting orbitals is small, and like the HF theory it provides ∼99% of the correct energy. The remaining kinetic energy is absorbed into an exchange-correlation term $E_{XC}[\rho]$. The Kohn-Sham model is thus closely related to the HF method, sharing identical formulas for the kinetic electron-nuclear and Coulomb electron-electron energies. A general DFT energy expression can be written as

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{1.50}$$

where

$$T_S[\rho] = \sum_{i=1}^{N_{elec}} \left\langle \phi_i \left| -\frac{1}{2}\nabla^2 \right| \phi_i \right\rangle \tag{1.51}$$

$$E_{ne}[\rho] = -\sum_{a}^{N_{nuclei}} \int \frac{q_a(R_a)\rho(r)}{|R_a - r|} dr \tag{1.52}$$

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' \tag{1.53}$$

The subscript S denotes that the exact kinetic energy of the non-interacting system is calculated from a slater determinant and $\rho(r)$ denotes the total electron density at a particular point r in space. $E_{ne}[\rho]$ is the attraction between electrons and nuclei and $J[\rho]$ the Coulomb energy. The kinetic energy of the

non-interacting reference system T_S, the attraction between electrons and nuclei and the Coulombic repulsion between electrons are thus calculated exactly in the Kohn-Sham approach. The problem is that the Hohenberg-Kohn theorems do not state the relationship between the functional and the density [82]. The challenge in DFT for a given density is to design a functional that models E_{XC} well. The definition of E_{XC} can be obtained by equating E_{DFT} to

$$E_{XC}[\rho] = \underbrace{(T[\rho] - T_S[\rho])}_{\text{A}} + \underbrace{(E_{ee}[\rho] - J[\rho])}_{\text{B}} \tag{1.54}$$

where A may be considered as the kinetic correlation energy, while B contains both potential correlation and exchange energy. It is also a usual approach to handle the exchange part and the correlation part separately ($E_{XC} = E_X + E_C$).

The model functionals in use may be classified into five different subgroups; the local (spin) density approximation (L(S)DA), the generalized gradient approximation (GGA), meta-GGA, hybrid (or hyper-GGA) functionals, and generalized random phase (GRP) methods. These methods differ in the way they approximate the E_{XC} part in Eq.1.50. In the LDA methods, the approximated value of the exchange-correlation term at position r is computed exclusively from the local density $\rho(r)$, and typically the density is treated as a uniform electron gas. In the GGA methods, the gradient of the electron density is also taken into account, to adjust for the general non-uniform electron density case. The hybrid functionals are named so because a part of the exchange contribution is taken from a Hartree-Fock method calculation and parameterized into the functional. Hybrid functionals have seen great success, and one of the most popular methods is the B3LYP [83]:

$$E_{XC}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{exact} + b\Delta E_x^{B88} + (1-c)E_x^{LSDA} + cE_c^{LYP} \tag{1.55}$$

The a, b and c parameters are determined by fitting to experimental data. Table 1.3 shows an overview of commonly used functionals given by their acronym, and placed in the Jacob's ladder classification[76].

Table 1.3: Perdew classification of exchange-correlation functionals (adapted from ref.[76]).

Level	Name	Variables	Examples
1	Local Density	ρ	LDS, LSDA, X_α
2	GGA	ρ, $\nabla\rho$	BLYP, OPTX, OLYP, PW86
3	Meta-GGA	ρ, $\nabla\rho$, $\nabla^2\rho$ or τ	BR, B95, PKZB, TPSS
4	Hyper-GGA	ρ, $\nabla\rho$, $\nabla^2\rho$ or τ HF exchange	H+H, B3LYP, O3LYP
5	Generalized RPA	ρ, $\nabla\rho$, $\nabla^2\rho$ or τ HF exchange Virtual orbitals	OEP2

One of the greatest advantages of DFT compared to *ab initio* (wavefunction methods) calculations is the low computational cost, especially for larger systems. The main disadvantage with DFT is that it is not possible to carry out systematic improvement by taking more electron configurations into account as

in wavefunction methods. The only way to improve the results is to use better functionals. However, even though the computational cost is lower compared with wavefunction methods, it is still not possible, in terms of computational cost, to model large systems like proteins, molecules in explicit water or membrane systems with DFT. In order to model larger systems it is desirable to use the less complicated molecular mechanics (MM) methods that ignore the motion of the electrons.

1.4.4 Molecular Mechanics and Force Fields

Without the assumption of the Born-Oppenheimer approximation it would not be possible to write the energy as a function of the nuclear coordinates, so this is one of the strongest and most important assumptions that makes molecular mechanics work. Molecular mechanics is further on based on more or less simple models of the interactions within a system. Typically the contributions come from processes such as bond stretching, rotation about single bonds, the opening and closing of angles, torsional terms and non-bonded interactions. In principle all force fields contain at least the five terms for describing bonds, angles, dihedrals, improper dihedrals and atom pairs as in;

$$\begin{aligned} V_{pot}(r^N) &= \sum_{bonds} \frac{1}{2}(r - r_0)^2 + \sum_{angles} (\theta - \theta_0)^2 \\ &+ \sum_{dihedrals} K_\psi(1 + \cos(n \cdot \psi - \delta)) + \sum_{improp.dihed.} \frac{1}{2} k_\zeta (\zeta - \zeta_0)^2 \\ &+ \sum_{atompairs} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \end{aligned} \tag{1.56}$$

where $V_{pot}(r^N)$ denotes the potential energy, which is a function of the positions (r) of N particles (usually atoms). Transferability is also an important concept of force fields as it enables a set of parameters developed and tested on a relatively small number of cases to be applied to a much wider range of problems. Transfarebility means that the same set of parameters can be used to model a series of related molecules, rather than having to define a new set of parameters for each individual molecule. Another point is that parameters developed from data on small molecules can be used to study much larger molecules.

When defining a force field one must not only specify the functional form but also the parameters to be used. In other words, two force fields may use the same functional form but different parameters, and force fields with different functional forms may give results of comparable accuracy [84]. A specific force field is further on designed to predict certain properties and will therefore be parameterized accordingly. Trying to predict other properties than the specific force field is designed for is in general not a good idea. Another important point to keep in mind is that force fields are empirical. There are in other words no "correct" form for a force field. One could claim that the principle of "survival of the fittest" applies to force fields as well. If one functional form is shown to perform better than another it is likely that form will be favored.

Bond Stretching

In molecular mechanics calculations it is rare that the bonds involved deviate significantly from their equilibrium values. It is therefore common to use the simple approach of the Hook's formula in which the energy varies with the square of the displacement from the reference bond length r_0:

$$v(r) = \frac{k}{2}(r - r_0)^2 \tag{1.57}$$

A true bond-stretching potential is not harmonic, but has a shape typically given by the Morse potential which resembles the dissociation limit in a better fashion:

$$\begin{aligned} v(r) &= D_e \left(1 - \exp[-\alpha(r - r_0)]\right)^2 \\ \alpha &= \omega\sqrt{\mu/2D_e} \end{aligned} \tag{1.58}$$

μ is the reduced mass and D_e is the depth of the potential energy minimum. Even though the Morse potential resembles the true energy curve in a better fashion than Hook's law, it is rarely used in molecular mechanics. This is mainly because it requires three parameters to be specified for each bond and the formula itself is not particularly amenable to efficient computation. Hook's law can thus be used to model the Morse potential more accurate by doing a Taylor like expansion and therebye including higher order terms:

$$v(r) = \frac{k}{2}(r - r_0)^2 \left[1 - k'(r - r_0) - k''(r - r_0)^2 - k'''(r - r_0)^3 - \cdots\right] \tag{1.59}$$

The MM2 [85] force field for example includes just the quadratic and cubic terms. Far from the reference value the cubic function passes through a maximum, which in turn can lead to catastrophic lengthening of bonds. Inclusion of the quartic term, as in MM3 [86], eliminates this inversion problem and leads to a better description of the Morse curve.

Angle Bending

The same arguments used for bond stretching is also applied for angle bending which also is described by Hook's law:

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \tag{1.60}$$

As with bond-stretching terms, the accuracy of the force field can be improved by the inclusion of higher order terms:

$$v(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \left[1 - k'(\theta - \theta_0) - k''(\theta - \theta_0)^2 - k'''(\theta - \theta_0)^3 - \cdots\right] \tag{1.61}$$

In general much less energy is required to distort an angle away from the equilibrium than to stretch a bond. Consequently, the force constants are proportionately smaller compared with the constants from bond-stretching.

Torsional Terms

The conformation of molecules are in general largely determined by torsional terms and non-bonded interactions. This is because angle bending and bond-stretching requires substantial energies to change from their initial values. Generally little energy is required to change the torsional terms, and this "softness" in turn leads to large variations. Moreover, the existence of barriers to rotation about chemical bonds is fundamental to understanding the structural properties of molecules and conformational analysis. Most force fields for organic molecules use explicit torsional potentials with contributions from each bonded quartet of atoms A-B-C-D. Ethane for example would thus involve nine individual torsional terms. The torsional terms are generally modelled using a periodic function on the form:

$$v(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}[1 + \cos(n\omega - \gamma)] \tag{1.62}$$

V_n is often referred to as the barrier height, ω the torsional angle, γ, the phase factor, which determines where the torsion angle passes through its minimum value, and n is the multiplicity determining the number of minimum points in the function as the bond is rotated 360^o. Ethane for example with a double bond between the two sp^2 carbon atoms would thus have $n = 2$ and $\gamma = 180^o$, giving minima at 0^o and 180^o. The value of V_n would also be larger for ethane compared with ethane for example.

Improper torsions

Many chemical systems are planar, and this planarity is taken care of by using either improper torsions or out-of-plane bending terms. In other words, in order to achieve the wanted geometry for planar systems it is necessary to incorporate additional terms in the force field that keeps the sp^2 carbon in the same plane as the three atoms bonded to it. Per definition improper atoms are between four atoms that are not bonded in the sequence 1-2-3-4. An example of a torsional potential used to maintain the improper torsion angle at 0^o or 180^o is given as:

$$v(\omega) = k(1 - \cos(2\omega) \tag{1.63}$$

Improper torsions can also be used to keep things at a certain degree out-of-plane, which can be modelled by using a harmonic potential on the form

$$v(\theta) = \frac{k}{2}\theta^2 \qquad \text{or} \qquad v(h) = \frac{k}{2}h^2 \tag{1.64}$$

θ is here the angle between the out-of-plane atom and the plane, and h in the second equation is the distance from the plain to the out-of-plane atom. It is however important to remember that out-of-plane terms may not always be necessary, and that to include such terms may have a deleterious effect on the performance of the force field [84].

Cross-terms

In theory cross terms between all possible contributions to a force field should be included due to coupling between the internal coordinates. For example, as a

bond angle is decreased it is found that the adjacent bonds stretch to reduce the interaction between the atoms that comes closer due to the decrease of the angle. It has however been found that only a few cross terms are generally necessary in order to reproduce structural properties accurately. A few important terms are bond-bond, bond-angle, angle-angle, bond-torsion and angle-torsion. The stretch-stretch cross term between bonds 1 and 2 can for example be modelled as:

$$v(r_1, r_2) = \frac{k_{r_1,r_2}}{2} \left[(r_1 - r_{1,0})(r_2 - r_{2,0})\right] \tag{1.65}$$

Even though not all cross terms are included in a force field one must remember that all the terms in reality are connected.

Non-bonded interactions

Non-bonded interactions are usually modeled using van der Waals and electrostatic interactions. The charge distribution can be represented in many ways, but one common approach being an arrangement of fractional point charges throughout the molecule. These charges are thus designed to reproduce the electrostatic properties of the molecules. Typically the electrostatics between two molecules or between different parts of the same molecule is calculated as a sum of interactions between pairs of point charges, using Coulomb's law:

$$v(r)_{elec.} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.66}$$

This expression correspond to monopoles which are important terms to include. Dipoles, octapoles, etc., goes extremely fast towards zero and can normally be neglected. The charge-charge energy thus decays as $\sim r^{-1}$ while the dipole-dipole energy decays as $\sim r^{-3}$ which is much faster. The electrostatic interactions are very important in terms of polarizability because charges move in response to other charges. The polarization effect is of crucial importance when simulating a molecule in water or when predicting ligand binding.

The van der Waals effects are in general small, but for larger systems, like proteins, they become non-negligible when summing them up. The most important interactions within van der Waals arises from induced dipole-dipole interactions between molecules that do not have a permanent dipole. Typically the attraction is calculated as being $\propto r^{-6}$ while the repulsion due to electron cloud overlap is calculated as being $\propto r^{-12}$. The best known of the van der Waals potential functions is the *Lennard-Jones 12-6* function:

$$v(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{1.67}$$

ϵ is the well depth and σ is the collision diameter, both being adjustable parameters. For polyatomic systems it is usual to calculate the van der Waals interaction energy between two molecules using a site model in which the interaction is determined as a sum of the interactions between all pairs of sites on the two molecules. The determination of van der Waals parameters can further on be difficult and time consuming process for larger systems. It is therefore

common to assume that the parameters for the cross interactions can be obtained from the parameters of the pure atoms using so called mixing rules. The Lorentz-Berthelot mixing is given as

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \tag{1.68}$$

$$\epsilon_{AB} = \sqrt{\epsilon_{AA}\epsilon_{BB}} \tag{1.69}$$

This formulation is most successfull when applied to similar species. The OPLS force field developed by Jorgensen and Tirado-Reeves in 1988[87] on the other hand calculates the collision diameter for mixed interactions as the geometric mean of the values for the two component atoms, which has seen great success.

1.4.5 Molecular Dynamics

Molecular dynamics (MD) is a computer simulation method in which atoms and molecules are allowed to interact for a period of time by the approximations of known physics. MD simulation is a commonly used method for studying the time dependent interactions of proteins and biomolecules. The analysis of a potential energy surface by locating the minima and saddle points corresponds to modelling the system at 0 Kelvin. Inclusion of a finite temperature can be done by means of statistical mechanics methods. In a molecular dynamics simulation, the macroscopic quantities derived from the partition function must be estimated from a representative sampling of the phase space. The word *simulation* in turn refers to methods aimed at generating a representative sampling of a system at a finite temperature. There are mainly two techniques for generating an ensemble, Monte Carlo (MC) and molecular dynamics (MD).

The successive configurations of the system in molecular dynamics are generated by integrating Newton's laws of motion. The resulting trajectory describes how the positions and velocities of the particles in the system vary with time. The trajectory in turn can be obtained by solving the differential equations embodied in Newton's second law ($F = ma$):

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{1.70}$$

m_i is the particle mass, x_i is the direction in which the particle is moving, and F_{x_i} is the force acting on the particle in the direction of x_i.

In the early days of MD simulations the hard-sphere model was used for the simulations. In such cases each particle is a sphere that moves at constant velocity in straight lines between collisions, where all collisions are perfectly elastic and occurs when the separation between the centres of the spheres equals the sphere diameter. After collision, the new velocities are calculated by applying the principle of conservation of linear momentum. The steps involved in the hard-sphere calculations are:

1. Identify the next pair of spheres to collide and calculate when the collision will occur.

2. Calculate the positions of all the spheres at the collision time.

3. Determine the new velocities of the two colliding spheres after collision.

4. Repeat from 1 until finished.

The simple model of hard-spheres suffer from many deficiencies, but has proven usefull to gain usefull insight into the microscopic nature of fluids.

Molecular Dynamics with continuous potentials

In a realistic model of intermolecular interactions, the force acting on each particle will change whenever the particle changes its position. However, under the influence of a continuous potential the motions of all particles are coupled together, which in turn gives rise to a many-body problem that can not be solved analytically. A solution to such problems is to integrate the equations of motion using a finite difference method.

The finite difference technique is used to generate molecular dynamics trajectories with continuous potential models, which is assumed to be pairwise additive. The idea is that the integration is broken down into many small steps, each separated in time by a fixed time δt. The total force on each particle in the configuration at time t is calculated as the vector sum of its interactions with the other particles. From this force, the accelerations of the particles can be determined, which in turn are combined with the positions and velocities at a time t to calculate the new positions and velocities at time $t+\delta t$. An important point here is that the force is assumed to be constant during the time step. Therefore small time steps are important. Many algorithms for integrating the equations of motion using finite difference methods have been developed, but all assume that the positions and dynamic properties can be approximated as Taylor series expansions:

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \cdots \quad (1.71)$$

$$v(t+\delta t) = v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \cdots \quad (1.72)$$

$$a(t+\delta t) = a(t) + \delta t b(t) + \cdots \quad (1.73)$$

$$\cdots = \cdots$$

v is here the velocity, a the acceleration, b the third derivative or the position with respect to time, and so on. There are many algorithms which uses different variants of this scheme. Some of the most common are the Verlet algorithm [88] and variants of the verlet algorithm.

The most demanding part of a molecular dynamics simulation is invariably the calculation of the force on each particle in the system. The integration algorithm used should obey some requirement where some important considerations are that the integration algorithm should conserve energy and momentum, be time-reversible, and it should permit a long time step to be used. A long time step will in computational terms require fewer iterations to cover a given amount of phase space.

Another important method is the predictor-corrector methods [89]. These methods have three basic steps:

1. New postions, velocities, accelerations and higher-order terms are predicted according to the Taylor expansions to order c in Eq.1.71.

2. The forces are in the second stage evaluated at the new positions to give accelerations $a(t + \delta t)$.

3. The accelerations obtained from stage 2 are then compared with the predicted ones from stage 1. The difference between these to is then used to "correct" the positions, velocities, etc., in the correction step:

$$\Delta a(t + \delta t) = a^c(t + \delta t) - a(t + \delta t) \tag{1.74}$$

From this it follows that:

$$\begin{aligned} r^c(t + \delta t) &= r(t + \delta t) + c_0 \Delta a(t + \delta t) && (1.75)\\ v^c(t + \delta t) &= v(t + \delta t) + c_1 \Delta a(t + \delta t) && (1.76)\\ a^c(t + \delta t) &= \frac{1}{2} a(t + \delta t) + c_2 \Delta a(t + \delta t) && (1.77)\\ \cdots &= \cdots \end{aligned}$$

The values of the coefficients c_i can vary from one algorithm to the other, but Gear suggested that the coefficients should depend on the order of the Taylor expansion [89].

Choosing time steps

Choosing the right time step for a molecular dynamics simulation is of crucial importance for obtaining well-behaved trajectories. It has been shown that for short time steps the predictor-corrector methods may be more accurate, while for longer time steps the Verlet algorithm may be better [90]. If the time step chosen is too small, the trajectory will cover only a limited proportion of the phase space. If the time step is too large, instabilities may arise in the integration algorithm due to high energy overlaps, and the calculation may "blow up". The aim is thus to find the correct balance between simulating the "correct" trajectory and covering the phase space. A typical rule of thumb to follow for MD simulations is that the time step should be approximately one order of magnitude smaller than the shortest motion of vibration. A C-H bond for example vibrates with a repeat period of approximately 10 fs. A suitable time step for system with C-H vibrations would then be 1 fs.

The restriction of the time step being dictated by the highest frequency motion present in the system is clearly a severe restriction, particularly since many of the high-frequency motions usually are of little interest as they have minimal effects on the overall system. One solution to this problem is to "freeze out" such vibrations by constraining the appropriate bonds to their equilibrium while still permitting the rest of the degrees of freedom to vary under the intramolecular and intermolecular forces present. This in turn enables longer time steps to be used and consequently larger parts of phase space to be covered.

The most common method for applying constraints in molecular dynamics is the so called SHAKE procedure proposed by Ryckaert, Ciccotti and Berendsen in 1977 [91]. The SHAKE constraint removes the bond stretching degrees of freedom in the force field. This is an iterative method where the constraints are treated one by one, and the coordinates are adjusted to satisfy each constraint in turn until all are within a specific tolerance. After the correction has been applied to all bonds in turn, every bond is checked. If there are any large

deviation, the process starts over. The process will be repeated until all are converged. The forces for the constrained bonds do not need to be computed when using this method. The equilibrium bond lengths are however needed since SHAKE uses them to reset the atomic postions. A rule of thumb says that if one can constrain all the C-H bonds, the time step can be doubled. A last note of warning, if SHAKE fails there is usually something wrong with the system and not related to problems with SHAKE.

Molecular dynamics can not be used to generate initial coordinates for the atomic positions, so this needs to be obtained somehow before starting a simulation. This can be done in several ways depending on what system one is working with. When studying ligand-protein interactions, the initial coordinates for for the protein atoms can be obtained obtained from experimental crystal structures. Starting coordinates for a protein-ligand complex can then be obtained from for example molecular docking experiments.

1.4.6 Molecular Docking

Molecular docking is an example of the global optimization problem where the goal is to determine the best alignment of two molecules with respect to each other. Molecular docking refers to the case where one tries to fit a small molecule into a large target structure. When docking a ligand into the active site of a rigid enzyme, one is dealing with a problem consisting of six degrees of freedom, three translational and three rotational, besides those arising from the ligand confirmations. The idea is that the three translational degrees of freedom can be sampled on a grid by for example placing the ligand center of mass within a central box with grid points every 1 Å. A small box of for example $10 \times 10 \times 10$ Å thus generates ~ 1000 possible points. At each point the overall rotational orientation of the ligand must be sampled, thereby typically generating a few hundred possibilities. A specific set of intermolecular translational and rotational variables is called a pose, and it is clear that there will be extremely many of them. Molecular docking is consequently a demanding task even though the majority of the poses can be rejected based on for example atom pair distances between the ligand and the receptor. Global optimization schemes such as genetic algorithms are therefore often employed for solving the optimization problem [84].

Using standard force fields which attempts to calculate the enthalpic interactions and an estimate of the free energy by simulation methods in molecular docking is far too expensive in terms of computational cost. The idea of force fields developed for molecular docking is that the non-bonded part of a force field function can be augmented with empirical terms and therebye hopefully capturing some of the entropy and solvent effects. The resulting scoring function can further on be parameterized against experimental binding data. A scoring function typically contains terms like [92]:

$$\Delta G_{score} = a_1 \Delta E_{vdW} + a_2 \Delta E_{el} + a_3 \Delta G_{rot} + a_4 \Delta G_{H-bond} + a_5 \Delta G_{solv} \cdots \quad (1.78)$$

where E_{vdW} are the van der Waals contribution to the interaction energies, E_{el} the Coulombic contribution, G_{rot} the energy of rotation, G_{H-bond} the energy based on hydrogen bond donors and acceptors and G_{solv} the energy required to displace water molecules when a ligand binds to a receptor. The a_i are

weighting factors that can be fitted to actual binding data for specific protein-ligand systems.

To date there more than 60 docking programs and more than 30 scoring functions have been disclosed [93]. Some of the most widely used programs include AutoDock [94], DOCK [95], GOLD [96], Glide [97] and FlexX [98]. The different programmes generally differ in the method used to explore the conformational space of the ligand and/or the protein target, and the scoring function used to evaluate the proposed binding mode (the pose). In theory the scoring function should guide the conformational sampling algorithm and thus assign the best score to the "correct pose" (the native pose observed in crystal structures). Whenever choosing a docking program, two inversely correlated factors are to be considered. First of all speed is essential for effective virtual high-throughput screening of large libraries, and secondly the accuracy is critical for lead optimization.

Some of the search algorithms used for conformational sampling include the genetic algorithm (GA), differential algorithm (DE), rigid-body docking (RBD), simulated annealing (SA), monte carlo (MC) and molecular dynamics (MD) [84]. MC and GA algorithms are examples of stochastic methods that explore the ligand conformational space on the fly. With MC, which is used in for example Glide, the pose of the ligand is sequentially modified through bond rotation, translation and/or rigid body rotation, one or more parameters at a time, and the new conformation is then evaluated. The new conformation is kept if it is lower in energy than the previous, else it is rejected or evaluated using a selection criterion such as the Metropolis criterion. This allows for higher energy conformation to exist by allowing a temperature dependence, resulting in the crossing of energy barriers on the potential energy surface. Generally the higher the temperature, the more likely it is that a higher energy conformation is kept [93]. The scoring functions on the other hand attempt to predict the binding free energy or to rank-order compound by their bioactivity. These functions are classified as empirical, force field (FF) based, and knowledge based [99–103]. Some of the most successfull scoring functions belong to the empirical class where the energetics of the ligand binding is decomposed into simpler, scalable contributions arising from, for example, metal ligation, hydrogen bonds, freezing of rotatable bonds and hydrophobic effects (Eq.1.79).

$$\begin{aligned} \Delta G_{bind} &= \Delta G^0 + \Delta G_{HB} \sum_{HB} f(\Delta r) f(\Delta 1\alpha) + \Delta G_{met} \sum_{met} f(\Delta r) \\ &\quad + \Delta G_{lipo} \sum_{lipo} f(\Delta r) + \Delta G_{rot} N'_{rot} \end{aligned} \tag{1.79}$$

The various scaling factors (ΔG_i) in Eq.1.79 are defined by regression to fit experimentally determined protein-ligand affinities. Some well known examples of such scoring functions include ChemScore[104] (implemented in GOLD among others), GlideScore[105, 106] (implemented in Glide) and AutoDock SF[107] (implemented in AutoDock).

Due to the high computational costs of docking, the receptor is normally treated as a rigid body. This means that the bond angles, torsion angles and bond lengths of the protein residues are not modified at any stage of the complex generation. In some cases conformational changes can be very important during the complex formation, and rigid-body docking is thus inadequate for dealing

with such problems. Alternatively one can use a method that scores possible conformational changes, called flexible docking. However, scoring all possible conformational changes for the whole receptor is still too expensive with respect to computer time, and one must thus select small subsets of possible conformational changes to be considered. It has previously been pointed out based on statistical analysis of the protein databank that ∼85% of the structures contain one to three flexible residues [108]. Small adjustment of the protein can thus have a significant effect on the molecular recognition process, and this is one of the major challenges within the development of new docking algorithms. Many approaches for dealing with protein flexibility have been proposed and recently reviewed by Cavasotto *et al.* [109]. However, very few of these algorithms have been made available to the greater medicinal chemistry community.

1.5 Aims of study

The main goal of this project is to identify tripeptides with antimicrobial activity that have high stability towards chymotryptic degradation and that potentially can serve as a lead for an oral drug. The mode of action for these tripeptides is further on not known, and it is therefore also a goal to identify a possible mechanism for the peptide membrane interaction. This knowledge is of great interest as it hopefully will enable us to design better and more specific antimicrobial peptides. As a step in the direction towards these goals, computational and experimental methods are going to be combined on a set of tripeptides with different unnaturally occurring side chains. The work is divided into the following parts:

1. Investigate the three dimensional structure of the tripeptides in solvent with computational methods for closer investigation of properties and behavior.

2. Perform molecular dynamics simulations of tripeptides in a membrane environment in order to propose a possible mode of interaction with a bacterial cell membrane.

3. Investigate the potential tripeptide binding mode and possibility for stability towards chymotryptic degradation with molecular modeling.

4. Perform isothermal titration calorimetry (ITC) studies to determine thermodynamic properties and verifying the computational modeling results.

5. Obtain a crystal structure of the peptide-enzyme complex for experimental determination of the binding mode and to verify the modeling results.

The tripeptides in this project are based on RWR-NHBn (Figure 1.12) which contain three cationic charges and two hydrophobic elements and displays a medium antibacterial activity [110]. The goal is to investigate the effect towards chymotryptic degradation of varying the different amino acids X and Y, and type of modification to the C-terminal (Z), as illustrated in Figure 1.12.

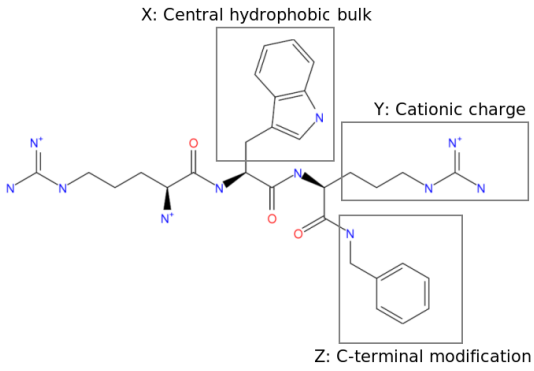


Figure 1.12: General scaffold of the peptides included in the study of RWR-NHBn and the variables X, Y and Z.

"Anyone who has never made a mistake has never tried anything new".

-Albert Einstein

2. Material & methods

Standard chemicals used in solutions were purchased from Sigma, AppliChem and Merck. α-chymotrypsin was purchased from Wartington and Sigma. All tripeptides used in this project were synthesized as described by Svenson *et al.* [111]. Molecular figures used for illustrations were generated using PyMol [112].

2.1 Peptide models

Molecular models of all tripeptides (Table 2.1 and Figure 2.1) were built with Maestro version 9.0 [113]. The geometries were first relaxed with the universal force field (UFF) minimiser in Maestro to eliminate distortions. Peptide geometries were further minimized with MacroModel version 9.6 [114] to prepare the peptides for the docking experiment. The OPLS2005 [115] force field was selected for all optimizations with the Polak-Ribiere-type conjugate gradient (PRCG) minimization scheme[116] and continuum solvent (water). Long-range

Table 2.1: Prepared tripeptide library and their corresponding mass and charge at physiological pH.

Peptide	Mass	Charge
RWR-NH_2	518.6	+3
RWR-NHBn	608.7	+3
RWR-NHEtPh	622.8	+3
RBipR-NH_2	655.7	+3
RBipR-NHBn	645.8	+3
RFR-NHBn	569.7	+3
RDipR-NH_2	555.7	+3
RDipR-NHBn	645.8	+3
RTbtR-NHBn	777.1	+3
RWHar-NHBn	622.8	+3
RWK-NHBn	580.8	+3
RWOrn-NHBn	566.7	+3
RWAgp-NHBn	580.7	+3
RWGpp-NHBn	656.8	+3
RWApp-NHBn	613.8	+2
RRW-NHBn	608.8	+3
RGppW-NHBn	656.8	+3
RAppW-NHBn	613.8	+2

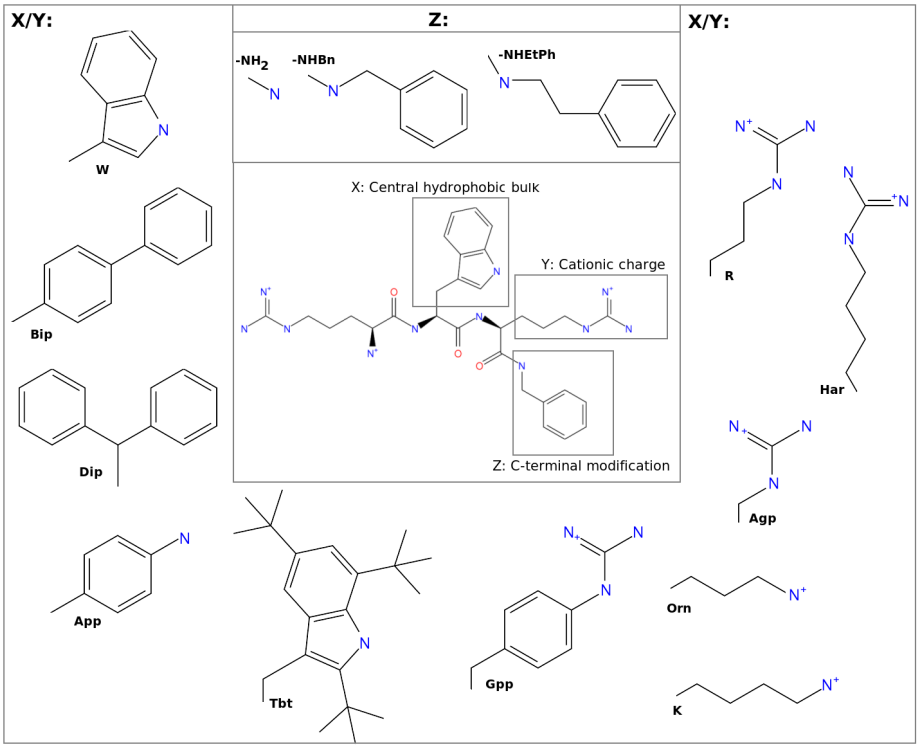


Figure 2.1: Illustration of residues and naming convention for the tripeptides.

interactions were treated with the extended cutoff method with truncation of the potentials after 4 Å, 8 Å and 20 Å for H-bonds, van der Waals and electrostatic interactions, respectively. The minimization was considered converged based on the gradient using a convergence threshold of 0.05. The minimized tripeptides were then prepared for the docking experiments with LigPrep [117] using the OPLS2005 force field. LigPrep was set to generate possible states at target pH between 4.0 and 8.0, at most 32 per ligand.

2.2 Electronic geometry optimization

Geometry optimizations of peptides were performed using density functional theory (DFT) with Gaussian03 [118] and Jaguar version 7.5 [119]. Gaussian03 and Jaguar offer different solvent models, and both were chosen to see if there were any difference in the performance of the programs and the solvent models. Calculations were run with the B3LYP [83, 120–122] and the M05-2X [123] functionals using the Pople basis set 6-31G(d,p) [124–133]. The charge of the molecules were set to 3 and the spin multiplicity to 1.

Calculations with Gaussian03 were performed with the self-consistent field (SCF) procedure which uses a combination of convergence schemes EDIIS [134] and CDIIS with no damping or Fermi broadening. The optimization was done in the presence of solvent (water) by placing the solute in a cavity within the solvent reaction field. The Polarizable Continuum Model (PCM) using the integral equation formalism variant (IEFPCM) was selected by including the

keyword SCRF=IEFPCM. Calculations with Jaguar were performed with the implemented PBF solvation model [119] and the DIIS [135] convergence scheme. The convergence criteria was specified with an energy change set to $5 \cdot 10^{-5}$ Hartree and a RMS density matrix change set to $5 \cdot 10^{-6}$. Thermochemistry was set to 1.00 atm pressure and a start temperature of 298.15 K.

2.3 Molecular Dynamics

2.3.1 Solvent simulation

All molecular dynamics simulations of peptides in solvent were carried out using the MD program package *Q* [136] developed by Åqvist and co-workers. The Q software is build up from the four programs Qprep, Qdyn, Qfep and Qcalc. These programs are used in three steps necessary for doing molecular dynamics simulation with Q, a preparation stage, a calculation stage and an analyzing stage, as illustrated in Figure 2.2. Before starting a molecular dynamics simula-

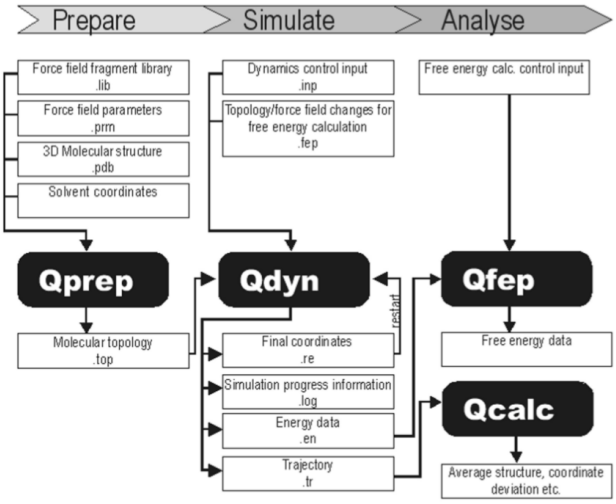


Figure 2.2: Overview of the procedure for calculations with Q. The white boxes represents files and typical file name extensions.

tions with Q one must prepare coordinates for the system that are in the correct format for Q, and a topology must be created for the system. The topology file is prepared with Qprep and contains all the information about the molecular system needed for running a simulation with Qdyn. Qcalc and Qfep can be used for analyzing the resulting trajectories, extracting energies, calculating free energies etc. depending on what type of simulation one has run and what type of information that is desirable.

Missing parameters and charges were assigned to all peptides according to the OPLS2005 all-atoms force field [137] with MacroModel Version 9.6 [114]. The topology was then prepared with Qprep5. The simulation center was defined as the C_α atom of the second amino acid residue counting from the N-terminal (the X residue in Figure 1.12). Each peptide was immersed into a spherical droplet of explicit water molecules with a radius of 25 Å centered at

the simulation center (C_α). Water molecules were described using the TIP3P model [138]. The C_α peptide atom (simulation center) and water atoms in the outermost 4.2 Å were weakly restrained to their initial positions with a harmonic potential of 5.0 kcal/(mol Å^2). The nonbonded potential was truncated at 10 Å for solvent-solvent interactions. Long-range electrostatics were treated using a multipole expansion method [139]. The peptide-solvent interactions were not truncated and the peptide was thus allowed to interact with the entire system. All systems were heated from 1 K to 300 K using a stepwise scheme, followed by an equilibration period of 500 ps. The step size was also gradually increased from 1 fs to 2 fs, whereas the bath coupling was increased from 1 to 10 during the heating. SHAKE [91] was used to constrain bonds and angles on solvent molecules, but not on the solute. A time step of 2 fs was used for the production phase, and the temperature was maintained at 300 K using a weak coupling to the bath. The production phase consisted of 50 ns and conformations were sampled every ps. A total of 3 such simulations were run for each tripeptide, differing in distributions of initial velocities, giving a final simulation time of 150 ns. All calculations were done on the NOTUR supercomputer STALLO [140]. Each calculation was submitted on eight processors and took around 12 days to finish.

The resulting trajectories were inspected with the Visual Molecular Dynamics program VMD [141]. To locate highly populated regions with stable conformations the RMSD (Eq.2.1) for each sampled conformation was calculated against the minimized starting structure.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} d_i^2}{N_{atoms}}}; \qquad d_i = |x_i - x_i'| \tag{2.1}$$

The RMSD plot was used to locate regions in the trajectory time line with favorable conformations. Once located, 10 ps from the populated regions were extracted with VMD. The current potential OPSL2005 energy for the corresponding conformations from these regions were then calculated with MacroModel using water as solvent. After calculating the current energy, the structures were energy minimized with OPLS2005 using water as solvent. All calculations were performed with the TNCG (Truncated Newton Conjugate Gradient) [142] optimization scheme and was chosen to converge on the gradient with a value of 0.001.

2.3.2 Membrane simulation

Peptide membrane interactions were studied with molecular dynamics simulations by using the parallel version of the MD program Desmond [143, 144]. The system was prepared with a 1-palmitoyl-2-oleoyl-glycero-3-phosphoethanolamine (POPE) membrane and explicit SPC [145] water molecules. The simulation was performed in an orthorhombic box ($\sim 30 \times 36 \times 100$ Å) where the box boundary was controlled by a buffer distance set to $10 \times 10 \times 10$ Å between the solute structure and the simulation box. The final peptide:lipid ratio was 1:42. The system was neutralized by adding Cl^- ions which were excluded within 7 Å of the cationic peptide to avoid peptide ion interactions. In order to study the mechanism of how the peptide approaches and enters the membrane, the peptide was manually placed outside the membrane with a peptide membrane

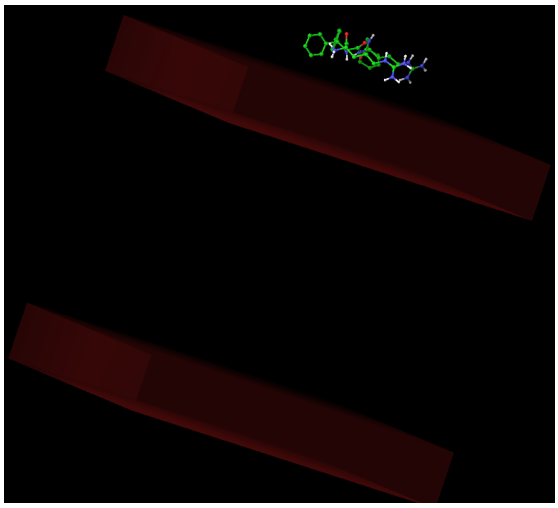


Figure 2.3: Illustration of the starting conformation for the membrane simulation. Peptide placed outside the POPE membrane (illustrated with red boxes).

distance of ∼3 Å as illustrated in Figure 2.3. All calculations were performed with the NPT ensemble at a temperature of 310 K and a pressure of 1 atm. The NPT ensemble was calculated with the Nose-Hoover [146, 147] thermostat method, with a relaxation time of 1.0 ps and a frequency update every second step, and the Martyna-Tobias-Klein [148] barostat method, using isotropic coupling with a relaxation time of 2.0 ps and a compressibility of $4.5 \cdot 10^{-5}$ (1/bar). Coulombic short range interactions were treated with a cutoff radius of 9 Å. Long range interactions were treated with the smooth Particle Mesh Ewald method [149] with a tolerance of 10^{-9}. Simulations of 50 ns were submitted using the RESPA (reference system propagator algorithm) time-stepping scheme [150] with time-steps set to 2 fs for bonded and near atoms and 6 fs for far atoms. Heavy atom-hydrogen covalent bonds were constrained with a shake tolerance of 10^{-8} with a maximum of 8 iterations.

2.4 Docking experiments

The docking program Glide [97] from Schrödinger was used to generate models of the peptide-chymotrypsin complex and the corresponding free energies of binding. Glide offers the possibility for docking and scoring a wide variety of ligands with different methods ranging from HTVS (high throughput virtual screening) for fast results to XP (extra precision) docking for more demanding calculations of the ligand-protein interaction. The search algorithm in Glide is based on a hierarchical filter, illustrated in Figure 2.4, that score hydrophobic and polar interactions followed by a Monte Carlo sampling. In general only a small number of the best refined poses is passed on to the third stage for energy minimization on the pre-computed OPLS-AA [137] van der Waals and electrostatic grid for the receptor. The minimized poses are in stage 4 re-scored

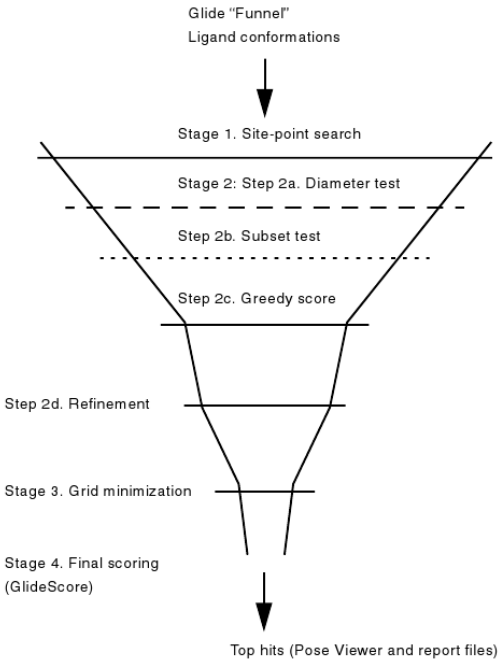


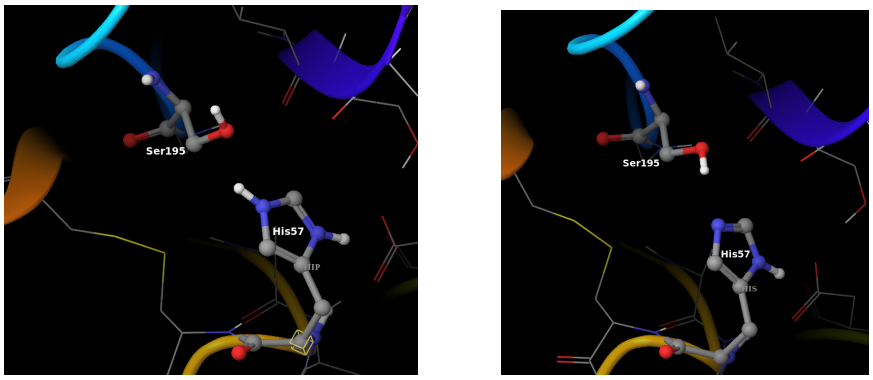
Figure 2.4: Illustration of the Glide docking hierarchy.

with Schrödingers *GlideScore* scoring function given as:

$$\begin{aligned} \Delta G_{GScore} &= 0.065 \cdot vdW + 0.130 \cdot Coul + Lipo + Hbond \\ &+ Metal + RotB + BuryP + Site \end{aligned} \tag{2.2}$$

Here Coul stands for Coulomb energy, Lipo the lipophillic contact term, vdW for van der Waals energy, HBond is the hydrogen-bonding term, Metal the metal-binding term where only the interactions with anionic acceptor atoms are include, BuryP is the penalty for buried polar groups, RotB is the penalty for freezing rotatable bonds, and site is the polar interactions in the active site term.

Coordinates for chymotrypsin were obtained from the crystal structure of the P1 Trp BPTI mutant- bovine α-chymotrypsin complex with the entry code 1t8o [46] from the Protein Data Bank, PDB [69]. The protein was prepared for the docking experiment in Maestro 9.0 [113] using the protein preparation wizard. The preprocess was run to assign bond orders, add hydrogens, treat metals and delete water molecules beyond 5 Å from protein heavy atoms. After the preprocess was run the protein was prepared using the Interactive Optimizer. The Analyze Network was run with the Include Current Orientations selected. The Optimize All was then run with high degree of sampling selected. His57 was found to be in the protonated state (Hip) with H-atoms bound to both Nδ and Nϵ. The hydroxyl hydrogen of Ser195 was also found to be oriented incorrectly as expected when His57 was in the protonated state. Hip57 was changed to Hid57 and the hydroxyl hydrogen from Ser195 was oriented towards Nϵ of Hid57, as illustrated in Figure 2.5. Finally the hydrogen bonding network



(a) Inital orientation **(b)** Final orientation

Figure 2.5: Orientation of the hydroxyl hydrogen from Ser195 in the active site of Chymotrypsin when changing Hip57 to Hid57.

was optimized with exhaustive sampling and sampling of water orientations prior to docking.

The receptor grid was prepared with Glide version 5.5 [97]. The P1 Trp BPTI ligand was truncated to only include the three residues 14-16 where residue 15 is Trp. These three residues are expected to be similar to the target peptides in this study, and were therefore specified as a ligand molecule in the grid generation. This truncated tripeptide defined the center of the grid box, and glide was set to dock ligands similar in size to this peptide. The van der Waals radius scaling was set to a scaling factor of 1.00 and a partial charge cutoff of 0.25. The grid was also prepared to allow rotation of the receptor hydroxyl groups Ser189 and Ser195.

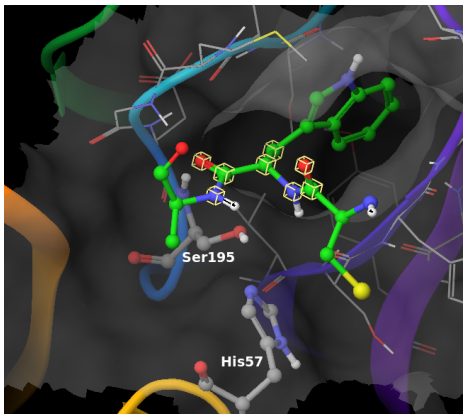


Figure 2.6: Core atoms selected from the reference ligand Trp BPTI. Yellow boxes illustrate the selected atoms for the SMART pattern.

All calculations were performed with the extra precision (XP) mode. The calculations were set to dock flexibly and to sample ring conformations. Non-planar amide bonds were penalized. Scaling of van der Waals radii on the ligands

was selected with a scaling factor of 0.80 and a partial charge cutoff of 0.15. For the Glide screens 10 000 poses per ligand were specified to be kept for the initial phase of the docking calculation. The scoring window for keeping the initial poses was set to 100 kcal/mol. The 1000 best poses per ligand were kept for energy minimization with a maximum number of conjugate gradient steps set to 5000. A maximum of one million ligand poses per docking run and 20 poses per ligand were collected. Glide was also set to perform post-docking minimization with 2000 poses per ligand included. The threshold for rejecting minimized poses was set to 0.50 kcal/mol. The docking calculations were performed both unconstrained and constrained. In the constrained docking calculation, 8 atoms from the truncated P1 Trp BPTI were selected as core atoms as illustrated in Figure 2.6. The selected atoms defined a SMARTS pattern and the calculations were restricted to dock to the reference position with a tolerance of 1 Å. The best poses were subjected to a refinement calculation where the key descriptors and per-residue interaction scores within 12.0 Å of the grid center were calculated.

2.5 Gelfiltration and SDS-PAGE

The α-chymotrypsin purity was inspected with gelfiltration and SDS-PAGE prior to the isothermal titration experiments. A reference calibration curve was prepared by mixing 40 μl each of aprotin (6.5 kDa), ribonuclease A (13.7 kDa) and carbonic anhydrase (29 kDa) from the GA Healthcare Gelfiltration Calibration kit LMW to a total volume of 200 μl in filtered and degassed 0.1 M NH_4HCO_3 buffer (pH 8.2). 0.2 ml of the mixed reference solution was injected into an ÄKTAExplorer (GE Healthcare) and eluated through a Superdex 200 10/300 GL size exclusion column with a flow rate of 0.5 ml/min (0.66 MPa). The absorption was measured at 260 nm and 280 nm. α-chymotrypsin was dissolved to a final concentration of 31 mg/ml in filtered and degassed 0.1 M NH_4HCO_3 buffer (pH 8.2) and run through the column in the same way as the reference solution. Fractions of 0.5 ml were collected between 12 ml and 33 ml based on the absorption curve for further analysis with SDS-PAGE.

15 μl of each collected sample was mixed with 15 μl of loading buffer. The loading buffer was prepared by mixing 0.5 ml of 4x LDS New-PAGE loading buffer, 0.4 ml mQ-H_2O and 0.1 ml Betamercapto EtOH reducing agent. The mixed solutions were then incubated for 5 minutes at 95^{o}C. The SDS-PAGE was run for 40 minutes with 150V together with 7μl molecular weight ladder Mark12 as reference.

2.6 Isothermal Titration Calorimetry

The heats of interaction were determined using a CSC 5300 nano-isothermal titration calorimeter III with a 1 mL cell volume (Calorimetry Sciences Corp., UT). The experiments were set up with 20 aliquots of 5 μL peptide (2.1 mM) added to a stirred (150 rpm) solution of α-chymotrypsin (0.1 mM) in an aqueous buffer at 25^{o}C. The reaction cell was allowed to equilibrate for 400 s before the titration was started. 400 s intervals were used between each injection to allow the interacting species to reach equilibrium. The buffer solution used was 0.1 M NH_4HCO_3 at pH 8.2. The heats of dilution were determined in a similar fashion

where the peptides were added to a stirred buffer solution without protein. Subtraction of the dilution heat yielded the heat of interaction and a binding isotherm from which the association constant and complex stoichiometry were calculated using NanoAnalyze software (TA Instruments, LLC, New Castle, DE).

2.7 Crystallization Experiments

2.7.1 Crystallization

Native crystals

Native crystals of α-chymotrypsin were grown by the hanging drop vapor diffusion method at room temperature. Drops consisted of a mixture of protein solution (30 mg/ml in distilled water), reservoir solution (10 mM sodium cacodylate pH 6.0 and 45-50% ammonium sulfate) and 1 M NaI. These were mixed in the ratio protein:reservoir:NaI = 5 μL:4 μL:1 μL, as described by Moulin *et al.* [151]. The protein solution was first pipeted onto a glass cover slip, then the reservoir solution was added, and finally the NaI solution was added. The drop was not mixed mechanically, but allowed to self-mix by simple diffusion. The well was filled with 700 μL of reservoir solution.

Soaking

Crystals of native α-chymotrypsin were, after growing to their final size, soaked in a solution of 20 mM sodium cacodylate (pH 7.4), 70% ammonium sulfate and 3 mM tripeptide for 2-3 weeks.

2.7.2 Data collection

Native crystals

Data of native crystal were collected at our home facilities with an R-AXIS IV image-plate detector (Rigaku, Japan) operating at 50 kV and 100 mA with a copper anode, to a final resolution of 1.8 Å. Crystals were harvested by passing them through a cryo solution of 22% PEG 3350 and 15% MPD in water before placing them in the beam where they were protected against radiation damage by a liquid nitrogen spray at a temperature of 100-120 K. Exposure times were 150.0 s with an incident wavelength of 1.5418 Å and an oscillation sweep of 0.5^o. A total of 300 images were collected. The data were indexed and integrated using XDS [152] and scaled with SCALA from the CCP4i [153] program suite (Collaborative Computational Project 4).

Soaking

Crystals left in soaking solution with RWAgp-NHBn were harvested and collected after two weeks at our home facilities as previously described to a final resolution of 2.1 Å. The exposure time was 300 s and a total of 360 images were collected.

Crystals soaked with RBipR-NHBn were collected after 3 weeks. They were passed through a cryo solution of 22% PEG 3350 and 15% MPD in water and

flash frozen in liquid nitrogen. Data was collected in Grenoble on the synchrotron beamline ID29 with a wavelength of 1.240 Å and an oscillation sweep of 0.5^o to a final resolution of 1.8Å. The exposure time was 0.20 s with a transmission of 19%. Data were indexed with Mosflm [154] and scaled with SCALA.

2.7.3 Refinement

Molecular replacement was performed with the maximum likelihood program Phaser [155]. Phases were derived from a starting model taken from the Protein Data Bank with entry code 1t8o [46]. A rigid body refinement followed by a restrained refinement was then performed with Refmac5 [156] applying default geometric parameters. The model quality was checked and adjusted manually in Coot version 0.6 [157] with the help of the electron density maps $F_O - F_C$ and $2F_0 - F_C$. Refmac5 was used to refine the structures after manual changes introduced in Coot in a reiterated procedure. Finally a restrained refinement was run with Refmac5 with Coot [157] set to automatically add/delete waters.

The structures were evaluated with PROCHECK [158] and by superimposing the structures on the deposited 1t8o structure with the program LSQKAB [159]. A temperature factor analysis was performed with BAVARAGE in the CCP4i program suite [153].

"If you're not part of the solution, you're part of the precipitate."

-Henry J. Tillman

3. Results and discussion

3.1 Conformational structures of tripeptides

It has previously been reported that one of the key features for antimicrobial activity of AMPs is that they must exhibit a suitable amphipathic structure in addition to a net positive charge for efficient membrane binding [11]. These features have however been suggested based on information from much larger AMPs (∼30 residues) than the small tripeptides from our research group. It was therefore of interest to examine the conformations of these peptides to see if they exhibit an amphipathic structure. As a step in investigating the properties of the tripeptides the three dimensional structure was examined using computational methods.

3.1.1 Geometry optimization with DFT fails

All tripeptides in this study have several aromatic regions which can give rise to internal stacking or packing. The chosen B3LYP [83] functional has previously been reported to fail in predicting the minima of stacked complexes [160]. The X3LYP functional [161] on the other hand was designed to describe non-covalent interactions, but Cerny and Hobza showed in their work that the X3LYP functional fails when locating the minima of dispersion-dominated stacked structures of nucleic acid pairs [160]. Recently Zhao *et al.* [123] reported a new functional called M05-2X that is claimed to greatly improve the performance for non-covalent interactions compared to previously DFT methods. It was therefore decided to geometry optimize the tripeptides with both M05-2X and B3LYP to see if there were any significant differences between the two functionals.

It was difficult to obtain convergence and most of the calculations crashed or died. Inspection of the output files showed very big fluctuations in the energies and forces, indicating problems with the cavity generation. Different ways of specifying the cavity were therefore tested with Gaussian03 where one or more of the following options for the cavity generation were used:

- Using sphere radii from the UFF force field, instead of the default UAO (RADII=UFF).
- Making the tessera larger to 0.4 instead of 0.2 which is default (TSARE=0.4).
- An older, and sometimes better, way of creating the cavity was tried by including the keyword OLDGEPOL.

- The NOADDSPH keyword meaning that spheres are not going to be added to smooth cavity was also tried.

However, this did not work either. In addition to the cavity problem, other methods like OLYP [120, 121, 162, 163] and HF [164–166] with both smaller and bigger basis sets were tried without any success. ADF [167] with OLYP and the STO-DZP basis set was also tested with the COSMO solvation model, but did again not converge.

Three protons from the tripeptides were also removed in an attempt to optimize uncharged molecules. The idea was that if the neutral molecule would converge to an optimized geometry in solvent, one could use that geometry and add one proton at a time between each new successive optimization. Unfortunately, this did not work either.

Peptides were however geometry optimized using both B3LYP and M05-2X along with the 6-31G(d,p) basis set in gas phase without any problems. It therefore appears to be quite clear that the solvent models offered in Gaussian03, Jaguar and ADF are not appropriate for dealing with molecules like these cationic tripeptides. We therefore decided to use other methods to investigate the peptide conformations.

3.1.2 Identification of stable structures in the MD simulations

As an alternative to geometry optimization with DFT, it was decided to investigate the phase space with molecular dynamics simulations. Simulations of 50 ns were run three times with the OPLS-AA force field for each peptide with the MD program Q. This was done by placing the peptide in a 25 Å simulation sphere of explicit water molecules. The stability of the simulations were inspected by checking the total energies and the solute-solvent interaction energies from the simulations. The total kinetic and potential energies were found with averages around 3800 and -19300 kcal/mol, respectively, for all peptides (Table 3.1). The total temperature for all simulations was 300 K ±5 indicating that the bath coupling is not too hard. The potential OPLS-AA interaction energies between the

Table 3.1: Q energies for selected petides (kcal/mol.

Peptide	Q-Surr.[a]		SUM Energies[b]		
	el.	vdW	Total	Potential	Kinetic
RBipRNH$_2$	-596.5	-11.2	-15582.0	-19415.5	3833.2
RBipR-NHBn	-586.3	-20.3	-15552.7	-19362.1	3809.4
RTbtR-NHBn	-585.7	-35.3	-15473.9	-19327.3	3853.4
RWR-NH$_2$	-595.7	-11.2	-15674.7	-19495.7	3821.0
RWR-NHBn	-586.2	-20.3	-15583.1	-19407.7	3824.7

[a] Peptide solvent interaction energies in kcal/mol taken as an average over 150 ns.
[b] Sum of energies in kcal/mol taken as an average over 2.5ns.

peptide and solvent have average electrostatic interaction energies between -580 to -600 kcal/mol and vdW energies between -15 to -20 kcal/mol. In Figure 3.1

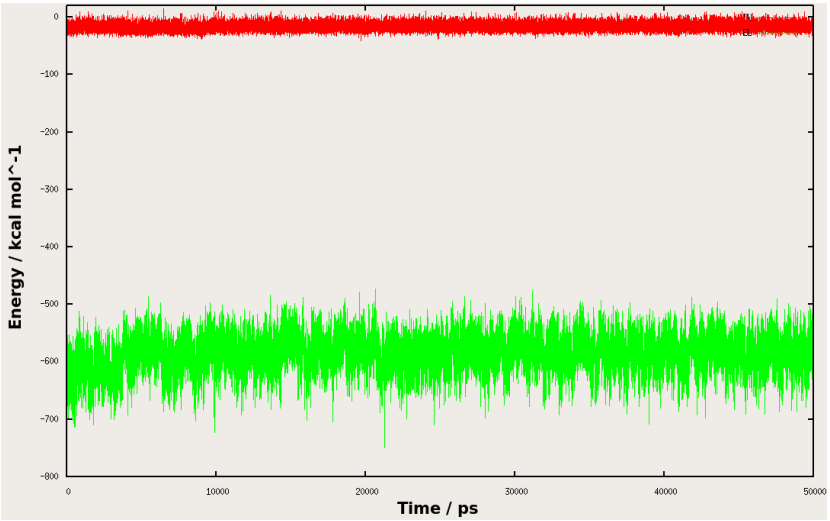


Figure 3.1: Peptide-solvent interaction energies in kcal/mol as observed in the MD simulations of RBipR-NH_2. Green is electrostatic interactions and red is vdW interactions.

the time development of the electrostatic and vdW solute solvent interaction energies are illustrated. The total energies, the temperature and the solute solvent interaction energies thus support that the simulations were stable.

In order to locate highly populated regions in the conformational space, the RMSD was calculated with the minimized starting structure as reference for each simulation. The RMSD plots from the 150 ns simulation time revealed that all the tripeptides had two or more highly populated regions in the explored conformational space, which indicates stable conformations. The RMSD plot for RBipR-NH_2 is illustrated in Figure 3.2 and shows how the simulation oscillates around 2 to 3 such conformations. This was also supported by visualizing the

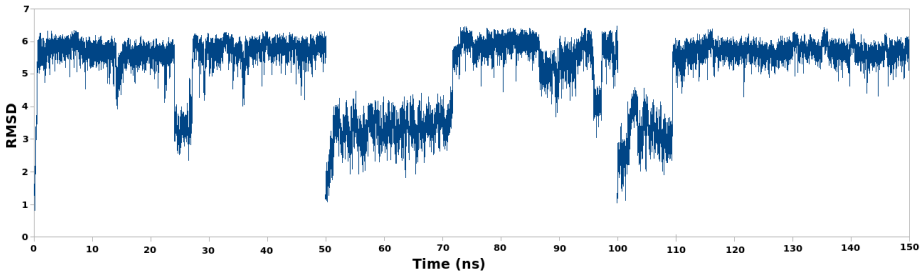
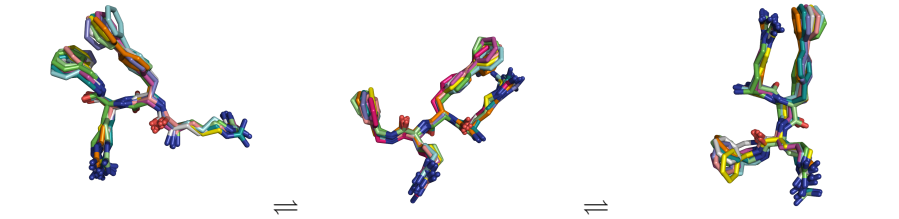


Figure 3.2: RMSD plot for RBipR-NH_2 from the 150 ns MD simulation in solvent (water).

MD trajectories with VMD [141] where it was found that the peptides stay more or less in one conformation before they completely change into the next.

After extracting and inspecting conformations from the most populated regions, it was found the tripeptides indeed do exhibit an amphipathic structure in at least one of the explored regions, but not in all of them. One of the most interesting results obtained from the MD simulations was that the peptides, especially those with two bulky groups, have at least two highly populated regions



(a) Amphipathic (two faces) ⇌ **(b)** "Transition" ⇌ **(c)** Not amphipathic (one face)

Figure 3.3: Face flipping: RBipR-NHBn going from an amphipathic conformation to a non-amphipathic conformation.

from the RMSD plot corresponding to conformations the peptide can switch between. In one region the peptides were found with an amphipathic structure (two faces) and in the other with non-amphipathic structure (one face). This indicates that the peptides are able to go from having only one face to having two faces, which will be referred to as face flipping. RBipR-NHBn is illustrated in Figure 3.3 with three different favorable conformations, observed in the simulation, resulting in face flipping. All peptides were found to alter between different conformations (Figure 3.4), but the face flipping property is most prominent for peptides with two bulky groups (in this case the additional benzyl group on the C-terminal). Interestingly, the MIC values (Table 3.2) also show that peptides

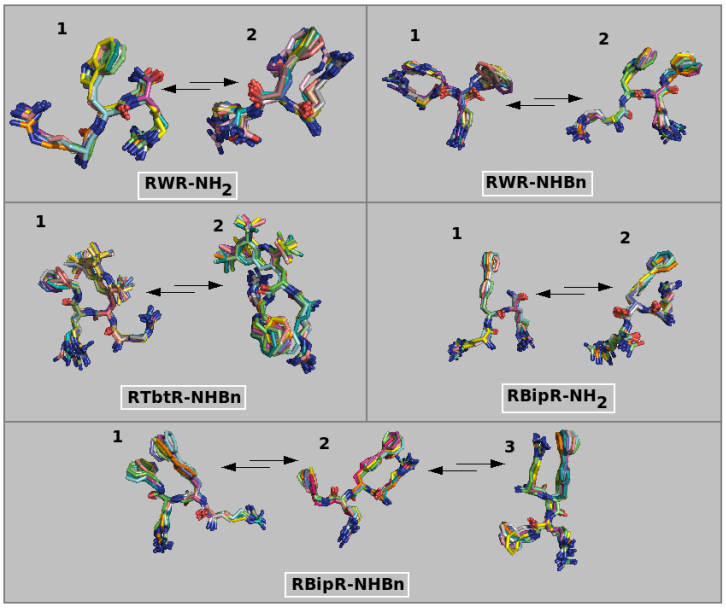


Figure 3.4: 10ps from the most obvious stable regions found from the RMSD plots superimposed on each other.

with benzyl as a C-terminal capping group are far more efficient than those

with the amino modified C-terminal. In some cases a ten-fold reduction in MIC values is observed when introducing benzyl as the C-terminal capping group [168]. Based on the simulations it is clear that the C-terminal benzyl group makes the difference between the amphipathic and non-amphipathic structures larger when compared to simulations with the C-terminal amide group.

Table 3.2: Antibacterial activity of selected peptides as MIC values[a]

Peptide	MIC (μg/mL)		
	S.aurerus[b]	*MRSA*[c]	*E. coli*
RWR-NH_2	>150	n.a	>150
RWR-NHBn	75	125	>150
RBipR-NH_2	>150	50	n.a[d]
RBipR-NHBn	10	15	>150
RBipGpp-NH_2	75	50	n.a[d]
RBipGpp-NHBn	5	5	100
GppBipR-NH_2	75	50	n.a[d]
GppBipR-NHBn	5	5	100
RTbtR-NH_2	7.5	7.5	60
RTbtR-NHBn	2.5	2.5	7.5

[a] Data from refs. 168, 169 and 110
[b] *S. aurerus* strain ATCC 25923
[c] Methicillin resistant *S. aurerus* strain ATCC 33591
[d] Data not available

In light of the few populated conformations observed for the peptides, the next question addressed was how easily the peptides can move between the different conformations. In order to obtain an estimate of the energy required for the peptides to change between different conformations, instantaneous structures from stable regions were extracted and the current energy was calculated with MacroModel [114]. After calculating the potential energy, the structures were energy minimized in order to optimize the geometry. The minimization resulted in fixing of non-planar hydrogens and bad torsions. The potential energies and the RMSD between the MD and the minimized conformation for selected peptides are summarized in Table 3.3. The minimization procedure did not change the overall conformation of the peptides, as revealed by the low RMSD values (Table 3.3). After minimization, the energies of the different conformations of each peptide were found to be surprisingly similar and only separated by 3.3 - 0.2 kcal/mol. The small energy difference between the conformations indicates that they are equally stable and thus quite flexible between each favorable conformation. RTbtR-NHBn goes through face flipping as illustrated in Figure 3.4, with an energy difference between the two conformations of approximately 0.2 kcal/mol. The low energy difference suggests that the peptide has an amphipathic and non-amphipathic conformation of equal stability. Interestingly, this particular peptide is also the most potent examined here.

The face flipping property and the low energy difference between different conformations suggest not only that the peptides are flexible, but also led us to believe that this might be a key features for the peptide to interact with a cell membrane. The CAPs may interact with a bacterial cell membrane with one face of the amphipathic structure and then switch conformation as it enters the

Table 3.3: Potential OPLS2005 energies (kcal/mol) of selected peptides before and after energy minimization and the corresponding RMSD (Å).

[a]Peptide	[a]Conf.	[b]Current Energy	[c]Minimized Energy	[d]RMSD
RWR-NHBn	1	-214.7	-286.6	0.9175
RWR-NHBn	2	-217.5	-287.1	1.4449
RWR-NH_2	1	-246.2	-299.7	1.9282
RWR-NH_2	2	-240.4	-299.3	0.5097
RBipR-NHBn	1	-203.6	-271.2	1.0646
RBipR-NHBn	2	-197.8	-267.9	1.4019
RBipR-NHBn	3	-189.3	-269.9	1.3215
RBipR-NH_2	1	-229.1	-285.5	1.5568
RBipR-NH_2	2	-217.2	-289.1	1.2725
RTbtR-NHBn	1	-186.7	-277.9	1.0671
RTbtR-NHBn	2	-180.1	-278.1	1.1811

[a] Peptide and numbers corresponding to those illustrated in Figure 3.4.
[b] Potential OPLS2005 energy for the unoptimized peptide extracted from the MD trajectory.
[c] Potential OPLS2005 energy for the energy minimized peptide extracted from the MD trajectory.
[d] RMSD between unoptimized and optimized peptide.

membrane with its flexible motions. These interesting results prompted us to investigate their mode of interaction with cellular model systems in more detail.

3.2 Face flipping and its implications for membrane interactions

Peptide membrane interactions were studied with molecular dynamics simulations using a 1-palmitoyl-2-oleoyl-glycero-3-phosphoethanolamine (POPE) membrane and explicit SPC water molecules. The simulations were initiated with the peptide placed outside the membrane (Figure 2.3) in order to study the mechanism of how the peptide approaches and potentially enters the membrane. The time development of the peptide-membrane vdW and electrostatic interaction energies observed in simulation of RBipR-NHBn are illustrated in Figure 3.5. The total peptide membrane interaction energies varied between -40 to -470 kcal/mol through the simulations with an average of around -260 kcal/mol. The vdW interactions were found to vary between 12 and -63 kcal/mol and the Coulombic interaction energies between 0 and -190 kcal/mol, depending on the peptide position relative to the membrane. During the first nano seconds of the simulations, the tripeptides were found to exhibit an amphipathic structure with the charged arginine groups oriented towards the head groups of the POPE membrane. After approximately 10 ns the peptides began, with their characteristic flexible movemens, to insert the benzyl group of the C-terminal into the membrane as illustrated in Figure 3.6 for RBipR-NHBn. This can also be seen from the plot for the vdW interaction energies for RBipR-NHBn in Figure 3.5 where the energies are close to zero until $\sim$10 ns. After that the vdW interaction energies decrease from around -5 kcal/mol to around -50 kcal/mol

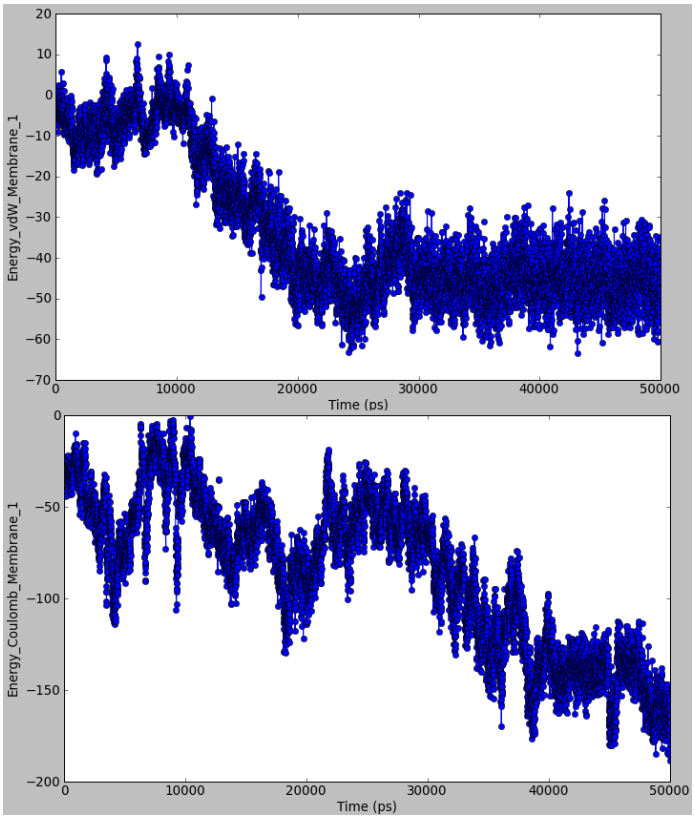


Figure 3.5: Peptide membrane vdW (top) and Coulombic (bottom) interaction energies for RBipR-NHBn (kcal/mol).

at ∼30 ns. In the interval between 25 and 30 ns RBipR-NHBn entered the membrane bilayer with the hydrophobic Bip residue pushing the hydrophilic head groups of the membrane out of the way (Figure 3.6). After approximately 30 ns RBipR-NHBn was found with both hydrophobic groups inserted into the membrane bilayer as illustrated in Figure 3.6. For the rest of the simulation, the peptide was found with an amphipathic structure with the hydrophobic groups buried in the membrane bilayer and the charged argenine groups anchored at the membrane surface interacting with the polar head groups. This can also be seen from the vdW plot in Figure 3.5 where the interaction energies becomes more or less stable around -50 kcal/mol during the last 20 ns of the simulation.

The rate of how fast the petides entered the membrane with both hydrophobic groups was found to vary between the simulations. All peptides studied successfully inserted the benzyl group of the C-terminal into the membrane bilayer quite fast (around 10 ns), but not all managed to perform the face flipping mechanism to get both hydrophobic groups into the membrane within the 50 ns simulation time. RDipR-NHBn, illustrated in Figure 3.7, for example showed the same mechanism as described above, but did not manage to bury the Dip residue in the membrane bilayer within the 50 ns of simulation. It is however not energetically favorable for such a hydrophobic residue as Dip to be left exposed

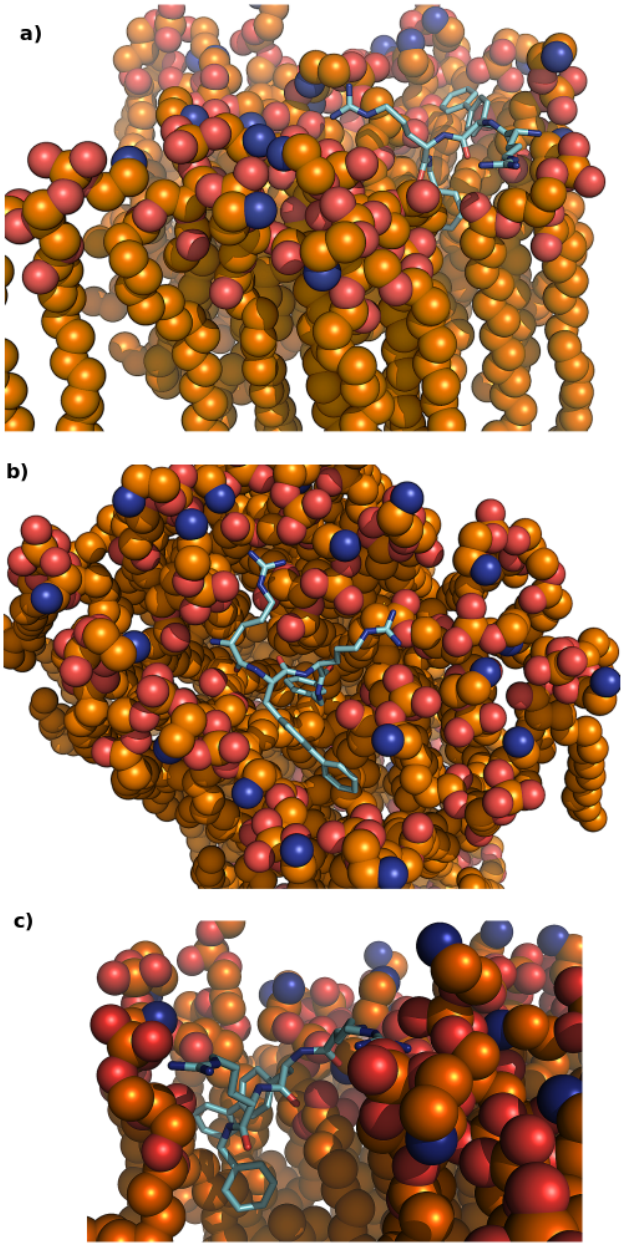


Figure 3.6: RBipR-NHBn entering the POPE membrane with the face flipping mechanism. a) Face flipping transition entering the membrane with the C-terminal benzyl group. b) Polar headgroups of the POPE membrane pushed out of the way as RBipR-NHBn flips the hydrophobic Bip residue into the membrane. c) Both hydrophobic groups buried in the membrane bilayer and the charged argenine residues anchored at the membrane surface.

to the polar environment of solvent (water) and the membrane head groups.

The simulations provide good support to the hypothesis that the peptides

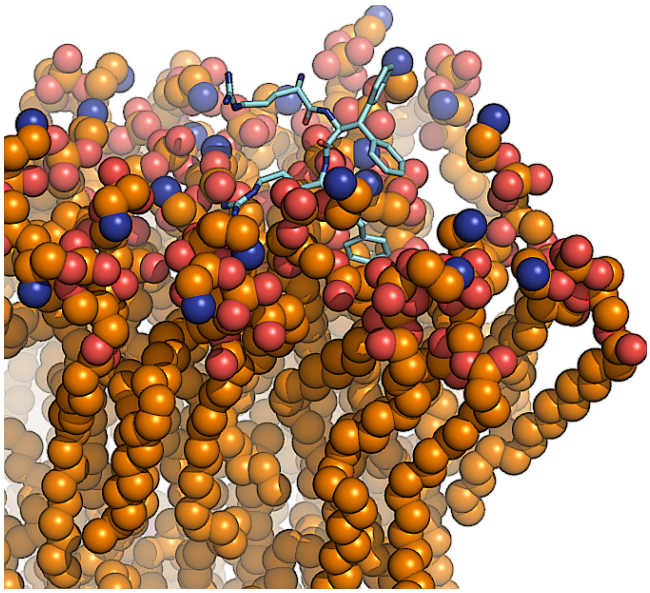


Figure 3.7: RDipR-NHBn with the C-terminal benzyl group buried in the membrane bilayer.

perform the face flipping mechanism as they enter the membrane. This means that they switch between amphipathic and non-amphipathic conformations in order to break through the polar membrane surface. The flexible peptide movements were also found to be quite effective in pushing the lipids out of the way as the peptide enters the membrane. The cationic charges seem to be very important as they interact strongly with the negatively charged parts of the membrane head groups. This will as a first step pull the peptide towards the membrane surface and make it stay there. This however leaves the peptide with its hydrophobic groups exposed to the solvent, which is not very favorable. It is therefore just a matter of time before the peptide flips its hydrophobic groups into the membrane bilayer. Previous studies have suggested the possibility for peptide interactions with intracellular targets [23, 24], but these simulations showed that the peptide is prevented from moving further into the membrane bilayer because the cationic charges work as an anchor at the polar membrane surface (Figure 3.6). If the peptides are going to cross the membrane bilayers and enter the cell, some sort of pore is required, but these preliminary studies indicate that the CAPs get "stuck" in the membrane bilayer and interactions with intracellular targets does not seem to be the primary mode of action. The results thus indicate that the flexible cationic tripeptides are pulled towards the membrane where they perform the face flipping mechanism to enter the membrane with their hydrophobic groups.

3.3 Refining the pharmacophore

All peptides studied here, except from RWR-NH_2, fulfil the previous reported minimum motif (two cationic charges and two hydrophobic bulk units) for an-

timicrobial activity, but the reported MIC values (Table 3.4) varies a lot. One

Table 3.4: Antibacterial activity of selected peptides as MIC[a] values compared with peptide volumes and surface areas.

Peptide	MIC (μM)		V[d]	FOSA[e]	PISA[f]	FISA[g]
	S.aurerus[b]	*MRSA*[c]	(Å^3)	(Å^2)	(Å^2)	(Å^2)
RTbtR-NHBn	3.2	3.2	2328	613	144	315
RBipR-NHBn	10	15	1901	240	368	306
RDipR-NHBn	25	10	1958	223	414	354
RWGpp-NHBn	50	10	1928	173	396	354
RGppW-NHBn	50	15	1909	204	356	380
RWAgp-NHBn	54	16	1739	210	289	405
RWR-NHEtPh	52	26	1950	246	340	428
RRW-NHBn	79	26	1809	258	274	383
RWHar-NHBn	78	52	1923	280	274	423
RWOrn-NHBn	110	55	1688	257	293	291
RWR-NHBn	79	132	1871	253	293	425
RWK-NHBn	163	54	1809	299	290	425
RWApp-NHBn	157	79	1791	195	398	291
RAppW-NHBn	157	79	1828	191	397	317
RFR-NHBn	165	83	1816	276	288	422
RWR-NH2	>200	>200	1544	223	158	415

[a] MIC data from refs.110 and 169
[b] *S. aurerus* strain ATCC 25923
[c] Methicillin resistant *S. aurerus* strain ATCC 33591
[d] V, Volume
[e] FOSA, Hydrophobic surface area
[f] PISA, π surface area
[g] FISA, Hydrophillic surface area

intersting question arising from this is what chemical features of the peptides determine how effective they are going to be in killing a microbe. A range of molecular properties were therefore calculated with QikProp [170] and compared with experimental MIC values. Some of the most interesting properties that correlate with the MIC values were the total solvent accessible volume, the hydrophobic (FOSA) and the hydrophilic (FISA) solvent accessible surface area (SASA) as shown in Table 3.4. The π component of the SASA (PISA) includes carbon and attached hydrogen, whereas the FOSA includes saturated carbon and attached hydrogen. FOSA, PISA, and FISA thus creates the total solvent accessible surface area where FOSA and PISA accounts for the hydrophobic and FISA the hydrophilic part of the SASA.

The total peptide volume seems to play an important role on the MIC values as illustrated in Figure 3.8 where the peptide volumes have been plotted against the avarage MIC values from Table 3.4. It also seems that the hydrophobic part (FOSA+PISA) of the SASA plays an important role on the MIC values (Figure 3.8). RTbtR-NHBn is in this study the peptide with the best MIC values, and also the one with both the largest volume and the largest hydrophobic surface area (approximately 71% of the total SASA). RWR-NH_2 is the only peptide in this study that display no antimicrobial activity against *S. aurerus*

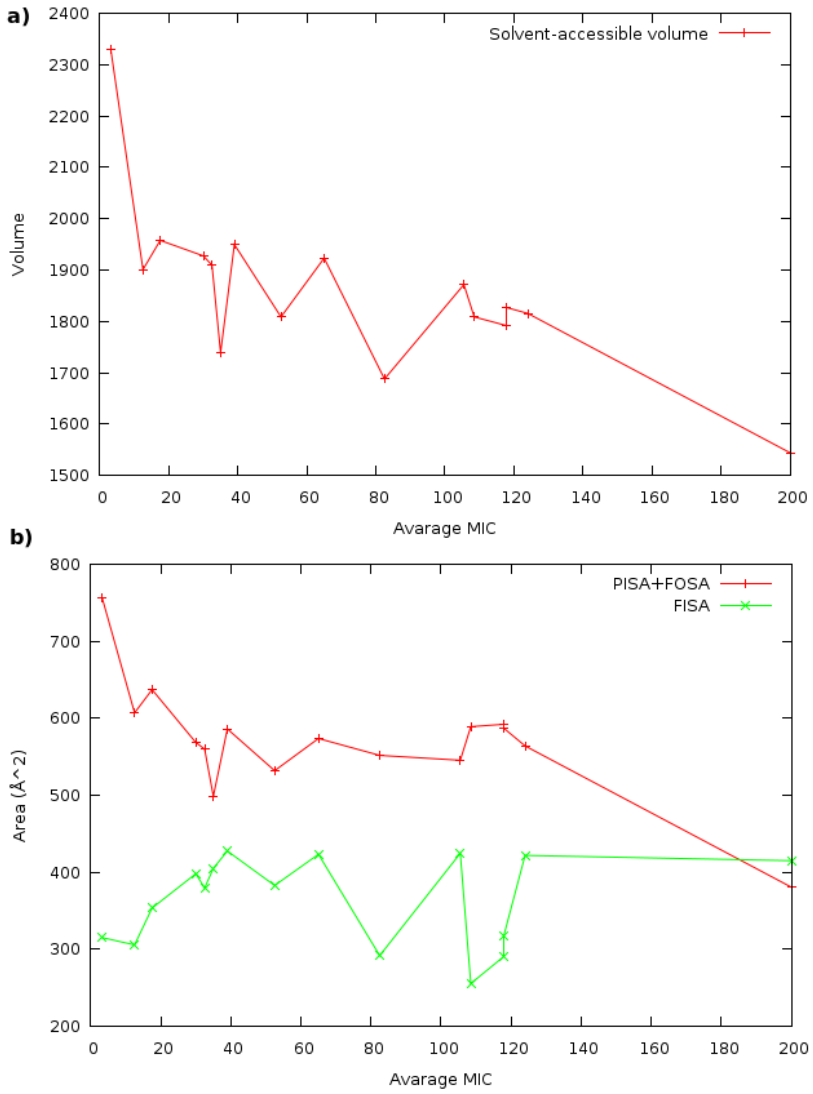


Figure 3.8: a) The calculated total solvent accessible peptide volume (Å^3) plotted against the average MIC values from Table 3.4. b) The total hydrophobic (FOSA+PISA) and hydrophilic part of the solvent accessible surface area plotted against the BAVARAGE MIC values from Table 3.4.

or *MRSA*. Interestingly, this peptide has not only the smallest volume, but also the smallest hydrophobic part of the SASA. Moreover, RWR-NH_2 is the only peptide from Table 3.4 that has a larger hydrophilic (∼52%) than hydrophobic (∼48%) solvent accessible surface area.

The results from the membrane simulations show that the tripeptides enter a bacterial cell membrane with the face flipping mechanism. These preliminary studies also suggest that the peptide volume and composition of hydrophobic and hydrophilic domains play a very important role on the antimicrobial prop-

erties. It thus seems that the hydrophobic solvent accessible surface area is required to be greater than the hydrophilic surface area for the peptide to exhibit antimicrobial properties.

3.4 Proteolytic stability

3.4.1 Protein purity

The protein purity was inspected prior to experiments to estimate the protein purity. that it was pure enough to work with. The UV 280 and 260 nm

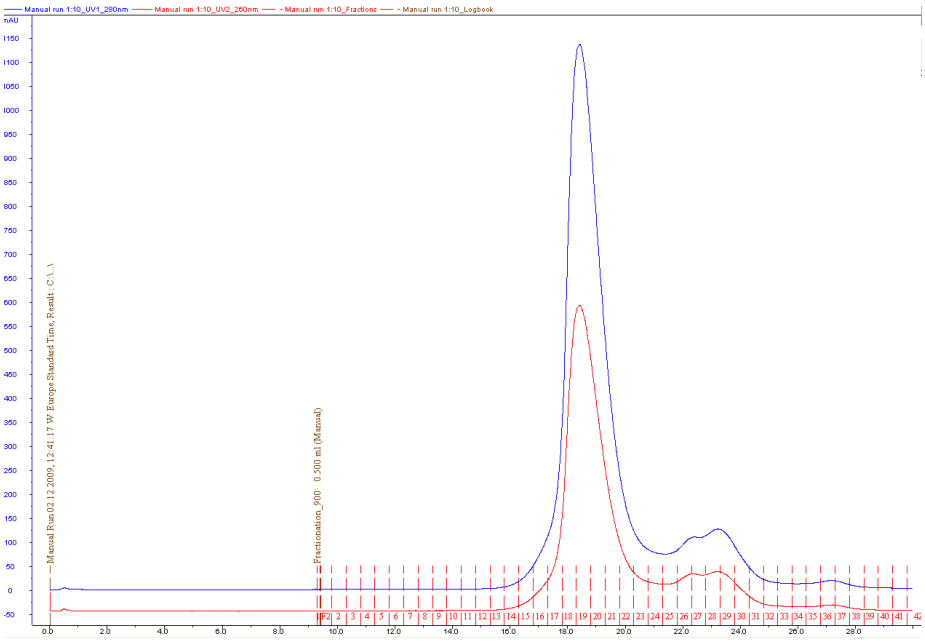


Figure 3.9: Absorption curves of α-chymotrypsin eluated through a Superdex 200 10/300 GL size exclusion column with a flow rate of 0.5 ml/min. Red line is UV 260 nm and blue curve is UV 280 nm. The collected samples (1-42) are illustrated on the x-axis.

gelfiltration spectrum of the elutant, illustrated in Figure 3.9, confirmed, after comparing with the reference molecular weight proteins, that the protein is pure and homogeneous. The first big absorption top at 18.4 ml is ∼25 kDa which is the molecular weight of chymotrypsin. The two small and connected absorption tops at 22.3 ml and 23.3 ml are both smaller than 6.5 kDa, because the absorption top of the reference protein with molecular weight 6.5 kDa came before 22.3 ml. The first top at 22.3 ml is ∼4 kDa and the second at 23.3 ml is ∼1.0 kDa. α-chymotrypsin is divided into the three chains A (Cys1-Leu13), B (Ile16-Tyr146), and C (Ala149-Asn245), which are linked together with disulfide bridges. The last two small absorption tops might be Cys1-Leu13 which has a weight of ∼1.2 kDa, either as monomers for the last top or dimers for the first top. These tops are however very small compared with the main top which is from α-chymotrypsin.

The collected fractions 15-37 from the gelfiltration were run on SDS-PAGE together with a Mark12 ladder (M) and a sample of chymotrypsin (R) that had not been gelfiltrated. The result from the SDS-PAGE is illustrated in Figure 3.10. The band between 31 kDa and 21.5 kDa corresponding to the

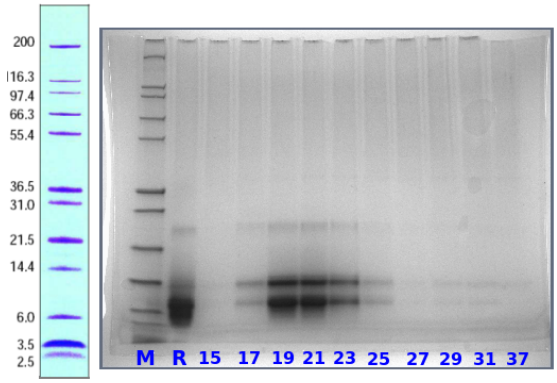


Figure 3.10: SDS-PAGE of α-chymotrypsin. M is the Mark12 ladder and R is a reference solution of α-chymotrypsin that has not been gelfiltrated. The numbers 15-37 are the respective numbers of the collected samples from the gelfiltration illustrated in Figure 3.9.

molecular weight of chymotrypsin of 25 kDa is very weak both for the referance chymotrypsin (R) and from the collected fractions (15-37). There are two strong bands both from the reference protein and the collected samples between 14.4 kDa and 6.0 kDa. These bands are most likely ∼14 kDa and ∼10 kDa which are good matches for chain B (Ile16-Tyr146, ∼13.9 kDa) and chain C (Ala149-Asn245, ∼10.1 kDa) from α-chymotrypsin. Chain A, B and C are connected with disulfide bridges, and the 5 minutes of incubating at 95 ^{o}C prior to SDS-PAGE with the reducing conditions of β-mercapto is probably breaking these connections. The SDS-PAGE results and the gelfiltration chromatogram support the protein being pure and homogenous.

3.4.2 Docking and ITC

Chymotrypsin is a well studied proteolytic enzyme that plays a major role in determining the fate of orally administrated peptides [171]. It has previously been reported that the substrate specificity of chymotrypsin is defined by the preferential binding of a hydrophobic residue in the S1 pocket which is additionally governed by a cationic side chain in the S1' pocket [172]. Bearing the substrate specificity in mind it is expected that the short cationic antimicrobial peptides with hydrophobic residues are expected targets for chymotryptic degradation. However, serine proteinases in general prefers at least six residues for optimal substrate binding and catalysis [45], something that in theory should make the tripeptides less suitable targets for chymotrypsin.

The binding modes of the peptides to chymotrypsin were investigated with docking experiments using the Glide software[97] and with isothermal titration calorimetry. The docking experiments were primarily carried out under the requirement that the main-chain of the peptide should match the observed mode

of the P2-P1' segment of the P1 Trp BPTI bound to chymotrypsin to reduce the number of false positives. BPTI is a transition state inhibitor of serine proteinases and the interactions and geometric arrangement around the scissile bond can be expected to resemble that of real substrates. Unconstrained docking gives similar results with respect to the energies, but with many structural solutions that can not lead to correct binding in the active site.

The free energies of binding obtained, presented in Table 3.5, ranged from -3.6 to -10.6 kcal/mol. The results presented in the table are the top ranked

Table 3.5: Constrained docking results from Glide in XP mode (kcal/mol) ranged by the experimental half life time ($t_{1/2}$).

Peptide	$t_{1/2}$[a]	ΔG_{bind}^{XP}	E_{Hbond}^{XP}	E_{Coul}^{XP}	E_{vdw}^{XP}	P1[b]
RWR-EtPh	0.5	-9.3	-1.9	-2.0	-6.2	W
RWHar-NHBn	1.1	-9.9	-1.3	-1.8	-7.7	W
RWR-NHBn	2.1	-10.6	-2.1	-2.0	-7.2	W
RFR-NHBn	3.3	-9.1	-1.8	-2.0	-6.3	F
RWK-NHBn	3.8	-10.4	-2.1	-2.0	-7.3	W
RWGpp-NHBn	4.5	-10.3	-1.5	-1.9	-7.5	W
RWR-NH_2	4.6	-6.1	-2.1	-2.0	-5.6	W
RRW-NHBn	7.2	-6.6	-2.2	-1.9	-5.5	R
RAppW-NHBn	17	-9.4	-1.7	-1.7	-6.7	App
RWOrn-NHBn	20	-10.6	-1.8	-2.0	-7.5	W
RBipR-NH_2	st.	-8.1	-1.4	-1.8	-5.4	Bip
RWAgp-NHBn	st.	-10.4	-2.1	-1.9	-7.4	W
RWApp-NHBn	st.	-9.4	-1.3	-1.0	-7.9	W
RBipR-NHBn	st.	-9.1	-1.3	-1.7	-7.0	Bip
RGppW-NHBn	st.	-6.9	-1.6	-1.6	-7.1	Gpp
RDipR-NHBn[c]	st.	-8.9	-2.6	-2.6	-6.2	R
RDipR-NH_2[c]	st.	-7.2	-1.7	-1.6	-4.8	R
RTbtR-NHBn[c]	n.a[d]	-3.6	-0.5	-1.5	-5.0	R

[a] $t_{1/2}$ values are in hours (data taken from ref.110).
[b] Residue in the S1 pocket of chymotrypsin.
[c] Results are from unconstrained docking.
[d] Data not available.

solutions of constrained docking except for RDipR-NHBn, RDipR-NH_2, and RTbtR-NHBn where no binding mode was observed. The top ranked solution for these are from the unconstrained docking. Based on the calculated free energies of binding it is clear that the choice of X residue (Figure 2.1) has varying impact on the binding to chymotrypsin. Going from RWR-NHBn to RAppR-NHBn reduces ΔG_{bind} from -10.6 to -9.4 kcal/mol respectively. With Bip as X (RBipR-NHBn) the free energy of binding is reduced to -9.1 kcal/mol, whereas X = Dip (RDipR-NHBn), it is further reduced to -8.9 kcal/mol. Tbt as X was found to bind in a much less favorable way compared to the other tripeptides with a calculated free energy of binding of only -3.6 kcal/mol for RTbtR-NHBn. Based on these results the following trend for the X residue in RXR-NHBn is suggested upon binding to chymotrypsin where X is ranked based on decreasing binding energies:

$$\text{W} > \text{App} > \text{Bip} > \text{Dip} > \text{Tbt} \tag{3.1}$$

Peptides with Tbt and Dip as X residues were however not found with any poses within the applied constraints, and they are consequently scored with poses from the unconstrained docking. Inspection of the results from docking RDipR-NHBn revealed that an arginine was placed at the entrance of the S1 pocket, where it interacts with the side-chain of Ser195 (Figure 3.11). The Dip

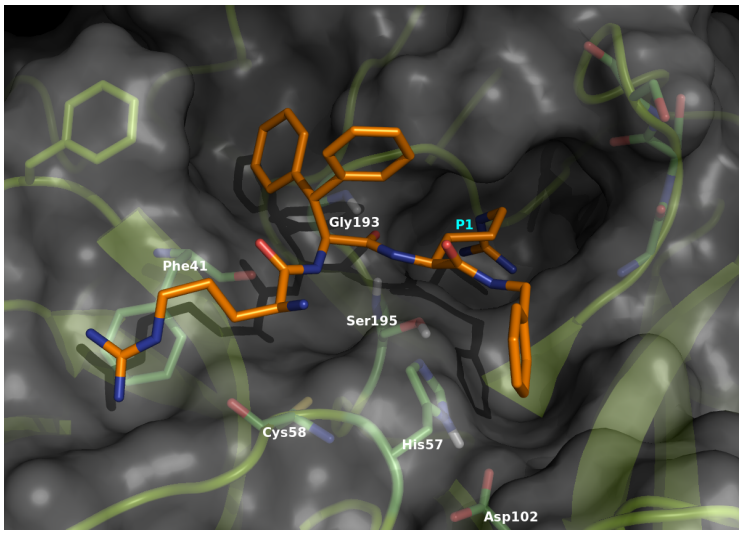


Figure 3.11: RDipR-NHBn oriented incorrectly in the active site of chymotrypsin for cleavage as revealed by docking experiments. The unfavorable arginine residue is in the S1 pocket and the Dip residue is placed in the S2' pocket.

residue is oriented in the S2' pocket and the peptide backbone going in the wrong direction for hydrolysis. ITC analysis of RDipR-NHBn did not yield a binding isotherm that differed from that of the heats of dilution which is in support of the results from the docking experiment. In a recent report, RDipR-NHBn was also found to be stable towards tryptic degradation [36]. RDipR-NHBn is thus a promising tripeptide for further development since it shows both good MIC values (Table 3.4) and stability towards tryptic and chymotryptic degradation.

RBipR-NHBn on the other hand binds well to chymotrypsin but is also oriented correctly for cleavage as illustrated in Figure 3.12. However, according to the experimental half life time, this tripeptide is stable towards chymotryptic degradation. After 24 hours only some trace amounts of product was observed suggesting that this is a bad substrate for chymotrypsin. This may also indicate that RBipR-NHBn is a potential inhibitor for chymotrypsin, something that is not revealed by the RP-HPLC experiment. The ITC experiments suggest that chymotrypsin binds RBipR-NH_2 in a 1:1 complex (Figure 3.13). Peptides that were sensitive to degradation by chymotrypsin could not be analyzed by ITC due to the interfering heat contribution generated from the hydrolysis. The calculated binding mode of RBipR-NHBn was however supported by the ITC experiments which yielded a strong ($K_D = 18.4\ \mu M$) favorable stoichiometric interaction (Table 3.6). One explanation for RBipR-NHBn being stable despite its good fit in the S1 pocket might be that the Bip residue is too long compared with the depth of the S1 pocket preventing the peptide from coming close

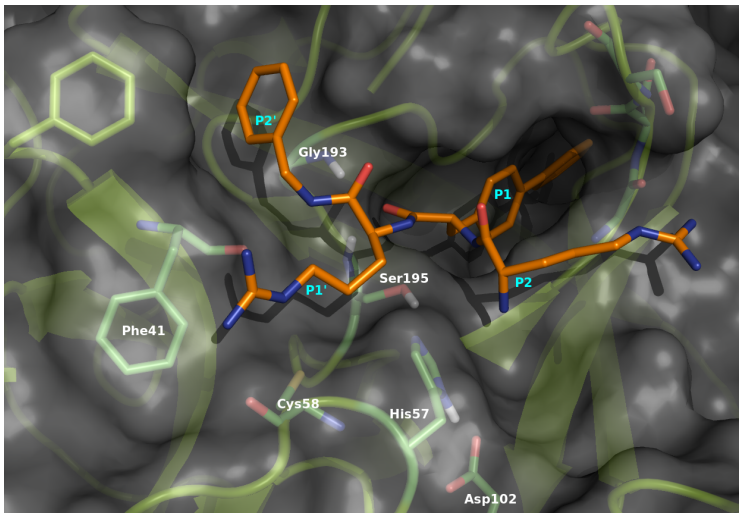


Figure 3.12: A docked model of RBipR-NHBn suggesting that it fits nicely in the active site of chymotrypsin acting as an inhibitor.

enough to the catalytic Ser195 for cleavage. In a recent report RBipR-NHBn was however found to be rapidly degraded by trypsin with a half life of only 2 hours [36], making it a less attractive candidate for oral administration.

Table 3.6: Binding data from ITC runs.

Peptide	$K_D{}^a$ (μM)	$n^{a,b}$	$\Delta H^{a,c}$ (kcal/mol)	$-T\Delta S^{a,c}$ (kcal/mol)	$\Delta G^{a,c}$ (kcal/mol)
RBipR-NHBn	18.4	0.8	-1.8	-4.7	-6.5
RBipR-NH_2	16.1	0.7	-5.4	-1.2	-6.6
RWAgp-NHBn	61.1	0.7	-1.0	-4.7	-5.7

[a]Data from the ITC experiments and the binding isotherms. Avarage values based on two parallel runs.
[b]Stoichiometry of the interaction; experimental error is ± 0.1.
[c]Experimental error is 15%.

The choice of C-terminal capping group (Z residue in Figure 2.1) was also found to be very important in tuning the stability towards chymotryptic degradation. The calculated free energy of binding for RBipR-NHBn is -9.12 kcal/mol, whereas replacing -NHBn with - NH_2, as in RBipR-NH_2, reduces the free energy of binding to -8.09 kcal/mol. The ITC experiment also showed a substantial decrease in $-T\Delta S$ from -4.7 kcal/mol to -1.2 kcal/mol upon the substitution of -NHBn with -NH_2 respectively for RBipR-Z (Table 3.6). It can also be seen from the half life times that the Z residue greatly affect the stability of the CAPs (Table 3.5). RWR-Z for example has a half life of 0.5 hours for Z = -NHEtPh, 2.1 hours for Z = -NHBn, and 4.6 hours for Z = -NH_2. Based on the results observed for RWR-Z the following trend for the Z residue is suggested upon binding to chymotrypsin:

$$\text{NHEtPh} > \text{NHBn} > \text{NH}_2 \tag{3.2}$$

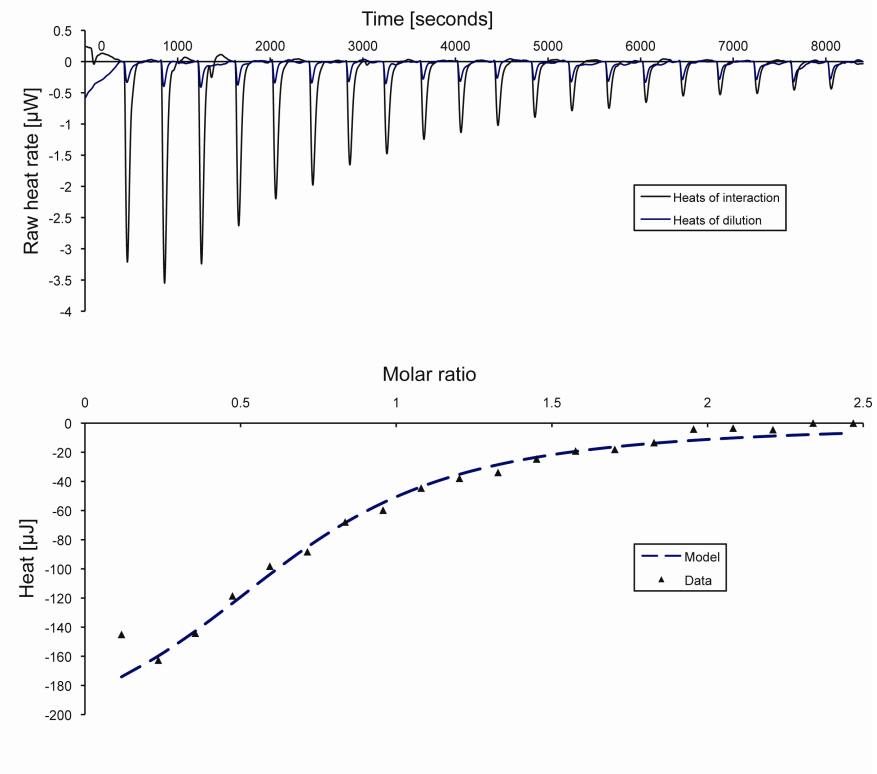


Figure 3.13: Representive raw data from the ITC analysis of RBipR-NH_2 (top figure) into a buffered solution of chymotrypsin and into buffer only. The lower figure illustrates the binding isotherm derived from the same run yielding a stoichiometric interaction and very good fit to the one-site binding model employed. The data from the first point is not included due to slow syringe leakage during the equilibration step prior to the first injection.

Upon binding to chymotrypsin it is thus a trend towards destabilization following incorporation of longer C-terminal additions. This behavior is in clear contrast to the results obtained from the previous trypsin study were the quite opposite trend was observed [36]. This illustrates how the same structural modifications easily can alter the stability towards similar enzymes differently.

The choice of the Y residues (Figure 2.1), primarily targeting the S1' pocket of chymotrypsin, was found to have a surprisingly large effect on the stability towards chymotrypsin. The calculated free energy of binding for RWR-NHBn was -10.7 kcal/mol, whereas the very similar Har (one additional methylene in its side chain) and Agp (one methylene less in its side chain) residues were calculated with binding energies of -9.9 and -10.4 kcal/mol respectively. The only difference between these three CAPs is the length of the Y residue (Arg $\pm$ 1 methylene) and consequently the calculated free energies of binding are very similar. At a first glimpse the docking results suggest that RWAgp-NHBn and RWHar-NHBn should be no less stable towards chymotryptic degradation than RWR-NHBn (Figure 3.14). However, according to experimental half life times RWY-NHBn is stable with Y = Agp, whereas for Har the half life is decreased to

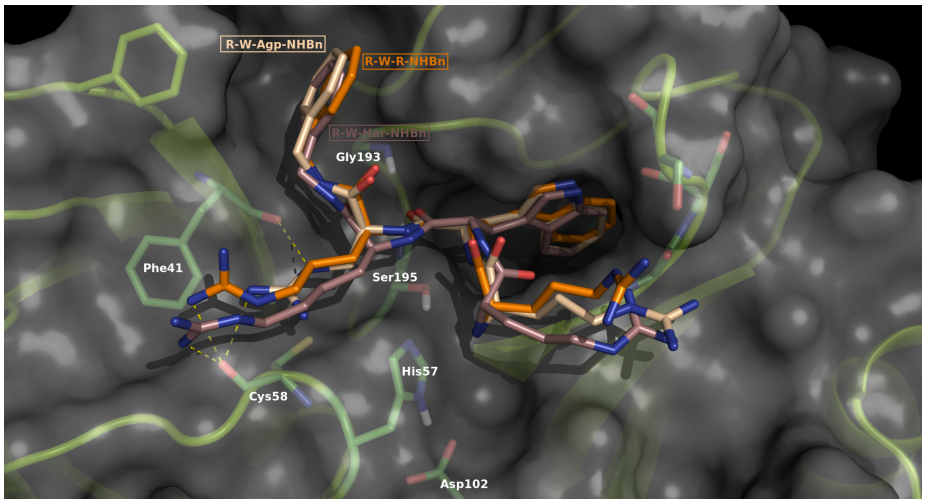


Figure 3.14: RWAgp-NHBn, RWR-NHBn and RWHar-NHBn superimposed in the active site of chymotrypsin. The hydrogen bonds between the main chain carbonyls of residues Phe41 and Cys58 in S1' pocket are illustrated by dotted lines.

only 1.1 hours compared with Arg which has a half life of 2.1 hours (Table 3.5). Based on the very similar binding modes of these three peptides, interactions in the S1' pocket seem to play a key factor in determining if the peptides are hydrolyzed or not. These interesting observations prompted us to investigate the residue specific interactions between the CAPs and chymotrypsin in more detail.

The key interactions between selected peptides with different P1' residues and the S1' pocket are presented in Table 3.7. The variation in the hydrophobic interactions (vdW) showed no correlation to the half life times and were virtually identical between the peptides and therefore only the Coulombic (electrostatic) is given. The main interactions found that differed between the peptides were

Table 3.7: Residue specific interaction energies (kcal/mol).

Peptide	$t_{1/2}$	ΔG^{XP}_{bind}	E^{Phe41}_{coul}	E^{Cys58}_{coul}	$\Delta E^{XP\ a}_{Coul}$
RWR-NHEtPh	0.5	-9.3	-4.3	-6.9	-2.6
RWHar-NHBn	1.1	-9.9	-4.3	-6.6	-2.3
RWR-NHBn	2.1	-10.6	-4.8	-6.1	-1.3
RWR-NH_2	4.6	-6.1	-5.3	-6.1	-0.8
RWAgp-NHBn	st.	-10.4	-5.4	-3.5	1.8

[a] $\Delta E^{XP}_{Coul} = E^{Cys58}_{coul} - E^{Phe41}_{coul}$

the Coulombic interactions with the main-chain carbonyl oxygens of residue Phe41 and Cys58, both part of the S1' pocket. Interactions with Phe41 pulls the peptide away from the catalytic Ser195, whereas interactions with Cys58 seems to counteract this. Based on the trend from the calculated energies it is suggested that if the Coulombic energy is greater for Cys58 than for Phe41, the peptide will be pulled towards the catalytic triad and cleaved, if not it will

be stabilized and act as an inhibitor. This difference (ΔE_{coul}) explains the difference of of interactions for the similar peptide side chains R, Har and Agp in the S1' pocket with good linearity (R^2 = 0.9, data not shown) despite the similar ΔG_{bind} values shown in Table 3.5. ITC control experiment on RWAgp-NHBn was performed and confirmes that this tripeptide binds to the active site in a stoichiometric manner similar to RBipR-NHBn with a dissociation constant of 61.1 μM (Table 3.6). Interestingly, RWR-NHEtPh is the CAP in this study with the shortest half life time and also the one with the largest difference in Coulombic interaction energies between Phe41 and Cys58 (ΔE^{XP}_{coul} = -2.55 kcal/mol). The calculated free energies of binding did not explain

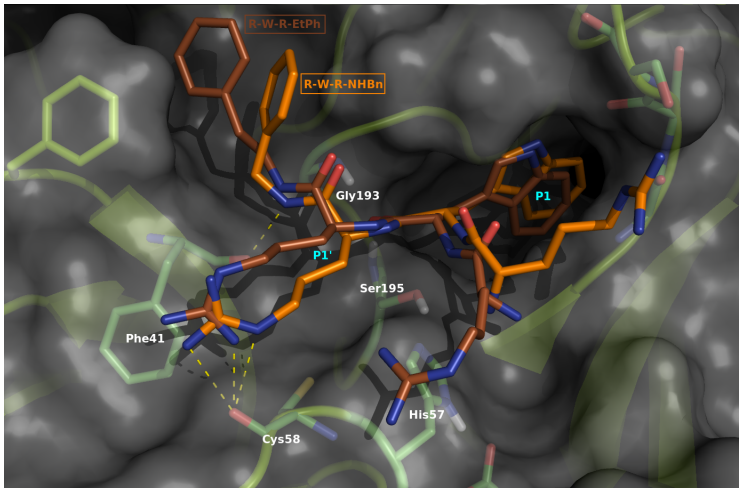


Figure 3.15: RWR-NHBn and RWR-NHEtPh superimposed in the active site of chymotrypsin. The C-terminal capping moiety binds to the S2' site giving rise to differences it the residue specific interactions between the R side-chain and the S1' pocket.

the destabilizing effect observed when replacing the Z capping group -NHBn with -NHEtPh, but the residue specific interactions in the S1' pocket did. It thus seems that longer capping groups give rise to different interactions in the S1' pocket, favoring Cys58 and consequently hydrolysis (Figure 3.15). These findings illustrates how the interactions in the S1' pocket plays a decisive role for the degradation of short cationic peptides with other residues available to fixate the scissile bond in place over the active site.

Concerning the very short unusual peptides included in this study, the sequence dependence is not as pronounced and recent studies have shown that confused peptides can yield the same antibacterial activities [168]. From a stability perspective this is an attractive drug design option and some peptides of the form RXW-NHBn was therefore included in addition to the "native" RXR-NHBn (X being a variable residue) providing peptides with a different overall distribution of charge and bulk. From the MIC values presented in Table 3.4 it can be seen that RRW-NHBn, RAppW-NHBn and RGppW-NHBn display similar antibacterial activities as the "native" RWR-NHBn, RWApp-NHBn and RWGpp-NHBn. From the calculated free energies of binding and the half lifes presented in Table 3.5 it can also be seen that the stability towards chymotryptic degradation thus is sequence sensitive. RWGpp-NHBn for example was

docked with a binding energy of -10.3 kcal/mol and has a half life of 4.5 hours, whereas the confused RGppW-NHBn was calculated with a binding energy of -6.9 kcal/mol and also found to be stable towards degradation. RAppW-NHBn and RWApp-NHBn on the other hand were calculated with quite similar binding energies around -9.4 kcal/mol, but RWApp-NHBn is stable whereas RAppW-NHBn has a half life of 17 hours, indicating that the combination of App in the S1 and W in S1' is a better fit for hydrolysis compared with the combination of W in S1 and App in S1'.

The calculated free energy of binding for RRW-NHBn was -6.6 kcal/mol suggesting a much less favorable binding mode by changing the peptide sequence from RWR to RRW. Docking of the peptide RRW-NHBn indicated a surprising binding mode with the central arginine residue placed in the S1 pocket of chymotrypsin (Figure 3.16) yielding a trypsin-like activity with a bulky tryptophan indole in the S1' pocket. This binding mode was also supported by the

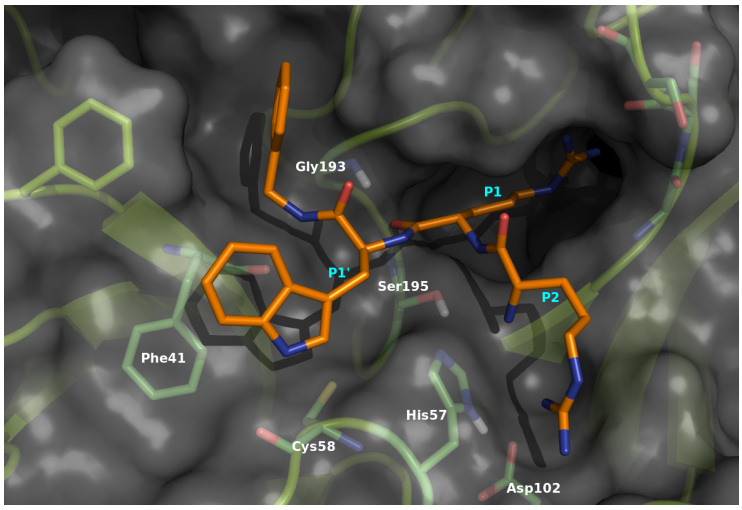


Figure 3.16: Docking of RRW-NHBn in the active site of chymotrypsin. Both modelling results and the degradation assay indicates that the S1 site interacts with the side chain of arginine in an unexpected manner yielding a trypsin-like activity with a bulky tryptophan indole in the S1'.

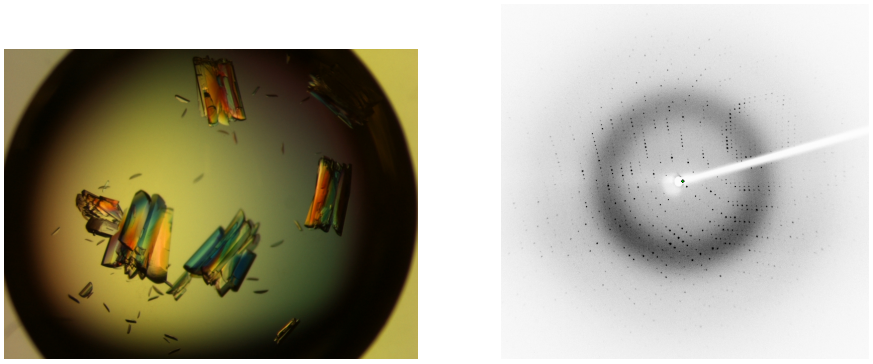
degradation assay which yielded the degradation products RR and W-NHBn with a half life of 7.2 hours. As illustrated in Figure 3.16, the P1 arginine of RRW-NHBn in S1 is bent in an attempt to minimize the interactions with the hydrophobic pocket in a binding mode previously described for chymotrypsin inhibitors containing lysine [49, 173]. This highligths that while sequence reversal may be employed to increase the stability, it is not always a general solution for these compounds.

3.4.3 Crystal structure

Molecular docking can provide important insight into the potential binding mode of the peptides to chymotrypsin, but there is no guarantee that the computational results resemble the real binding mode. It is therefore of great interest to obtain experimental structures of the peptide chymotrypsin complex. An experimental verified binding mode is thus of importance for further development of the peptides with respect to stability design.

It was initially decided to try to co-crystallize the peptide with chymotrypsin through screening different conditions such as buffers (Tris-HCl, MES, Hepes, Sodium Cacodylate etc.), precipitants (ammonium sulfate, LiCl, PEG 3-6 K, glycerol etc.), pHs (4-8), concentrations and temperatures (4, 20 and 37 ^{o}C). Co-crystallization did not result in crystals, and due to the time limitation of this project, it was in the end decided to try obtaining native chymotrypsin crystals for doing soaking experiments.

Native crystals of α-chymotrypsin were crystallized with the hanging drop vapor diffusion method and grew to their final size within 3-5 days (Figure 3.17). The drops consisted of protein (20-30 mg/ml) : reservoir solution (10 mM



(a) Native CHT crystals **(b)** Diffraction pattern

Figure 3.17: α-chymotrypsin crystals and diffraction pattern. (a) Crystals of native α-chymotrypsin grown by the hanging drop vapor diffusion method at room temperature with drops consisting of protein (30mg/ml):reservoir(10mM sodium cacodylate, pH 6.0, and 47% ammonium sulfate):NaI (1M) mixed in the corresponding ratio 5μL:4μL:1μL. The well consisted of 700μL reservoir solution. (b) Diffraction pattern of native crystals collected with a Rigako RAXIS-IV image plate detector.

sodium cacodylate, pH 6.0, and 44-49 % ammonium sulfate) : NaI (1M) mixed in the corresponding ratio 5 μL : 4 μL : 1 μL. Many of the larger crystals had grown together with other crystals, but these were quite easily separated with either the cryo loop or a needle and still diffracted well (Figure 3.17). A reference data set was collected on the native crystals at our home facilities to a final resolution of 1.8 Å (data in Table 3.8). Native crystals of α-chymotrypsin (native) were left in a soaking solution consisting of 20 mM sodium cacodylate (pH 7.4), 70% ammonium sulfate and 3 mM tripeptide for 2-3 weeks. The crystals seemed unaffected by the soaking solution, but dissolved in lower concentrations of ammonium sulfate than ∼65% after a couple of days. Crystals soaked with RWAgp-NHBn (Agp) were collected at our home facilities after 2 weeks to a

Table 3.8: Statistics in data collection and processing

Parameter	CHT structure:		
	Native	Agp[a]	Bip[b]
Overall Resolution range (Å)	40.00-1.80	38.05-2.20	35.36-1.80
Unit cell (Å)	a=42.12	a=65.10	a=41.32
	b=75.68	b=76.10	b=75.53
	c=65.25	c=84.78	c=61.39
	$\beta = 109.41$	β=108.49	β=104.44
Space group	$P2_1$	$P2_1$	$P2_1$
Wavelenght (Å)	1.5418	1.5418	1.2400
No. reflections	140210	146700	147081
No. unique reflections	33936	40065	33784
R_{merge} (%)[c]	5.5	7.7	6.3
R_{meas} (%)[d]	6.3	9.0	7.2
Completeness (%)	94.4	100	99.7
Mean $(I/\sigma I)$	9.32	5.28	8.24
Mean $(I/Sd(I))$	16.69	10.40	14.68
Wilson B ($Å^2$)	19.6	23.0	17.29
Multiplicity	4.1	3.7	4.4
Resolution of outer shell (Å)	1.80-1.89	2.32-2.20	1.90-1.80
No. of reflections	15811	21189	21480
No. of unique reflections	4299	5857	4896
R_{merge}	29.2	23.3	26.4
R_{meas}(%)	33.9	27.5	30.1
Completeness	81.9	100	99.9
Mean $(I/\sigma I)$	2.45	3.12	2.84
Multiplicity	3.7	3.62	4.4

[a] Agp: CHT soaked with RWAgp-NHBn
[b] Bip: CHT soaked with RBipR-NHBn
[c] $R_{merge} = \sum \sum_j |I_{hj} - \langle I_h \rangle| / \sum \sum_j I_{hj}$
[d] $R_{meas} = \sum (n/(n-1)) \sum_j |I_{hj} - \langle I_h \rangle| / \sum \sum_j I_{hj}$

final resolution of 2.1 Å. Crystals soaked with RBipR-NHBn (Bip) were collected after 3 weeks in Grenoble on the synchrotron beamline ID29 to a final resolution of 1.8 Å. The statistics of the data collection are summarized in Table 3.8.

Native and Bip crystals belonged to the monoclinic space group $P2_1$ with two molecules in the asymmetric unit. The Agp crystal also belonged to the space group $P2_1$, but had 4 molecules in the asymmetric unit. The calculated solvent content was 37, 38 and 34 % with matthew coefficients of 1.97, 1.99 and 1.89 $Å^3$ for native, Agp and Bip, respectively. The rather low solvent content indicate that the packing structure is rather tight in the crystals. This in turn may make it difficult to soak the tripeptides into the active site.

The solution of the structures were obtained using the molecular replacement program Phaser with Z-scores as high as 16-23, indicating clear solutions (the boundary value for a clear solution is 8). After obtaining the solutions from Phaser, a rigid body refinement followed by a restrained refinement was performed with Refmac5 before the model quality was checked and adjusted manually in Coot with the help of the electron density maps $F_O - F_C$ and

Table 3.9: Final refinement statistics

Parameter	CHT structure:		
	Native	Agp	Bip
R_{work} (%)[a]	15.9	19.8	17.6
R_{free} (%)[b]	22.2	28.2	24.7
No. protein atoms	4096	7600	3949
No. Solvent molecules			
SO_4^{2-}	2	0	0
H_2O	578	584	433
Avarage B factor (Å^2)			
overall	18.67	17.84	19.97
Main chain	15.98	16.86	17.69
Side chain	17.79	17.90	19.95
solvent atoms	35.12	27.56	34.22
rms deviations			
Bond lenghts (Å)	0.022	0.651	0.022
Bond angles (^{o})	1.936	1.548	1.940
Coordinate error			
DPI (Å)[c]	0.141	0.270	0.152
Luzzati (Å)	0.156	0.256	0.180
Ramachandran plot:			
Res. in favored regions (%)	98.5	95.9	96.8
Res. in allowed regions (%)	1.1	3.6	2.6
Res. in outlier regions (%)	0.4[d]	0.4[e]	0.6[f]

[a,b] $R = \sum |F_o - F_c| / \sum F_o$
[c] Diffraction-component precision index
[d] Ser77 and Glu78 chain A
[e] Ser77 and Glu78 chain B and D
[f] Glu78 chain A and Ser77 and Glu78 chain B

$2F_0 - F_C$. The native structure was refined with Refmac5 after manually introducing changes in Coot in a iterative procedure to a final R_{work} of 15.9 and the R_{free} of 22.2. The ramachandran plot of the native chymotrypsin crystal structure (Figure 3.18 a) shows that the overall geometry is good with 98.5 % in favored regions and only 0.4 % in outlier regions (Ser77 and Glu78). By comparing the crystal structure with PDB entry 1t8o it was found that the overall geometry was very similar with a calculated RMSD between the structures of 0.954 (Figure 3.18 b). As can be seen from Figure 3.18, it is the expected loop regions that are differing the most, whereas the core and secondary elements are conserved.

From the inspection of the electron density maps in Coot it became clear that there were no signs of RWAgp-NHBn or RBipR-NHBn in the active site of chymotrypsin from the soaking experiments. Chymotrypsin is a well-studied protein and additional efforts to improve the statistics were not made without the peptide in the active site. As previously mentioned, the crystal packing structures are rather tight, and this might be one reason for the soaking to fail. Crystallization and soaking conditions involved very high concentrations of ammonium sulfate and the sulfate anions may interact with the cationic pep-

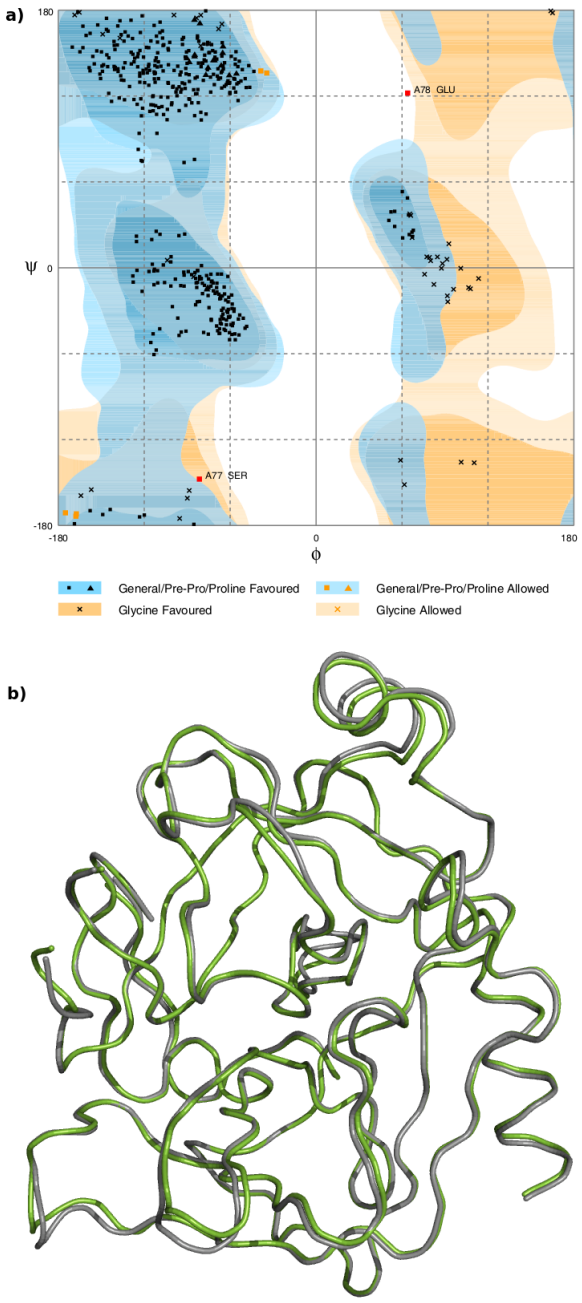


Figure 3.18: a) Ramachandran plot of native α-chymotrypsin. b) Superimposition of backbone between crystal structure of native α-chymotrypsin (green color) and PDB entry code 1t8o (grey color) with an RMSD of 0.954.

tides making it difficult for them to interact with the protein and the active site. Interestingly, crystals of native chymotrypsin grew easily, but trying to co-crystalline around the same conditions gave no crystals. The drops stayed

either clear, or with increased concentrations of ammonium sulfate it precipitated. It thus seems that ammonium sulfate is a bad choice for crystallization experiments with these cationice tripeptides. This indicates that future crystallization experiments should screen with uncharged precipitants such as PEG, since salts may interfere with the cationic tripeptides.

"Shall I refuse my dinner because I do not fully understand the process of digestion?"

-Oliver Heaviside

4. Concluding remarks

4.1 Face flipping, a membrane destabilizing property

From the molecular dynamic simulations of the tripeptides in solvent it was found that the peptides exhibit 2-3 stable conformations where at least one is amphipathic. These conformations were further found to be equally stable based on the low energy differences between them. Visual inspection of the corresponding trajectories shows that the peptides stay in one conformation and rapidly undergoes a complete structural rearrangement yielding the second conformation. The peptides thus switch between stable conformations that are amphipathic and non-amphipathic, a property introduced as face flipping. The dynamic flexibility between equally stable peptide conformations may be a reason why geometry optimization with DFT fails when including solvent (continuum). It was thus suspected based on the energies that something within the cavity generation failed during the optimization.

Tripeptides with two bulky groups (Bn as capping group) display a more distinct difference in structure, and face flipping is consequently most prominent for these peptides. Interstingly, the MIC values (Table 3.4) also show that peptides with benzyl as the C-terminal capping group are far more efficient than those with an amide group as C-terminal. The face flipping property was based on these findings proposed as a potential mechanism for membrane interaction.

Peptide membrane interactions were studied with MD simulations using a POPE membrane as model system. By starting the simulation with the peptide placed outside the membrane, it was possible to examine the mechanism of how the peptide approaches and enters the membrane. The simulations indicate that the peptides indeed do perform the face flipping mechanism as they enter the membrane. All the studied peptides successfully inserted the benzyl group of the C-terminal into the membrane bilayer (around 10 ns), but not all managed to perform the complete face flipping mechanism to get both hydrophobic groups into the membrane within the 50 ns simulation time. It might be speculated in that this is the reason why peptides with -NHBn as C-terminal capping group are more efficient than those with the polar -NH_2, since it seems that the benzyl group enters the membrane more easily compare with the central X bulk (Figure 2.1).

Ideally the membrane system used for the MD simulations should have been composed as a mix of several other phospholipids such as PG, DPG, and L-lysl PG in order to represent a more realistic bacterium membrane (see Table

1.2). Due to time limitations this was not possible to accomplish in the context of this project, and must thus be studied in more detail in future projects. The results from the membrane simulations must therefore be considered as a preliminary study of the CAPs membrane interaction mechanism based on the face flipping property. However, the simulations provide strong support to the hypothesis that the peptides perform the face flipping mechanism as they enter the membrane. Based on these preliminary studies it is suggested that the CAPs switch between amphipathic and non-amphipathic conformations in order to break through the polar membrane surface. Once the hydrophobic residues are inserted into the membrane the peptide stay more or less in the amphipathic conformation. This has also been confirmed by recent unpublished NMR data that show more interaction between the C-terminal capping and the hydrophobic X-group, suggesting amphipathic dominans.

4.2 A refined pharmacophore for antimicrobial activity

The included CAPs in this project, except from RWR-NH_2, fulfill the previous reported minimum motif (two cationic charges and two hydrophobic bulk) for antimicrobial activity, but the reported MIC values (Table 3.4) varies a lot. By comparing the available MIC values with the calculated peptide properties it was found that the ratio between the peptide hydrophilic and the hydrophobic surface accessible surface area seems to play an important role in determining the peptide antimicrobial properties. The total surface accessible volume was also found to correlate strongly with the MIC values. RTbtR-NHBn has MIC values superior to the other peptides, and is also the CAP in this study with both the largest volume and the largest hydrophobic surface area (approximately 71% of the total solvent accessible surface area). RWR-NH_2 is the only peptide that display no antimicrobial activity against *S. aurerus* or *MRSA*, and interestingly it has both the smallest volume and the smallest hydrophobic part of the solvent accessible surface area (48%).

A refined minimum pharmacophore for antimicrobial activity is in light of these results suggested. A minimum motif consisting of cationic charges, a hydrophobic surface accessible area larger than 50% and a solvent accessible volume greater than $\sim$1500 Å^3 seem to be required for these small peptides to display antimicrobial properties. The cationic charges are required for interaction with the anionic parts of the bacterial cell membrane, and thus act as an attracting force between CAP and membrane. A large hydrophobic solvent accessible surface area seem to be important as it is not energetically favorable for large hydrophobic residues to be exposed to the polar environment of the solvent. It is therefore energetically favorable to bury these residues in the hydrophobic membrane. Larger peptide volumes seem to be important as they may have a more disrupting effect on the cell membrane, e.g taking more space. It may also be speculated in the requirement for peptide flexibility. Face flipping might be required for the CAPs to enter and potentially destabilize a cell membrane properly, and this dynamic property could therefore also be included in the refined pharmacophore.

4.3 Tripeptides with chymotryptic stability

Through extensive studies, combining molecular docking and ITC with experimental half life assays and MIC values, it has become clear that the interactions between the tested short antibacterial peptides and the chymotrypsin subsites S1, S1' and S2' can all be used to tune the proteolytic stability while maintaining a high antibacterial activity. For the S1 pocket, the bulk provided by Dip, Bip and Tbt are sufficient for stability which opens for significant freedom in peptide design. It has also been shown that the stability of peptides containing arginine analogs as S1' binders invoke a dramatic increase in the stability when including short cationic side chains. Through residue specific interaction studies between similar peptides with arginine analogs and chymotrypsin, it has been suggested that Coulombic interactions with the main chain carbonyls of Phe41 and Cys58, both part of the S1' pocket, play an important role with respect to peptide stability. CAPs displaying a stronger interaction with Cys58 than with Phe41 was found to be hydrolyzed, whereas those that did no were stable. Bulky C-terminal capping groups, that are essential for activity, act as an additional residue and lowers the stability due to substrate stabilizing interactions with S2'. It was also found that interactions with the S1' pocket were altered as a result of the enhanced S2' interactions in favor of Cys58 leading to instability.

Collectively the S1, S1' and S2' interactions represent important findings that allows for the design of short stable peptides, inhibitors and peptidomimics. In addition to these more readily applicable findings it also provides fundamental insights into the structural determinants of several unnatural amino acid side chains and into the nature of the active site of chymotrypsin.

4.4 Further work & development

The overall goals of this project were to identify tripeptides with antimicrobial properties exhibiting stability towards chymotryptic degradation and to investigate the potential peptide membrane interaction mechanism. The new and detailed insight into the peptide interactions with the S1, S1' and S2' pockets of chymotrypsin should be used for further development of potential CAPs with a proteolytic stability design. Similar studies should also be done on other potential peptide degrading enzymes in order to design a potent drug for oral administration. An experimental structure of the peptide binding mode in the active site of chymotrypsin has still not been obtained. This is of utmost importance for verifying the docking results and in general the reliability of the docking algorithm. The search for the right crystallization conditions should therefore be continued to complete the structural study.

The peptide membrane interactions were studied with molecular dynamics simulations using a POPE membrane as model system. However, in order to simulate a realistic bacterial membrane, more complex model membranes should be used. Future projects should investigate the peptide membrane interactions using model systems resembling the phospholipid composition found in more specific bacteria. Knowing that the membrane composition varies from one organism to another, detailed interaction studies could ultimately be used to fine tune the CAPs towards specific bacteria species. Future computational studies should also be combined with experimental studies, such as NMR, in

order to verify the simulations.

It has also been suggested that the minimum motif for antimicrobial activity consisting of two cationic charges and two hydrophobic bulk residues is not a very accurate definition based on the MIC values included in this project. Future work should include larger libraries of antimicrobial peptides with available MIC values in order to further refine the new pharmacophore suggested from this project. The MIC values for the CAPs vary between different bacteria, and this in turn may ultimately lead to pharmacophores for antimicrobial properties fine tuned towards more specific species.

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not Eureka! (I found it!) but rather, "hmm.... that's funny...."

-Isaac Asimov

References

[1] M. T. Madigan and J. M. Martinko. *Brock biology of microorganisms 11.ed*. Pearson, Prentice Hall, 11 edition, 2006.

[2] `http://www.who.int/drugresistance/en/`, 7.Feb 2009. WHO.

[3] `http://www.who.int/drugresistance/WHO_Global_Strategy_English.pdf`, 8.Feb 2009. WHO global strategy.

[4] R. E Hancock and R. Lehrer. Cationic peptides: a new source of antibiotics. *Trends Biotechnol*, 16:82–88, 1998.

[5] E. Martin, T. Ganz, and R. I. Lehrer. Defensies and other endogenous peptide antibiotics of vertebrates. *J. Leukoc. Biol.*, 58:128–136, 1995.

[6] Z. Wang and G. Wang. Apd: the antimicrobial peptide database. *Nucleic Acids Res.*, 32:D590–D592, 2004.

[7] Antimicrobial peptide database, Aug 2009. `http://aps.unmc.edu/AP/main.php`.

[8] R. E. Hancock and Diamond G. Role of membranes in the activities of antimicrobial cationic peptides. *FEMS Microbiol. Lett.*, 8:402–410, 2000.

[9] H. G. Boman. Peptide antibiotics and their role in innate immunity. *Annu. Rev. Immunol*, 13:61–92, 1995.

[10] R. E. Hancock. Peptide antibiotics. *Lancet*, 349:418–422, 1997.

[11] M. Zasloff. Antimicrobial peptides of multicellular organisms. *Nature*, 415:389–395, 2002.

[12] H. W. Huang. Action of antimicrobial peptides: Two-state model. *Biochemistry*, 39(8347-8352), 2000.

[13] Y. Shai. Mode of action of membrane active antimicrobial peptides. *Biopolymers*, 66:236–248, 2002.

[14] M. N. Melo, R. Ferre, and M. A. R. B Castanho. Animicrobial peptides: linking partition, activity and high membrane-bound concentrations. *Microbio. Nat. Rev.*, 7:245–250, 2009.

[15] K. Matsuzaki. Why and how are peptide-lipide interactions utilized for self-defense? magainins and tachyplesins as archetypes. *Biochem. Biophys. Acta.*, 1462(1-10).

[16] S. Ludtke and H. Huang. Membrane thinning caused by magainin 2. *Biochemistry*, 35:16764–16769, 1995.

[17] R. E. W. Hancock and A. Bell. Antibiotic uptake into gram-negative bacteria. *Eur. J. Clin. Microbio. Infect. Dis.*, 7:713–720, 1998.

[18] Y. Shai. Mechanism of binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochem. Biophys. Acta*, 1462(55-70), 1999.

[19] M. N. Melo and M. A. Castanho. Interaction with bacterial membranes and cell wall models. assigning a biological role to saturation. *Biochim. Biophys. Acta.*, 1768:1277–1290, 2007.

[20] L. Zhang, A. Rozek, and R. E. Hancock. Interaction of cationic antimicrobial peptides with model membranes. *J. Biol. Chem.*, 276:35714–35722, 2001.

[21] A. Giacometti, O. Cirioni, G. Greganati, M. Quarta, and G. Scalise. In vitro activities of membrane-active peptides against gram-positive and gram-negative aerobic bacteria. *Antmicrob. Agebts Chemother.*, 42:3320–3324, 1998.

[22] A. Giacometti et al. In vitro susceptibility tests for cationic peptides: Comparision of broth microdilution methods for bacteria that grow aerobically. *Antimicrob. Agents Chemother.*, 44:1694–1696, 2000.

[23] R. E W. Hancock and D. S Chapple. Peptide antibiotics. *Antimicrob. Agents Chemother*, 43:1317–1323, 1999.

[24] M. H. 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 cytoplasmic membrane of *Escherichia coli*. *Biochemistry*, 38:7235–7242, 199.

[25] H. Leontiadou, A. E. Mark, and S. J. Marrink. Antimicrobial peptides in action. *J. Am. Chem. Soc.*, 128:12156–12261, 2006.

[26] J. C. Hsu and C. M Yip. Molecular dynamics simulations of indolicidin association with model lipid bilayers. *Biophys. J.*, 92:L100–L102, 2007.

[27] D. Sengupta, H. Leontiadou, and A. E Mark. Toroidal pores formed by antimicrobial peptides show signifacant disorder. *Biochem. Biophys. Acta*, 10:2308–2317, 2008.

[28] G. Perron, M. Zasloff, and G. Bell. Experimental evolution of resistance to an antimicrobial peptide. *Proc. Biol. Sci.*, 273:251–256, 2006.

[29] M. R. Yeaman and N. Y. Yount. Mechanism of antimicrobial peptide action and resistance. *Pharmacol. Rev.*, 55(27-55), 2003.

[30] K. Lohner, E. Sevcsik, and G. Pabst. Liposome-based biomembrane mimetic systems: Implications for lipid-peptide interactions. *Advances in Planar Lipid Bilayers and Liposomes*, 6:103–138, 2008.

[31] S.M Hammond, P.A Lambert, and A.N Rycroft. The bacterial cell surface. *Croom Helm, London*, 1984.

[32] R.T Ambron, R.A Picringer, G.B Ansell, J.N Hawthorne, and R.M.C Dawson. Form and function of phospholipids in micro-organisms. *Elsevier, Amsterdam*, pages 289–331, 1973.

[33] S. G. Wilkinson and C. Ratledge. Microbial lipids in gram-negative bacteria. *Academic press, London*, pages 299–488, 1988.

[34] W. M. O'Leary, S.G. Wilkinson, and C. Ratledge. Microbial lipids in gram-positive bacteria. *Academic press, London*, pages 117–201, 1988.

[35] H. Steiner, D. Hultmark, Å. Engstrm, H. Bennich, and H. G. Boman. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature*, 13:61–92, 1981.

[36] J. Svenson, W. Stensen, B. Brandsdal, B. E. Haug, J. Monrad, and J. S. Svendsen. Antimicrobial peptides with stability toward tryptic degradation. *Biochemistry*, 47(12), 3777-3788 2008.

[37] C.G. Wermuth, C.R. Ganellin, P. Lindberg, and L.A. Mitscher. Glossary of terms used in medicinal chemistry. *Pure Appl. Chem.*, 10(5):1129–1143, 1998.

[38] B. Strom, B. E. Haug, M. L. Skar, W. Stensen, T. Stiberg, and J. S Svendsen. The pharacophore of short cationic antibacterial peptides. *Med. Chem.*, 46(9):1567–1570, 2003.

[39] BE. Haug, W. Stensen, T. Stiberg, and JS. Svendsen. *J. Med. Chem.*, 47:4159–4162, 2004.

[40] M. Werle and A. Bernkop-Schnurch. Strategies to improve plasma half life time of peptide and protein drugs. *Amino Acids*, 30:351–367, 2006.

[41] I. Schechte and A. Berger. *Biochem. Biohys. Res. Commun.*, 27(2):157, 1967.

[42] M. Bergmann and W. F. Ross. On proteolytic enzymes. the enzymes of papain and their activation. *J. Biol. Chem.*, 114:717–726, 1936.

[43] Carl Brandon and John Tooze. *Introduction to protein structure*. Gerland Publishing, 2 edition, 1999.

[44] T. Vajda and T. Szabo. *Acta Biochim. Biophys. Hung*, 11(4):287–294, 1976.

[45] L. Hedstrom. *Chem. Rev.*, 102(12):4501–4523, 2002.

[46] H. Czapinska, R. Helland, A.O Smalå s, and J. Otlewski. Crystal structures of five bovine chymotrypsin complexes with p1 bpti variants. *J. Mol. Biol*, (344):1005–1020, 2004.

[47] D. M. Blow. *The Enzymes*, volume 3. Academic press: Boca Raton, 3 edition, 1971.

[48] J. J. Perona and C. S. Craik. *Protein Sci.*, 4:337–360, 1995.

[49] H. Czapinska and J. Otlewski. *Eur. J. Biochem.*, 260:571–595, 1999.

[50] J. R Knowles. *J. Theor. Biol.*, 9:213–228, 1965.

[51] V. N. Dorovskya, S. D. Varfolomeyev, N. F. Kaznskaya, A. A. Klysov, and K. Martinek. *FEBS lett.*, 23:122, 1972.

[52] J. J. Perona and C. S. Craik. *J. Biol. Chem.*, 272(48):29987–29990, 1997.

[53] J. Kraut. *Annu. Rev. Biochem.*, 46:331–358, 1977.

[54] G. P. Hess, J. McConn, E. Ku, and G. McConkey. *Philos. Trans. R. Soc. London*, B 257:89–104, 1970.

[55] M. L. Bender and J. V. Killhefer. *Crit. Rev. Biochem*, 1:149, 1973.

[56] T. A. Steitz and R. G. Shulman. *Annu. Rev. Biophys. Bioeng.*, 11:419–444, 1982.

[57] L. Polgar. Mechanism of protease action. *CRC Press: Boca Raton*, 1989.

[58] L. H. Brinckerhoff, V. V Kalashnikiv, L. W. Thompson, G. V Yamschhikov, R. A Pierce, H.S Galavotti, V. H Engelhard, and C. L. Slinkluff. Terminal modification of an immunogenic mart-1(27-35) peptide: Implications for peptide vaccines. *Int. J. Cancer*, 83:326–334.

[59] L. Aurelio, R. T. C. Brownlee, and A. B Hughes. *Chem. Rev.*, 104(12):5823–5846, 2004.

[60] A. G. Harris. *Gut*, 35:S1–S4, 1994.

[61] G. Osapay, L. Prokai, H. S. Kim, K. F Medzihradszky, D. H. Coy, G. Lipakis, T. Reisine, G. Melacini, Q. Zhu, S. H. H. Wang, R. H. Mattern, and M. Goodman. *J. Med. Chem.*, 40(14):2241–2251, 1997.

[62] A. Rozek, J. P. S. Powers, C. L. Friedrich, and R. E. W. Hancock. *Biochemistry*, 42(48):14130–14138, 2003.

[63] J. J. Nestor, R. Tahilramani, T. L. Ho, G. I. McRae, and B. H. Vickery. *J. Med. Chem.*, 24(9):1170–1174, 1984.

[64] D. Leung, G. Abbenante, and D. P. Fairlie. *J. Med. Chem.*, 43(3):305–341, 2000.

[65] I. Morishita, M. Morishita, K. Takayama, Y. Machida, and T. Nagai. *Int. J. Pharm.*, 78(1):9–16, 1992.

[66] B. E. Haug, M. B Strom, and J. S. M. Svendsen. *Curr. Med. Chem.*, 14(1):1–18, 2007.

[67] J. Gante. Peptidomimics - tailored enzyme-inhibitors. *Angew. Chem. Int. Ed.*, 33:1699–1720, 1994.

[68] R. P. Hicks, J. B. Bhonsle, D. Venugopal, B. W. Koser, and A. J. Magill. De novo design of selective antibiotic peptides by incorporation of unnatural amino acids. *J. Med. Chem.*, 50:3026–3036, 2007.

[69] The protein data bank. `http://www.rcsb.org/pdb/home/home.do`.

[70] G. Rhodes. *Crystallography made crystal clear.* Academic press, 3 edition, 2006.

[71] R. Perozzo, G. Folkers, and L. Scapozza. Thermodynamics of protein-ligand interactions: History, presence, and future aspects. *J. Rec. Sig. Trans.*, 24(1 & 2):1–54, 2004.

[72] W. H. J. Ward and G. A. Holdgate. *Prog. Med. Chem.*, 38:309–376, 2001.

[73] J. E. Ladbury and M. L. Doyle. *Biocalorimetry 2: Applications of calorimetry in the biological sciences.* Wiley, 2004.

[74] M. M. Pierce, C. S. Raman, and B. T Nall. Isothermal titration calorimetry of protein-protein interactions. *Methods*, 19:213–221, 1999.

[75] T. Wiseman, S. Williston, and J.F. Brandts L.N Lin. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, 179:131–137, 1989.

[76] Frank Jensen. *Introduction to computational chemistry.* Wiley, 2 edition, June 2008.

[77] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Ann. der. Physik*, 84:24, 1927.

[78] C. J. Cramer. *Essentials of computational chemistry. Theories and models.* Wiley, 2 edition, Januar 2008.

[79] G. G. Hall. The molecular orbital theory of chemical valency & a method of calculating ionization potentials. *Proc. Royal Soc. London A.*, 205:541–552, 1951.

[80] J-L. Lagrange. Theorie des fonctions acceleration. *J. Comp. Chem.*, 3(4):556–560, 1982.

[81] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, 1965.

[82] P. Hohenber and W. Kohn. *Phys. Rev.*, 136:B864, 1964.

[83] A. D. Becke. Density-functional thermochemistry iii. the role of exact exchange. *J. Chem. Phys*, 98:5648–5652, 1993.

[84] R. A. Leach. *Molecular modelling: Principles and applications.* Pearson, Prentice Hall, 2 edition, 2001.

[85] N. L. Allinger. Conformational analysis. 130. mm2. a hydrocarbon force field utilizing v1 and v2 torsional terms. *J. Am. Chem. Soc.*, 99:8127–8134, 1977.

[86] N. L. Allinger, Yuh Y. H, and J.H Lii. Molecular mechanics. the mm3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, 111:8551–8565, 1989.

[87] W. L. Jorgensen and J. Tirado-Rives. The opls force field for proteins. energy minimmizatios for crystals of cyclic peptides and crambin. *J. Am. Soc.*, 10:1657–1666, 1988.

[88] L. Verlet. *Phys. Rev.*, 159(1):98–103, 1967.

[89] C. W. Gear. *Numerical initial value problems in ordinary differential equations*. Prentice Hall, Englewood Cliffs, 1971.

[90] D. Fincham and D. M. Heyes. Integration algorithms in molecular dynamics. *CCP5 Quarterly*, 6:4–10, 1982.

[91] J-P. 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-alkenes. *J. Comput. Phys*, 23:327–341, 1971.

[92] A. Ajay, W. P. Walters, and M. A. Murcko. Computational methods to predict the binding free energy in liand-receptor complexes. *J. Med. Chem.*, 38:4951–4967, 1995.

[93] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C.R. Corbeil. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *British Journal of Pharmacology*, 153:S7–S26, 2008.

[94] M. F. Sanner. Python: A programming language for software integration and development. *J. Mol. Graphics mod.*, 17:57–61, Feb 1999.

[95] I. D. Kuntz, J. M. Blaney, S. J. Oatly, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 162:269–288, 1982.

[96] G. Jones, P. Willet, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267:727–748, 1997.

[97] New York NY Schrödinger, LLC. Glide version 5.5, 2009.

[98] B. Kramer, M. Rarey, and T. Lengauer. Evaluation of the flexx incremential construction algorithm for protein-ligand docking. *Proteins: Structure, Function and Genetics*, 37:228–241, 1999.

[99] J.R.H Tame. Scoring functions: a view from the bench. *J. Comput. Aided Mol. Des.*, 13:99–108, 1999.

[100] T. Schulz-Gasch and M. Stahl. Scoring functions for protein-ligand interactions: a critical perspective. *Drug discov. Today: Technologies*, 1:231–239, 2004.

[101] V. Mohan, A.C Gibbs, M.D Cummirigs, E.P Jaeger, and R.L DesJarlais. Docking: Successes and challenges. *Curr. Pharm.*, 11:323–333, Des 2005.

[102] J.R.H Tame. Scoring functions - the first 100 years. *J. Comput. Aided Des.*, 13:99–108, 2005.

[103] A.N. Jain. Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.*, 7:407–420, 2006.

[104] M.D. Eldridge, C.W Murray, T.R. Auton, G.V Paolini, and R.P Mee. Empirical scoring functions: The development of fast empirical scoring functions to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.*, 11:425–445, 1997.

[105] X. Friesner, J.L Banks, R.B Murphy, T.A Halgren, J.J Klicic, and D.T Mainz *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1.methods and assessment of docking accuracy. *J. Med. Chem.*, 47:1739–1749, 2004.

[106] R.A Fresner, R.B. Murphy, M.P Repasky, L.L. Frye, J.R Greenwood, and T.A. Halgren *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.*, 49:6177–6196, 2006.

[107] G.M. Morris, D.S. Goodsell, R.S Halliday, R. Huey, W.E. Hart, and R.K. Belew *et al.* Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19:1639–1662, 1998.

[108] R. Najmanovich, J. Kuttner, V. Sobolev, and M. Edelman. Side-chain flexibility in proteins upon ligand binding. *Proteins struct. Funct. Genet.*, 39:261–268, 2000.

[109] C.N Cavasotto, A.J.W. Orry, and R.A Abagyan. The challenge of considering receptor flexibility in ligand docking and virtual screening. *Curr. Comput. Aided Drug Des.*, 1:423–440, Des 2005.

[110] R. Karstad, G. Isaksen, B. Brandsdal, J. S. Svendsen, and J. Svenson. Probing the s1, s1' and s2' binding pockets of chymotrypsin with non-naturally occuring amino acid side chains: Generatoin of short cationic antimicrobial peptides with a high stability towards chymotrypsin degradation. *submitted to J. Med. Chem.*, n.a(n.a):n.a, March 2010.

[111] J. Svenson, R. Karstad, G.E. Flaten, BO. Brandsdal, M. Brandl, and J.S Svendsen. Altered activity and physiochemical properties of short cationic antimicrobial peptides by incorporation of arginine analogues. *Molecular pharmaceutics*, 6(3):996–1005, 2009.

[112] LLC Schrödinger. *The PyMol Molecular graphics system 1.0.* New York, NY, 2 edition, 2009.

[113] Maestro, in., 9.0 ed., schrödinger, llc, new york, ny, 2009.

[114] NY Schrödinger LLC, New York. Macromodel version 9.6, 2008.

[115] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. J. Jorgensen. Opls2005. *J. Phys. Chem. B*, 105(6474), 2001.

[116] E. Polak and G. Ribiere. Revenue francaise informat. *Recherche Operationelle*, 16(39), 1969.

[117] Ligprep. in., 2.3 ed., schrödinger, llc, new york, ny. 2009.

[118] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr. amd T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, J. E. Knox X. Li, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, G. Liu B. B. Stefanov, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, revision e.01. *Gaussian, Inc*, 2004.

[119] New York NY Schrödinger, LLC. Jaguar version 7.5, 2008.

[120] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 2+99.

[121] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. Results obtained with the correlation-energy density functionals of becke and lee, yang and parr. *Chem. Phys. Lett*, 157:200–206, 1989.

[122] S. H. Vosko, L. Wiik, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis. *Can. J. Phys*, 58:1200–1211, 1980.

[123] Yan Zhao and Donald G. Truhlar. Density functionals for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput*, 3:289–300, 2007.

[124] R. Ditchfield, W. J. Hehre, and J. A. Pople. Self-consistent molecular orbital methods. 9. extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys*, 54(724), 1971.

[125] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-consistent molecular orbital methods. 12. further extensions of gaussian-type basis sets for use in molecular-orbital studies of organic-molecules. *J. Chem. Phys.*, 56(2257), 1972.

[126] P. C. Hariharan and J. A. Pople. Influence of polarization functions on molecular-orbital hydrogenation energies. *Theor. Chem. Acc.*, 28:213–222, 1973.

[127] P. C. Hariharan and J. A. Pople. Accuracy of ah equilibrium geometries by single determinant molecular-orbital theory. *Mol. Phys.*, 27:209–214, 1974.

[128] M. S. Gordon. The isomers of silacyclopropane. *Chem. Phys. Lett.*, 76:163–168, 1980.

[129] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, D. J. DeFrees, J. A. Pople, and M. S. Gordon. Self-consistent molecular orbital methods. 23. a polarization-type basis set for 2nd-row elements,. *J. Chem. Phys.*, 77:3654–3665, 1982.

[130] R. C. Binning Jr. and L. A. Curtiss. Compact contracted basis-sets for 3rd-row atoms - ga-kr. *J. Comp. Chem.*, 11:1206–1216, 1990.

[131] J.-P. Blaudeau, M. P. McGrath, L. A. Curtiss, and L. Radom. Extension of gaussian-2 (g2) theory to molecules containing third-row atoms k and ca. *J. Chem. Phys.*, 107:5016–5021, 1997.

[132] V. A. Rassolov, J. A. Pople, M. A. Ratner, and T. L. Windus. 6-31g* basis set for atoms k through zn. *J. Chem. Phys.*, 109:1223–1229, 1998.

[133] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss. 6-31g* basis set for third-row atoms. *J. Comp. Chem.*, 22:976–984, 2001.

[134] K. N. Kudin, G. E. Scuseria, and E. Cancs. A black-box self-consistent field convergence algorithm: One step closer. *J. Chem. Phys.*, 116:8255–8261, 202.

[135] P. Pulay. Improved scf convergence acceleration. *J. Comp. Chem.*, 3:556–560, 1982.

[136] Marelius J., Kolmodin K., Feierberg I., and Åqvist J. Q: a molecular dynamics program for free energy calculations and emperical valence bond simulation in biomolecular systems. *J Mol Graph Model*, 16(4-6):213–225, 261, 1998.

[137] Jorgensen W. L., Maxwell D. S, and Tirado-Rives J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[138] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W Impey, and M.L. Klein. Comparision of of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[139] F.S. Lee and A. Warshel. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 97(5):3100–3107, 1992.

[140] NOTUR. Stallo. `http://www.notur.no/hardware/stallo/`.

[141] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[142] J. W. Ponder and F. M. J. Richards. Truncated newton conjugate gradient. 8(1016), 1987.

[143] KJ. Bowers et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. *ACM press, New York*, 2006.

[144] DE. Shaw. *Desmond molecular dynamics system, Version 2.2.* Schrödinger, New York, NY, 2009.

[145] H.J.C. Berendsen, J.P.M. Postma, W.F van Gunsteren, and J. Hermans. Intermolecular forces. *Reidel, Dordrecht*, page 331, 1981.

[146] W.G. Hoover. Canonical dynamics - equilibrium phase-space distributions. *Phys. Rev.*, 31(3):1695–1697, 1985.

[147] S. Nose. A molecular dynamics method for simulations in the canonical ensamble. *Mol. Phys.*, 52(2):255–268, 1984.

[148] G. J. Martyna, D. J. Tobias, and M. L. Klein. *Chem. Phys.*, 101:4177, 1994.

[149] T. Darden, D. York, and L. Pedersen. Particle mesh ewald - an n.log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[150] M. E. Tuckerman and B. J. Berne. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.*, 94(2):1465–1469, 1991.

[151] A. Moulin, J. H. Bell adn R. F. Pratt, and D. Ringe. Inhibition of chymotrypsin by a complex of ortho-vanadate and benzohydroxamic acid: Structure of the inert complex and its mechanistic interpretation. *Biochemistry*, 46(20):5982–5990, 2007.

[152] W. Kabsch. Automatic processing of rotation diffraction data from crystals of initially unknown symmentry and cell constants. *J. Appl. Cryst.*, 26:795–800, 1993.

[153] Number 4. 1994 Collaborative Computational Project. The ccp4 suite: Programs for protein crystallography. *Acta Cryst.*, D50:760–763, 1994.

[154] A.G.W. Leslie. Recent changes to the mosflm package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, (26), 1992.

[155] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L.C. Storoni, and R.J. Read. Phaser crystallographic software. *J. Appl. Cryst.*, 40:658–674, 2007.

[156] G.N. Murshudov, A.A.Vagin, and E.J.Dodson. Refinment of macromelecular structures by the maximum-likelihood method. *J. Acta. Cryst.*, D53:240–255, 1997.

[157] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallographica*, D60:2126–2132, 2004.

[158] Laskowski R. A., MacArthur M.W., Moss D.S., and Thornton J.M. Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26:283–291, 1992.

[159] Kabsch W. Lsqkab. *Acta. Cryst.*, A32:922–923, 1976.

[160] J. Cerny and P. Hobza. *Phys. Chem. Chem. Phys.*, 7(1624), 2005.

[161] X. Xu and W. A. Goddard III. The x3lyp extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. USA*, 101:2673–2677, 2004.

[162] N. C. Handy and A. J. Cohen. Left-right correlation energy. *Mol. Phys.*, 99:403–412, 2001.

[163] W.-M. Hoe, A. Cohen, and N. C. Handy. Assessment of a new local exchange functional optx. *Chem. Phys. Lett.*, 341:319–328, 2001.

[164] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 69:69, 1951.

[165] J. A. Pople and R. K. Nesbet. Self-consistent orbitals for radicals. *J. Chem. Phys.*, 22:571–72, 1954.

[166] R. McWeeny and G. Dierksen. Self-consistent perturbation theory. 2. extension to open shells. *J. Chem. Phys.*, 49:4852, 1968.

[167] E.J. Baerends, J. Autschbach, A. Brces, F.M. Bickelhaupt, C. Bo, P.M. Boerrigter, L. Cavallo, D.P. Chong, L. Deng, R.M. Dickson, D.E. Ellis, M. van Faassen, L. Fan, T.H. Fischer, C. Fonseca Guerra, S.J.A. van Gisbergen, A.W. Gtz, J.A. Groeneveld, O.V. Gritsenko, M. Grning, F.E. Harris, P. van den Hoek, C.R. Jacob, H. Jacobsen, L. Jensen, G. van Kessel, F. Kootstra, M.V. Krykunov, E. van Lenthe, D.A. McCormack, A. Michalak J. Neugebauer, V.P. Nicu, V.P. Osinga, S. Patchkovskii, P.H.T. Philipsen, D. Post, C.C. Pye, W. Ravenek, J.I. Rodrguez, P. Ros, P.R.T. Schipper, G. Schreckenbach, J.G. Snijders, M. Sol, M. Swart, D. Swerhone, G. te Velde, P. Vernooijs, L. Versluis, L. Visscher, O. Visser, F. Wang, T.A. Wesolowski, E.M. van Wezenbeek, G. Wiesenekker, S.K. Wolff, T.K. Woo, A.L. Yakovlev, and T. Ziegler. Adf2008.01, scm, theoretical chemistry, vrije universiteit, amsterdam, the netherlands, 2008. `http://www.scm.com`.

[168] J. Svenson, R. Karstad, G. E Flaten, B. Brandsdal, M. Brandl, and Svendsen J. S. Altered activity and physiochemical properties of short cationic antimicrobial peptides by incorporation of arginine analogues. *Molecular Pharmaceutics*, 6(3):996–1005, 2009.

[169] B. E Haug, W. Stensen, M. Kalaji, Ø . Rekdal, and J. S Svendsen. Synthetic antimicrobial peptidomometics with therapeutic potential. *J. Med. Chem.*, 51:4306–4314, 2008.

[170] LLC Schrödinger. *QikProp version 3.1.* New York, 2009.

[171] V. H. L. Lee. *Crit. Rev. Ther. Drug carrier Syst.*, 5(2):69–97, 1982.

[172] T. Kurth, D. Ullman, H. D. Jakubke, and L. Hedstrom. *Biochemistry*, 36(33):10098–10104, 1997.

[173] A. J. Scheidig, T. R. Hynes, L. A. Pelletier, J. A. Wells, and A. A. Kassiakoff. *Prot. Sci.*, 6(9):2794–2799, 1997.

To be continued...