Investigation of the dynamic nature of proteins with enhanced sampling techniques

Zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

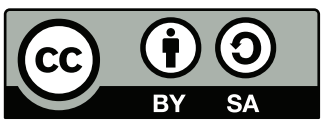
von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Marina Putzu

Referent:	Prof. Dr. Marcus Elstner
Korreferentin:	Prof. Dr. Anne Ulrich
Tag der mündlichen Prüfung:	19. April 2018



Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet, sowie die Regeln zur Sicherung guter wissenschaftlicher Praxis am Karlsruher Institut für Technologie (KIT) beachtet habe.

Marina Putzu
30.06.2018

Zusammenfassung

In dieser Arbeit werden zwei Aspekte der dynamischen Natur von Proteinen mit Hilfe von Enhanced Sampling Methoden untersucht.

Zunächst wurden Mikrosekunden-Molekulardynamiksimulationen von Harzianin HK VI (HZ) in Wechselwirkung mit einer Dimyristoylphosphatidylcholin-Doppelschicht unter der Bedingung eines niedrigen Peptid-zu-Lipid Verhältnisses durchgeführt. Zwei Orientierungen des HZ-Moleküls in der Doppelschicht wurden gefunden und charakterisiert. In der Ausrichtung senkrecht zur Doppelschicht-Oberfläche induziert HZ eine lokale Ausdünnung der Doppelschicht. Wird HZ parallel zu seiner Oberfläche in die Doppelschicht eingesetzt, befindet es sich nahezu vollständig im hydrophoben Bereich der Doppelschicht. Eine ausgedehnte Sampling lieferte qualitative Ergebnisse und zeigte, dass die letztgenannte Orientierung ein globales Minimum der freien Energie ist. Die sekundäre Struktur von HZ wurde charakterisiert, und es wurde festgestellt, dass sie sich in der 3_{10}-helikalen Familie befindet.

Zweitens wurde die Thiol-Disulfid-Austauschreaktion in Modellsystemen und kleinen Peptiden mit Hilfe eines kombinierten QM/MM-Metadynamikschemas untersucht. Die freien Energielandschaften dieser Systeme wurden generiert, und liefern die Strukturen von Reaktanden und Produkten mit atomaren Details, sowie die Höhen der freien Energiebarrieren (oder Aktivierungsenergien), die dem spontanen Austausch entgegenwirken. Ein QM/MM-Schema mit rein klassischem Wasser erwies sich als effiziente und präzise Kompromisslösung. Die Berechnungen ergaben den erwarteten symmetrischen Trisulfid-Übergangszustand und seine Struktur und Energiebarriere waren für die intramolekularen Thiol-Disulfid-Reaktionen in Modellpeptiden sehr ähnlich. Während CXC-Disulfidbindungen als sterisch ungünstig eingestuft wurden, wurde CXXC gegenüber längerfristigen Disulfidbindungen entlang des Peptidrückgrats bevorzugt, was der hohen Häufigkeit von CXXC-Motiven in Redox-Proteinen entspricht. Eine direkte Anwendung auf ein reales Protein wurde ebenfalls in Form von Force-Clamp Simulationen durchgeführt. Das Protokoll ermöglichte die Simulation der Disulfidbindungsisomerisierung in einem einzelnen Protein und reproduzierte die in einem AFM-Experiment beobachtete Regioselektivität.

Abstract

In this work two aspects of the dynamic nature of proteins are investigated using enhanced sampling techniques.

Firstly, microsecond molecular dynamics simulations of harzianin HK VI (HZ) interacting with a dimyristoylphosphatidylcholine bilayer were performed at the condition of low peptide-to-lipid ratio. Two orientations of HZ molecule in the bilayer were found and characterized. In the orientation perpendicular to the bilayer surface, HZ induces a local thinning of the bilayer. When inserted into the bilayer parallel to its surface, HZ is located nearly completely within the hydrophobic region of the bilayer. An extended sampling simulation provided qualitative results and showed the latter orientation to be a global minimum of free energy. The secondary structure of HZ was characterized, and it was found to be located in the 3_{10}-helical family.

Secondly, the thiol–disulfide exchange reaction in model systems and small peptides was investigated by means of a combined QM/MM metadynamics scheme. The free energy landscapes of these systems were generated, providing the structures of reactants and products with atomic detail, as well as the heights of free energy barriers (or, activation energies) opposing the spontaneous exchange. A QM/MM scheme with purely classical water turned out to be an efficient and accurate compromise solution. The calculations yielded the expected symmetric trisulfide transition state, and its structure and energy barrier were very similar for the intramolecular thiol–disulfide reactions in model peptides. While CXC disulfide bonds were found sterically unfavorable, CXXC were favored over longer-range disulfide bonds along the peptide backbone, in line with the high abundance of CXXC motifs in redox proteins. A direct application in a real protein was also performed through force-clamp simulations. The protocol allowed for simulation of the disulfide bond isomerization in a single protein, reproducing the regioselectivity observed in an AFM experiment.

Table of contents

Abbreviations

AA	amino acid
AFM	atomic force microscopy
Aib	α-aminoisobutyric acid
ALM	Alamethicin
AMP	antimicrobial peptides
CG	coarse-grained
COM	center of mass
CVs	Collective variables
DFT	density functional theory
DFTB	density functional tight binding
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
FES	free energy surface
Grx	glutaredoxin
H-REMD/HREX	hamiltonian-replica exchange molecular dynamics
HF	Hartree-Fock
HZ	Harzianin HK VI
LCAO	linear combination of atomic orbitals
Leuol	leucinol

MAP	membrane active peptide
MD	molecular dynamics
MM	molecular mechanics
MO	molecular orbital
NMR	nuclear magnetic resonance
PDI	protein disulfide isomerase
PME	particle–mesh Ewald
PMF	potential of mean force
POPC	1-palmitoyl-2-oleoyl-phosphatidylcholine
QM	quantum mechanics
REMD	replica exchange molecular dynamics
RESP	restrained electrostatic potential
REST	replica exchange solute tempering
SCC-DFTB	self-consistent-charge density functional tight binding
SCF	Self-consistent-field
SRCD	synchrotron radiation circular dichroism
ssNMR	solid state nuclear magnetic resonance
T-REMD	temperature-replica exchange molecular dynamics
Trx	thioredoxin
Valol	valinol

Chapter 1

Introduction

The function of a protein is determined by its native conformation, which is strictly related to the amino acid sequence, as pointed out in the classic work of Anfinsen [1]. The folded shape is stabilized by the combined strength of large numbers of noncovalent interactions, i.e. hydrogen bonds, salt bridges, and hydrophobic contacts, and by covalent intra- and inter-molecular disulfide bridges between a pairs cysteine residues.

Conformational transitions are the hallmarks of protein dynamics and are often intimately related to its functions. The weak noncovalent bonds are easily broken and reformed in response to various external stimuli. An example is the behaviour of many membrane active peptides. They are usually completely unfolded in solution but change conformations when in contact with lipid membranes.

Recently, it was shown that also disulfide bonds may have a chemically labile dynamic nature. They do not only stabilize protein structure statically, but also rearrange by spontaneous disulfide shuffling.

This two aspects of the dynamic nature of proteins are investigated in this way: In chapter 3, the study of interactions of an antimicrobial peptide with a bilayer membrane and the changes in structure and orientation is presented. In chapter 4, the mechanism of disulfide bond shuffling in proteins is studied, analyzing the free energy profiles of the reaction in some model systems. Finally, an application in protein is presented in chapter 5. In the following, a background description of the two process is given.

1.1 Antimicrobial Peptides

The rapid increase in drug-resistant infections has presented a serious challenge to antimicrobial therapies. The development of new families of antibiotics that can overcome the resistance problem has become a very important task . The mechanism of action of antimicro-

bial peptides (AMPs) is so different from that of the conventional antibiotics that AMPs are considered attractive substitutes or additional drugs. Rather than targeting a specific site on a protein or ribosome, they act directly on the bacterial cell membrane [2]. It is therefore less likely for bacteria to develop resistances against AMPs. The elucidation of the mechanisms of AMP action and their specific membrane damaging properties is supposed to be a key for the rational design of novel types of peptide antibiotics [3].

Hundreds of AMPs have been isolated so far. They are produced by all organisms, be it unicellular microorganisms, plants, insects or mammals. Secretions from frog skin, for example, contain more than 300 different AMPs [4]. AMPs range from 10 to 50 amino acids [5], and are generally cationic, which allows them to bind to the negatively charged phospholipids of bacterial membranes by electrostatic attraction [6]. Generally, helical AMPs are unstructured in solution and fold once bound to the membrane.

Several different models have been proposed to describe the mode of membrane permeabilization by AMPs. However, most AMPs have been found to act according to either one of the so called “barrel-stave” and “carpet” models (figure 1.1)[7–11].

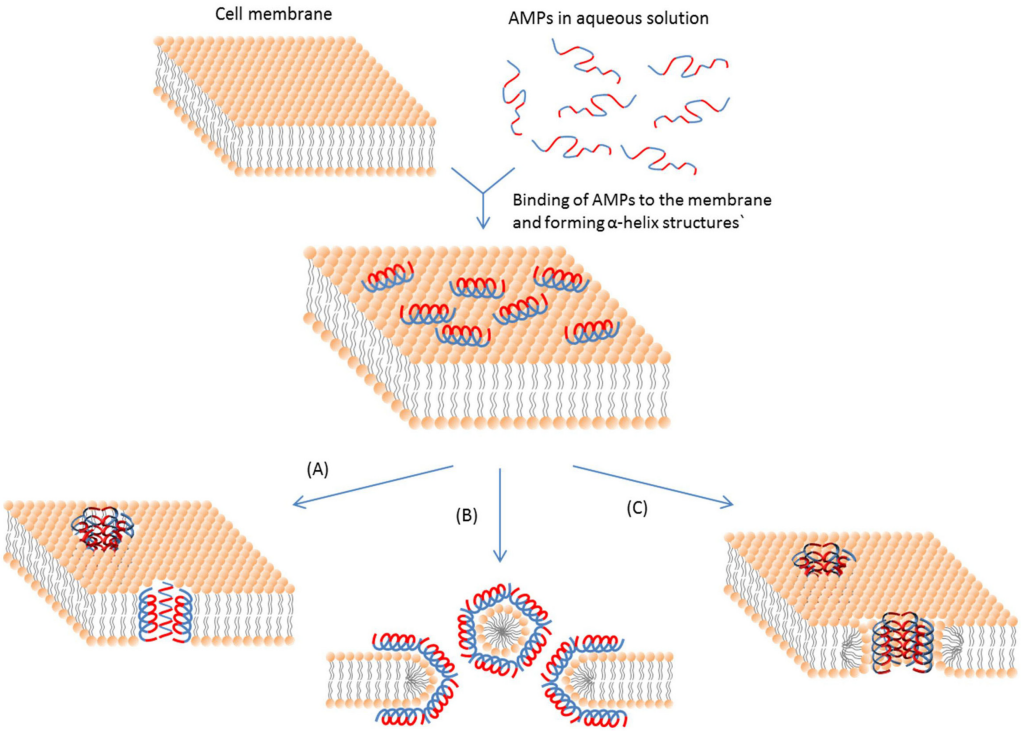


Fig. 1.1 Schematic representation of some action mechanisms of membrane-active AMPs.(A) Barrel-Stave model. (B) Carpet model. (C) Toroidal pore model. This model resembles the Barrel-stave model, but AMPs are always in contact with phospholipid head groups of the membrane. Figure from [2].

In the barrel-stave model, peptide molecules initially bind on the membrane surface, parallel to it. When the concentration of the membrane-bound peptide increases, reaching a certain threshold, peptides insert into the bilayer, forming transmembrane pores. These pores consist of aggregated helices, in which the molecules interact laterally, and are arranged like the staves of a barrel. These structures are favored by the amphiphilic character of the peptide helices: in the aggregate, the hydrophilic side of the helix points toward the aqueous lumen of the pore, while the hydrophobic side interacts with the hydrophobic core of the membrane. The overall structure of the cylindrical, water-filled pore, resembles that of protein ion channels (figure 1.1). In this mechanism, the formation of the pore is a cooperative process: after the insertion of one peptide molecule into the bilayer, the uptake of other monomers to form the aggregates is induced. Since the barrel-stave mechanism foresees peptide insertion in the hydrophobic core of the bilayer, and peptide aggregation, AMPs acting according to this model are usually highly hydrophobic and electrically neutral. One of the most studied pore-forming peptides acting according to this mechanism is alamethicin [12].

The carpet mechanism is typical for cationic peptides in anionic bilayers: in this case, peptides interact strongly with the phospholipid polar headgroups by electrostatic attraction, covering the membrane surface like a carpet. This causes a perturbation to the surface tension of the bilayer, leads to the formation of defects and eventually to membrane micellization. In contrast to the barrel-stave mechanism, the peptide interacts with the phospholipid headgroups throughout the entire process. In specific cases, a very specific structure was proposed for the peptide-generated membrane defects, called “toroidal pores” [13, 14]: in these structures, characterized by a toroidal shape, the lipid bilayer bends back onto itself and the two leaflets connect.

A physico-chemical interpretation of the mechanism of pore formation by AMPs, which can explain the driving forces for both the barrel-stave and the carpet models, is the two-state mechanism, proposed by Huang [15]. It suggests that, for low peptide/lipid (P/L) ratios, peptides bind to the membrane lying parallel to its surface. Above a certain P/L threshold, which depends both on the specific peptide and on the lipid composition of the bilayer, peptides attain a transmembrane helical orientation. The surface-parallel orientation is called *S-state*, while the transmembrane, inserted orientation is called *I-state*. The I-state is associated with the formation of both toroidal pores or barrel-stave pores, while the S-state is considered “functionally inactive”. When peptides bind to the membrane surface, they cause a stretching of the membrane, and this distortion has a cost in term of energy. Thus, the free energy of adsorption on the membrane surface is composed by a negative term of binding energy and a positive term of bilayer deformation energy. The bilayer deformation is proportional to P/L and the energy cost to $(P/L)^2$, according to the elasticity theory. Thus, the

energy of the S-state is lower than that of the I-state, but at sufficiently high concentrations the energy for S-state exceeds the energy of the I-state, and the S-to-I transition occurs [15].

These models can provide an explanation to the bactericidal activity of some AMPs, but they cannot be easily applied to all of them. For many AMPs the mechanism of pore formation has not been established yet. The short peptides of the peptaibol family Fall into this category.

1.1.1 Peptaibols

Peptaibols are a special class of short AMPs with a length from seven to twenty amino acids (AAs). They are secondary metabolites in soil filamentous fungi, including *Trichoderma* and related genera such as *Emericelopsis* and *Gliocladium* [16], and exhibit pronounced antibiotic, antifungal, and anticancer activities, which render them as potentially interesting leads for pharmacological and agricultural applications. Unlike the cationic AMPs, their content of charged residues is very low or even absent altogether.

Peptaibols are characterized by the presence of a non-proteinogenic amino acid, α-aminoisobutyric acid (Aib) (figure 1.2), an N-terminal-acetylated AA and a 1,2-amino alcohol, i.e., leucinol (Leuol) or valinol (Valol), which is linked by an amide bond at their C-terminus.[12] Peptaibols, with their unusual amino acid content, are the product of non-ribosomal biosynthesis.[17] The original name 'peptaibol' has been constructed from the names of the three characteristic components: **pep**tide, **Aib**, and amino alcoh**ol**. A peptaibol

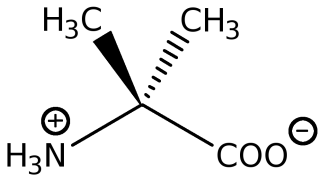


Fig. 1.2 α-aminoisobutyric acid (Aib)

online database has been compiled as a freely accessible public resource at peptaibiotics-database.boku.ac.at [18], which contains the sequence and structure information for more than 1300 peptaibols.

The backbone of a peptaibol molecule generally forms a helical structure due to the conformational constraints imposed by the presence of abundant Aib. [19] Furthermore, the helix-stabilizing effect of Aib in a sequence even overrides the effects of helix-breaking AAs, such as proline.[20]

Due to their amphipathic nature, the biological activity of peptaibols is principally attributed to the formation of voltage-dependent ion channels in lipid bilayer membranes.

Peptaibols are classified into:

(i) long-sequence peptaibols (18–20 AA, typically with a centrally located Pro, and Gln residues near both termini);

(ii) short-sequence peptaibols (11–16 AA, with several Aib-Pro motives, and typically either Ac-Aib-Asn- or Ac-Aib-Gln- as N-terminus);

(iii) ultrashort lipopeptaibols (7 or 11 AA, with a high content in Gly and N-terminal AA acylated by a C_8–C_{15} fatty acid).

The archetypal representative of long-sequence peptaibols are alamethicins (ALMs), which contain 20 AAs. The most studied variant is ALM F30/3 (Ac-Aib^1-Pro-Aib-Ala-Aib-Ala-Gln^7-Aib-Val-Aib-Gly-Leu-Aib-Pro^{14}-Val-Aib-Aib-Glu-Gln^{19}-Pheol), whose cytolytic activity was recognized already in the 1960s.[21] ALM adopts a predominantly α-helical conformation, including a few 3_{10}-turns, with a small bend generated by Pro14. The overall conformation of ALM results in an amphipathic helix with the C=O groups of Aib10 and Gly11 as well as the side-chains of Gln7 and Glu18 being located on the same face of the helical structure.

For this peptide it has been conclusively demonstrated through a combination of a large set of biophysical techniques that it forms pores.[22, 23] ALM may either bind to the surface of a lipid bilayer, or/and it may penetrate into the interior of the hydrophobic core; as soon as inserted, it may form oligomers and, thus, create barrel-stave or helix-bundle channels, with the more hydrophilic side of the helices facing the channel lumen.[22, 23] The channels may be filled with water and conduct ions consequently.

The ALM helix has a length corresponding almost exactly to the thickness of biological membranes (fig. 1.3). However, it is also one of the longest members of the peptaibols family. Considering the closely related amino acid composition and physico-chemical properties of all members of the peptaibols family, it is conceivable that they could form pores in a similar way. However, a minimal peptide length of 18–20 residues is usually required for the formation of transmembrane helix monomers. Thus, in the case of short-sequence peptaibols, this model may not be relevant. Short sequence peptaibols were also reported to form pores and channels in planar lipid bilayer membranes.[24] However, most of the studies of short-sequence peptaibols have focused on structural features in solution, and not so much is known as for the behavior of these peptaibols inside a membrane.

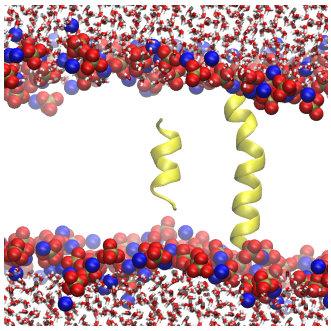


Fig. 1.3 Comparison of the length of the Alm (20 AA) and Harzianin HK VI (11 AA) helices with the thickness of a DOPC bilayer.

Name	Sequence												Source
Harzianin HK VI	Ac	Aib	Asn	Leu	Leu	Aib	Pro	Ile	Ile	Aib	Pro	Leuol	T. pseudok.
Trichobrachin A-I[26]	Ac	Aib	Asn	Lxx	Lxx	Aib	Pro	Lxx	Aib	Aib	Pro	Leuol	T. longibrach.
Trichobrachin A-II[19]	Ac	Aib	Asn	Lxx	Lxx	Aib	Pro	Val	Lxx	Aib	Pro	Leuol	T. longibrach.
Trichorovin TV Ia[27]	Ac	Aib	Asn	Val	Lxx	Aib	Pro	Lxx	Lxx	Aib	Pro	Valol	T. viride
Trichorovin TV IIb[27]	Ac	Aib	Asn	Lxx	Val	Aib	Pro	Lxx	Lxx	Aib	Pro	Valol	T. viride
Trichorozin I[28]	Ac	Aib	Asn	Ile	Leu	Aib	Pro	Ile	Leu	Aib	Pro	Valol	T. harzianum

Table 1.1 Some 11-residue peptaibols HZ-analogue.

1.1.2 Harzianin HK VI

The subject of this work is the short-sequence peptaibol harzianin HK VI (hereafter called HZ). HZ is produced naturally by various terrestrial and marine fungi of the *Trichoderma sp.*, and was first isolated from *T. pseudokoningii* and characterized by Bodo et al. in 1996.[25] Within its 11-residue sequence, Ac-Aib^{1}-Asn-Ile-Ile-Aib^{5}-Pro-Leu-Leu-Aib^{9}-Pro-Leuol, HZ contains two Aib-Pro pairs at positions 5-6 and 9-10, the Ac-Aib-Asn run at the N-terminus and several bulky hydrophobic Leu/Ile residues. Notably, due to mass spectrometry being the prime sequencing method, inability to distinguish Leu from Ile leads to the ambiguity in identification of HZ analogues, e.g. when Lxx is reported as an aliphatic residue.

Functional studies showed HZ, its homologues (Ac-Aib-Asn-Lxx-Lxx-Aib-Pro-Lxx-Lxx-Aib-Pro-Lxxol) or mixtures containing HZ to exert embryotoxicity against *Crassostrea gigas*[29] as well as chlorpromazine level neuroleptic effects in male mice.[30] Crude extracts from HZ-producing fungi inhibited motility of the boar spermatozoa and quenched the mitochondrial transmembrane potential of the sperm cells at low exposure concentrations.[30, 31] HZ is a membrane-active peptaibol as it was reported to cause leakage of the carboxyfluorescein-loaded lecithin/cholesterol liposomes and, as other short-sequence peptaibols, forms pores and channels on planar lipid bilayer membranes.[25, 24, 31] HZ

can act as a cell-penetrating peptide – uptake of covalently conjugated oligonucleotides was shown.[32, 33]

Notably, there are no particularly strong in vitro cytotoxic effects demonstrated specifically for HZ or any other short-sequence peptaibols. Taken together with the fact that peptaibols are produced in mixtures as natural combinatorial libraries, this finding challenges putative exclusive membranolytic cytotoxicity as a prime mechanism of biological action and the main purpose of their biosynthetic production. Particularly interesting as alternatives are recently suggested synergism between hydrolytic enzymes and membrane affecting peptaibols indirectly leading to concerted fungicidal effects.[34] Furthermore, a recent work proposes the synergistic enhancement of the membrane-permeabilizing activity of canonical 20-mer peptaibols (like ALM) by the 11-mer ones, i.e., by the short-sequence peptaibols like HZ. [31] Should any of these alternatives or the direct membranolytic action be the prime mechanism of HZ biological action, determination of HZ structure in the membrane-bound state is of key relevance.

From the general structure perspective, all harzianins (and their sequential analogues table 1.1) including HZ were suggested to assume a β-bend ribbon spiral-like secondary structure. The β-bend ribbon spiral has been described in the case of peptide regions consisting of a succession of Aib–Pro motifs, such as the $(\text{Aib-Pro})_n$ peptide [35]. The presence of the Pro residues, which disrupt the conventional H-bonding schemes observed in helices (lacking the usual N–H donor group), is compensated by helix-forming residues such as Aib. This structure is characterized by two sets of (ϕ,ψ) angles: $\phi_i = -54°$, $\psi_i = -40°$; $\phi_i = 78°$, $\psi_i = 10°$ associated with the Aib–Pro motif.

The α- and 3_{10}-helices are characterized by their intramolecular hydrogen bonds, and by the values of backbone torsion angles. Since the ϕ/ψ for α- and 3_{10}-helices partially overlap in the Ramachandran space, their intramolecular H-bonding patterns which differ from each other, are a more effective criterion to distinguish between the two. For the α-helix, the H-bonds are formed between the amides of the $(i+4)$th residues as donors and the carbonyls of the ith residues as acceptors (type $1 \leftarrow 5$ H-bond). For the 3_{10}-helix, the NH groups of the $(i+3)$th residues are the donors and the CO groups of the ith residues are the acceptors (type $1 \leftarrow 4$ H-bond).

The β-bend ribbon spiral is a subtype of the 3_{10}-helix because it is also stabilized by the $1 \leftarrow 4$ type H-bonds. However, HZ as well as its analogue and other (longer) peptaibols are not exclusively composed of alternating Aib and Pro residues, rather of Aib–Pro motif alternating with dipeptides, forming the (Xaa–Yaa–Aib–Pro) tetrapeptide subunit. A particular type of β-bend ribbon spiral was determined for the 14-mer harzianin HC IX, which specifically occurs in the case of $(\text{Xaa–Yaa–Aib–Pro})_n$ peptides. This structure was

Residue	ϕ	ψ
Xaa(i)	$(-90\pm8)^\circ$	$(-27\pm18)^\circ$
Yaa($i+1$)	$(-98\pm11)^\circ$	$(-17\pm14)^\circ$
Aib($i+2$)	$(-49\pm3)^\circ$	$(-50\pm2)^\circ$
Pro($i+3$)	$(-78\pm1)^\circ$	$(+3\pm19)^\circ$

Table 1.2 Backbone dihedral angles of the fragment forming a (Xaa–Yaa–Aib–Pro)–β-bend ribbon spiral structure.[25]

termed (Xaa–Yaa–Aib–Pro)–β–bend ribbon spiral and it is described by the four sets of (ϕ,ψ) dihedral angles given in Tab. 1.2. [25] It appears to be intermediate between the 3_{10}–helix and the β–bend ribbon spiral. It mainly differs from both of them by the number of stabilizing $1 \leftarrow 4$ hydrogen bonds for a given number of amino acids. This is related to the number and location of the prolines in the peptide sequence, as this residue interrupts the hydrogen-bonding network.

An α-helical conformation appears not to be relevant for HZ, as it seems to prevail only for Aib-poor long-sequence peptaibols like ALMs.

1.2 Disulfide-bond shuffling

Disulfide bonds are important building blocks of the secondary and tertiary structures of proteins, serving as inter- and intra-subunit cross links. A disulfide is formed by the oxidation of the sulfhydryl groups of two cysteine residues. Previously, these bonds have been thought to serve two functions. First, they influence the thermodynamics of protein folding: disulfide bonds stabilize the native conformation of a protein by destabilizing the unfolded form [36, 37]; they lower the entropy of the unfolded form, making it less favorable compared with the folded form. However, disulfide bonds can decrease the stability of folded proteins by restricting energetically favorable conformational changes. Second, they maintain protein integrity: oxidants and proteolytic enzymes in the extracellular environment can inactivate proteins. By stabilizing protein structure, disulfide bonds can protect proteins from damage and increase their half-life.

Previously, the disulfide bonds present in mature proteins were thought to be inert, that is, once formed they remain unchanged for the life of the protein. However, it now appears that this is not necessarily the case; in contrast to other covalent bonds, disulfides can be highly dynamic in physiological conditions as a consequence of a set of reactions known as thiol–disulfide exchanges. [38]

Reactions that result in thiol–disulfide exchange have pivotal roles in biology. During protein folding disulfide crosslinks are formed de novo by external proteins or small molecules. This disulfide arrangement is not necessarily stable, and the folding intermediate undergoes disulfide bond reshuffling to reach the stable native fold with correct disulfide crosslinks. [39–41] For a long time these reactions were thought to only have a protein stabilizing structural purpose. In fact, thiol–disulfide exchange reactions are central to oxidative protein folding and key to the mechanism of almost all enzymes that generate and isomerize disulfide bonds [42, 38, 43, 38], such as protein disulfide isomerase (PDI).

However, intramolecular disulfide shuffling can also occur upon an external stimulus and thereby can change protein conformation and possibly function. [44, 45] For example, α–keratin, a structural protein abundant in hair and nails, reshuffles the interchain disulfide bonds upon mechanical stress (e.g. curling your hair), leading to protein conformational changes. [46–48] A similar example of intramolecular shuffling occurs in the von Willebrand factor protein. This enormous multimer is critically involved in primary hemostasis and is particularly rich in cysteines and disulfide bonds. Current data points toward disulfide shuffling triggered by shearing of the protein. [46, 49, 50]

This suggests that disulfide bonds can readily react with thiols within proteins at physiologically relevant time scales even in the absence of enzymes. Given the multitude of

possible thiol-disulfide exchange reactions, an important but hitherto unresolved question is how specificity is achieved.

1.2.1 Thiol-Disulfide exchange reaction

Thiol–disulfide interchange is an S_N2 reaction. A thiolate anion (RS^-) is a nucleophile that attacks one of the sulfur atoms of the disulfide to produce a new disulfide bond while a new free thiolate is released.[38, 51] Theoretical and experimental evidence suggest a linear trisulfide-like transition state,[52] with the negative charge being delocalized over all of the three sulfur atoms and mostly abundant on the terminal sulfur atoms, see Fig. 1.4.

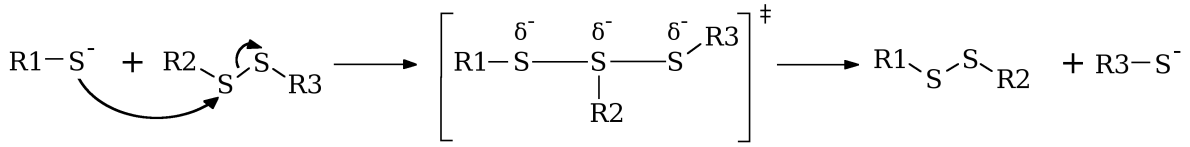


Fig. 1.4 Thiol–disulfide exchange: a thiolate anion (R–S^-) is a potent nucleophile which attacks a disulfide bond, displacing one sulfur atom and forming a new bond with the other sulfur atom (nucleophilic substitution).

The noncatalyzed reaction is slow (in the order of $k = 10^{-1}$–10 $M^{-1}s^{-1}$ at pH 7) [53–56], but oxidoreductase enzymes [such as thioredoxin (Trx), glutaredoxin (Grx), protein disulfide isomerase (PDI)] are designed to accelerate the process up to $k = 10^4$-10^6 $M^{-1}s^{-1}$. A general feature of these enzymes is that they contain a Trx fold domain with a redox active CXXC motif , which was shown to be essential for their oxidoreductase activities,[57]. They catalyze not only the oxidation of thiols to disulfides and the backward reaction but also disulfide isomerization.

The reactivity of the sulfhydryl group is dominated by that of its deprotonated form. The protonated forms of simple alkyl thiols are practically unreactive as nucleophiles under normal conditions, reacting some 10^{10}–fold slower than their corresponding thiolates [58, 43]. As the concentration of the thiolate is derived from the Henderson-Hasselbalch equation, the pH dependency of the reaction rate for thiol-disulfide exchanges is [59]:

$$k_{obs} = \frac{k}{1 + 10^{pK_a - pH}} \tag{1.1}$$

Here k_{obs} is the observed rate constant at a given pH, and k is the corresponding limiting rate constant for the thiolate at high pH. Thus, k_{obs} is one half of the limiting rate constant at the pK_a, but falls to $1/10^4$ of the maximal reactivity at 4 pH units below the pK_a.

Biological thiols show a very wide range of pK_a values from about 3 to 11, thus corresponding to an 8-order of magnitude shift in the deprotonation equilibrium [60]. The factors contributing to this profound modulation of thiol pK_a are under active investigation, and include solvation, electrostatic effects with neighboring charges and dipoles, as well as H-bonding interactions. [51, 61, 62] It is important to note that the pK_a of thiols has two distinct effects on reactivity. Obviously, as noted above, a lower pK_a increases the fraction of thiol in its reactive thiolate form, however, a thiol with a high pK_a is expected to be a strong nucleophile, whereas one with a low pK_a is expected to be a weak nucleophile, thus less reactive. From the balance between these two factors it emerges that the most reactive thiols are those with a pKa close to the pH of the solution.[63, 64] Some enzymes have evolved to have a thiol pK_a far below the pH of a typical cellular environment. Such low pK_a values might, however, suppress oxidative side reactions that would otherwise compromise catalysis. Another reason is that marked differences in acidity allow the equilibrium constant for thiol disulfide exchange to be tuned by thiol pK_a values. Thus, in the thiol-disulfide exchange reaction:

$$\mathrm{R1-SH} \quad + \quad \mathrm{R2-S-S-R3} \quad \rightleftharpoons \quad \mathrm{R1-S-S-R2} \quad + \quad \mathrm{R3-SH} \tag{1.2}$$

lowering the pK_a of R3-SH with respect to R1-SH will improve the leaving-group properties of R3-SH and bias the equilibrium to the right. [51, 65]

An important aspect of thiol–disulfide reactions is the transition state geometry. The nucleophilic attack by the thiolate is driven along the disulfide axis and this establishes the orientation necessary for interactions between well-structured redox partners. [66, 67, 65, 68]. For this reason disulfide–exchange reactions have significant steric requirements that must be accommodated by the enzymes catalyzing oxidative protein folding. Additionally, these stringent requirements influence the rates by which disulfides can be reduced and rearranged nonenzymatically. The transition state structure is approximately symmetrical relative to the attacked sulfur atom, [68–70] with an S–S–S angle that ranges from 180°[70] to 160° [71].

An experimental study of a disulfide exchange reaction between glutathione and dithiothreitol in an aqueous environment yielded a height of energy barrier of ca. 14 $\mathrm{kcal \cdot mol^{-1}}$.[72] In gaseous phase the energy barrier of this reaction is inverted, thus the potential energy surface has a parabolic shape with only one minimum, which belongs to the trisulfide anion (transition state in aqueous environment). Since the transition structure has a charge density more largely delocalized than a free thiolate, the energy barrier is lower (or, equivalently, the thiolate is more weakly stabilized) in hydrophobic environments. The effect can be so dramatic that the disulfide exchange taking place in the active site of the ribonucleotide reductase enzyme is estimated to run 1000× faster than the same reaction in water.[68]

Chapter 2

Computational methods

The first dynamical simulation of a protein based on a detailed atomic model was reported in 1977 [73]. Since then, the uses of various theoretical and computational approaches have contributed tremendously to our understanding of complex biomolecular systems such as proteins, nucleic acids, and bilayer membranes. By providing detailed information that is often experimentally inaccessible, computational approaches based on detailed atomic models can help in the current efforts to understand the relationship of the structure of biomolecules to their function. For that reason, they are now considered to be an integrated and essential component of research in modern chemistry, biology, biochemistry, and biophysics.

In this chapter will be presented an overview of the computational methods used in the present work and the theoretical background that support them.

2.1 Molecular Dynamics

Molecular dynamics (MD) simulation provides the methodology for detailed microscopical modeling on the atomic scale. This technique is able to monitor time-dependent processes in molecular systems in order to study their structural, dynamic, and thermodynamic properties by numerically solving the equation of motion.

The time evolution of a set of interacting particles is followed via the solution of Newton's second law

$$\mathbf{F}_i = \frac{\mathrm{d}\mathbf{p}_i}{\mathrm{d}t} \tag{2.1}$$

where $\mathbf{p}_i(t) = m_i\mathbf{v}_i$ and $\mathbf{F}_i$ are the momenta and the force of ith particle at time t. m_i and $\mathbf{v}_i$ are the mass of the particle and the velocity vector respectively. To integrate the above second-order differential equations the instantaneous forces acting on the particles and their initial positions and velocities need to be specified. The force on a particle at position r_i, is

determined from the potential energy of the system:

$$\mathbf{F}_i = -\nabla_i V(\mathbf{r}) \tag{2.2}$$

Under the influence of a continuous potential the motions of all the particles become coupled together, giving rise to a many body problem that cannot be solved analytically. Thus, the equations of motion need to be solved numerically. Several algorithms have been designed to do this. The most common is the Verlet algorithm [74] which is based on the Taylor expansion of the positions $\mathbf{r}_i$ around time t:

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i + \frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t}\Delta t + \frac{1}{2}\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}(\Delta t)^2 + \frac{1}{3!}\frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}t^3}(\Delta t)^3 + \mathscr{O}(\Delta t^4) \tag{2.3}$$

$$\mathbf{r}_i(t-\Delta t) = \mathbf{r}_i - \frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t}\Delta t + \frac{1}{2}\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}(\Delta t)^2 - \frac{1}{3!}\frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}t^3}(\Delta t)^3 + \mathscr{O}(\Delta t^4) \tag{2.4}$$

Summing these two equations and substituting the velocity $\mathbf{v}_i$ and the acceleration $\mathbf{a}_i$ in place of the first and the second derivative respectively, we obtain:

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \mathbf{a}_i(\Delta t)^2 \tag{2.5}$$

The velocities can be calculated subtracting (2.4) from (2.3)

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t} \tag{2.6}$$

This is necessary for the calculation of the kinetic energy. Eq. (2.5) is accurate to order Δt^4 while eq. (2.6) is accurate to order Δt^2. The Verlet algorithm is simple and compact to code and the time-reversal symmetry leads to good energy conservation. However, the velocities are not well handled. Also (2.5) involves the addition of a small number ($\mathbf{a}_i(\Delta t)^2$) onto a large number ($2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t)$), which can introduce numerical inaccuracies.
One of the alternative proposed to remedy this shortcoming is the Leap-Frog algorithm[75].

$$\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \mathbf{v}_i(t-\frac{1}{2}\Delta t) + a(t)\Delta t \tag{2.7}$$

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t+\frac{1}{2}\Delta t)\Delta t \tag{2.8}$$

Here the mid-step velocities are calculated using (2.7) and then these are using to calculate the new positions using (2.8). The velocity at time t may then be calculated from

$$\mathbf{v}_i(t) = \frac{1}{2}\left(\mathbf{v}_i(t+\frac{1}{2}\Delta t) + \mathbf{v}_i(t-\frac{1}{2}\Delta t)\right) \tag{2.9}$$

The time step Δt is an important control parameter for a simulation. The largest value of Δt is determined by the fastest process occurring in the system, typically being an order of magnitude smaller than the fastest process. For simulating nuclear motions, the fastest process is the motion of hydrogens, being the lightest particles. Hydrogen vibrations occur with a typical frequency of 3000 cm $^{-1}$, corresponding to 10^{14} s^{-1}, and therefore necessitating time steps of the order of one femtosecond (10^{-15} s).

Molecular dynamics simulations generate information at the microscopic level. The conversion of this microscopic information to macroscopic observables requires statistical mechanics.

The *thermodynamic state* of a system is defined by the set of its macroscopic parameters, like temperature T, pressure P and number of particles N, while the microscopic state is defined by the positions, **r**, and momenta, **p**, of all the particles in the system. The collection of these phase space points that belong to a particular thermodynamic state is called ensemble and corresponds to all the possible conformations of the system. The macroscopical observables are then defined as averages over all possible state of the system and they are called *ensemble averages*. For a generic physical property A the ensemble average is described in equation 2.10.

$$\langle A(\mathbf{r},\mathbf{p})\rangle_{ensemble} = \iint \rho(\mathbf{r},\mathbf{p})A(\mathbf{r},\mathbf{p})\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p} \tag{2.10}$$

where $\rho(\mathbf{r},\mathbf{p})$ is the probability density of finding the system at each point in phase space, which is given by:

$$\rho(\mathbf{r},\mathbf{p}) = \frac{\exp[-\beta E(\mathbf{r},\mathbf{p})]}{Q} \tag{2.11}$$

with $\beta = 1/(k_\mathrm{B}T)$, k_B being the Boltzmann constant, T being the absolute temperature and Q is the partition function, which is an integral of $\exp(-\beta E(\mathbf{r},\mathbf{p}))$ along all possible points in phase space. The integration in eq. 2.10 is over all possible variables of **r** and **p**. This integral is generally extremely difficult to calculate because one must calculate all possible states of the system. In a molecular dynamics simulation, the points in the ensemble are calculated sequentially in time, so to calculate an ensemble average, the molecular dynamics simulations must pass through all possible states corresponding to the particular thermodynamic constraints.

The way MD simulations determine the observable values is, however, trough a *time average*

of A as described in equation 2.12

$$\langle A \rangle_{time} = \frac{1}{M} \sum_{1}^{M} A\left(\mathbf{r}(t), \mathbf{p}(t)\right) \tag{2.12}$$

where M is the number of time steps in the simulation, t is the simulation time and $A\left(\mathbf{r}(t), \mathbf{p}(t)\right)$ is the instantaneous value of A. A key concept that ties the two averaging strategies together is the *ergodic hypothesis*. This hypothesis states that for an infinitely long trajectory the ensemble average and the time average become equivalent to each other,

$$\lim_{t \to \infty} \langle A \rangle_{time} = \langle A \rangle_{ensemble} \tag{2.13}$$

In other words, the ergodic hypothesis claims that when the trajectory becomes long enough, the points that it generates will eventually cover all of phase space, so the two averages become identical. One goal, therefore, of a molecular dynamics simulation is to generate enough representative conformations such that this equality is satisfied.

As the macroscopic experiments might be performed in different conditions so do MD simulations, and the choice of the ensemble does make a difference when computing the observables averages. However, techniques exist to relate fluctuations in different ensembles allowing a comparison of the results of MD simulations performed in different conditions. Several algorithms have been developed to mimic the different ensembles properties, such as the Berendsen, Velocity Rescaling, Nosé–Hoover and Parinello-Rahman algorithms.[76–80] With these methods is thus possible to perform MD simulations in different ensembles, such as: *NVE* (micro-canonical) ensemble, in which the number of particles N, the volume V and the total energy E are constant; *NVT* (canonical) ensemble, with conserved number of particles N, volume V and temperature T; *NPT* (isothermal–isobaric) ensemble, with constant number of N particles, pressure P and temperature T.

2.2 Molecular Mechanics

Deriving the total potential energy is one of the main tasks in molecular dynamics. Molecular mechanics (MM) describes molecules in terms of *bonded atoms*. The electronic structure of the system is only described effectively and the potential energy is formulated as a parametrized function of the nuclear coordinates where the parameters are fitted to experimental or high level computational data.

The potential energy function can be written as a sum of bonded terms and non-bonded

terms:

$$V(\mathbf{R}) = V_{\text{bonded}} + V_{\text{non-bonded}}$$

The non-bonded energy terms account for interactions between non-bonded atoms and atoms that are more than three covalent bonds away from each other in a molecule. These are modelled by the van der Waals energy and electrostatic energy, the first and second summation terms in equation [(2.14)], respectively:

$$V_{\text{non-bonded}} = \sum_{\text{non-bonded pairs}} 4\varepsilon \left[\left(\frac{\sigma_{ik}}{r_{ik}} \right)^{12} - \left(\frac{\sigma_{ik}}{r_{ik}} \right)^{6} \right] + \sum_{\text{non-bonded pairs}} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ik}} \qquad (2.14)$$

The van der Waals energy is described by the Lennard-Jones potential, which includes both a short-range, repulsive component due to electronic clouds overlap, $\left(\frac{\sigma_{ik}}{r_{ik}} \right)^{12}$, and an attractive component due to dispersion forces, $-\left(\frac{\sigma_{ik}}{r_{ik}} \right)^{6}$. The two empirical parameters ε and σ_{ik} represent respectively the potential well depth and the zero-potential distances between atoms i and k (figure 2.1).

The calculation of the non-bonded terms in the potential function is the most time-consuming part of the MD simulation. In principle, the interaction between every pair of atoms should be calculated explicitly, meaning that for an N-atom system, N^2 calculations would be required. For the van der Walls interactions the contribution of remote particles can be considered insignificant, as depicted in figure 2.1. Thus, in order to save computational time, methods were developed to ignore the interactions between two atoms separated by a distance greater than a specified cutoff. However, for the electrostatic interactions, cutoff schemes have been shown to be inaccurate ([81]), since they have considerably longer range (they decay as a function of r^{-1}). A more efficient approach is the Ewald summation technique ([82]), which separates the potential into a slowly decaying long-range component and also a quickly varying short-range component.

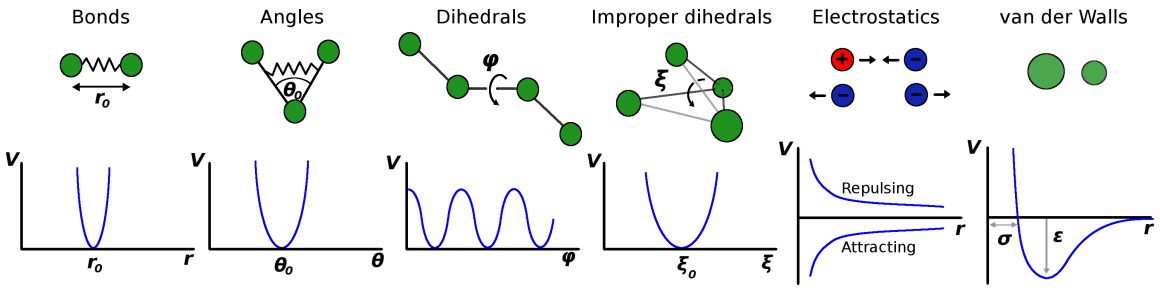


Fig. 2.1 Bonded and non-bonded potential energy terms.

The bonded terms include the contributions of the bond, angles and dihedral interactions. They have the following functional form:

$$V_{\text{bonded}} = \sum_{\text{bonds}} k_b(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedral pairs}} V_{n,\psi}(1-\cos(n\psi)) \qquad (2.15)$$

The terms are calculated using approximations for each type of motion. The bond length (stretching) between two particles is often modelled using a harmonic spring potential (first term in equation (2.15)). In this approximation the interacting particles are described as rigid bodies, connected by a spring with an empirical k_b constant and an equilibrium length r_0. Analogue to bonds, angle θ between three particles also fluctuates around a reference value θ_0, by an harmonic potential with force constant k_θ.

Dihedral angles describe the rotation around covalent bonds. Four atoms are needed to define the dihedral angle (ψ). The dependence of potential energy on the dihedral angle is modelled by a simple periodic function, and a usual way to express this function is with a cosine (third term in eq. (2.15)). $V_{n,\psi}$ is the height of the rotational barrier and n is the periodicity.

In most cases the bonded potential energy functions include an *out-of-plane* term, usually in the form of an improper dihedral, where the potential energy is again harmonic:

$$V_{\text{improper}} = \sum_{\text{improper}} k_\xi(\xi-\xi_0)^2 \qquad (2.16)$$

Molecular mechanics models differ both in the number and specific nature of the terms incorporated in the potential energy function, as well as in the details of their parameterization. Taken together, functional form and parameterization, constitute what is called a *force field*.

2.3 Quantum Chemical Methods

A different way to obtain the potential energy acting on the nuclei is based on quantum mechanics (QM), which can describe the electronic structure of the system explicitly. With this framework it is possible to derive properties that depend upon the electronic distribution and, in particular, to investigate chemical reactions in which bonds are broken and formed. QM methods are based on the time-independent Schrödinger equation (eq. 2.17), which cannot actually be solved for any but a one-electron system (the hydrogen atom), and approximations need to be made. Quantum chemical models differ in the nature of these approximations. There are essentially two radically different approaches: the wave function based *ab initio* methods and density functional theory (DFT).

An essential part of solving the Schrödinger equation is the *Born–Oppenheimer approximation*, where the coupling between the nuclei and electronic motion is neglected. This allows the electronic part, also called *electronic Schrodinger equation*, to be solved as a function of the nuclear coordinates:

$$\mathbf{H}_{el}\Psi_{el} = E_{el}(\mathbf{R})\Psi_{el} \tag{2.17}$$

Here, Ψ_{el} is the electronic wavefunction, $\mathbf{R}$ are the nuclear coordinates and $\mathbf{H}_{el}$ is the electronic Hamiltonian operator, which in atomic units is given by:

$$\mathbf{H}_{el} = -\sum_{i}^{electrons} \nabla^2 - \sum_{i}^{electrons}\sum_{k}^{nuclei} \frac{z_k}{r_{ik}} + \sum_{i<j}^{electrons} \frac{1}{r_{ij}} \tag{2.18}$$

where z_k is the number of protons in the nucleus k, r_{ik} is the distance between nucleus k and electron i and r_{ij} is the distance between electrons i and j.
The solution, E_{el}, gives the potential energy experienced by the nuclei also called *potential energy surface* (PES), which forms the basis for solving the nuclear motion.
The best approximation of the wave function is determined by the variational principle, which states that any approximate wave function has an energy above or equal to the exact energy. The equality holds only if the wave function is the exact function. By constructing a trial wave function containing a number of parameters, one can generate the "best" trial function of the given form by minimizing the energy as a function of these parameters. The energy of an approximate wave function can be calculated as the expectation value of the Hamiltonian operator, divided by the norm of the wave function.

$$E_e = \frac{\langle \Psi_{el}|\mathbf{H}_{el}|\Psi_{el}\rangle}{\langle \Psi_{el}|\Psi_{el}\rangle} \tag{2.19}$$

2.3.1 Hartree-Fock

The Hartree-Fock (HF) method is a wave function based approach and it is the basis of molecular orbital (MO) theory. In the HF model the motion of one electron is considered to be independent of the dynamics of all other electrons. The interaction between the electrons is approximated by taking all interactions into account in an average fashion.
Neglecting the electron-electron potential in equation (2.18) the electronic Schrodinger equation would be separable and the total electronic wave function could be written as a product of single-electron molecular orbitals (MOs) describing each electron on the molecule. Since electrons are indistinguishable fermions (particles with a spin of ½), however, the

overall wave function must be antisymmetric (change sign) with respect to interchange of any two electron coordinates. This is conveniently achieved by arranging the orbitals in a *Slater determinant*:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \dots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \dots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \dots & \chi_N(\mathbf{x}_N) \end{bmatrix} \tag{2.20}$$

$\chi_i(\mathbf{x})$ is termed a *spin orbital* and is the product of a spatial function, $\phi_i(\mathbf{r})$, and a spin function, α or β. The spin orbitals are functions of the space-spin coordinate $\mathbf{x} = \{\mathbf{r}, \omega\}$ where ω represents the spin coordinate which can be either α or β.
The one-electron spin-orbital functions are ortho-normalized:

$$\langle \chi_i | \chi_j \rangle = \delta_{ij} \tag{2.21}$$

The spatial functions, which are also termed molecular orbitals (MO), are built by a linear combination of atomic orbitals η (LCAO).

$$\phi_i = \sum_j c_{ij} \eta_j \tag{2.22}$$

The best set of molecular orbitals determined by the variational principle are obtained by a process called *self-consistent-field* (SCF). These functions are solutions of one-dimensional eigenvalue equations with a special *Fock operator* $\hat{f}(i)$:

$$\hat{f}(i)\chi(\mathbf{x}) = \varepsilon\chi(\mathbf{x}) \tag{2.23}$$

with

$$\hat{f}(i) = -\frac{1}{2}\nabla_i^2 - \sum_k^{nuclei} \frac{z_k}{r_{ik}} + v^{\mathrm{HF}}(i) \tag{2.24}$$

Here v^{HF} is an effective mean-field potential that is obtained from the combined charge densities of all other electrons. Thus, in order to solve this equation, one needs an initial guess for the wave functions, which are used to construct the mean-field potential v^{HF} that is used to solve for new wave functions. The new wave functions then are used to construct a new mean-field potential, and the process is repeated until no significant change in the orbital energies or functions occurs.

Hartree-Fock model provides good descriptions of equilibrium geometries and conformations. However, it does not provide chemical accuracy. The failures of Hartree-Fock models can be traced to an incomplete description of *electron correlation*. Electrons in the same spatial orbital (but obviously with different spin state) tend to avoid each other and a proper description of the two-electron wave function should take the electron correlation into account, leading to a lower energy. This is also the case for electrons in different orbitals. In fact, the London dispersion interaction between far-away electrons is based on electron correlation and will be entirely neglected in the HF approximation.
Two fundamentally different approaches for improvement of Hartree-Fock models have emerged. One approach is to construct a more flexible description of electron motions in terms of a combination of Hartree-Fock descriptions for ground and excited states. Configuration interaction (CI) and Møller-Plesset (MP) models are two of the most commonly used models of this type. The so-called second-order Møller-Plesset model (MP2) is the most practical and widely employed. It generally provides excellent descriptions of equilibrium geometries and conformations, as well as thermochemistry, including the thermochemistry of reactions where bonds are broken and formed. An alternative approach to improve upon Hartree-Fock models involves including an explicit term to account for the way in which electron motions affect each other. In practice, this account is based on an *exact* solution for an idealized system, and is introduced using empirical parameters. As a class, the resulting models are referred to as density functional models.

2.3.2 DFT

Density Functional Theory (DFT) in the Kohn-Sham version [83] can be considered as an alternative to HF theory, where the many-body effect of electron correlation is modelled by a function of the electron density. DFT is, analogously to HF, an independent-particle model, and is comparable to HF computationally, but provides significantly better results. The central quantity in DFT is the *electron density*, $\rho(\mathbf{r})$, as opposed to the many-electron wave function of HF method. The basic idea comes from the two theorems of Hohenberg and Kohn. [84] The first theorem states that the ground-state properties of an atom or molecule are uniquely determined by its electron density function.Thus, the ground state energy can be expressed as a functional of the density $\rho(\mathbf{r})$:

$$E^{\mathrm{DFT}}[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + E_{\mathrm{ne}}[\rho(\mathbf{r})] + E_{\mathrm{ee}}[\rho(\mathbf{r})] \tag{2.25}$$

where

$$E_{\mathrm{ne}} = \int \rho(\mathbf{r}) V_{\mathrm{n}}(\mathbf{r}) d\mathbf{r} \tag{2.26}$$

is the electron-nuclear interaction energy, with V_{n} the potential due to the nuclei, T is the kinetic energy of the electrons, and E_{ee} is the electron-electron interaction which can be written as:

$$E_{\mathrm{ee}}[\rho(\mathbf{r})] = J[\rho(\mathbf{r})] + E_{\mathrm{ncl}}[\rho(\mathbf{r})] \tag{2.27}$$

The second term on the right hand side of (2.27) is the non-classical exchange-correlation energy, while the first is the classical electron-electron electrostatic interaction, given by:

$$J[\rho(\mathbf{r})] = \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|} d\mathbf{r} d\mathbf{r}' \tag{2.28}$$

Now, the second theorem of Hohenberg and Kohn states that for any density distribution $\rho' \neq \rho$ (where ρ is the exact ground state density), the energy is never smaller than the true ground state energy E:

$$E[\rho'] \geq E[\rho] \tag{2.29}$$

Thus the minimization of E by varying $\rho(\mathbf{r})$ would provide the ground state energy and the charge distribution. The problem is that the exact functionals $T[\rho(\mathbf{r})]$ and $E_{\mathrm{ncl}}[\rho(\mathbf{r})]$ are not known.

A practical solution was provided by Kohn and Sham [83], who suggested to calculate the exact kinetic energy of a reference non-interacting electrons system with the same density as the real system of interacting electrons. The total antisymmetrized wave function would be a Slater determinant made of Kohn-Sham orbitals $\phi(\mathbf{r})$ which are similar as in HF theory. The one-electron Schrödinger equation for the reference non-interacting system is:

$$\left[-\frac{1}{2}\nabla^2 + V_s(\mathbf{r})\right]\phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{2.30}$$

with the requirement that

$$\rho_s(\mathbf{r}) = \sum_i^N \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}) \qquad \text{and} \qquad \int \rho(\mathbf{r})\mathrm{d}\mathbf{r} = N \tag{2.31}$$

where N is the total number of electrons.

The kinetic energy of such a non-interactive system can be written as:

$$T_s[\rho(\mathbf{r})] = \sum_i \int \phi_i^*(\mathbf{r})\left(-\frac{\nabla^2}{2}\right)\phi_i(\mathbf{r})\mathrm{d}\mathbf{r} = \sum_i \varepsilon_i - \int V_s(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} \tag{2.32}$$

In order to have a formally exact theory the difference between T_s and the kinetic energy of an interacting electrons system T is added to the non-classical exchange-correlation energy:

$$E_{\mathrm{xc}}[\rho(\mathbf{r})] = E_{\mathrm{ncl}}[\rho(\mathbf{r})] + (T[\rho(\mathbf{r})] - T_s[\rho(\mathbf{r})]) \tag{2.33}$$

The exchange and correlation energy $E_{\mathrm{xc}}[\rho(\mathbf{r})]$ is the functional that contains everything that is unknown. The total energy is then given by

$$\begin{aligned} E^{\mathrm{DFT}}[\rho(\mathbf{r})] =& T_s[\rho(\mathbf{r})] + E_{\mathrm{ne}}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{\mathrm{xc}}[\rho(\mathbf{r})] \\ =& \sum_i \varepsilon_i - \int V_s(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} + \int V_{\mathrm{ne}}(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} \\ &+ \frac{1}{2}\iint \frac{\rho(\mathbf{r}')\rho(\mathbf{r})}{|r'-r|}\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r} + E_{\mathrm{xc}}[\rho(\mathbf{r})] \end{aligned} \tag{2.34}$$

The (2.34) is obtain substituting (2.32), (2.26) and (2.28) in (2.25). The next step would be to find a potential $V_s(\mathbf{r})$ such that it provides a total wave function with the same density as the real system. In order to get it the energy $E^{\mathrm{DFT}}[\rho]$ is minimized in a variational search in the space of $\rho(\mathbf{r})$.

$$\left.\frac{\delta E^{\mathrm{DFT}}[\rho]}{\delta\rho}\right|_{\rho=\rho(\mathbf{r})} = -V_{\mathrm{s}}(\mathbf{r}) + \left[V_{\mathrm{n}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|r'-r|}\mathrm{d}\mathbf{r}' + V_{\mathrm{xc}}(\rho(\mathbf{r}))\right]_{\rho=\rho(\mathbf{r})} = 0 \tag{2.35}$$

The exchange-correlation potential, $V_{\mathbf{xc}}[\rho(\mathbf{r})]$ is defined as the functional derivative of $E_{\mathrm{xc}}[\rho(\mathbf{r})]$ with respect to $\rho(\mathbf{r})$:

$$V_{\mathrm{xc}}[\rho(\mathbf{r})] = \frac{\delta E_{\mathrm{xc}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \tag{2.36}$$

Thus, the potential of the non-interacting system is given by:

$$V_{\mathrm{s}}(\mathbf{r}) = V_{\mathrm{n}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|r'-r|}\mathrm{d}\mathbf{r}' + V_{\mathrm{xc}}(\rho(\mathbf{r})) \tag{2.37}$$

Now, this potential can be used to solve the one-electron Schrödinger equation (2.30), which in turn determines the Kohn-Sham orbitals and hence the ground state density and the ground state energy. Note that $V_{\mathrm{s}}(\mathbf{r})$ depends on the density, and therefore the equations have to be solved iteratively in a self-consistent manner. It is very important to note that if the exact forms of E_{xc} and V_{xc} were known, the Kohn-Sham strategy would lead to the exact ground-state energy.

As long as no theory is available to derive the form of the exchange-correlation potential

from first principles, approximations must be made. The first attempts to find a form for the exchange-correlation functional started from the exact result for a uniform electron gas, in which case the exchange potential is only dependent on the electron density. This *Local density approximation* (LDA) works well for uniform electron density distributions, such as found in metals. However, it is not accurate enough for atoms and molecules. More sophisticated corrections include at least the gradient of the density as in the *generalized gradient approximation* (GGA). A different approximation are the so called hybrid functionals. These include a portion of Hartree-Fock (exact) exchange using a few empirical parameters. One of the most popular is the B3LYP functional.[85]

Nowadays, DFT is widely used for the calculation of properties of molecular systems of up to hundreds of atoms. One disadvantage of DFT is the large amount of computational time needed for the evaluation of integrals during runtime. Especially when calculating electronic structures along MD trajectories, DFT becomes too expensive in computational cost.

2.3.3 DFTB

Density Functional Tight Binding (DFTB) denotes a series of semi-empirical models derived from DFT. They require element and element-pair specific parameters, however the number of them is much smaller than in MM approaches. DFTB is a quantum chemical method and therefore is able to describe bond breaking and formation events. In terms of computational efficiency, DFTB are midway between ab initio/DFT and MM methods: they are roughly three orders of magnitude faster than DFT (with a medium sized basis set) and three orders of magnitude slower than MM methods. Without further algorithmic improvements, both DFTB and DFT methods show an $\mathcal{O}(N^3)$ scaling. This means that, with a given computational equipment, DFTB is able to treat systems with 10 times more atoms or can perform Molecular Dynamics (MD) at a three orders of magnitude longer time scale than DFT. For example, it is common to use DFT for systems up to 100 atoms to carry out MD simulations on the picosecond time scale, while using DFTB the time scale is readily extended to nanoseconds.

The starting point is the use of a reference density ρ_0, which is a superposition of neutral atomic densities and deviates from the DFT total energy as introduced by $\Delta\rho$. After expanding the exchange-correlation energy in a Taylor series and some rearrangements the DFT total energy can be rewritten in a different form. In that form one can approximate the DFT total energy just by the assumption of a reference density (DFTB) and also by an approximation of the density fluctuation (DFTB2).

The DFT ground-state energy can be explicitly written as:

$$\begin{aligned} E^{\mathrm{DFT}}[\rho(\mathbf{r})] = &\sum_i \int \phi^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{ne}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|} \mathrm{d}\mathbf{r}' + V_{\mathrm{xc}}[\rho(\mathbf{r})] \right) \phi(\mathbf{r}) \mathrm{d}\mathbf{r} \\ &- \frac{1}{2} \iint \frac{\rho(\mathbf{r}')\rho(\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|} \mathrm{d}\mathbf{r}' \mathrm{d}\mathbf{r} - \int V_{\mathrm{xc}}[\rho(\mathbf{r})] \mathrm{d}\mathbf{r}' \\ &+ E_{\mathrm{xc}}[\rho(\mathbf{r})] + E_{\mathrm{nn}} \end{aligned} \tag{2.38}$$

which is the electronic energy shown in equation (2.38) plus the nuclear-nuclear repulsion energy $\sum_{lk} \frac{z_l z_k}{R_{lk}}$.

The starting point of DFTB is to obtain the reference density $\rho_0(\mathbf{r})$ from a superposition of pre-calculated neutral atomic densities.

$$\rho_0(\mathbf{r}) = \sum_a \rho_a \tag{2.39}$$

The DFT density deviates from the reference density by $\Delta\rho(\mathbf{r})$.

$$\rho(\mathbf{r}) = \sum_i \phi(\mathbf{r})\phi^*(\mathbf{r}) = \rho_0(\mathbf{r}) + \Delta\rho(\mathbf{r}) \tag{2.40}$$

The DFT total energy can then be written as [1]:

$$\begin{aligned} E^{\mathrm{DFT}}[\rho_0 + \Delta\rho] = &\sum_i \int \phi^* \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{ne}} + \int' \frac{\rho_0' + \Delta\rho'}{|\mathbf{r}'-\mathbf{r}|} + V_{\mathrm{xc}}[\rho_0 + \Delta\rho] \right) \phi \\ &- \frac{1}{2} \iint' \frac{\left(\rho_0' + \Delta\rho'\right)\left(\rho_0 + \Delta\rho\right)}{|\mathbf{r}'-\mathbf{r}|} - \int V_{\mathrm{xc}}[\rho_0 + \Delta\rho] \\ &+ E_{\mathrm{xc}}[\rho_0 + \Delta\rho] + E_{\mathrm{nn}} \end{aligned} \tag{2.41}$$

In the next step the first term of eq (2.41) is modified such that it is only dependent on the reference density, i.e. the $\Delta\rho$ term of the classical electron-electron interaction J is extracted from this integral and written in a separate term. Further the exchange-correlation term dependent on $\rho_0 + \Delta\rho$ is removed from the first term and inserted separately, at the same time an exchange-correlation contribution dependent only on the reference density, $V_{\mathrm{xc}}[\rho_0]$ is

[1] For an easier readability in some cases the abbreviations $\int = \int \mathrm{d}\mathbf{r}$, $\int' = \int \mathrm{d}\mathbf{r}'$, $\int = \int'' \mathrm{d}\mathbf{r}''$, $\rho = \rho(\mathbf{r})$, $\rho' = \rho(\mathbf{r}')$, $\rho'' = \rho(\mathbf{r}'')$ and $\phi = \phi(\mathbf{r})$ are used.

added into the first term and again subtracted in an additional term.

$$\begin{aligned} E^{\mathrm{DFT}}[\rho_0+\Delta\rho] = &\sum_i \int \phi^* \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{ne}} + \int' \frac{\rho_0'}{|\mathbf{r}'-\mathbf{r}|} + V_{\mathrm{xc}}[\rho_0] \right) \phi \\ &- \frac{1}{2} \iint' \frac{\rho_0'\rho_0}{|\mathbf{r}'-\mathbf{r}|} - \frac{1}{2} \iint' \frac{\Delta\rho_0'\Delta\rho_0}{|\mathbf{r}'-\mathbf{r}|} \\ &- \int V_{\mathrm{xc}}[\rho_0]\rho_0 - \int V_{\mathrm{xc}}[\rho_0]\Delta\rho \\ &+ E_{\mathrm{xc}}[\rho_0+\Delta\rho] + E_{\mathrm{nn}} \end{aligned} \tag{2.42}$$

With this rearrangement it is not clear anymore how to obtain the exact Kohn-Sham wave functions ϕ_i. However, within the DFTB formalism one can obtain approximate ϕ_i as will be shown in the following subsections. In a last step the exchange-correlation energy is expanded in a Taylor series :

$$\begin{aligned} E_{\mathrm{xc}}[\rho_0+\Delta\rho] = E_{\mathrm{xc}}[\rho_0] &+ \int \left[\frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho} \right]_{\rho_0} \Delta\rho \\ &+ \frac{1}{2} \int' \int \left[\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'} \right]_{\rho_0\rho_0'} \Delta\rho\Delta\rho' \\ &+ \frac{1}{6} \int'' \int' \int \left[\frac{\delta^3 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'\Delta\rho''} \right]_{\rho_0\rho_0'\rho_0''} \Delta\rho\Delta\rho'\Delta\rho'' \\ &+ \cdots \end{aligned} \tag{2.43}$$

Applying (2.43) together with (2.36) the total energy can be written as:

$$\begin{aligned} E^{\mathrm{DFT}}[\rho_0+\Delta\rho] = &\sum_i \int \phi^* \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{ne}} + \int' \frac{\rho_0'}{|\mathbf{r}'-\mathbf{r}|} + V_{\mathrm{xc}}[\rho_0] \right) \phi \\ &- \frac{1}{2} \iint' \frac{\rho_0'\rho_0}{|\mathbf{r}'-\mathbf{r}|} - \int V_{\mathrm{xc}}[\rho_0]\rho_0 + E_{\mathrm{xc}}[\rho_0] + E_{\mathrm{nn}} \\ &+ \frac{1}{2} \int' \int \frac{1}{|\mathbf{r}'-\mathbf{r}|} + \left[\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'} \right]_{\rho_0\rho_0'} \Delta\rho\Delta\rho' \\ &+ \frac{1}{6} \int'' \int' \int \left[\frac{\delta^3 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'\Delta\rho''} \right]_{\rho_0\rho_0'\rho_0''} \Delta\rho\Delta\rho'\Delta\rho'' \\ &+ \cdots \end{aligned} \tag{2.44}$$

Approximations of different levels of sophistication can be introduced by truncation of the Talyor series. Standard (nonself-consistent) DFTB [86] neglects density fluctuation and only the first two lines of eq (2.44) are considered. DFTB2 [87] (formerly referred to as SCC-

DFTB) approximates the second order terms in the density fluctuations, thus the third line of eq (2.44) is also considered. Consequently, the newly developed DFTB3 [88] considers also the third order terms for the density fluctuation, thus the fourth line in eq 2.44. The terms included in each level are subject to further approximations.

The first line of eq (2.44) is the energy contribution from an atomic orbital Hamiltonian that depends only on the reference density, ρ_0. To solve it the single-particle wave functions are represented with a linear combination of a minimal basis set of atomic valence orbitals ϕ_μ centered on the nuclei.

$$\psi_i(\mathbf{r}) = \sum_a \sum_{\mu \in a} c_{i\mu} \phi_\mu (\mathbf{r} - \mathbf{R}_a) \tag{2.45}$$

where ϕ_μ is a basis function of orbital μ centered at atom a, $\mathbf{R}_a$ is the coordinate of the nucleus of atom a. The term in the first line of eq (2.44) can be written as:

$$E^{H_0} = \sum_i \sum_{ab} \sum_{\mu \in a} \sum_{\nu \in b} c_{\mu i} c_{\nu i} H^0_{\mu\nu} \tag{2.46}$$

where i are the occupied molecular orbitals, and a and b are indices for atoms, μ and ν for atomic orbitals. The Hamilton matrix elements, $H^0_{\mu\nu} = \int \phi^*_\mu H^0 \phi_\nu \mathrm{d}\mathbf{r}$, are only dependent on the reference density and they are precalculated in DFT calculations of neutral atoms and stored together with the corresponding overlap matrix elements in a parameter file. This way, none of the integrals have to be computed at runtime.

The second line of eq. 2.44 is the repulsive energy contribution, E^{rep}, which includes the double-counting terms, the exchange-correlation contributions and the nucleus-nucleus repulsion. In DFTB this term is usually approximated as a sum of pairwise potentials.

$$E^{\mathrm{rep}} = \sum_{ab} V^{\mathrm{rep}}_{ab}[\rho^0_a, \rho^0_b, r_{ab}] \tag{2.47}$$

The pair potentials V^{rep}_{ab} are fitted to atom types and distances using several reference systems.

The energy for the DFTB1 only has these two terms. This early truncation of the Taylor series is sufficient for describing unpolar systems with no charge transfer between the atoms. For systems with heterogeneous electron density distribution, including biomolecules, one has to include the density fluctuations $\Delta\rho$. The SCC-DFTB method (DFTB2) also includes the second term of the Taylor series (third line of equation 2.44). Like the reference density

$\rho_0(\mathbf{r})$ in eq. 2.39, also the density variation $\Delta\rho(\mathbf{r})$ is divided into atomic contributions:

$$\Delta\rho_0(\mathbf{r}) = \sum_a \Delta\rho_a \tag{2.48}$$

Furthermore, the atomic-like density fluctuations are expanded in a multipole expansion, however, only keeping the monopole term:

$$\Delta\rho_a(\mathbf{r}) = \Delta q_a F_{00}^a Y_{00} \tag{2.49}$$

where F_{00} denotes the normalized radial distribution function of the density and Y_{00} is the angular distribution, which is spheric. Introducing an analytical function γ, which corresponds to the interaction between the charge fluctuations and further approximations within the second order terms, one can rewrite the second order energy as:

$$E^{\text{2nd}} = \frac{1}{2}\sum_{a,b} \Delta q_a \Delta q_b \gamma_{ab} \tag{2.50}$$

The γ-function has two limiting cases: (i) For large interatomic distances r_{ab}, γ_{ab} basically reduces to $1/r_{ab}$ thus the energy term describes a pure Coulomb interaction of the partial charges Δq_a and Δq_b. (ii) If $a = b$, γ describes the on-site self-repulsion, i.e. the electron-electron interactions on atom a:

$$\gamma_{aa} = U_a \tag{2.51}$$

where U_a is the Hubbard parameter, which describes the chemical hardness. It is a measure for the energy change of a system upon addition or removal of electrons.

The function γ_{ab} assumes a specific inverse relationship of the chemical hardness U of an element with its atomic size. Although this relation holds well within one row of the periodic table, it does not for elements of different rows. In particular, hydrogen turned out to deserve special attention. A modified γ function for hydrogen was proposed, which improves hydrogen bonding significantly. This modified function can be applied within DFTB2, but it is not the default option. Within DFTB3, it has become default.

Further improvement with DFTB3 is made with the inclusion of the third order terms (fourth line of 2.44). The same approximation as for E^{2nd} is applied for E^{3rd}, that is application of equations 2.48 and 2.49. After rearrangement the third order energy contribution becomes:

$$E^{\text{3rd}} = \frac{1}{3}\sum_{ab} (\Delta q_a)^2 \Delta q_b \Gamma_{ab} \tag{2.52}$$

It contains the Γ-matrix, which accounts for the changes of chemical hardness with respect to the atomic charge.

2.4 QM/MM

Most biochemical systems are too large to be described at any level of ab initio theory and even semi-empirical methods such as DFTB are to expensive. At the same time, the available molecular mechanics force fields are not sufficiently flexible to model processes in which chemical bonds are broken or formed. To overcome the limitations of a full quantum mechanical description on the one hand, and a full molecular mechanics treatment on the other hand, methods have been developed that treat a small part of the system at the level of quantum chemistry (QM), while retaining the computationally cheaper force field (MM) for the larger part. Within this approach, chemical reactivity can be studied in large systems, such as enzymes. The justification for dividing a system into regions that are described at different levels of theory is the local character of most chemical reactions in condensed phases. Most reactions in solution involve the reactants and only the first solvation shells. The bulk solvent is hardly affected by the reaction, but can influence the reaction via long-range interactions. The same is true for most enzymes, in which the catalytic process is restricted to an active site located somewhere inside the protein. The rest of the protein provides an electrostatic background that may effect the reaction.

2.4.1 QM/MM energy

The hybrid QM/MM potential energy contains three classes of interactions: interactions between atoms in the QM region, between atoms in the MM region and interactions between QM and MM atoms. The interactions within the QM and MM regions are relatively straightforward to describe, that is at the QM and MM level, respectively. The interactions between the two subsystems are more difficult to describe, and several approaches have been proposed. These approaches can be roughly divided into two categories: subtractive and additive coupling schemes. In the *subtractive scheme*, the QM/MM energy of the system is obtained in three steps. First, the energy of the total system, consisting of both QM and MM regions, is evaluated at the MM level. The QM energy of the isolated QM subsystem is added in the second step. Third, the MM energy of the QM subsystem is computed and subtracted. The last step corrects for including the interactions within the QM subsystem twice:

$$E^{\text{sub}}_{\text{QM/MM}} = E_{\text{MM}}(\text{MM}+\text{QM}) + E_{\text{QM}}(\text{QM}) - E_{\text{MM}}(\text{QM}) \tag{2.53}$$

The subscripts indicate the level of theory at which the potential energies are computed. The main advantage of the subtractive QM/MM coupling scheme is that no communication is required between the quantum chemistry and molecular mechanics routines. This makes the implementation relatively straightforward. However there are also disadvantages. A major disadvantage is that a force field is required for the QM subsystem, which may not always be available. In addition, the force field needs to be sufficiently flexible to describe the effect of chemical changes when a reaction occurs. A further drawback of this method is absence of polarization of the QM electron density by the MM environment. This shortcoming can be particularly problematic for modelling biological charge transfer processes, since these are usually mediated by the protein environment. For a realistic description of such reactions a more consistent treatment of the interactions between the electrons and their surrounding environment is needed.

In *additive schemes*, the QM system is embedded within the larger MM system, and the potential energy for the whole system is a sum of MM energy terms, QM energy terms and QM/MM coupling terms:

$$E^{\text{add}}_{\text{QM/MM}} = E_{\text{MM}}(\text{MM}) + E_{\text{QM}}(\text{QM}) + E_{\text{coup}}(\text{QM/MM}) \tag{2.54}$$

Here, only the interactions within the MM region are described at the force field level, $E_{\text{MM}}(MM)$. In contrast to the subtractive scheme, the interactions between the two subsystems are treated explicitly: $E_{\text{coup}}(\text{QM/MM})$. These interactions can be described at various degrees of sophistication.
The essential task in the development of QM/MM algorithms is the description of the coupling between QM and MM regions. The QM-MM interaction is composed of the bonded (stretching, bending, torsion) and non-bonded (electrostatic interaction and van der Waals interacion) components:

$$E_{\text{coup}}(\text{QM/MM}) = E^{\text{bonded}}_{\text{coup}} + E^{\text{vdw}}_{\text{coup}} + E^{\text{el}}_{\text{coup}} \tag{2.55}$$

2.4.2 Non-bonded coupling

Many techniques are available to deal with the non-bonded interaction terms. In the most basic approach, called *mechanical embedding*, those terms are handled at the force field level. For the calculation of the electrostatic interaction is either used a fixed set of charges for the QM region, for example, those given by the force field, or re-compute the partial charges on the QM atoms at every integration step of the simulation. In both cases the electronic wave function is evaluated for an isolated QM subsystem. Therefore, the MM environment cannot

induce polarization of the electron density in the QM region. A second approach, called *electrostatic embedding*, includes polarization effects by means of an additional operator in the QM Hamiltonian:

$$H^{\mathrm{el}}_{\mathrm{QM/MM}} = \sum_{i}^{\text{electrons}} \sum_{M} \frac{q_M}{|\mathrm{r}_i - \mathrm{R}_M|} + \sum_{\alpha} \sum_{M} \frac{q_M Z_\alpha}{|\mathrm{R}_\alpha - \mathrm{R}_M|} \tag{2.56}$$

where R_M and q_M are the position and the partial charge of MM atom M respectively. r_i and R_α are the positions of the QM electron i and the QM nucleus α with charge Z_α.

The van der Waals interactions are described at the force field level. Thus the additive and the subtractive scheme have the same description with respect to the QM/MM van der Waals coupling. The introduced errors are quite small and can be neglected due to the short range nature of van der Waals interactions.

2.4.3 Bonded Interactions

If the QM and MM subsystems are connected by covalent bonds, care has to be taken when evaluating the QM wave function. A straightforward cut through the QM/MM bond creates one or more unpaired electrons in the QM subsystem. In reality, these electrons are paired in bonding orbitals with electrons belonging to the atom on the MM side. A number of approaches to remedy the artefact of such open valences have been proposed. The easiest solution is to introduce a monovalent link atom at an appropriate fixed position along the bond vector between the QM and MM atoms. Hydrogen is most often used, but there is no restriction on the type of the link atom. These atoms are present only in the QM calculation, and are invisible for the MM atoms. In principle each link atom introduces three additional degrees of freedom to the system. However, in practice the link atom is placed at a fixed position along the bond in every step of the simulation by means of a constraint that fix the distance between the QM and MM atoms to the original bond length. In this way the additional degrees of freedom are removed again but the distance between the link atom and the MM atom results very short. This leads to an overpolarization of the electron density at the QM-MM boundary. Various solutions have been attempted to minimize or even avoid this overpolarization effect. To avoid it and conserve the overall charge, the charges around the QM/MM boundary have to be redistributed. To this end, the charges of the MM boundary atoms are set to zero and redistributed evenly among the other MM boundary atoms. This procedure follows the *divided frontier charge* scheme, which was found to be especially suitable for QM/MM simulations with DFTB.

2.5 Free Energy Calculations

The calculation of free-energy differences is a central task in computational chemistry. The free-energy difference is the driving force of any process, such as a chemical reaction, conformational rearrangement and binding process.

The Helmholtz free energy is the free energy associated with the canonical(NVT) ensemble and is given by the logarithm of the canonical partition function Q_{NVT} by:

$$A(N,V,T) = -\frac{1}{\beta}\ln(Q_{\mathrm{NVT}}) \tag{2.57}$$

with $\beta = 1/(k_{\mathrm{B}}T)$, k_{B} being the Boltzmann constant and T being the absolute temperature. The canonical partition function can be calculated via an integral over the whole phase space, i.e., configuration space and momentum space. If the potential energy E is independent of the latter, the integral over the momentum is a multiplicative constant, which can be ignored. Then, Q_{NVT} is obtained as:

$$Q_{\mathrm{NVT}} = \int \exp[-\beta E(\mathbf{r})]\mathrm{d}\mathbf{r} \tag{2.58}$$

If the pressure, rather than the volume, is kept constant, the Gibbs free energy, denoted as G, is obtained:

$$G(N,P,T) = -\frac{1}{\beta}\ln(Q_{\mathrm{NPT}}) \tag{2.59}$$

where the isothermal-isobaric partition function Q_{NPT} can be written as the weighted sum of the canonical partition function (the laplace transform of the canonical partition function):

$$Q_{\mathrm{NPT}} = \int Q_{\mathrm{NVT}} \exp[-\beta \mathrm{pV}]\mathrm{dV} \tag{2.60}$$

Apart from the change in the ensemble, the formalisms and derivations are equivalent for A and G. In the condensed phase, which is relevant for most applications, the systems are hardly compressible; so ΔA and ΔG are numerically very similar.

Free energies of substates can be evaluated directly from completely equilibrated trajectories that contain all accessible regions of configurational space. In practice it is hard to generate such complete ensembles when there are many low-lying states separated by barriers. Using the concept of ensemble average and probability density (equations (2.10)

and (2.11)), equation (2.57) can be rewritten as:

$$\begin{aligned} A(N,V,T) &= -\frac{1}{\beta}\ln Q \\ &= \frac{1}{\beta}\ln\frac{1}{Q} \\ &= \frac{1}{\beta}\ln\frac{\int \exp[-\beta E(\mathbf{r})]\cdot\exp[\beta E(\mathbf{r})]\mathrm{d}\mathbf{r}}{Q} \\ &= \frac{1}{\beta}\ln\int \exp[\beta E(\mathbf{r})]\rho(\mathbf{r})\mathrm{d}\mathbf{r} \\ &= \frac{1}{\beta}\ln\langle\exp[\beta E(\mathbf{r})]\rangle \end{aligned} \tag{2.61}$$

This expression illuminates the fast growth of $\exp[\beta E(\mathbf{r})]$ with the total energy, $E(\mathbf{r})$, of the system. It should, therefore, be expected that the weight of the high–energy regions of phase space be significant when evaluating the integral. In simulations of finite length, sampling of these regions is likely to be insufficient to guarantee a correct estimate of free energy. A way to improve the sampling is therefore necessary.

2.5.1 Potential of mean force

In biomolecular simulations it is conventional to use *potential of mean force* (PMF) which is a free energy with respect to certain defined variables, called *collective variables* or *reaction coordinates*. In order to calculate the PMF along a reaction coordinate ξ in an ensemble, one needs to integrate out all degrees of freedom but ξ in configuration space. In the case of the canonical ensemble, the PMF along ξ can be calculated as:

$$A(\xi) = -\frac{1}{\beta}\ln\int \delta(\xi(\mathbf{r})-\xi)\exp[-\beta E(\mathbf{r})]\mathrm{d}\mathbf{r} \tag{2.62}$$

where $\delta(\xi(\mathbf{r})-\xi)$ is the Dirac delta function.

In chemistry, one is typically interested not in absolute values of thermodynamic functions but in their changes over the course of a chemical process. Then the energy difference of two states A and B, along a reaction coordinate ξ, is:

$$\Delta A = A(\xi_B) - A(\xi_A) = -\frac{1}{\beta}\ln\frac{\int \delta(\xi(\mathbf{r})-\xi_B)\exp[-\beta E(\mathbf{r})]\mathrm{d}\mathbf{r}}{\int \delta(\xi(\mathbf{r})-\xi_A)\exp[-\beta E(\mathbf{r})]\mathrm{d}\mathbf{r}} \tag{2.63}$$

Recalling the definition of probability density of state in eq. 2.11 the previous can be written as:

$$\Delta A = A(\xi_B) - A(\xi_A) = -\frac{1}{\beta} \ln \frac{P(\xi_B)}{P(\xi_A)} \tag{2.64}$$

Where $P(\xi_0)$ is the probability distribution that the reaction coordinate ξ takes the value ξ_0 and it is given by:

$$P(\xi_0) = \int \delta(\xi(\mathbf{r}) - \xi_0)\rho(\mathbf{r})\mathrm{d}\mathbf{r} \tag{2.65}$$

In computer simulations, the direct phase-space integrals in 2.65 is impossible to calculate and all we have is a trajectory in phase-space. However, if the system is ergodic, i.e., if every point in phase space is visited during the simulation, one can apply eq. (2.13), thus the ensemble average becomes equal to the time average. In this case $P(\xi_0)$ can be evaluated simply counting the occurrences when the reaction coordinate is in the vicinity of the given value of ξ_0.

$$P(\xi_0) \simeq \frac{1}{n\mathrm{d}\xi} \sum_{t=1}^{n} \chi(\xi(t)) \tag{2.66}$$

where the characteristic function $\chi(\xi(t)) = 1$ if $\xi(t) \in [\xi_0, \xi_0 + \mathrm{d}\xi]$ and zero otherwise. n is the number of steps in the trajectory.

However, simulations are only run for finite time. Regions in configuration space around a minimum in E(r) are typically sampled well, whereas regions of higher energy are sampled rarely. For rare events, those with an energy barrier significantly larger than $k_\mathrm{B}T$, direct sampling is unfeasible. To obtain a free energy profile, however, also those high-energy regions, those rare events, are required. Different methods have been developed to sample such rare events. These are often referred to as enhanced sampling techniques. In the next paragraphs, three of these techniques will be introduced: Umbrella Sampling, Metadynamics and Replica Exchange Molecular Dynamics (REMD). The first two belong to a class of methods in which sampling is facilitated by the introduction of an additional bias potential that acts on a selected number of collective variables (CVs). In REMD, several replicas of the same system run in parallel, and coordinates are exchanged from time to time with a Monte Carlo procedure. The sampling is accelerated by increasing the temperature or by artificially modifying the Hamiltonian of all the replica but one from which the unbiased statistics can be extracted.

2.5.2 Umbrella Sampling

Umbrella sampling [89, 90] is one of the most broadly applied methods to calculate the free energy profile along a reaction coordinate.

A reaction coordinate, also called collective variables (CVs), is a function of the atomic coordinates that is capable of describing the physics behind the process under investigation. In short the CVs represent a sort of coarse description of the system, which can be used to analyse a given process is low dimensionality.
A bias potential is applied to drive the system from one thermodynamic state to another. The effect of the bias potential is to get sampling as uniform as possible along the whole reaction coordinate connecting the two states. Thus, the optimal choice would be $V_{\mathrm{b}}(\xi) = -F(\xi)$,[2] with which $V_{\mathrm{b}}(\xi)$ completely compensate the free energy surface and therefore the simulation could uniformly sample the whole surface. Unfortunately in real systems $F(\xi)$ is not known but rather what we aim to calculate, so the main problem that arises is how to construct $V_{\mathrm{b}}(\xi)$ without a detailed knowledge of the system. In order to solve this problem, an efficient strategy is to split the path along the reaction coordinate into windows. In each window a simulation is performed with an auxiliary potential $V_{\mathrm{b}}^{i}(\xi)$, in most cases harmonic, that applies a restraint to the reaction coordinate:

$$V_{\mathrm{b}}^{i}(\xi) = \frac{k}{2}(\xi - \xi_i)^2 \tag{2.67}$$

where i denotes the index of windows, ξ_i is the equilibrium position of the harmonic restraint defining the center of window and k is the force constant.
The biased probability distribution can be written as:

$$P_{\mathrm{b}}^{i}(\xi) = \frac{1}{Q_{\mathrm{b}}^{i}} \int \mathrm{d}\mathbf{r}\delta(\xi(\mathbf{r}) - \xi) \exp[-\beta(E(\mathbf{r}) + V_{\mathrm{b}}^{i}(\xi)] \tag{2.68}$$

where Q_{b}^{i} is the canonical partition function for the potential $E(\mathbf{r}) + V_{\mathrm{b}}^{i}(\xi)$. Because the bias depends only on ξ and the integration is performed over all degrees of freedom but ξ,

$$\begin{aligned} P_{\mathrm{b}}^{i}(\xi) =& \frac{1}{Q_{\mathrm{b}}^{i}} \exp[-\beta V_{\mathrm{b}}^{i}(\xi)] \int \mathrm{d}\mathbf{r}\delta(\xi(\mathbf{r}) - \xi) \exp[-\beta(E(\mathbf{r})] \\ =& \frac{Q}{Q_{\mathrm{b}}^{i}} \exp[-\beta V_{\mathrm{b}}^{i}(\xi)] \frac{1}{Q} \int \mathrm{d}\mathbf{r}\delta(\xi(\mathbf{r}) - \xi) \exp[-\beta(E(\mathbf{r})] \\ =& \frac{Q}{Q_{\mathrm{b}}^{i}} \exp[-\beta V_{\mathrm{b}}^{i}(\xi)] P^{i}(\xi) \end{aligned} \tag{2.69}$$

and consequently the unbiased probability, $P^{i}(\xi)$, is given by:

$$P^{i}(\xi) = \frac{Q_{\mathrm{b}}^{i}}{Q} \exp[\beta V_{\mathrm{b}}^{i}(\xi)] P_{\mathrm{b}}^{i}(\xi) \tag{2.70}$$

[2] F can be both Helmholtz and Gibbs free energy

thus:

$$\begin{aligned} F^i(\xi) &= -\frac{1}{\beta}\ln P^i(\xi) \\ &= -\frac{1}{\beta}\ln P^i_{\mathrm{b}}(\xi) - V^i_{\mathrm{b}}(\xi) + f^i_{\mathrm{b}} \end{aligned} \tag{2.71}$$

$P^i_{\mathrm{b}}(\xi)$ is obtained from the MD simulations of the biased system and $V^i_{\mathrm{b}}(\xi)$ is given analytically. $f^i_{\mathrm{b}} = \frac{1}{\beta}\ln\frac{Q^i_{\mathrm{b}}}{Q}$ is a constant which does not depend on ξ but it rather depends on $V^i_{\mathrm{b}}(\xi)$. Thus, the free energy curves for two subsequent windows differ by a constant shift, which depends on the different values of f^i_{b}. An efficient way of estimating these constants and combining the free energy $F^i(\xi)$ of many windows is by using the weighted histogram analysis method (WHAM).[91]

2.5.3 Metadynamics

Metadynamics [92] is another powerful technique able to enhance sampling and reconstruct the free-energy surface (FES) as a function of a few selected degrees of freedom, referred to as collective variables (CVs).

In metadynamics, a history-dependent bias potential is built during the simulation as a sum of repulsive Gaussians. These Gaussians are centered on the explored points in the CVs space, have a preassigned width (σ) and height (ω_G), and are deposited every τ_G time units as the simulations proceeds. The bias potential at time t thus reads:

$$V_G(\mathbf{s},t) = \sum_{i=1}^{t/\tau_G} \omega_G \exp\left[-\sum_{\alpha}^{N_{CV}} \frac{(s_\alpha - s_\alpha(i\cdot\tau_G))^2}{2\sigma_\alpha^2}\right] \tag{2.72}$$

Note that the width needs to be fixed for each of the CVs (σ_α). Intuitively, Gaussians are repeatedly added to the potential according to the explored states, such that they discourage the system from visiting again already visited configurations.

It can be understood intuitively that such a procedure not only discourages the exploration of the already visited states in the CV space, but also provides an immediate estimate for the underlying free energy surface. As more and more Gaussians are added to the bias, the system will explore a larger and larger part of the CV space. Because Gaussians are most likely added at points where the effective total free energy $F(S)+V(s)$ is lower, their effect will tend to flatten the $F(s)+V(s)$ function (figure 2.2). After a suitable 'filling' time, the bias will start growing parallel to itself, and one can expect to directly estimate $F(s)$ as

$-V(s)$, but for an additional arbitrary constant:

$$V_G(\mathbf{s}, t \to \infty) = -F(\mathbf{s}) + C \tag{2.73}$$

The correctness of relation Eq. (2.73) has been shown empirically by extensive testing on simplified model and by comparison with results obtained with other free-energy methods on complex systems [93, 94]. Furthermore, a formal proof of Eq. (2.73) was provided under the assumption that in the absence of bias, the stochastic dynamics in the CVs space is memoryless. For a detailed explanation see ref. [95].

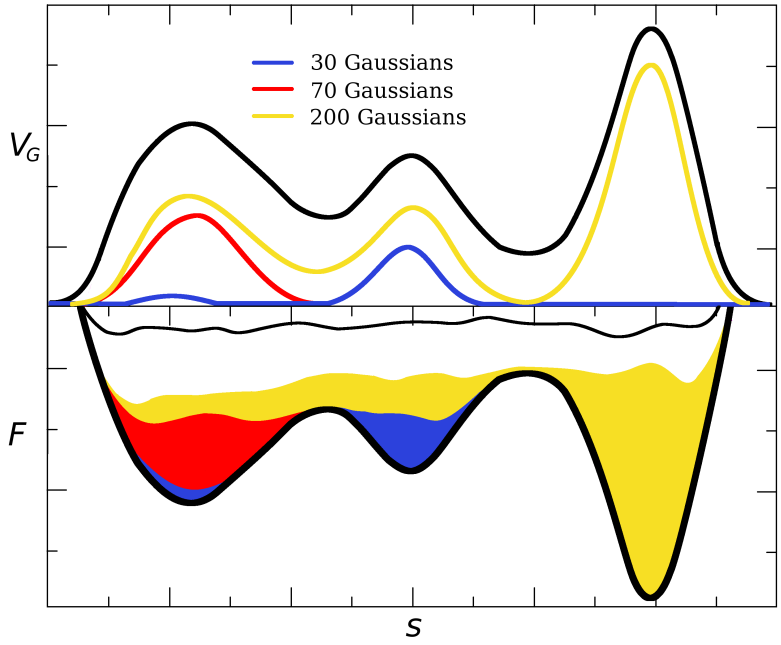


Fig. 2.2 Example of metadynamics simulation along one CV. Upper pannel: Time evolution of the bias potential. The time is measured by counting the number of Gaussians deposited. Lower panel: Schematic representation of the progressive filling of the underlying free energy surface by means of the Gaussians deposited along the trajectory.
The simulation is started from the minimum at the center. After 30 gaussians this minimum is completely filled (blue area), and the system evolves through the lowest saddle point into the left minimum. The system remain there until the underlying free-energy basin is completely filled (red area). At this point, the system diffuses in the region between the first two minima until it reach the minimum in the right. Finally, when this basin is also compensated by the bias potential the system evolution resembles a random walk on the flattened FES.

In standard metadynamics, described above, the height of deposited Gaussians is constant for the entire course of a simulation. This possibly results in the overfilling of the free energy surface and pushes the system to high energy region of the CVs space. Consequently, the

estimated free energy from the bias potential oscillates around the real value. A solution to this problem is provided by *well-tempered metadynamics*[96], in which the heights of the Gaussians decrease with the proceeding of the simulation as follows:

$$\omega_G(t) = \omega \exp\left[-\frac{V_G(\mathbf{s},t)}{k_B \Delta T}\right] \tag{2.74}$$

Equivalently, each Gaussian, when deposited, is scaled down by a factor $\exp[-\frac{V_G(s,t)}{k_B \Delta T}]$, where the bias potential has been evaluated at the same point where the Gaussian is centered and ΔT is an input parameter measured in temperature units. This implies that, after the initial filling, Gaussians of different height are added in different regions of the CV space. In particular, on top of deep wells, where a sizable bias has been already accumulated, the additional Gaussians have small height. In contrast, at the border of the explored region, where the bias is still small, the additional Gaussians have large height. With this rescaling of the Gaussian height, the bias potential smoothly converges in the long time limit. Still, it does not fully compensate the underlying free energy but it rather converges to:

$$V_G(\mathbf{s}, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\mathbf{s}) + C \tag{2.75}$$

It must be noted that in the long time limit the system will explore the biased canonical distribution:

$$P(\mathbf{s}) \propto \exp\left[-\frac{F(\mathbf{s}) + V_G(\mathbf{s},t)}{k_B T}\right] \propto \exp\left[-\frac{F(\mathbf{s})}{k_B (T + \Delta T)}\right] \tag{2.76}$$

Therefore, for $\Delta T \to 0$, ordinary MD is recovered, whereas the $\Delta T \to \infty$ limit corresponds to standard metadynamics. In between one can regulate the extent of FES exploration by tuning ΔT.

In metadynamics literature and in the program PLUMED, used in this work, ΔT is determined by the *bias factor* term, which is defines as:

$$\gamma = \frac{T + \Delta T}{T} \tag{2.77}$$

The bias factor should thus be carefully chosen in order for the relevant free-energy barriers to be crossed efficiently in the time scale of the simulation.

A common challenge in executing metadynamics is selecting the proper CVs. To guarantee an effective application of metadynamics, the CVs should be able to distinguish between the different stable states of a system and capture the slowest degrees of freedom for the transitions of interest. While in some systems this can be an intuitive decision, oftentimes the proper selection of all of the relevant CVs is not obvious and requires a significant level of

knowledge regarding the different states and transitions that exist in a given mechanism. In particular neglecting a slow variable that cannot be sampled in the timescale of the simulation sufficiently may lead to a wrong free energy surface.

2.5.4 Replica Exchange Molecular Dynamics

Replica Exchange Molecular Dynamics (REMD) is one of the most successful enhanced sampling methods. It has two advantages: First, unlike Umbrella Sampling and Metadynamics, it does not require the choice of a set of CVs (i.e. it does not require any a-priori knowledge of the investigated events). Second, REMD can intuitively be combined with many other enhanced sampling methods.

In REMD several replicas of the system are simulated independently in parallel, under different conditions. At predefined intervals, neighboring pairs of replicas are exchanged with a specific transition probability. Based on the type of REMD, replicas can span temperature space (T-REMD) or Hamiltonian space (H-REMD).

T-REMD

The standard temperature formulation of REMD is based on *Tempering*, which refers to a class of methods based on increasing the temperature of an MD system to overcome barriers. In T-REMD, N noninteractive replicas of the system are simulated at different temperatures ($T_0 < T_1 < \cdots < T_N$), where T_0 in the desired temperature, from which the unbiased statistics can be extracted. Replicas evolve independently, and after a certain time interval, exchange between neighboring replicas are attempted, according to the Metropolis Criterion. The probability of exchange between temperatures $T_i < T_j$ is defined as:

$$P(i \longleftrightarrow j) = \begin{cases} 1 & \text{if } U_j < U_i \\ \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)\left(U(\mathbf{r}_j) - U(\mathbf{r}_i)\right)\right] & \text{otherwise} \end{cases} \tag{2.78}$$

where r_i are the coordinate of the i-th replica.

A very well-known issue of T-REMD is that the number of replicas required to span a preassigned temperature range grows with the system size. In fact the overall Hamiltonian grows with system size and the acceptance probability for the exchange depends exponentially on the change in energy. Therefore, for a larger system, one must choose smaller temperature difference to obtain viable acceptance probabilities.

H-REMD

A more efficient alternative to T-REMD is the Hamiltonian replica exchange molecular dynamics (H-REMD)[97], where different replicas evolve according to different Hamiltonians.

Many recipes for H-REMD have been proposed (see, among others,[98, 99]). In the present work the Replica exchange Solute Tempering in its REST2 variant was employed. Here, only the Hamiltonian of the solute is modified.
One could notice that, since the ensemble probability only depends on $H/(k_{\mathrm{B}}T)$, a scaling of the Hamiltonian by a factor λ is completely equivalent to a scaling of the temperature by a factor λ^{-1}. The advantage of scaling the potential energy instead of the temperature is related to the fact that the energy is an extensive property, whereas the temperature is an intensive one. Thus, one can limit the scaling to the portion of the system which is considered to be interesting and which has the relevant bottlenecks. With solute tempering, the solute energy is scaled whereas the solvent energy is left unchanged. This is equivalent to keeping the solute at a high effective temperature and the solvent at the physical temperature. Since in the simulation of solvated molecules most of the atoms belong to the solvent, only a small part of the Hamiltonian is affected, and the energy differences between replicas remain small leading to relatively large $P(i \longleftrightarrow j)$. Therefore, the advantage of H-REMD is that even a relatively small number of replicas may be sufficient to spam a broad range of effective temperatures, resulting in largely improved sampling.

Chapter 3

Structural behavior of the peptaibol harzianin HK VI in a DMPC bilayer

3.1 Introduction

3.1.1 Simulation of membrane-active peptides

Molecular dynamics (MD) simulation has been a useful tool for the investigation of membrane-active peptides (MAPs) for two decades [100, 101]. It faces multiple challenges regarding the parametrization of lipids or more generally the development of force fields uniform for bilayers and molecules embedded into them, or the principal limitations of potential energy functions lacking electronic polarizability [102–104]. Even if these challenges are tackled, there is a real danger of undersampling in all-atom MD simulations of even simple peptide–bilayer systems [105–107]. The reason is that the hydrocarbon tails of lipids rearrange very slowly; as an illustration, note the (lateral) diffusion coefficient in DMPC of 1–10 $\mu m^2/s$ [108, 109], as compared to the value for water of 2000–3000 $\mu m^2/s$ [110].

A partial solution of this problem is to pass to coarse-grained (CG) representation [111], e.g. with the Martini force field [112], which improves the sampling by a factor of 100 to 1000. That makes CG simulation undoubtedly useful for the initial docking and refinement of orientation with low resolution at least, while the lower level of detail may be of limited use for complex lipid environments, specific interactions, and fine conformational changes [111].

A promising simulation strategy seems to be to combine CG and all-atom representations for large-scale sampling and more detailed description of relevant structural states, respectively, to take advantage of the strengths of both of the approaches [113].

Generally, the reliability and suitability of MD simulation should be tested by comparison with the outcome of biophysical experiments. The factors that affected the quality of the results were mostly the choice of the force field and the simulation parameters like inclusion or omission of counterions and the treatment of long-range electrostatics [114, 115].

MAPs have been also a subject of work applying advanced simulation techniques. The temperature replica exchange protocol (REMD)[116] was exploited in structural studies [117–120]. The dissipative particle dynamics [121], which is a variant of CG MD, was applied in a study of pore formation and other bilayer deformations caused by several different MAPs [122]. Umbrella sampling with a novel reaction coordinate was used in a work on antimicrobial lipopeptides forming a micelle and fusing with a membrane [123]. Finally, a series of modern techniques were employed in a thermodynamic study of the WALP peptide [124]: REMD was combined with metadynamics with newly designed collective variables to yield the free energy of folding, and CG simulations used the Hamiltonian replica-exchange approach (HREX). These studies illustrate the flexibility and the broad range of options available for MD simulations of peptide/bilayer complexes.

3.1.2 Simulation of peptaibols

Approaching the current topic, MD simulations were previously employed to investigate the interaction of the long-sequence peptaibol ALM with lipid bilayers consisting of POPC mostly. Sometimes, ALM channels were built artificially first, and then simulated over tens of nanoseconds. The stability and dynamics properties of the helical bundles, the secondary structural elements, as well as the intra- and intermolecular hydrogen-bonding (H-bonding) patterns were examined. First, rather short all-atom simulations concentrated on a single ALM molecule, its structural stability and insertion pathway [125, 107], and the latter point was analyzed in longer simulations [126]. Further studied was the aggregation of several ALM molecules to form channels,[127] a complex topic that was also approached with a combination of CG and all-atom simulations [128]. The formation of aggregates was simulated on a CG level, and then all-atom simulations were applied to focus on the secondary structure of ALM as well as the interaction of water with the ALM clusters. The structure of the clusters was rather irregular, different from previous suggestions of symmetric channels formed by perfect α-helices. The observations would not have been possible with either CG or all-atom simulations alone.

Regarding short-sequence peptaibols, of which harzianin HK VI (hereafter called HZ) is a representative, not much simulation work seems to have been done. Merely conformational analyses of short-sequence peptaibols in aqueous solution were reported, investigating trichobrachin A and B [129] and hypomurocin A [130]. Further, a MD simulation study of

the lipopeptaibol trichogin GA IV showed that such a short lipidated molecule interacts better with bilayers that are thinner by themselves, or, alternatively, the binding of the lipopeptaibol may induce local thinning of the membrane [131]. It was noted, however, that the extent of simulation was insufficient to obtain global thermodynamics equilibrium due to the slow lateral diffusion of the lipid.

3.1.3 Aim of this work

The aim of this work is to simulate, for the first time, an entirely non-polar peptide, a short-sequence peptaibol interacting with a lipid bilayer under conditions of low P/L. The primary goals are (i) to determine the mode of interaction of HZ with the bilayer by combination of the simulations with NMR experiments, and (ii) to establish the secondary structure on the basis of the simulations, also producing structural models for a subsequent solid-state NMR analysis. This will provide a comprehensive picture of the HZ molecule interacting with a DMPC bilayer at low P/L. In addition, free energies of the different possible orientations of HZ will be estimated. Also demonstrated will be any possible difficulties and challenges of application of MD simulation for such a system involving a non-polar peptide, as opposed to applications to more usual polar or cationic peptides.

3.2 Methods

3.2.1 Force fields, general simulation parameters, and starting structures

The molecular system to be simulated consisted of one HZ molecule of the sequence Ac-Aib-Asn-Ile-Ile-Aib-Pro-Leu-Leu-Aib-Pro-Leuol, a bilayer formed by 128 DMPC molecules, and ca. 3500 water molecules. The system was enclosed in a rectagonal periodic box sized ca. $6.3 \times 6.3 \times 6.0$ nm^3. In the simulation with HZ starting outside the bilayer, the box size was increased by ca. 1 nm in the direction of the bilayer normal (z), and the number of water molecules was increased to 4300. A combination of the following force fields was used to describe the system: Amber parm14SB [132–134] for the HZ molecule, Slipids [135–137] for the DMPC molecules, and TIP3P [138] for the water.

The needed sets of atomic charges for the non-standard residues Aib and Leuol were obtained by restrained electric potential fitting method (RESP)[139] with molecular electric potentials obtained on the HF/6-31G* level of theory, and the parameters for torsions along the backbone dihedral angles of Aib were obtained by fitting to data generated on the

B3LYP/TZVP level. Quantum chemical calculations were performed with Gaussian 09 [140]. Long-range electrostatic interactions were treated with the particle–mesh Ewald scheme (PME) [82], using a real-space cutoff of 1.0 nm, Fourier spacing of 0.12 nm and a fourth-order interpolation to the Ewald mesh. Van der Waals interactions were cut off at 1.0 nm. Long-range dispersion correction to the pressure and potential energy were considered [141].

The leap-frog integrator was applied with a time step of 2 fs, and all of the covalent bonds were constrained by means of the LINCS algorithm [142]. All production simulations were performed in the NPT ensemble, where the temperature was kept constant with the Nosé–Hoover thermostat [77, 78], with solvent, peptide and lipid coupled to separate heat baths with coupling constants of 0.5 ps. The pressure was maintained constant with a semi-isotropic scheme, so that the pressure in the membrane plane was controlled separately from the pressure in the membrane normal direction, and the Parrinello–Rahman barostat [79] was applied with a reference pressure of 1 bar, a coupling constant of 2 ps, and a compressibility of 4.5×10^{-5} bar^{-1}. The coordinates of atoms were saved every 2 ps.

To build the structure of the system, use was made of a DMPC bilayer containing 128 molecules (64 per leaflet) pre-equilibrated at 303 K, which was downloaded from the Slipids web site [135]. Two different protocols were employed to create the initial structures of the HZ–bilayer system: (i) The HZ molecule was placed in the aqueous phase and the pre-equilibrated DMPC bilayer was left intact. First, 5 ns of MD simulation were performed at 480 K, which led to the insertion of HZ into the bilayer within 2 ns. Then, the system was cooled down with further 40 ns of simulation with a reference temperature of 300 K. This procedure will be referred to as the heating–cooling protocol [143]. (ii) The structure of the HZ molecule immersed in the DMPC bilayer was generated directly with Inflategro [100], and an equilibration simulation of 40 ns was performed.

All of the simulations were performed with the Gromacs package, versions 5.0.1 and 4.6.7 [144, 145]. The cumulative sampling in this study amounts to 104 μs for the simulations presented in the main text (16 μs of free simulations, 24 μs of HREX, 60 μs of umbrella sampling and 4 μs of metadynamics). These are state-of-the-art time scales, approaching those in contemporary super-computer studies of other authors like e.g. that in Ref. 146.

3.2.2 Free MD simulations

A series of free MD simulations of HZ molecule interacting with the DMPC bilayer was performed, considering four different continuous initial conformations of the peptide: α-helix, 3_{10}-helix, β-bend ribbon helix and fully extended conformation. The corresponding structures of HZ were constructed with the xLeap program of the AmberTools package [147], and the initial backbone dihedral angles applied are shown in Tab. 3.1.

Conformation	Residue	ϕ	ψ	ω
α-helix		$-60°$	$-45°$	$180°$
3_{10}-helix		$-50°$	$-25°$	$180°$
β-bend	Xaa1	$-90°$	$-27°$	$180°$
	Yaa2	$-98°$	$-17°$	$180°$
	Aib3	$-49°$	$-50°$	$180°$
	Pro4	$-78°$	$+3°$	$180°$
extended		$180°$	$180°$	$180°$

Table 3.1 Idealized backbone dihedral angles of the considered initial conformations of the HZ molecule. The values for the β-bend ribbon spiral taken from Ref. 25.

Some simulations involved additional restraints in order to keep the conformation of HZ close to the initial one. For the helical conformations, 3_{10} and α, these took the form of harmonic distance restraints on all of the backbone hydrogen bonds, $V(d) = \frac{1}{2}k(d-d_0)^2$ with d being the length of the hydrogen bond, and the parameters $k = 50$ kJ mol^{-1} Å^{-2} and $d_0 = 2.4$ Å. The two prolines in the HZ molecule cannot act as hydrogen donors, and in these cases, the restraints were applied to the distance b between the unsatisfied hydrogen acceptor (the amido oxygen atom of the amino acid three or four residues upstream, for 3_{10} and α, respectively) and the nitrogen of the proline, with $b = 4.0$ Å. For the β-bend ribbon structure, harmonic restraints of the form $V(\zeta) = \frac{1}{2}k(\zeta-\zeta_0)^2$ were applied on the backbone dihedral angles of the two motifs Xaa–Yaa–Aib–Pro, i.e., amino acid residues 3–10, with $k = 1.523$ kJ mol^{-1} deg^{-2}. Restrained simulations were performed with three different initial conformations (which were also reference conformations for the restraints, at the same time), the α-helix, the 3_{10}-helix, and the β-bend ribbon. Unrestrained simulations were performed with two different initial conformations, the 3_{10}-helix and the fully extended conformation. For each of these five options regarding the conformation of HZ molecule, three different initial positions/orientations were considered: (i) HZ in the aqueous solvent, subject to heating–cooling protocol subsequently, (ii) HZ inside the bilayer, with its molecular axis perpendicular to the bilayer surface, and (iii) HZ inside the bilayer, immersed in the hydrophobic region, with the HZ axis parallel to the bilayer surface. Thus, in total 15 individual free MD simulations were performed, and every of them was extended to 1 μs.

In addition, a cationic analog of HZ interacting with the DMPC bilayer was simulated. That peptide had an AA sequence Ace-Aib-Asn-Lys-Lys-Aib-Pro-Lys-Lys-Aib-Pro-Leuol and will be referred to as HZ4+. Its initial conformation was considered to be 3_{10}-helical, and the starting structure of the complete system was constructed with the heating–cooling

protocol (see the Supporting Information). Then, an unrestrained MD simulation of 1 μs was performed.

3.2.3 Hamiltonian replica-exchange simulations (HREX)

The Hamiltonian replica-exchange method with solute tempering REST2[99] was applied as implemented in Plumed 2.1.2[148] interfaced to Gromacs 4.6.7. Two HREX simulations were performed differing in the initial position of the HZ molecule, and each of them involved eight replicas, with the force field terms of the HZ molecule scaled down by factors that followed a geometric series. The amplitudes of dihedrals in the HZ molecule were scaled by λ, and for the non-bonded interactions, the charges and Lennard-Jones depths of the atoms of HZ were scaled by $\sqrt{\lambda}$ and λ, respectively.

Each of the HREX simulations ran for 2 μs, exchange was attempted every 2 ps for neighboring replicas, and the atomic coordinates were saved every 20 ps. The first simulation was started with HZ in the aqueous phase, and the scaling factors were in the interval $1 \ldots 0.5$, corresponding to effective temperatures from 300 K to 600 K; the cut-off distance for Lennard-Jones interactions was increased to 1.4 nm in this simulation. The second simulation was started with HZ inside the DMPC bilayer, with its axis perpendicular to bilayer surface, and used scaling factors of $1 \ldots 0.3$, or effective temperatures of $300 \ldots 1000$ K. Both simulations were started with a 3_{10}-helical HZ molecule. No barostat was applied in the HREX simulations, thus an NVT canonical ensemble was sampled.

Each of the trajectories generated by the unperturbed replica (scaling factor of 1) was used to generate a plot of free energy as a function of two variables. These two collective variables were defined in the cartesian coordinate system, with the DMPC bilayer oriented in the xy-plane and the z-axis being perpendicular to it, in the following way: The first variable was the distance in the z-direction (i.e., projection of the distance on the z-axis) between the center of mass (COM) of the HZ molecule and the COM of all lipid atoms of the entire bilayer. The second variable was the angle between the major axis of the HZ molecule, and the z-axis. To obtain the axis vector of HZ, the backbone atoms of HZ (all of the N, Cα, C_{CO}, and O atoms) were divided into two equally sized groups, one being N-terminal and the other being C-terminal, and a COM was obtained for each of them, $\vec{R}(\mathrm{N})$ and $\vec{R}(\mathrm{C})$. The helix axis was considered to pass through these two COMs, $\vec{R}(\mathrm{NC}) = \vec{R}(\mathrm{C}) - \vec{R}(\mathrm{N})$. Finally, the tilt angle was obtained from the z-component of $\vec{R}(\mathrm{NC})$ as $\tau = \arccos \frac{R_z(\mathrm{NC})}{R(\mathrm{NC})}$. The two-dimensional space in z and τ was divided into equally spaced bins of width of 0.20 Å and 1.1°, respectively, and the occurrences of (z, τ) from the replicas running with the unperturbed Hamiltonian were counted. The resulting probability distributions $\mathscr{P}(z, \tau)$ were

re-calculated to free energies as $\Delta G = -k_B T \log(\mathscr{P})$, with k_B being the Boltzmann constant and $T = 300$ K.

3.2.4 Metadynamics Simulation

Metadynamics simulations [92], optionally in its well-tempered variant [96], were performed with Plumed 2.1.2 interfaced to Gromacs 5.0. The molecular mechanics parameters were the same as used for the free MD and HREX simulations, as were the initial positions and orientations (HZ placed outside the DMPC bilayer, with its axis parallel to the bilayer surface). The collective variables z and τ were considered, much the same as in the HREX simulations.

In the normal metadynamics simulations, Gaussian potentials of 1.5 kJ·mol^{-1} height were deposited every 1 ps or every 2 ps. The well-tempered metadynamics used two different values of the bias factor, 6 and 10, and Gaussians were deposited every 2 ps with an initial height of 1.2 kJ·mol^{-1}. For all of the simulations, the Gaussian width was set to 0.35 rad for τ and to 0.35 nm for z.

In all of the attempts, one of four different initial conformations of the peptide were employed: α-helix, 3_{10}-helix, fully extended and central-α. The "central-α" conformation was constructed by modelling only the central part of the peptide in an α-helical conformation; that involved only two hydrogen bonds, Ile3–Leu7 and Ile4–Leu8. The considered hydrogen bonds were restrained during the entire simulations. This unusual conformation was motivated by preliminary NMR results.

3.2.5 Umbrella Sampling Simulation

One-dimensional umbrella sampling (US) simulations were employed to estimate the free energy, or potentials of mean force (PMF) along the distance in the z-direction of the center of mass (COM) of the HZ molecule from COM of the bilayer. This variable describes the transfer of the HZ molecule from the DMPC bilayer to the aqueous phase, and also from one bilayer/solvent interface to the center of the bilayer.

The range of the z-coordinate from 0 (corresponding to the center of the bilayer) to 30 Å (which is already far in the aqueous phase) was divided into 61 windows spaced by 0.5 Å. Two series of US simulations were performed, with different starting coordinates in all of the windows. In the first set, the starting structures were created by pulling COM of HZ placed initially in the center of the bilayer (I-state) along the z-direction. In the second set, the initial pulling simulation was started with the HZ molecule placed at the edge of the hydrophobic tail region parallel to the surface (S-state). The initial structures of the I- and S-states were

taken from the output of the previously performed free MD simulations. Harmonic biasing potential with a force constant of 10 kJ mol^{-1} Å^{-2} was applied, and the simulation in each of the windows ran for 500 ns. The Gromacs `g_wham` tool was used to analyze the trajectories and construct the PMF.

3.3 Results and Discussion

3.3.1 Orientation of HZ from MD simulations

All-atom MD simulations were performed for one HZ molecule interacting with a DMPC bilayer. 15 independent trajectories over 1 μs were generated, using different initial conformations, positions and orientations of HZ in the bilayer. Some of the simulations employed distance or dihedral restraints to force a given secondary structure of the HZ molecule. The HZ molecule was always found embedded into the lipid bilayer in every simulation, and the dominant orientation of observed in each of the simulations is presented in Tab. 3.2.

		Initial position of HZ		
		in solvent	in bilayer	in bilayer
			HZ axis to bilayer	plane initially
Initial conformation	Restraints		perpendicular	parallel
α-helix	α	S	I	I
3_{10}-helix	3_{10}	I	I	I
β-bend ribbon	β-bend	I	I	I
fully extended	none	S	S	I
3_{10}-helix	none	**S**	S	**I**

Table 3.2 Resulting orientation of HZ in each of the individual free MD simulations performed with the different initial conditions. S – surface-bound orientation with HZ at the inner edge of the bilayer surface and the HZ axis parallel to the surface; I – transmembrane orientation with HZ axis perpendicular to the bilayer surface; bold – MD trajectories considered in the following analyses.

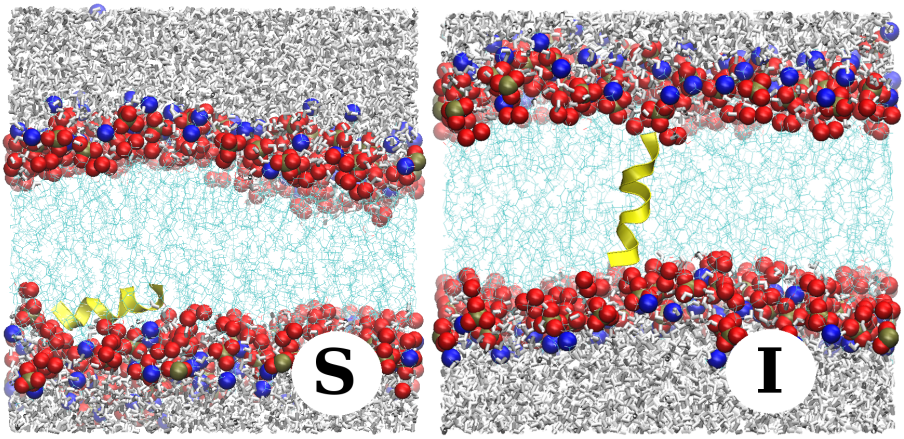


Fig. 3.1 Selected snapshots from free MD simulation illustrating the two distinct orientation of the HZ molecule in a lipid bilayer that were observed: S-state (parallel to the surface, at its inner edge), I-state (transmembrane, along the bilayer normal).

A striking observation is that each of the many simulations started from different initial conditions converged to one of just two distinct structures. In what follows, these will be designated as the *S-state* – the HZ molecule is in the interfacial layer between the non-polar and head-group regions of the bilayer, with its main axis parallel to the bilayer surface – and the *I-state* – the HZ molecule spans most of the hydrophobic bilayer width, with its main axis perpendicular to the bilayer surface; the representative instances of the S- and I-states are shown in Fig. 3.1.

It is also interesting that the outcome of MD simulations is different whenever the conformation of the HZ molecule is restrained to a certain structure by means of distance restraints imposed on H-bonds or backbone dihedral angles. Specifically, all of the simulations with HZ restrained to 3_{10}-helical or β-bend conformation actually led to an I-state orientation of HZ, no matter what the initial orientation in the simulation was. Thus, it seems that the flexibility of the HZ molecule may play a certain role either in the relative propensities of the different orientations, or in the process of insertion of HZ into the membrane.

All of the structural analyses below are based on the trajectories from two of the performed MD simulations, shown in bold in Tab. 3.2; these were unrestrained simulations with the HZ molecule in a 3_{10}-helical structure initially, placed either inside the bilayer perpendicular to its surface, which led to the I-state, or outside of the membrane and subjected to a heating–cooling protocol, which provided the S-state. Additionally, a simulation of a hypothetical charged mutant of HZ in an orientation resembling the S-state have been performed. This molecule, HZ4+ has Ile3, Ile4, Leu7 and Leu8 replaced by lysines and carries four positive charges. To analyze the location of the peptide molecule within the bilayer in the different states in more detail, one-dimensional densities of atoms across the bilayer were obtained

from MD trajectories, see Fig. 3.2. Also, the flexibility of conformation of the HZ molecule was quantified by means of the root-mean-squared fluctuation, Fig. 3.3; no noticeable features or differences between the I- and S-states were found.

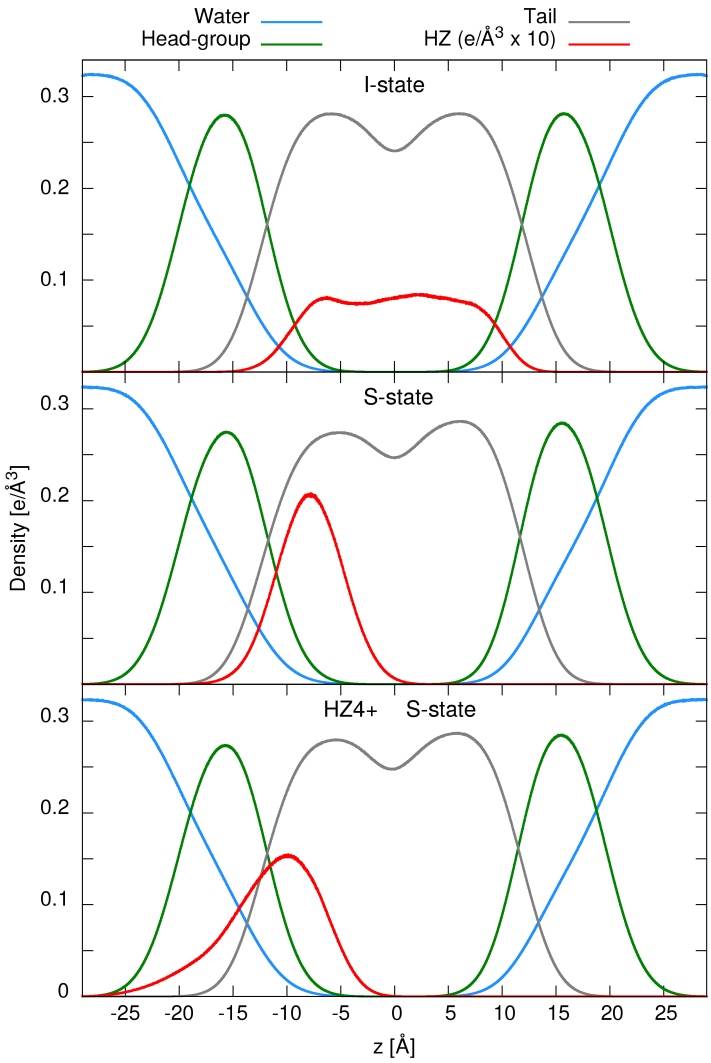


Fig. 3.2 Density profiles along the direction perpendicular to the bilayer plane, corresponding to the HZ molecule, water as well as the non-polar tails and polar head groups of the lipid. Electron density per unit volume (e/Å^3) is plotted for the I- and S-states of HZ as well as the S-state of the charged mutant HZ4+.

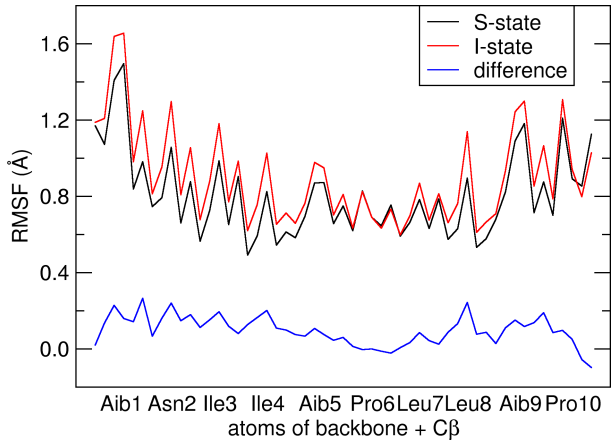


Fig. 3.3 Root mean square fluctuation (RMSF) of the individual atoms of backbones and Cβ of the HZ molecule in the S- and I-states, obtained from free MD simulations.

I-state

The HZ molecule is oriented with its main axis along the bilayer normal direction, as reflected by the tilt angle fluctuating close to zero, of 22 ± 13 degrees (mean $\pm$ std. dev.). The rather broad distribution may be due to the flexibility of the structure of the molecule, which affects the helical axis direction within its simple definition employed here. HZ spans the region of hydrophobic lipid tails symmetrically, and the displacement of the COM of HZ from the center of the bilayer is 0.0 ± 1.7 Å (mean $\pm$ std. dev.). Thus, HZ is only slightly mobile along the direction perpendicular to the bilayer surface.

One of the striking properties of the HZ molecule is its small length, and a meaningful question to ask is if and how such a short molecule may span the entire width of a lipid bilayer. Notably, the considered DMPC bilayer is a rather thin one: its hydrophobic thickness of ca. 25 Å in the current simulations is well in the range of 23–27 Å found in the experimental literature.[149–151] In the simplest idea of the peptide–bilayer interaction, the length of a hydrophobic peptide should match the hydrophobic thickness of the bilayer, in order to span the bilayer with no need for unfavourable interactions. To assess such a match, the length of the HZ molecule represented by the distance of the atoms C$_\alpha$ of Aib1 and Leuol11 was measured. With the length of 17.6 ± 1.3 Å, HZ is only slightly longer in the I-state than in the S-state (16.3 ± 1.3 Å, mean $\pm$ std. dev.); the corresponding distributions are shown in Fig. 3.4. Apparently, the HZ molecule is too short to match the hydrophobic region of even such a thin bilayer as DMPC when immersed in the I-state.

Still, the polar termini of HZ might reach into the head-group region of the bilayer if the bilayer is thinner in the vicinity of the HZ molecule; this would not be visible in the averaged density profiles in Fig. 3.2. To test this possibility, vertical density profiles were resolved

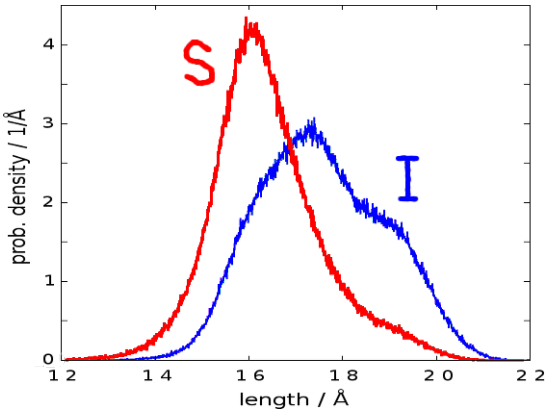


Fig. 3.4 Distribution of end-to-end length of the HZ molecule measured as the distance of the C_α of the first AA (Aib1) and the C_α of the last AA (Leuol11) in the MD trajectories of the S and I orientations.

by the horizontal distance from the COM of the HZ molecule, see Fig. 3.5. At distances above 15 Å, the thickness of the hydrocarbon tail region of 24 to 25 Å corresponds to the unperturbed value. In the range of 5–10 Å from the HZ molecule, however, the observed thickness decreased locally to ca. 20 Å, and is similar in the range below 5 Å. Thus, the hydrophobic region of the bilayer becomes thinner considerably in the vicinity of the HZ molecule. The termini of the HZ molecule in the I-state are capable of reaching into the lipid head-group region and may participate in favorable interactions like H-bonding.

Previously, it was reported that a rather thick D(22:1)PC bilayer thinned with added ALM, the long-sequence peptaibol [152], and a POPC bilayer thinned considerably in the vicinity of an inserted molecule of the lipopeptaibol trichogin GA IV [131]. In the current study, we have yet a different situation: there is a short-sequence peptaibol inserted into an inherently thin DMPC bilayer. Still, local thinning is observed, as an effort to reach a hydrophobic match of the inserted peptide and the hydrophobic thickness of the lipid bilayer.

S-state

A distinct observation is that the HZ molecule resides completely within the hydrophobic region of the bilayer, formed by the hydrophobic tails of the DMPC molecules. This is at odds with the usual behavior of amphiphilic membrane-active peptides, which often exhibit a strong interaction with the interface between the non-polar and the polar regions due to the presence of one or more cationic amino acids [153]. Such an effect is also observed in this study for the cationic mutant of HZ carrying four positive charges, HZ4+. Compared with HZ, HZ4+ is located much further from the center of the bilayer, roughly at the polar/non-polar interface. This is reflected by the observed distance of HZ4+ from the center of the bilayer

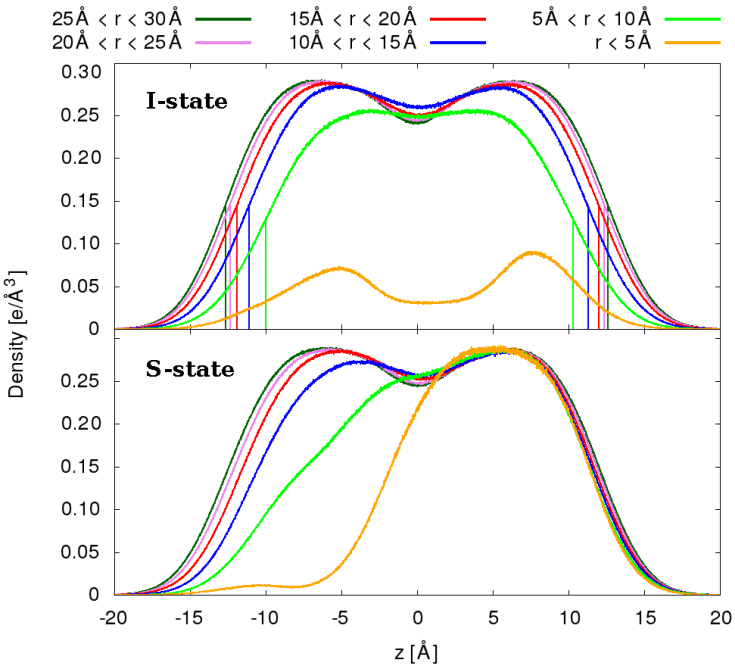


Fig. 3.5 Density profiles of the non-polar hydrocarbon tails of DMPC in the simulation of HZ in the I-state (top), and S-state (bottom). The different profiles were obtained considering lipid atoms in the given ranges of horizontal distances from the center-of-mass of the HZ molecule. (Horizontal distance is the distance projected in the plane of bilayer surface.)

of (11.5 ± 2.7) Å being significantly larger than (7.9 ± 1.7) Å for the uncharged HZ (mean $\pm$ std. dev.). Let us rephrase this important observation: HZ is a strongly non-polar peptide with no charged groups and few H-bonding donors or acceptors, and it is immersed nearly completely within the hydrophobic region of the bilayer, instead of interacting with the polar head-groups region in the way cationic peptides do.

Another structural feature of HZ is the orientation of the individual AA side chains, which describes the rotational state of HZ in the S-state along its helical axis. To this end, the z-cordinate of the terminal heavy atom of each of the AA side chains (relative to the z-coordinate of the COM of the HZ molecule) was calculated. These data, presented in Fig. 3.6, show which AA side chains point towards the hydrophobic volume preferentially, and which tend to interact with polar interface. Three of the five non-polar side chains (Ile4, Leu7 and Leu8) point towards the hydrophobic membrane region on average, and the two remaining ones (Leuol11 and Ile5) are rather indifferent. By contrast, the only polar side chain (Asn2) points clearly towards the head groups of the lipid. Note that the system is dynamic, and it would be wrong to expect all of the AA side chains to assume the described orientation for 100 % of the time. On the other hand, the clear shape of the distributions obtained from an MD trajectory of 1 μs indicates that the result may be considered converged.

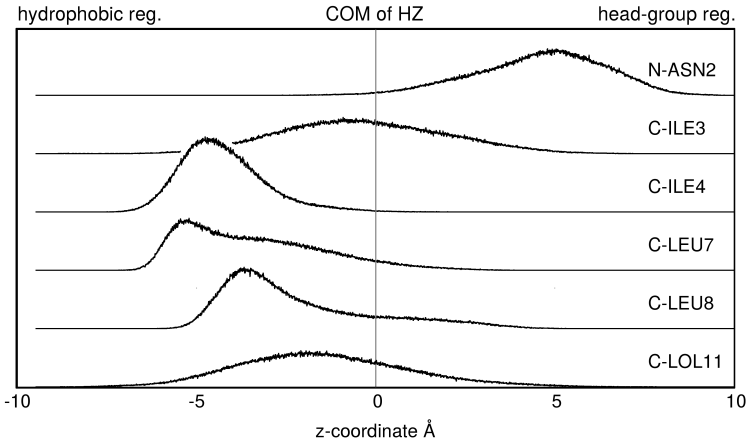


Fig. 3.6 z-coordinate of the terminal heavy atoms of the AA side chains except Aib and Pro. The value $z = 0$ corresponds to the center of mass of the HZ molecule; the side chain points towards the polar lipid head groups if $z > 0$, while it points into the hydrophobic tail region if $z < 0$.

3.3.2 Structure of HZ from MD simulations

The structural ensembles of HZ were assessed. Besides the general goal, to characterize the conformation of the molecule, a specific question was whether HZ assumes a β-bend ribbon (spiral), which can be considered a subclass of the 3_{10}-helical family. To this end, a structural analysis was performed for the MD trajectories of HZ inserted in the DMPC bilayer both in the S-state and the I-state. A Ramachandran plot was obtained separately for each of the AAs 3 through 10 in both MD trajectories, see Fig. 3.7, which also shows the averaged structures of HZ from the unrestrained simulations in the S- as well as I-states.

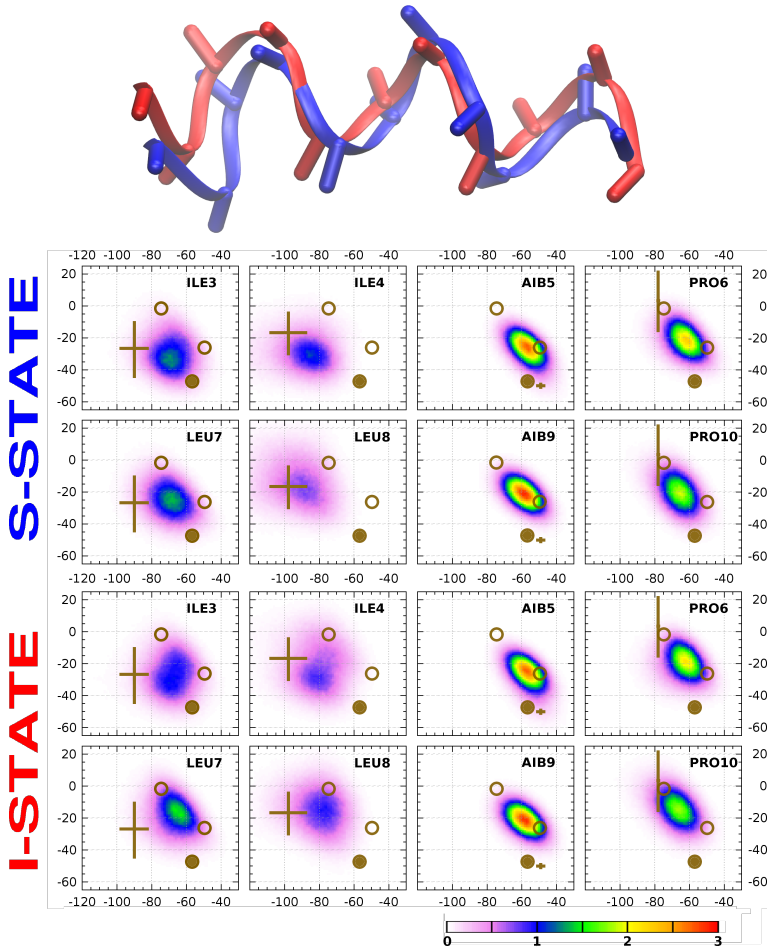


Fig. 3.7 Top: Dominant conformations of the HZ molecule in the S-state (blue) and the I-state (red). Structures were obtained with a PCA procedure run on MD trajectories, described below. Bottom: Ramachandran plots for the amino acids 3 through 10 of the HZ molecule interacting with a DMPC bilayer, in the S-state and in the I-state. Usual convention employed (horizontal axis – ϕ, vertical axis – ψ, both in degrees); empty circles – 3_{10}-helical conformations with $(-74°, -4°)$ and $(-49°, -26°)$; solid circles – α-helical conformation with $(-57°, -47°)$; locations and sizes of crosses – the ranges reported for β-bend ribbons in Ref. 25. See the Supporting Information for numerical values of (φ, ψ) of the dominant conformations.

The distributions of (ϕ, ψ) of both Leu and of both Ile are rather broad, located in the helical region of the plot. There is also a certain overlap with the typical β-bend values; note here the large fluctuation of the values reported in Ref. 25, reprinted in Tab. 3.1. The (ϕ, ψ) distributions for Aib and Pro are more distinct, located along the line typical for 3_{10}-helices, and there is also a certain match with the β-bend for Pro and a modest one for Aib. In short, no clear conclusion on the conformation of HZ is possible based on the analysis of Ramachandran plots; it may only be stated that the conformations of HZ in the S-

and I-states seem to be similar. The structures of HZ accumulated along the MD trajectories were also compared to static, idealized conformations by means of the root-mean-squared displacements of atomic positions (RMSD), see table 3.3 (appendix A for the time evolution). The idealized structures were $(\varphi,\psi) = (-74^\circ, -4^\circ)$ for 3_{10}, $(\varphi,\psi) = (-57^\circ, -47^\circ)$ for α, and considering the values of angles from Ref. 25 for β-bend ribbon. Only the fragment of the HZ molecule that may form a β-bend ribbon, consisting of the AAs Ile3–Pro10, was considered for this analysis, which was further restricted to the heavy atoms of the backbone.

reference	S-state	I-state
3_{10}-helix	1.25 ± 0.27	0.93 ± 0.30
β-bend ribbon	1.07 ± 0.24	1.10 ± 0.26
α-helix	1.40 ± 0.24	1.51 ± 0.28
average structure[a]	0.52 ± 0.20	0.58 ± 0.20

Table 3.3 Root-mean-squared deviation (RMSD, in Å) of atomic positions of the HZ molecule with respect to selected reference structures, obtained from the MD trajectories with HZ molecule in the S and I orientations. The calculation of RMSD involved non-hydrogen atoms of the backbones of AAs 3 through 10. [a] Average structure taken from each respective MD trajectory.

The RMSD values with respect to β-bend ribbon as well as 3_{10}-helix are low, indicating a large similarity to the idealized structures, while α-helical conformation is less likely.

Just one conformational element that differs between HZ in the S- and I-states is apparent in the Ramachandran plots as well as visually in the averaged structures. The Ramachandran plots show ψ of Leu7 and ϕ of Leu8 to be shifted clearly if slightly, by ca. 15°. Accordingly, the average structures of HZ in the S- and I-states differ at the position 7–8. Apart from this particular feature, no other structural distinctions are apparent, and the conformations of both the up- and down-stream segments are hardly distinguishable visually.

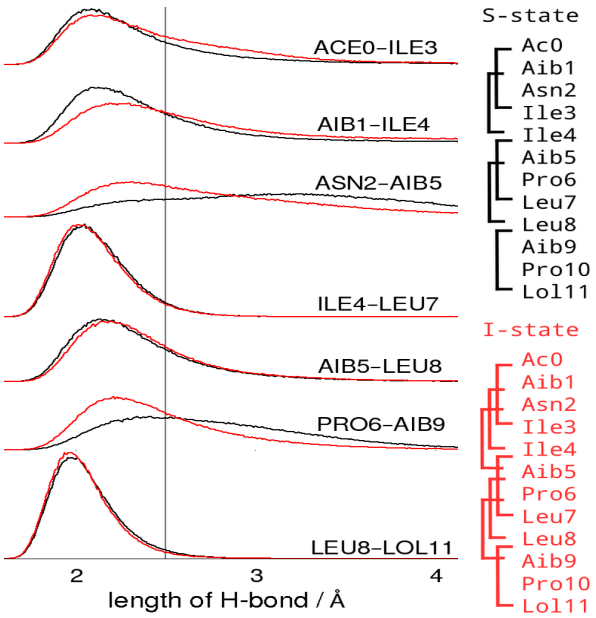


Fig. 3.8 Left: Distribution of lengths of putative backbone hydrogen bonds in HZ, corresponding to the 3_{10}-helical conformation, in the S-state (black) and in the I-state (red). ResX–ResY is the distance between the carboxy-O atom of ResX and the imino-H atom of ResY. Vertical line at 2.5 Å is a guide to the eye. Right: H-bonds present in the HZ molecule in the S and I-states.

Additional information on the structure of the HZ helix may be provided by the analysis of intra-backbone H-bonding patterns. To this end, the lengths of putative backbone H-bonds typical of a 3_{10}-helical conformation were measured, and their distributions were obtained, see Fig. 3.8 (left). Also here, there are very similar pictures of H-bonding for the HZ molecule in the S and I-states. The most of the monitored putative H-bonds exhibit a distribution of length located close to 2 Å, showing good, stable H-bonding. However, two of the putative H-bonds, Asn2 $\leftarrow$ Aib5 and Pro6 $\leftarrow$ Aib9, show distributions that are shifted to longer distances in the S-state, meaning that these H-bonds cease to exist over certain periods of the respective MD trajectory. A qualitative picture of H-bonding in the HZ molecule in the S- and I-states that follows is presented in Fig. 3.8 (right).

In the I-state, there is a continuous pattern of $i \leftarrow i+3$ H-bonding along the entire HZ molecule, except the cases where a proline cannot be a hydrogen donor. By contrast, two more H-bonds are missing in HZ in the S-state: Asn2 $\leftarrow$ Aib5 and Pro6 $\leftarrow$ Aib9. Strikingly, this observation correlates with the idea of two possible β-bends of the type Xaa-Yaa-Aib-Pro stretching over the residues Ile3–Pro6 and Leu7–Pro10. These two bends might constitute rather independent structural elements, with a flexible mutual arrangement. Thus, a β-bend ribbon conformation is possible for the HZ molecule in the S-state, while the conformation is rather 3_{10}-helical in the I-state.

To summarize, the structure of HZ immersed in a DMPC bilayer was characterized. The interpretation shall concentrate on the structure in the S-state, which was revealed as the more likely orientation by the experimental study. The I-state, which may be an intermediate in, e.g., the oligomerization of HZ, shall also serve as a reference to accessible structural space. The Ramachandran analysis corroborated what a simple visual inspection indicated, namely that the molecule assumes a structure in the 3_{10}-helical family, and a tendency to form a β-bend ribbon was confirmed by the analysis of H-bonding. The Leu7–Leu8 segment was shown to be flexible: While it assumes a rather stretched conformation in the (less likely) I-state orientation, it is bent slightly in the dominant S-state orientation.

3.3.3 Convergence of HZ structure

The sampling of structural ensembles is a persistent issue in MD simulations of lipid bilayer systems. Therefore, a procedure based on the principal component analysis (PCA) was employed to examine the convergence in the MD trajectories of HZ in the S- and I-states. Then, each of the two trajectories was projected into a two-dimensional space spanned by the first (PC1) and second eigenvectors (PC2) obtained from the PCA, see Fig. 3.9 for the projections. Note that the combination of PC1 and PC2 covers 72 % of structural variation of HZ in the structural ensemble combined from the simulations of S- and I-states. Also, we introduce a transformation of PC1 and PC2 into polar coordinates, as depicted in Fig. 3.9 also.

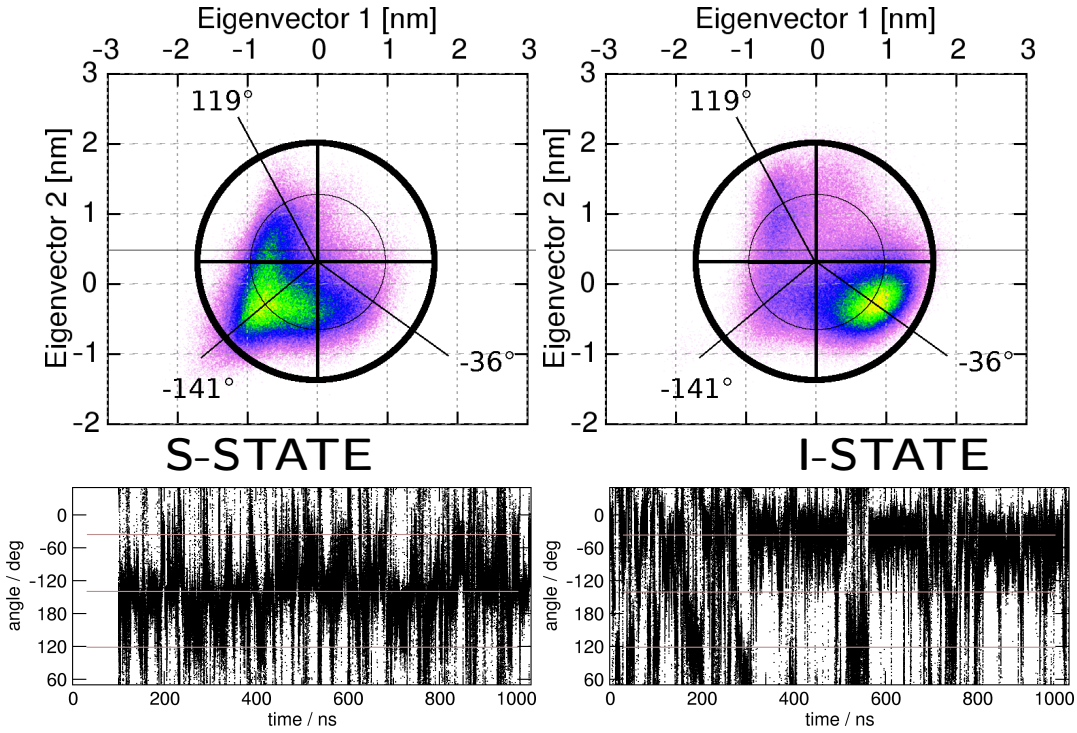


Fig. 3.9 Results from the principal component analysis. Top: Projection of each trajectory onto the first two eigenvectors, which were obtained from a PCA on the combined trajectories for the S- and I-states; Definition of the angular coordinate composed of PC1 and PC2, which makes it possible to distinguish between three structural states. Bottom: Time course of the angular coordinate in each of the trajectories.

The PCA plots reveal three different structural domains. Then, the structure at -141° is dominant in the S-state, and the structure at -36° dominates in the I-state, while the third structure at 119° is minor in both states. We use the angular coordinate obtained from PC1 and PC2 to assess the convergence of HZ structure in the free MD simulations performed.

On the assumption that three structural domains are possible, as identified above, it shall be checked whether interconvertions between them are observed during the simulation period. To this end, the angular coordinate is plotted along the respective MD trajectories in Fig. 3.9 (bottom). For the S-state, there are a number of transitions between -141° and -36°, and 119° is populated transiently. The histogram of PC1/PC2 for the I-state features more clearly separated maxima, and the transitions between them are correspondingly more scarce; still, the dominantly populated -36° is interrupted by interconversions to -141° and 119° on multiple occasions. These observations confirm that the structural ensembles of HZ in the bilayer in both the S- and I-states are largely converged on the simulated time scale of 1 μs.

3.3.4 Free energies from replica exchange simulations

The HREX method was employed to determine the relative free energies of the different orientations of the HZ molecule in a DMPC bilayer. Within the HREX framework, the force field contributions localized in the HZ molecule were scaled down in all but one of the eight replicas of the molecular complex. Two HREX simulations were performed differing in the initial position of the HZ molecule.

In the fist HREX simulation, the starting structure in all replicas had the HZ molecule inside the bilayer, perpendicular to the bilayer plane, i.e., close to the I-state, and the HZ molecule remained within the bilayer for the entire simulation time of 2 μs. The resulting surface of free energy as a function of two variables, the z-coordinate of the COM of HZ relative to the COM of bilayer, and the tilt angle between the long axis of HZ and the z-axis, is presented in Fig. 3.10.

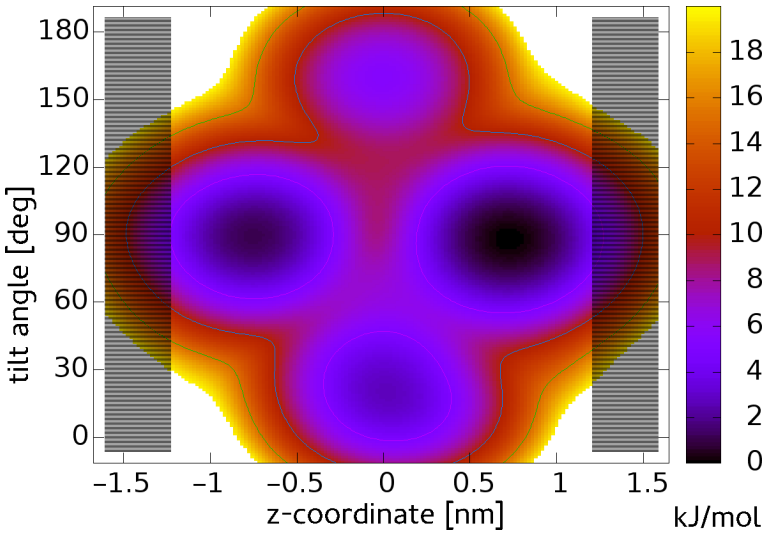


Fig. 3.10 Free energy surface from the HREX simulation with HZ molecule placed in the hydrocarbon tail region initially. Grey shaded area – region of polar lipid head groups.

A pair of each of the two structures, S- and I-states, observed in free simulations are visible in the free energy plot clearly. Both of the S-states at $z \approx \pm 8$ Å and tilt angle of ca. 90° are equally deep global minima of energy; they are set to zero free energy in Fig. 3.10. The I-states at $z = 0$ and tilt of 15° and 165° are secondary minima, which lie higher in energy than those corresponding to the S-states. Since both leaflets of the bilayer are equivalent, the I-state minima should be equally deep much like the S-state minima, therefore we consider this feature of the plot not converged. An achievement to be appreciated is that the global minima have been reproduced, and the secondary minima are present and higher in energy. Several interconversion between the S- and I-states are observed. This seems to be enough to

provide a qualitatively correct free energy landscape but even longer time scales would be required to obtain a statistically converged result.

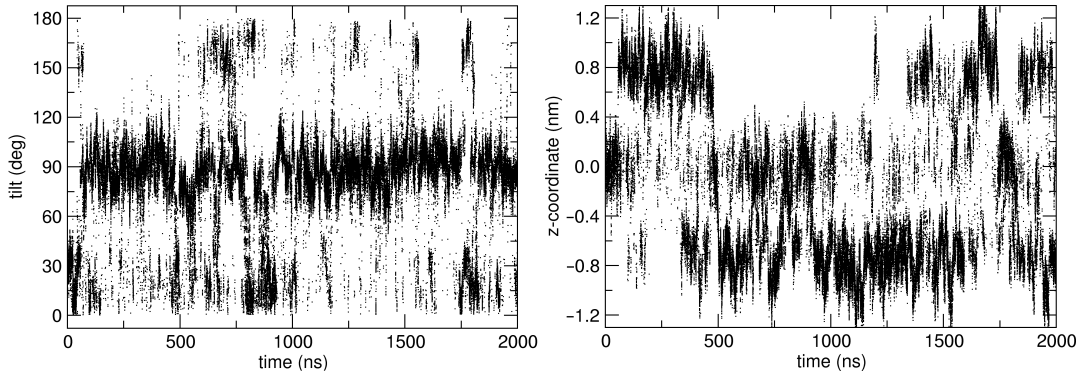


Fig. 3.11 The collective variables that describe the orientation of HZ in the DMPC bilayer, followed along the unperturbed replica ($\lambda = 1$) in the HREX simulation. The S-state is characterized by a tilt of 90° and z-coordinate of ca. $+0.8$ or -0.8 nm; the I-state exhibits a tilt close to 180° or $-180°$ and z-coordinate of ca. 0.0 nm.

While the current simulation of 2 μs already sampled several interconversions between the S- and I-states (Fig. 3.11), an extension of the simulation would be necessary in order to have the secondary minima converged as well. Conclusion that may be made at this point is that both the S-state and the I-state have been confirmed as the relevant structures, with the S-state being the most likely structure, in accordance with ssNMR and SRCD experiments, performed by Ulrich's group [154, 155].

A technical point worth a discussion is the efficiency of sampling with the current HREX simulations. Recall that the difficult convergence is due to the very slow rearrangements of lipid hydrocarbon tails, leading to very slow equilibration in MD simulations. The idea of HREX is to reduce energy barriers by way of scaling down appropriate contributions to the force field energy of the molecular system. Now, it is not obvious if the scaling of interactions involving a non-polar molecule like HZ (with, prominently, non-polar lipid components) may improve the rate of sampling or not. A typical application of HREX would rely on the weakening of polar interactions or rotational barriers, which both do not apply in the current study. This may be the reason for the failure of HREX to provide largely improved sampling efficiency.

The second HREX simulation of 1 μs was performed with the HZ molecule placed in the aqueous phase initially, aiming at an estimate of the free energy profile of the HZ molecule passing from the solution to the bilayer. The resulting free energy surface (FES) is shown in Fig. 3.12. Evidently, the deepest free energy minima are found in the aqueous phase, on both sides of the lipid bilayer. A minor minimum is found inside of the bilayer, corresponding to

the S-state. In spite of the extended length of the simulations of 1 μs, convergence has not been reached apparently, for two reasons: Firstly, the free energy plot should be symmetric, meaning that if one of the S-states (with $z \approx 8$ Å) was found, the other one ($z \approx -8$ Å) would be expected as well. And, more importantly, the HZ molecule in the aqueous phase is predicted as the most likely state, which is entirely unexpected considering the purely non-polar character of HZ.

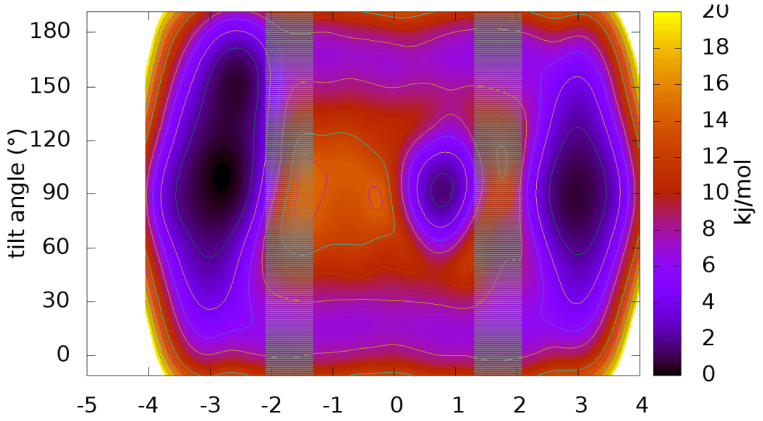


Fig. 3.12 Free energy surface from the HREX simulation with HZ molecule placed in the aqueous phase initially.

It seems that there is a rather high energy barrier opposing the insertion of HZ into the lipid bilayer, through the strongly polar lipid head group region. Then, the HZ molecule only passed into the interior of the bilayer once, in a single HREX replica, and the free energy minimum at $z \approx 8$ Å is only shallow. Thus, it seems that obtaining convergence in such a simulation is far beyond the possibilities of current HREX methodology and computational equipment. The event of insertion that was observed may still constitute a starting point of a follow-up study of the mechanism of insertion. For the time being, we can merely state that the HZ molecule was inserted in the bilayer with its Leuol-terminus first. It has to be noted again that this event occurred once only, and took an amount of time that extended over several successful replica exchange instances; in other words, it took place in a simulation running with artificial and variant potential energy function.

3.3.5 Free energies from metadynamics simulations

Relative free energies of the various orientations of the peptide molecule in the bilayer represent some of the most desirable quantities in a study like this. Unfortunately, this goal is difficult to achieve even with extended sampling techniques. This is best illustrated by our application of metadynamics simulation, with the intention to reach convergence of

free energies in a two-dimensional representation. The orientation of the HZ molecule is described with the z-coordinate and the tilt angle, much the same as in the HREX simulations. In recent work by other authors, metadynamics simulations have been successful in the reconstruction of water/membrane partitioning free energy surfaces, in particular for small solutes [156, 157].

The same kind of calculation with a peptide, even a small one like HZ, in an atomistic lipid-water system is more challenging, and we are going to describe the problems that we were confronted with in these calculations. All of the attempted metadynamics simulations were either unstable or inefficient. Almost all of the normal metadynamics simulations crashed after ca. 100 ns. At this point, the underlying free energy surface was already filled with biasing Gaussians largely, and in fact, the motion along the collective variables (CV) became diffusive, see e.g. Fig. 3.13. A possible cause of the problem may lie in the singularities of the tilt angle at 0° and 180°; the diffusive dynamics along tilt may have led to an accelerated motion along the tilt CV and instabilities whenever the tilt became close to 0° or 180°. Unfortunately, even an analysis performed for the period of simulation preceding the crash never produced any reliable free energy surfaces.

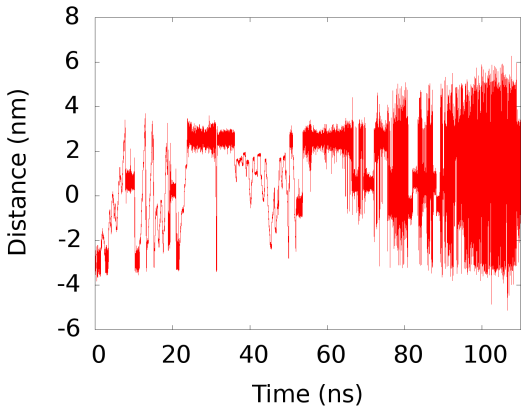


Fig. 3.13 Motion of the z CV during one of the normal metadynamics simulations of unrestrained HZ in a DMPC bilayer. The motion becomes diffusive after ca. 70 ns.

The well-tempered metadynamics simulations behaved somewhat better, but still no convergence was reached. The bias deposition rate decreased over simulation time, and probably the bias factors used were too low, so that the system was unable to explore all of the FES. Changing the bias factor value from 6 to 10 did not bring any improvement, and it is difficult to choose the suitable bias factor because the relevant barrier heights are unknown. Still, the best results from the many attempts exhibited two minima associable to the S- and I-states, see Fig. 3.14.

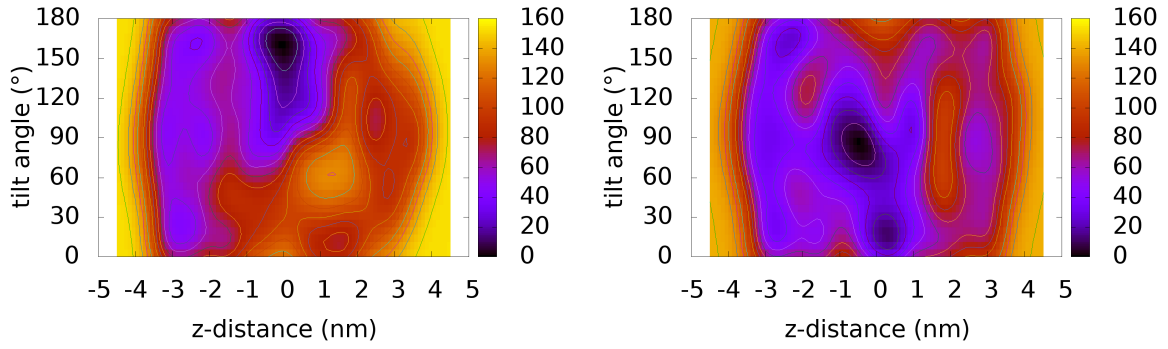


Fig. 3.14 Free energy surface from metadynamics well-tempered simulations with HZ molecule 3_{10}-helix restrained. The bias factor was set to 6 (left) and 10 (right).

There might be various reasons for such unsatisfactory results. The first concerns the representation of the two CVs. The tilt angle as defined here exhibits singularities at 0° and 180°. While this is no problem in a post-processing of a free or HREX simulation, it might become critical in a simulation with biasing potentials like metadynamics. A somewhat better choice might be the cosine of the tilt angle, and our further studies go in that direction. passing to the z-distance, this was defined relative to COM of all of the lipid. Such a definition may clearly lead to artifacts due to the undulations of the bilayer, which generate local deformations. Then, the COM of the entire bilayer does not describe the local neighborhood of the HZ molecule effectively. A better definition might be based on the lipid atoms within a cylinder aligned along the bilayer normal and centered on the solute [158].

As mentioned above, even though the well-tempered simulations did not reach convergence, were be able to sample two minima of the FES, Surprisingly, only the peptide restrained to 3_{10}-helical conformation entered the hydrophobic region of the bilayer and really reached these minima. This means that the conformation of the peptide may be another important degree of freedom. Importantly, a dangerous caveat of metadynamics is that whenever such degrees of freedom, which are orthogonal to the biased CVs, are not sampled properly then the convergence becomes difficult, possibly leading to incorrect free energies.

In addition, another important phenomenon that is uncontrolled in the metadynamics simulation is the deformation or disruption of the bilayer that takes place as the peptide is being transferred from the aqueous phase to the interior of the bilayer. It seems to be very difficult to define a CV to control such structural changes, while it would be clearly orthogonal to the existing CVs, in the metadynamics language. A possible solution to that issue could be to combine metadynamics (with the present CVs) with HREX.

3.3.6 Free energies from umbrella sampling simulations

In another attempt to estimate free energies, we turned to a one-dimensional representation and performed umbrella sampling simulations. The aim was to obtain the free energy profile for the HZ molecule approaching the center of the lipid bilayer and, on the other hand, being pulled out across the polar head-group layer into the aqueous solvent. The reaction coordinate was represented by the difference of the z-coordinate of COM of HZ and the z-coordinate of COM of the lipid bilayer. Two series of simulations were performed, each starting in on of the previously identified stable orientations of HZ, the S- and I-states. For each of the series, the reaction coordinate was divided into 60 windows spaced by 0.5 Å, and the starting structures for each of the windows were created from the respective stable state (S or I), by means of a rather fast non-equilibrium pulling simulation. Thus, the starting structures in each of windows were different in the S- and I-series of US simulations. It will turn out that this is an important point.

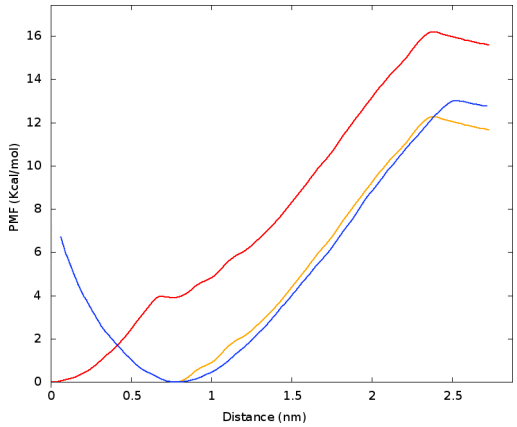


Fig. 3.15 Free energy profiles obtained from 1D umbrella sampling simulations of pulling HZ out of the bilayer into the aqueous phase. Starting structures generated by pulling simulations from two different structures; Red – starting from the I-state; orange – the same data, shifted vertically; blue – starting from the S-state.

The free energy plots shown in Fig. 3.15 originate in series of US simulations of 500 ns per window, or 30 μs in total, for each of the I- and S-series, Apparently, even such an extensive sampling is insufficient to obtain a result that would be converged in the hydrocarbon tail region of $z = 0 - 1$ Å. The reason is that the orientation of the HZ molecule hardly changes during a simulation in a given window, and no interconversions between the S- and I-states are observed. Thus, each of the two series of US simulations provide its respective starting structure as the more stable state – just because the respective other state is not sampled correctly.

On the other hand, both free energy profiles are very similar in the range of reaction coordinate above 10 Å. Seemingly, this provides a certain confidence to a rough estimate of the free energy cost of pulling the HZ molecule out of the membrane into the aqueous phase, which is around 11–13 kcal/mol on the basis of the US simulations performed. It would little surprise that such a largely non-polar molecule like HZ prefers the non-polar environment of the hydrocarbon tails to the polar solvent strictly. However, there might be a very similar problem as in the above reported metadynamics simulations – the deformation of the bilayer possibly leading to poor convergence, and thus to wrong free energies and hysteresis.

3.4 Discussion and summary

We simulated a molecule of harzianin HK VI (HZ), a short-sequence peptaibol interacting with a DMPC bilayer. The goal was to assess the structure and orientation of HZ in the bilayer on the basis of unrestrained MD simulations. The simulations were complemented with ssNMR and SRCD experiments to facilitate a decision on the orientation. Further, it was attempted to estimate free energies from extended sampling simulations.

Two possible orientations of HZ in the bilayer were identified on the basis of free MD simulations: the I-state, with HZ perpendicular to the bilayer surface, and the S-state, where HZ lies in the hydrophobic part of the bilayer, close and parallel to its surface. The ensemble of simulations that were performed involved largely varying initial structures and orientations. In spite of that, each of the simulations performed ended up in one of those two states. However, the sampling achieved in such free MD simulations was by far too insufficient to decide which of the orientations is preferred.

HREX simulations were performed to enhance the sampling of the different orientations of HZ in the bilayer. Even with this modern technique, however, it was still impossible to obtain a number of interconversions between the S- and I-states that would be sufficient to obtain conclusive results in terms of converged populations (or, equivalently, free energies). Calculation of free energies with extended metadynamics and umbrella sampling simulations turned out to be even less practicable.

The issue of unclear dominant orientation was resolved by means of ssNMR and SRCD experiments, performed by Ulrich's group [154, 155]. Their analysis revealed that the HZ molecule is embedded in the surface of the bilayer preferably, i.e., it assumes the S-state. On the other hand, it is conceivable that the I-state may act as an important, low-energy intermediate in the insertion into bilayer, transfer between the bilayer leaflets, or in a process of aggregation of HZ into oligomers at conditions of higher P/L. Therefore, the following analysis of structure addressed both the S- and the I-state.

The structure of HZ, for which no hard experimental evidence is available, was characterized on the basis of free MD simulations. These were converged with respect to HZ structure, as confirmed by an analysis based on PCA. Ramachandran plots were inspected, and intra-backbone H-bonding patterns were investigated. The conformation of the HZ molecule in both the S- and I-state orientations falls into the 3_{10}-helical family. Apparently, the HZ molecule is somewhat longer along the helical axis in the I-state than in the S-state. This goes to the account of the slightly different conformation of the Leu7–Leu8 moiety.

In the more relevant S-state, the HZ molecule is located completely within the hydrophobic region of the bilayer, in contrast to the behavior of cationic amphiphilic peptides which usually dock into the hydrophobic/hydrophilic interface. In other words, HZ is immersed deeper into the bilayer than an amphiphilic peptide of comparable size. In addition, the HZ molecule in the S-state assumes an oriented rotational state: The polar side chain of Asn2 is pointing toward the head group region, while the non-polar side chains – all of the others except both Aib side chains – are oriented into the interior of the bilayer.

In the I-state, the HZ molecule is too short to match the hydrophobic thickness of an unperturbed DMPC bilayer. However, our analysis of local thickness profiles revealed that the HZ molecule present in the I-state induces a local thinning of the DMPC bilayer. Then, the length of the HZ molecule closely matches the (decreased) thickness of the hydrophobic region of the bilayer in the immediate vicinity of HZ. Therefore, although the S-state is the dominant orientation, the I-state may represent a conceivable, viable intermediate in processes like insertion or oligomerization of HZ.

This study illustrated the grand challenge of simulation of non-polar peptides in lipid bilayers. Extended sampling techniques proved to be of limited use: HREX simulations, and even more so biasing potential methods (metadynamics and umbrella sampling simulations) hardly produced any useful results in spite of the large computational effort taken. Note that extremely low efficiency of umbrella sampling simulations of AMP binding was reported previously [159]. A viable alternative may be an application of a coarse-grained force field, which would require the secondary structure of HZ as an input for the parametrization. The results from the current work could be used for this purpose, and we are currently working on an application for the oligomerization of HZ at elevated P/L.

Chapter 4

Mechanism of Spontaneous Thiol–Disulfide Exchange in Proteins

4.1 Introduction

Thiol–disulfide exchange reactions play important roles in many aspects of cellular function. For a long time, these reactions were thought to only serve to stabilize the protein structure. Recently, however, increasing evidence has suggested that many disulfide bonds have a chemically labile, dynamic nature, and may rearrange upon conformational changes of proteins easily [160, 161].

Overall, the mechanism and catalysis of thiol–disulfide exchange reactions in active sites of proteins were analyzed in depth in a large body of experimental and computational studies [162–164]. It is much less clear how this type of reaction occurs in proteins outside of the active site, and thereby decides on the spontaneous shuffling of disulfide bonds. Single molecule force spectroscopy experiments and molecular simulations showed a spontaneous yet regioselective intramolecular disulfide isomerization in an immunoglobulin domain [165, 166]. Proximity and accessibility were pointed out as the likely primary factors of thiol–disulfide exchange in proteins. Interestingly, CXC motifs were previously shown to mimic the active site of protein disulfide isomerases, suggesting this motif to be well suited to undergo continuous thiol–disulfide exchange [167], as the disulfide bond is less stable than that in the canonical CXXC motif of the thioredoxins.

It seems that to understand the mechanism of spontaneous disulfide shuffling in full detail, knowledge of the microscopic mechanism on the atomic scale is required but missing.

Here, It is presented a study of thiol–disulfide exchange reaction performed by means of extended sampling QM/MM molecular dynamics (MD) simulations. The application of

approximative density-functional method DFTB3 for the quantum-chemical calculations made it possible to reach multi-nanosecond time scales, which turned out to be necessary to achieve semi-quantitative results. First, a representative model system is considered, consisting of a dimethyldisulfide and a methylthiolate molecule, see Fig. 4.1. Free energy surfaces were obtained for the disulfide shuffling reaction taking place both in the gas phase and in an aqueous solution. Then, several model peptides containing three cysteines were considered, of which two are oxidized, forming a disulfide bond, while the remaining one is present in the reduced anionic form. The results are discussed in relation to the disulfide shuffling reaction in proteins, and implications are formulated.

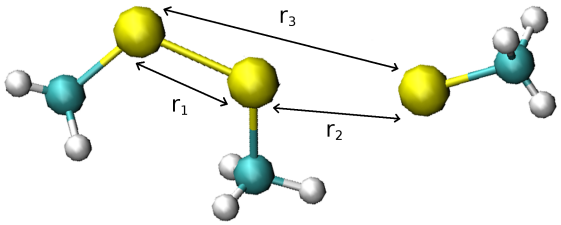


Fig. 4.1 The disulfide exchange reaction. The sulfur–sulfur distances r_1, r_2 and r_3 were considered as the collective variables.

4.2 Materials and Methods

4.2.1 Protein Models

The first step was to simulate a simple model system that supports a disulfide exchange reaction. It consists of one methylthiolate and one dimethyldisulfide, see Fig. 4.1. The simulations were performed in three different environments: gas phase, MM water solvent and partial QM water solvent. In the simulations involving solvent, the system was enclosed in a rectangular periodic box sized ca. $2.5\times2.5\times2.5$ nm^3 and solvated with 482 water molecules. The negative charge of the methylthiolate was neutralized with one sodium ion.

The following simulations were performed with peptides that contained three cysteine residues separated by short runs of one, two or three alanine residues, see Fig. 4.2. Their structures were built with the xLeap tool from the AmberTools package [147]. In the starting structure of each simulation, there was a disulfide bond between the side chains of two of the cysteine residues, while the third cysteine side chain was in the reduced deprotonated form. Each of the peptide systems was solvated with 2895 water molecules and enclosed in a

rectangular periodic box sized ca. 4.5×4.5×4.5 nm^3. The negative charge of the reduced cysteine was neutralized with one sodium ion.

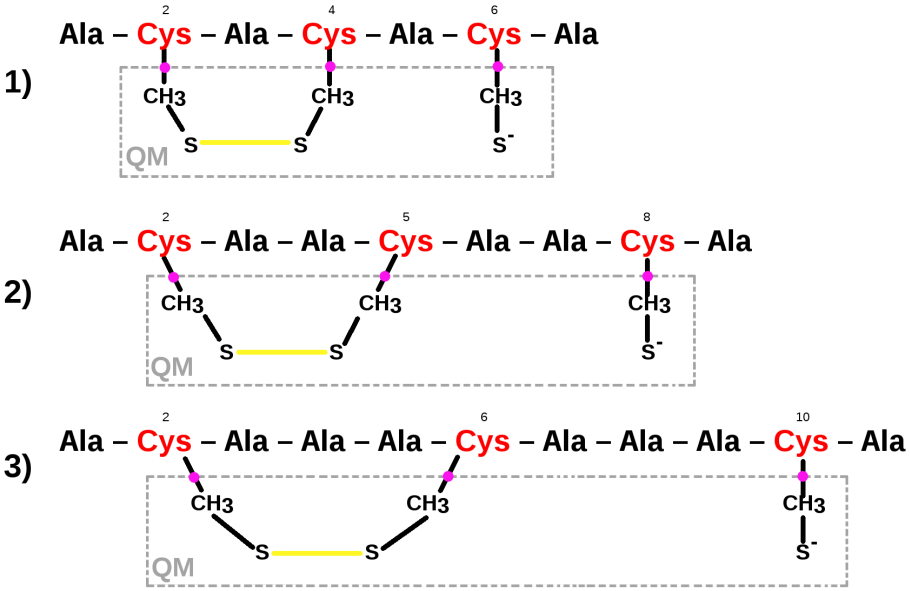


Fig. 4.2 The employed model peptides for disulfide exchange reactions. Boxes – QM regions in QM/MM simulations; pink – link hydrogen atoms.

4.2.2 Simulation Details

All of the simulations were performed with Gromacs 5.0.1 [144, 145] patched with our previous DFTB3 implementation [168] and Plumed 2.1.1 [169, 148]. The gas-phase simulation was performed with the stochastic dynamics integrator with a time step 0.5 fs maintaining a constant temperature of 300 K. There were no periodic boundary conditions and no QM/MM. The simulations of solvated systems were performed with the leap-frog integrator with a time step of 0.5 fs, and periodic boundary conditions were employed. Regarding the MM and QM/MM interactions, the electrostatics were treated with the particle–mesh Ewald method, while the Lennard-Jones were cut-off at 1 nm; the neighbor list was updated every 10 MD steps (5 fs). A constant temperature of 300 K and a constant pressure of 1 bar were maintained, respectively, by means of the Nosé–Hoover thermostat and the Parrinello–Rahman barostat.

In the dimethyldisulfide–methylthiolate simulations, the sulfuric system with the methyl groups was treated quantum chemically (QM). The simulations of disulfide exchange in the gas phase employed pure QM calculation of forces, and the semi-empirical density functional theory method DFTB3 [88] with the 3OB parametrization was employed [170, 171]. The

simulations of disulfide exchange in solvated systems were performed with an additive QM/MM scheme using the electrostatic embedding as implemented in Gromacs. The MM component used the TIP3P representation of water [138] and Åqvist's parameters for the sodium counterion [172]. In the first series of QM/MM simulations, all of the aqueous solvent was treated as MM atoms. Then, the QM region was expanded to include also 30 water molecules; these were the 30 water molecules closest to the solute in the final structure of the simulation with pure MM water. The electrostatic interaction between the negatively charged QM region and the MM system was scaled down by the factor of 0.75 corresponding to the inverse square root of the optical dielectric constant. This is an effective approach to compensate for the missing electronic polarization of the MM environment as recommended by Stuchebrukhov for MD simulations of charged solutes [173].

In order to keep the two different sulfur-containing molecules close together in the solvated simulations, flat-bottomed position restraints were applied to the QM region. These were designed to keep the QM region within the distance of r_{fb} from the center of the simulation box. There are no forces if the QM molecules are up to r_{fb} away from box center, while it is beyond the radius of r_{fb} that the restraint acts as a harmonic potential with a force constant of k_{fb} =1000 kJ·mol^{-1}·nm^{-2}, which pushes the QM atoms back towards the box center,

$$V_{\mathrm{fb}}(\mathbf{r}_i) = \begin{cases} 0 & \text{if } |\mathbf{r}_i - \mathbf{R}_{\mathrm{c}}| < r_{\mathrm{fb}} \\ \frac{1}{2} k_{\mathrm{fb}} \left(|\mathbf{r}_i - \mathbf{R}_{\mathrm{c}}| - r_{\mathrm{fb}}\right)^2 & \text{if } |\mathbf{r}_i - \mathbf{R}_{\mathrm{c}}| > r_{\mathrm{fb}} \end{cases} \tag{4.1}$$

Here, $\mathbf{r}_i$ and $\mathbf{R}_{\mathrm{c}}$ are the instantaneous coordinates of the QM atoms and the coordinates of the box center, respectively. In order to not interfere with the reaction investigated, the sodium counterion was subjected to a similar restraint that pushed the ion away from the box center; in the implementation, this is designated with a negative value of r_{fb}.

Prior to the production simulations, an equilibration procedure was performed with the same set-up as above. The first phase consisted of heating to 300 K over the period of 10 ps, and was followed by further 100 ps of simulation at a constant temperature of 300 K. Finally, a simulation of 1 ns was performed with the Berendsen thermostat at 300 K and the Berendsen barostat at 1 bar. In the equilibration, flat-bottomed restraints were used with $r_{\mathrm{fb}} = 0.6$ nm for the protein system and –1.5 nm for the sodium ion.

The QM water system was generated from the fully equilibrated MM water system. From the water molecules within 0.4 nm of the solute, 30 molecules were selected and added to the QM region. To equilibrate the new configuration, a steepest-descents energy minimization was followed by 50 ps of MD simulation at a constant temperature of 300 K. In this simulation, flat-bottomed restraints were applied with $r_{\mathrm{fb}} = 0.4$ nm for the solute, 0.7 nm for the QM waters, –0.6 nm for the MM waters and –1.5 nm for the sodium ion.

In the QM/MM simulations of peptide systems, the QM region to be described with DFTB3/3OB included the three cysteine side chains. The QM/MM boundary passed through the C_α–C_β bonds, and the QM region was saturated by link hydrogen atoms. The remainder of the peptide including the alanine residues and the entire backbone was treated as part of the MM region and was described with the Amber 99SB-ILDN force field [132, 174].

Prior to production QM/MM simulations, an MM equilibration procedure was performed. The initial velocities of atoms were assigned from a Maxwell–Boltzmann distribution at 10 K, and the system was heated up to 300 K linearly over the interval of 10 ps. Then, a simulation at a constant temperature of 300 K for 100 ps was performed, followed by a simulation of 1 ns employing both Berendsen thermostat at 300 K and Berendsen barostat at 1 bar. During these steps, harmonic position restraints were applied to the heavy atoms of the peptide, with a force constant of 1000 $kJ \cdot mol^{-1} \cdot nm^{-2}$. No restraints were employed in production simulations.

4.2.3 Metadynamics simulations

Free energy calculations were performed with the metadynamics algorithm [92, 96]. A general challenge in metadynamics is selecting the right degrees of freedom, or collective variables (CV) to bias: only if a proper set of CV is introduced, then the movements along the slowest degrees of freedom are captured, the different stable states of the system are characterized and the transitions between them are observed. To meet this requirement, metadynamics simulations were performed with three different CV biased, and these were the distances between each pair of the sulfur atoms, see Fig. 4.1. Trivially, this set of CV describes the disulfide exchange reaction space completely.

The metadynamics simulations of the dimethyldisulfide–methylthiolate system were performed in two steps with the following settings. Gaussian biasing potentials of height of 0.05 $kJ \cdot mol^{-1}$ and width of 0.02 nm were deposited every 100 MD steps (0.05 ps). To prevent the molecular system from leaving the region of interest in the configurational space, a restraining potential was applied to all of the CV; this ‘upper wall’ was placed at 2.0 nm in the gas-phase simulation, and at 0.6 nm in the simulations with aqueous solvent. In the first step, simple metadynamics simulations were carried out in order to obtain a rough estimate of the free energy barriers. From this, a bias factor was estimated for more accurate well-tempered metadynamics simulations to be performed in the second step; a value of 7 was considered [96]. The multiple walker metadynamics approach was employed [175], which involves several different metadynamics simulations of the same system running in parallel and sharing the generated biasing potentials with each other. A total of 15 replicas of the same system were considered, utilizing most of the processor cores available on the computer

used in an efficient and convenient way. The initial structures of the 15 replicas were taken from the first metadynamics simulations randomly, and each of the three possible disulfide bonds was represented in five structures. The simulations with QM region containing water molecules were rather computationally demanding due to the large size of the QM region, and not very stable, therefore a single step of normal metadynamics was performed also in this case with the multiple walker approach.

To generate the free energy profiles of disulfide exchange in the peptides ACA_nCA_nCA, several different settings of metadynamics simulations were tested, and the convergence was generally more difficult than in the above simulations of the much simpler model system. A certain dependence on the starting structure was experienced in spite of the sampling being increased to 20 ns. The goal in metadynamics is to maximize the number of transitions between the minima of free energy, and best performing in this respect were the simulations in which Gaussian biasing potentials of height of 0.1 $kJ \cdot mol^{-1}$ and width of 0.02 nm were added every 100 MD steps. The results from these simulations will be presented and discussed below. Longer simulation times and/or a larger number of replicas in the multiple-walker metadynamics protocol would be needed to decrease the statistical uncertainty of the free energies obtained. To this end, subsequent multiple-walkers metadynamics simulations were carried out. Also in this case, the initial structures of the 15 replicas were randomly chosen from the previous metadynamics simulations.

4.3 Results

4.3.1 Benchmark of DFTB3 for Reactions Involving Sulfur

The general ability of DFTB3 to describe molecular systems containing sulfur has been reviewed [171]. Still, it is necessary assess the performance of DFTB3 with the 3OB parameter set for the specific reaction considered. To this end, we carried out gas-phase calculations of (i) the proton affinity of a thiol, (ii) the reaction energy of proton transfer between a thiol and a water molecule or a hydroxyl ion, and (iii) the energy of the transition state in a disulfide exchange reaction. DFTB3/3OB results were compared to calculations performed with the Gaussian G3B3 theory [176] and with the previous MP2/CBS and CCSD(T)/CBS data [70].

The model system for the disulfide exchange was taken from Ref. 70, and the structures of reactant, product and transition state were considered. The model system consisted of one methylthiolate and one dimethyldisulfide, see Fig. 4.1. The structures for the G3B3 calculations of proton affinity and proton transfer energy were generated with Molden [177].

The respective optimized structures were then used for the DFTB single-point calculations. Those were performed with DFTB+ [178], while G3B3 was calculated with Gaussian09 [140]. The proton affinities were computed as $E^{\mathrm{PA}} = E^{\mathrm{CH_3-S^-}} + E^{\mathrm{H+}} - E^{\mathrm{CH_3-SH}}$, and the reaction energies of proton transfer were obtained as $\Delta E = (E^{\mathrm{H_3O^+}} + E^{\mathrm{CH_3-S^-}}) - (E^{\mathrm{H_2O}} + E^{\mathrm{CH_3-SH}})$, and $\Delta E = (E^{\mathrm{H_2O}} + E^{\mathrm{CH_3-S^-}}) - (E^{\mathrm{HO^-}} + E^{\mathrm{CH_3-SH}})$. The results are shown in Tab. 4.1.

Reaction	DFTB3/3OB	G3B3	MP2/CBS	CCSD(T)/CBS
CH_3–S–S–CH_3 + CH_3–S^- → CH_3–S–S(CH_3)–S–$CH_3^{(-)}$	–0.7		–4.1	–1.8
CH_3–SH → CH_3–S^- + H^+	363.4	362.5		
H_2O + CH_3–SH → H_3O^+ + CH_3–S^-	191.1	191.2		
OH^- + CH_3–SH → H_2O + CH_3–S^-	–36.7	–36.0		

Table 4.1 Energetics of the relevant reactions that involve molecules containing sulfur. Values obtained with DFTB3/3OB are compared to higher-level calculations. All reaction energies in kcal·mol^{-1}.

The DFTB3 values of proton affinity are within 1 kcal·mol^{-1} of the G3B3 references, and also the DFTB3 reaction energy of the proton transfer is nearly identical with G3B3. Note that the energy barrier of disulfide exchange is inverted in the gas phase, and the trisulfide state is a global minimum of energy [68]. Excellent agreement is found for the energy of the trisulfide state: The DFTB3 value differs by 1.2 kcal·mol^{-1} from CCSD(T)/CBS, which is a much better agreement than the error of 2.2 kcal·mol^{-1} with MP2/CBS. Thus, it may be concluded that DFTB3/3OB describes disulfide exchange reactions adequately.

Also assessed was the potential energy surface in the vicinity of the trisulfide anion in vacuo; note that this species is the transition state in a disulfide exchange reaction taking place in an aqueous or protein environment. The DFTB3/3OB surface generated with DFTB+ is compared with a reference B3LYP/def2-TZVPP surface obtained with TurboMole 6.5 [179, 180] in Fig. 4.3.

The minimum-energy bond length of 2.73 Å with DFTB3/3OB is larger than reference (e.g., 2.49 Å with B3LYP/def2-TZVPP, and 2.41 Å with MP2/aug-cc-pVTZ reported by Neves et al.) [70]. Also, the flat region around the minimum – broad on the B3LYP surface – is rather narrow with DFTB3/3OB. Further, the overall topology of the surface is somewhat different with DFTB3/3OB: On the one hand, it is shallower along the shortening of one of the S–S bonds. This means that the structure with one shortened S–S is easily accessible, approaching the minimum-energy structure observed on the reference PES. On the other hand, it is steeper along the elongation of one of the S–S bonds, and that may affect the

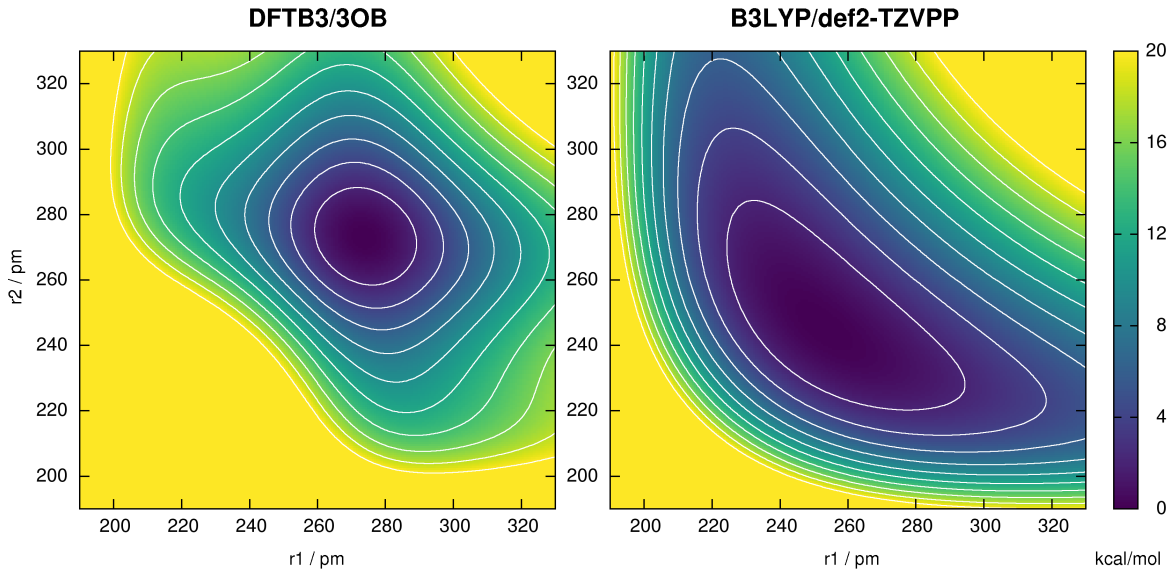


Fig. 4.3 The potential energy surfaces of CH_3–S–S(CH_3)–S–$CH_3^{(-)}$ on the levels DFTB3/3OB and B3LYP/def2-TZVPP. Linear configuration is considered, and the coordinates r_1 and r_2 represent the S–S distances; potential energy in kcal·mol^{-1} is coded by color, and contour lines are plotted every 2 kcal·mol^{-1}. (The structure was energy-minimized with each respective method first, and then r_1 and r_2 were varied with all of the other internal coordinates kept intact, a.k.a. unrelaxed scan.)

surroundings of the transition state of the system with environment: the slopes leading from the transition states downwards may be less steep.

Clearly, the observed deviations may lead to inaccurate details of the topology of the transition state in applications. However, since the energetics of the reaction is described by DFTB3/3OB in good accordance with reference, the overall appearance of free energy landscapes, with an exception of the narrow vicinity of transition states, may be expected to be reproduced well.

4.3.2 Dimethyldisulfide–Methylthiolate System

Metadynamics simulations were performed for the system composed of a dimethyldisulfide molecule and a methylthiolate anion. That system undergoes a series of disulfide-bond exchange reaction to rearrange in all three different ways. The three different sulfur–sulfur distances were employed as CV, allowing for a complete description of the relevant configurational space. Recall that in the gas phase, the free energy surface (FES) is expected to have a global minimum for the trisulfide anion structure, in which the negative charge is delocalized between all of the three sulfur atoms. Fig. 4.4 shows the three-dimensional

FES in a two-dimensional representation with an explicit dependence on r_1 and r_2, while the dependence on r_3 is integrated out. Also presented are the molecular structures found in the minima of free energy.

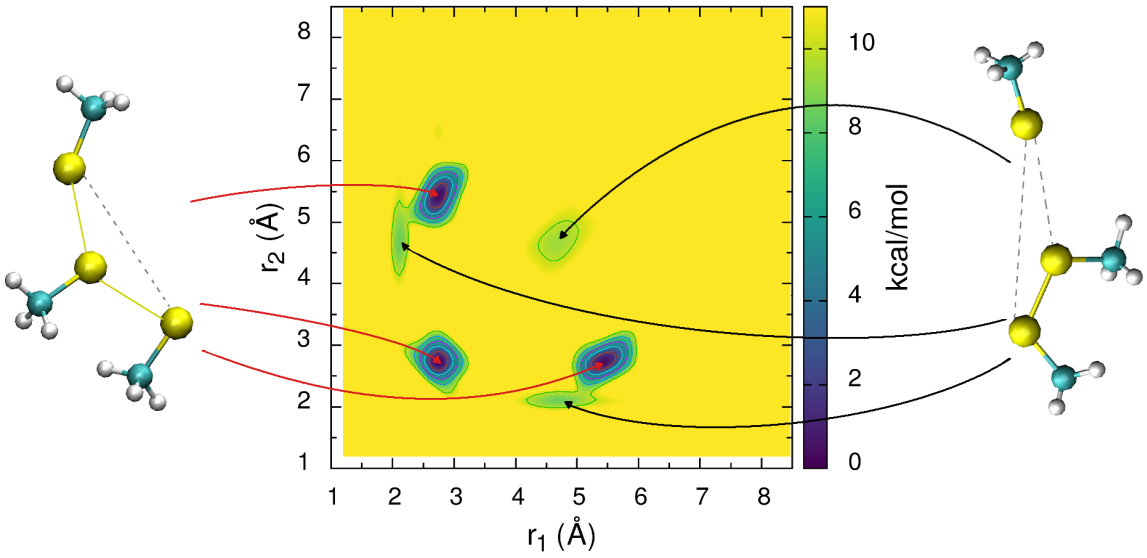


Fig. 4.4 Free energy surface of the disulfide-bond exchange reaction in the gas phase. Plotted is the projection of free energy as function of three variables (all of the S–S distances, r_1, r_2 and r_3) into two dimensions, r_1 and r_2. Left – the trisulfide anion is the global minimum; right – the system composed of disulfide and separate thiol lies higher in energy.

The global minimum of free energy is the trisulfide anion with S–S lengths of 2.7 Å and an S–S–S angle of ca. 165°. The slight deviation of geometry from a perfectly linear arrangement may be explained by a weak steric repulsion between the central methyl group and the two terminal sulfur atoms, and agrees with the previous results [71]. Note that the trisulfide anion species becomes a transition state for the S_N2 reaction as soon as the system is placed in a polar solvent like water, and a (nearly) linear geometry is expected for an S_N2 transition state. The structures consisting of a disulfide and a separate thiolate molecule are present as minima of free energy as well, however ca. 10 kcal·mol^{-1} higher in energy. Needless to note that each of the minima appears three times in the FES because of symmetry. The two other possible representations (r_1 or r_2 integrated out, shown in appendix B) were also analyzed and allowed for the same observations to be made.

While the trisulfide anion is the global minimum of energy in the gas phase, a polarizable environment stabilizes the disulfide–thiolate state of the molecule. Then, it becomes meaningful to consider a disulfide exchange reaction. The most practicable model of a polarizable

environment is liquid water, thus the reaction was investigated in an aqueous environment. The water was described with an explicit MM model TIP3P within a QM/MM framework, so that it was still possible to describe the molecules undergoing the reaction with a QM method, DFTB3. Again, the metadynamics protocol employing three CV was applied, and the resulting FES projected on r_1 and r_2 is shown in Fig. 4.5.

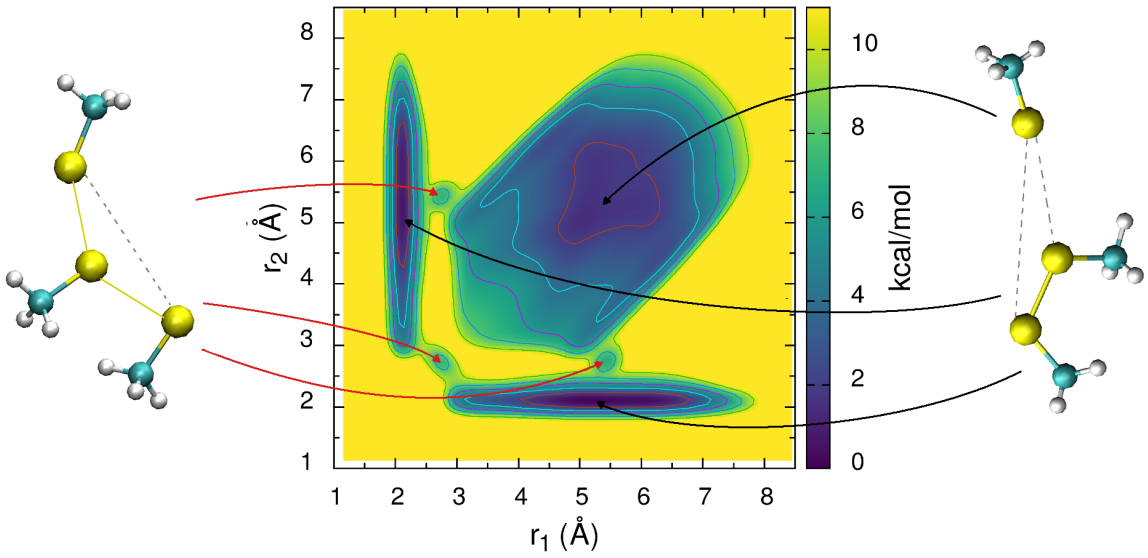


Fig. 4.5 Free energy surface of disulfide-bond exchange reaction in MM water projected on two S–S distances. Left – the trisulfide anion is a transition state; right – the disulfide–thiol system is a (global) minimum.

Each of the three minima features a dimethyl disulfide with a bond length of 2.1 Å, and a non-covalently interacting methylthiolate molecule is located at a distance of ca. 4.5 Å. The basin is rather broad and shallow along the two coordinates corresponding the distance of the methylthiolate, which illustrates a weak dependence of interaction energy on geometry of the non-covalent complex. The trisulfide anion structures with S–S bond lengths of 2.7 Å are saddle points with a very small curvature, and the height of the barriers is ca. 7 kcal·mol^{-1}. Note that the three minima correspond to the same state of the molecule by symmetry, as do the three transition states; therefore, their energies should be identical. In fact, it is apparent from the visual comparison with the other two possible projections, shown in appendix B, that these values are very similar actually. This fact points at a good convergence of the metadynamics simulation.

To eliminate any effects of the QM/MM boundary on the reaction free energies, the above simulations were repeated with an expanded QM region, which involved 30 water molecules

in addition to the disulfide and the thiolate. By virtue of the specifically designed set of restraints, the QM region remained compact throughout the simulation length, ensuring good convergence of DFTB3 calculation of forces. The resulting FES is shown in Fig. 4.6 in the same representation as in the previous simulations.

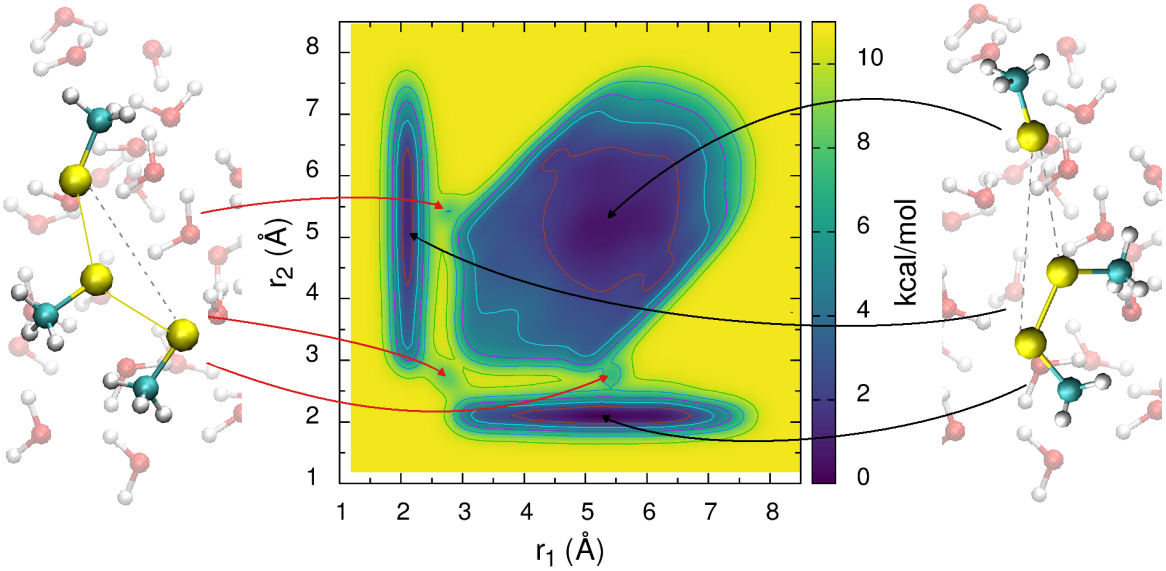


Fig. 4.6 Free energy surface of disulfide-bond exchange reaction in QM water projected on two S–S distances. The representative snapshots shown include the QM water molecules.

Clearly, the energies are the same as in the simulation with pure MM water. The basins of minimal free energies are located in the same regions, their spatial extension is the same, and the height of energy barriers of ca. 8 kcal·mol^{-1} agrees well also. The picture of the reaction energetics rendered by the simulations involving the extended QM region is qualitatively the same as that obtained with MM-only water. Identical structures correspond to global minima of energy, and the transition states are also placed identically. The relative free energies of the individual stable states of the molecule may be alternatively obtained by re-calculating the FES to a probability distribution, summing the probabilities for the stable states, and re-calculating these probabilities back to free energies. This was done here as well, with the stable states defined as $r_1 < 2.5$ Å, $r_2 < 2.5$ Å and $r_3 < 2.5$ Å; the overlap of these regions is small, hardly populated and was neglected. The free energies obtained in this way are shown schematically in Fig. 4.7 together with barrier heights read off from the above FES plots. Although the energy barriers obtained from simulations with QM-water molecules appear up to 1 kcal·mol^{-1} higher than in the MM-only-water case, such a difference is within the

margin of uncertainty of the DFTB method. Thus, it may be concluded that the disulfide exchange reaction is unaffected by any specific interactions with the aqueous environment.

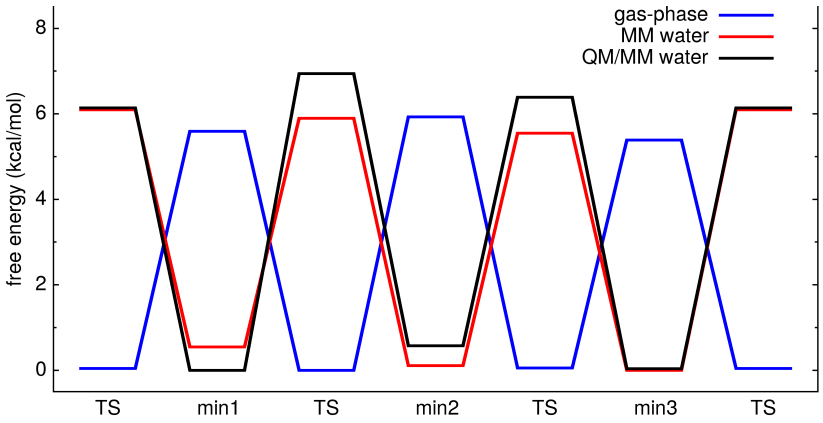


Fig. 4.7 Free energies of the disulfide shuffling reactions in all of the considered different kinds of environment. The energies of the minima were obtained by summation of the basins in the 3D histogram; the barrier heights were read off in the 2D projections of the free energy landscapes.

4.3.3 Disulfide-Bridge Exchange in Small Peptides

The disulfide-bond exchange reactions that occur between cysteine side chains embedded in peptide strands are chemically exactly the same as in the above studied model systems. However, the constraining effects of the peptide backbones reduce the configurational flexibility of the disulfide–thiol system. In other words, the entire process is affected by another degree of freedom which is the peptide conformation. Also, the three sulfur atoms are no longer indistinguishable, thus no symmetric FES is expected any longer. There will be still three minima of free energy, corresponding to the dimethyldisulfide–methylthiolate configuration, but their energies will be most likely different due to the different contributions of steric strain of the backbones. A series of short peptides is considered here, with sequences ACA_nCA_nCA (n=1,2,3), so that each peptide contains three cysteines separated by one, two and three alanines, respectively, recall Fig. 4.2.

In the first peptide, the cysteines are separated by single alanines, ACACACA. The cysteines are located rather close to each other, so that a strong constraining effect of the backbone may be expected for disulfide bridges Cys2–Cys4 and Cys4–Cys6. The FES obtained from the metadynamics simulation projected to two dimensions is shown in Fig. 4.8, as is its one-dimensional representation, which may be conveniently compared with that for the model system in Fig. 4.7.

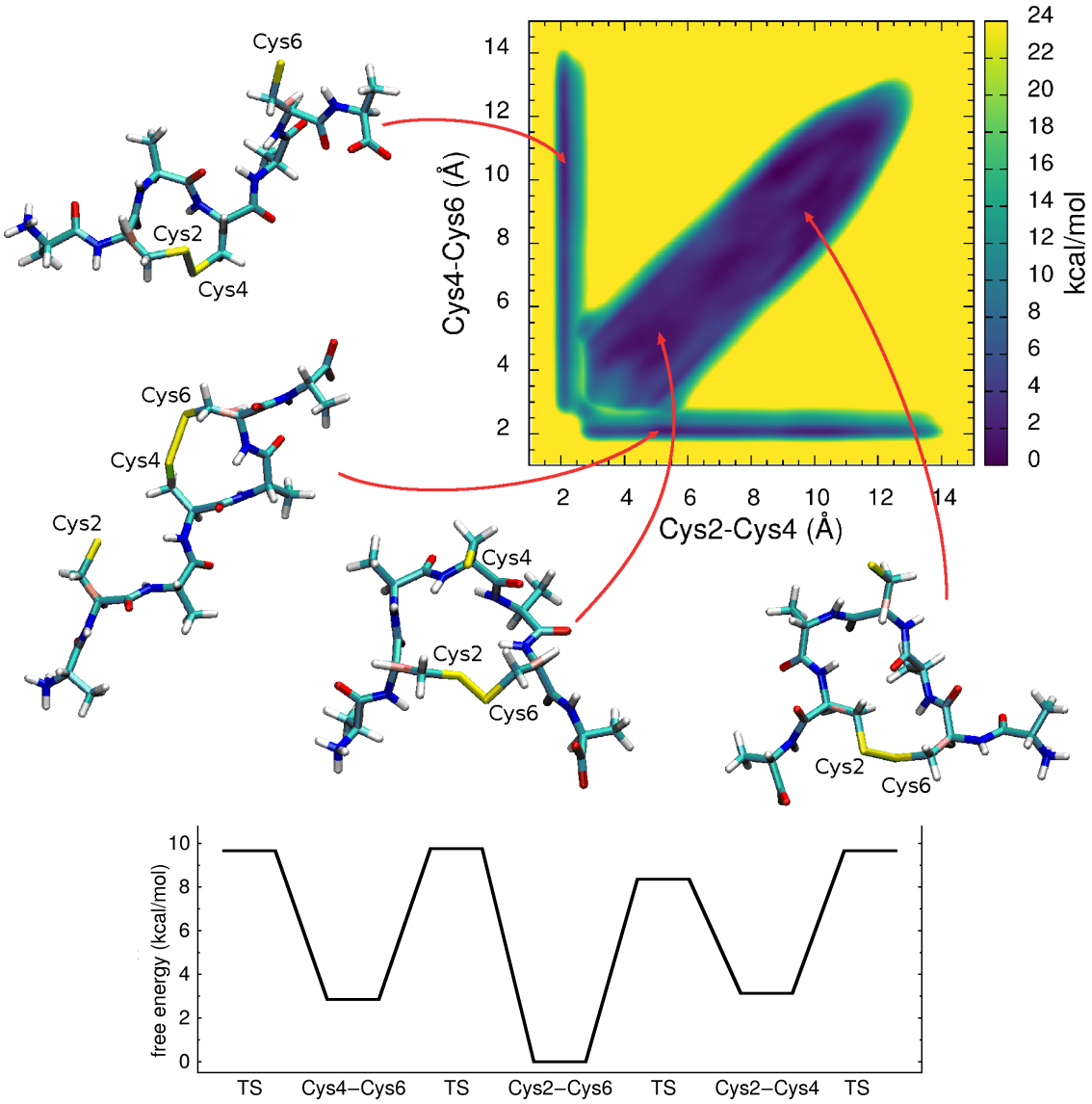


Fig. 4.8 Top: Free energy surface for the disulfide exchange reactions in the solvated ACACACA peptide. The 3D surface is projected to two dimensions as labeled in the plot. Representative snapshots are shown for all of the three disulfide bonds. Bottom: The energies of minima obtained by summation of basins in the 3D histogram, and the barrier heights read off from the 2D projections of free energy landscape.

There is a striking difference between this FES and that for the above model system: One of the three minima is clearly the deepest, and it corresponds to the disulfide bond between the two terminal cysteines, Cys2–Cys6. The other two minima are considerably less populated, and correspond to disulfide bridges between cystines that are next-neighbors on the peptide strand. This observation is apparent in the other representations of the FES

shown in appendix B, too. To estimate the relative free energies of the three disulfide-bonded situations, the FES was re-calculated to probabilities, and those were then summed over the three major basins; finally, the summed probabilities were converted back to free energies. It turns out that the disulfide bonds Cys2–Cys4 and Cys4–Cys6 lie ca. 2.5 kcal·mol^{-1} higher in the free energy than Cys2–Cys6. The height of the barriers between the minima amounts to 7–8 kcal·mol^{-1}, marginally higher than observed for the model system above.

The disulfide bonds Cys2–Cys4 and Cys4–Cys6 imply a ring of 11 atoms, which embraces two peptide bonds. Such a structure is strained largely, and that is reflected by a higher free energy. The global minimum corresponds to the state with the Cys2–Cys6 disulfide bond, in which the peptide forms a β-turn around Cys4, with the sulfur atom of Cys4 pointing outwards (right-bottom structure in Fig. 4.8). Any steric strain is clearly weaker than that in the structures with Cys2–Cys4 and Cys4–Cys6 disulfide bonds. Another point that becomes apparent already in these very simple peptides is how complex the energy landscape may become. For instance, the peptide backbone in the state with Cys2–Cys6 may conceivably assume many different conformations, and while these may differ little in energy, they may be separated by energy barriers. This makes the metadynamics simulation less efficient because (i) the additional barriers increase the time needed to sample any relevant transitions, and (ii) these degrees of freedom are not biased in the metadynamics simulations, therefore the corresponding transitions are not accelerated.

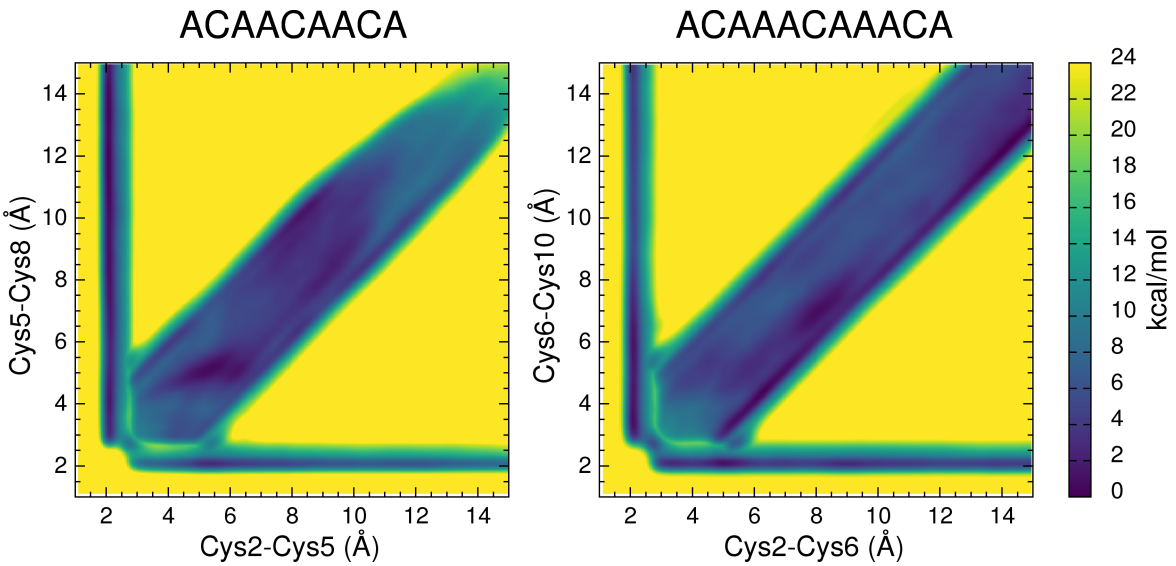


Fig. 4.9 Free energy surfaces for the disulfide exchange reaction in the peptides ACAACAACA and ACAAACAAACA, obtained from multiple-walker metadynamics simulations. Visible are the basins of minimal energy, corresponding to the three disulfide bonds, as well as the transition states. Still, no convergence is reached due to the conformational flexibility of the peptides and the sampling limited to tens of nanoseconds.

The metadynamics simulations of disulfide exchange in the longer peptides converged very slowly, making merely qualitative observations possible. There are two minima of similar depths in ACAACAACA (Fig. 4.9, left), corresponding to the "shorter" disulfide bonds Cys2–Cys5 and Cys5–Cys8. The disulfide bond spanning the longest segment of the peptide, Cys2–Cys8, lies several kcal·mol^{-1} higher in energy. Apparently, the steric strain affecting the relative orientation of two cystines with two amino acids in between is weak enough to allow for a relatively stable disulfide bond between those cystines. Further, the geometries of transition states are very similar to those in ACACACA, deviating from linearity to some extent. The heights of the corresponding energy barriers are in a similar range as in the case of ACACACA.

Basic qualitative features are also apparent in the FES obtained for the disulfide exchange in ACAAACAAACA (Fig. 4.9, right): There are the same kinds of minima as in the shorter peptides – those corresponding to the three disulfide bonds – and three transition states between them. It is not quite possible to formulate more detailed findings. Additional information on the longer peptides (n=2,3) is presented in appendix B.

Note again that the sampling of 75 ns each was insufficient to converge the FES for disulfide exchange in ACAACAACA, and even more so, ACAAACAAACA. Evidently, the volume of conformational space accessible to the peptide grows largely with its increasing length, as seen in Fig. 4.9 directly: The basins for the Cys5–Cys8 and Cys6–Cys10 bonds in the longer peptides reach much further in the north-east direction than that for Cys4–Cys6 in ACACACA in Fig. 4.8. The slow convergence of the metadynamics simulations is a direct consequence of that.

In summary, the underlying chemistry in the small peptides that were investigated here is the same as in the model methylthiolate system, and the only difference is caused by the steric effects induced by the peptide backbones. Note that these steric effects are described by the MM force field in the QM/MM simulations. Consequently, if a parameterized model for the disulfide exchange reaction (e.g., a reactive force field) should be constructed, it is sufficient to consider a very small molecular system like in dimethyldisulfide–methylthiolate for the construction of the model. Any large-scale effects like, prominently, the steric strain due to the peptide backbone would be accounted for by the standard MM force field.

4.4 Conclusions

We find the combined DFTB/metadynamics scheme to reproduce the major features and expected free energy landscapes of the S_N2 mechanism that has been established for thiol/disulfide exchange. A QM/MM scheme with classical water molecules is an effi-

cient and accurate mimic of a solvated and purely quantum mechanically described system. Interestingly, disulfide bonded CXC motifs very readily undergo isomerization reactions with nearby cysteines, with a barrier of 7–8 kcal·mol^{-1}, which is quite similar to that observed in a hydrated model system in this work, however lower than 14 kcal·mol^{-1} measured for another reaction in water experimentally. We attribute the lability of a CXC disulfide bond to the inherent strain of the short loop formed by the three involved residues. Our simulations explain the comparably low abundance of CXC compared to other (e.g. the mostly reducing CXXC) motifs in known protein structures [181, 182], and their ability to mimic the high isomerization propensity in protein disulfide isomerases.

This work showed several aspects of the energetics of the disulfide shuffling reaction: In the first part, the inherent free energy landscape of the system was characterized, both in vacuum and in water representing a highly polar environment. The energetics in these systems is driven by the properties of the 'tri-sulfur system' itself, with little external restraining force. The second part illustrated how that inherent energetics of the disulfide exchange reaction changes in response to the structural constraints that appear in a real molecule, represented here by a short peptide. The thermodynamics of the disulfide exchange is affected, with steric effects making some of the disulfide bridges unfavorable. On the other hand, there were little changes to the kinetics in terms of energy barriers opposing the exchange reaction.

The free energy maps generated may be used to build a reactive force field for the disulfide bridge exchange reaction. Such a force field may be deployed for large-scale simulations of various phenomena in which the dynamics of the exchange reaction in proteins plays a role. Such a development may well be both superior in accuracy and more practicable than our previous approach [166]. Work along these lines is underway in our laboratories.

Chapter 5

Force-clamp QM/MM simulations of intramolecular thiol-disulfide isomerization

5.1 Introduction

Pulling experiments, e.g. force-clamp atomic force microscopy (AFM), are widely used to study the mechanical unfolding of proteins [183–186]. In such experiments, a single chain of protein domains is stretched with a constant force able to induce a rupture of structures maintained by non–covalent bonds. The resulting stepwise increase in the length of the protein chain is registered. Thus, the unfolding events are determined by the measurements of the chain elongation. The same protocol is applied in simulations. The technique is called steered molecular dynamics (SMD)[187, 188] and involves a constant external force along the vector between the termini of the protein.

These techniques are also used to investigate the effect of mechanical force on chemical reactions [189, 71, 190]. Many studies have been performed to investigate the mechanism and kinetics of thiol–disulfide exchange reaction under mechanical force. Force clamp AFM experiments and molecular simulations showed that a pulling force triggered the thiol-disulfide exchange reaction by activating the disulfide bond. It was demonstrated that the rate of disulfide bond reduction by various chemical reducing agents, such as dithiothreitol (DTT) and thioredoxin (Trx), is dependent on the applied force [189, 71, 190].

Mechanical unfolding experiments proved to be also very useful to quantify the spontaneous disulfide isomerization in a single protein, which is a difficult task with the current condensed phase techniques. As a matter of fact, these techniques have serious limitations

in discriminating parallel intramolecular pathways, because they are not able to distinguish isomers with the same oxidation state. Furthermore, they have to deal with the interference from the reverse reaction.

In a recent study, Fernandez and co-workers [165] reported a direct isomerization of disulfide bond in a single protein using mechanical unfolding of a mutated I27 immunoglobulin domain (I27*). The I27* domain was engineered to have two oxidized cysteines at positions 24 and 55, which form a disulfide, and a free reactive cysteine at position 32. With a pulling constant force of 250 pN, far below the force necessary to break covalent bonds (above 1 nN [189, 191]), the unfolding extends up to the disulfide bond , which can only be cleaved by a reducing agent. The reducing agent is Cys32, which remains in the vicinity of the disulfide bond after the first unfolding step. Thus, disulfide Cys24-Cys55 can be cleaved by the reaction of any of its sulfur atoms with Cys32, figures 5.1 and 5.2. The two possible reaction pathways are detected and distinguished as the elongation of the protein chain. With this setup, the regioselectivity of the reaction was determined, with the result that the attack of Cys32 on Cys55 is 3.8 times more frequent than that on Cys24.

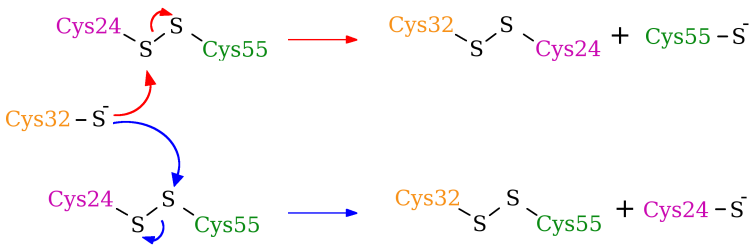


Fig. 5.1 The two possible reaction pathways for the reduction of disulfide Cys24-Cys55 by the nucleophilic attack of Cys32

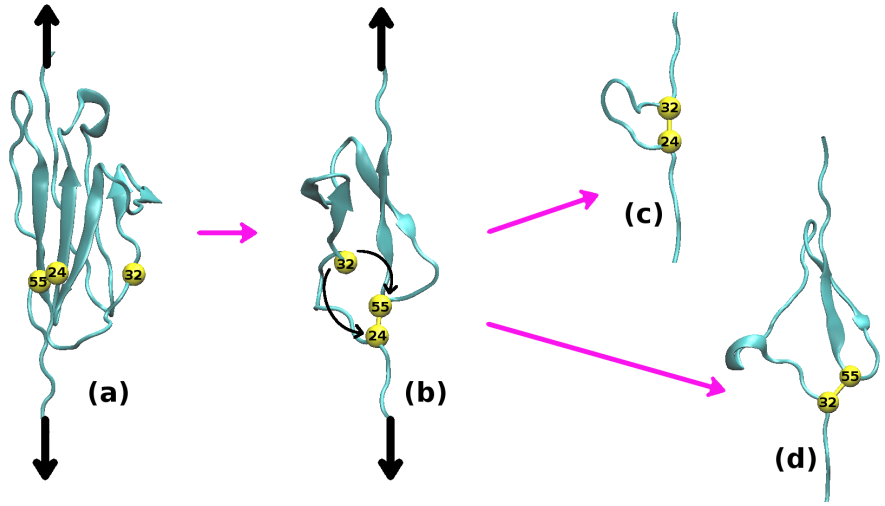


Fig. 5.2 **a** Unfolding of the I27* domain under constant force. The sulfur atoms are depicted with yellow spheres. The first unfolding step uncaged Cys32 **(b)**, which remains in the vicinity of the disulfide 24-55. The latter can be cleaved by the nucleophilic attack of Cys32 at any of its sulfur atoms, following one of the two possible reaction pathway **(c)(d)**.

Later on, Gräter et al. [166] performed force-clamp MD simulations on the same mutated protein (I27*) , using a purely classical framework that allows the disulfide rearrangements through a Monte Carlo-based topology swap (hereafter called ”reactive MD“ approach). The simulation proceeds with a normal force field description until the distance of the free cysteine (Cys32) from the disulfide bond becomes lower than a threshold value, then the simulation stops and an energy-based Metropolis criterion is applied. If the shuffling is accepted the conformation with the new disulfide bond will be constructed and relaxed by performing local minimization. The simulation will then restart with the new conformation and topology. This approach was able to reproduce the regioselectivity observed by the experimental AFM setup with the advantage of an atomistic description of the process. However, it has at least two drawbacks: (i) it does not provide information about the kinetics of the process. Every time a swap is accepted, the simulation is stopped and a new one is started (ii) It does not incorporate information on the free energy barrier or transition state structure of the chemical reaction.

In order to improve upon that previous work, a QM/MM description may be employed. The application of DFTB3 for the quantum-chemical calculations, successfully used for the free energy description of thiol-disulfide exchange (chapter 4) , seems to be a good compromise between the fast but not very accurate reactive MD approach and QM/MM with more sophisticated but computationally expensive quantum chemistry methods. In the

following, the preliminary results of the study of thiol-disulfide isomerization in a single molecule by means of force clamp QM/MM simulations are presented.

5.2 Materials and Methods

5.2.1 Protein Models

The protein structures were obtained from Gräter laboratory from the reactive MD simulations [166]. A set of 100 conformations were randomly selected from the snapshots in which a shuffling was attempted (50 for each pathway (Fig. 5.1)). The protein is already elongated up to the disulfide in these structures. To reduce the system size, the unfolded termini were removed and only residues 20-65 were kept.

A further mutation was made in the polypeptide chain in order to make the system neutral and avoid the use of counterions, which can interfere with the investigated reaction. To this end, residues 62, 64 and 65 were converted into lysines. The xLeap tool from the AmberTools package [147] was used.

The protein was then centered in a rectangular box sized $15.0 \times 4.8 \times 4.8$ nm^3 and solvated by 11125 water molecules.

5.2.2 Simulation Details

The simulations were performed with Gromacs 5.0.1[144, 145] patched with our previous DFTB3 implementation[168] and Plumed 2.1.1 [169, 148]. The leap-frog integrator was used with a time step of 0.5 fs, and periodic boundary conditions were employed. Regarding the MM and QM/MM interactions, the electrostatics were treated with the particle–mesh Ewald method, while the Lennard-Jones were cut-off at 1 nm; the neighbour list was updated every 10 MD steps (5 fs). A constant temperature of 300 K and a constant pressure of 1 bar were maintained, respectively, by means of the Nosé–Hoover thermostat and the Parrinello–Rahman barostat. Trajectory was written with an output frequency of 0.5 ps.

The semi-empirical density functional theory method DFTB3[88] with the 3OB parametrization [170, 171] was employed for the QM interactions.

The QM region to be described with DFTB3/3OB included the three cysteine side chains (24,32 and 55). The QM/MM boundary passed through the C_α–C_β bonds, and the QM region was saturated by link hydrogen atoms. The remainder of the protein as well as the water solvent were treated as part of the MM region and were described with the Amber 99SB-ILDN force field [132, 174] and the TIP3P model[138] respectively.

The electrostatic interaction between the negatively charged QM region and the MM system was scaled down by the factor of 0.75 corresponding to the inverse square root of the optical dielectric constant. This is an effective approach to compensate for the missing electronic polarization of the MM environment as recommended by Stuchebrukhov [173].

Prior to production QM/MM simulations, an MM equilibration procedure was performed. The initial velocities of atoms were assigned from a Maxwell–Boltzmann distribution at 10 K, and the system was heated up to 300 K linearly over the interval of 10 ps. Then, a simulation at a constant temperature of 300 K for 100 ps was performed, followed by a simulation of 1 ns employing both v-rescale thermostat at 300 K and Parrinello-Rahman barostat at 1 bar. During these steps, harmonic position restraints were applied to the heavy atoms of the peptide, with a force constant of 1000 $kJ \cdot mol^{-1} \cdot nm^{-2}$. The QM/MM system was generated from the fully equilibrated MM system, and a further step of 100 ps of simulation at constant temperature and pressure was finally performed. During the pulling simulations, the two centers of mass of the terminal amino acids were pulled away from each other with a constant force of 500 $kJ \cdot mol^{-1} \cdot nm^{-1}$ (830 pN).

Altogether, 100 force-clamp simulations of 5 ns each were performed, yielding 500 ns of accumulated simulation time.

The simulations were performed without any additional bias but the mechanical constant force applied at the termini of the chain.

5.3 Results

Force-clamp QM/MM simulations were performed for a mutated I27 immunoglobulin domain in aqueous solution. The protein contains a disulfide bond between Cys24 and Cys55 and a free deprotonated cysteine at position 32.

A disulfide isomerization reaction was observed in 9 out of 100 pulling simulations (9%), in which Cys32 reacted 3 times with Cys24 (Fig. 5.2-c) and 6 times with Cys55 (Fig. 5.2-d), resulting in a 2-fold regioselectivity of 55 versus 24. Surprisingly, even with such a small sample size the preference of Cys32 attacking Cys55 reproduces the AFM experimental results. [165]

The primary factors of thiol–disulfide exchange in protein are proximity and accessibility of the nucleophile sulfur to the disulfide, as was pointed out in the previous studies. Upon unfolding, the tertiary structure of the $I27^*$ domain disintegrated almost completely, and only the loop within the disulfide bond, which contains the Cys32, remained protected from a complete stretching. The intrinsic dynamics of the loop defines the accessibility of Cys32 to react with the disulfide. In order to check the ability of Cys32 to come into contact

with the disulfide, the distributions of distances of the attacking sulfur from the disulfide were obtained. The broad range from 4 Å to 20 Å, reported in figure 5.3, denotes a highly mobility/flexibility of the loop. Although few, the distributions include samples located around the transition state distance, which our previous calculation with DFTB3 (chapter 4) revealed to be 2.7 Å.

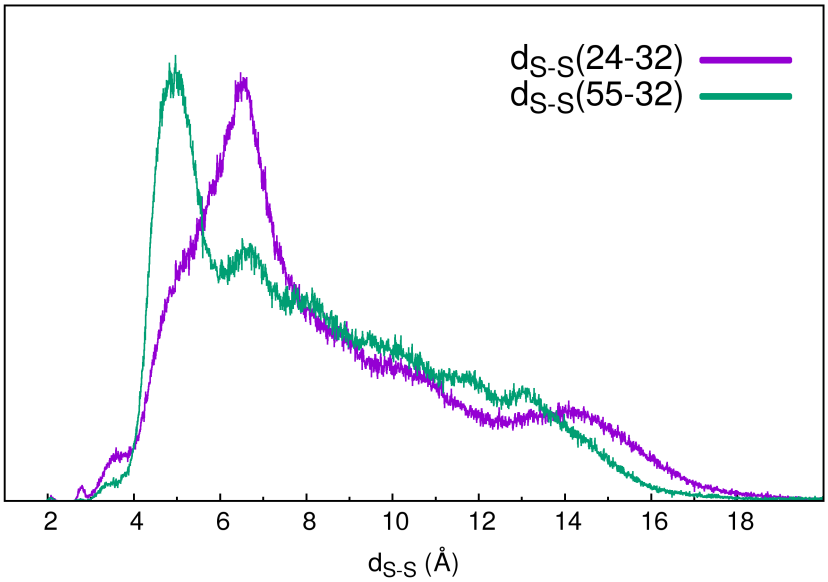


Fig. 5.3 Distribution of distances between the sulfur of the Cys32 and the sulfurs of Cys24 (purple) and Cys55 (green)

The average distances between the sulfur of Cys32 and the sulfurs of Cys24 and Cys55 are 8.7 Å and 8.3 Å, respectively. This very small difference does not justify the preference of Cys55 over Cys24 as attacked sulfur.

The orientation of the attacking sulfur is another important factor in the reaction. For an efficient thiol-disulfide reaction, the attacking thiolate should approach along the disulfide axis, with an S–S–S angle that ranges from 180° to 160°. This leads to significant steric requirements. We analyzed the angle distributions (see Fig. 5.4) for the two reaction pathways. Only the occurrences where the distance between the Cys32 and the considered attacked sulfur was smaller than 5 Å were considered in the calculation.

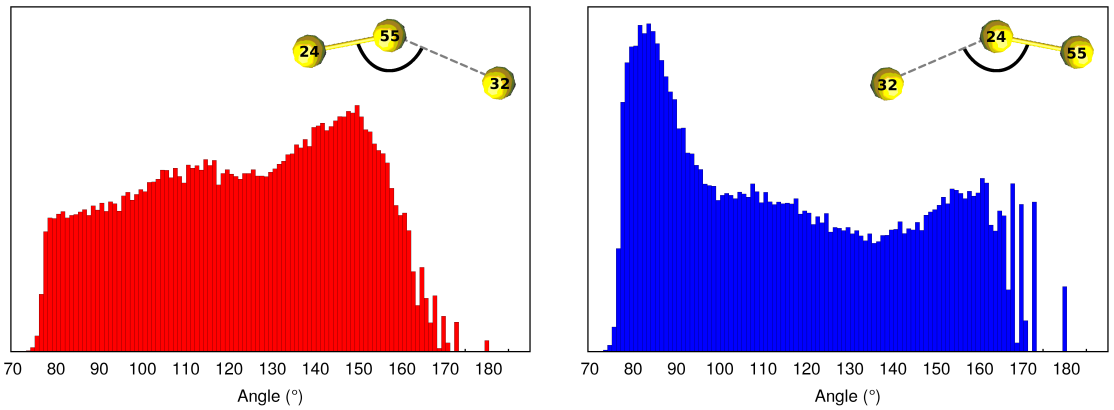


Fig. 5.4 Distributions of angles of the three sulfur atoms involved in the reaction for the two isomerization reaction pathways.

Both distributions cover a range from 75° to 170°. However, While for the reaction between Cys32 and Cys24 (figure 5.4 right) small angles occur more often, there are more frequent occurrences of large angles involving Cys55. In other words, Cys32 approaches Cys55 more often at the correct angle than it does with Cys24. That could be an explanation for the regioselectivity in favor of the reaction involving Cys55.

Finally the transition state structure of the 9 observed disulfide shuffling events was characterized. All the reaction showed a symmetric geometry with the bond lengths of approximately 2.75 Å and an angle that ranges from 160° to 170°. These values are in accordance with those found in the free energy calculations in our previous work on model molecules as well as small peptides(chapter 4). Representatives conformations of transition structures for each of the two reactions are depicted in figure 5.5.

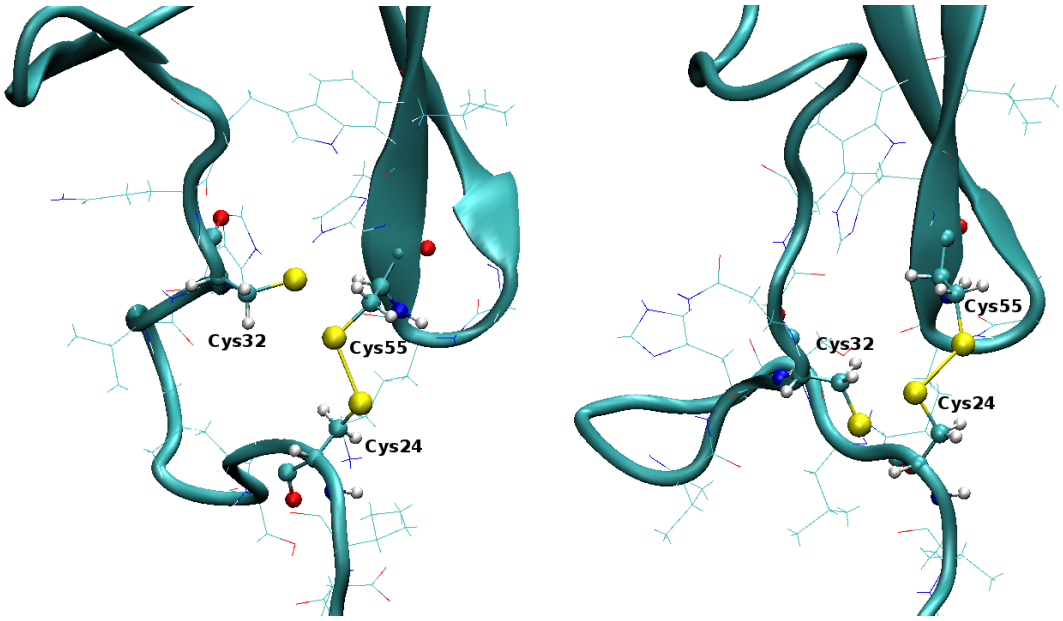


Fig. 5.5 Representative conformations of the transition state for the two thiol-disulfide exchange reactions, in which Cys32 attacks the disulfide at Cys55 (*left*) and at Cys24 (*right*)

5.4 Conclusion

The force-clamp QM/MM framework with DFTB3 as quantum chemical method was able to simulate the disulfide bond isomerization in a single protein. Note that, no bias potentials were applied to enhance the reaction. The isomerizations occurred spontaneously and the regioselectivity observed in the AFM experiment was reproduced.

However, of the 100 simulations performed, only 9 reactions were observed, too few for extrapolate good statistics. Therefore, a cumulative simulation time of 500 ns was not sufficient, and more sampling would be required.

In the recent work of Gräter and co-workers, a new purely classical approach was implemented, which allows the disulfide switching in a normal MD simulation through a proximity criterion. This approach has the great advantage of being able to sample the large-scale protein motions at the atomistic level, however, is limited in the sense that it does not provide information on the kinetics of the process and is not able to monitor the interactions and the conformational changes of the transition state.

On the other hand, the QM/MM approach used in this work was shown to characterize the spontaneous reaction well, but it can hardly reach the time-scale necessary to sample conformational dynamics of the protein.

A more complete picture of non-enzymatic thiol-disulfide isomerization in proteins could only be obtained by coupling the large-scale conformational changes, addressed by MM, to the QM/MM simulations. A possible implementation could be through a switching simulation tool between an MM and QM/MM description based on a distance criterion. The results from this work could be used for this purpose, and we are currently working on that direction.

Chapter 6

Summary

Two different aspect of the dynamic nature of proteins are investigated in this work.

First, the structure and orientation of an antimicrobial peptide in lipid bilayer were studied. We performed simulations of Harzianin HK VI (HZ), a short-sequence peptaibol, interacting with a DMPC bilayer. On the basis of free MD simulations, two possible orientations of HZ in the bilayer were identified: the I-state, with HZ perpendicular to the bilayer surface, and the S-state, where HZ lies in the hydrophobic core of the bilayer, close and parallel to its surface. However, the sampling achieved in such free MD simulations was by far too insufficient to decide which of the orientations is preferred.

The HREX method was then employed to determine the relative free energies of the two different orientations found. Both the S-state and the I-state were confirmed as the relevant structures, with the S-state being the more likely one. However, it was still impossible to obtain a number of interconversions between the S- and I-states that would be sufficient to obtain conclusive results in terms of converged populations (or, equivalently, free energies). Calculation of free energies with extended metadynamics and umbrella sampling simulations turned out to be even less practicable. The issue of unclear dominant orientation was resolved by means of ssNMR and SRCD experiments, performed by Ulrich's group. Their analysis revealed that the HZ molecule is embedded in the surface of the bilayer preferably, i.e., it assumes the S-state.

The structure of HZ was characterized on the basis of free MD simulations. The conformation of the HZ molecule in both the S- and I-state orientations falls into the 3_{10}–helical family. The HZ molecule is somewhat longer along the helical axis in the I-state than in the S-state.

In the more relevant S-state, the HZ molecule is located completely within the hydrophobic region of the bilayer, in contrast to the behavior of cationic amphiphilic peptides, which usually dock into the hydrophobic/hydrophilic interface.

The analysis of local thickness profiles revealed that the HZ molecule in the I-state induces a local thinning of the DMPC bilayer. In that way, the length of the HZ molecule, otherwise too short, closely matches the (decreased) thickness of the hydrophobic region of the bilayer in the immediate vicinity of HZ. Therefore, although the S-state is the dominant orientation, it is conceivable that the I-state may act as an important, low-energy intermediate in the insertion into bilayer, transfer between the bilayer leaflets, or in a process of aggregation of HZ into oligomers at conditions of higher P/L.

This study illustrated also the grand challenge of simulation peptides in lipid bilayers. The slow rearrangement of the hydrocarbon lipid tails can lead to undersampling with all-atom MD simulation. Even, the extended sampling techniques proved to be of limited help. A partial solution of this problem would be to pass to a coarse-grained representation, which would require the secondary structure of HZ as an input for the parametrization. The results from this work could be used for this purpose, with a direct application for the oligomerization of HZ at elevated P/L.

In the second part, a study of thiol-disulfide exchange reaction was performed by means of extended sampling QM/MM molecular dynamics simulations. The application of approximative density-functional method DFTB3 for the quantum chemical calculations made it possible to reach multi-nanosecond time scale, which turned out to be necessary to achieve semi-quantitative results.

This work showed several aspects of the energetics of the disulfide shuffling reaction: In the first part, the inherent free energy landscape of the system was characterized, both in vacuum and in water representing a highly polar environment. The energetics in these systems is driven by the properties of the 'tri-sulfur system' itself, with no external restraining force. The second part illustrated how that inherent energetics of the disulfide exchange reaction changes in response to the structural constraints that appear in a real molecule, represented here by a short peptide. The thermodynamics of the disulfide exchange is affected, with steric effects making some of the disulfide bridges unfavorable. On the other hand, there were little changes to the kinetics in terms of energy barriers opposing the exchange reaction.

The method was able to reproduce the major features and expected free energy landscapes of the S_N2 mechanism that had been established for thiol/disulfide exchange. A QM/MM scheme with classical water molecules is an efficient and accurate mimic of a solvated and purely quantum mechanically described system.

A direct application in a real protein was also performed through force-clamp simulations. The protocol allowed for simulation of the disulfide bond isomerization in a single protein, and reproduced the regioselectivity observed in the AFM experiment. However, the simulation time of half a microsecond turned out to be insufficient statistically reliable results.

The time scale necessary to sample the conformational dynamics of a protein can hardly be reached with a QM/MM approach. Two possible solutions are highlighted: (i) the development of a reactive force field for the disulfide bridge exchange reaction, using the free energy maps generated from model systems, and (ii) coupling of the large-scale conformational changes, addressed by MM, to QM/MM simulations. Work along these lines is underway.

References

[1] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences*, 47(9):1309–1314, 1961.

[2] Ali Adem Bahar and Dacheng Ren. Antimicrobial peptides. *Pharmaceuticals*, 6(12):1543–1575, 2013.

[3] K. Lohner. *Development of novel antimicrobial agents: emerging strategies.* Horizon Scientific Press, Wymondham, UK, 2001.

[4] John Conlon and Ágnes Sonnevend. Antimicrobial peptides in frog skin secretions. *Methods in molecular biology*, 618:3–14, 2010.

[5] Erik Strandberg and Anne S. Ulrich. Nmr methods for studying membrane-active antimicrobial peptides. *Concepts in Magnetic Resonance Part A*, 23A(2):89–120, 2004.

[6] W.William Wilson, Mary Margaret Wade, Steven C. Holman, and Franklin R. Champlin. Status of methods for assessing bacterial cell surface charge properties based on zeta potential measurements. *Journal of Microbiological Methods*, 43(3):153 – 164, 2001.

[7] B. Bechinger. Structure and functions of channel-forming peptides: Magainins, cecropins, melittin and alamethicin. *J. Membrane Biol.*, 156:197–211, 1997.

[8] Kim A. Brogden. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology*, 3:238–250, 2005.

[9] Steve J. Ludtke, Ke He, William T. Heller, Thad A. Harroun, Lin Yang, and Huey W. Huang. Membrane pores induced by magainin. *Biochemistry*, 35(43):13723–13728, 1996.

[10] Yechiel Shai. Mode of action of membrane active antimicrobial peptides. *Peptide Science*, 66(4):236–248, 2002.

[11] Lin Yang, Thad A. Harroun, Thomas M. Weiss, Lai Ding, and Huey W. Huang. Barrel-stave model or toroidal model? a case study on melittin pores. *Biophysical Journal*, 81(3):1475–1485, 2001.

[12] B. Leitgeb, A. Szekeres, L. Manczinger, C. Vágvölgyi, and L. Kredics. The history of alamethicin: A review of the most extensively studied peptaibol. *Chem. Biodivers.*, 4(6):1027–1051, 2007.

[13] K. Lohner and S. E. Blondelle. Molecular mechanisms of membrane perturbation by antimicrobial peptides and the use of biophysical studies in the design of novel peptide antibiotics. *Combinatorial Chemistry & High Throughput Screening*, 8(3):241–256, 2005.

[14] Ziv Oren and Yechiel Shai. Mode of action of linear amphipathic β-helical antimicrobial peptides. *Peptide Science*, 47(6):451–463, 1998.

[15] Huey W. Huang. Action of antimicrobial peptides: Two-state model. *Biochemistry*, 39(29):8347–8352, 2000.

[16] Juliana F. de S. Daniel and Edson Rodrigues Filho. Peptaibols of trichoderma. *Nat. Prod. Rep.*, 24:1128–1141, 2007.

[17] Xianying Wei, Fuquan Yang, and David C Straney. Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in trichoderma virens. *Canadian Journal of Microbiology*, 51(5):423–429, 2005.

[18] Nora K. N. Neumann, Norbert Stoppacher, Susanne Zeilinger, Thomas Degenkolb, Hans Brückner, and Rainer Schuhmacher. The peptaibiotics database – a comprehensive online resource. *Chemistry & Biodiversity*, 12(5):743–751, 2015.

[19] Nicolas Ruiz, Gaeetane Wielgosz-Collin, Laurence Poirier, Olivier Grovel, Karina Ethel Petit, Mustapha Mohamed-Benkada, Thibaut Robiou du Pont, John Bissett, Philippe Verite, Gilles Barnathan, and Yves Francois Pouchus. New trichobrachins, 11-residue peptaibols from a marine strain of *trichoderma longibrachiatum*. *Peptides*, 28(7):1351–1358, 2007.

[20] A. Szekeres, B. Leitgeb, L. Kredics, Z. Antal, L. Hatvani, L. Manczinger, and C. Vágvölgyi. Peptaibols and related peptaibiotics of trichoderma. *Acta Microbiol. Imm. H.*, 52(2):137–168, 2005.

[21] C. E. Meyer and F. Reusser. A polypeptide antibacterial agent isolated from *Trichoderma viride*. *Experientia*, 23(2):85–86, 1967.

[22] R. O. Fox Jr. and F. M. Richards. A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-å resolution. *Nature*, 300(5890):325–330, 1982.

[23] M. K. Mathew and P. Balaram. A helix dipole model for alamethicin and related transmembrane channels. *FEBS Lett.*, 157(1):1–5, 1983.

[24] Shun-ichi Wada, Akira Iida, Koji Asami, and Tetsuro Fujita. Ion channel-forming property of trichorovin-XII, an 11-residue peptaibol from the fungus *Trichoderma viride*, in planar lipid bilayer membranes. *Bioorg. Med. Chem. Lett.*, 6(19):2275–2278, 1996.

[25] I. Ségalas, Y. Prigent, D. Davoust, B. Bodo, and S. Rebuffat. Characterization of a type of β-bend ribbon spiral generated by the repeating (xaa-yaa-aib-pro) motif: The solution structure of harzianin hc ix, a 14-residue peptaibol forming voltage-dependent ion channels. *Biopolymers*, 50(1):71–85, 1999.

[26] Mustapha Mohamed-Benkada, Monique Montagu, Jean-François Biard, Florence Mondeguer, Philippe Verite, Michèle Dalgalarrondo, John Bissett, and Yves François Pouchus. New short peptaibols from a marine trichoderma strain. *Rapid Communications in Mass Spectrometry*, 20(8):1176–1180, 2006.

[27] Shun ichi Wada, Akira Iida, Naoshige Akimoto, Michiko Kanai, Nobuo Toyama, and Tetsuro Fujita. Fungal metabolites. xix structural elucidation of channel-forming peptides, trichorovins-i-xiv, from the fungus trichoderma viride. *Chemical & Pharmaceutical Bulletin*, 43(6):910–915, 1995.

[28] Akira Iida, Masaki Sanekata, Shun ichi Wada, Tetsuro Fujita, Hiromi Tanaka, Akio Enoki, Goro Fuse, Michiko Kanai, and Koji Asami. Fungal metabolities. xviii. new membrane-modifying peptides, trichorozins i-iv, from the fungus trichoderma harzianum. *Chemical & Pharmaceutical Bulletin*, 43(3):392–397, 1995.

[29] Laurence Poirier, Françoise Quiniou, Nicolas Ruiz, Monique Montagu, Jean-Claude Amiard, and Yves François Pouchus. Toxicity assessment of peptaibols and contaminated sediments on *crassostrea gigas* embryos. *Aquat. Toxicol.*, 83(4):254–262, 2007.

[30] Zsuzsanna Antal, L. Kredics, J. Pakarinen, Ilona Dóczi, Maria Andersson, Mirja S. Salkinoja-Salonen, L. Manczinger, A. Szekeres, L. Hatvani, C. Vágvölgyi, and Elisabeth Nagy. Comparative study of potential virulence factors in human pathogenic and saprophytic *trichoderma longibrachiatum* strains. *Acta Microbiol. Imm. H.*, 52(3-4):341–350, 2005.

[31] Raimo Mikkola, Maria A. Andersson, László Kredics, Pavel A. Grigoriev, Nina Sundell, and Mirja S. Salkinoja-Salonen. 20-residue and 11-residue peptaibols from the fungus *trichoderma longibrachiatum* are synergistic in forming na^+/k^+-permeable channels and adverse action towards mammalian cells. *FEBS J.*, 279(22):4172–4190, 2012.

[32] Shun-ichi Wada and Reiko Tanaka. A novel 11-residual peptaibol-derived carrier peptide for in vitro oligodeoxynucleotide delivery into cell. *Bioorg. Med. Chem. Lett.*, 14(10):2563–2566, 2004.

[33] Shun-ichi Wada, Yasunari Hitora, Saori Yokoe, Osamu Nakagawa, and Hidehito Urata. Cellular uptake of covalent conjugates of oligonucleotide with membrane-modifying peptide, peptaibol. *Bioorg. Med. Chem.*, 20(10):3219–3222, 2012.

[34] M Schirmböck, M Lorito, Y L Wang, C K Hayes, I Arisan-Atac, F Scala, G E Harman, and C P Kubicek. Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *trichoderma harzianum* against phytopathogenic fungi. *Appl. Environ. Microb.*, 60(12):4364–4370, 1994.

[35] B Di Blasio, V Pavone, M Saviano, A Lombardi, F Nastri, C Pedone, E Benedetti, M Crisma, M Anzolin, and C Toniolo. Structural characterization of the β-bend ribbon spiral – crystallographic analysis of 2 long (L-Pro-Aib)n sequential peptides. *J. Am. Chem. Soc.*, 114(16):6273–6278, 1992.

[36] J.M. Thornton. Disulphide bridges in globular proteins. *Journal of Molecular Biology*, 151(2):261 – 287, 1981.

[37] Ronald Wetzel. Harnessing disulfide bonds using protein engineering. *Trends in Biochemical Sciences*, 12:478 – 482, 1987.

[38] H. F. Gilbert. Thiol/disulfide exchange equilibria and disulfide bond stability. *Methods Enzymol*, 251:8–28, 1995.

[39] Sarah Lewney and Lorna J. Smith. Characterization of an alternative low energy fold for bovine α-lactalbumin formed by disulfide bond shuffling. *Proteins: Structure, Function, and Bioinformatics*, 80(3):913–919, 2012.

[40] Kenta Arai, Fumio Kumakura, and Michio Iwaoka. Characterization of kinetic and thermodynamic phases in the prefolding process of bovine pancreatic ribonuclease a coupled with fast ss formation and ss reshuffling. *Biochemistry*, 49(49):10535–10542, 2010.

[41] Meng Qin, Wei Wang, and D. Thirumalai. Protein folding guides disulfide bond formation. *Proceedings of the National Academy of Sciences*, 112(36):11241–11246, 2015.

[42] Feras Hatahet and Lloyd W. Ruddock. Protein disulfide isomerase: A critical evaluation of its function in disulfide bond formation. *Antioxidants & Redox Signaling*, 11(11):2807–2850, 2009.

[43] Péter Nagy. Kinetics and mechanisms of thiol–disulfide exchange covering direct substitution and thiol oxidation-mediated pathways. *Antioxidants & Redox Signaling*, 18(13), 2013.

[44] Salima Patel, Alain F. Chaffotte, Batt Amana, Fabrice Goubard, and Emmanuel Pauthe. In vitro denaturation–renaturation of fibronectin. formation of multimers disulfide-linked and shuffling of intramolecular disulfide bonds. *The International Journal of Biochemistry & Cell Biology*, 38(9):1547 – 1560, 2006.

[45] Jasimuddin Ahamed, Nathalie Burg, Keiji Yoshinaga, Christin A. Janczak, Daniel B. Rifkin, and Barry S. Coller. In vitro and in vivo evidence for shear-induced activation of latent transforming growth factor-β1. *Blood*, 112(9):3650–3660, 2008.

[46] T. Garderton, M. C. Berndt, C. N. Chesterman, and P. J. Hogg. Hypothesis for control of von willebrand factor multimer size by intra-molecular thiol-disulphide exchange. *Journal of Thrombosis and Haemostasis*, 5(1):204–206, 2007.

[47] Hans-Dietrich Weigmann, Ludwig Rebenfield, and Cornelia Dansizer. Kinetics and temperature dependence of the chemical stress relaxation of wool fibers. *Textile Research Journal*, 36(6):535–542, 1966.

[48] Max Feughelman. *Mechanical properties and structure of alpha-keratin fibres : wool, human hair and related fibres*. Sydney : UNSW Press, 1997.

[49] S. E. Shapiro, A. A. Nowak, C. Wooding, G. Birdsey, M. A. Laffan, and T. A. J. McKinnon. The von willebrand factor predicted unpaired cysteines are essential for secretion. *Journal of Thrombosis and Haemostasis*, 12(2):246–254, 2014.

[50] Barbara A. Solecka, Christoph Weise, Birte Fuchs, and Christoph Kannicht. Free thiol groups in von willebrand factor (vwf) are required for its full function under physiological flow conditions. *Thrombosis Research*, 137:202 – 210, 2016.

[51] H. F. Gilbert. Molecular and cellular aspects of thiol-disulfide exchange. *Adv. Enzymol. Relat. Areas Mol. Biol.*, 63:69–172, 1990.

[52] R. D. Bach, O. Dmitrenko, and C. Thorpe. Mechanism of thiolate-disulfide interchange reactions in biochemistry. *J. Org. Chem*, 73:12–21, 2008.

[53] Antonino Fava, Antonio Iliceto, and Ettore Camera. Kinetics of the thiol-disulfide exchange. *Journal of the American Chemical Society*, 79(4):833–838, 1957.

[54] David A. Keire, Erin Strauss, Wei Guo, Bela Noszal, and Dallas L. Rabenstein. Kinetics and equilibria of thiol/disulfide interchange reactions of selected biological thiols and related molecules with oxidized glutathione. *The Journal of Organic Chemistry*, 57(1):123–127, 1992.

[55] Ze'ev Shaked, Richard P. Szajewski, and George M. Whitesides. Rates of thiol-disulfide interchange reactions involving proteins and kinetic measurements of thiol pka values. *Biochemistry*, 19(18):4156–4166, 1980.

[56] Richard P. Szajewski and George M. Whitesides. Rate constants and equilibrium constants for thiol-disulfide interchange reactions involving oxidized glutathione. *Journal of the American Chemical Society*, 102(6):2011–2026, 1980.

[57] Shu Quan, Irmhild Schneider, Jonathan Pan, Annekathrin Von Hacht, and James C. A. Bardwell. The cxxc motif is more than a redox rheostat. *Journal of Biological Chemistry*, 282(39):28823–28833, 2007.

[58] Rodney A. Bednar. Reactivity and ph dependence of thiol conjugation to n-ethylmaleimide: detection of a conformational change in chalcone isomerase. *Biochemistry*, 29(15):3684–3690, 1990.

[59] Jeffrey W. Nelson and Thomas E. Creighton. Reactivity and ionization of the active site cysteine residues of dsba, a protein required for disulfide bond formation in vivo. *Biochemistry*, 33(19):5974–5983, 1994.

[60] Surapong Pinitglang, Aaron B. Watts, Manij Patel, James D. Reid, Michael A. Noble, Sheraz Gul, Anjumon Bokth, Akavish Naeem, Hasu Patel, Emrys W. Thomas, Suneal K. Sreedharan, Chandra Verma, and Keith Brocklehurst. A classical enzyme active center motif lacks catalytic competence until modulated electrostatically. *Biochemistry*, 36(33):9968–9982, 1997.

[61] Rosa E. Hansen, Henrik Østergaard, and Jakob R. Winther. Increasing the reactivity of an artificial dithiol–disulfide pair through modification of the electrostatic milieu. *Biochemistry*, 44(15):5899–5906, 2005.

[62] Goedele Roos, Nicolas Foloppe, and Joris Messens. Understanding the pk_a of redox cysteines: The key role of hydrogen bonding. *Antioxidants & Redox Signaling*, 18(1):94–127, 2013.

[63] George M. Whitesides, Jennifer E. Lilburn, and Richard P. Szajewski. Rates of thiol-disulfide interchange reactions between mono- and dithiols and ellman's reagent. *The Journal of Organic Chemistry*, 42(2):332–338, 1977.

[64] Todd V. DeCollo and Watson J. Lees. Effects of aromatic thiols on thiol–disulfide interchange reactions that occur during protein folding. *The Journal of Organic Chemistry*, 66(12):4244–4249, 2001.

[65] Rasmus Iversen, Peter Anders Andersen, Kristine Steen Jensen, Jakob R. Winther, and Bent W. Sigurskjold. Thiol–disulfide exchange between glutaredoxin and glutathione. *Biochemistry*, 49(4):810–820, 2010.

[66] Jakob R. Winther and Colin Thorpe. Quantification of thiols and disulfides. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1840(2):838 – 846, 2014. Current methods to study reactive oxygen species - pros and cons.

[67] Jennifer A. Codding, Benjamin A. Israel, and Colin Thorpe. Protein substrate discrimination in the quiescin sulfhydryl oxidase (qsox) family. *Biochemistry*, 51(20):4226–4235, 2012.

[68] P. A. Fernandes and M. J. Ramos. Theoretical insights into the mechanism for thiol/disulfide exchange. *Chem. Eur. J.*, 10:257–266, 2004.

[69] Maria Francesca Iozzi, Trygve Helgaker, and Einar Uggerud. Influence of external force on properties and reactivity of disulfide bonds. *The Journal of Physical Chemistry A*, 115(11):2308–2315, 2011.

[70] R. P. P. Neves, P. A. Fernandes, A. J. C. Varandas, and M. J. Ramos. Banchmarking of density functionals for the accurate description of thiol-disulfide exchange. *Journal of Chemical Theory and Computation.*, 10:4842–4856, 2014.

[71] Wenjin Li and Frauke Gräter. Atomistic evidence of how force dynamically regulates thiol/disulfide exchange. *Journal of the American Chemical Society*, 132(47):16790–16795, 2010.

[72] D. M. Rothwarf and H. A. Scheraga. Equilibrium and kinetic constants for the thiol-disulfide interchange reaction between glutathione and dithiothreitol. *Proceedings of the National Academy of Sciences*, 89(17):7944–7948, 1992.

[73] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[74] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[75] W. F. Van Gunsteren and H. J. C. Berendsen. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3):173–185, 1988.

[76] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[77] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52(2):255–268, 1984.

[78] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[79] S. Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, 50(5):1055–1076, 1983.

[80] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[81] Patra M, Karttunen M, Hyvönen MT, Falck E, Lindqvist P, and Vattulainen I. Molecular dynamics simulations of lipid bilayers: Major artifacts due to truncating electrostatic interactions. *Biophysical Journal*, 84(6):3636–3645, 2003.

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

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

[84] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[85] Axel D. Becke. A new mixing of hartree-fock and local density-functional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, 1993.

[86] D. Porezag, Th. Frauenheim, Th. Köhler, G. Seifert, and R. Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B*, 51:12947–12957, 1995.

[87] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998.

[88] Michael Gaus, Qiang Cui, and Marcus Elstner. Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb). *Journal of Chemical Theory and Computation*, 7(4):931–948, 2011.

[89] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187 – 199, 1977.

[90] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[91] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[92] A. Laio and M. Parrinello. Escaping free energy minima. *Proc. Natl. Acad. Sci. USA*, 99(20):12562–12566, 2002.

[93] Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the accuracy of metadynamics. *The Journal of Physical Chemistry B*, 109(14):6714–6721, 2005.

[94] Francesco Luigi Gervasio, Alessandro Laio, and Michele Parrinello. Flexible docking in solution using metadynamics. *Journal of the American Chemical Society*, 127(8):2600–2607, 2005.

[95] Giovanni Bussi, Alessandro Laio, and Michele Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.*, 96:090601, 2006.

[96] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100(2):020603, 2008.

[97] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, 116(20):9058–9067, 2002.

[98] Pu Liu, Byungchan Kim, Richard A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA*, 102(39):13749–13754, 2005.

[99] Lingle Wang, Richard A. Friesner, and B. J. Berne. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B*, 115(30):9431–9438, 2011.

[100] Christian Kandt, Walter L. Ash, and D. Peter Tieleman. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, 41(4):475–488, 2007.

[101] Erik Lindahl and Mark S. P. Sansom. Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 18(4):425–431, 2008.

[102] Scott Feller. *Molecular Dynamics Simulation of Phospholipid Bilayers*, pages 89–107. Springer, Berlin, Heidelberg, 2001.

[103] Herman J. C. Berendsen and D. Peter Tieleman. *Molecular Dynamics: Studies of Lipid Bilayers*. John Wiley & Sons, Ltd, 2002.

[104] Alexander P. Lyubartsev and Alexander L. Rabinovich. Force field development for lipid membrane simulations. *Biochim. Biophys. Acta Biomembranes*, 1858(10):2483–2497, 2016.

[105] C. Chipot and A. Pohorille. Structure and dynamics of small peptides at aqueous interfaces: A multi-nanosecond molecular dynamics study. *J. Mol. Struct. Theochem*, 398-399:529–535, 1997.

[106] C. Chipot and A. Pohorille. Folding and translocation of the undecamer of poly-L-leucine across the water-hexane interface. a molecular dynamics study. *J. Am. Chem. Soc.*, 120(46):11912–11924, 1998.

[107] D. P. Tieleman, H. J. C. Berendsen, and M. S. P. Sansom. Surface binding of alamethicin stabilizes its helical structure: Molecular dynamics simulations. *Biophys. J.*, 76(6):3186–3191, 1999.

[108] M R Alecio, D E Golan, W R Veatch, and R R Rando. Use of a fluorescent cholesterol derivative to measure lateral mobility of cholesterol in membranes. *Proc. Natl. Acad. Sci. USA*, 79(17):5171–5174, 1982.

[109] L.K. Tamm and H.M. McConnell. Supported phospholipid bilayers. *Biophys. J.*, 47(1):105–113, 1985.

[110] R. Mills. Self-diffusion in normal and heavy water in the range 1–45 deg. *J. Phys. Chem.*, 77(5):685–688, 1973.

[111] Peter J. Bond, John Holyoake, Anthony Ivetac, Syma Khalid, and Mark S. P. Sansom. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J. Struct. Biol.*, 157(3):593–605, 2007.

[112] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.*, 4(5):819–834, 2008.

[113] A. J. Rzepiela, D. Sengupta, N. Goga, and S.-J. Marrink. Membrane poration by antimicrobial peptides combining atomistic and coarse-grained descriptions. *Faraday Discuss.*, 144:431–443, 2010.

[114] Andrey A. Gurtovenko, Jamshed Anwar, and Ilpo Vattulainen. Defect-mediated trafficking across cell membranes: Insights from in silico modeling. *Acc. Chem. Res.*, 110(10):6077–6103, 2010.

[115] Jochen S. Hub, Bert L. de Groot, Helmut Grubmüller, and Gerrit Groenhof. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.*, 10(1):381–390, 2014.

[116] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, 1999.

[117] W. Im and C. L. Brooks III. Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*, 102(19):6771–6776, 2005.

[118] Hugh Nymeyer, Thomas B. Woolf, and Angel E. Garcia. Folding is not required for bilayer insertion: Replica exchange simulations of an α-helical peptide with an explicit lipid bilayer. *Proteins*, 59(4):783–790, 2005.

[119] Alexander Vogel, Matthew Roark, and Scott E. Feller. A reinterpretation of neutron scattering experiments on a lipidated ras peptide using replica exchange molecular dynamics. *Biochim. Biophys. Acta Biomem.*, 1818(2):219–224, 2012.

[120] Jianguo Li, Rajamani Lakshminarayanan, Yang Bai, Shouping Liu, Lei Zhou, Konstantin Pervushin, Chandra Verma, and Roger W. Beuerman. Molecular dynamics simulations of a new branched antimicrobial peptide: A comparison of force fields. *J. Chem. Phys.*, 137(21):215101, 2012.

[121] R. D. Groot and K. L. Rabone. Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic surfactants. *Biophys. J.*, 81(2):725–736, 2001.

[122] Licui Chen, Nana Jia, Lianghui Gao, Weihai Fang, and Leonardo Golubovic. Effects of antimicrobial peptide revealed by simulations: Translocation, pore formation, membrane corrugation and euler buckling. *Int. J. Mol. Sci.*, 14(4):7932–7958, 2013.

[123] Dejun Lin and Alan Grossfield. Thermodynamics of micelle formation and membrane fusion modulate antimicrobial lipopeptide activity. *Biophys. J.*, 109(4):750–759, 2015.

[124] Tristan Bereau, W. F. Drew Bennett, Jim Pfaendtner, Markus Deserno, and Mikko Karttunen. Folding and insertion thermodynamics of the transmembrane WALP peptide. *J. Chem. Phys.*, 143(24):243127, 2015.

[125] D. P. Tieleman, M. S. P. Sansom, and H. J. C. Berendsen. Alamethicin helices in a bilayer and in solution: Molecular dynamics simulations. *Biophys. J.*, 76(1 I):40–49, 1999.

[126] D. P. Tieleman, H. J. C. Berendsen, and M. S. P. Sansom. Voltage-dependent insertion of alamethicin at phospholipid/water and octane/water interfaces. *Biophys. J.*, 80(1):331–346, 2001.

[127] D. P. Tieleman, B. Hess, and M. S. P. Sansom. Analysis and evaluation of channel models: Simulations of alamethicin. *Biophys. J.*, 83(5):2393–2407, 2002.

[128] Lea Thøgersen, Birgit Schiøtt, Thomas Vosegaard, Niels C. Nielsen, and Emad Tajkhorshid. Peptide aggregation and pore formation in a lipid bilayer: A combined coarse-grained and all atom molecular dynamics study. *Biophys. J.*, 95(9):4337–4347, 2008.

[129] Zoltán Násztor, János Horváth, and Balázs Leitgeb. Structural characterization of the short peptaibols trichobrachins by molecular-dynamics methods. *Chem. Biodivers.*, 10(5):876–886, 2013.

[130] Zoltán Násztor, János Horváth, and Balázs Leitgeb. In silico conformational analysis of the short-sequence hypomurocin A peptides. *Int. J. Pept.*, 2015:281065, 2015.

[131] S. Bobone, Y. Gerelli, M. De Zotti, G. Bocchinfuso, A. Farrotti, B. Orioni, F. Sebastiani, E. Latter, J. Penfold, R. Senesi, F. Formaggio, A. Palleschi, C. Toniolo, G. Fragneto, and L. Stella. Membrane thickness and the mechanism of action of the short peptaibol trichogin GA IV. *Biochim. Biophys. Acta Biomem.*, 1828(3):1013–1024, 2013.

[132] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, 2006.

[133] Lauren Wickstrom, Asim Okur, and Carlos Simmerling. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys. J.*, 97(3):853–856, 2014.

[134] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, 11(8):3696–3713, 2015.

[135] Joakim P. M. Jämbeck and Alexander P. Lyubartsev. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B*, 116(10):3164–3179, 2012.

[136] Joakim P. M. Jämbeck and Alexander P. Lyubartsev. An extension and further validation of an all-atomistic force field for biological membranes. *J. Chem. Theory Comput.*, 8(8):2938–2948, 2012.

[137] Joakim P. M. Jämbeck and Alexander P. Lyubartsev. Another piece of the membrane puzzle: Extending slipids further. *J. Chem. Theory Comput.*, 9(1):774–784, 2013.

[138] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Phys. Chem.*, 79(2), 1983.

[139] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, 97(40):10269–10280, 1993.

[140] M. J. Frisch and colleagues. Gaussian 09 Revision C.01. Gaussian Inc. Wallingford CT 2009.

[141] Gromacs reference manual, version 2016, section 4.9.1. http:/manual.gromacs.org/documentation/2016/manual-2016.pdf, last accessed 7 Sep 2016.

[142] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[143] J. P. Ulmschneider, J. C. Smith, M. B. Ulmschneider, A. S. Ulrich, and E. Strandberg. Reorientation and dimerization of the membrane-bound antimicrobial peptide PGLa from microsecond all-atom MD simulations. *Biophys. J.*, 103(3):472–482, 2012.

[144] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.*, 91(1-3):43 – 56, 1995.

[145] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.

[146] Yukun Wang, Charles H. Chen, Dan Hu, Martin B. Ulmschneider, and Jakob P. Ulmschneider. Spontaneous formation of structurally diverse membrane channel architectures from a single antimicrobial peptide. *Nat. Commun.*, 7:13535, 2016.

[147] AmberTools 1.4 (2010). http://ambermd.org.

[148] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Comp. Phys. Commun.*, 185(2):604–613, 2014.

[149] Barbara A. Lewis and Donald M. Engelman. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *J. Mol. Biol.*, 166(2):211–217, 1983.

[150] Pavol Balgavý, Martina Dubničková, Norbert Kučerka, Mikael A. Kiselev, Sergey P. Yaradaikin, and Daniela Uhríková. Bilayer thickness and lipid interface area in unilamellar extruded 1,2-diacylphosphatidylcholine liposomes: a small-angle neutron scattering study. *Biochim. Biophys. Acta, Biomembr.*, 1512(1):40 – 52, 2001.

[151] Norbert Kučerka, Yufeng Liu, Nanjun Chu, Horia I. Petrache, Stephanie Tristram-Nagle, and John F. Nagle. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.*, 88(4):2626 – 2637, 2005.

[152] Jianjun Pan, D. Peter Tieleman, John F. Nagle, Norbert Kučerka, and Stephanie Tristram-Nagle. Alamethicin in lipid bilayers: Combined use of x-ray scattering and MD simulations. *Biochim. Biophys. Acta Biomem.*, 1788(6):1387–1397, 2009.

[153] Mei Hong and Yongchao Su. Structure and dynamics of cationic membrane peptides and proteins: Insights from solid-state NMR. *Protein Sci.*, 20(4):641–655, 2011.

[154] Marina Putzu, Sezgin Kara, Sergii Afonin, Stephan L. Grage, Andrea Bordessa, Grégory Chaume, Thierry Brigaud, Anne S. Ulrich, and Tomáš Kubař. Structural behavior of the peptaibol harzianin hk vi in a dmpc bilayer: Insights from md simulations. *Biophysical Journal*, 112(12):2602 – 2614, 2017.

[155] Stephan L. Grage, Sezgin Kara, Andrea Bordessa, Véronique Doan, Fabio Rizzolo, Marina Putzu, Tomáš Kubař, Anna Maria Papini, Grégory Chaume, Thierry Brigaud, Sergii Afonin, and Anne S. Ulrich. Orthogonal 19f-labeling for solid-state nmr spectroscopy reveals the conformation and orientation of short peptaibols in membranes. *Chemistry - A European Journal*, 2018.

[156] Joakim P. M. Jämbeck and Alexander P. Lyubartsev. Exploring the free energy landscape of solutes embedded in lipid bilayers. *J. Phys. Chem. Lett.*, 4(11):1781–1787, 2013.

[157] Davide Bochicchio, Emanuele Panizon, Riccardo Ferrando, Luca Monticelli, and Giulia Rossi. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *J. Chem. Phys.*, 143(14):144108, 2015.

[158] Hugo A. L. Filipe, Maria João Moreno, Tomasz Róg, Ilpo Vattulainen, and Luís M. S. Loura. How to tackle the issues in free energy simulations of long amphiphiles interacting with lipid membranes: Convergence and local membrane deformations. *J. Phys. Chem. B*, 118(13):3572–3581, 2014.

[159] Chris Neale, Jenny C. Y. Hsu, Christopher M. Yip, and Régis Pomès. Indolicidin binding induces thinning of a lipid bilayer. *Biophys. J.*, 106(8):L29–L31, 2014.

[160] D. Butera, K. M. Cook, J. Chiu, J. W. H. Wong, and P. J. Hogg. Control of blood proteins by functional disulfide bonds. *Blood*, 123(13):2000–2007, 2014.

[161] Clive Metcalfe, Peter Cresswell, Laura Ciaccia, Benjamin Thomas, and A. Neil Barclay. Labile disulfide bonds are common at the leucocyte cell surface. *Open Biol.*, 1(3):110010, 2011.

[162] D. Narzi, S. W. Siu, C. U. Stirnimann, J. P. Grimshaw, R. Glockshuber, G Capitani, and R. A. Böckmann. Evidence for proton shuffling in a thioredoxin-like protein during catalysis. *J. Mol. Biol.*, 382(4):978–986, 2008.

[163] L. E. Netto, M. A. de Oliveira, C. A. Tairum, and J. F. da Silva Neto. Conferring specificity in redox pathways by enzymatic thiol/disulfide exchange reactions. *Free Radic. Res.*, 50(2):206–245, 2016.

[164] R. P. P. Neves, P. A. Fernandes, and M. J. Ramos. Mechanistic insights on the reduction of glutathione disulfide by protein disulfide isomerase. *Proc. Natl. Acad. Sci. USA*, 114(24):E4724–E4733, 2017.

[165] Jorge Alegre-Cebollada, Pallav Kosuri, Jaime Andrés Rivas-Pardo, and Julio M. Fernández. Direct observation of disulfide isomerization in a single protein. *Nature Chemistry*, 3:882–887, 2011.

[166] Katra Kolšek, Camilo Aponte-Santamaría, and Frauke Gräter. Accessibility explains preferred thiol-disulfide isomerization in a protein domain. *Scientific Reports*, 7:9858, 2017.

[167] Kenneth J. Woycechowsky and Ronald T. Raines. The CXC motif: a functional mimic of protein disulfide isomerase†. *Biochemistry*, 42(18):5387–5394, 2003.

[168] T. Kubař, K. Welke, and G. Groenhof. New qm/mm implementation of the dftb3 method in the gromacs package. *J Comput Chem*, 36(26):1978–89, 2015.

[169] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A. Broglia, and Michele Parrinello. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Commun.*, 180(10):1961–1972, 2009.

[170] Michael Gaus, Albrecht Goez, and Marcus Elstner. Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.*, 9(1):338–354, 2013.

[171] M. Gaus, X. Lu, M. Elstner, and Q. Cui. Parameterization of dftb3/3ob for sulfur and phosphorus for chemical and biological applications. *Journal of Chemical Theory and Computation*, 10:1518–1537, 2014.

[172] Johan Åqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Chem. Phys.*, 94(21):8021–8024, 1990.

[173] Igor Leontyev and Alexei Stuchebrukhov. Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys.*, 13:2613–2626, 2011.

[174] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, 2010.

[175] Paolo Raiteri, Alessandro Laio, Francesco Luigi Gervasio, Cristian Micheletti, and Michele Parrinello. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B*, 110(8):3533–3539, 2006.

[176] Anwar G. Baboul, Larry A. Curtiss, Paul C. Redfern, and Krishnan Raghavachari. Gaussian-3 theory using density functional geometries and zero-point energies. *J. Chem. Phys.*, 110(16):7650–7657, 1999.

[177] G. Schaftenaar and J.H. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput.-Aided Mol. Design*, 14:123–134, 2000.

[178] B. Aradi, B. Hourahine, and Th. Frauenheim. Dftb+, a sparse matrix-based implementation of the dftb method. *The Journal of Physical Chemistry A*, 111(26):5678–5684, 2007.

[179] Reinhart Ahlrichs, Michael Bär, Marco Häser, Hans Horn, and Christoph Kölmel. Electronic structure calculations on workstation computers: The program system TURBOMOLE. *Chem. Phys. Lett.*, 162(3):165–169, 1989.

[180] Filipp Furche, Reinhart Ahlrichs, Christof Hättig, Wim Klopper, Marek Sierka, and Florian Weigend. Turbomole. *WIREs Comput. Mol. Sci.*, 4(2):91–100, 2014.

[181] Dmitri E. Fomenko and Vadim N. Gladyshev. Genomics perspective on disulfide bond formation. *Antioxid. Redox Sign.*, 5(4):397–402, 2003.

[182] Jun Lu and Arne Holmgren. The thioredoxin superfamily in oxidative protein folding. *Antioxid. Redox Sign.*, 21(3):457–470, 2014.

[183] Julio M. Fernandez and Hongbin Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 303(5664):1674–1678, 2004.

[184] The study of protein mechanics with the atomic force microscope. *Trends in Biochemical Sciences*, 24(10):379 – 384, 1999.

[185] Julia R Forman and Jane Clarke. Mechanical unfolding of proteins: insights into biology, structure and folding. *Current Opinion in Structural Biology*, 17(1):58 – 66, 2007.

[186] David J. Brockwell, Godfrey S. Beddard, Emanuele Paci, Dan K. West, Peter D. Olmsted, D. Alastair Smith, and Sheena E. Radford. Mechanically unfolding the small, topologically simple protein l. *Biophysical Journal*, 89(1):506 – 519, 2005.

[187] Hui Lu, Barry Isralewitz, André Krammer, Viola Vogel, and Klaus Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, 75(2):662 – 671, 1998.

[188] Jonathan J. Booth and Dmitrii V. Shalashilin. Fully atomistic simulations of protein unfolding in low speed atomic force microscope and force clamp experiments with the help of boxed molecular dynamics. *The Journal of Physical Chemistry B*, 120(4):700–708, 2016.

[189] Arun P. Wiita, Sri Rama Koti Ainavarapu, Hector H. Huang, and Julio M. Fernandez. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proceedings of the National Academy of Sciences*, 103(19):7222–7227, 2006.

[190] Arun P Wiita, Raul Perez-Jimenez, Kirstin A. Walther, Frauke Gräter, B. J. Berne, Arne Holmgren, Jose M. Sanchez-Ruiz, and J. S. Fernandez. Probing the chemistry of thioredoxin catalysis with force. *Nature*, 450 7166:124–7, 2007.

[191] Michel Grandbois, Martin Beyer, Matthias Rief, Hauke Clausen-Schaumann, and Hermann E. Gaub. How strong is a covalent bond? *Science*, 283(5408):1727–1730, 1999.

[192] Marina Putzu, Frauke Gräter, Marcus Elstner, and Tomáš Kubař. On the mechanism of spontaneous thiol–disulfide exchange in proteins. *Physical Chemistry Chemical Physics*, 20(23), 2018.

Appendix A

Appendices of Chapter 3

	S-state		I-state	
	φ	ψ	φ	ψ
Aib1	-46°	-16°	-62°	-17°
Asn2	-65°	-22°	-71°	-25°
Ile3	-69°	-30°	-68°	-26°
Ile4	-86°	-31°	-83°	-15°
Aib5	-57°	-26°	-57°	-23°
Pro6	-61°	-23°	-61°	-15°
Leu7	-71°	-26°	-62°	-16°
Leu8	-99°	-17°	-76°	-14°
Aib9	-59°	-21°	-59°	-22°
Pro10	-66°	-20°	-68°	-14°

Table A.1 Backbone dihedral angles of the dominant conformations of HZ observed in the unrestrained simulations of HZ immersed in the DMPC bilayer. Structures within 0.15 of each of the first two principal components were selected from the projection of each trajectory on the components obtained from principal component analysis performed on the combined trajectories. Presented are the averaged values of (φ, ψ) in these structural ensembles.

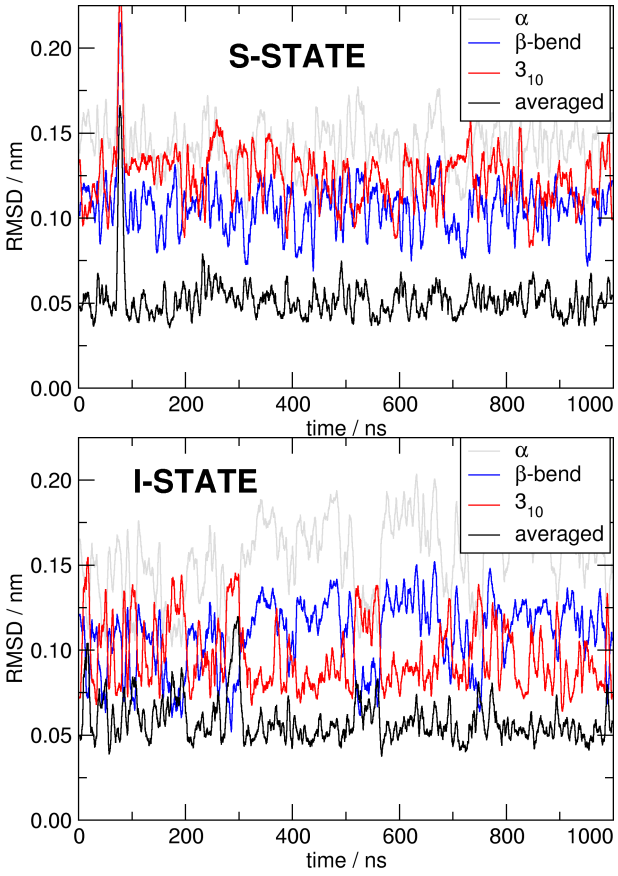


Fig. A.1 Time course of RMSD of the HZ molecule with respect the selected reference structures, in the unrestrained simulations of the S-state (top) and the I-state (bottom). The calculation involved the heavy atoms of the backbones of amino acids 3 through 10.

Appendix B

Appendices of Chapter 4

Peptide	Min1		Min2		Min3	
	Popul.	ΔG kcal/mol	Popul.	ΔG kcal/mol	Popul.	ΔG kcal/mol
ACACACA	0.005	3.13	0.008	2.85	0.987	0
ACAACAACA	0.056	0	0.001	3.85	0.342	0.39
ACAAACAAACA	0.072	1.50	0.021	2.23	0.907	0

Table B.1 Populations of energy minima for the disulfide exchange in peptides ACA_nCA_nCA.

Peptide	TS12 deg	TS23 deg	TS13 deg
ACACACA	159	160	151
ACAACAACA	157	154	160
ACAAACAAACA	156	164	156

Table B.2 Geometry of transition states for the disulfide exchange in peptides ACA_nCA_nCA.

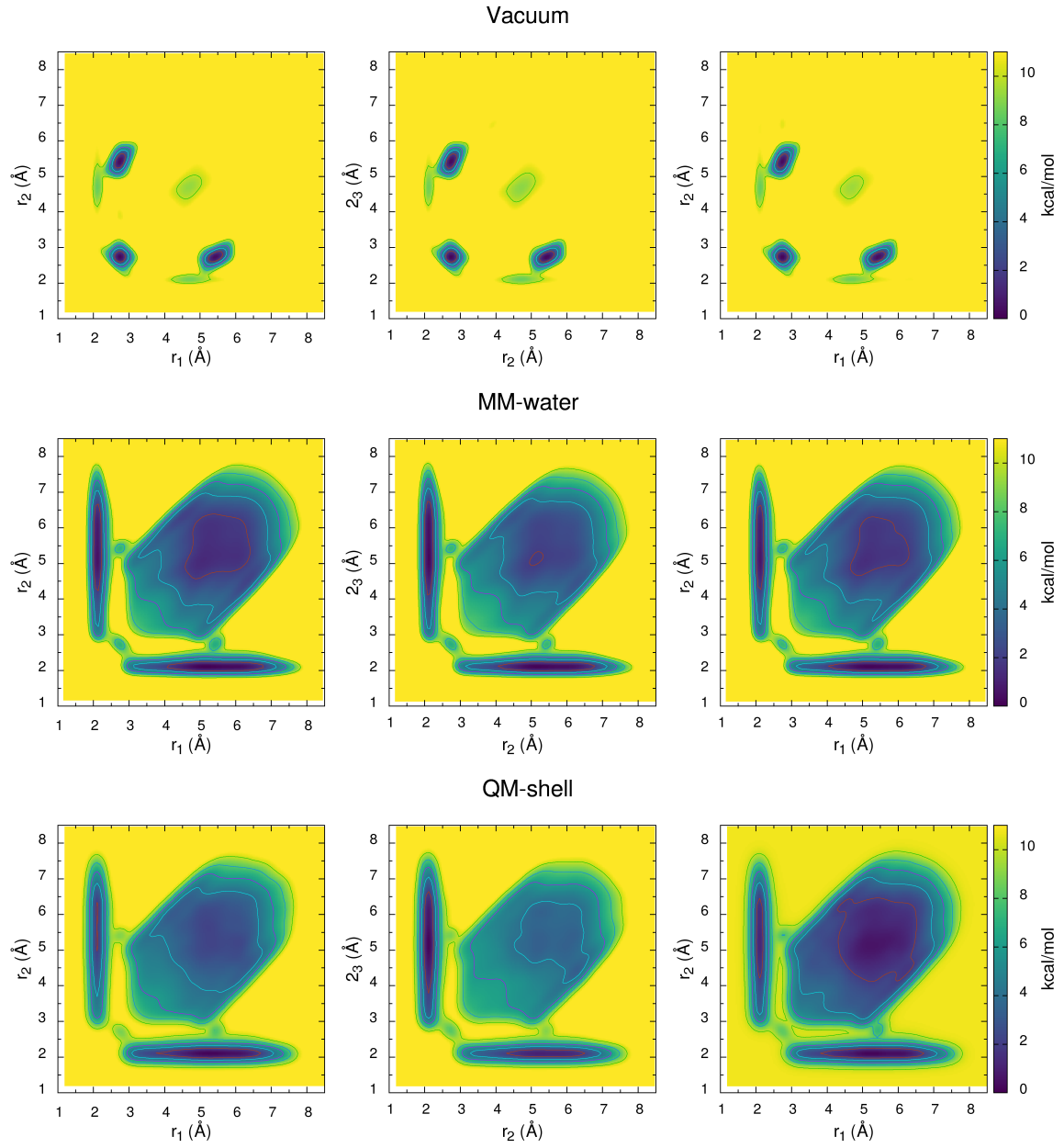


Fig. B.1 Additional representations of free energy surfaces for the dimethyldisulfide–methylthiolate system.

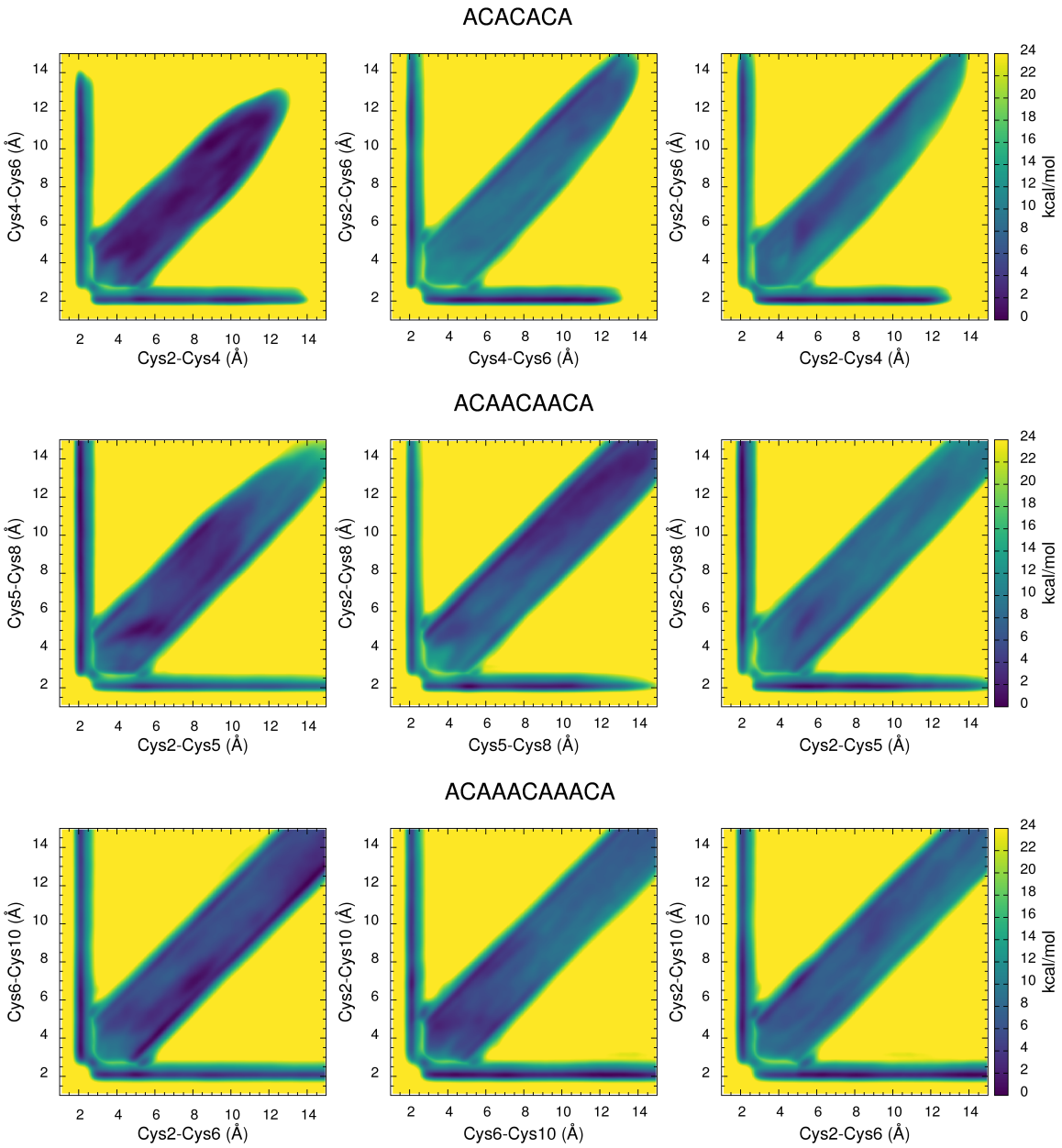


Fig. B.2 Additional representations of free energy surfaces for all of the small peptides considered.

List of Publications

1. Marina Putzu, Sezgin Kara, Sergii Afonin, Stephan L. Grage, Andrea Bordessa, Grégory Chaume, Thierry Brigaud, Anne S. Ulrich, and Tomáš Kubař. Structural behavior of the peptaibol harzianin hk vi in a dmpc bilayer: Insights from md simulations. *Biophysical Journal*, 112(12):2602 – 2614, 2017

2. Stephan L. Grage, Sezgin Kara, Andrea Bordessa, Véronique Doan, Fabio Rizzolo, Marina Putzu, Tomáš Kubař, Anna Maria Papini, Grégory Chaume, Thierry Brigaud, Sergii Afonin, and Anne S. Ulrich. Orthogonal 19f-labeling for solid-state nmr spectroscopy reveals the conformation and orientation of short peptaibols in membranes. *Chemistry - A European Journal*, 2018

3. Marina Putzu, Frauke Gräter, Marcus Elstner, and Tomáš Kubař. On the mechanism of spontaneous thiol–disulfide exchange in proteins. *Physical Chemistry Chemical Physics*, 20(23), 2018

Acknowledgements

First of all, I would like to express my gratitude to Prof. Marcus Elstner for giving me the opportunity to work on his group and for his support over the entire time of my PhD.

Second, but not for importance, a special thank goes to Dr. Tomáš Kubař, who guided me through the work done in this thesis. Tomáš door was always open for discussion and he always provided help and advice on innumerable occasions. He also carefully read through the manuscript of this thesis. I can't thank him enough for all he did.

Thanks to Dr. Sergii Afonin and Dr. Sezgin Kara for the productive discussions and the collaboration regarding the Harzianin project.

Thanks to Prof. Frauke Gräter from the Heidelberg Institute for Theoretical Studies for the fruitful discussions about the thiol-disulfide exchange project and for providing the structures of the unfolded I27* domain.

A big thank is owed to all my colleagues past and present in the theoretical chemical biology group. I would especially like to thank Franziska Wolff, Mila Andreeva and Julian Kranz for their help in the German translation of the abstract, and Thilo Mast, who provided me valid assistance in many situations during the last three years.

Finally on a more personal note I would like to thank all of my friends and my family for their support during the course of my PhD studies. And last, and certainly not least, I thank Alessandra for always being there, and for believing in me even when I didn't. This work would not have been remotely possible without her love and support.