Structural Studies of Biomolecular Systems with Molecular Dynamics Simulations

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

M.Sc. Violetta Franziska Roswitha Schneider

aus Grünstadt (Pfalz)

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

To my parents and my brother Jonathan
for their tireless support

"...if we were to name the most powerful assumption of all,
which leads one on and on in an attempt to understand life,
it is that all things are made of atoms,
and that everything that living things do
can be understood in terms of the jigglings and wigglings of atoms."

— *Richard Feynman (1918-1988)*

Die vorliegende Arbeit wurde im Zeitraum von Dezember des Jahres 2014 bis April des Jahres 2018 im Arbeitskreis Theoretische Chemische Biologie, unter der Leitung von Prof. Dr. Marcus Elstner, am Institut für Physikalische Chemie des Karlsruher Instituts für Technologie (*KIT*) angefertigt.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst habe sowie die benutzten Hilfsmittel und Quellen vollständig angegeben habe. Darüber hinaus versichere ich, dass ich die Regeln zur Sicherung guter wissenschaftlicher Praxis am Karlsruher Institut für Technologie (*KIT*) in der gültigen Fassung beachtet habe.

Karlsruhe, den 7. März 2018

Violetta Schneider

Danksagungen

Zusammenfassung

In vorliegender Arbeit wurden computerbasierte Simulationen eingesetzt, um verschiedene strukturelle Eigenschaften und Dynamiken von Biomolekülen aufzuklären.

Die bearbeiteten Themen resultieren aus Kooperationen innerhalb des Graduierten-kolleges 2039 *Molekulare Architekturen für die fluoreszente Bildgebung von Zellen.* Dabei wurden sowohl klassische Molekulardynamik (MD)-Simulationen verwendet als auch Simulationen mit Verwendung sogenannter *coarse-grained*-Ansätze, bei denen mehrere Atome zu größeren Einheiten zusammengefasst werden.

Aus dem Gebiet der Peptid-Membran-Wechselwirkungen wurden in Zusammenarbeit mit dem Arbeitskreis Ulrich (KIT) zwei hydrophobe Peptide aus Typ I Toxin-Antitoxin-Systemen bearbeitet.

Im Mittelpunkt von Kapitel 3 steht das stress-induzierte Peptid tisB (29 Aminosäuren) aus *E. coli*, welches die Formation von Biofilmen bewirkt und den elektrochemischen Gradienten an der Bakterienzellmembran stilllegen kann.

Es wurden Studien der Assemblierung des Peptides tisB als paralleles Dimer und Te-tramer in der Lipidmembran durchgeführt. Die anschließenden langen Simulationen ausgewählter Strukturen wurden mit experimentell gefundenen Ergebnissen verglichen, wobei das Hervortreten des *charge zipper*-Motivs als Interaktionsmotiv bestätigt werden konnte. Für das Dimer und das Tetramer konnten erstmals Strukturen für einen mög-lichen Protonentransfer über die Membran aufgespürt werden. Das Tetramer zeichnet sich durch seinen Aufbau aus zwei antiparallelen tisB-Dimeren aus, welches einen stän-digen Wasserfaden über das entstandene polare Interface erlaubt, und erscheint als sehr vielversprechende Ausgangsstruktur für mögliche Protonentransferstudien.

In Kapitel 4 sind erste strukturelle Studien zum Peptid bsrG (38 Aminosäuren) aus *B. subtilis* dargestellt. Es konnte gezeigt werden, dass bsrG als Monomer nicht in Lipidmem-branen integriert, aber an die Oberfläche assoziiert ist. Die polaren Aminosäuren sind dabei zum Wasser hin ausgerichtet. Mittels Assemblierungsstudien des Peptides als Dimer in der Lipidmembran konnten stabile parallele und antiparallele Strukturen eluiert werden, die mit experimentellen Ergebnissen übereinstimmen und das Bindungsmotiv des *hydrogen bond zipper* zeigen.

In Zusammenarbeit mit den Arbeitskreisen Wagenknecht, Nienhaus und Schepers (alle KIT) wurden MD-Simulationen zur Aufklärung des unterschiedlichen Verhaltens zweier Doppelstrang-RNA-Konstrukte mit FRET-Farbstoffen (arabino- und ribo-Konfiguration) bei der Effizienz des FRET-Transfers eingesetzt (Kapitel 5). MD-Simulationen konnten aufzeigen, dass dieser Unterschied in der Effizienz auf Grundlage der erhöhten Mobilität der arabino-konfigurierten Farbstoffe zu erklären ist, im Gegensatz zu den sehr stark wechselwirkenden und unbeweglichen ribo-konfigurierten Farbstoffen. Jene Mobilität ermöglicht einen effizienten Energietransfer zwischen den Farbstoffen, der zu FRET führt.

Abstract

In this work, computer-based simulations were performed to elucidate different structural properties and dynamics of biomolecules.

The topics resulted from collaborations within the Research Training Group 2039 *Molecular architectures for fluorescent imaging of cells.* Classical Molecular Dynamics (MD) simulations and coarse-grained simulations were used, in which several atoms are combined into larger units.

In the field of peptide-membrane interactions, two hydrophobic peptides from type I Toxin-antitoxin systems were investigated in cooperation with the Ulrich research group (KIT).

Chapter 3 focuses on the stress-induced peptide tisB (29 amino acids) from *E. coli*, which causes the formation of biofilms and can incapacitate the electrochemical gradient on the bacterial cell membrane.

Assembly studies of the peptide tisB as a parallel dimer and tetramer in the lipid membrane were carried out. The subsequent long simulations of selected structures were compared with experimentally found results, whereby the emergence of the *charge zipper* motif could be confirmed as important interaction pattern.

For the first time, structures for a possible proton transfer via the membrane could be detected for the dimer, and also for the tetramer. The tetramer is characterised by its structure consisting of two antiparallel tisB dimers, which allow a constant water filament over the polar interface, and appears as a very promising starting structure for possible proton transfer studies.

The first structural studies on the peptide bsrG (38 amino acids) from *B. subtilis* are presented in chapter 4.

It has been shown that bsrG as a monomer is not integrating into lipid membranes, but is associated to the surface with the polar amino acids directed to water. By means of dimer assembly studies of the peptide in the lipid membrane stable parallel and antiparallel structures could be elucidated which coincides with experimental results and shows the binding motif of the *hydrogen bond zipper*.

In cooperation with the working groups Wagenknecht, Nienhaus and Schepers (all KIT), MD simulations were used to elucidate the different behaviour of two double-stranded RNA constructs with FRET dyes (arabino- and ribo-configuration) in the efficiency of FRET transfer (Chapter 5).

MD simulations have shown that this striking difference in efficiency can be explained by the increased mobility of the arabino-configured dyes, as opposed to the highly interactive and stacking ribo-configured dyes. This mobility of the arabino-configured dyes enables an efficient energy transfer between the dyes which leads to FRET.

Contents

Abbreviations

aa all-atom or amino acid

AMBER Assisted Model Building with Energy Refinement, a force field

arabino Arabinose

B. subtilis *Bacillus subtilis*, a gram-positive bacterium

bsrG bsrG is a peptide (toxin) found in *B. subtilis* from a type I Toxin-antitoxin system

C. albicans *Candida albicans*, a yeast

C. elegans *Caenorhabditis elegans*, a nematode

CD Circular Dichroism

CG Coarse-Grained

CHARMM Chemistry at Harvard Macromolecular Mechanics, a force field

DAFT Docking Assay For Transmembrane peptides

DEPC 1,2-dierucoyl-sn-glycero-3-phosphocholine

DFT Density Functional Theory, a quantum chemical calculation method

DMPC 1,2-dimyristoyl-sn-glycero-3-phosphocholine

DMPG 1,2-dimyristoyl-sn-glycero-3-phosphoglycerol

DNA Desoxyribonucleic Acid

DOPC 1,2-dioleoyl-sn-glycero-3-phosphocholine

DPC n-dodecylphosphocholine

DPPC 1,2-d̲i̲palmitoyl-sn-glycero-3-p̲hosphoc̲holine

ds d̲ouble s̲tranded

E. coli *Escherichia c̲oli*, a gram-negative bacterium

ESP E̲lectrostatic S̲urface P̲otential

FISH F̲luorescence *In S̲itu* H̲ybridisation

FRET F̲örster R̲esonance E̲nergy T̲ransfer

GAFF G̲eneral A̲mber F̲orce F̲ield

GPU G̲raphic P̲rocessing U̲nit

GROMACS G̲roningen M̲achine for C̲hemical S̲imualtions, Software package for MD-simulations

HBV H̲epatitis B̲ V̲irus

HCV H̲uman C̲ V̲irus

HeLa H̲uman cervix carcinoma cells, taken from H̲enriette L̲ack

HIV H̲uman I̲mmunodeficiency V̲irus

in vitro from latin for "in glass", means reactions and processes in laboratory vessels respectively in an artificial environment

in vivo from latin for "in one that is living", means reactions and processes which take place in the living organism under physiological conditions

KIT K̲arlsruhe I̲nstitute of T̲echnology

MD M̲olecular D̲ynamics

mRNA m̲essenger R̲NA

NMR N̲uclear M̲agnetic R̲esonance

nt n̲ucleot̲ide

OCD Oriented Circular Dichroism

PBC Periodic Boundary Conditions

PME Particle-Mesh Ewald

POPC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

POPE 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine

POPG 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol

QM Quantum Mechanics

RESP Restrained ElectroStatic Potential

ribo Ribose

RMSD Root Mean Square Deviation

RMSF Root Mean Square Fluctuation

RNA Ribonucleic acid

RNAi Ribonucleic acid interference, RNA-silencing

S. aureus *Staphylococcus aureus*, a gram-positiv bacterium

SDS-PAGE sodium dodecyl sulfate–polyacrylamide gel electrophoresis

siRNA small interfering RNA

SLIPIDS Stockholm Lipids, a force field

TA Toxin-antitoxin

tisB toxity induced by SOS response B, type I Toxin-antitoxin peptide from *E. coli*

1 Introduction and Biological Background

1.1 Introduction to Biomolecular Simulations

Due to the increasing computational power over the last years it became possible to simulate large biomolecular systems. Therefore, the role of computation in the field of fundamental research has expanded. Since only a few properties of biological systems are accessible by experiment, simulations are a versatile tool to shed light on structural behaviour over time, conformational distributions and interactions. Previously, they were used as a graphical tool to illustrate experimental data sets.

Molecular Dynamics simulations (MD simulations) provide a microscopic view at timescales, space and energy which cannot be detected in experiments. Relevant timescales of biological properties range from femtoseconds (bond length vibration) to milliseconds (protein folding) and beyond. These timescales are nowadays accessible via atomic-detail simulations (see figure 1.1) [1].
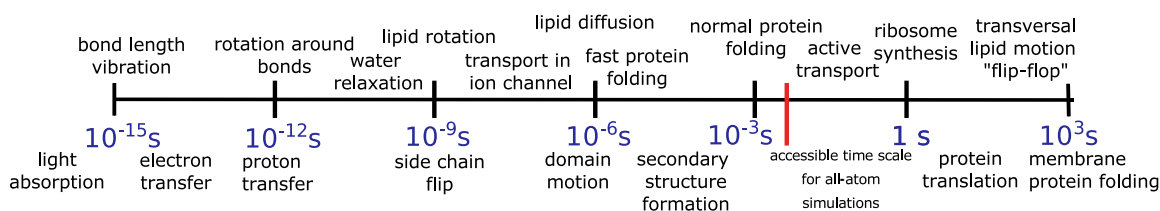
Figure 1.1: Range of timescales in biomolecular simulations in comparison with timescales in biological systems. Picture modified from [1].

Techniques like Electron Microscopy and X-Ray Microscopy can only provide static information. Other techniques, which provide dynamical information like Förster Resonance Energy Transfer (FRET) and Electrophysiology, can only reach timescales up to microseconds. Nuclear magnetic resonance (NMR) methods offer a wide range from milli- to nanoseconds but MD simulations are the appropriate technique for an extended time range from milli- to femtoseconds which cover all fundamental molecular

processes [2]. One important step for this progress was the parallelisation of computer processes which enabled computational power with fewer costs. Parallelisation means using multiple computer processors, which are networked, at the same time to calculate forces. Also, the use of GPUs (graphic processing units) - initially designed for video games - were responsible for the great leap forward in terms of timescale and size of simulations [3].

Because of the fundamental questions in biomolecular simulations based on weak, non-bonded interactions, not electron transfer and bond breaking, it is possible to describe these on atomic level with classical simulations (atoms described as balls and bonds as springs) by a classical potential-energy function. Also, these systems are microscopic at non-zero temperatures and can be depicted with classical statistical mechanics. Thus, the state of the biomolecular system can be described by an ensemble of configurations. A fixed combination of particles, volume and temperature is a canonical ensemble. Therefore, the goal of the simulations is not to find a single representative structure, but to find the minimum on the potential energy surface (PES) [4].

In addition to the mentioned benefits Molecular Dynamics simulations also have some limitations. Van Gunsteren (ETH Zürich) has identified four main problems for biomolecular modelling [5]:

- the force field problem

- the sampling problem

- the ensemble problem

- the experimental problem

The force field problem
The force field consists of potential energy terms which represent the covalent and non-covalent interactions. The critical step is the force field parametrisation. The bottleneck for improvement is composed of the dealing with electronic polarisation and entropic effects. Furthermore, the force field should be transferable to a variety of compounds. This requirement contrasts with the approach to keep the force fields as simple as possible.

The sampling problem
Biomolecular systems exhibit a large number of degrees of freedom. The potential energy surface is very rugged and complicates the search for the

global minimum. Hence, it is difficult to obtain a good sampling of the surface, and the goal is to smooth the potential energy surface. Therefore, it is possible to start from X-ray structures or develop new techniques which allow a larger sampling in shorter time or the smoothing of the surface. Applied methods are *Metadynamic simulations, Replica Exchange MD simulations* or *Temperature Annealing*.

The ensemble problem

The behaviour of biomolecular systems is governed by statistical mechanics. The system is not generally characterised by a single structure but by a Boltzmann ensemble of structures. It is easier to obtain a single structure from simulations instead of a representative ensemble.

The experimental problem

Experimental data is not easy to compare because it is an average over time and space and does not yield the direct data. Also, the accuracy of the experimental data is insufficient to be compared with results from simulations.

With the right choices for the prospect of interest and with respect to the available computer power, accuracy and methods, MD simulations are a powerful tool to study biomolecular systems. In this work, the structural behaviour of RNA, tagged with a fluorescent FRET-dye pair, is examined.

Another important field of interest in the biomolecular context is the behaviour of proteins inside membranes. Approximately 30 % of proteins in eukaryotes are membrane proteins [6] and from these only circa 1 % are available as crystal-structure. Astonishingly, more than half of all drug targets are membrane proteins [7]. These proteins can interact as transporters or channels, can be membrane-spanning, dipping to the hydrophobic core or be attached in a variety of conformations to the biomembrane surface. For experimentalists, it is difficult to map the structure of a protein inside the two-dimensional liquid-crystalline structure of a biomembrane.

The timescales for processes inside the membrane exhibit a wide range. The range differs from femtoseconds for bond vibrations, several picoseconds for the rotation around their long axis, nanoseconds for the lateral diffusion of the lipids and up to seconds (protein-catalysed) or even days (uncatalysed) for the *flip-flop* transverse-diffusion (see figure 1.1) [8]. The dynamics of the protein is slowed down in the membrane environment, as well.

Concerning the huge atomic system size and the required quite a few microseconds short timescale to sample, all-atom MD simulations are not sufficient. The computational costs and the required time are too high, even with the increased computational speed due to GPUs and parallelisation. Therefore, the development of simplified *coarse-grained* models (see section 2.2) to describe proteins and lipids has been advanced. Such models consist of particles which describe three to five heavy atoms at the same time. In addition, the non-bonded interactions are parametrised to represent the overall thermodynamic properties. The advantages of these models are for instance the decreased particle number, the reduced interactions and the longer simulation time steps. Hence, microsecond to second timescales are attainable which provides the opportunity to shed light on the membrane protein interactions [9].

In this work, two hydrophobic membrane proteins from bacteria *tisB* and *bsrG* are in the focus of interest. The goal is to illuminate and screen the structural arrangement inside the membrane with all-atom simulations and a new coarse-grained screening method, called DAFT, combined with all-atom simulations based on the best found structure afterwards.

Over the last decades the evolution in the field of Molecular Dynamics simulations has expanded in accuracy, method development, speed and simulation size more than anyone could imagine. As a result, the Nobel Prize in Chemistry 2013 was awarded "for the development of multiscale models for complex chemical systems" to Martin Karplus, Michael Levitt and Ariel Warshel [10].

In the near future, the combination of enhanced sampling techniques (algorithms which speed up the exploration of molecular conformations by altering the physics of the system), computational effort (new supercomputers and cheaper CPUs) and improved accuracy of the force fields (as an example *reactive* force fields which can describe bond breaking) will lead to the design of new therapeutics and protein designs for applications in medicine as drugs. Perhaps, dreams will come true and the simulation of whole cells, not only simplified models, and the folding of proteins or RNA become possible [3, 11].

1.2 The Toxin-antitoxin Systems bsrG/SR4 and tisB/IstR

Two of the major challenges of modern medicine is the fight against chronic bacterial films and the demand for new antibiotics to address the problem of multidrug resistance. Each year, 2 million people in the United States of America become infected with bacteria that are resistant to antibiotics. As a result, about 23,000 people die immediately, and many more die from long-term conditions which are related to antibiotics and resistance [12].

In 1928, Alexander Fleming discovered penicillin as the first antibiotic agent [13]. Two other scientists, Howard Florey and Ernst Chain, took an interest in penicillin large-scale production and devised methods right in time for the Second World War. Together, they were awarded the Nobel Prize in 1945 [14].

One year before, in 1944, Joseph Bigger, a medical military doctor, was experimenting with penicillin. Joseph Bigger found out that on addition of *Staphylococcus aureus* (*S. aureus*) the desired effect of lysis ensued. Then, he plated the serum and observed surviving colonies. These colonies were treated again with penicillin and a lysis occurred. In the process he also detected a small new subpopulation of remaining living cells. He called them *persister*, sleeping cells, to distinguish them from mutants [15]. He concluded that penicillin cannot sterilise infections. Later, this observation was confirmed for all antibiotics, but forgotten in the following years as antimicrobial resistance was discovered, and some new antibiotics were established.

Approximately forty years after Alexander Flemings groundbreaking discovery, Harris Moyed received this topic, but it was only in the early 1990s that the victory over infections was revealed as an illusion since chronic infections are usually connected with biofilms (which are densely packed ensembles of microbial cells that are associated with a surface and enclosed in a matrix of polysaccharide material) and in consequence with a small fraction of persister cells [16].

Unfortunately, the mechanism of how persister cells form and how resistance occurs is not yet fully understood. With this knowledge, a specific treatment would be accessible. Chronic infections like dental diseases, cystitis, encocarditis and cystic fibrosis are often accompanied by the formation of biofilms. These are indicated as the source of the problem and responsible

for resistance against antibiotics [17, 18]. Analysis of the biofilm showed that the bulk of cells in the biofilm are killed by antibiotics, but a small fraction of persister cells remain alive [16, 19].

1.2.1 Persister Cells

Persister cells are not mutants, but phenotypes of wild type cells which grow very slowly. They are mostly found in bacteria, but could also be found in fungus *Candida albicans* (*C. albicans*). One of their major feature is the high tolerance to antibiotics [16]. In presence of antibiotics, or other external processes like carbon source transitions, persister cells slow down their metabolism and reduce their growth. Accordingly, they were called *dormant* cells. There is also a hint that they pre-exist or stochastically form as non growing cells from the stationary phase [20].

Studies have revealed that persister cells are responsible for chronic infections due to biofilms [16, 21]. But persisters cannot be used as antibiotic-resistant *pharmaceuticals*. They differ in phenotype, have sub-populations with phenotypic heterogeneity, and these phenotypes have a high tolerance towards antibiotics and are non-dividing. Until the initial stress (environmental stress, nutrition reduction or antibiotic treatment) is removed, they stay in the *dormant* state [22]. More importantly, persister cells do not interact with or influence other cells, they act only inside themselves to pass into the *dormant* state [23].

Up to now, the mechanism of their formation is not fully understood, but there are strong hints that the Toxin-antitoxin systems (TA systems) (chapter 1.2.2) are involved in persister formation. Especially in *E. coli* these TA systems are associated with the formation of persister cells.

With more knowledge about the TA systems drugs could be developed which prevent or at least reduce persistence/resistence. Also, biological processes, which regulate the TA levels, could be influenced. With the knowledge about the molecular basis new therapy forms against chronic diseases and multi-drug resistance can evolve [22].

1.2.2 Toxin-antitoxin Systems

Toxin-antitoxin (TA) genes are found in bacterial plasmids, chromosomes and phages. Nearly all TA genes are encoded in *E. coli* (15 TA loci) and *Mycobacterium tuberculosis* (80 TA loci) [24]. Toxin-antitoxin modules exhibit different influences on physiology, molecular mechanisms, cellular targets or the nature of their gene products.

They have something in common: they interact simultaneously as a toxin and an antitoxin. The toxin is a protein which causes zero growth by interfering with vital cellular processes. On the other hand, the antitoxin neutralises the toxin and is degraded. The antitoxin can be a non coding RNA or a low molecular weight protein. The antitoxins regulate their toxins with a variety of different mechanism such as: translational repression, inhibition of ribosome standby sites and mRNA (messenger RNA) degradation [20].

Toxin-antitoxin systems are divided into six classes I-VI, whereof classes IV-VI were adopted recently [22]. The function and regulatory mechanism of the six types are summarised in figure 1.2. The best studied TA system is the type I system *hok/sok* from *E. coli* [25].

- Type I: a toxic protein and a small non coding RNA (sRNA) as antitoxin.

  The sRNA base pairs to the mRNA of the protein and inhibits its translation. This leads to inhibition of the ribosome binding. Therefore, the mRNA will be degraded by RNase III. Upon stress the antitoxin sRNA is reduced, and this results in the translation of the toxin. Type I toxins are small hydrophobic peptides < 60 aa (amino acids) with transmembrane domains. These α-helical transmembrane domains form pore like structures which force the loss of the membrane potential and interrupt cells growth [26].

- Type II: a toxic protein and a protein as antitoxin.

  The type II Toxin-antitoxin system is well-studied. One prominent system is *relB/relE* [27]. On the one hand the antitoxin protein can bind directly to DNA and on the other hand directly to the toxin and inhibit the activity. Both proteins form a complex which regulates the *Operon* that encodes the proteins and is the toxin co-repressor. The balance between toxin and antitoxin is regulated by a distinct

difference in the cellular lifetime. Overall, the toxin is stable, and the antitoxin is proteolysed. Exposed to stress the antitoxin is degraded which leads to zero growth, inhibition of replication and translation as well as activation of the ribosome elongation factor.

- Type III: a toxic protein and a small-RNA antitoxin.

  Instead of duplexing with the toxin mRNA the antitoxin binds directly to the toxic protein via pseudo-knots. The toxic protein is an endoribonuclease. Two functions were described for Type III systems. First, an abortive infection ability, which protects bacterial populations from invading bacteriophages, and secondly a role in the stabilisation of plasmids [28].

- Type IV: a toxic protein and an antitoxin protein.

  In contrast to Type II, these proteins never interact. The toxin prevents growth of the cytoskeleton by binding and inhibiting of the polymerisation. This hinders cell division. On the other hand, the antitoxin as antagonist promotes and stabilises the cytoskeleton proteins.

- Type V: a toxic protein and a RNase as antitoxin.

  Only one system is known. The antitoxin RNase cleaves the toxin mRNA. Upon stress the antitoxin RNase is degraded. The toxin, which is a small hydrophobic peptide, damages the cell membrane.

- Type VI: a toxic protein and an antitoxin protein.

  In 2012, the *socAB* operon was identified as Toxin-antitoxin system [29]. The toxin blocks and represses the replication elongation by binding directly to the sliding clamp. On the other hand, the antitoxin is a proteolytic adapter protein which forces a protease-mediated degradation of the toxin.

Two examples for Type I TA systems are *tisB/IstR1* and *bsrG/SR4* [27]. Both are in the focus of this work. Interestingly, the biological mode of action is different. *tisB* is known for formation of persister cells and interacts as membrane-spanning peptide inside the inner bacterial membrane which leads to depletion of the cell's ATP level. For *bsrG* no clear mechanism is verified yet (see section 1.2.4).
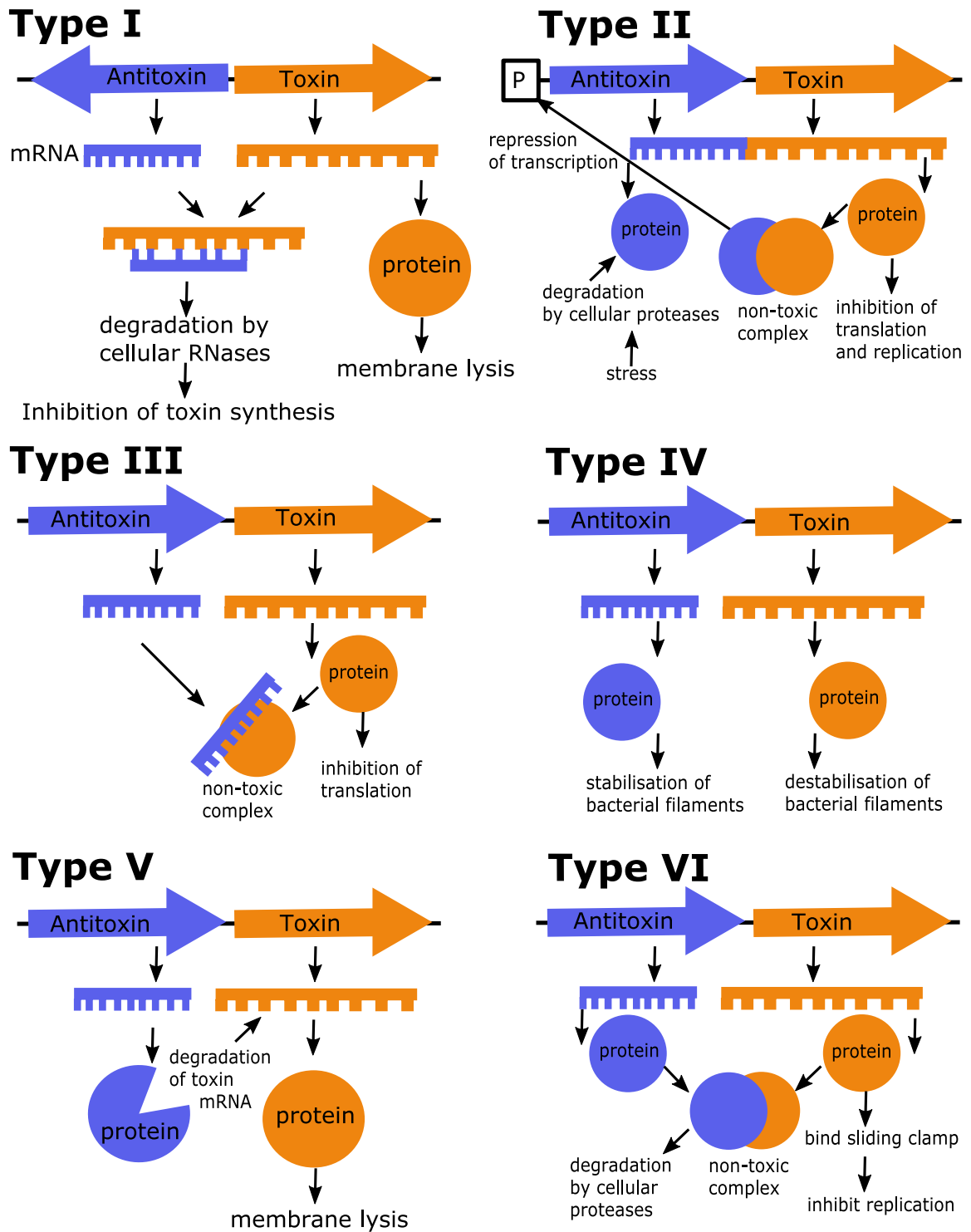
Figure 1.2: Schematic depiction of type I-VI toxins and antitoxins mechanisms. Toxins are coloured in orange, antitoxins in blue. Detailed information are given in the text. The picture is modified from [30].

1.2.3 The tisB Peptide

tisB is an extraordinarily hydrophobic membrane spanning peptide from *E. coli* with twenty-nine amino acids [31]. It is involved in stress response via the *TisB-IstR* Toxin-antitoxin (TA) system from type I. Hence, the name derives from toxicity induced by SOS response B. In this system the tisB protein is the toxin and IstR (Inhibitor of SOS-induced toxicity by RNA) is the antitoxin, a non-coding small RNA.

The sequence is:

$$\text{MNLVDIAILILKLIVAALQLLDAVLKYLK}$$

There are only few polar and charged amino acids. Hydrophobic amino acids (coloured in black) form the main part. Three positively charged lysines are coloured in dark blue, two charged aspartic acids in red, and one polar glutamine and one polar asparagine in light blue. This results in a total charge of **+1**. In figure 1.3 the representation of the tisB sequence as helical wheel plot and helical mesh plot is depicted. These were generated with the web-based tool *Protein Origami* [32]. The helical wheel plot suggests that tisB is an amphiphilic alpha-helical protein. In addition, the helical mesh plot illustrates that all charged amino acids are located on a narrow strip along the helical axis.

1.2.3.1 Functional Hints from Biological Experiments

TisB is an unusual Type I TA system, which is chromatosomally encoded and is involved in the SOS-stress response and persister formation [23].

The tisB genome exhibits two bi-directional reading frames and a LexA binding site. LexA is involved in the SOS response and is the master regulator of the DNA damage thereof. Under normal conditions, LexA represses the transcription of tisB and expresses IstR1 [23]. DNA damage leads to activation of the SOS-response which degrades LexA via self-cleavage and leads to transcription of the tisB mRNA. The tisB mRNA is subject to a complex translational control with several active/inactive transcripts.

Two regulated bio-functional elements follow: first, the depolarisation of the membrane, and second, the switching of some cells to the dormant state due to DNA damage during exponential growth which leads to persister formation and also to stress-adaption [20].
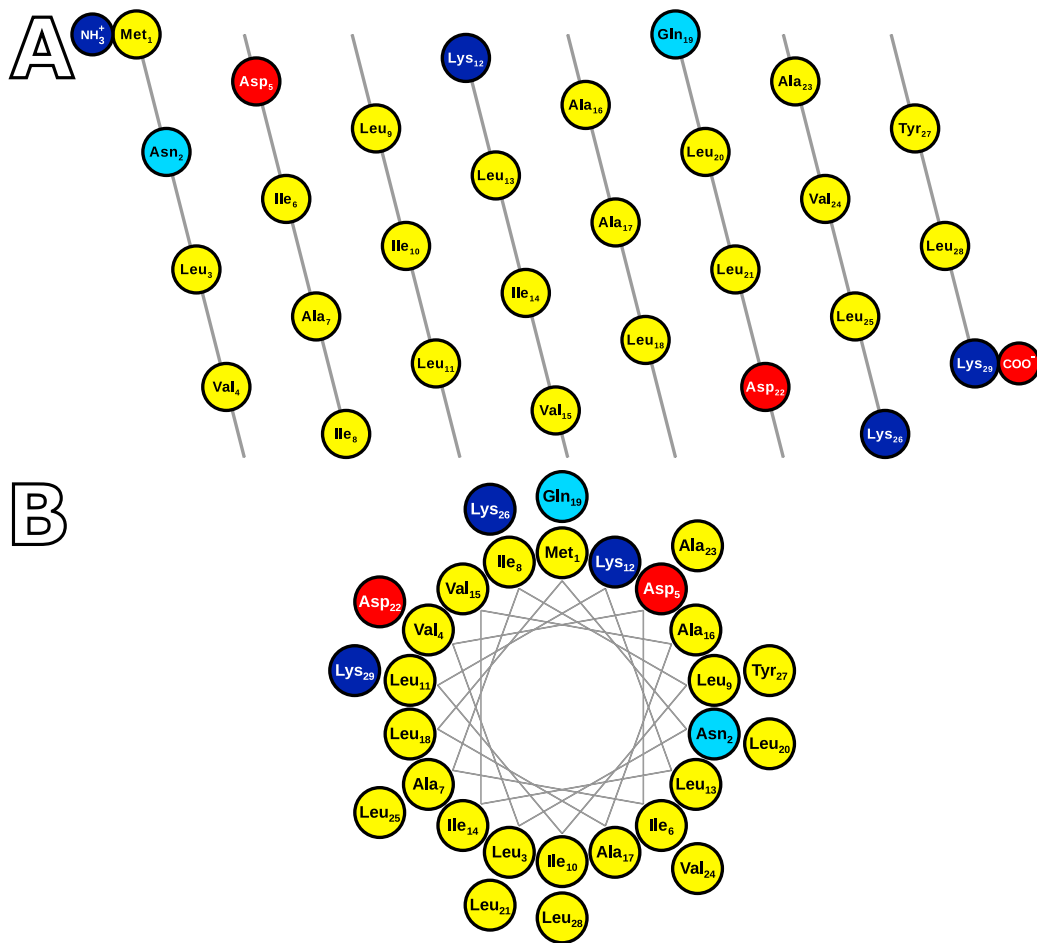
Figure 1.3: Representation of tisB as helical wheel [A] and helical mesh plot [B]. Suggesting that tisB is an amphiphilic alpha-helical peptide with few charged amino acids. The charged and polar amino acids are located on a narrow strip along the helical axis. Hydrophobic amino acids are coloured in yellow, negatively charged in red, positively charged in blue and the polar amino acids in light blue. The amino acids are depicted in three letter code.

TisB directly interacts with the cytoplasmic side of the inner bacterial membrane. This leads to the depletion of the pH gradient and the ATP level of the cells, which disables the macromolecular synthesis, replication, translation and transcription [21, 33]. The tisB protein forms multi-state ion conductive pores with a low diameter. The ionic selection correlates with the net charge of the protein. Thus, the positive net charge is important for the action as anionic selective pore [34]. This intervention into the metabolism and the degradation of the energy level results in the formation of persister cells and the occurrence of biofilms.

First coarse-grained simulations combined with oriented circular dichroism experiments showed that tisB favourably inserts in a transmembrane alignment [35]. Furthermore, it does not prefer pore formation, but slowly equilibrates a proton gradient in vesicle permeation assays.

Notably, experiments to prove these assumptions were performed by overexpression which leads to cell death. The mild expressions during SOS-response may lead to active DNA repair and persister formation, and that will protect the cells from antibiotics [16].

1.2.3.2 *Charge Zipper* Mechanism of tisB

Several years ago, the group of Anne Ulrich (KIT) proposed a mechanism for the folding and the self-assembly of pore forming proteins. This mechanism is based on electrostatic interactions between the protein chains and suggests that the protein can be *zipped up* by a ladder consisting of several salt bridges. Proteins with these *charge zipper* motifs are involved in diverse cellular processes such as protein translocation, proton transport, antimicrobial action and molecular recognition [36].

For tisB an assembly as an antiparallel dimer, stabilised with four salt bridges, is proposed. With this polar interface the dimer could pull water into the hydrophobic bilayer core, and this could explain the triggering of biofilm formation. The sequence reveals that the positive and negative charges are complementary ordered: so the negatively charged aspartic acids on positions 5 and 22 can form salt bridges with the positively charged lysines on positions 12 and 26. Furthermore, the glutamines on position 19 stabilise the dimer with an additional hydrogen bond.

MNLVDIAILILKLIVAALQLLDAVLKYLK

KLYKLVADLLQLAAVILKLILIAIDVLNM

This is a possible *charge zipper* structure motif of tisB. Feasible amino acids for intermolecular salt bridges are coloured in green. The possible intermolecular hydrogen bond is coloured in orange.

1.2.3.3 Current structural knowledge

Based on the experimental results of the Ulrich's Group (KIT) [35], mainly from the doctoral thesis of Benjamin Zimpfer [37], the following assumptions can be made:

- Fluorescence experiments showed that tisB and several of its charge mutants can break down the proton gradient on lipid vesicles.

- Circular dichroism studies demonstrate that tisB is a highly alpha-helical protein in all studied environments. The helical conformation is not sensitive for denaturation processes, which is remarkable for a membrane peptide. Most membrane proteins adopt their functional structural state only in contact with the membrane.

- FRET-experiments proved the assumption that an active unit of tisB is (at least) an antiparallel dimer.

- Solid-state NMR experiments revealed that tisB exhibits a transmembrane alignment with a tilt of 20°-30° to the membrane normal. The lipid to protein ratio is not dependent on the orientation of tisB inside the membrane. Further experiments with charge mutants showed that the array of complementary charges, as hypothesised in chapter 1.2.3.2 (*charge zipper* model), is mandatory for biological function and precise structural arrangement. Besides, experiments at low pH, which have led to gradual reduction of the transmembrane orientation, support this hypothesis. In addition, the GLN^{19} is essential for functionality. Moreover, the observed GLN^{19}-GLN^{19} distance of less than 10 Å confirms the proximity of the helices.

1.2.4 The bsrG Peptide

BsrG is a 38 amino acid long peptide with a central transmembrane domain and a charged C-terminus from the gram-positive bacterium *Bacillus subtilis*. It is a toxin of the type I Toxin-antitoxin system (see section 1.2.2) bsrG/SR4 where bsrG is the toxic peptide and SR4 the antitoxin as small RNA. This TA system belongs to the group of *cis-encoded* sRNA-antitoxins that function via post-transcriptional regulation. Up to now, only 14 type I Toxin/Antitoxin systems are postulated for gram-positive bacteria. The gene is located on the bacterial chromosome in the SPβ-prophage region of *B. subtilis* [38]. In contrast to the tisB Toxin-antitoxin system described above this system is not induced by the SOS stress-response [39].

The bsrG RNA is 294 nucleotides (nt) long, the antisense SR4 RNA 180 nt. At the 3'-end the SR4 RNA overlaps with the RNA of bsrG by 123 nt. This interaction of both RNAs at the 3' promotes the degradation of the bsrG mRNA. RNase III cleaves this bsrG/SR4 duplex, but is not involved in the degradation. Then, the endonuclease Y and exonuclease R undertake the degradation. The RNase III is not essential for this process [38, 40].

When the bsrG/SR4 complex is formed, the SR4 induces structural alteration of the ribosome binding side of bsrG. Subsequently, the double stranded region is extended, and the translation of the bsrG mRNA is inhibited. It is the first known antitoxin with two functions: first, the inhibition of the translation, and second, the provoked degradation [41]. In figure 1.4 the two independent functions for the antitoxin are depicted schematically. The promotor for SR4 is 6 to 10 fold stronger than the bsrG promotor. Remarkably, it was found that a heat-shock of 48 °C leads to a faster degradation and a decreasing level of bsrG mRNA. That indicates a novel type I TA system: the first temperature sensitive Toxin-antitoxin system in prokaryotes [39].

The heat-shock leads to refolding of the central region of the toxin mRNA. Hence, it is easier for the RNase Y and J1 to get in contact and degrade the mRNA. This might play a role in the onset of stationary phase when the antitoxin SR4 can no longer prevent the toxin synthesis [40].
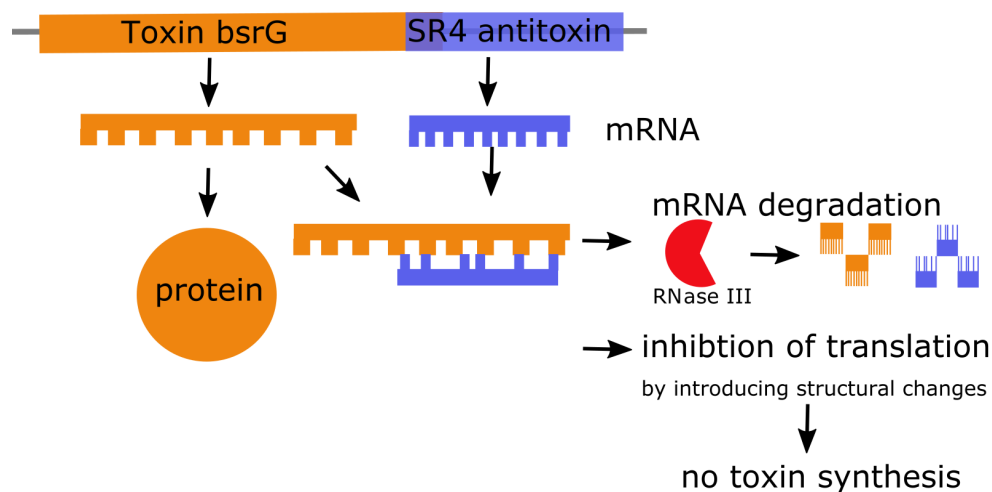
Figure 1.4: Scheme of the two independent functions of the antitoxin SR4, a *cis-encoded* sRNA, from *B. subtilis*. The illustration is adopted from [38].

The protein sequence is:

MTVYESLMIMINFGGLILNTVLLIFNIMMIVTSSQKKK

There are only few polar and charged amino acids. Hydrophobic amino acids (coloured in black) form the main part. Three positively charged lysines at the C-terminus are coloured in dark blue, the one charged glutamic acid in red, and three polar asparagines, threonines and serines in light blue. This results in a total charge of **+2**.

The predicted transmembrane domain (with the tool *SOSUI* [42]) suggests a 23 amino acids long part from positions 7 to 29. Two other algorithms, *TMpred* and *TMHMM*, suggested a transmembrane domain from positions 9 to 31. This leads to the suggestion that the three positively charged lysines with four polar amino acids before them are not part of the transmembrane domain. Furthermore, the negatively charged aspartic acid is not involved in the transmembrane domain which is in accordance to the observation that the transfer and placement of charged and polar amino acids inside the hydrophobic membrane environment is unfavourable [43].

The bsrG peptide is shown as a helical wheel plot and a helical mesh plot (generated with the web-based tool *Protein Origami* [32]) in figure 1.5. With the assumption, that bsrG is α-helical (like most of the other type I Toxin-antitoxin systems) and in an oligomeric state, the location of the three asparagines is important as a possible hydrogen-bonded pattern.

Lear *et al.* [44] assume that the hydrogen bonding network of asparagine side chains mediates the association of helices in helix bundles and leads to oligomerisation of transmembrane helices. Hence, asparagine side chains are crucial for the structural stability of oligomeric transmembrane helices. In bsrG, three asparagines follow seven amino acids apart from each other, so that they form a polar strip along the helical axis, ascending like a ladder. In combination with the proposed asparagine dimerisation concept of Lear *et al.* a possible hydrogen-bonding pattern with N^{12}, N^{19} and N^{26} similar to the proposed *charge zipper* mechanism is implied (described in chapter 1.2.3.2). Therefore, bsrG may be a dimeric transmembrane peptide.

Figure 1.5: Representation of bsrG as helical wheel. First CD experiments suggest that bsrG is an alpha-helical peptide with a few charged amino acids and mainly positive charges on the C-terminus. The three polar asparagines are located in an ascending ladder along the helical axis. Hydrophobic amino acids are coloured in yellow, negatively charged ones in red, positively charged ones in blue, polar amino acids in light blue and glycines in green. The amino acids are depicted in three letter code.

1.2.4.1 Functional Hints from Biological Experiments

In contrast to tisB, bsrG exhibits no destabilisation of the cell membrane with subsequent depletion of the cell energy (ATP-level) level. It is also not involved in persister cell formation.

Instead, bsrG strongly interferes with the cell envelope biosynthesis which is disturbed. This induces the invagination of the cytoplasmic cell membrane. Also, the delocalisation of the cell wall cytoskeleton and the associated cell wall synthesis could be observed. The consequence is the inhibition of protein synthesis. Altered cell morphology and aberrant cell division could be noticed as well [45].

1.2.4.2 Current Structural Knowledge

From the recent experimental studies in the Group of Anne Ulrich (KIT) and mainly from the doctoral thesis of Katharina Becker (2017) [46] and preliminary work of Annika Hespelt (Diploma thesis, 2015) [47], it is known that:

- Reconstitution of bsrG.

  Reconstitution in pure phosphocholine environment was unsuccessful. The best reconstitution was established in a POPE/POPG (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine /1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoglycerol) mixture of 2:1. Furthermore, this mixture imitates perfectly the natural membrane of a gram-positive bacterium like *B. subtilis* [48].

- Proven α-helical secondary structure by CD spectroscopy.

  In DPC detergents the α-helical part is 60 % which fits perfectly to the predicted 23 amino acid long transmembrane domain. For POPE/POPG 2:1 the α-helical part is 52 %. No other structural motif was found. Hence, the α-helical transmembrane part is the dominant secondary structure.

- Orientation inside the membrane.

  NMR and OCD measurements show a slightly tilted transmembrane orientation for bsrG.

- Parallel dimeric orientation in the membrane.

  SDS-PAGE provided that bsrG is in a dimeric state. No monomers or higher ordered oligomers were found. Moreover, the transfer and placement of the four polar amino acids and the three charged lysines at the C-terminus inside the hydrophobic membrane environment including the negatively charged head-groups of the lipids is unfavourable. Additionally, ToxR-assays verify this assumption. Due to the design of the constructs, these assays are only functional, and results can be obtained when bsrG is in a parallel orientation.

- Analysis of bsrG mutants.

  To get insight into the underlying structural interactions between the helices, 20 amino acid mono-/double- and triple-mutants were designed. All of them exhibited an α-helical (partly) tilted transmembrane orientation. Exchange of the negatively charged aspartic acid to alanine showed a reduced helicity, even for the double and triple asparagine mutants. The mutation of single polar amino acids and of glycines revealed oligomeric bsrG in SDS-PAGE. Single mutants of asparagines lead to the formation of higher ordered oligomers where the double and triple mutants show a monomeric state for bsrG in the SDS-PAGE.

- Cell leakage and toxic effects induced by bsrG.

  In contrast to the results of Jahn *et al.* [45] a weak leakage of POPE/POPG 2:1 vesicles was observed. In comparison with pore inducing antimicrobial peptides the cell leakage is very low. Interestingly, for the multiple asparagine mutants no leakage was found. A small toxic effect was found which indicates that bsrG can translocate through the gram-positive peptidoglycane layer.

1.2.4.3 Hydrogen Bond Zipper of BsrG

As shown by the Ulrich group, the type I Toxin-antitoxin system bsrG/SR4 from *B. subtilis* displays a completely new structural dimerisation motif: the *hydrogen bond zipper*. Three asparagines form a ladder-shaped hydrogen-bonding pattern which stabilises the two helices: the possible *hydrogen bond zipper* motif of bsrG. The three asparagines N^{12}, N^{19} and N^{26} form the feasible intermolecular hydrogen bonds, they are coloured in green in the following depiction. The amino acids, in the predicted transmembrane domain, from position 7 to 29 are coloured in grey. This assumption, based on the currently available information, needs further studies and has to be proved to be true.

MTVYESLMIMINFGGLILNTVLLIFNIMMIVTSSQKKK

MTVYESLMIMINFGGLILNTVLLIFNIMMIVTSSQKKK

2 Methods

2.1 Molecular Dynamics Simulations

Computer Simulations can bridge microscopic length and timescale to the macroscopic world [5]. The microscopic state can be characterised in terms of position and momenta of particles, here atoms and molecules. As described in chapter 1.1 one method, that can shed light on the needed time scales for biomolecular systems on atomistic level, are Molecular Dynamics simulations (MD simulations) [49].

The first biomolecular MD simulation was a 8.8 ps long simulation of a protein in vacuum in the 1970s [50].

Nowadays, the development has progressed so far, that realistic representations, larger timescales, realistic surroundings (lipid membranes, solvent, counter-ions) are standard conditions for simulations. Natural dynamics can be simulated as well as thermal averages of properties and accessible conformations [51]. This chapter outlines the principles, approximations and techniques on which MD simulations are based.

2.1.1 Basic Approximations

In principle on atomic level the dynamics of every system can be described only by the time-dependent Schrödinger equation:

$$H\psi = i \cdot \hbar \frac{\partial \psi}{\partial t}$$

where H is the Hamiltonian (sum of potential and kinetic energy), ψ is the wave function, and $\hbar$ is the Planck constant divided by 2π.

By applying the wave function the dynamics of the complete system (nuclei and electrons) is described. With the assumption of Born-Oppenheimer this function can be split in two terms. The Born-Oppenheimer approximation describes that the motion of electrons can be treated independently

from the motion of the nuclei, due to the smaller mass and the larger velocities of the electrons [52]. Therefore, the wave function can be separated in a nuclear and an electronic part. The nuclear part is time-dependent, and the electronic part depends parametrically on the position of the nuclei, not on dynamics. Hence, the wave function is a function of nuclear coordinates with time-independent electronic energies. The wave function can be separated by this *ansatz*:

$$\psi(R, r, t) = \psi_{nuclei}(R, t)\psi_{electrons}(r, R)$$

where R are the positions and momenta of the nuclei and r the positions and momenta of the electrons.

The electronic wave function is only parametrically dependent on the nuclei positions not on their dynamics. As a result, the function above splits into two parts: a time-dependent Schrödinger equation for the motion of the nuclei and a time-independent Schrödinger equation for the electronic dynamics. Accordingly, by solving the time-independent Schrödinger equation with parametrically contained nuclei positions, the obtained eigenvalues describe the *potential energy surface*, which is the potential the nuclei feel upon motion. All these approximations can only be applied for electronic ground state dynamics [53].

For large biomolecules and biomolecular systems in general it is impossible to solve the time-independent Schrödinger equation. That resulted in a second approximation where the potential energy of this system is treated classically as a function of nuclear coordinates, that means simple empirical expressions, which are only approximations, but cheap to compute. This energy function is therefore fitted to the quantum-mechanical ground state energy, which is termed classical force field. Within the force field the atoms are approximated as balls with point charges covalently bound via springs [54, 55].

2.1.2 The Classical Force Field

The classical force field calculates the potential energy as a function of nuclear coordinates. MD Simulations are sometimes named *Molecular Mechanics* simulations, due to the applied classical mechanics. Each atom is represented by a single particle with a point charge. These charged spheres

are connected via models of springs and simple interactions between non bonded atoms. Due to the electrons, described in an approximate/effective way, no bond forming/bond breaking or polarisation can be represented. Only ground state electrostatics are described [56].

Common biomolecular force fields, which are used in this work, are CHARMM (Chemistry at Harvard Macromolecular Mechanics) [57] and AMBER (Assisted Model Building with Energy Refinement) [58]. More specified force fields are GLYCAM (for sugar) [59], χOL_3 (a variant of AMBER for RNA) [60]and SLIPIDS (Stockholm Lipids) [61] specifically designed for Lipid bilayer description. Each force field has its own parametrisation strategy for the energy function, but all of them define atom types for atoms in distinct structural units. These atom types describe structurally similar atoms with the same atomic number and type of chemical bonding.

The parameters are derived from *ab initio* calculations, experimental data, quantum chemical calculations, from experimental thermodynamic data, spectroscopic data, bond lengths, angles and force constants.

In general, the energy function is composed of two terms describing the two main interactions: the bonded interactions between atoms and the non-bonded interactions. Both are based on the distance between the interacting atoms.

$$E_{total} = E_{bonded} + E_{non-bonded}$$

Bonded interactions are the stretching of the bond between the atoms, the angle bending (where three atoms are involved) and the dihedral angle which describes the rotation about the two central of four atoms. Non-bonded interactions are pairwise and a function of the distance. The Lennard-Jones interactions take part of two interactions and describe the short-range repulsion and the long-range attraction of dispersion. Particles with charge or partial charge are described by using Coulomb interactions [62, 63].

$$E_{total} = E_{bond} + E_{angle} + E_{torsion} + E_{Coulomb} + E_{Lennard-Jones}$$

In figure 2.1 these basic interactions for the force field are depicted schematically.

Bonded interactions
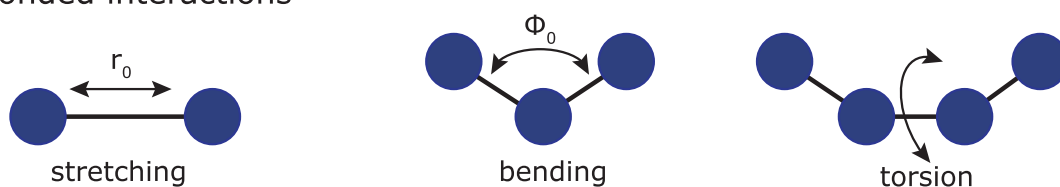
Non-bonded interactions

Figure 2.1: The main bonded and non-bonded interactions of atoms in the force field description. Bonded interactions consist of stretching of bonds, opening/closing of angles and out of plane rotation of bonds. Non-bonded interactions are Coulomb electrostatics and non-polar (van der Waals) interactions described as Lennard-Jones interactions.

2.1.2.1 Bonded Interactions

E_{bond} - **stretching**

The energy of a covalent bond between two atoms A-B by stretching around the equilibrium bond-length can be described by a simple Taylor expansion truncated after the second derivative. This leads to a harmonic potential (sometimes called Hooke's law) of the form:

$$E_{bond} = \sum_{bonds} \frac{1}{2} \cdot k \cdot (r_i - r_0)^2$$

where k is the force constant, r is the bond length, and r_0 is the reference equilibrium bond-length.

E_{angle} - **bending**

Between three atoms A-B-C, where A and C are bound to B, the angle can be calculated similar to the bond by a harmonic potential

$$E_{angle} = \sum_{angles} \frac{1}{2} \cdot k \cdot (\phi_i - \phi_0)^2$$

where k is the force-constant, ϕ is the angle, and ϕ_0 is the reference equilibrium angle.

$E_{torsion}$ - **dihedral angle torsion**

The torsional motion between four atoms A-B-C-D is a rotation about the B-C bond. The energy function exhibits a rotational barrier, and it must be periodic in angle ϕ which needs to have the same value after 360°. Therefore, the function is described as a Fourier series.

$$E_{torsion} = \sum_{torsions} V_n \cdot \cos(n \cdot \omega)$$

where V_n is the barrier for rotation, n is the rotation periodicity, and ω is the torsional angle.

2.1.2.2 Non-Bonded Interactions

The non-bonded interactions are split in 1-body, 2-body ... many body terms.

Lennard-Jones potential

The Lennard-Jones potential describes the repulsion and attraction of non-bonded particles. For large distances the potential is zero. For short

and intermediate distances there is a slight attraction (because of electron correlation). In accordance to the Lennard-Jones potential [64] the repulsive part is r^{-12} and the attractive part r^{-6}, dependent on the distance r:

$$E_{LJ}(r) = 4 \cdot \epsilon_{ij} \cdot \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right]$$

where σ is the diameter/minimum energy distance, ϵ is the well depth and r the distance.

Coulomb potential, charged interactions

For charged atoms and partially charged atoms the Coulomb potential is applied. Partial charges occur because of the internal redistribution of electrons which introduces positive and negative charges in parts of the molecule:

$$E_{Coulomb}(r) = \frac{q_1 \cdot q_2}{4 \cdot \pi \cdot \epsilon_0 \cdot r}$$

where ϵ_0 is the permittivity (electric constant), r is the distance and q is the charge.

2.1.3 The *Complete* Force Field Function

Combination of the inter- and intramolecular energy functions described before result in the complete energy function of the classical force field:

$$E_{total} = \sum_{bonds} \frac{1}{2} \cdot k \cdot (r_i - r_0)^2 + \sum_{angles} \frac{1}{2} \cdot k \cdot (\phi_i - \phi_0)^2 + \sum_{torsions} V_n \cdot \cos(n \cdot \omega)$$

$$+ \frac{q_1 \cdot q_2}{4 \cdot \pi \cdot \epsilon_0 \cdot r} + 4 \cdot \epsilon_{ij} \cdot \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right]$$

2.1.4 Propagation of the System

The dynamics of motion can be described classically by Newton's second law [65]:

$$F_i = m_i \cdot a_i$$

$$-\frac{\partial U}{\partial r_i} = m_i \cdot \frac{\partial^2 r_i}{\partial t^2}$$

where U is the potential energy (describing the full potential energy) which is the function of all atomic coordinates r_i.

The force F_i, interacting on an atom at position r_i, determines the acceleration a_i. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. This can be achieved by changing the velocity and position of atom *i* in the time step *dt*. Here, no analytic solution is possible, so, this has to be solved numerically. A finite difference algorithm is needed [66]. These different algorithms should conserve energy and momentum and should be computationally efficient and allow a long time step.

All of them are based on a Taylor series expansion:

$$r(t + \delta t) = r(t) + v(t) \cdot \delta t + \frac{1}{2} \cdot a(t) \cdot \delta t^2 + \dots$$

$$v(t + \delta t) = v(t) + a(t) \cdot \delta t + \frac{1}{2} \cdot b(t) \cdot \delta t^2 + \dots$$

$$a(t + \delta t) = a(t) + b(t) \cdot \delta t + \dots$$

where r = position, v = velocity (first derivative with respect to time) and a = acceleration (second derivative with respect to time).

To keep the integration step as large as possible, the fastest motion is selected as lower boundary. For biomolecular simulations this is the bond vibration with a period of 10 fs for X-H bonds. Accordingly, the time step was set to 1 fs, stable with the numerical integration.

To increase the time step from 1 fs to 2 fs, several constraints for bond lengths have been designed. The most used constraints in Molecular dynamics simulations and also in this work are: LINCS (Linear Constraint Solver, small molecules) [67], SHAKE (bond geometry restraint, large molecules) [68] and SETTLE (to constrain rigid water) [69].

2.1.4.1 Verlet

The Verlet algorithm uses position and acceleration at time t and time $t - \delta t$ to calculate the new position $t + \delta t$. This is a straightforward derivation from the Taylor-expansion. For the calculation no velocities were used, although, they can be calculated from $t - \delta t$ and $t + \delta t$. Still, the velocities on time $t + \delta t$ must be calculated before the velocities on time t [70]:

$$r(t + \delta t) = r(t) + v(t) \cdot \delta t + \frac{1}{2} \cdot a(t) \cdot \delta t^2$$

$$r(t - \delta t) = r(t) - v(t) \cdot \delta t + \frac{1}{2} \cdot a(t) \cdot \delta t^2$$

The sum of the two is:

$$r(t + \delta t) = 2 \cdot r(t) + r(t - \delta t) + a(t) \cdot \delta t^2$$

Variants of the Verlet method with higher numerical accuracy are the leap-frog scheme and the Velocity-Verlet algorithm.

2.1.4.2 Leap-frog

In this case, the velocities were first calculated at time $t + \frac{1}{2}\delta t$. These were used to calculate positions r at time $t + \delta t$. Afterwards, the velocities leap over positions, then positions leap over velocities, and so on. Within this approach, velocities are explicitly calculated, but not at the same time as the velocities. But they can be calculated approximately for a distinct time t [71].

$$r(t + \delta t) = r(t) + v(t) \cdot \delta t$$

$$v(t + \frac{1}{2}) = v(t - \frac{1}{2}\delta t) \cdot a(t) \cdot \delta t$$

2.1.4.3 Velocity-Verlet

Explicit atom position, speed and stability can be achieved [72].

$$r(t + \delta t) = r(t) + v(t) \cdot \delta t + \frac{1}{2} \cdot a(t) \cdot \delta t^2$$

$$v(t + \delta t) = v(t) + \frac{1}{2} \cdot [a(t) + a(t + \delta t)] \cdot \delta t$$

2.1.5 Periodic Boundary Conditions

For systems with around 10^6 atoms the modelled boundary surface has a large effect on the system properties. Periodic boundary conditions (PBC) are used, so that the system does not have an abrupt border.

Hence, the system is surrounded by infinite copies of itself, and each molecule will move the same way as the original (see figure 2.2) [73]. Errors caused by artificial boundaries can be removed, however, errors might occur when the box is too *small*, so particles can *see* each other. The choice of the box size is crucial and depends on the size of the studied system and the proposed behaviour e.g. unfolding [74].
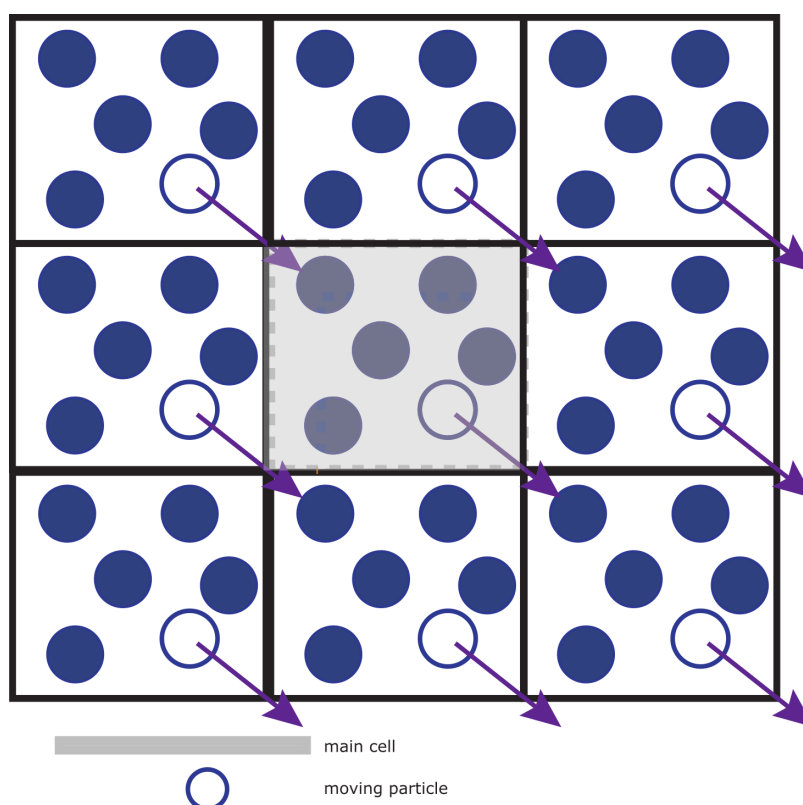
main cell

moving particle

Figure 2.2: The periodic boundary condition to minimise artefacts through abrupt borders. An infinitive repetition of itself surrounds the simulation box (coloured in light grey). When a particle moves out, an image particle moves in.

2.1.6 Intramolecular Interactions

The computationally most expensive part is the evaluation of non-bonded interactions which is a summation of N^2 for N atoms in the system. For biological systems with system sizes of 10^6 atoms a comparison of all pair-atoms should be avoided. Long-range interactions can be restricted with a cut-off distance. This is a good approximation for the Lennard-Jones potential with the fast decaying r^{-6} function, however, for the Coulomb potential with the slow decay of r^{-1} this leads to errors [75, 76].

2.1.6.1 Ewald Summation

The most popular methods to overcome the problems of cutting off the Coulomb potential are lattice summation methods such as the Ewald summation [77–79]. Coulomb interactions are treated with periodic boundary

conditions. The energy for the main cell is summarised over all pairs of charged particles (in the main cell and periodic images):

$$U = \frac{1}{2} \cdot \sum_{i,j=1}^{N} \sum_{n \in \mathbb{Z}^3} \frac{q_i \cdot q_j}{r_{ij,n}}$$

where N = number of charged particles, $n(n_x, n_y, n_z)$ = box index vectors, q = charges of the particles and $r_{ij,n}$ = the real distance between charges.

The sum is conditionally convergent and converges very slowly. The basis for the Ewald summation is the splitting of the slowly converging sum into two series which converge faster. Here, $\frac{1}{r}$ is split into error functions:

$$\frac{1}{r} = \frac{erfc(\alpha \cdot r)}{r} + \frac{erf(\alpha \cdot r)}{r}$$

where α is termed the *Ewald splitting parameter*.

The slowly converging series is transformed into a short range term (converging quickly in real space), a long range, smoothly varying term (converging quickly in reciprocal space) and a correction term. So the Ewald formula is summarised to:

$$U = U_{direct} + U_{reciprocal} + U_{correction}$$

$$U_{direct} = \frac{1}{2} \cdot \sum_{i,j=1}^{N} \sum_{n \in \mathbb{Z}^3} q_i \cdot q_j \frac{erfc(\alpha \cdot r_{ij,n})}{r_{ij,n}}$$

$$U_{reciprocal} = \frac{1}{2 \cdot V_{box}} \sum_{k \neq 0} \frac{4 \cdot \pi}{k^2} \cdot e^{-\frac{k^2}{4 \cdot \alpha^2}} \cdot |\rho(k)|^2$$

$$U_{correction} = -\frac{\alpha}{\sqrt{\pi}} \cdot \sum_{i=1}^{N} q_i{}^2$$

Where V_{box} = volume of unit cell, α = *Ewald splitting parameter*, determines the relative contributions to the direct and reciprocal sum, $k = (k_x, k_y, k_z)$, a vector in Fourier space.

The complex full mathematics can be looked up in source [80].

The Ewald-summation [81] has a computational expensive scaling of $N^{\frac{3}{2}}$ that can be improved by the Particle-Mesh Ewald method (PME) [82] which has an effective scaling of $N \cdot \log N$. Furthermore, the computation is fast and uses fast Fourier transformations for the reciprocal sum.

The Ewald-summation requires a neutral simulation box (according to the assumption made for the model that n = 0). Net charge leads to artefacts where a background charge occurs. This is mainly important for simulations with different dielectric constants. For example proteins with lipid membrane simulation will result in strong artefacts. Therefore, it is recommended to use counter-ions to neutralise the system or add an additional neutralising background charge [83].

2.1.7 Temperature and Pressure Control

The natural time evolution of a classical system with N particles in a distinct volume has a constant of motion which is the Energy E. Due to the fact, that time averages are equivalent to ensemble averages, the system can be described as a microcanonical ensemble NVE. To compare the results of simulations with experimental data, other ensembles are more convenient. Often, experiments are under constant pressure or temperature. Thus, the canonical ensemble NVT or isothermal-isobaric ensemble NPT is the goal.

Due to the rough calculation of non-bonded interactions and the numerical integration of the propagator the total energy is *not* conserved [80, 84].

2.1.7.1 Thermostats

Berendsen

To adjust the temperature, the system is coupled to an external heat bath of reference temperature T_0 with coupling constant τ_T. In the Berendsen approach the atomic velocities are rescaled, and the system is weakly coupled to the heat bath with the desired temperature T_0 [85].

$$v^{new} = v \cdot \sqrt{1 + \frac{\delta t}{\tau_T} \cdot \left(\frac{T_0}{T} - 1 \right)}$$

where δt is the integration time step and the coupling strength (rate of this correction) is described by τ_T. Large temperature differences are removed, but fluctuations of the kinetic energy are suppressed. No canonical ensemble can be achieved.

Parrinello-Bussi or Velocity rescaling

This thermostat is equal to the Berendsen barostat with an additional stochastic term that ensures a correct kinetic energy ensemble [86].

Nosé-Hoover

This thermostat yields the correct canonical ensemble [87, 88]. For the calculation an extended Lagrangian is used which contains additional artificial coordinates and velocities.

2.1.7.2 Barostats

In the same manner as for the temperature the system can be coupled to a external pressure bath. At each time step the size of the simulation box is adjusted, not only in size, but also in shape.

Berendsen

The pressure is controlled by adding an extra-term to the equation of motion like the thermostat [85]:

$$\frac{dP}{dt} = \frac{P_d - p}{t_p}$$

where P_d is the desired pressure and t_p is the time constant for exchange.

At each time step the volume of the box (coordinates and box vector) is rescaled by the factor:

$$\chi = 1 - \beta_\tau \cdot \frac{\delta t}{t_p} \cdot (P_d - P)$$

where β_τ is the isothermal compress ability and δt is the simulation time step.

Parrinello-Rahman

This barostat approach is similar to the Nosé-Hoover temperature coupling, and a canonical NPT ensemble can be generated [89].

2.2 Coarse-Grained Simulations

For MD simulations of large biomolecular systems the required time and length scales are not feasible with all-atom simulations at this point in time, see also chapter 1.1. So, a simple model, that is computationally fast, easy to apply and which can be used for a broad range of applications without the need to reparametrisation, became necessary.

Several coarse-grain (CG) methods were developed with different approaches for the simulation of solvent free models, models with explicit water or models for (chemical) specific compounds, as well as for the application to lipid systems [90].

2.2.1 The Martini (Coarse-Grain) Force Field

One of the most prominent and widely used CG force fields is the Martini force field which has a broad range of applications from self-assembly of lipid membranes, protein studies up to interactions of nano-particles with membranes and polymer simulations [91–93]. *Martini* is the nickname of the city of Groningen in the Netherlands where the force field was developed by the groups of Siewert J. Marrink and Peter D. Tielemann. The town's landmark is a tower called *Martini*, and the cocktail with the same name emphasises the universality of the force field.

A complex task can be described by the simplification of the molecule to building blocks with distinct properties and behaviour. Nowadays, a range of biomolecules from lipids, amino acids, sugars, DNA, collagenes to dendrimers, notably proteins and peptides can be described. The CG model is implemented in major simulation packages like GROMACS [94], NAMD [95], GROMOS [96] and Desmond [97].

2.2.1.1 Mapping

In the Martini model, the real molecules are treated as building blocks which represent groups of atoms. On average, four heavy atoms (and associated hydrogens) are represented by a CG bead. This bead has a single interaction centre. For ring systems and some amino acids this mapping is insufficient, so a mapping with a higher resolution of two to one is used. Figure 2.3 shows the mapping of a structure on atomistic level in contrast to the CG

one. As an example, water is represented by one bead which represents four real water molecules.

There are four building block main types which represent the main interactions:

- polar (P)

- non-polar / intermediate (N)

- apolar (C)

- charged (Q)

Each of the main types consists of several subtypes to describe the hydrogen bonding-capabilities (d = donor, a = acceptor, 0 = none) and the degree of polarity (from 1 = low polarity to 5 = high polarity). The outcome of this is a total of 18 particle types which are often called *building blocks*.

The mass of the beads is overall set to 72 u which corresponds to the mass of four water molecules. Only the beads in ring structures were set to a mass of 45 u. Figure 2.3 shows the different particle types described above jointly with the coarse-grained description of amino acids and the representation of water.

2.2.1.2 Parametrisation

The parametrisation was done systematically in close contact with atomistic models and will not focus on the fine reproduction of structural detail as in all-atom simulations. This opens up the possibility to use a broader range of compounds without the need to re-parametrise.

For the parametrisation of the non-bonded interactions a Lennard-Jones (LJ) 12-6 potential for dispersion and overlap is used.

$$U_{LJ}(r) = 4 \cdot \epsilon_{ij} \cdot \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right]$$

Here, σ_{ij} is the distance (r) between two particles and ϵ_{ij} is the strength of their attraction. For all particles the same effective size σ is used. The electrostatic behaviour is described by the Coulomb energy function.

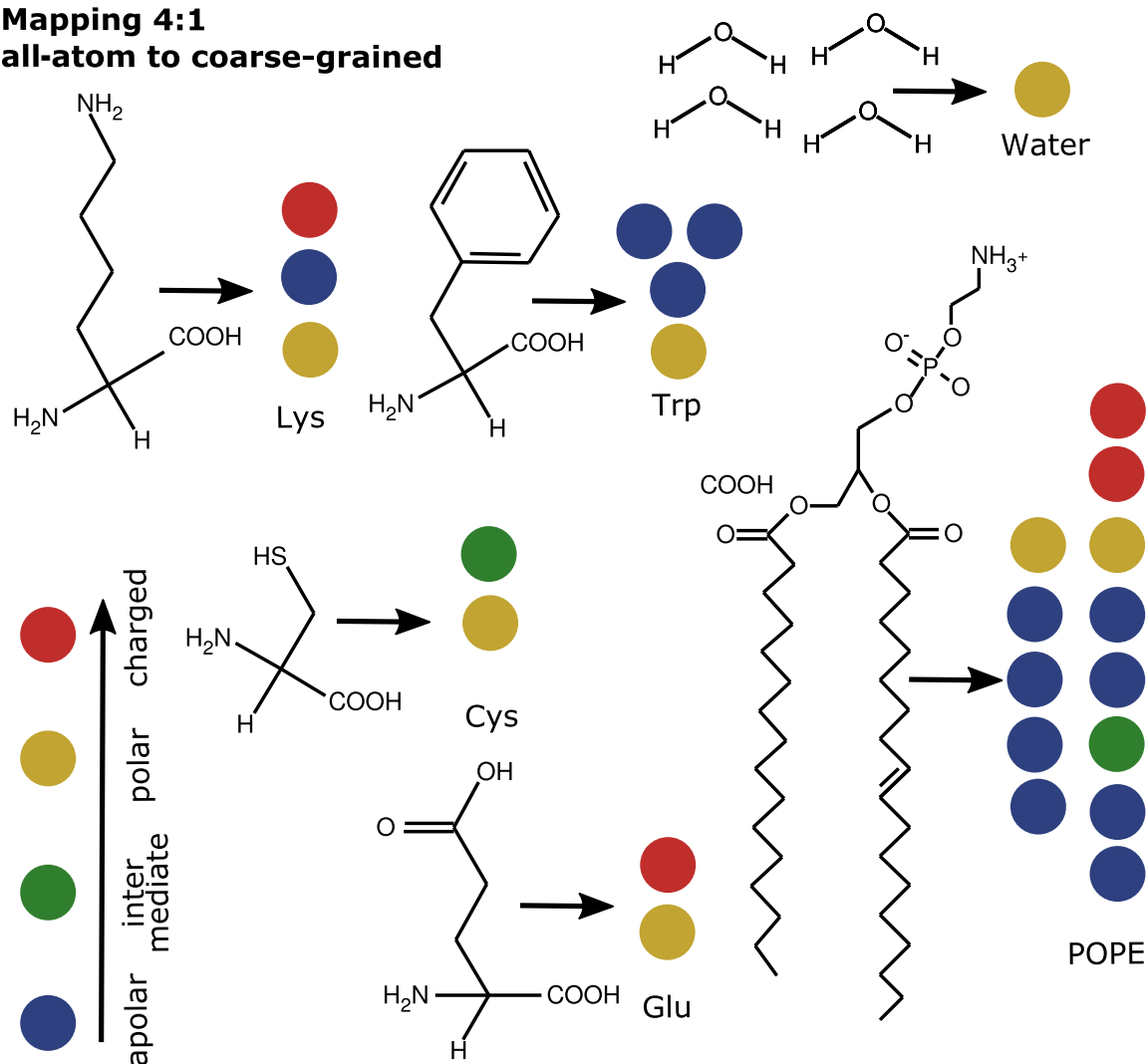$$U_{el}(r) = \frac{q_i q_j}{4\pi \epsilon_0 \epsilon_r r}$$

Figure 2.3: Representation of the different particle types in the Martini force field. As an example the representation of water, lipid and several amino acids is depicted. Picture is modified from [98].

Here, *r* represents the distance between the particles, q is the charge of the particles, ϵ_0 is the dielectric constant; the relative dielectric constant is ϵ_r, which is set to 15 for explicit screening.

The parameters are the reproduction of experimental partitioning energies, especially the free energy of hydration, the free energy of vaporisation and also the partitioning free energies for several sets of water to organic phases. This is the most important point because almost every biological question of interest for simulation is based on these interactions, for example peptide-membrane binding, protein-protein aggregation or lipid self-assembly. Bonded interactions were parametrised by mapping to all-atom simulations or structural experimental data (bond lengths) and described by a weak harmonic potential. This was also applied for angles and torsional interactions.

$$V_{bond}(r) = \frac{1}{2} \cdot K_{bond} \cdot (r - r_{bond})^2$$

Here, r_{bond} is a fixed equilibrium distance and K_{bond} is a force constant especially chosen for the Martini force field. Lennard-Jones interactions are excluded between bonded beads and for the description of angles, but not for the second-nearest neighbours [93].

2.2.1.3 Advantages and Disadvantages

Computational simulations are accelerated with the Martini force field. Due to the reduction of the number of particles, fewer interactions are computed, so the reachable timescale is larger in comparison to all-atom simulations. However, some atomistic degrees of freedom are neglected. The interactions between the particles are only the effective ones, the energy landscape is simplified. This results in fewer details but allows for time steps of 20-40 fs in comparison to 2 fs in all-atom (aa)-simulations. Furthermore, the energy landscape is not sampled exactly, but entirely [99].

As a result of the parametrisation, the properties of molecules in the solid and the gas phase are overestimated because they are only parametrised for the fluid phase [91]. Martini solvents are too structured because a softer form of the Lennard-Jones potential is needed, which leads to freezing of the water molecules amongst others. Also, the change in the environment (from water to the lipid bilayer for example) cannot be detected by charged

beads, implying that the electrostatic screening of water is only implicit. The solvation free energies are not reproduced, so water freezes too easily. A solution is the use of anti-freeze particles, softer non-bonded potentials or a polarisable water model [100]. Currently, long range electrostatic interactions are not parametrised, which means that interactions two to three CG beads away will not be considered.

A feature of the Martini force field is the fixed secondary structure of proteins. Structure transformations are not modelled in the current parametrisation, so protein folding processes cannot be simulated, and the secondary structure is adopted from experimental knowledge or forecast algorithms. However, it allows tertiary structure changes [98].

On the positive side, due to the fewer particles the effective timescale is much faster compared to all-atom simulations. Furthermore, the energy landscape is simplified because of the neglected degrees of freedom. However, this results in an increased sampling speed of the energy landscape three- to eight-fold larger compared to aa-simulations. As a standard rule of thumb a conversion factor of four is used to compare CG results with experimental data. The factor is based in comparison between the diffusion constant of CG water with the experimental value.

To conclude, the following points could be improved [92, 99]:

- new functional form for the non-bonded interactions with softer longer-ranged attraction

- structural flexibility and the ability to model structural changes and transitions (between secondary structures)

- multiscale approach analogous to QM/MM simulations where the *important* part is treated all-atomistic and the surroundings on the base of coarse-graining. In this way, the advantages of cg and aa would be combined: Faster simulation and the needed better simulation resolution for the *active* part.

2.2.2 Martini Tools

One of the main advantages of the Martini package is the easy and adjustable way of handling a huge number of tools. There are many *out of the box* tools to use: *martinize* (convert all-atom models to coarse-grained ones) [91], *insane* (acronym for *INSert membrANE*, a multicomponent membrane builder) [101], *martinate* (run CG-simulations and insert membrane proteins) but only the two most important tools for this work will be outlined: the <u>D</u>ocking <u>A</u>ssay <u>F</u>or <u>T</u>ransmembrane peptides (DAFT) (see 2.2.2.1) [102] and *backward* (see 2.2.2.2) [103].

2.2.2.1 <u>D</u>ocking <u>A</u>ssay <u>F</u>or <u>T</u>ransmembrane peptides (DAFT)

The main problem of membrane protein association simulations is proper sampling, and the same holds true for <u>C</u>oarse-<u>G</u>rained (CG) simulations.

A *High-throughput peptide self-assembly assay* is required. With such an assay the peptide behaviour under self-assembling conditions can be observed, and meaningful structures can be obtained at the end. Several hundred of unbiased association simulations were run to get a view of the energy landscape of the proteins inside the membrane with the main focus on the interactions between the membrane proteins [102]. An innovative feature is the docking with a coarse-grained method with a specially designed set of periodic boundary conditions. These special layouts were designed for a number of zero to nine interacting proteins. This leads to several of *unbiased* non-interacting structures which evolve collectively. Their evaluation enables the detection of the underlying energy landscape in an easy to use, flexible and fast way. This includes a workflow from *scratch* to the results of the simulations which is described and shown in figure 2.4.

At the end of the approach the probability of two (or more) components to encounter (which depends on the size and the diffusion rate) and the probability to bind during the encounter (which depends on the interaction and the relative orientation) can be obtained. One disadvantage during the simulation is the stability of the complexes, so the rate constant for the dissociation is too low, and the sampled time scale is not sufficient.

As a rule of thumb the simulations should reach 300 ns or longer, and the number of simulations per assay should be 400 to get a high resolution view on the studied topic.

2.2.2.2 Backward - Reverse Transformation

Backmapping or *reverse transformation* tools are needed for the conversion of coarse-grained structures to all-atom representations. With CG-simulations it is possible to scan the potential energy surface faster, but the resulting structure exhibits less detail. In combination with the DAFT approach (see section 2.2.2.1) this less detailed structure is a powerful combination. To study atomistic interactions in high resolution, a simulation of all-atom structures is useful afterwards. An approach, which bases on a geometric projection that is followed by a force field based relaxation, is used to handle this conversion [103].

The first stage is the generation of an atomistic structure followed by the second stage, the relaxation of the structure to reach a stable structure close to equilibrium. In this concept, atoms are first placed on an interpolated position along a line through the coarse-grain beads. Afterwards, the peripheral atoms are added at an offset from the line, close to the connecting atom. Then the structure is optimised with several short energy minimisation steps followed by constrained relaxation simulations with increasing simulation time steps from 0.2 fs to 2.0 fs.

DAFT - simulation approach

Input

peptide sequence

all-atom structure

martinize

cg-structure
topology

DAFT

combine
arrange
PBC conditions

insane

Output

trajectory
conformations
(coarse-grained)

MD-
simulation

*GROMACS
+ martinate*
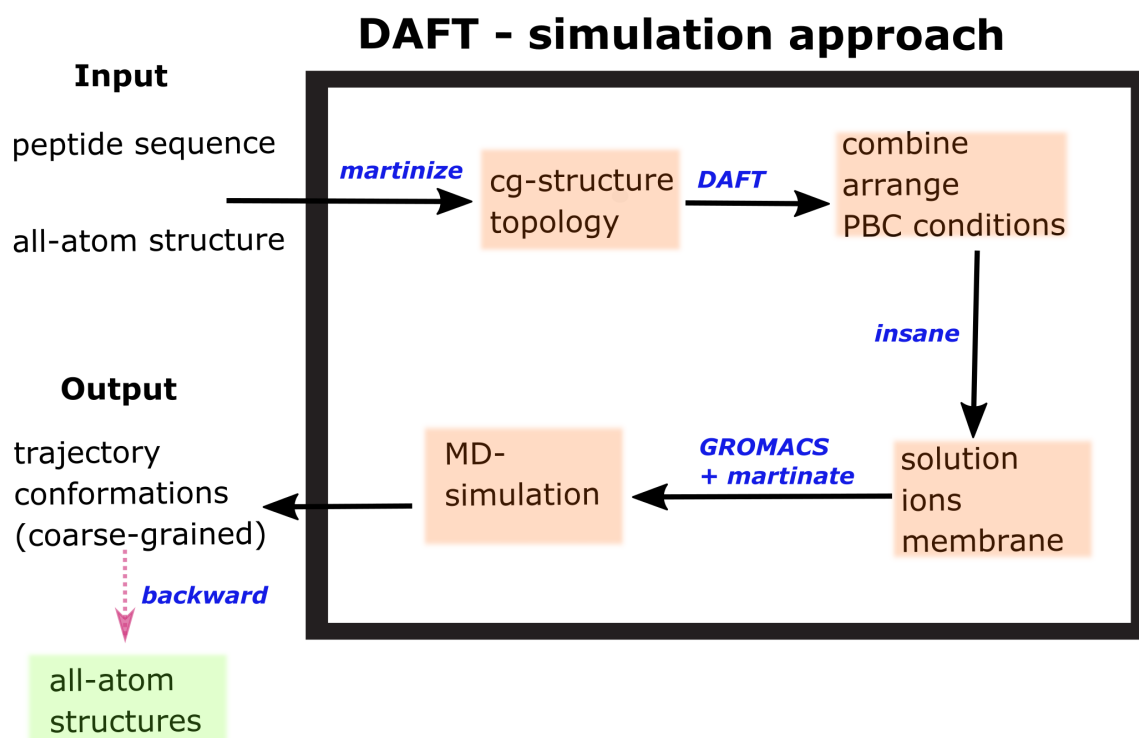
solution
ions
membrane

backward

all-atom
structures

Figure 2.4: Workflow of the DAFT approach. Peptide sequences, atomic structures or coarse-grained units with topologies serve as input at different stages of the approach. All structures are converted to CG-structures with the tool *martinize*. Afterwards, the DAFT layout is combined with the structures and the specified starting structures are generated. Subsequently, the systems are processed with *insane* to build the lipid bilayer and surround it with solvent and, if necessary, ions. Simulations were performed with the tool *martinate* in combination with *GROMACS*. The desired number of energy-minimisation steps, NVT and NPT equilibrations are performed with position-restraints and finally the production simulations. These simulations are performed for the whole number of sets with different starting structures separately. Obtained structures from the trajectories and simulations can be translated into all-atom via the *backward* script. The picture is taken in modified form from [104].

3 Structural Investigations of the Peptide TisB from *E. coli*

3.1 Introduction

This chapter focuses on the structural elucidation of the hydrophobic twenty-nine amino acid long peptide tisB from *E. coli*. Despite the experimental investigations in the group of Prof. Anne S. Ulrich (KIT) open questions concerning the appearance and quantity of tisB in the membrane associated state remain. Furthermore, it is unclear how the peptide can mediate the passage of protons across the hydrophobic bilayer and if a structure can be recovered by simulation. The manifestation of the structural shape, combined with the proposition of the *electrostatic charge zipper* of a dimeric antiparallel tisB, needs further investigations (see 1.2.3.3).

Preceding studies of tisB by coarse-grained simulations were not able to shed light on these structural questions due to their known disadvantages. To cite a few, the fixed secondary structure and the main idea of the 4:1 mapping cannot describe the exact position and angles of the side chain. In this work all-atom and coarse-grained simulations were combined to obtain a maximum of sampled configuration space and speeding up the simulations regarding the limited computational power.

First of all, long coarse-grained MD simulations with the binding association screening method DAFT were executed to choose reliable binding distributions as starting points for the following 1 µs all-atom simulation.

As starting assumptions for the coarse-grained docking approach, the helicity of tisB, incorporated in an hydrophobic bilayer (with a tilt of 20°-30°), was verified as well as the *active unit* being at least an antiparallel dimer. For this reason, two starting arrangements were chosen, tisB as an antiparallel dimer and as an antiparallel tetramer inside a hydrophobic bilayer with appropriate thickness, in this case POPC (16:0/18:1).

3.2 Methods

All simulations were performed and analysed using the molecular simulation package *GROMACS* in the versions 4.6.7, 5.0.4 and 2016.3 [94, 105–110].

3.2.1 Coarse-Grained Simulations

Structure of the tisB peptide

The 29 amino acid long tisB peptide was constructed using the *xleap* tool from the AmberTools package [111] based on the peptide sequence shown in section 4.1. For the N-terminus an amino-group (NH_3^+) and for the C-terminus a carboxyl-group (COO^-) were attached. The secondary structure of the peptide was assumed as a complete ideal α-helix with the backbone dihedral angles $\Phi = -57°$ (dihedral $C_{i-1}, N_i, C_{\alpha,i}, C_i$) and $\Psi = -47°$ (dihedral $N_i, C_{\alpha,i}, C_i, N_{i+1}$).

The DAFT approach

The DAFT approach (see also section 2.2.2.1) [102] was executed for tisB as an antiparallel dimer and an antiparallel tetramer. The number of DAFT simulations at once was set to 400, and the simulation time was set to 1 μs for the dimer and to 3 μs for the tetramer. A distance between the protein chains of 3.5 nm was selected. The proteins were embedded in a pure POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer solvated with water and neutralised with chloride ions. The simulation temperature was set to 300 K and the pressure to 1 bar in order to keep the POPC bilayer in the fluid phase. The temperature and the choice of the lipid bilayer were consistent with experimental measurements. For the whole system the coarse-grained force field *Martini 22* [84, 90–93, 100] was adopted.

The dimer and tetramer simulations were performed according to the standard DAFT protocol. First, the helical tisB model was aligned along the z-axis using the tool *editconf* from the *GROMACS* package. This all-atom protein model was coarse-grained using the tool *martinize* and afterwards duplicated and distributed as an antiparallel dimer rather than a tetramer over the corresponding DAFT layout. The total system size was ~3,000 atoms for the dimer simulations and ~5,400 for the tetramer simulations.

Then each system (from the different 400 starting configurations) was energy minimised with the steepest-descent algorithm and afterwards

shortly equilibrated via 10 ps of NVT simulation using a 2 fs time step, and a 100 ps NPT simulation using a 20 fs time step. Finally, the long production run (400 x 1 µs dimer / 3 µs tetramer) was performed. The time step was set to 20 fs. To regulate the pressure, the Berendsen barostat [85] was applied with a semi-isotropic pressure coupling, a compressibility of $\kappa_{xy} = \kappa_z = 3 \cdot 10^{-4} bar^{-1}$ and a time constant of 3.0 ps. The temperature was controlled via the V-Rescale thermostat [86] with a coupling constant of 1.0 ps. For temperature and pressure control three energy groups were used, one for the protein, one for the membrane, and one for the solvent including the ions. Electrostatic interactions were computed using the particle-mesh-Ewald (PME) algorithm [77] with a direct space cut-off at 1.2 nm and the group cut-off scheme. Dispersion interactions and short-range repulsion were described by a Lennard-Jones shift-potential (potential-shift-Verlet modifier) with a cut-off of 1.2 nm and a shift from 0.9 to 1.2 nm.

3.2.2 All-Atom Simulations

Conversion from coarse-grained to all-atom

The two most appropriate dimer and tetramer structures from the DAFT-approach were chosen (selected after clustering) and embedded into a lipid bilayer. Therefore, the obtained coarse-grained structures were converted to all-atom with the tool *backward* [103], see section 2.2.2.2. The backmapping is performed using a library of mapping definitions for the geometric reconstruction. For description of the protein the *AMBER14SB* force field [112] was used. The resulting structures were subsequently relaxed through energy minimisation and short MD-simulations with the standard protocol from the *backward* tool.

Protein embedding into the lipid bilayer

Afterwards the protein dimer/tetramer was aligned along the z-axis, which corresponds to the lipid bilayer normal. The dimer/tetramer was then inserted into a pre-equilibrated and fully hydrated POPC bilayer consisting of 288 lipids (144 per leaflet) with ~11,500 TIP3P water [113] molecules (~50 water molecules per lipid, see also section 4.2.1.1). For the description of the lipid Bilayer the *SLIPIDS* force field [61, 114–116] was adopted. The embedding was performed with the tool *g_membed* [117] with standard parameters. Here, the protein is first reduced in size in the membrane

plane, and overlapping lipids and water molecules were removed. Then, the protein was restored followed by a short MD-simulation. This approach has the advantage that the system is near the equilibrium afterwards. Finally, the positive charge of the system due to the net charge of the protein was neutralised with chloride ions described by the Cheatham and Young parameters [118]. The simulation size for the two dimer simulations was ~72,700 atoms in total (2 tisB proteins, 278 POPC lipids, ~11,500 TIP3P water and 2 chloride ions) and ~73,700 atoms in total (4 tisB proteins, 278 POPC lipids, ~11,500 TIP3P water and 4 chloride ions) for the two tetramer simulations.

Simulation protocol, AMBER14SB and SLIPIDS

The simulations were performed with a time step of 2 fs by integration using the leap-frog algorithm [119]. All bonds were restrained using the p-LINCS algorithm [67]. Electrostatic interactions were computed using the particle-mesh-Ewald (PME) algorithm [77] with a direct space cut-off at 1.4 nm and the Verlet [120] cut-off scheme. Dispersion interactions and short-range repulsion were described by a Lennard-Jones potential with a cut-off of 1.4 nm. For the thermostat and the barostat three coupling groups were considered: Protein, POPC-bilayer and the solvent including the chloride ions. During equilibration the positions of the heavy atoms of the peptide were constrained by a force of 1,000 kJ/nm in three-dimensional directions.

The system was first energy minimised using the *steepest-descent* integrator with a step size of 0.01 nm, until the ultimate tolerance of 1,000 kJ/nm or the maximum step number of 1,000 steps was reached.

Afterwards, the system was equilibrated at a constant volume in the NVT (N: number of atoms, V: volume and T: temperature) ensemble. Random initial velocities from the Maxwell-Boltzmann distribution at 310 K were assigned, followed by 1 ns of constant volume simulation at 308 K using the V-Rescale thermostat [86] with a coupling time of 0.1 ps. Subsequently, 1 ns of equilibration at a constant pressure in the NPT (N: number of atoms, P: pressure and T: temperature) ensemble at 308 K and 1 bar was executed. The temperature control was conducted with the Nosé-Hoover thermostat [87] with a coupling time of 2.0 ps. For the regulation of the pressure the Parinello-Rahman barostat [89] was applied with a semi-isotropic pressure coupling with the compressibility $\kappa_{xy} = \kappa_z = 4.5 \cdot 10^{-5} bar^{-1}$ and a time

constant of 5.0 ps. Lastly, for the production simulation, 1,010 ns of NPT simulation ($5 \cdot 10^8$ steps) were performed at 308 K using the Nosé-Hoover thermostat with a coupling time of 0.5 ps and 1 bar using the Parinello-Rahman barostat with a coupling time of 10 ps.

3.2.3 Analysis

The visualisation and the graphical analysis was illustrated with the tools *Gnuplot* [121], *Grace* [122] and *VMD 1.9.2* [123]. The first 10 ns of the all-atom simulations were omitted for equilibration of the system. To ensure periodic boundary conditions (PBC) for visualisation and analysis the trajectory was processed in several steps with the *GROMACS* tool *trjconv*.

Cluster analysis

With the *GROMACS* tool *cluster* a cluster analysis on the DAFT simulation results was performed. In this way, similar structures and the application of the two most populated structures for all-atom simulations should be found. For the analysis, a matrix of RMSDs of the molecule coordinates of the last snapshot at 1,000 ns for the dimer and 3,000 ns for the tetramer from the 400 different simulations was calculated. A cut-off radius of 0.4 nm in combination with the single linkage algorithm [124] was used to obtain clusters of structural similarity.

Particle distribution

With the *GROMACS* tool *confrms* the RMSD of two structures was calculated after least-square fitting the second structure to the first. That way, the particle distributions of the 400 different final snapshots of the DAFT-simulations were compared. For each assay (dimer, tetramer) the last frames were collected, and the protein structure was least square fit to the first protein chain.

Distance maps

To obtain a more detailed view of the interactions evolving over simulation time of the DAFT approach contact maps were calculated. With the tool *dist_Mat* [125] the average (over the 400 simulations) minimum distance matrix of the dimer protein chains was calculated for every 100 ns of simulation time. The used cut-off distance for the contact map was 0.5 nm.

Helicity calculation

To calculate the helicity H_i for each residue i averaged over time, the dihedral angles Φ (dihedral $C_{i-1}, N_i, C_{\alpha,i}, C_i$) and Ψ (dihedral $N_i, C_{\alpha,i}, C_i, N_{i+1}$) were calculated for every snapshot of the trajectory. A residue is counted as helical when Φ and Ψ are within 30° around their ideal values $\Psi = -47°$ and $\Phi = -57°$. So, a residue is 100 % helical if the condition is achieved for all evaluated timesteps.

$$H_i = \frac{1}{N_{\text{time steps}}} \sum^{N_{\text{time steps}}} h_i$$

$$h_i = \begin{cases} 1, & \begin{aligned} &\text{if } (\Psi_0 - 30°) < \Psi_i < (\Psi_0 + 30°) \\ &\text{ and } (\Phi_0 - 30°) < \Phi_i < (\Phi_0 + 30°) \end{aligned} \\ 0, & \text{otherwise} \end{cases}$$

Calculation of other protein structural parameters

To obtain information about the flexibility of the constructs the *GROMACS* tool *rmsf* was used. This tool calculates the root mean square fluctuation RMSFs (for all heavy atoms). In addition, the root mean square deviation was computed with the tool *rms.* The RMSD describes the structural stability in comparison with a reference structure over the simulation time. Distances between specific atoms of the side chains were calculated with the tool *distance.*

For the calculation of the tilt angle between the helical axis and the membrane normal (z-axis) the tool *bundle* was used. Beforehand, the helicity for each chain was calculated, and only the transmembrane part with >90 % helicity over simulation time was taken into account for the tilt angle calculation. Besides, the tilt angle was calculated for the overall protein complex and for each chain separately.

3.3 Results

3.3.1 DAFT - Coarse-Grained Docking Approach

To consider the binding patterns of the antiparallel tisB dimer and tetramer, a new coarse-grained docking approach (DAFT) was used. Herein, tisB was placed as dimer or tetramer with 400 different starting structures (obtained from rotation along the helical axis) with a 3.5 nm gap placed in a POPC membrane environment, solvated with water and neutralised with chloride ions. For each structure, the expanded simulation time was 1 µs for the dimer and 3 µs for the tetramer.

3.3.1.1 Dimer

Particle distribution

First, the formation of dimers was analysed by examination of the particle distributions (shown in figure 3.1). This provides a quick classification of the preferred binding distributions and the specificity of binding. Overall, it can be seen that the majority of the particles exhibit the same binding motion. On the left side (the view from top, where the backbone of peptide 2 is coloured in brown), a binding is shown that is mainly on one side where the density of particles is much higher than on the other side. On the right panel, the side view clearly shows a strong binding of the alternating charged amino acids forming the dimer interface. A preferred orientation is also evident. The peptides orient themselves like a *X* with a subtle distinction in the tilt-angle to the membrane normal, and that leads to the broader distribution of the particles. This rotation of the peptide chains towards each other was a consequence of the polar side chain snorkelling through the water environment and the penetration of the water molecules (not shown here). During the simulation all peptides remained in a transmembrane state. No peptide became oriented towards the lipid bilayer or inverted its orientation towards a parallel dimerisation state.

Distance maps

A more detailed view of the interactions and their involvement is shown in figure 3.2. The distance maps of tisB at the points of time 0 ns (start of the simulation), 100 ns, 500 ns and 1 µs (end of simulation) are plotted. Each of these maps is obtained by averaging the 400 individual maps over
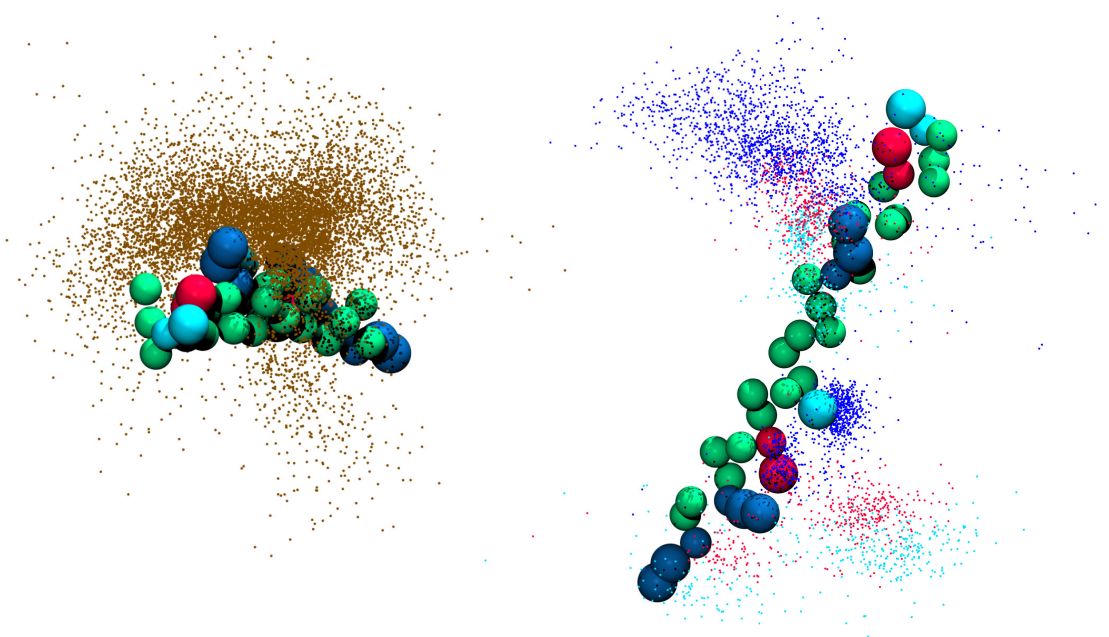
Figure 3.1: Particle distribution of one component around the other from the final ensembles. For each assay, the last frames were collected and the positions were least square fit to the first peptide of the tisB dimer. This chain is represented by balls in the coarse-grained description. Only the polar (cyan) and charged side chains (positive ones coloured in dark blue and negative ones in red) are represented. The backbone of chain one is coloured in green. The view from the top (left) shows the distribution of the backbone atoms from chain 2 in brown and reveals a preferred binding on one side. The side view indicates an interaction of the charged and polar side chains mainly differing in the tilt angle towards each other resulting in divergent distances at the N- and C-terminal ends.

time from the corresponding simulation time frames. The cut-off radius for interactions was set to 4 Å to consider bindings such as salt bridges and hydrogen bonds (in the range of 3 Å). One can clearly see that the distance between the side chains decreases over time, and binding occurs between the peptides. By comparing the distance maps over time the interaction patterns between 500 ns and 1,000 ns are almost not changing. However, the provided simulation time is long enough in order to obtain an interaction pattern and a sufficient sampling for a qualitative assessment of the binding distribution. The interaction pattern is well-defined, and a

characteristic profile can be recognised. Both peptides are in close contact, and the interaction as an antiparallel dimer is reliable. It can be concluded that the dimer interface is build up by the polar and charged amino acid interactions.
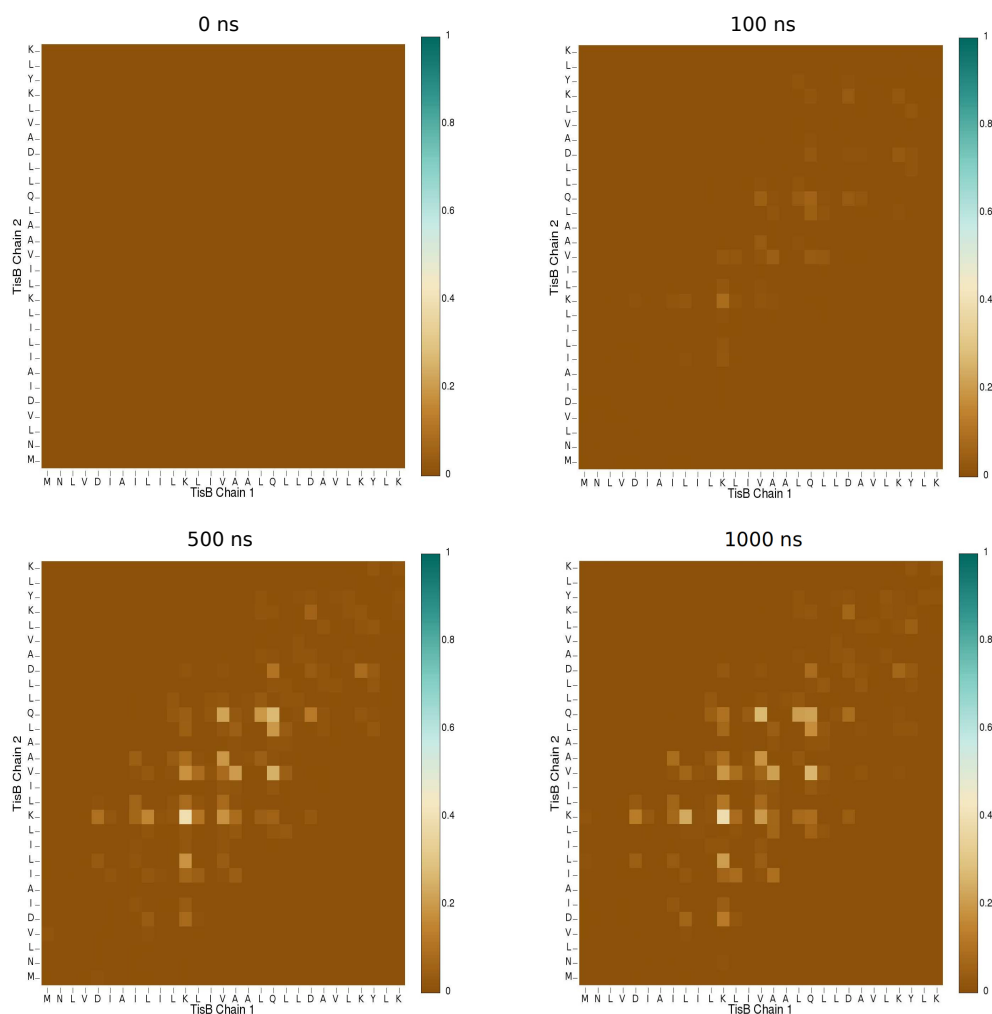

Figure 3.2: Distance maps over time. The distance maps were calculated for each residue of the peptides at each time by averaging over the individual contact maps for the 400 simulations for the time points 0, 100, 500 and 1,000 ns. At the starting time 0 no interaction could be observed which is in line with the start setup of the DAFT approach. Only a few interactions, mainly between the polar and charged amino acids, occur after 100 ns. The evolution of the distinct patterns reaches its maximum at 500 ns. In the following 500 ns the interaction pattern remains stable which indicates that the expanded simulation time of 1 µs was long enough to reach a qualitative assessment of the preferred binding state.

Cluster analysis and *backmapping*

Finally, a cluster analysis was performed on the last snapshots at 1 μs simulation time. Aligning the 400 different ending structures and identifying the preferred binding options were desired. Overall, the cluster analysis confirms the previous results that all structures end up in a bound interface indicating that the peptide is able to behave as an antiparallel structure inside the hydrophobic bilayer.

Two dimer conformations dominate which are shown in figure 3.3. For both structures the interface is built up via interactions between the polar glutamines and the charged aspartic acids and lysines. Cluster 1 exhibits a mirror-symmetric orientation of the dimer interface interactions. The strands are slightly displaced against each other. A triade interaction between a lysine, an aspartic acid and a glutamine occurs. Both ends are moved marginally apart from each other interacting with the polar water environment to shield the charged side chains. Besides, the dimer is slightly tilted to the membrane normal.

Cluster 2, however, is in a straight-upright orientation. On one side, the two proteins are in close contact with the polar amino acids tending to each other and stabilising the end. The overall structure is *Y*-like with the peptide chains being further away from each other. It is notable that the glutamines in both clusters are directed to each other. Thus, in the central part of the interface the hydrophobic interactions stabilise the binding.

Since the helical structure of the peptide chain is fixed, a feasible unwinding of the helix is not possible. Furthermore, the precise orientation of the side chain cannot be carried out.

Also, as a result of the coarse-grained water mapping of four water molecules to one big the localisation of water inside the dimer interface is not beneficial. Only if the protein dimer chains are apart from each other, in this case at the unbound end structure, water can be incorporated. Yet, the benefit of the coarse-grained docking approach definitely outweighs the drawbacks, i.e. shorter simulation time and reasonable preformed starting structures for the extended all-atom simulation.

At last, the two promising coarse-grained structures were converted to all-atom via the *backward* protocol (see 2.2.2.2).
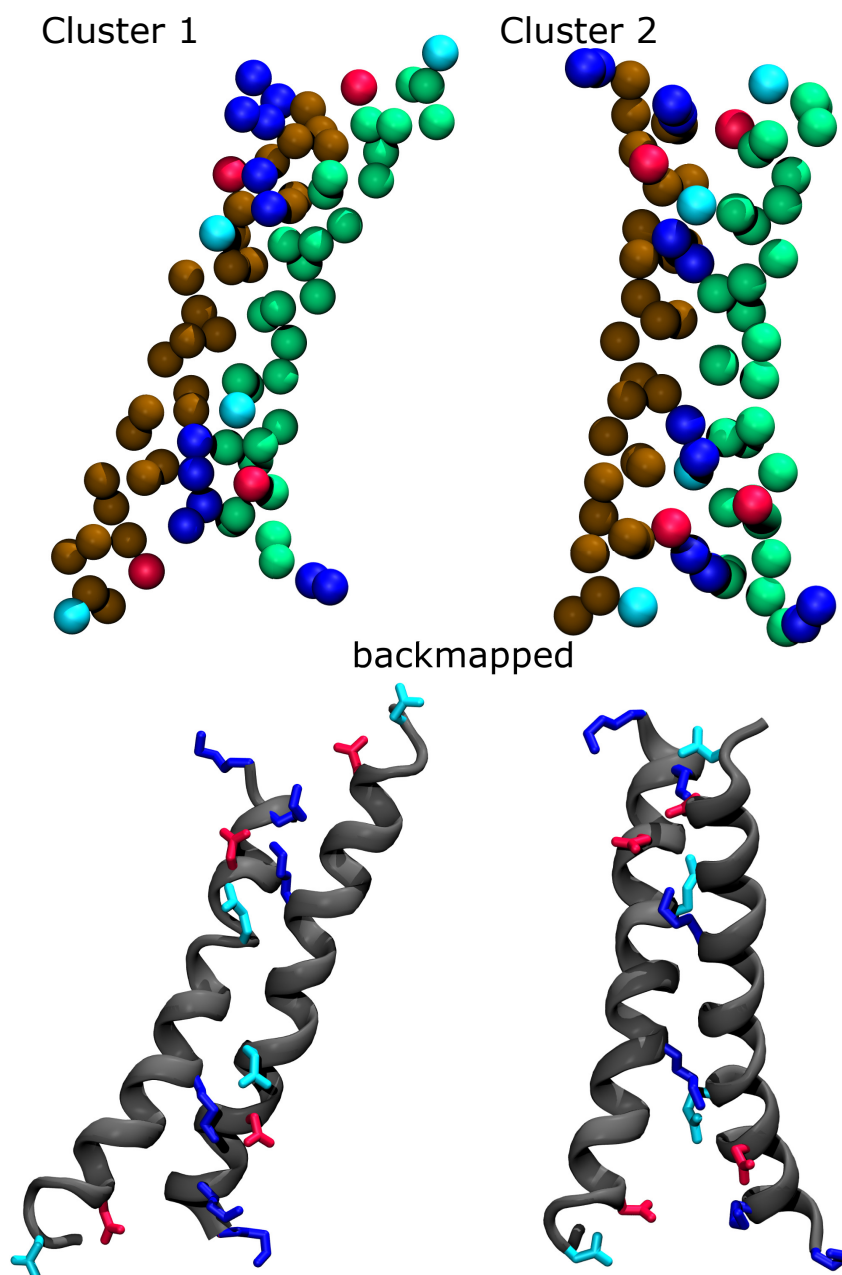
Figure 3.3: At the top, the two most populated structures from the cluster analysis of the last snap-shots simulation time of the DAFT coarse-grained approach at 1 µs are illustrated. At the bottom, the converted structures to the all-atom description via the *backward* protocol are depicted. The protein backbone is coloured in grey (all-atom) and brown/green (coarse-grained). Polar amino acids are coloured in light blue, positively charged ones in dark blue and negatively charged ones in red.

3.3.1.2 Tetramer

Particle distribution

The formation of tetramer binding assemblies was analysed according to the previously described analysis of the dimer DAFT simulations. The investigation of the preferred binding distributions was performed via particle distributions. Therefore, the final ensembles of the extended 3 µs simulations were collected and least square fit to the first peptide of the tisB tetramer (see fig. 3.4).

From the distribution of the peptide main chains it can be concluded that a bulk of structures were able to find a binding position. At the same time, more simulations compared to the dimer DAFT simulations were not interacting despite the extended simulation time of 2 µs. This might be due to the doubled size of the simulation box with four different starting structures, each 3.5 nm apart from the other. If the structure at the beginning is too far away for a possible interaction, the rotation towards each other and the rotation around the helix axis will need longer simulation scales. For the qualitative analysis the result is sufficient.

The side view of the interaction between the polar and charged side chains provides a good view of the driving binding forces. As in the dimer approach, the positive and negative side chains tend to each other and a preferred *X*-like structure could be observed. The tetramer is built up by two proteins interacting in a dimer state, and those sub-dimers combine again as a superordinate dimer with the overall tetramer shape. During the simulation all peptides remained in a transmembrane state. No peptide oriented towards the lipid bilayer or inverted its orientation towards a parallel dimerisation state.
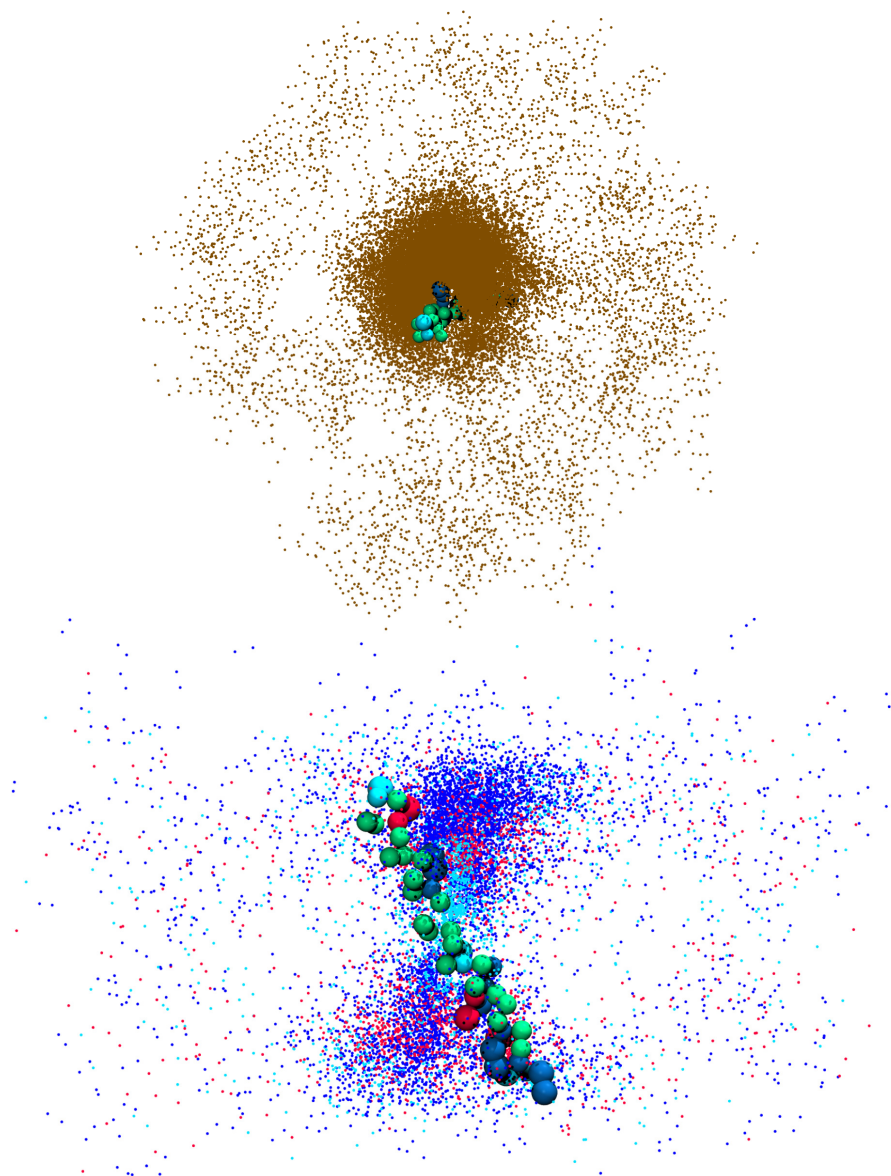
Figure 3.4: Particle distribution of one component around the other from the final ensembles. For each assay, the last frames were collected and least square fit to the first peptide of the tisB tetramer. This chain is represented by balls in the coarse-grained description. Only the polar (light-blue) and charged side chains (positive in dark blue and negative coloured in red) are represented. The backbone of chain 1 is coloured in green. The upper figure (view from the top) shows the distribution of the backbones of the chains 2, 3 and 4 coloured in brown. The majority of the peptides assembles during the 3 μs approach and reveals a preferred binding position. In addition, the side view, where only the polar and charged side chains of the peptides 2,3 and 4 are shown, demonstrates that the predominate interactions for the binding are the same as for the dimer. The tetramer is built up via two dimer units.

Cluster analysis and *backmapping*

After the analysis of the particle distribution a cluster analysis was performed. Therefore, the last snapshots of the 400 different simulations at the time 3 µs were chosen. The cluster analysis shows preferred binding orientations, but in a greater diversity and with smaller quantity. Also, a lot of structures were found, where only two or three of four peptides were interacting, or all are far apart from each other. In contrast to the tetramers no formation of dimers or trimers occurred. This gives a slight hint that the formation of the tetramer is preferred under the conditions of the coarse-grained simulation.

The two most populated clusters were selected (see figure 3.5 and 3.6). Both clusters are stabilised via the interactions between the positively and the negatively charged amino acids. The overall shape is similar to the dimer simulation with the distinction that the X-like shape is formed by two dimers interacting. Two peptides in antiparallel orientation interact via the possible salt bridges extending the glutamines in the central part of the tetramer interface. This leads to the stacking of the four glutamates building a polar strip which leads to a continuous polar interface. In both clusters the antiparallel peptides were slightly shifted against each other for the dimer formation. This leads to the interaction of the C-terminal lysine with the N-terminal aspartic acid.

On both sides of the tetramer pattern the polar amino acids are pointed out of the hydrophobic membrane core and are surrounded with water. The main difference between cluster 1 and 2 is the interaction of the N- and C-terminal ends. In cluster 1 one dimer is marginally distorted, which leads to the drifting of the ends apart from each other and a greater surface for solvent accessibility, while the second cluster has C-terminal lysines pointing towards each other.

Finally, the two selected coarse-grained tetramer structures were converted via the *backward* protocol to all-atom structures for subsequent simulations.
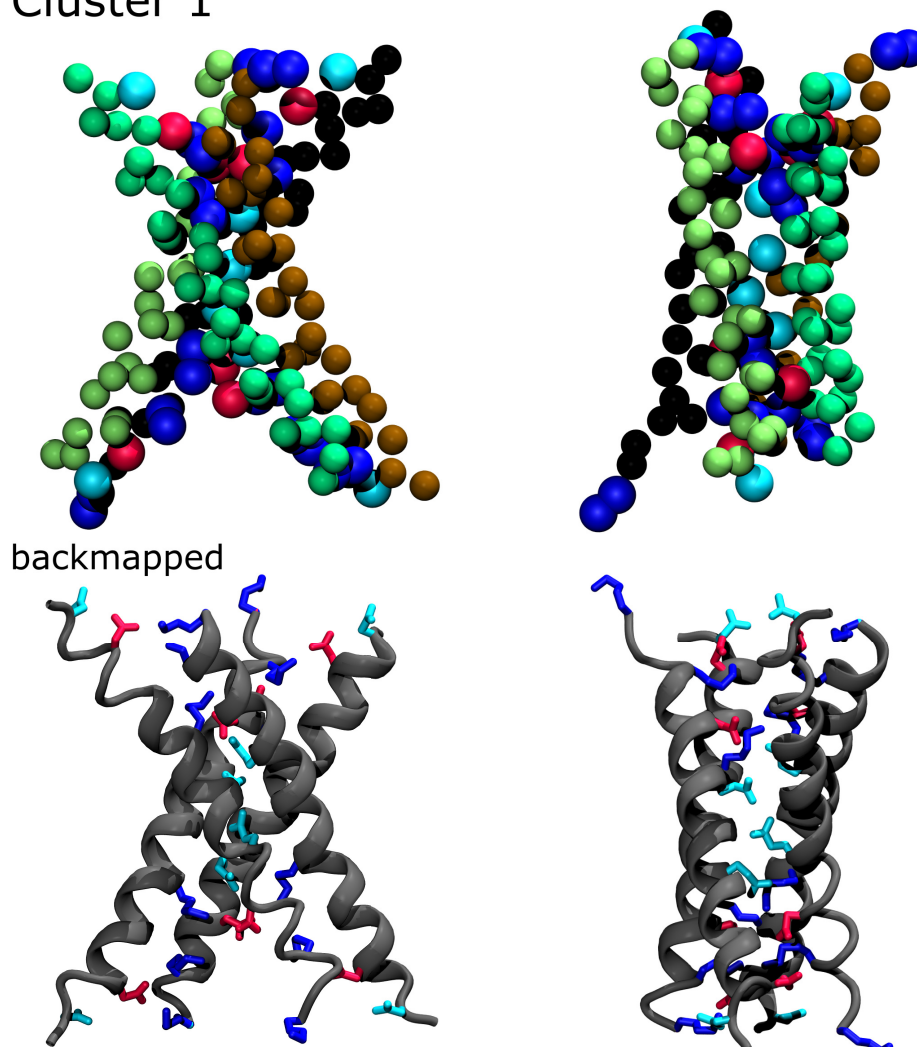
Cluster 1

Figure 3.5: At the top, cluster 1 of the two most reasonable and populated structures from the cluster analysis of the last snap-shots at 3 µs simulation time of the DAFT coarse-grained approach is illustrated. On the left, the tetramer is depicted from the broader side. On the right, the tetramer is shown from the smaller side with the interface strip. At the bottom, there is the same representation in the all-atom description converted via the *backward* protocol. The tetramer is built up of two peptides interacting as dimer via salt bridges, and this dimer interacts with the second dimer to form the tetramer. In the middle part, the four glutamines form a polar strip which is a result of the stacked polar residues. The protein backbone is coloured in grey (all-atom) and brown/lighter green/green and black (coarse-grained) for a direct visual comparison. Polar amino acids are coloured in light blue, positively charged in dark blue and negatively charged in red.
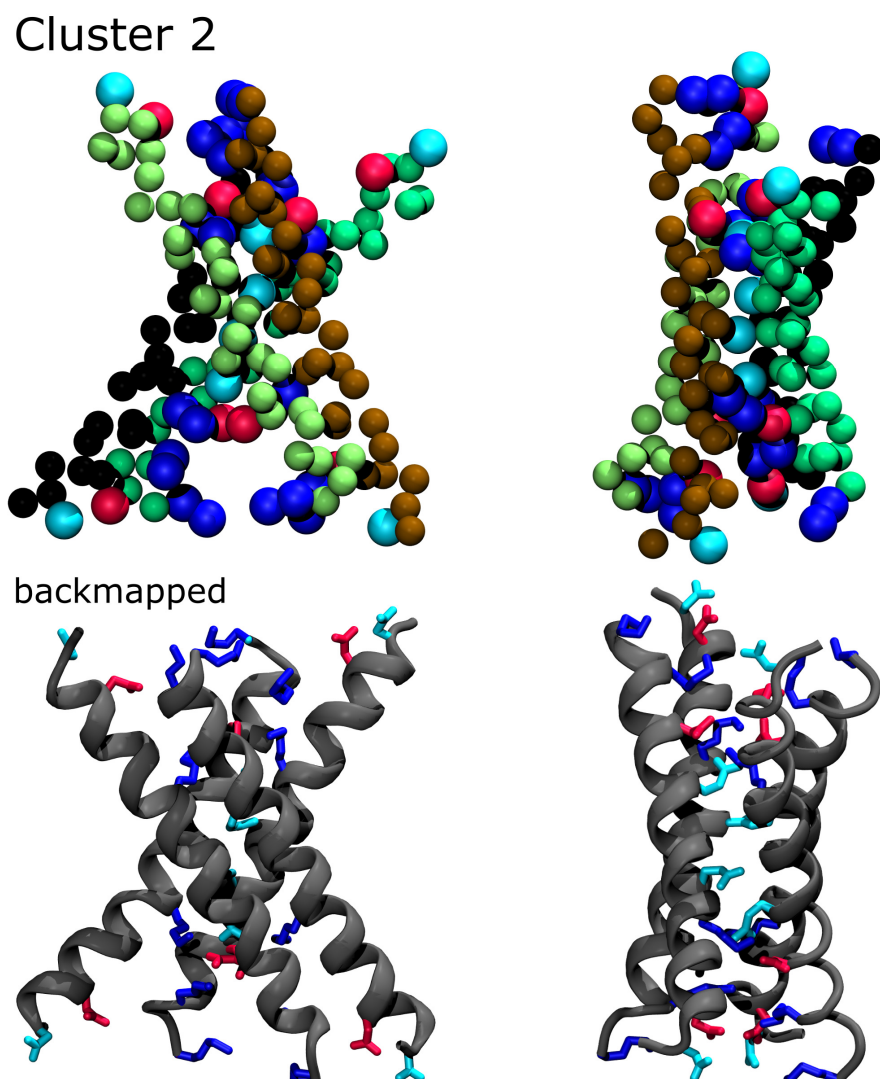
Cluster 2

backmapped

Figure 3.6: At the top, cluster 2 of the two most reasonable and populated structures from the cluster analysis of the last snap-shots at 3 µs simulation time of the DAFT coarse-grained approach is illustrated. On the left, the tetramer is depicted from the broader side. On the right, the tetramer is shown from the smaller side with the interface strip. At the bottom, there is the same representation in the all-atom description converted via the *backward* protocol. On the first view, the overall structures seem to be similar to the structure of cluster 1. The main difference is the interaction at the protein ends. Here the lysines at the peptide C-terminal end are closer together, while the C-terminal end is further apart. The protein backbone is coloured in grey (all-atom) and brown/lighter green/green and black (coarse-grained) for a direct visual comparison. Polar amino acids are coloured in light blue, positively charged in dark blue and negatively charged in red.

3.3.2 Dimer Simulations (All-Atom)

Each of the two selected antiparallel dimer structures were converted to all-atom from the previous DAFT-approach and embedded in a POPC bilayer. After several equilibration steps, an unrestrained production run of 1 μs length was progressed.

In both simulations of the chosen DAFT cluster the assembled dimer structure remains in a transmembrane state, close to each other and in a tilted state against the membrane normal. The last snapshots (shown in figure 3.7 at the top) demonstrate the overall shape of the binding pattern. Here, the ends of the dimer are fully solvated and membrane overhanging. Mainly, all charged amino acids are covered with water, permeating inside the membrane to shield the polar amino acids from the hydrophobic lipid environment and stabilising the dimer construct. The inner core of the dimer interface is hydrophobic and not filled with water. This is because the polar amino acids in this region are apart and provide the boundary for the water.

The averaged structures over simulation time are shown in figure 3.7 at the bottom. Comparing the constructs the basic construction of both is similar and stable over simulation time. Only the averaged shape of the aspartic acids perceives a higher flexibility during the simulation. In cluster 1 the two peptides are in closer contact, and the spreading on one side of the cluster is less. In contrast, cluster 2 is only interacting via the glutamines and lysines. In the following sections more structural parameters were calculated to get a deeper insight on the interactions and the stabilising forces of the dimer and the main differences of both structures.

3.3.2.1 Helicity

The evaluation of the helicity over 1 μs simulation time is shown in figure 3.8. Because of the helical similarity of the peptides inside the dimer cluster, only one chain is plotted and compared with the second cluster. For the N- and C-terminal ends the peptide unravels starting from positions 7 and 24. The central part of the protein between positions 7 and 24 remains helical.
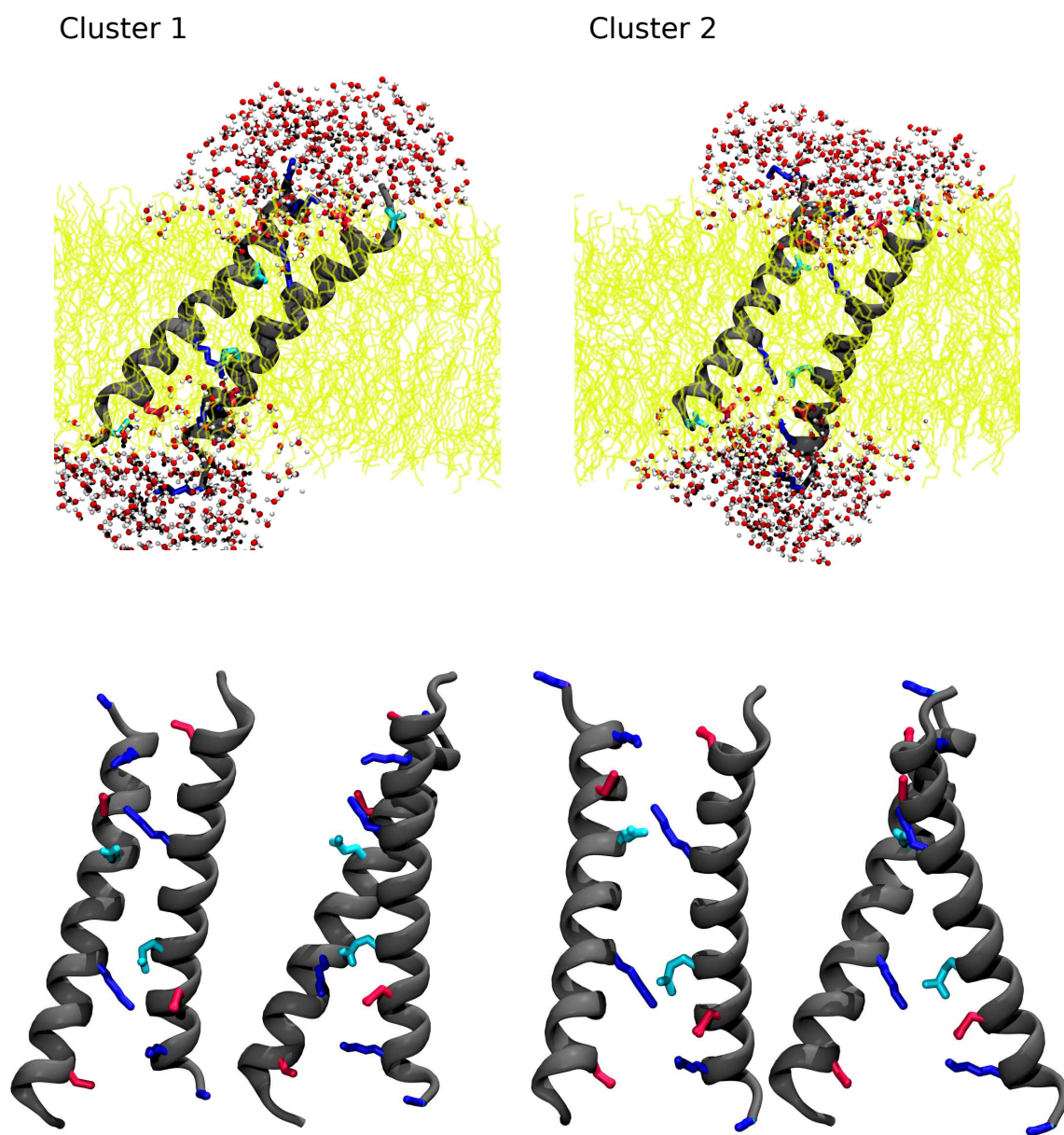
Cluster 1   Cluster 2

Figure 3.7: The last snapshots of the 1 μs simulation of the clusters are shown at the top. The POPC membrane is coloured in yellow. Water with a maximum distance of 8 Å to the protein is illustrated. The backbone of the peptide is coloured in grey, only the polar (light blue) and charged amino acids (blue: positively charged, red: negatively charged) are plotted. At the bottom, the averaged structure of the assembled dimer is depicted from two points of view.
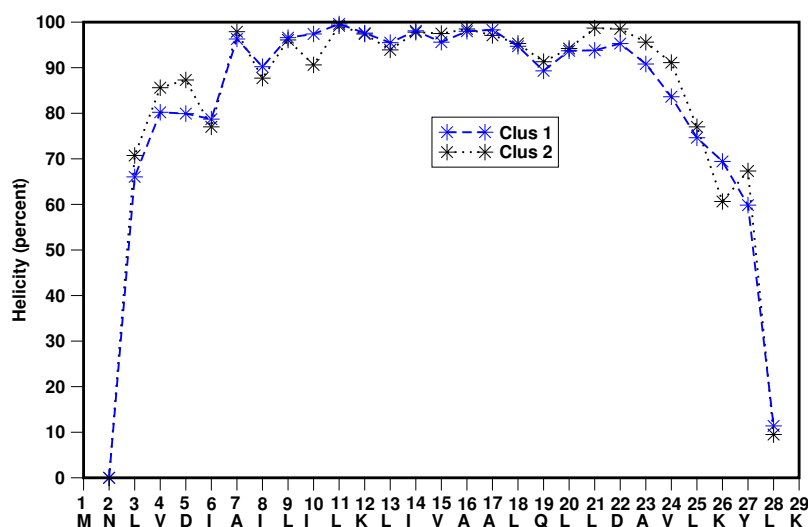
Figure 3.8: Helicity along the peptide sequence as percentage of time averaged over the 1 μs simulation. Both dimer structures exhibit mainly the same helicity. At the N- and C-terminal ends they unravel. The main part of the protein (from positions 7-24) remains stable and fully helical.

3.3.2.2 Tilt-angle

For the calculation of the tilt-angle only the fully helical part of the peptide from positions 7 to 24 in the centre of the tisB peptide was considered. The tilt angle was determined for each peptide chain, and the overall tilt of the dimer complex against the membrane normal was estimated (see figure 3.9).

The evaluation of the tilt angle of cluster 1 reveals that the tilt increases over the first 200 ns and then fluctuates with 10° around the value of 40°. In this central part inside the membrane both peptides exhibit the same orientation which results in an overall angle with the same value.

In contrast, the tilt angle of cluster 2 varies a lot, while both peptides have a difference of around 30°. The overall tilt-angle increases over the simulation time. During the last 100 ns both peptides orient around the value of cluster 1 with the same tilt angle. Thus, the view on the last snapshot of the simulation depicts the same result.

3.3.2.3 Sidechain Distances

To get a deeper insight into the main interaction pattern, the distances between the polar glutamines and the alternating possible salt bridges
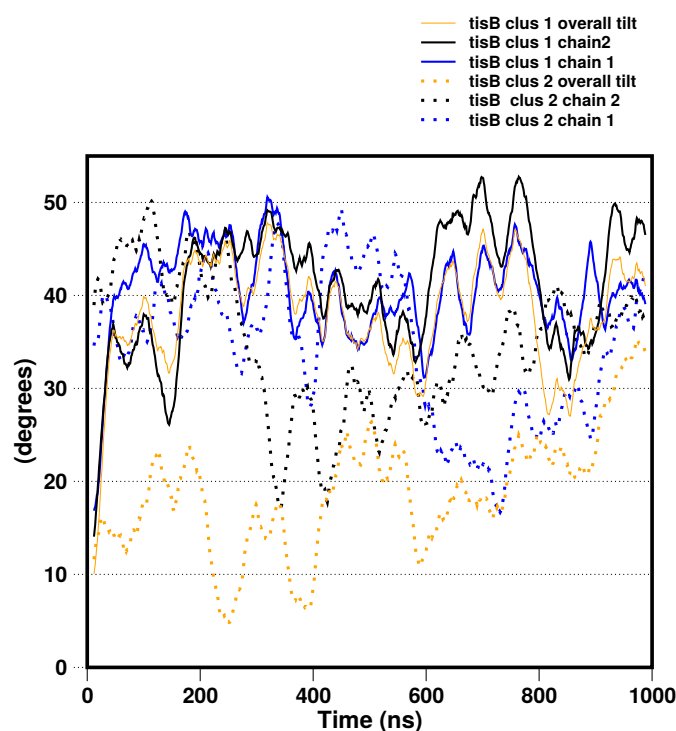
Figure 3.9: Evaluation of the tilt angle over simulation time for clusters 1 (solid line) and 2 (dotted line). The tilt angle was calculated for the helical part from positions 7-24 of the tisB peptide. The overall tilt angle describes the tilt of the dimer association state against the membrane normal.

between the positively charged lysines and the negatively charged aspartic acids were investigated. For the salt bridges the distance between the side chain nitrogen and the carboxyl oxide is measured, whereas, for the possible two hydrogen bonds the distance between the carbon of the amide group was used. The distance limit for a salt bridge is ~4 Å and for a hydrogen bond ~3 Å. Because of the measured distance between the amide carbons the distance limit for an interaction between the glutamines is around 6 Å (additional distance for the carbon oxid/nitrogen bond-length).

The side chain distance of cluster 1 (see 3.10) over simulation time indicates that mainly for the $Asp^{22} - Lys^{12}$ and the $Lys^{12} - Asp^{22}$ interaction a salt bridge is possible. Only in section 150- ~200 ns the distance between the glutamines is in a range for a possible interaction. Around 200 ns all possible interaction patterns can occur, side chains are pointing towards each other.
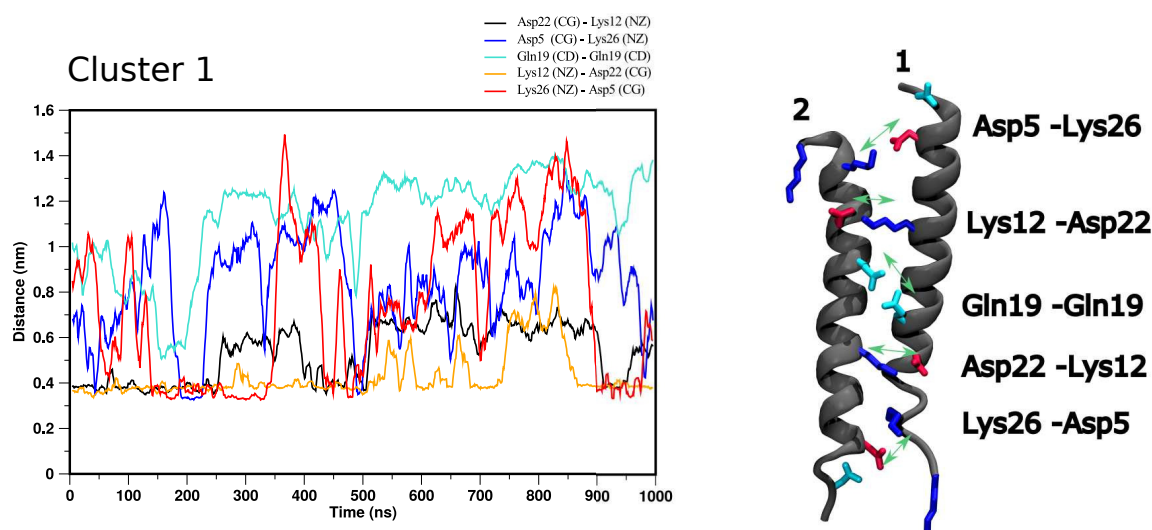
Figure 3.10: Calculated distances between the charged and polar side chains of the antiparallel tisB dimer cluster 1. Salt bridges of the aspartic acid and lysines have a boundary distance of 4 Å, the interaction between the glutamines amounts to 6 Å. Not any interaction between the side chains remains stable over the simulation time. Most of the time the $Lys^{12} - Asp^{22}$, and at distinct points in time the $Asp^{22} - Lys^{12}$ as well as the $Lys^{26} - Asp^5$ side chain, can make salt bridges. During a small time frame from 150 - ~200 ns the glutamines are in direct contact.

In contrast, for cluster 2 (see 3.11) rather different results are obtained. During very short points in time the side chains of the tisB dimer are in a possible distance for salt bridges. The glutamines are far away from each other over the whole simulation time.

3.3.2.4 Water permeability

The end structures and the averaged structures of the tisB dimer showed that the polar side chains of the peptides are shielded with water and that the water can permeate inside the hydrophobic bilayer core via the peptide dimer. Due to the ability of tisB to reduce the ATP-level at the bilayer and the mediation of the passage of protons across the bilayer the simulations were screened for a water wire. There are two possibilities for the proton passage. Either a proton can travel along a water wire, or a hydronium ion could travel along a continuous polar interface.

The main border for a possible proton transfer is the hydrophobic part between the two glutamines. The water can pass until the glutamine, but
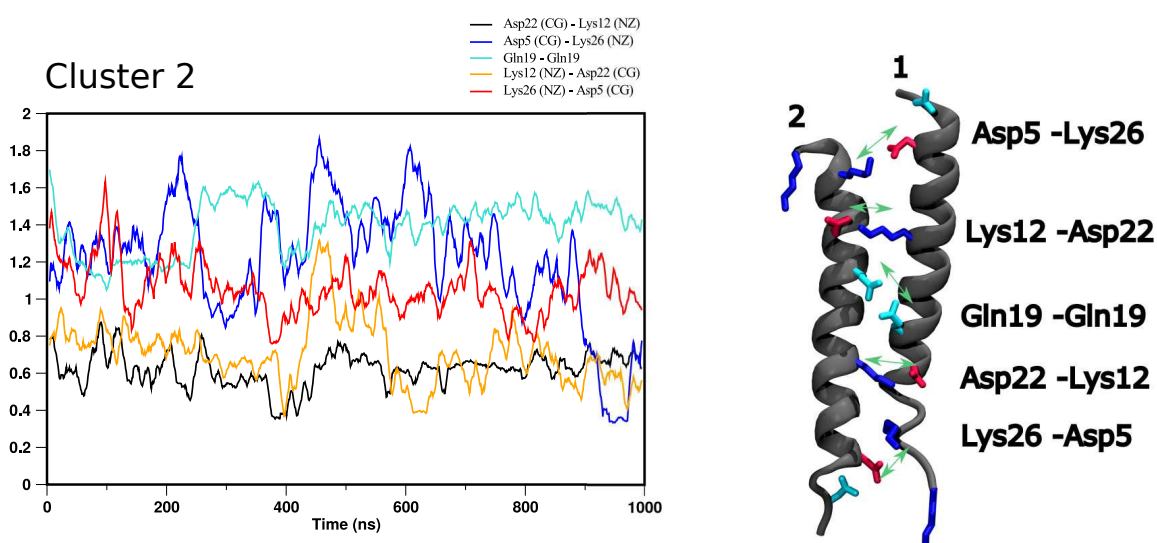
Figure 3.11: Side chains distances over time. Over the whole simulation time the side chains are only rarely at distinct points of time in a possible distance for interactions. The glutamines always point to other directions.

the part between is stacking together, thus it is unfavourable to pass a positive charge as a hydronium ion or as a polar water molecule.

Overall there is no possible structure present for a proton passage with 1 µs of simulation of cluster 2. Considering cluster 1, two structures were detected (shown in figure 3.12 A and B). Structure A is stable for 2 ns at 200 ns simulation time, here the dimer is nearly upright in the lipid bilayer, and only one of the glutamines limits a water wire. Structure B occurs for only a few pico seconds around 800 ns simulation time. One water molecule managed to bridge the two glutamines pointing to each other. So, a proton would be able to cross the membrane along the dimer interface.

To detect similarities between the two structures, where a proton transfer may be possible, the two were aligned (see figure 3.12 C). In both cases the glutamines are pointing to each other, and the charged amino acids are close together.

A comparison between the structure A (darker grey) and the averaged structure over simulation time of cluster 1 (lighter grey) is shown in panel D. The averaged structure is higher tilted, and the glutamines are further apart despite the same rotation of the side chain.
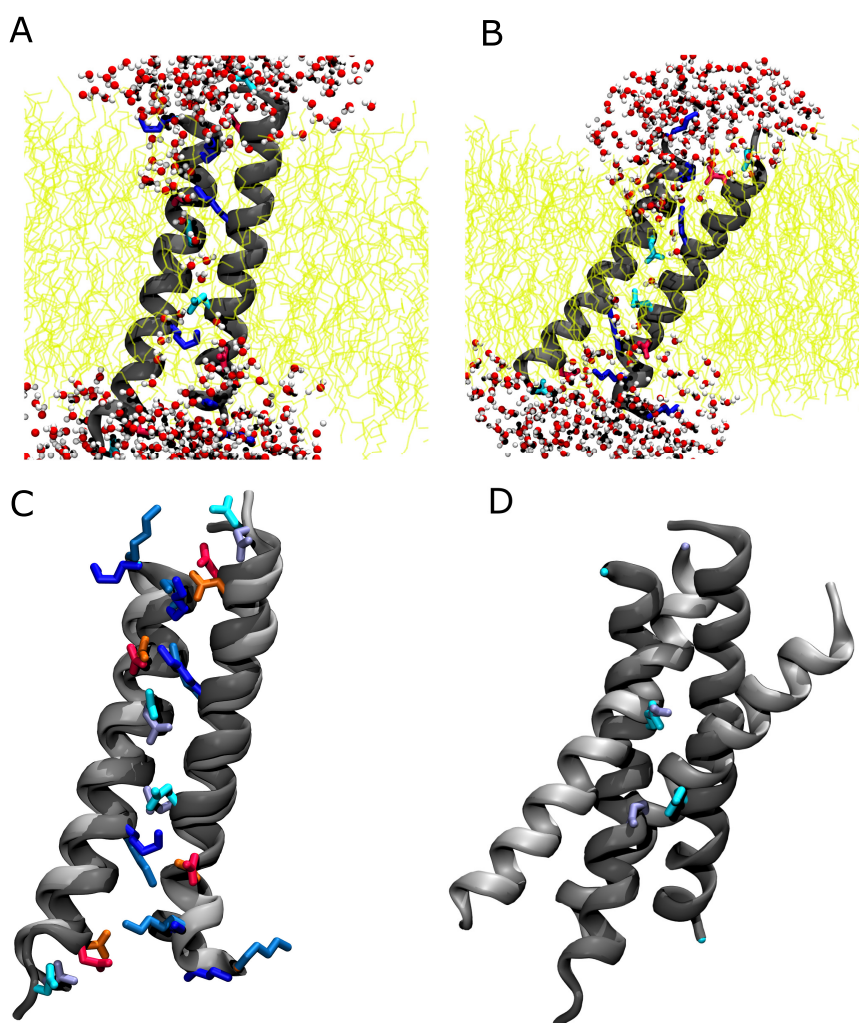
Figure 3.12: A: Arrangement of a possible structure for proton transfer detected at ~200 ns of simulation time (cluster 1). The water was able to get into the dimer interface only limited by one glutamine side chain. The structure was only stable for 2 ns. B: Structure of a possible proton transfer (cluster 1). For a short period in a picosecond time-scale one water molecule was able to bridge the hydrophobic part between the two glutamines, and a proton could be transferred. C: Overlaid and aligned structures of the two possible proton transfer structures (A: dark grey, B: light grey). The illustration shows possible interactions. In both cases the glutamines are pointing to each other, and the salt bridges are stabilised because the helices are in close contact. D: Comparison between the possible proton transfer structure A (dark grey) with the averaged structure over the simulation time of cluster 1 (light grey). The main difference is the tilt angle and consequently the distance of the glutamines. Positively charged amino acids are coloured in dark blue, negatively charged in red, polar ones in light blue and the main chain is coloured in grey. The lipid bilayer is represented as lines and coloured in yellow.

3.3.3 Tetramer Simulations (All-Atom)

The antiparallel tetramer simulations were executed and analysed in the same way as the dimer. First, the two most promising and populated structures from the DAFT coarse-grained approach were converted to all-atom simulations and embedded in a POPC-bilayer. They were solvated with water and neutralised with chloride ions. After an equilibration protocol, a free MD-simulation of 1 μs length was performed.

Both simulations demonstrate an assembled tetramer structure which is stable over the whole simulation time. The tetramer structures remain tilted against the membrane normal. The built up tetramer structures behave like a narrow channel structure, and they are filled with water. This will be reviewed in one of the next sub chapters (see 3.3.3.4).

The averaged structure over simulation time is shown in figure 3.13. Cluster 1 is depicted in chart A and cluster 2 in chart B. They have in common that the backbone structures are in a distinct position over time, whereas the shape of the polar side chains indicates that they fluctuate around their averaged value. The interaction between two antiparallel dimers seems to be crucial. Two of the peptides build a dimeric structure with the aspartic acids and lysines pointing to each other. Also, the glutamine side chains are in an orientation pointing to each other. These sub dimer structures are interacting over the hydrophobic central part of the peptides supported by the glutamines.

To clarify the subtle differences between the two tetramer simulations, the averaged helical backbone structures were superimposed (shown in figure 3.13 C, cluster 1 dark grey, cluster 2 blue). There is a clear difference between the orientation of the sub dimer structures against each other. Cluster 2 is highly symmetrical; the peptides of the sub dimers are slightly shifted against each other, but the sub dimers are mirrored to each other. In cluster 1, not only the peptides of the sub dimer are shifted against each other, but also the sub dimers themselves. Besides, it is notable that a kink is visible at the end of the strands. These overall structural differences have also an impact on the tetramer interface. The four glutamines of cluster 2 are all close together, bridging the hydrophobic inner core of the peptides, so that the tetramer interface is polar overall. In cluster 1 the shifting of

the peptides leads to intermolecular contact of only two glutamines. So, a small part of the tetramer interface is not covered by polar side chains.
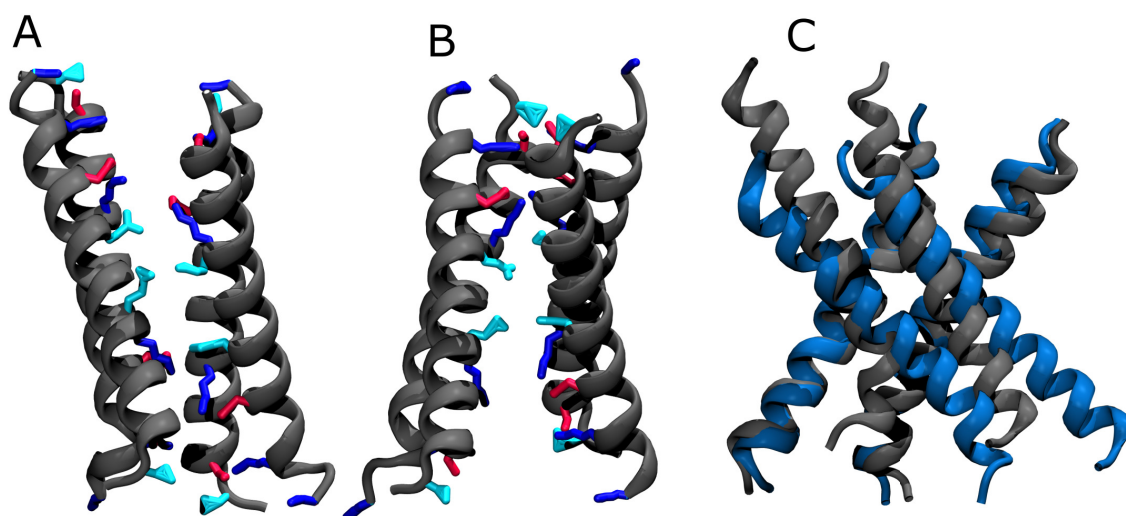

Figure 3.13: Averaged structures over the 1 μs simulation time of the antiparallel tisB tetramers. For both simulations the assumed tetramer structure is stable and built up of two sub dimers (charged residues are pointing towards each other) which are interacting via the hydrophobic inner core, respectively, the glutamine side chains (cluster 2). A describes cluster 1, B describes cluster 2. The polar amino acids are coloured in light blue, aspartic acid in red and the lysines in dark blue. The backbone is coloured in dark grey. In part C the structures of cluster 1 (dark grey) and 2 (blue) are superimposed. A clear difference between the associated tetramers can be detected. Cluster 2 is highly symmetrical with both sub dimers interacting, whereas, in cluster 1 the sub dimers are shifted against each other and a kink at the end of one of the peptides is detectable.

3.3.3.1 Helicity

The evaluation of the peptide helicity over 1 μs simulation time is shown for clusters 1 and 2 in figures 3.14a and 3.14b. It should be noted that chain 1-2 and chain 3-4 are the associated sub dimers. For both clusters each peptide is, with small differences in the exact percentage over time, shaped similar. At the N- and C-terminal ends the peptides' secondary structure unfolds. The region from positions 6 to 22 of cluster 1 is nearly perfect helical, while for cluster 2 a three amino acids longer helical part in the centre occurs from positions 5 to 24. In conjunction with the averaged structure of cluster 1 and the viewable kink the abrupt decrease of the

helicity at position 24 indicates that the kink is located at this position. This coincides also with the shifted sub dimers. Because of the shift, the amino acids are not incorporated into the lipid bilayer. They unfold in the hydrophilic environment and alter the alpha-helical secondary structure in comparison with the symmetrical cluster 2.
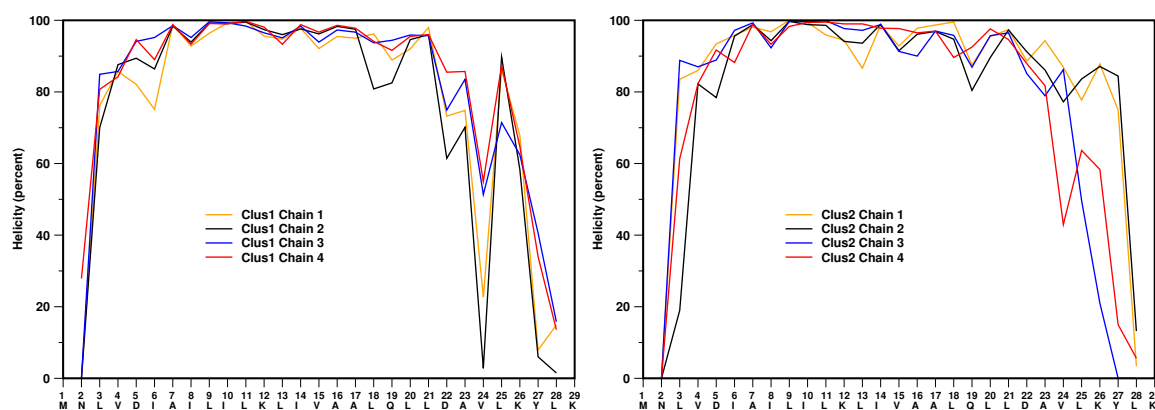
(a) Cluster 1: The region from positions 6 to 22 is almost perfectly helical. The abrupt change at position 24 indicates a kink.

(b) Cluster 2: The region from positions 5 to 24 is almost perfectly helical. At the C- and N-terminal ends the structure unravels.

Figure 3.14: Helicity along the peptide sequence as percentage of time for the antiparallel tisB tetramer.

3.3.3.2 Tilt angle

Only the full helical central part of the tetramer peptides (positions 6 to 22 in cluster 1, and positions 5 to 24 in cluster 2) were adopted for the calculation of the tilt angle. The tilt angle describes the orientation of the helical part against the z-axis of the lipid bilayer membrane. Apart from the calculation of the tilt angle for each peptide the overall tilt angle of the assembled tetramer complex against the membrane normal was determined. It should be noted that chain 1-2 and chain 3-4 are the associated sub dimers.

A distinguishable behaviour of the tilt angle between the clusters can be noticed. While the peptides assembled in cluster 2 show a continuous behaviour over the simulation time with a fluctuation around 15° in the range between 35° and 45°, the range of the fluctuation for cluster 1 is larger. Moreover, the behaviour of each peptide changes over time. The peptides incorporated in the sub dimers (1-2, 3-4) are tilted nearly over the whole

time with a difference of ~20° which indicates that both chains are not closely packed with a tight interaction pattern.
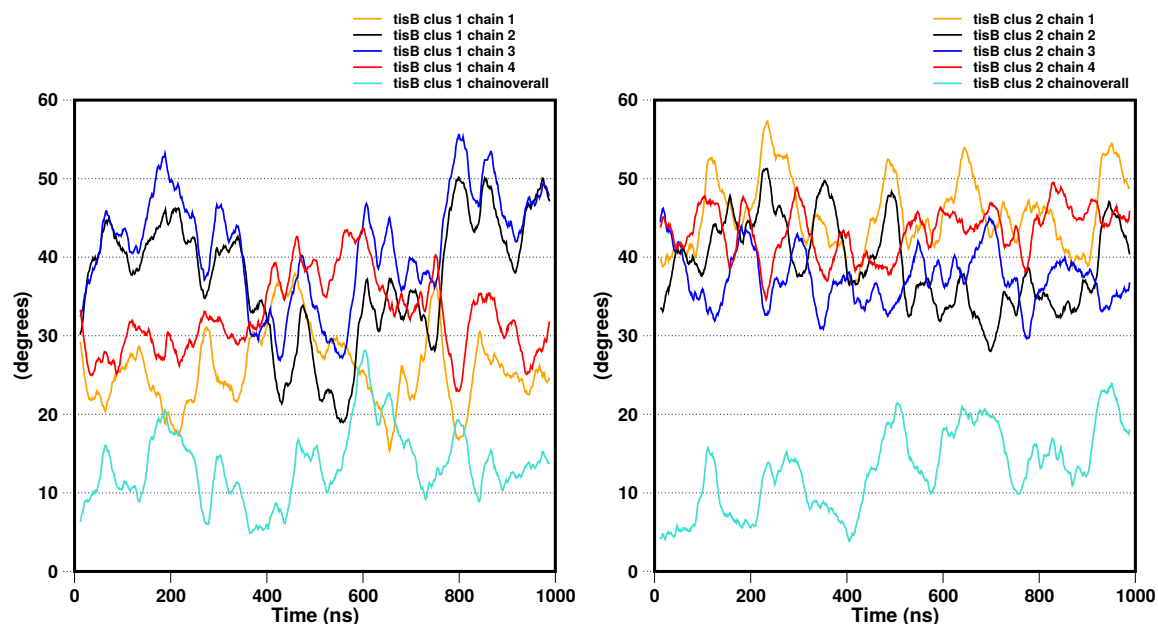
(a) Cluster 1: The tilt angle was calculated for the determined helical part from positions 6 to 22. When comparing the tilt of each chain, it is notable that the tilt angle is not stable over time and fluctuates around 20° in both directions. The overall tilt angle fluctuates around a value of 15°.

(b) Cluster 2: The tilt angle was calculated for the determined helical part from positions 5 to 24. The overall tilt angle slightly increases over the first 500 ns of simulation time and has an averaged value of ~15°. Considering the behaviour of the single peptides each of them fluctuates around ~10°.

Figure 3.15: Evaluation of the tilt angle over 1 μs simulation time. The tilt angle was calculated for each peptide and the assembled tetramer complex.

3.3.3.3 Side Chain Distances

To provide a more precise view on the interactions between the peptides assembled in the tetramer, the distances between the side chains were measured over time. To reduce the wide range of possible amino acids, only specific interactions were chosen. The main stabilisation occurs through intermolecular interactions inside the sub dimers, and also the glutamine distance plays a crucial role for the continuous polar interface. The assessed distances and the evaluation scheme for clusters 1 and 2 are depicted in figure 3.16.

For the salt bridges between the aspartic acid and the lysines the distance between the side chain nitrogen and the acidic carbon is measured. For the two possible hydrogen bonds of the glutamine interaction the distance between the carbon of the amide groups was utilised. The maximum distance for a possible salt bridge is ~4 Å and ~3 Å for a hydrogen bond. Due to the measured distance for the glutamine bonding pattern, the limit is around 6 Å.

The analysis of the side chain distances reveals a clear difference between clusters 1 and 2. For cluster 1 salt bridges may occur only for the first 500 ns of simulation time between Lys12-Asp22 and vice versa. Afterwards the distance between the side chains is increasing. Furthermore, the distance between the glutamines shows that only in the sub dimer of peptide 1-2 the glutamines are in the range for a hydrogen bond until 800 ns of simulation time. Interestingly, with increasing distance between the salt bridges the distance between the glutamines of the sub dimers (1/2 - 3/4) decreases.

Cluster 2 indicates a homogeneous picture for the interactions. The maximum distance between the charged amino acids inside the sub dimers is 1 Å. Furthermore, the Lys22-Asp12 distance is more or less stable over the simulation time and in contact. Also, the salt bridge between Lys26-Asp5 is intact at distinct points of simulation time. Therefore, an antiparallel dimer, which perfectly stacks the peptide chains, can be assumed. In addition, the glutamine interactions are all in the range between hydrogen bonds and the maximum distance of 12 Å. All glutamines are tightly packed in the centre of the tetramer, and both sub dimers are close together in this region. Moreover, the glutamines inside the sub dimer are most of the time pointing towards each other.

Figure 3.17 illustrates a snapshot of cluster 2 separated into several points of view to connect the abstract values of the side chain distance with the structure. The perfect mirrored structure of the sub dimer 1/2, 3/4 can be detected with slight difference of the tilt angle between the peptides 1 and 2, 3 and 4. Both sub dimers are arranged in such a way that a perfect *charge zipper* pattern can occur that stabilises them. Additionally, the glutamines are pointing towards each other.
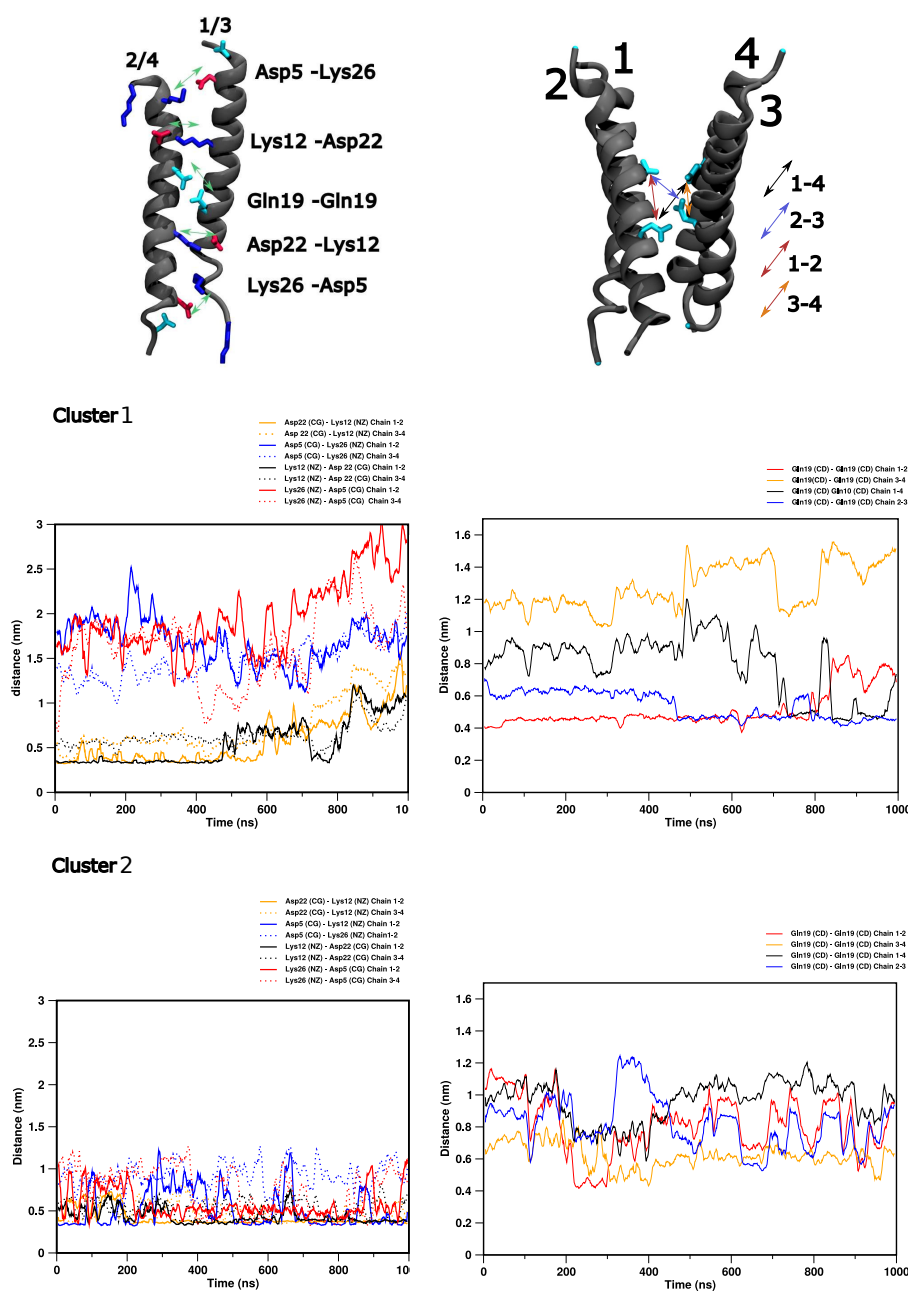
Figure 3.16: Side chain distances over time. The measured distances are shown schematically at the upper part of the figure. A clear difference between the tetramer clusters can be determined. For cluster 2 the maximum distance between the charged side chains of the sub dimer is 1 Å, and the salt bridge near the core of the peptide Asp22-Lys12 and vice versa is stable over the whole time. The same picture can be adopted to the distance of the four glutamines. They exhibit a maximum difference slightly above 1 Å, and the glutamines of the peptide 1-2/3-4 are rotated towards each other. Cluster 1, however, is increasingly developing a larger distance between the side chains over time. Until 500 ns the distances between Lys12-Asp22 are in a range for a possible salt bridge. Afterwards all side chains are far apart from each other.
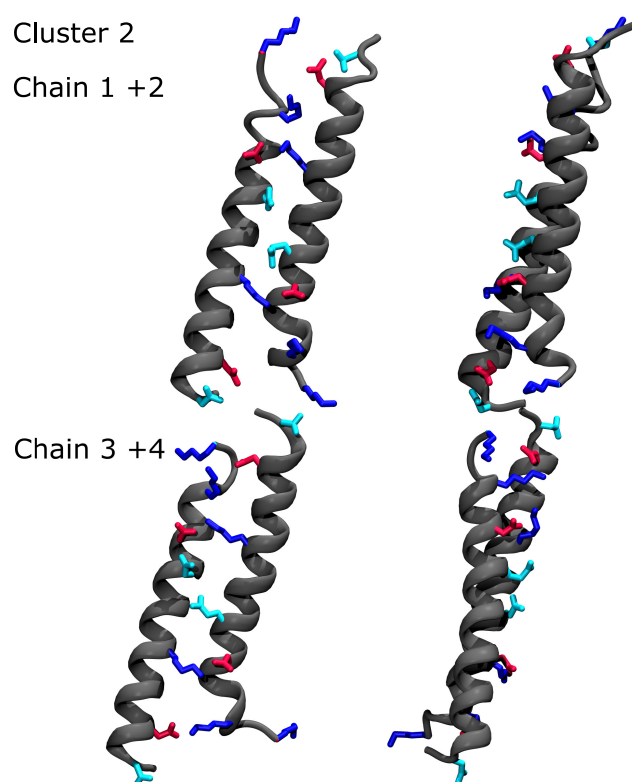
Figure 3.17: Representation of the side chain interactions of cluster 2. The tetramer struc-
ture is sliced into several points of view towards the sub dimers. It is shown
that the sub dimers arrange in the *charge zipper* pattern which stabilises the
structure. Additionally, the glutamines are pointing towards each other and
assemble with the second sub dimer to gain a continuous polar part. The
peptides 1 and 2, as well as 3 and 4, are tilted in a slightly different way. It
should be noted that the sub dimers are arranged mirror-like.

3.3.3.4 Water Permeability

Because of the tisB property to reduce the ATP-level at the bilayer through
proton transport, the simulations were screened for a water wire. Protons
can cross in two ways: Either a proton can move along a water wire, or a
hydronium ion can travel along a continuous polar interface.

Figure 3.18 shows two representative impressions of water existing in the
lipid bilayer. On the left side, the cluster 1 is shown, a similar illustration
to the dimer simulations can be identified. Water cannot pass through the
complete tetramer interface. The non-polar part in the centre, confined by
two glutamines, obstructs a passage. Here, only a proton passage, mediated
by the glutamines, is possible. Although the cluster 2 simulation provides a

quite different picture through the complete long simulation time of 1 μs, water can pass through the polar tetramer interface. This water wire is possible by the medium of the glutamines pointing towards each other to bridge the non-polar gap. The water flow is additionally mediated by the polar side chains of the *charge zipped* sub dimers. In figure 3.19 the movement of a single water molecule through the antiparallel tisB dimer is illustrated.
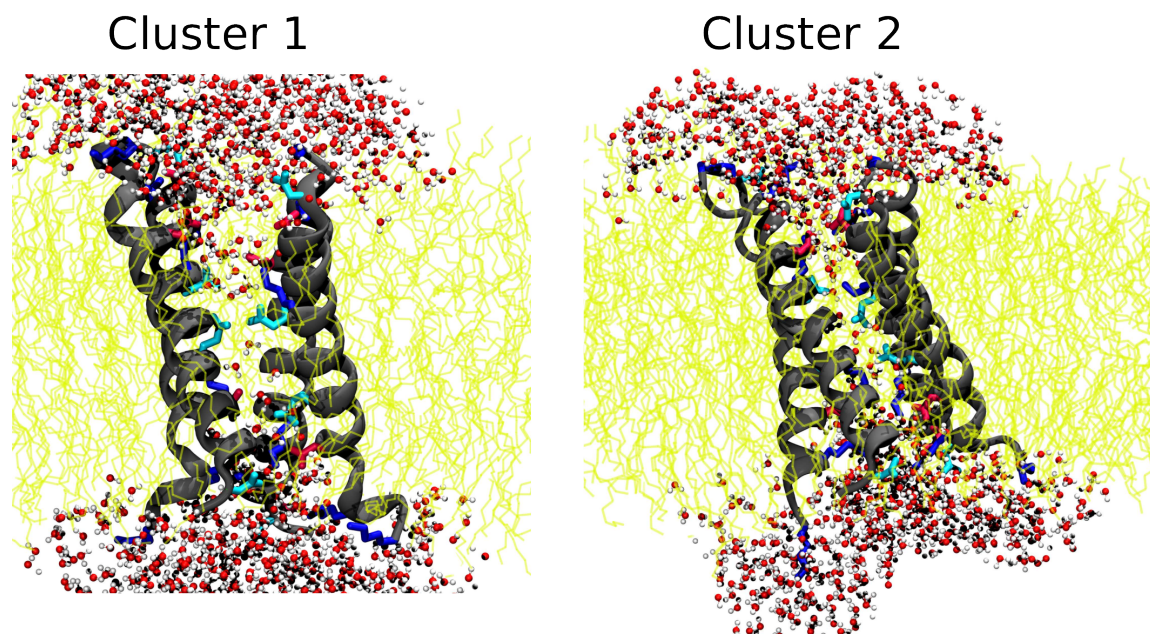
Cluster 1   Cluster 2

Figure 3.18: Representative pattern for the water movement towards the lipid bilayer induced by the antiparallel tisB tetramer. During the simulation of cluster 1 no passing of a water molecule through the tetramer could be observed. The passage is limited by the glutamines and the non-polar central part. In the simulation of cluster 2 over the complete 1 μs simulation time water molecules can exchange along the continuous polar tetramer interface. Supported by the four glutamines pointing towards each other and the stabilisation of the sub dimers via salt bridges a constant water wire was maintained.
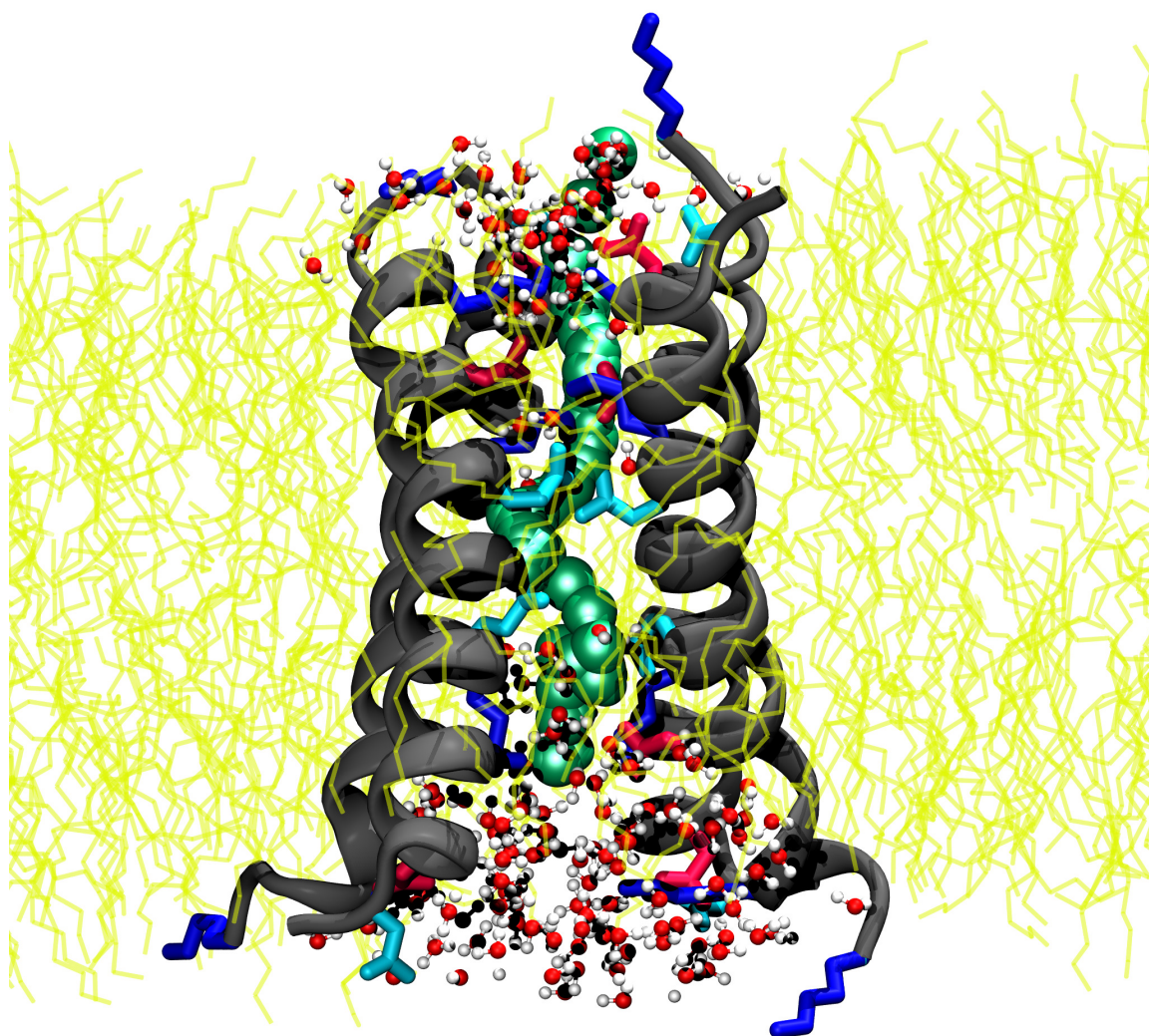
Figure 3.19: Movement of a single water molecule (coloured in green) along the polar tetramer interface of cluster 2. The passage is mediated via the charged side chains and the glutamines.

3.4 Discussion

This chapter focuses on the structural elucidation of tisB inside the lipid bilayer due to the unavailability of the protein crystal structure. Published coarse-grained simulations in 2012 [35] were not able to depict a clear picture for the behaviour of tisB due to the shortcomings of the coarse-grained description. It was outlined that the detailed side chain behaviour is the essential driving force for the specific interaction pattern and the structural alignment in the membrane. In particular, the crucial biological action, the depletion of the ATP-level, which is connected with the proton transport over the hydrophobic layer, is in the focus.

Compared to the published simulations, which were performed in DOPC (18:1) bilayer, the simulations shown here were calculated with the POPC (16:0/18:1) bilayer due to extended experimental experiments which have been applied in this environment. Previous all-atom simulations of $2\,\mu s$ of the antiparallel dimer (not published), which start from an ideal alpha-helical *charge zipper* state, were not able to display a structure with a continuous water wire or a polar interface for hydronium ion travelling. This may be a problem of sampling time. So, either extended simulation times or a starting structure, which is closer to a feasible assembled state, were needed. The problems of slow protein dynamics inside the lipid bilayer will be overcome by the coarse-grained docking assay which enables the opportunity to gain the most promising binding distribution as starting point for the all-atom simulations. Additionally to the antiparallel dimer, an antiparallel tetramer was set up.

3.4.1 Coarse-Grained Simulations

The DAFT coarse-grained approach has proven to be a *fast* method to achieve associated binding arrangements qualitatively. Also, the applied simulation length of $1\,\mu s$ for the dimer and $3\,\mu s$ for the tetramer appeared to be sufficient. For the dimer the majority of 400 different simulations ended in an associated state and in the same *X-like* binding motion with the hint for a *charge zipper* binding motion. The main difference is the tilt of the peptides against each other. Because of the favoured association state two dominating structures could be observed, mainly differing in the tilt angle and the overall compactness. For the tetramer approach the same results

were obtained. During the simulation time a huge amount of tetramer ensembles could arise, less than in the dimer simulation, but sufficient for a cluster analysis of the preferred state. Notably, the tetramer state was favourable for the antiparallel dimers. Since the active unit of tisB is an antiparallel dimer, the last snapshots of the DAFT approach were screened for these states. No associated single dimers could be observed, though all tetramers were built up via two antiparallel dimers interacting. In both approaches, all peptides remained in a transmembrane state which was also determined in experiments. Due to the coarse-grained representation with 4 atoms converting to 1 coarse-grained bead for tisB as dimer/tetramer with the narrow polar interface diameter no water was able to pass. So it can be concluded that this approach is easy to use, fast, and provides the ability to gain sufficient starting structures.

3.4.2 All-Atom Simulations

During our extensively long MD simulations (with regard to the available computer power) we could observe in each case (dimer/tetramer) for one of the clusters a possible dimer structure and a tetramer structure where protons could be transported across the lipid bilayer. The dimer structure exists only a short time, and no water directly passes, but a hydronium ion could travel along the dimer interface. This was remarkable because both clusters differ only in slight interactions.

Thus, the focus was to point out the crucial parameters of these structures. As with the other simulations before, the main advantage was to bridge the hydrophobic inner part of the protein. This was only conceivable when the glutamines were in close contact where either a continuous polar interface occurred (tetramer) or a water molecule could bridge (dimer) between the two glutamines for hydronium transport. Additionally, the occurrence of the *charge zipper* motif was crucial for the proton transfer. The salt bridges stabilise the interface and result in a directed transfer of the water/proton. Besides, the tilt angle is more upright for this pattern.

For the structures, where the zipper motion did not occur, the evolution of the side chain distances and the tilt angle, as well as the overall structure, suggest that the interface is filled with water. Therefore, it tends to tilt more and bend the helices to solvate the charged side chains in the polar

environment. Simultaneously, the hydrophobic core sticks together with the non-polar lipid environment and stabilises the assembled peptides which in turn can result in the hydrophobic border for the water passage. This assumption underscores that the peptide chains of the non-proton passage clusters are further apart. To prove this assumption the simulation should be extended, and with increasing time the unravelling of the assembly may occur or the peptides orienting toward the lipid bilayer surface.

The comparison of the briefly occurring dimer structure and the stable tetramer structure (where water molecules can pass at full length of the simulation) shows that in the tetramer assembly the two antiparallel dimers can reveal a nearly perfect alignment and a continuous dimer interface. For the only dimer structure, however, the possible interface is too narrow surrounded by the non-polar environment and not stable. Thus, this might be an artefact of the MD simulation description and needs further improvements or comparison with the experiment.

3.4.3 Comparison to Experimental Findings

The results for the dimer/tetramer structure, which support the biological action of proton transport, were briefly compared to experimental results [37]. CD measurements demonstrated that tisB is highly helical (~80 %). The calculation for the dimer shows a helical part from positions 7 to 24 which results in a value of ~60 % helicity. For the tetramer the helical part from positions 5 to 24 results in ~66 % helicity. This is in good match with the experimental results because only the central helical part was used here, whereas in the experiment the overall helicity was calculated. For the tetramer structure the transmembrane domain with 19 amino acids x 1.5 Å rise per residue = 28.5 Å matches perfect to the hydrophobic thickness of the POPC membrane of 28.8 Å.

The tilt angle of 40° in the simulation for the single peptides and around 15° for the overall tilt of the tetramer is not in a reasonable agreement with the experimental value of 20-30°. This may arise from the different estimating methods.

The idea of the *charge zipper* is in perfect match with the outcome of the MD simulations. The occurrence of a water wire and the stabilisation of

the structure goes hand in hand with the occurrence of the salt bridges for the dimer and for the tetramer structure.

Finally, for the distance between the glutamines a maximum distance of 10 Å was revealed. This is also in perfect agreement with our results. The simulation of the dimer structure shows this pattern as well as the tetramer where all glutamines are closer than 10 Å over the entire simulation time.

3.4.4 Conclusion

To conclude, our combined procedure of coarse-grained docking followed by classical all-atom MD simulations of the antiparallel tisB dimer/tetramer determines a possible dimer structure and a tetramer structure where protons could be transported across the lipid bilayer. The dimer structure persists only a short time, and no water directly passes, but a hydronium ion could travel along the dimer interface. The tetramer structure is very stable, and water can pass along the continuous polar interface. Both structures show that the proposed *charge zipper* motif and the glutamines pointing to each other are crucial for stabilisation and a feasible proton transport. It was not possible to clearly determine if the dimer or the tetramer structure is more reliable.

Further investigations are necessary. The next step is a calculation of the order parameter and NMR splittings for the dimer and tetramer and to compare them with the experimental values of the ssNMR studies. Besides, QM/MM (quantum-mechanics/molecular-mechanics) simulations of the long-range proton transfer may give a hint, which mechanism (the water wire or the hydronium ion) is more favourable. Pulling simulations in combination with free energy calculation of chloride ions may confirm the experimental findings that tisB can mediate chloride ions and whether this is possible for the dimer and tetramer structures.

4 Structural Investigations of the Peptide BsrG from *B. subtilis*

4.1 Introduction

This chapter focuses on the structural elucidation of the hydrophobic thirty-eight amino acids long type I Toxin-antitoxin peptide bsrG from *B. subtilis* and its behaviour in the lipid bilayer.

Initially, the structural behaviour of this peptide was unknown in 2015. Experimental studies were carried out simultaneously in the Ulrich working group (KIT).

At the beginning, simulations of the monomeric bsrG peptide were performed with an assumed helical folding, since Toxin-antitoxin systems are preferably α-helical, and they likely interact with the lipid membrane.

Two different groups were used for the C-terminal end to test different behaviour. On the one hand, a negative carboxyl group was used, as it is frequently applied in peptide synthesis (mainly in the laboratory). On the other hand, the uncharged amide group, which is attached to many proteins as a post-translational modification, was applied.

Initial simulations of bsrG in four diverse phosphocholine lipid bilayers (DMPC, DOPC, POPC, DEPC) with different hydrophobic thicknesses were executed. These were performed at high temperature to observe the embedding process on or into the membrane.

The first experimental results indicated that bsrG is preferably used in negatively charged lipid layers of the type POPE/PG (2:1) (which represents very well the cell membrane of *B. subtilis*). BsrG can be found in a membrane inserted state and is presumably in a higher level of oligomerisation.

Therefore, coarse-grained docking simulations were performed for bsrG in a parallel and antiparallel state as dimer and tetramer. In this way, preferred binding motifs should be found and, subsequently, all-atom sim-

ulations with the most promising candidates from the docking approach were performed.

4.2 Methods

All simulations were performed and analysed using the molecular simulation package *GROMACS* in the versions 4.6.7 and 2016.3 [94, 105–110].

4.2.1 Monomer and Dimer Simulations (All-Atom)

The 38 amino acid long bsrG peptide was constructed using the *xleap* tool from the AmberTools package [111] based on the peptide sequence shown in section 1.2.4. For the N-terminus an amino-group (NH_3^+) was attached, and for the C-terminus two different varieties were chosen, one with a *standard* carboxy-group (COO^-) and one with an amidated uncharged end (NH_2). The peptide secondary structure was assumed as a complete ideal α-helix with the backbone dihedral angles Φ =-57° (dihedral C_{i-1}, $N_i, C_{\alpha,i}$, C_i) and Ψ =-47° (dihedral N_i, $C_{\alpha,i}$, C_i, N_{i+1}).

4.2.1.1 Lipid Bilayer Equilibration

A multitude of different lipid bilayer systems were equilibrated for usage in protein membrane simulations. This was necessary because large membrane peptides like bsrG (38 aa) or tisB (29 aa) need sufficient membrane sizes in particular when simulated as higher ordered oligomers. Since these large systems were not available, it was necessary to build and equilibrate them.

As lipid bilayer 12 x 12 lipids per leaflet in a total of 288 lipids were chosen. The systems were established with CHARMM-GUI [126–130]. The lipid bilayer systems were chosen in consultation with the experimental results of the Ulrich group (KIT) to select the (*right*)-one, used in the experiments. Pure phosphocholine lipid bilayers were built up with the lipids DMPC (14:0), DOPC (16:1), POPC (16:0/18:1) and DEPC (22:1).

The negatively charged phosphoglycerol:phosphoethanolamine/phosphocholine bilayers with a 2:1 mixture were set up with DMPC/DMPG (14:0), DPPC/DPPG (16:1), POPC/POPG (16:0/18:1) and POPE/POPG (16:0/18:1). As water to lipid ratio 40:1 (except DMPC 30:1) was chosen. TIP3P water [113] was

selected, and the negatively charged membrane systems were neutralised with sodium ions. The temperature of each system was set close to room temperature and the transition temperature from gel to liquid-crystalline phase. Specific information about the lipids and equilibrated systems including the equilibration time is shown in figure 4.1.

The equilibration of the systems was performed for two different lipid force fields, the CHARMM36 [131] and SLIPIDS force field [61, 114, 115, 132]. After the *standard* membrane building and equilibration steps, following the protocol from the CHARMM-GUI package, the membrane systems were equilibrated for several hundred ns (detailed length is shown in figure 4.1). The distinct differences in the equilibration settings are described below. The system size was dependent on the lipid acyl-chain length between ~60,000 atoms (for the shortest lipid DMPC) and ~80,000 atoms (for the longest lipid DEPC).

In general, the equations of motion were integrated by means of the leap-frog algorithm [119] with a time step of 2 fs. All bond lengths were constrained with p-LINCS [67]. For temperature and pressure control two coupling groups were used: the membrane and the solvent including ions.

CHARMM36 - Equilibration protocol

Electrostatic interactions were computed using the particle-mesh-Ewald (PME) algorithm [77] with a direct space cut-off at 1.2 nm and the Verlet cut-off scheme [120]. Dispersion interactions and short-range repulsion were described by a Lennard-Jones shift-potential (force-switch modifier) with a cut-off of 1.2 nm and a shift from 1.0 to 1.2 nm. The temperature control was performed with the Nosé-Hoover thermostat [87] with a coupling time of 1.0 ps. For regulation of the pressure the Parinello-Rahman barostat [89] was applied with a semi-isotropic pressure coupling with compressibility $\kappa_{xy} = \kappa_z = 4.5 \cdot 10^{-5} bar^{-1}$ and a time constant of 5.0 ps.

SLIPIDS - Equilibration protocol

Electrostatic interactions were computed using the particle-mesh-Ewald (PME) algorithm [77] with a direct space cut-off at 1.5 nm and the Verlet cut-off scheme [120]. Dispersion interactions and short-range repulsion were described by a Lennard-Jones shift-potential (switch modifier) with a cut-off of at 1.5 nm and a shift from 1.4 to 1.5 nm. The temperature control was performed with the Nosé-Hoover thermostat [87] with a coupling time of 0.5 ps. For regulation of the pressure the Parinello-Rahman barostat [89]
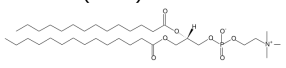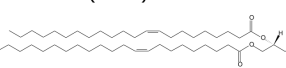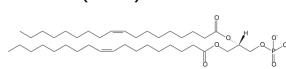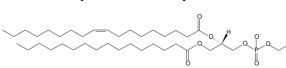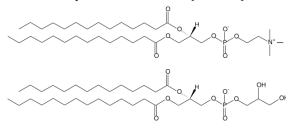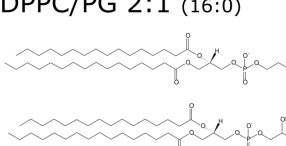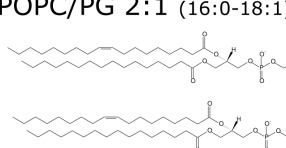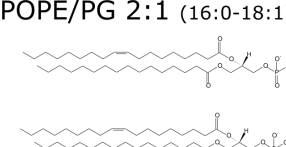
Name	net charge	T_m [a]	hydrophobic thickness [a]	force field	T_{equi}	equi length
DMPC (14:0)	0	297 K	25.7 Å	CHARMM 36	303 K	544 ns
				SLIPIDS	303 K	200 ns
DErPC (22:1)	0	286 K	31.5 Å	CHARMM 36	303 K	495 ns
DOPC (18:1)	0	256 K	27 Å	CHARMM 36	303 K	534 ns
				SLIPIDS	303 K	157 ns
POPC (16:0-18:1)	0	271 K	28.8 Å	CHARMM 36	303 K	539 ns
				SLIPIDS	303 K	200 ns
DMPC/PG 2:1 (14:0)	-1		25.4 Å	SLIPIDS	303 K	200 ns
DPPC/PG 2:1 (16:0)	-1			SLIPIDS	303 K	200 ns
POPC/PG 2:1 (16:0-18:1)	-1			SLIPIDS	298 K	200 ns
POPE/PG 2:1 (16:0-18:1)	-1		28.31 Å	SLIPIDS	303 K	200 ns

Figure 4.1: An overview of the equilibrated lipid bilayer systems. The crucial specifications of the systems are depicted in the table. Two categories of lipid bilayers were equilibrated, pure PC and negatively charged PE:PG/PC (2:1). The equilibrations were performed for a system size with in total 288 lipids (144 per leaflet). The water to lipid ratio was 30:1 for DMPC and for all other systems 40:1. SLIPIDS and CHARMM36, were used for equilibration. T_m gives the transition temperature from the gel to the liquid-crystalline phase of the lipid bilayer. T_{equi} is the chosen simulation temperature, above the transition temperature of the lipid-bilayer. The hydrophobic thickness is the size of the hydrophobic core of the bilayer (excluding the head-groups). These experimental values [a] were obtained from [133].

was applied with a semi-isotropic pressure coupling with compressibility $\kappa_{xy} = \kappa_z = 4.5 \cdot 10^{-5} bar^{-1}$ and time constant of 10.0 ps.

After the equilibration, the quality of the equilibrated lipid bilayer was verified. Therefore, the lipid bilayer was visually checked and, additionally, the area per lipid was calculated and compared to experimental values [133] and benchmarks for the CHARMM36 force field [134]. All systems were equilibrated well.

4.2.1.2 Monomer Insertion - Simulation

To study the peptide membrane interactions and the mechanism of insertion into the lipid bilayer, a high temperature simulation was performed followed by a long simulation at lower temperatures. The fully α-helical bsrG peptide was described in two configurations with an amino-group (NH_3^+) as N-terminus and two different C-terminus varieties a *standard* carboxyl-group (COO^-) and an amidated uncharged end (NH_2). For the lipid bilayer environment four pure phosphocholine bilayers were chosen: DMPC, POPC, DOPC and DEPC. These bilayers differ in the size of the hydrophobic membrane core (for more information see figure 4.1). This leads to seven simulations in total (for the DEPC membrane only the carboxyl N-terminus was simulated). The CHARMM36 [131] force field was used for the peptide and the lipid bilayer.

Simulation Parameter

The equations of motion were integrated by means of the leap-frog algorithm [119] with a time step of 2 fs. All bond lengths were constrained with p-LINCS [67]. For the temperature and pressure control two coupling groups were used: the protein and the solvent including ions. The temperature was set to 303 K and the pressure to 1 bar with an isotropic compressibility of $\kappa_{xy} = \kappa_z = 4.5 \cdot 10^{-5} bar^{-1}$. During the equilibration the heavy atoms of the peptide were restrained by a force of 1.000 kJ/nm in three-dimensional directions. For the production runs the Nosé-Hoover thermostat with a coupling constant of 0.8 ps was applied. Electrostatics were treated with the particle-mesh Ewald algorithm [77] and the Verlet [120] cut-off scheme. The particle-mesh Ewald algorithm was utilised with a direct-space cut-off of 1.2 nm. The Lennard-Jones interactions were cut off at 1.2 nm and from 1.0-1.2 nm a switching function was applied (using the force-switch modifier).

Unrestrained membrane binding simulations of 500 ns at 303 K were performed to obtain the membrane-peptide complexes. First of all, the peptide was placed parallel to the pre-equilibrated lipid bilayers at a distance of approximately the length of the bsrG peptide. Afterwards, the system was solvated with ~30,000 TIP3P water and neutralised with sodium ions. Due to the large box side with the peptide far apart from the membrane the total system size was around 100,000 atoms. The temperature was set to 480 K in order to speed up insertion. During this high temperature simulation the hydrogen bonds in the helical peptide were restrained to the ideal helical structure using distance restraints of 10,000 kJ/nm^2 to prevent unfolding. The system was minimised using the steepest-descent algorithm, followed by 100 ps NVT equilibration, 1 ns NPT equilibration and 20 ns production run. Afterwards, the system was slowly cooled down to 303 K in 20 ns.

In order to reduce the simulation size after protein absorption 20,000 water atoms were deleted from every simulation, which result in box sizes with around 70,000 atoms in total. Once again, a 100 ps NVT equilibration and a 1 ns NPT equilibration followed with position-restraints on the peptide heavy-atoms with a force of 10,000 kJ/nm in three-dimensional directions. Eventually, a 500 ns simulation was performed.

4.2.2 DAFT - Coarse-Grained Docking Approach

The DAFT approach (see also section 2.2.2.1) [102] was executed for bsrG as a parallel/antiparallel dimer and a parallel/antiparallel tetramer. The number of DAFT simulations was set to 400, and the simulation time was set to 1 μs for the dimer and to 3 μs for the tetramer. As distance between the protein chains 3.5 nm were chosen. The proteins were embedded in a POPE/POPG (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine /1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoglycerol) 2:1 bilayer solvated with water and neutralised with chloride ions. The simulation temperature was set to 300 K and the pressure to 1 bar in order to keep the POPE/POPG bilayer in the fluid phase. The temperature and the choice of lipid bilayer was in accordance to the experimental approaches and necessary to mimic the bacterial membrane of *B. subtilis.* For the whole system the coarse-grained force field *Martini 22* [84, 90–93, 100] was adopted. The

total system size was ~3,700 atoms for the dimer simulations and ~6,300 for the tetramer simulations.

The dimer and tetramer simulations were performed according to the standard DAFT protocol. First, the helical bsrG model was aligned along the z-axis using the tool *editconf* from the *GROMACS* package. This all-atom protein model was coarse-grained using the tool *martinize*, duplicated and distributed as antiparallel dimer rather than tetramer over the corresponding DAFT layout.

Then, each system (of the different 400 starting configurations) was energy minimised with the steepest-descent algorithm and afterwards shortly equilibrated via 10 ps of NVT simulation using a 2 fs time step and a 100 ps NPT simulation using a 20 fs time step. Finally, the long production run (400 x 1 μs dimer/3 μs tetramer) was performed. The time step was set to 20 fs. For regulation of the pressure the Berendsen barostat [85] was applied with a semi-isotropic pressure coupling, a compressibility of $\kappa_{xy} = \kappa_z = 3 \cdot 10^{-4} bar^{-1}$ and a time constant of 3.0 ps. The temperature was controlled via the V-Rescale thermostat [86] with a coupling constant of 1.0 ps. For temperature and pressure control three energy groups were used, one for the protein, one for the membrane and one for the solvent including the ions. Electrostatic interactions were computed using the particle-mesh-Ewald (PME) algorithm [77] with a direct space cut-off at 1.2 nm and the group cut-off scheme. Dispersion interactions and short-range repulsion were described by a Lennard-Jones shift-potential (potential-shift-Verlet modifier) with a cut-off of 1.2 nm and a shift from 0.9 to 1.2 nm.

4.2.3 Dimer Simulations (All-Atom) after DAFT

Conversion from coarse-grained to all-atom

The two most appropriate parallel/antiparallel dimer structures from the DAFT-approach were chosen (selected after clustering) and embedded into a lipid bilayer. Therefore, the obtained coarse-grained structures were (re-) converted to all-atom with the tool *backward* [103], see section 2.2.2.2. The backmapping is performed using a library of mapping definitions for the geometric reconstruction. For description of the protein the *AMBER14SB* force field [112] was used. The resulting structures were subsequently

relaxed through energy minimisation and short MD-simulations with the standard protocol from the *backward* tool.

Protein embedding into the lipid bilayer

Afterwards, the protein dimer/tetramer was aligned along the z-axis, which corresponds to the lipid bilayer normal. The dimer/tetramer was then inserted into a pre-equilibrated and fully hydrated POPE/POPG (2:1) bilayer consisting of 288 lipids (144 per leaflet) with ~11.500 TIP3P water [113] molecules (~40 water per lipid, see also section 4.2.1.1) and 91 sodium ions (to neutralise the bilayer charge). For the description of the lipid bilayer the *SLIPIDS* force field [61, 114–116] was adopted. The embedding was performed with the tool *g_membed* [117] with standard parameters. Here, the proteins were first reduced in size in the membrane plane, overlapping lipids and waters were removed. Then, the proteins were grown back in size followed by a short MD-simulation. This approach has the advantage that the system is near equilibrium afterwards. Finally, the positive charge of the system due to the net charge of the protein was neutralised, deleting sodium ions (due to the positive charge of the peptide). The sodium ions were described by the Cheatham and Young parameters [118].

For the four dimer simulations (two parallel, two antiparallel) the simulation size was ~71,000 atoms in total (2 bsrG proteins, 187 POPE lipids, 93 POPG lipids, ~11,500 TIP3P water and 92 sodium ions).

Simulation protocol, AMBER14SB and SLIPIDS

The simulations were performed with a time-step of 2 fs by integration using the leap-frog algorithm [119]. All bonds were restrained using the p-LINCS algorithm [67]. Electrostatic interactions were computed using the particle-mesh-Ewald (PME) algorithm [77] with a direct space cut-off at 1.5 nm and the Verlet [120] cut-off scheme. Dispersion interactions and short-range repulsion were described by a Lennard-Jones potential with a cut-off of 1.5 nm and a force-switch function from 1.4 to 1.5 nm. For the thermostat and barostat three coupling groups were considered: Protein, POPE/POPG-bilayer and the solvent including the chloride ions. During equilibration the positions of the heavy atoms of the peptide were constrained by a force of 1,000 kJ/nm in three-dimensional directions.

First, the system was energy minimised using the *steepest-descent* integrator with a step size of 0.01 nm, until the ultimate tolerance of 1,000 kJ/nm or the maximum step number of 1,000 steps was reached.

Then, the system was equilibrated 1 ns at constant pressure in the NPT ensemble (N: number of atoms, P: pressure and T: temperature) at 300 K and 1 bar. The temperature was controlled with the Nosé-Hoover thermostat [87] with a coupling time of 0.5 ps. For the regulation of the pressure the Parinello-Rahman barostat [89] was applied with a semi-isotropic pressure coupling with compressibility $\kappa_{xy} = \kappa_z = 4.5 \cdot 10^{-5} bar^{-1}$ and a time constant of 10.0 ps. Lastly, for the production simulation, 110 ns of NPT simulation ($5.5 \cdot 10^5$ steps) were performed at 300 K using the Nosé-Hoover thermostat with a coupling time of 0.5 ps and 1 bar using the Parinello-Rahman barostat with a coupling time of 10 ps.

4.2.4 Analysis

The visualisation and graphical analysis was performed with the tools *Gnuplot* [121], *Grace* [122] and *VMD 1.9.2* [123]. The first 10 ns of the all-atom simulations were omitted to allow for equilibration of the system. To account for periodic boundary conditions (PBC) for visualisation and analysis, the trajectory was processed in several steps with the *GROMACS* tool *trjconv*.

Cluster analysis

The *GROMACS* tool *cluster* was used to perform a cluster analysis on the DAFT simulation results. In this way, similar structures should be found and the two most populated structures for all-atom simulations should be obtained. A matrix of RMSDs between the molecule coordinates of the last snapshot at 1,000 ns for the dimer and 3,000 ns for the tetramer from the 400 different simulations was calculated. A cut-off radius of 0.4 nm in combination with the single linkage algorithm [124] was used to obtain clusters of structural similarity.

Particle distribution

With the *GROMACS* tool *confrms* the RMSD of two structures after least-square fitting the second structure to the first one was calculated. Hence, the particle distributions of the last snapshots from the 400 different DAFT-simulations were compared. For each assay (dimer, tetramer) the last protein structure was collected and then least square fit to the first protein chain.

Helicity calculation

To calculate the helicity H_i for each residue i averaged over time the dihedral angles Φ (dihedral $C_{i-1}, N_i, C_{\alpha,i}, C_i$) and Ψ (dihedral $N_i, C_{\alpha,i}, C_i, N_{i+1}$) were calculated for every snapshot of the trajectory. A residue is counted as helical when Φ and Ψ are within 30° around their ideal values $\Psi = -47°$ and $\Phi = -57°$. So a residue is 100 % helical if the condition is achieved for all evaluated time steps.

$$H_i = \frac{1}{N_{\text{time steps}}} \sum^{N_{\text{time steps}}} h_i$$

$$h_i = \begin{cases} 1, & \begin{aligned} &\text{if } (\Psi_0 - 30°) < \Psi_i < (\Psi_0 + 30°) \\ &\text{and } (\Phi_0 - 30°) < \Phi_i < (\Phi_0 + 30°) \end{aligned} \\ 0, & \text{otherwise} \end{cases}$$

Hydrogen bonds

With the *GROMACS* tool *hbond* the hydrogen bonds between the polar asparagines (Asn) in the dimer interface of bsrG were evaluated over simulation time by measuring the distance between the hydrogen donor acceptor pairs (nitrogen and oxygen of the side chain). All hydrogen bonds between the side chains were evaluated and the number of hydrogen bonds plotted over time. The distance for a pair of hydrogen-bonded atoms had to be below 0.3 nm.

Calculation of other protein structural parameters

To obtain information about the flexibility of the constructs the *GROMACS* tool *rmsf* was used. This tool calculates the root mean square fluctuation RMSFs (for all heavy atoms). In addition, the root mean square deviation was computed with the tool *rms*. The RMSD describes the structural stability in comparison with a reference structure over the simulation time.

For the calculation of the tilt angle between helix axis and the membrane normal (z-axis) the tool *bundle* was used. Prior, the helicity for each chain was calculated and only the transmembrane part with > 90 % helicity over simulation time was taken into account for the tilt angle calculation. Besides, the tilt angle was calculated for the overall protein complex and for each chain separately.

4.3 Results

4.3.1 BsrG Monomer - Insertion

To investigate the behaviour and the interaction of the bsrG monomer respectively into the lipid bilayer, all-atom simulations of the insertion process at high temperatures were executed. This method has already been used to demonstrate the insertion of peptides into the membrane [135, 136]. For small peptides, such as PGLa and Magainin2, it is an appropriate tool to shed light on the membrane embedding. Also, comparable results to the experiment could be obtained.

The four chosen phosphocholine membranes for the monomer insertion simulation of bsrG have a different hydrophobic thicknesses: 25.7 Å for DMPC (14:0), 27 Å for DOPC (18:1), 28.8 Å for POPC (16:0/18:1) and 31.5 Å for DEPC (22:1). BsrG is a very long peptide with a predicted transmembrane length of twenty-three amino acids which corresponds to 34.5 Å. Therefore, the membrane thickness might influence the performance of the peptide.

Furthermore, the behaviour of the different C-terminal ends should be determined.

In the simulations, the peptide was assumed helical and restrained at the high temperature phase. Otherwise, these high temperatures would lead to unfolding of the peptide. The high temperature phase was simulated 20 ns at 480 K, followed by 20 ns at 303 K (so that all lipid membranes were in the liquid phase) and 500 ns of unrestrained simulation.

4.3.1.1 Consideration of the Simulations

In figure 4.3 the last snapshots of the simulations are depicted after a total of 540 ns simulation time. All simulations, regardless of the selected C-terminal end and the choice of the lipid bilayer, resulted in the same picture. The monomer is attached to the lipid membrane with the polar amino acids and charged ones dip into the polar environment of the water.

At the beginning, the monomer is aligned a few nanometers parallel to the membrane, but envolves quickly aligned in an upright orientation to the membrane. The first interactions of the positively charged N-terminal end with the lipid head groups occur. Subsequently, the hydrophobic amino

acids are added gradually to the hydrophobic surface of the lipid membrane. The polar amino acids turn towards the polar water.

After the absorption of the hydrophobic amino acids to the surface the peptides remain in this position and only move over the surface.

4.3.1.2 Root Mean Square Fluctuation (RMSF)

The root mean square fluctuation (see figure 4.2), which indicates the mobility of the helix over the simulation time in comparison to the start structure, was computed along the residues over the 500 ns simulation time (after docking to the membrane at 303 K). It can be seen that the peptides move very sparsely over time according to the RMSF. The structures are in a very balanced position, however, at the peptide ends larger fluctuations appear, since the helical structure is dislocated at the ends (as shown in the pictures).

When embedding the bsrG monomer with the amide end group in DOPC and POPC, a kink in the helix emerges. However, the kink was present since the end of the adsorption process and did not lead to embedding over the following 500 ns of simulation time. The comparison of the RMSDs shows that these two simulations exhibit the largest fluctuations.
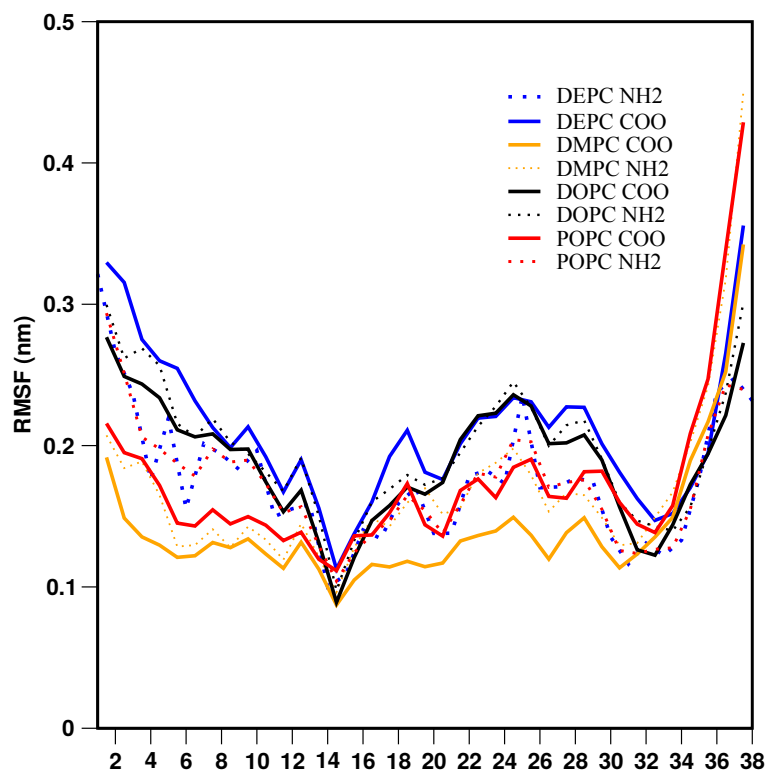
Figure 4.2: Representation of the structural flexibility of the peptides by calculated RMSF values for each amino acid residue for the 500 ns simulation at 303 K. Overall, only a few structural changes of the adsorbed peptide helix to the membrane surface exist. The very polar ends with the positive and negative charges fluctuate slightly more (also due to the unfolding of the helix peptide, as can be seen in the representative structures).

DMPC COO⁻   DMPC NH$_2$

DOPC COO⁻   DOPC NH$_2$

POPC COO⁻   POPC NH$_2$
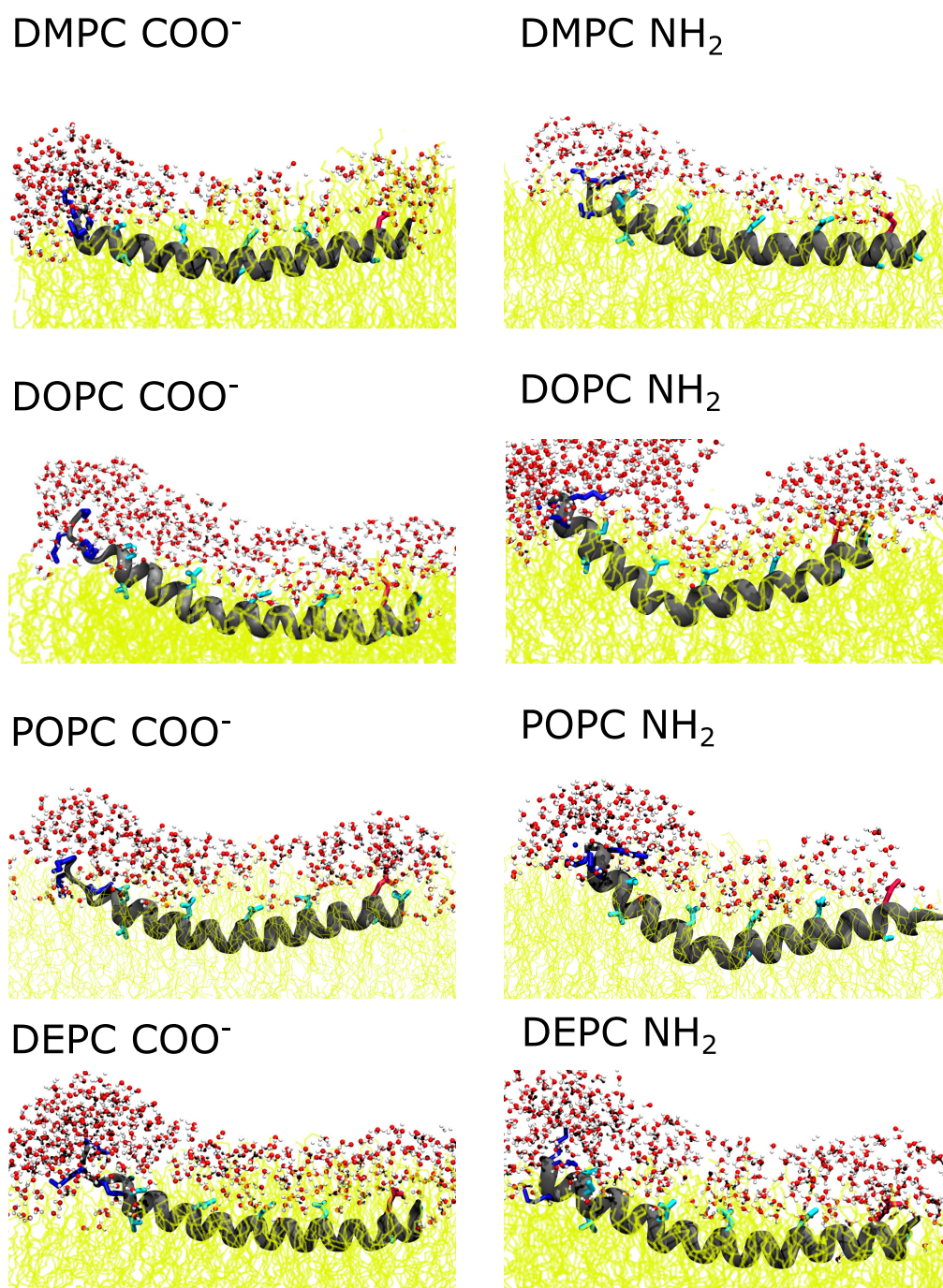
DEPC COO⁻   DEPC NH$_2$

Figure 4.3: Representative pattern of the last snapshots of the bsrG monomer after the high-temperature simulation and the following 500 ns simulation at 303 K. All simulations result in the same picture regardless of the lipid environment or the C-terminal end. The peptide is attached to the membrane surface, whereby the polar and charged amino acids point towards the polar water environment, and the hydrophobic part attaches to the membrane surface. The peptide backbone is illustrated in grey, negatively charged amino acids are coloured in red, positively charged in blue. The polar amino acids are coloured in light blue, the lipid bilayer in yellow. Only the hydrophobic core is depicted.

4.3.2 DAFT - Coarse-Grained Docking Approach

To consider the binding patterns of the bsrG dimer in parallel and antiparallel orientation, a new coarse-grained docking approach (DAFT) was used. Herein, the bsrG peptide was placed in a parallel or antiparallel state with 400 different starting structures (obtained from rotation along the helical axis) and with a 3.5 nm gap placed in a POPE/PG (2:1) membrane environment, solvated with water and neutralised with chloride ions. For each dimer structure the expanded simulation time was $1\,\mu s$ and $3\,\mu s$ for the tetramer.

4.3.2.1 Dimer

Particle distribution

First, the formation of dimers was analysed by examination of the particle distributions (shown in figure 3.1). With this visual representation the preferred binding distribution and specificity of binding can be identified. Considering the parallel and the antiparallel distributions the majority of peptides assembles in a dimer orientation. The view from the top (where the backbone of peptide 2 is coloured in brown) shows that peptide 1 completely wraps around the other peptide 2. No specific orientation on one side of the helix could be determined. The view from the side (with depiction of the polar and charged amino acids (lower half of the image)) reveals that the arrangement is made in such a way that interactions between the polar amino acids can occur. Furthermore, the peptides attach themselves in such a way that they can reduce the undesired position of the polar amino acids within the lipid bilayer.

Dimer parallel

Moreover, the simulation shows that the helix is curved like a banana. This has the advantage that the polar amino acids can point to the membrane surface. So, they are able to come in contact with the polar water environment.

Dimer antiparallel

In the simulation with the antiparallel peptides, the helix dimer tilt in the membrane is only slight and not shaped like the parallel one. The interaction between the positively charged lysines and the negatively charged aspartic acid is evident.
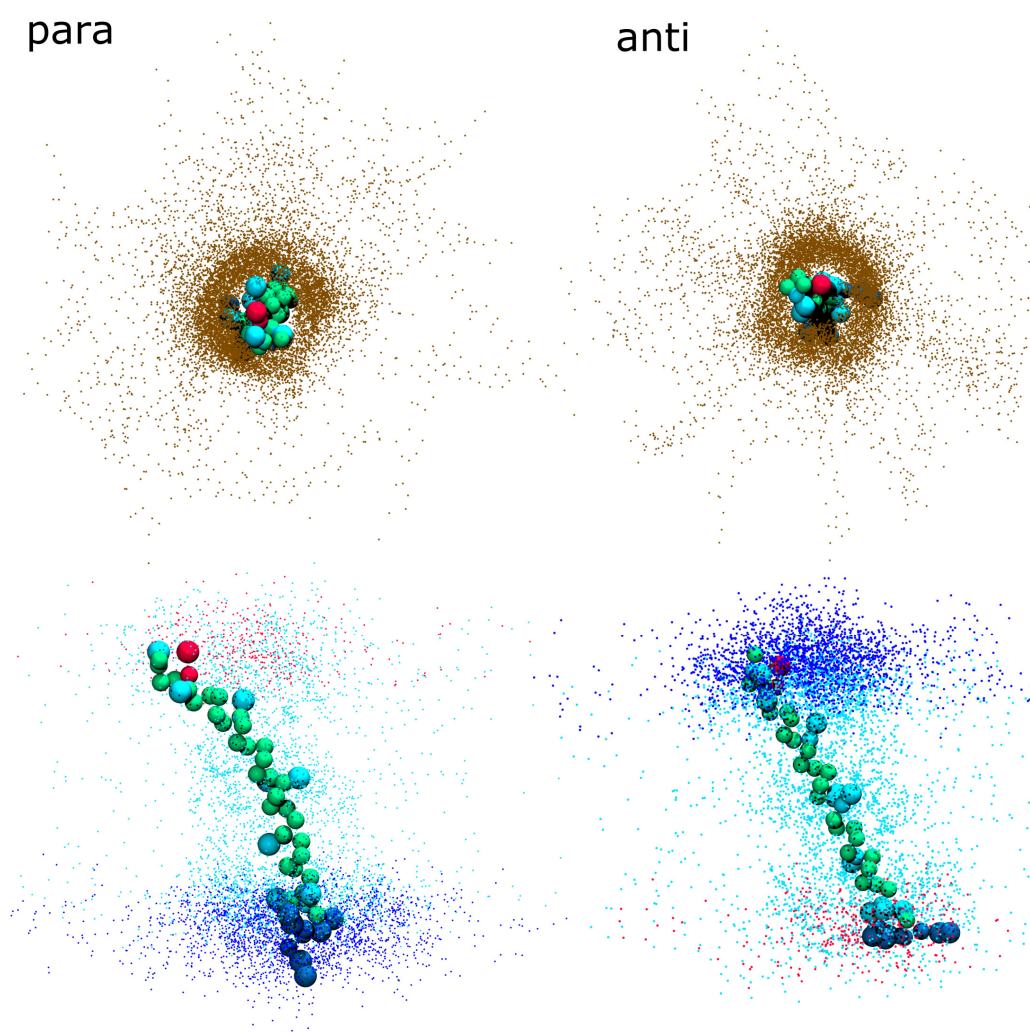
Figure 4.4: Particle distribution of one component around the other from final ensembles. For both assays, the last frames (at 1 μs) were collected, and the positions were least square fit to chain 1 of the bsrG dimer. The polar side chains (threonine, glutamine, serine and asparagine) are coloured in light blue, the negatively charged glutamic acid in red and the positively charged lysines in dark blue. The backbone of chain 1 is represented in green. On the left side, the results for the parallel dimer are depicted, on the right side for the antiparallel dimer. The upper illustration shows the distribution of the backbone atoms from chain 2 in brown, the upper illustration provides a side view of the interaction of the charged and polar side chains. In both approaches the majority of the peptide chains interact as dimer at the end of the simulation. The illustration of the specific interactions between the polar and the charged side chains shows a high preferred orientation due to the polar side chains, which are spread over the whole surface of the helical axis. Some of the polar side chains are interacting and cover themselves from the unlikely position inside the hydrophobic bilayer.

Cluster analysis and *backmapping*

Finally, a cluster analysis was performed on the last snapshots at 1 µs simulation time for each of the bsrG parallel and antiparallel dimers. Overall, the cluster analysis confirms the previous results that mainly all structures end up in a bound interface. In addition, no dominating structure was found by clustering. A lot of structures differ in the formation of the polar interface (by the polar side chains), which results in a different turn of the helices towards each other.

For the parallel and the antiparallel case the two most populated dimer conformations were chosen. These two structures are illustrated for the parallel bsrG in figure 4.6 and for the antiparallel bsrG in figure 4.5. All four structures generate a polar interface between the two peptides due to the interaction of the polar asparagines with each other and the peptide backbone.

Dimer parallel

In the case of the parallel dimers the analysis of the two cluster structures shows that cluster 2 is very compact. Both peptides are tightly packed, and a banana-shaped tilt leads to the membrane normal. In cluster 1, only the C-terminal ends are tightly packed, and there is a slight expansion of the N-terminal ends. The overall structure is similar to a "V".

Dimer antiparallel

Regarding the antiparallel dimer both found clusters align straightly to the membrane normal. The polar interface is not as densely packed as in cluster 2 of the parallel dimers. Cluster 1 of the antiparallel dimer exhibits a structure similar to a "V" (same as cluster 2 of the parallel DAFT-approach). There is also an interaction between the glutamic acid and a polar side chain. In cluster 2, the interface between the dimers is not consistently polar. Some polar amino acids are turned away from the interface, mainly due to one of the two ends of the dimer (e.g. between lysine and glutamic acid).

It should be noted that, due to the coarse-graining, the precise orientation of the side chain cannot be carried out. Also, the localisation of water inside the dimer interface is not beneficial, because of the 4:1 mapping procedure.

At last, the four promising coarse-grained structures (two parallel and two antiparallel) were converted to all-atom via the *backward* protocol (see 2.2.2.2).
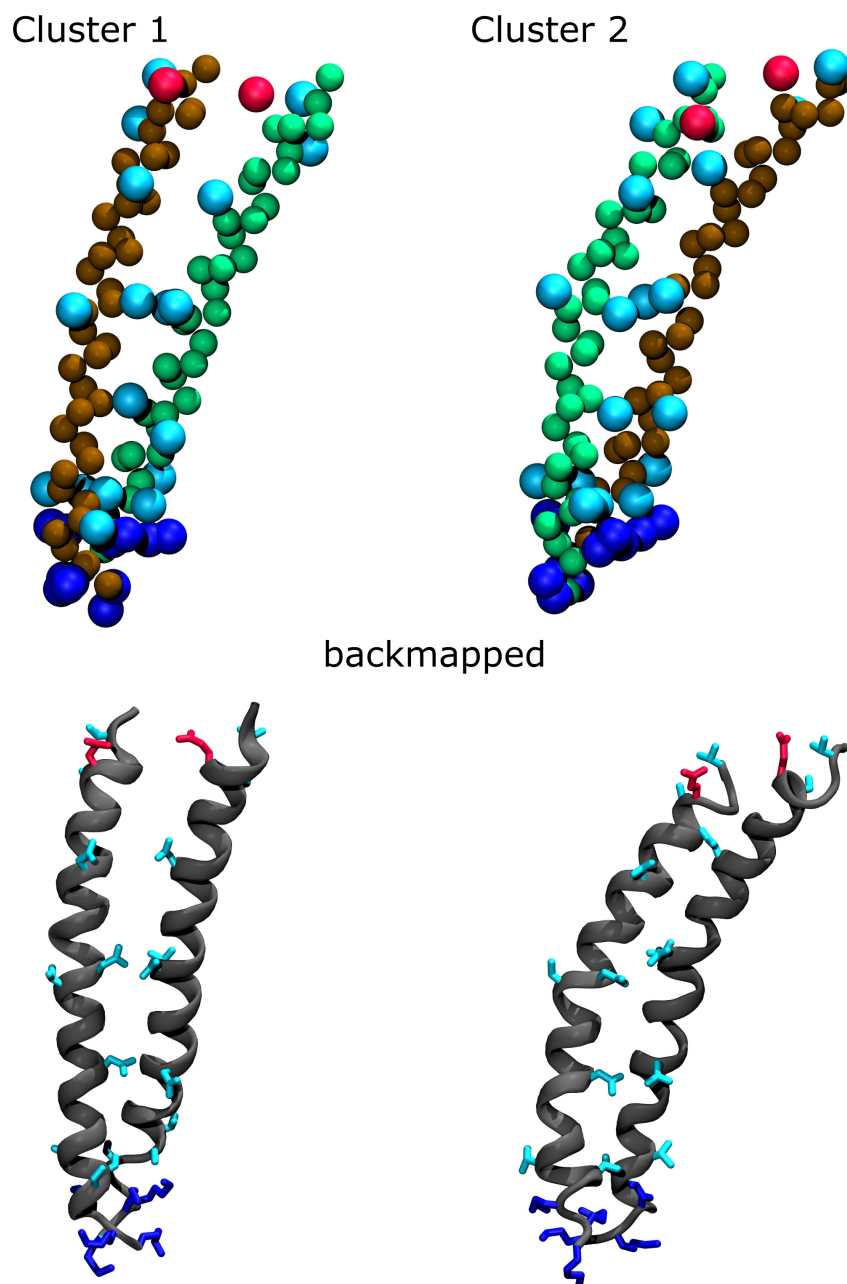
Figure 4.5: At the top, the two most populated structures from the cluster analysis of the last snap-shots simulation time of the DAFT coarse-grained approach at 1 μs for the parallel bsrG are shown. At the bottom, the converted structures of the all-atom description via the *backward* protocol are depicted. The protein backbone is coloured in grey (all-atom) and brown/green (coarse-grained). Polar amino acids are coloured in light blue, positively charged in dark blue and negatively charged in red.

Figure 4.6: At the top, the two most populated structures from the cluster analysis of the last snap-shots simulation time of the DAFT coarse-grained approach at 1 μs for the antiparallel bsrG are shown. At the bottom, the converted structures of the all-atom description via the *backward* protocol are depicted. The protein backbone is coloured in grey (all-atom) and brown/green (coarse-grained). Polar amino acids are coloured in light blue, positively charged in dark blue and negatively charged in red.
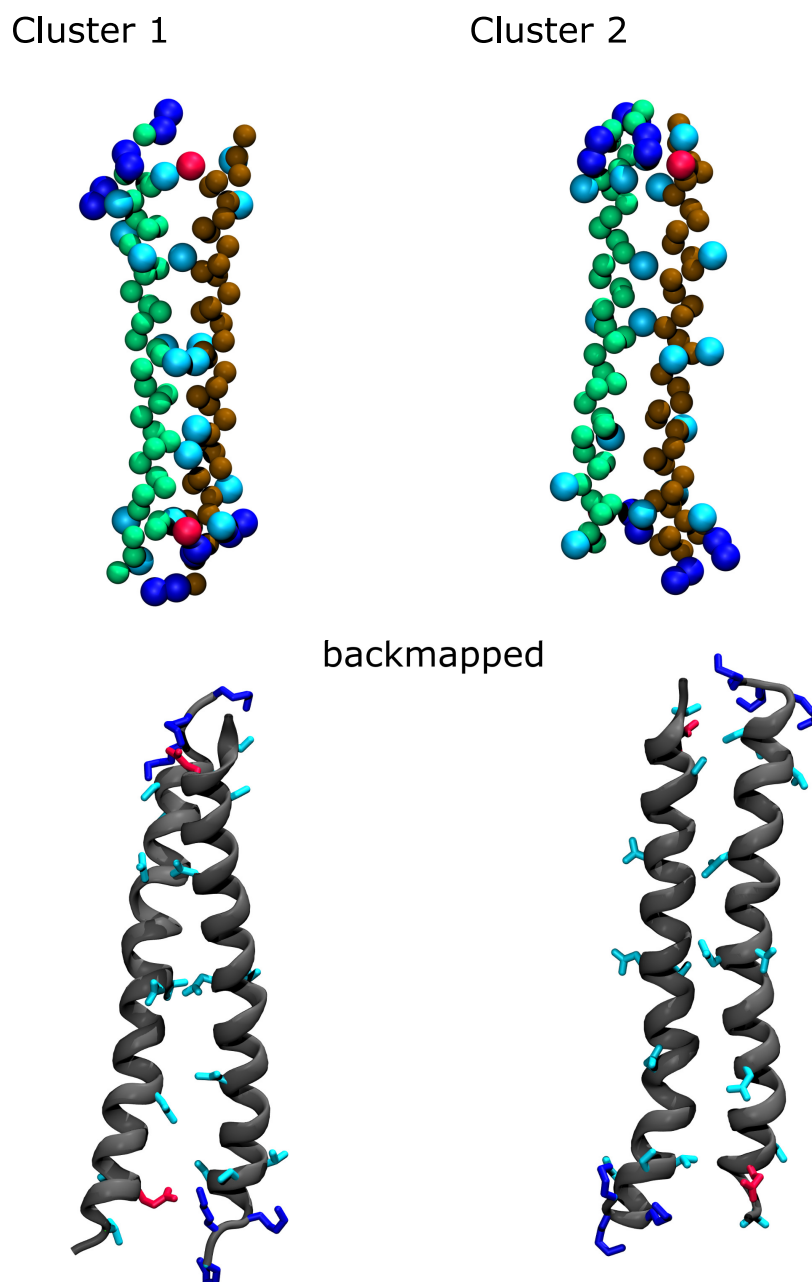
4.3.2.2 Tetramer

The tetramer simulations were analysed in the same way as the aforementioned dimer simulations. The particle distribution was used to get a quick overview of the preferred bindings. With the subsequent cluster analysis, the preferred tetramers should be found and the two most popular ones selected for the projected all-atomic simulation.

Particle distribution

The final ensembles of the extended 3 µs simulations were collected and least square fit to the first peptide of the bsrG tetramer (see figure 4.7).

The two DAFT approaches for parallel and antiparallel bsrG peptides as tetrameric binding assays produced completely different results.

Tetramer parallel

Regarding the distribution of the tetramer from above (backbone of peptides 2-4 are shown in brown) it can be seen that the majority of peptides accumulate. Adding the side view (where the polar and loaded side chains of peptides 2-4 are shown), they can be detected as tetramer states, suggesting that this association is mainly caused by the assembling of the polar amino acids. A large amount of the peptides are found in an associated state. However, the particle distribution shows no preferred orientation.

Tetramer antiparalell

For the antiparallel tetramer, the result is completely different. Considering the distribution from above no specific motive for binding can be found. There are only a couple of interactions with the represented peptide 1. Also, a large amount of peptides are arranged in a second semicircle around the referenced peptide 1. By adding the view from the side (where the polar and charged amino acids are represented) the image can be refined. Apart from a very slight tendency to attach a second peptide to peptide 1 as a dimer (as slightly inclined antiparallel binding motif), no preferred orientation can be found. Furthermore, the peptides do not remain upright in the membrane, instead they move partially to the membrane surface and occupy several different orientations. In summary, the antiparallel tetramer does not exhibit tetrameric assembly motive.

Cluster analysis and *backmapping*

After the analysis of the particle distribution a cluster analysis was performed. Therefore, the last snapshots of the 400 different simulations at the time $3\,\mu$s were taken.

Tetramer antiparalell

As already suggested, when considering the particle distribution, no preferred orientation was found in the cluster analysis. No motive assembling of tetramers increased. Therefore, no structures were selected for all-atomic simulations.

Tetramer parallel

The analysis of particle distribution showed a great amount of parallel peptides assembled to a tetramer and that the polar amino acids are rotated to each other. Cluster analysis confirms and refines this assumption. During the cluster analysis, many different clusters were found in which the bsrG peptide is assembled as a tetramer. However, no clusters with an increased population appeared. All these clusters differed only in the formation of the polar interface. Since no preferred binding motifs were found on the basis of the DAFT approaches, no specific structures were selected for subsequent all-atom simulations.
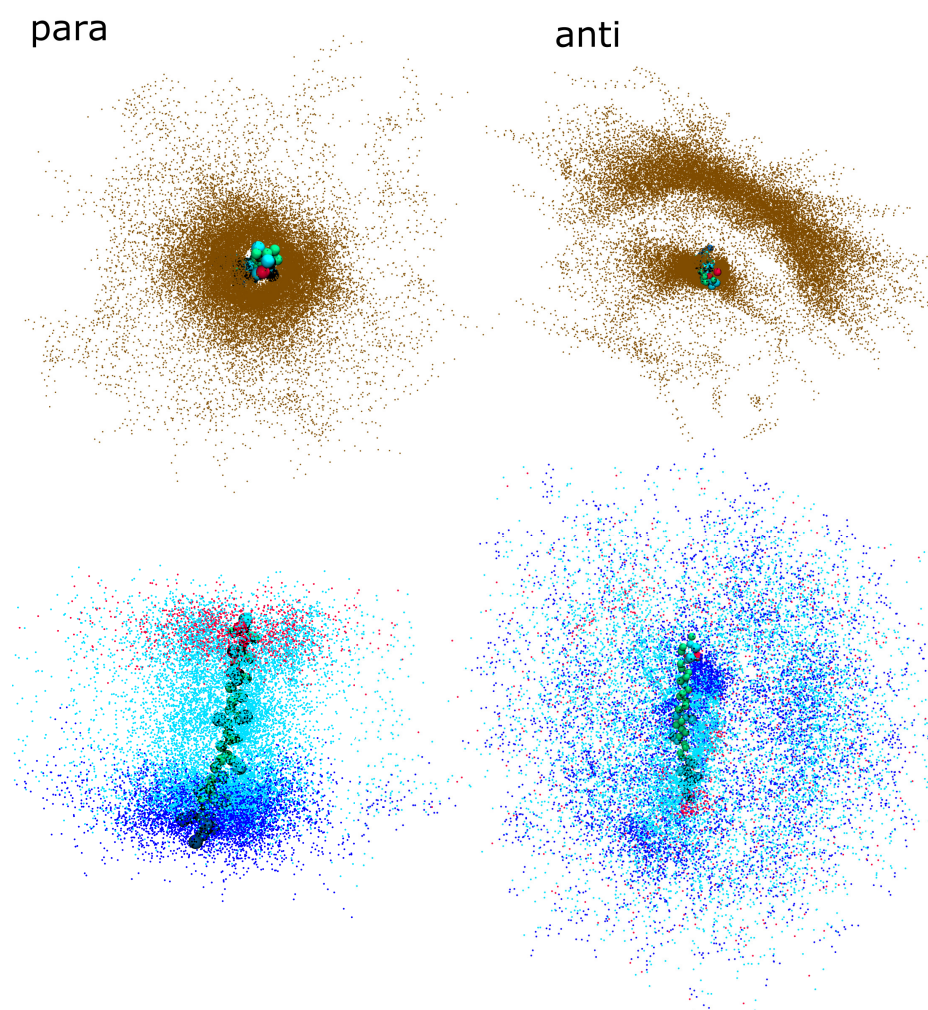
Figure 4.7: Particle distribution of one component around the other from the final ensembles at 3 µs. For each assay, the last frames were collected and least square fit to the first peptide of the bsrG tetramer. This chain is represented by balls in the coarse-grained description. Only the polar (light-blue) and charged side chains (positive in dark blue and negative in red) are represented. The backbone of chain 1 is coloured in green. The upper figure (view from the top) shows the distribution of the backbones of the chains 2, 3 and 4 coloured in brown. For the parallel approach a preferred assembly of the peptide chains can be detected, but no distinct specific binding motif. The antiparallel tetramer shows a completely different image. No assembling of the peptides as tetramer took place. In addition, the polar amino acids show that without assembling to the tetramer, the peptides are also reorientated to the membrane normal, and some peptides move to the membrane surface.

4.3.3 Dimer Simulations (All-Atom)

For bsrG as a parallel and antiparallel dimer in the all-atomic description, 150 ns simulations were performed for both selected clusters of the previous coarse-grained docking approach. The peptide was embedded in a POPE/POPG (2:1) membrane, which is negatively charged and imitates the membrane of gram-positive bacteria like *Bacillus subtilis* very well [137].

4.3.3.1 Dimer parallel

A representative structure for clusters 1 and 2 at the end of the simulation according to 150 ns is depicted in the upper half of figure 4.8. Both simulations show that bsrG is oriented upright in the membrane. The helices are slightly inclined to the membrane normal. The polar asparagines in the middle of the helix point towards each other. Also, both clusters demonstrate that some water molecules can enter the polar interface. However, this does not lead to the formation of a continuous water thread. The C- and N-terminal ends, protruding from the membrane, exhibit the unfolding of the helix in polar solvent. Comparing clusters 1 and 2, a very dense, uniform arrangement of the two peptides can be detected in cluster 2. Cluster 1, on the other hand, leads to an oval behaviour. This could be caused by the slightly deeper penetration of water molecules. That leads to a drift of the two helices, so that water can get through on the basis of their size.

In the lower half of figure 4.8 the initial and final structures of each simulation are represented by the superimposition of the backbone structure. The start structure of the simulation is illustrated in light grey and the final structure in dark grey.

Cluster 1

For the parallel dimer cluster 1, the structure of the dimer changes over the simulation time in such a way that the two strands are located at the end in a slightly crossed, x-like structure. This results in a V-shaped recess at both ends. As illustrated in the representation of the final structure with lipid and water, it can be filled with water.

Cluster 2

In cluster 2, which exhibits a banana-like appearance at the beginning of the simulation, the structure changes to two densely packed helices that are slightly crossed and stand upright in the membrane. The C-terminal end is more densely packed than the N-terminal. The representation with lipids and water indicates that water dips into the polar interface.

4.3.3.2 Dimer Antiparallel

The results of the 150 ns simulation for the antiparallel dimer is depicted in figure 4.9. During the simulation both structures are in a transmembrane state and slightly inclined to the membrane normal. Neither structure contains any water inside the membrane. The ends above the membrane are solvated in water and no longer helically folded.

In the lower half of figure 4.8, the initial and final structures of each simulation are represented by the superimposition of the backbone structure. The starting structure of the simulation is illustrated in light grey and the final structure in dark grey.

Cluster 1

The two bsrG peptides form a polar interface in which the asparagines point towards each other. In this way, they can stabilise the dimerised peptide via contacts. Considering the variation of the two peptide chains over the simulation time, it can be seen that from the original x-shaped arrangement, the peptides find their way to an erected arrangement, and the two peptides are less inclined towards each other.

Cluster 2

Compared to cluster 1, the polar amino acids do not point towards each other. Both helices have the same orientation. The polar amino acids point from one peptide to the impurity backbone of the second peptide. In case of peptide 2 the polar amino acids point to the hydrophobic membrane environment. A comparison of the dimerised peptide complex from the beginning with that of the end of the simulation shows that the dimer takes on a large structural change. One of the two peptides moves in the membrane around the second peptide. This leads to an erection of the two peptides and to an enlargement of the interaction interface between the two peptides.
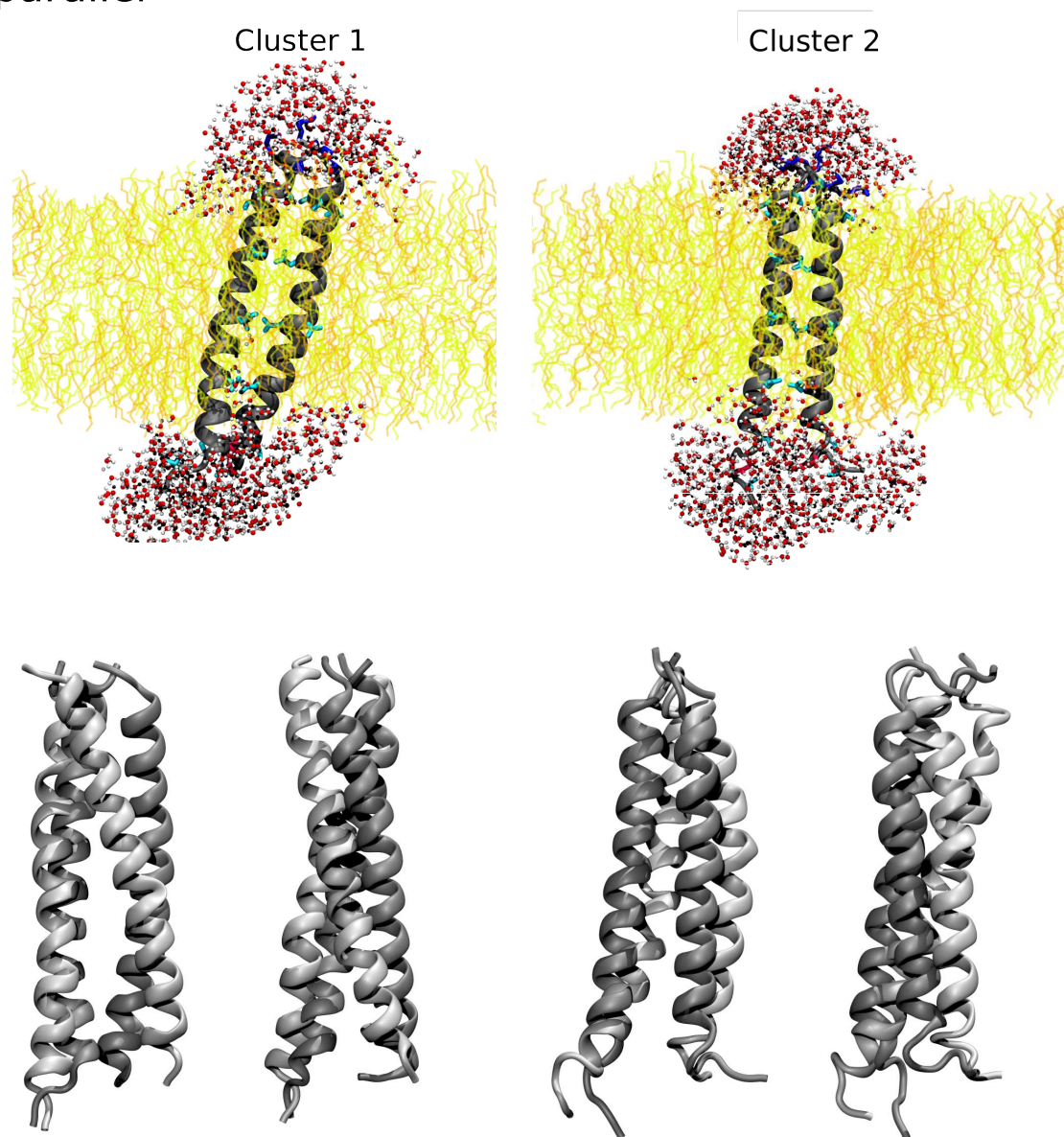
parallel

Cluster 1   Cluster 2

Figure 4.8: The last snapshots of the 150 ns simulation of clusters 1 and 2 of the bsrG parallel dimer are shown at the top. The POPE/POPG membrane is coloured in yellow/orange. The protein backbone is coloured in grey. Polar side chains appear in light blue, positively charged in dark blue and negatively charged in red. Water with a maximum distance of 8 Å to the protein is illustrated. At the bottom, the starting and the end structures of the simulations are overlaid. The starting structure is coloured in light grey, the end structure in dark grey. On the left side, the view from the front is depicted, on the right side the side view.

antiparallel

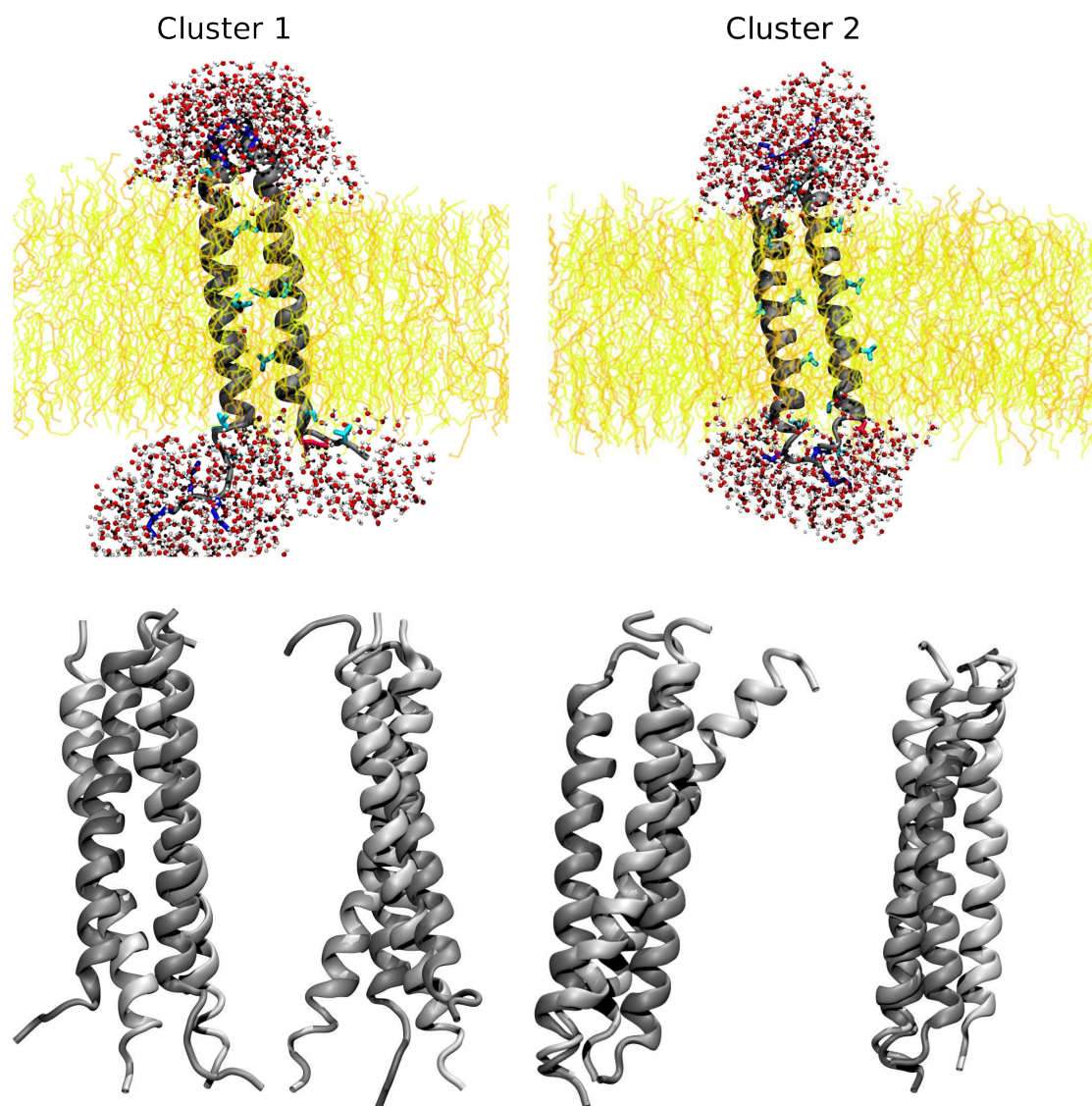Cluster 1   Cluster 2

Figure 4.9: The last snapshots of the 150 ns simulation of cluster 1 and 2 of the bsrG antiparallel dimer are shown at the top. The POPE/POPG membrane is coloured in yellow/orange. The protein backbone is coloured in grey. Polar side chains appear in light blue, positively charged in dark blue and negatively charged in red. Water with a maximum distance of 8 Å to the protein is illustrated. At the bottom, the starting and the end structures of the simulations are overlayed. The starting structure is coloured in light grey and the end structure in dark grey. On the left side, the view from the front is depicted, on the right side the side view.

4.3.3.3 Comparison between Cluster 1 and 2

Figure 4.10 shows the final structures of the two selected clusters of the bsrG parallel and antiparallel simulations. Cluster 1 is illustrated in the lighter and cluster 2 in the darker colours.

Dimer parallel

The final structures of clusters 1 and 2 are very similar. The polar amino acids point to each other and form a polar interface with interactions of the asparagines in this area, following a ladder. Only the rotation of the two helices differs from each other. The interface of cluster 1 is inverted in comparison to cluster 2, which leads to a mirror image with the polar interface on the front instead of the back.

Dimer antiparallel

The comparison of the two cluster structures reveals a major difference. The orientation and erection of the dimerised peptides is very similar, only on one side the peptides of cluster 1 are arranged upright to each other. However, the formation of a polar interface is different. Similar to the two structures of the parallel dimer, cluster 1 forms a polar interface through the interaction of the asparagine. Cluster 2 does not form a polar interface as both peptides do not point in the same direction with the polar amino acids. As a result, the polar amino acids in a peptide show up in the hydrophobic lipid membrane.

4.3.3.4 Helicity

The evaluation of the peptide helicity over 150 ns simulation time is shown in figure 4.11 for all four simulations in total. Only the helicity for one peptide of every simulation is depicted since both peptides showed the same result in the evaluation. All four simulations exhibit a similar overall shape. The N-terminal and C-terminal ends unravel, whereas the main part of the peptide is helically folded.

The helical part at the C-terminal end stops in all four simulations at position 34. However, it is different for the N-terminal end. Cluster 1 para has a helical part from positions 3 to 34 and cluster 2 para from 4 to 34. Cluster 1 anti forms the helical main part from positions 7 to 34 and cluster 2 anti from 10 to 34. This results in a total transmembrane helical length of 24 to 31 amino acids for the four different simulations.
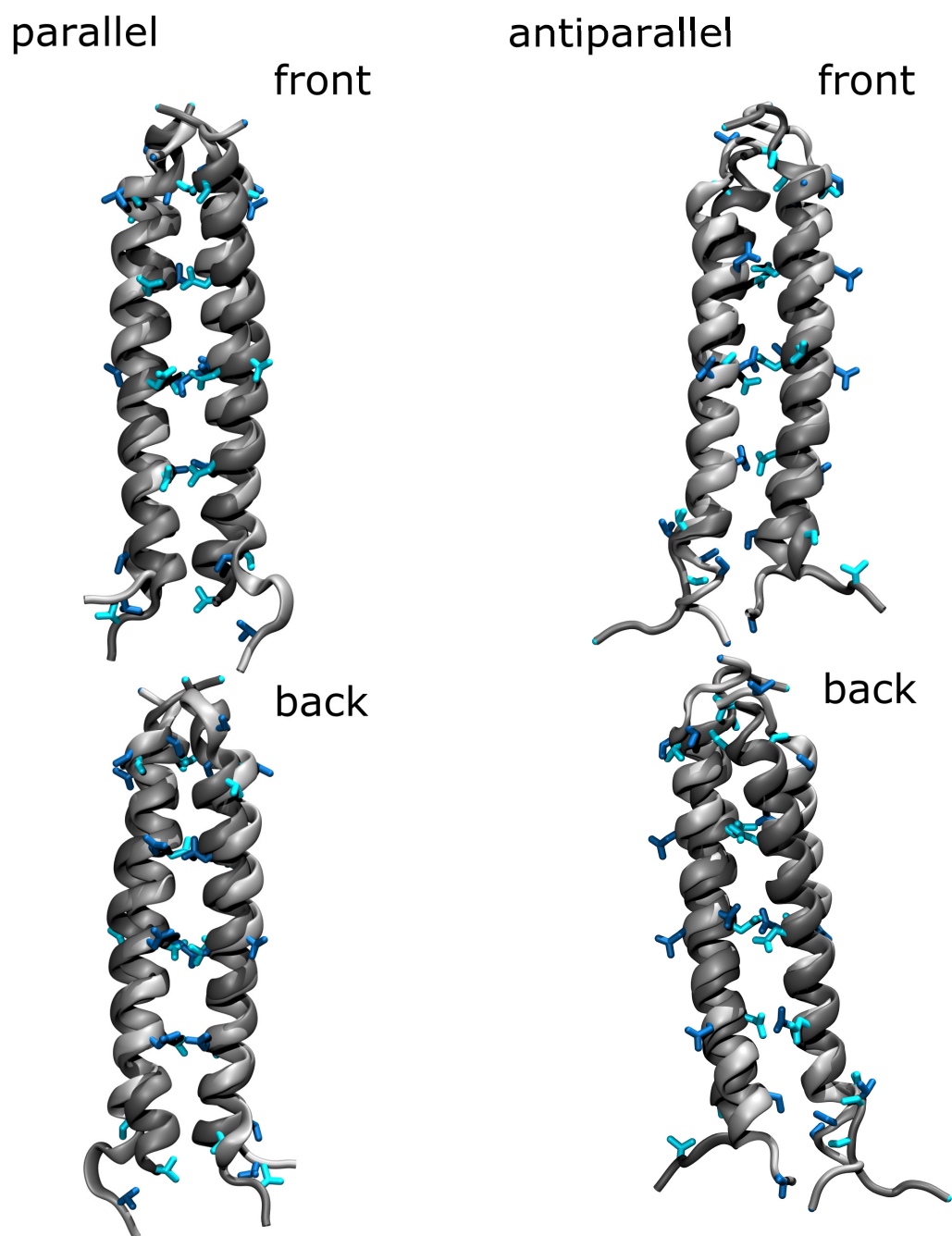
Figure 4.10: Comparison between the last structures of clusters 1 and 2 of the bsrG parallel and antiparallel dimer. The protein backbone is coloured in light grey for cluster 1 and in dark grey for cluster 2. The front and the backside views are depicted. The polar amino acids, which represent the interaction interface, are coloured in cyan for cluster 1 and in blue for cluster 2.
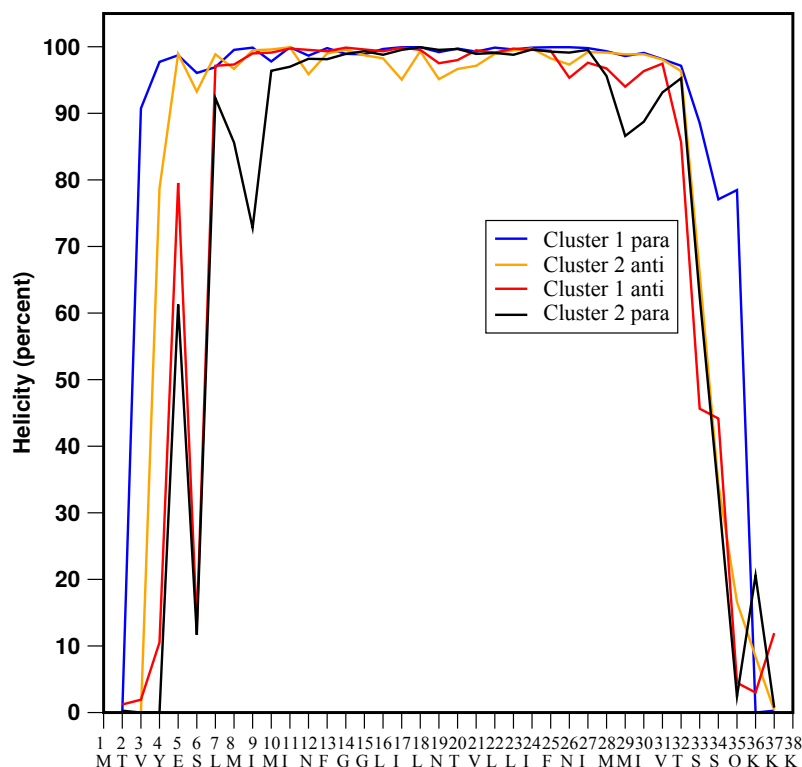
Figure 4.11: Helicity along the peptide sequence as percentage of time averaged over the 150 ns simulation. For all four simulations the helicity of peptide 1 is shown, since the helicity is equal for both peptides. The C-terminus unravels at position 34 in all four simulations. The helical behaviour of the N-terminus is different and differs by six amino acid positions.

4.3.3.5 Tilt-angle

Only the full helical central part of the parallel and antiparallel dimer (Cluster 1 para positions 3 to 34, cluster 2 para positions 4 to 34, cluster 1 anti positions 7 to 34 and cluster 2 anti positions 10 to 34) were adopted for the calculation of the tilt angle. This angle describes the orientation of the helical part against the z-axis of the lipid bilayer membrane.

All peptides show a very upright position within the lipid membrane (see figure 4.12). The maximum tilt angle is ~25°, the minimum ~ 5°. Within the 150 ns simulation, the tilt angle also varies between these values. The mean value is about 15°.
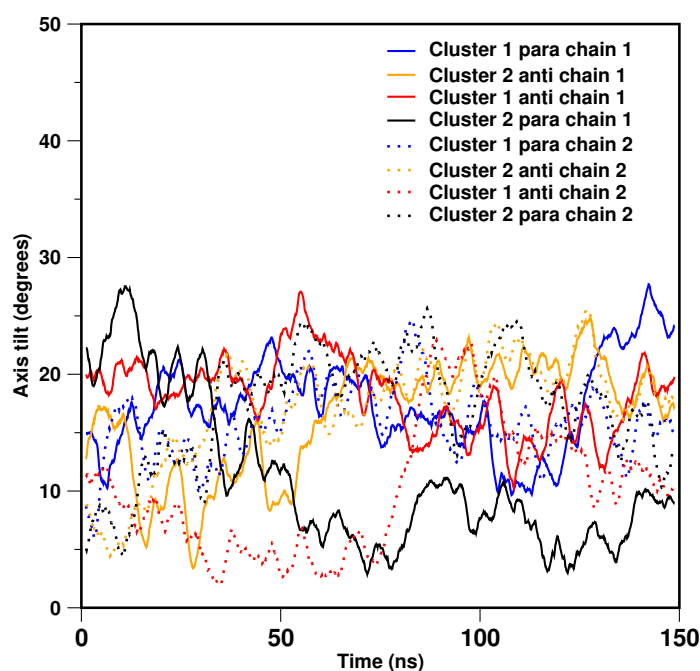
Figure 4.12: Evaluation of the tilt angle over simulation time for cluster 1 and cluster 2 of the bsrG parallel and antiparallel dimer. Peptide chain 1 is illustrated with a solid line, chain 2 with a dotted line. The tilt angle was calculated for the previously calculated helical part for each peptide.

4.3.3.6 Hydrogen Bonding Network

To provide a more precise view on the interactions between the peptides assembled in the dimer, the total amount of hydrogen bonds between the asparagines in the dimer interface is plotted over time. In this way, the assumption, that the hydrogen bonds between the asparagines are the main stabilising part of the bsrG peptide in the dimer state, could be proven. These three asparagines are 7 amino acids apart from each other and proposed to build an ascending hydrogen bond ladder (see 1.2.4.3).

Different results are obtained for the four simulations (see figure 4.13). Cluster 2 anti detects no hydrogen bonds between the two peptides. This is not unpleasant, as the polar amino acids do not point to each other. Cluster 1 para shows one to two hydrogen bridges between the peptides, while cluster 1 anti and cluster 2 para form up to five hydrogen bonds between the asparagines. The stabilising interactions for the dimerisation of the bsrG peptide are thereby the hydrogen bonds between the polar asparagines.
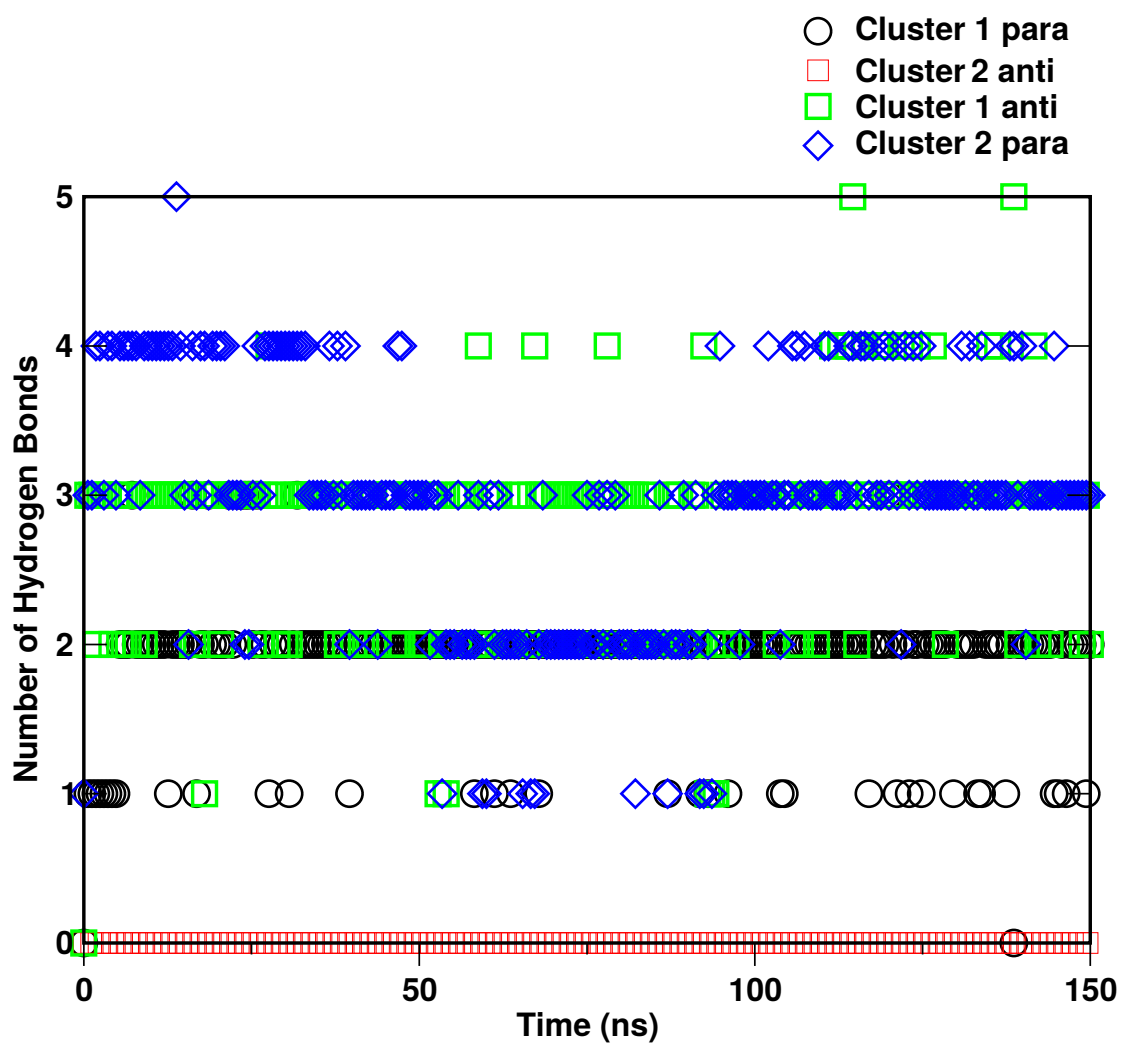
Figure 4.13: Total amount of hydrogen bonds between the asparagines in the central part of the bsrG peptide over time. For cluster 2 anti no hydrogen bonds between the two dimers can be found. Cluster 1 para is stabilised by one to two hydrogen bonds between the asparagines, while cluster 2 para and cluster 1 anti are stabilised by up to five hydrogen bonds.

4.4 Discussion

This chapter focusses on the structural elucidation of the hydrophobic 38 amino acid long peptide bsrG and its behaviour in the lipid bilayer membrane. Since no crystal structure and only a few experimental results (group of Anne Ulrich (KIT)) were available at the beginning of this work, these are the first studies to elucidate the structural behaviour of the bsrG peptide.

4.4.1 Monomer - Insertion

The studies on embedding the bsrG monomer in lipid membranes revealed that the monomer spontaneously binds flat to the hydrophobic membrane surface, but does not integrate. The polar and charged amino acids point into the polar solvent, and the hydrophobic amino acids are in contact with the membrane. Consequently, bsrG embedded in the membrane is an unfavourable state. This is consistent with the experimental results of the Ulrich Group who predicted at least one state as a dimer for bsrG.

4.4.2 Coarse-Grained Simulations

The DAFT coarse-grained approach has proven to be a *fast* method to achieve associated binding arrangements qualitatively. In addition, the applied simulation length of 1 µs and 3 µs appeared to be sufficient. This could also be shown by the studies of the tisB peptide.

4.4.2.1 Dimer

By means of the DAFT-approach, no specification between the parallel and the antiparallel orientation can be determined. In both simulations, interactions occur in almost all starting structures within the microsecond simulation. These interactions are mainly caused by the interaction between the polar side chains and (in the antiparallel case) by interaction between the positively and negatively charged ends.

4.4.2.2 Tetramer

For the parallel tetramer the formation of a tetramer is likely, but no distinct interactions can be found. All peptides are stored together and stably integrated into the membrane, but no specific binding pattern can be detected.

Based on the cluster analysis, the polar amino acids attach themselves in such a way that they assemble to protect themselves from the non-polar lipid environment, which yields energetic advantages. However, the peptides are very mobile in this conformability.

Regarding the antiparallel tetramer, on the other hand, the peptides are not stable in the membrane as antiparallel tetramer, and there is no build-up.

4.4.3 Dimer - All-Atom Simulations

Recent studies have shown that, due to the slow oriental and conformational process, timescales in the range of μs are necessary for peptide membrane systems to achieve more than a qualitative view of the simulations [138, 139]. The first results of the 150 ns simulations provide a good view of the structural parameters of the bsrG peptide, due to the preformed initial structures by the coarse-grained simulations.

Analysis of some structural parameters of the parallel and the antiparallel dimers showed for two of the four structures (cluster 1 para and cluster 1 anti) a good match with the experimental results.

In accordance with the experiment, the assembled dimers exhibit a slightly inclined transmembrane orientation within the membrane. This orientation is stable over time.

In theoretical determination (see 1.2.4) of the transmembrane part a 23 amino acid long alpha-helix was predicted; in the experiment a helical fraction of 52 % was found. The results of the simulations are much higher than this experimental value. The transmembrane helix of cluster 1 anti has a higher helicity with a percentage of 81 %. In cluster 2 para, the helical fraction lies nearer to the experimental value with 63 %. All simulations indicate that the helix folds up in the C-terminal end. Furthermore, the three polar lysines with the previous four polar amino acids are not embedded in the membrane. It can be assumed that this area interacts very well with the negatively charged lipid head groups. In this way, they form a stable anchorage in the cell membrane.

The binding patterns between the polar asparagines predicted by Lear *et al.* [44] could be found in this work as a structural binding motif for assembling of the peptide dimer. In cluster 2 para and cluster 1 anti up to five hydrogen bonds could be found between the asparagines. Since these two structures assort excellently with the other specific parameters, they represent very well the performance of bsrG in the membrane as a dimer. On the basis of the MD simulations it is not possible to differentiate between them. If the behaviour of bsrG *in vivo* (see 4.1) is considered, it is unlikely that bsrG will attach antiparallel, as the highly charged C-terminal end has to be transported through the hydrophobic membrane. However, a recent publication showed that the transport of highly charged peptides via the membrane would be possible [138].

4.4.4 Conclusion

The first structural results from calculations were obtained for bsrG. As a monomer alone, it will not dive into the membrane due to its very polar side. The polar amino acids are deposited towards the polar solvent and the hydrophobic part of the membrane towards the lipid surface.

Assembly studies of the dimer and the tetramer showed, that a higher oligomerisation, in which the polar side forms a polar interface, is stable in the membrane environment. For the dimer in the antiparallel and parallel state, stable dimer structures could be found in each case, where the polar asparagines form hydrogen bonds and thus stabilise the dimerisation. This supports the thesis that bsrG forms a *hydrogen bond zipper* as dimer. In addition, the upright position of the dimers in the membrane corresponds to the experimental results. The MD simulations do not show any preferred arrangement of bsrG as parallel or antiparallel dimer.

The coarse-grained studies of the bsrG tetramers showed that the occurrence of an antiparallel tetramer in the membrane can be excluded. Assembling occurs for the parallel tetramers, however, the high mobility of the peptides prevents identified specific structures.

Extended studies with longer simulation times could indicate which structure of the bsrG dimer cluster 2 para and cluster 1 anti is more stable. The measured solid-state NMR parameters can also be compared with the simulations.

5 siRNA "Traffic Lights": Fluorescent siRNA with Dual Colour Readout

5.1 Introduction

The concept of siRNA (small interfering RNA) traffic lights was designed in the group of Hans-Achim Wagenknecht (KIT). Here, two fluorescent dyes, which can interact as base surrogates or 2'-modifications, are also energy transfer pairs and are used to monitor cellular processes such as siRNA transport in real-time.

The characterisation and synthesis of three different RNA traffic light constructs was one of the main topics of the RTG 2039 *Molecular architectures for fluorescent cell imaging*. Optical spectroscopy (group of Hans-Achim Wagenknecht (KIT)), fluorescence lifetime measurements (Group of Ulrich Nienhaus (KIT)) as well as fluorescence imaging (groups of Hans-Achim Wagenknecht (KIT) and Ute Schepers (KIT)) were used in combination with MD simulations (shown here) to shed light on the optical behaviour of the three RNA constructs. Parts of this chapter have been submitted [140].

5.2 Biological Background

In 2006, the Nobel Prize in Physiology or Medicine was awarded to Andrew Fire and Greg Mello for the discovery of ribonucleic acid (RNA) interference. RNA interference (RNAi) is found in most eukaryotes and is a conserved self-defence mechanism which leads over several biochemical processes to sequence-precise suppression of gene expression. This mechanism, whereby RNA molecules can regulate the activity of genes by degradation of the complementary messenger RNA (mRNA), is one of the most powerful tools in modern biology to study the effect of gene-knockout directly.

The academic research of RNA interference (RNAi) has increased, taking big steps since its first characterisation in the nematode *Caenorhabditis elegans* (*C. elegans*) [141]. Shortly afterwards, it was shown that RNAi occurs in response to the introduction of double-stranded small interfering RNAs (siRNAs) into a cell. siRNA consists of ~21 nucleotides (nt) in mammalian cells, can induce the specific post-transcriptional degradation of messenger RNA with a distinct complementary sequence and leads to sequence-specific gene silencing [142]. Recently, advances have been made in understanding the mechanism by which RNAi regulates the gene expression.

There is a glimmer of hope that RNAi can be used for therapeutic intervention of human diseases. The benefit of using RNAi lies in its ability to a specific knock-down of the gene expression of disease-causing genes [143].

Therapeutic potential is reported for treatment of diseases such as autoimmune diseases, human immunodeficiency virus (HIV), ocular and neurodegenerative diseases as well as cancer [144].

Clinical trials have started for HIV, hepatitis B virus (HBV), hepatitis C virus (HCV) and Huntington disease [145]. Initial descriptions of RNAi activity from siRNA therapy in humans have been evaluated [146].

Despite the consistent enthusiasm of fundamental research results the clinical trails (mainly in cancer treatment) have faced a couple of obstacles for *in vivo* gene silencing [139]:

- Delivery obstacles *in vivo*: Specific delivery of siRNA into cells and to the place of action is still a challenge. Biological barriers like endogenous ribonuclease, kidney filtration, cell membranes and survival from reticuloendothelial clearance are merely a few examples. In addition, not only the penetration into the tumour is important, but also the internal tumour distribution.

- Choice of RNAi target: An ideal target should be thermodynamically stable, exhibit a bio marker that can predict biological response, cause tumour regression and express preferentially in tumours. To date, most of the targets fall short of these criteria [147].

- Safety issues: Toxic side effects may arise due to non-specific uptake by immune and endothelial cells or non-specific accumulation in liver and spleen.

To summarise, the most severe bottleneck of clinical application of siRNA systems is the development of synthetic siRNA for delivery systems *in vivo*. Therefore, tracking of the siRNA pathway and its cellular processing is crucial.

Nowadays, fluorescence imaging in real time is an optical technique suitable to visualise siRNA location. Several new techniques have been developed in the last years [148]. One of these methods, denominated FISH (fluorescence *in situ* hybridisation), is the most significant approach. With this method RNA or DNA are hybridised to labelled short nucleic acid probes in fixed cells [149]. Therefore, one disadvantage is the lack of dynamic information on the RNA movement. Another drawback is the usage of a single fluorophore: signals from an intact fluorophore and from one, which occurs as a consequence of RNA degradation, cannot be distinguished. Besides, undesired fluorescence quenching by cellular compounds can also happen.

The conjugation of siRNA to two dyes, which form donor-acceptor pairs for Förster Resonance Energy Transfer (FRET), is an appropriate choice. FRET is a non-radiative process where an excited state donor dye transfers energy to a nearby ground state acceptor dye through long-range dipole-dipole interactions. For that purpose the dyes have to be in a close proximity, and the efficiency of FRET depends strongly on the distance [150]. This method has the advantage that the fluorescence is mainly dependent on the dye-dye distance quenched, and two different emission colours *in vivo* can be detected [151, 152].

In recent years, several research groups have developed and synthesised FRET-dye pairs for siRNA imaging. These include, among others, base surrogates with attached dyes such as pyrrolo-cytosine [153] and thiazole orange [154].

In the next section, the RNA-labelling concept called "RNA traffic lights" will be highlighted which was first proposed in the group of Hans-Achim Wagenknecht (KIT) [155].

5.3 "Traffic Lights" - Concept and Dyes

The "traffic lights" concept allows for real-time monitoring of RNA or DNA *in vivo* by following the fluorescence colour change from red to green. As mentioned before (see section 5.2), this approach is based on FRET which is a powerful tool to detect the intracellular pathway of siRNA. The application of FRET dyes might overcome the disadvantages of other optical methods like FISH. The main advantage, however, is the wave-length shifting readout with focus on two distinct emission colours (here red and green) and the ability to track the *in vivo* pathway of the siRNA construct [155].

The dual colour readout of the concept allows a "two-state" tracking of the siRNA. Both dyes are placed in a diagonal interstrand orientation into a part of the siRNA which is not the recognition part to ensure proper gene silencing. The donor dye is placed to the sense strand, the acceptor to the antisense strand (the antisense strand is the one which binds the target RNA and leads to gene silencing). If the double strand siRNA is intact, FRET can occur, and a red fluorescence readout follows. If the siRNA dissociates or binds to target mRNA, it shows a green fluorescence readout. The concept is visualised in figure 5.1.

Primarily, the "traffic-lights" concept was designed for DNA with the fluorescent dyes thiazole orange and thiazole red. Those were not covalently bound to one of the DNA bases forming strands, but assimilated as artificial base surrogates. For energy transfer, these dyes form an interstrand pair which is exposed to the solvent [156]. To improve the optical properties, such as brightness (defined as extinction coefficient multiplied by the quantum yield), photostability and fluorescence intensity (in the presence of nucleic acids), an alternative class of dyes was examined. The cyanine-styryl dyes seem promising: a whole set of new photostable dyes was developed with a broad range of colour spectrum (blue, green, yellow and red) and enhanced photostability in comparison with the thiazole dyes [157, 158].
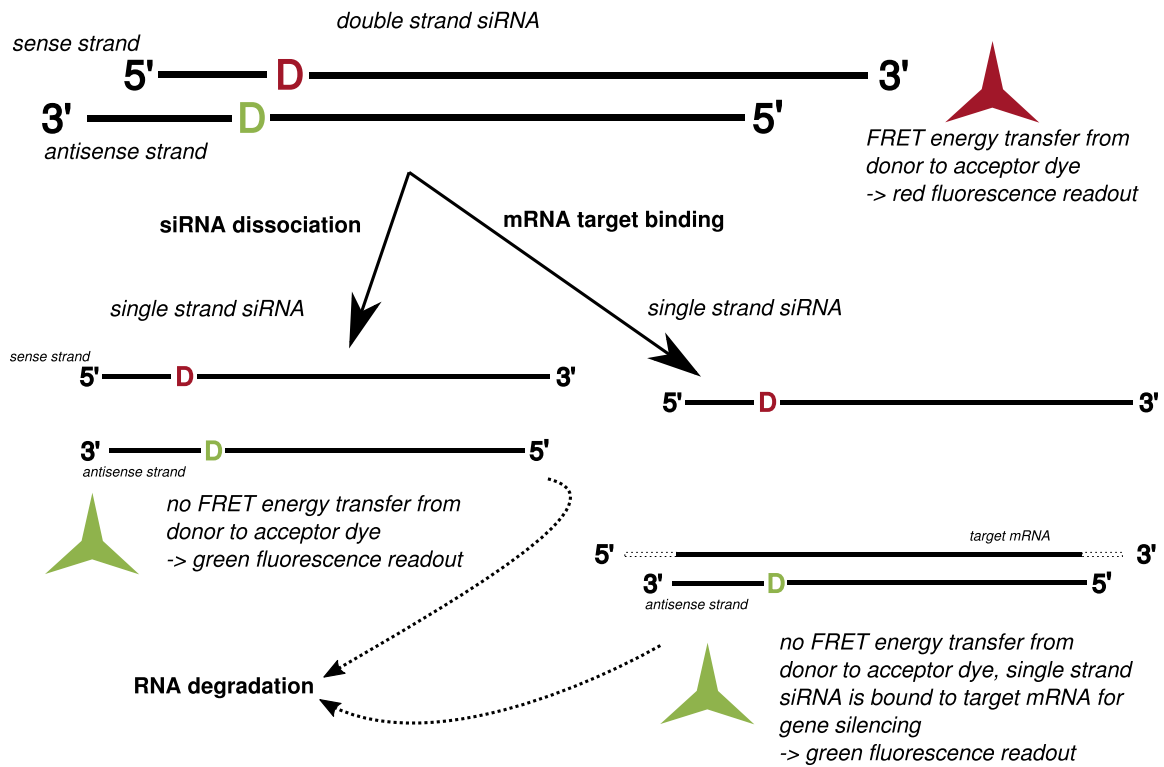
Figure 5.1: Concept of siRNA "traffic lights". Two fluorescent dyes, which can interact as FRET pair, are attached in an interstrand orientation to double strand (ds) siRNA. The donor dye (green readout) is attached to the antisense strand and the acceptor dye (red fluorescence) to the sense strand. Both are located in a region which does not interfere with the gene silencing mechanism. To monitor the siRNA integrity inside the cells the fluorescent colour change from red to green can be detected. If the siRNA is intact as double strand, red fluorescence can be detected. Otherwise, if the ds-siRNA is degraded or the sense strand binds to the target mRNA, green fluorescence occurs. By tracking the red to green readout the siRNA integrity *in vivo* and in real-time can be visualised.

5.4 siRNA "Traffic Lights" - Aim

Three diverse RNA constructs were designed to find the construct with the best fluorescence readout. The sequence of the chosen siRNA double-strand is shown in figure 5.2.

sense strand
5' G-C-U-A-D-A-C-G-A-C-C-C-U-G-A-A-G-U-U-C-A-U 3'
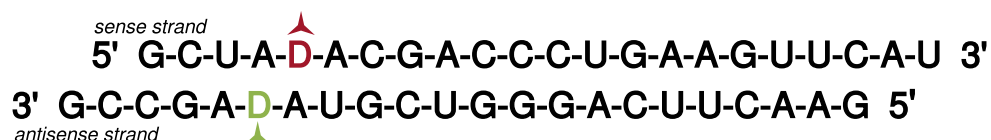3' G-C-C-G-A-D-A-U-G-C-U-G-G-G-A-C-U-U-C-A-A-G 5'
antisense strand

Figure 5.2: Sequence of the double strand siRNA. The interstrand position of the fluorescence dyes is tagged with a capital D in the colour of the respective fluorescence emittance. RNA contains 4 different bases. Two purine bases, Adenine (A) and Guanine (G), and two pyrimidine bases, Cytosine (C) and Uracil (U).

From the broad range of new photostable cyanine-styryl dyes two promising candidates for siRNA tracking were chosen. The choice fell on a green emitting donor dye and a red emitting acceptor dye (shown in figure 5.3).
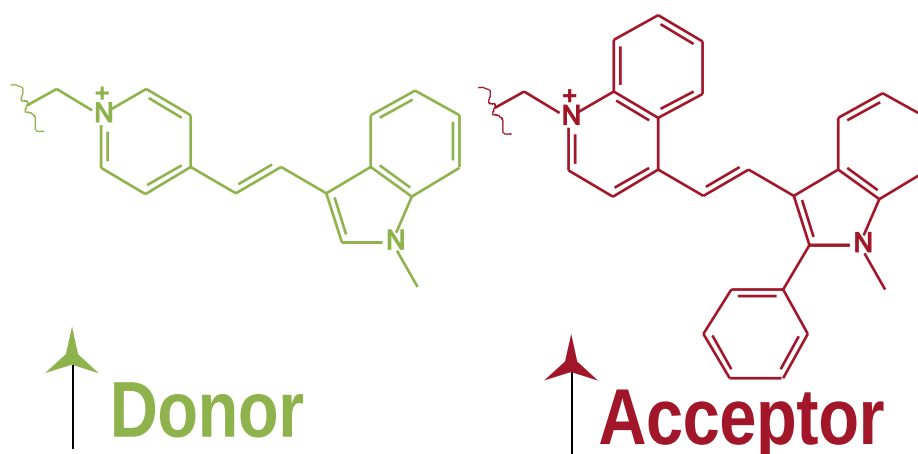

Figure 5.3: Structure of the cyanine-styryl dyes which were used as promising pair for FRET siRNA imaging. The donor dye shows a green readout, the acceptor dye a red fluorescence.

One of the constructs contains the dyes as base surrogates with (S)-3-amino-1,2-propanediol as a substitute for the ribofuranoside. This replacement of the 2'-deoxy-ribofuranosides had to be compensated by a linker which lacks the hydrolytically labile glycosidic bond and provides sufficient chemical stability (see figure 5.4) [140].

For the other two constructs the dyes are applied as 2'-modifications. As a consequence of this modification, a remarkable increase of the ob-
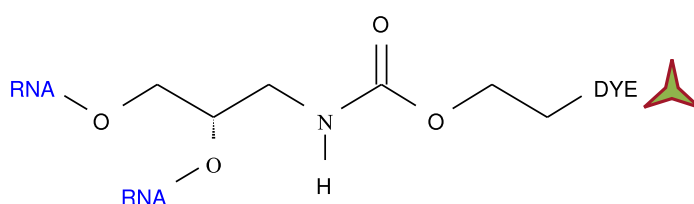
Figure 5.4: Structure of the dye attachment in the first construct. The ribofuranoside is replaced by (S)-3-amino-1,2-propanediol.

tained emission colour contrast could be established [159, 160]. The dyes were covalently bound after the synthesis of the RNA strand via a copper catalysed "click"-reaction (cyclo-addition) between the 2'-propargylated oligonucleotides and the corresponding azide-functionalized dyes [161]. One of the 2'-modifications uses the natural ribofuranoside (construct 2 see figure 5.5 A), and the other one uses an arabino-configured furanoside (construct 3 B). The 2'-propargylated RNA uracil building block is shown in figure 5.5 (B) in comparison to the natural-like ribofuranoside uracil (A).

The inversion of the 2'OH-group in the arabino-configuration suggests an A-RNA double helix more similar to a DNA-like configuration [162].
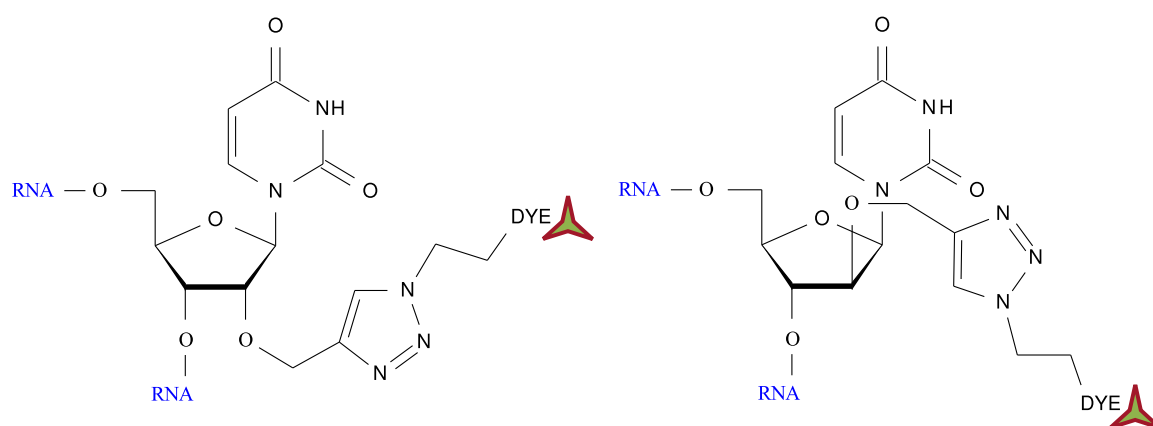
Figure 5.5: Structures of the constructs 2 and 3, as uracil, with the dyes covalently attached via 2'-modifications. Construct 2 is based on the natural ribofuranoside. The arabino-configured furanoside in construct 3 might affect the overall structural properties of the siRNA.

5.4.1 Results from Optical and Cell Imaging Experiments

Steady state fluorescence measurements of the constructs *in vitro* showed that the arabino- and the base-surrogate construct have a much higher

FRET-efficiency (93 % respectively 90 %) than the ribo-configured construct with 69 %. Moreover, the siRNA construct with arabino-configured dyes has a melting temperature decreased by ~2 °C.

In addition, live cell imaging experiments show that a biological application is feasible. All three siRNA constructs were detected in endosomal compartments after transfection to HeLa cells (Human cervix carcinoma cells) and imaging via live cell imaging by confocal microscopy. The *in vitro* found difference of the FRET efficiency can also be detected by observing the "red-to-green" fluorescence read-out of the constructs *in vivo*. For the constructs 2 and 3 only a little amount of green fluorescence in the donor channel can be detected which means that FRET is possible, the dyes are nearby, and the acceptor is excited. Finally, the arabino-configured one is the most suitable of these constructs because the high FRET efficiency is stable over several hours after transfection [140].

The experiments have shown that all three siRNA constructs can be used as siRNA tracking tools via optical detection of the wave-length shifting fluorescence. The most promising solution is the non-natural arabino-configured construct which has the highest FRET-efficiency.

However, this large difference in FRET-efficiency is not explainable by experiment.

Long classical MD-simulations are able to solve this question and find structural causes in the time-dependent behaviour of the siRNA constructs with the fluorescent dyes. Because of the preserved base-pairing in constructs 2 and 3, which is more natural-like, and the synthetically time saving postsynthetic "click"-reactions, compared to the preparative advantage of base-surrogates synthesis, only simulations for these two constructs were required to shed light on the striking difference in FRET-efficiency.

5.5 Methods

The MD-simulations in this part were executed with the *GROMACS 5.0* simulation package [94, 105–110]. Quantum mechanical calculations were performed with the *Gaussian09* quantum chemistry package [163].

5.5.1 Parametrisation of the Dyes and the Linker

The two dyes (donor and acceptor) and the linker, which connects the cyanine-styryl dyes and the RNA building block, were treated as individual molecules for parametrisation. To avoid artificial behaviour during the calculations in vacuum, protecting groups were attached to the cross unbound ends. For the dye ends ethyl groups were used, and for the linker a methoxy group on one side and another ethyl group on the other side (structures shown in figure 5.6) was attached.
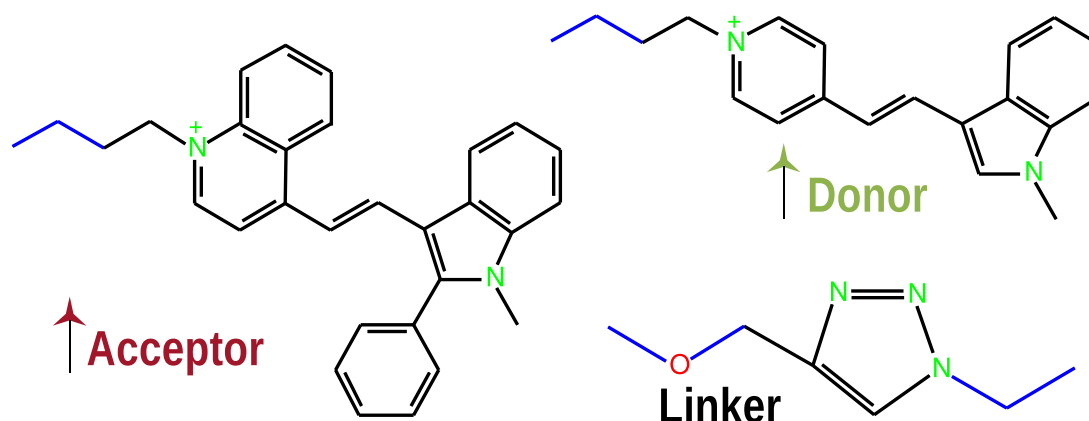

Figure 5.6: Structures of the dyes and the linker used for parametrisation. The protecting groups are marked in blue. Hydrogen atoms are not shown explicitly. The heterogeneous nitrogen atoms are marked in green, the oxygen in red.

The molecules were constructed using *xleap* from the AmberTools package [111]. Then, geometries were optimised in vacuum with DFT (Density functional theory) using the hybrid functional B3LYP [164, 165] and the basis set 6-31G* [166]. Atomic charges were calculated from the electrostatic potential surface (ESP) with the Hartree-Fock method [167] using the basis set 6-31G*. Afterwards, the final charges were obtained with *antechamber* [168] applying the Restrained Electrostatic Potential methodology (RESP) [169]. For the calculation of the charges, the charges for the capping groups of the dyes were set to zero. For the linker the ethyl capping group was set

to zero, and the methoxy capping group was set to 0.1953 which depicts the influence of the sugar to the linker. As force field parameters, the General Amber force field (*GAFF*) was taken. Missing improper and dihedral angles were adopted from available parameters which match the central atoms.

5.5.2 Parametrisation of a Missing Dihedral Angle

Due to the significant effects on the structural behaviour of the acceptor dye it was necessary to parametrise one dihedral angle which was *lacking* in the GAFF force field. As shown in figure 5.7, only the rings of the dye themselves were taken to calculate the dihedral between the two rings.
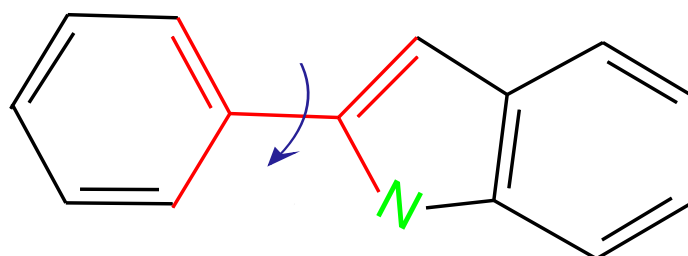

Figure 5.7: Structure of the used segment of the acceptor dye for parametrisation of the dihedral (marked in red) between the two rings. The heterogeneous nitrogen atom is marked in green. The arrow (blue) illustrates the rotation between the rings which will be scanned to obtain the potential surface.

First, the potential energy curve along the torsional angle was calculated quantum mechanically with *Gaussian09* using the Hartree-Fock method with the basis set TZVP [170, 171]. To obtain the profile, the dihedral was rotated from 0 to 180 degrees in steps of 10 degrees while the rest of the structure was minimised. Astonishingly, the minimum for the PES was around 25 degrees, so the ground structure of the dye is not flat. This obtained PES served as reference.

Afterwards, the PES was also calculated on the molecular mechanical force field level. The parameters of the remaining structure were converted with the *ambconv* software. For the PES calculation all potential terms, in which the wanted torsional angle is involved, were set to zero. The profile was received from single-point energy calculations of structures which were restrained to get the rotation by steps of 10 degrees. The obtained MM profile was subtracted from the QM profile. Then, the torsional potential

was fitted to the functional form of the dihedral angle in the GAFF force field with the *XMGRACE* software.

$$y = A_0 + A_1 \cdot \cos\left(2 \cdot x \cdot \frac{\pi}{180}\right) + A_2 \cdot \cos\left(4 \cdot x \cdot \frac{\pi}{180}\right)$$

Since there are four potentials, which describe the rotation of the two rings to each other, the A_1 and A_2 parameters were divided by four. This results in the following parameters: $A_0 = -1.5610^{-6}$ kJ/mol, $A_1 = 0.20$ kJ/mol and $A_2 = -3.20$ kJ/mol. Finally, the PES Scan was repeated with the newly parametrised dihedral angle on force field level. In figure 5.8 the result of the scan is shown in comparison with the QM scan. The newly parametrised dihedral is reasonably accurate.
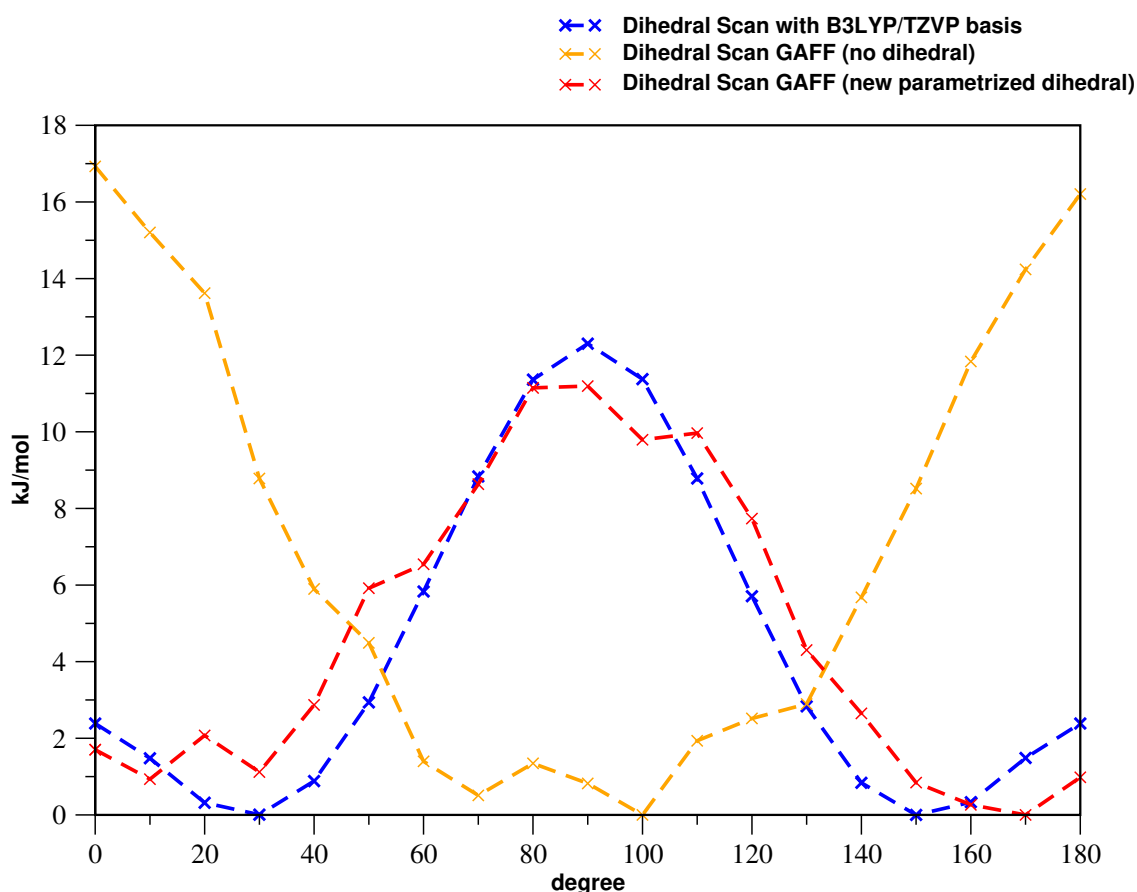

Figure 5.8: Potential energy surface comparison between the quantum chemical scan on Hartree-Fock/TZVP basis, the scan with the dihedral parameters set to zero on basis of the GAFF force field and the newly parametrised dihedral angle introduced to the GAFF force field.

5.5.3 RNA Building

The double-stranded RNA species with the sequence shown in figure 5.2 was constructed with the web tool *make_na* [172] which is based upon the NAB (Nucleic Acid Builder) [173]. As helix type the canonical A-RNA conformation was chosen, and the RNA was described with the Amber ff99bsc0+parmχOL3 parameter set with the recently added corrections for torsions [58–60, 174, 175]. Then, the arabino-configured uracil building block was built upon the basis of the epimeric ribose described in the force field. The spatial orientation of the 2'-OH group was manually converted from *down*, the natural ribose configuration, to the *up* configuration. No other values were changed. Assembly of the helical RNA with the non-natural ribose/arabinose uracil with the linker and the dyes were done manually with the *antechamber* utility. The dyes were placed in a manner that they could fluctuate freely with no sterical hindrance. Starting structures are shown in figure 5.9.

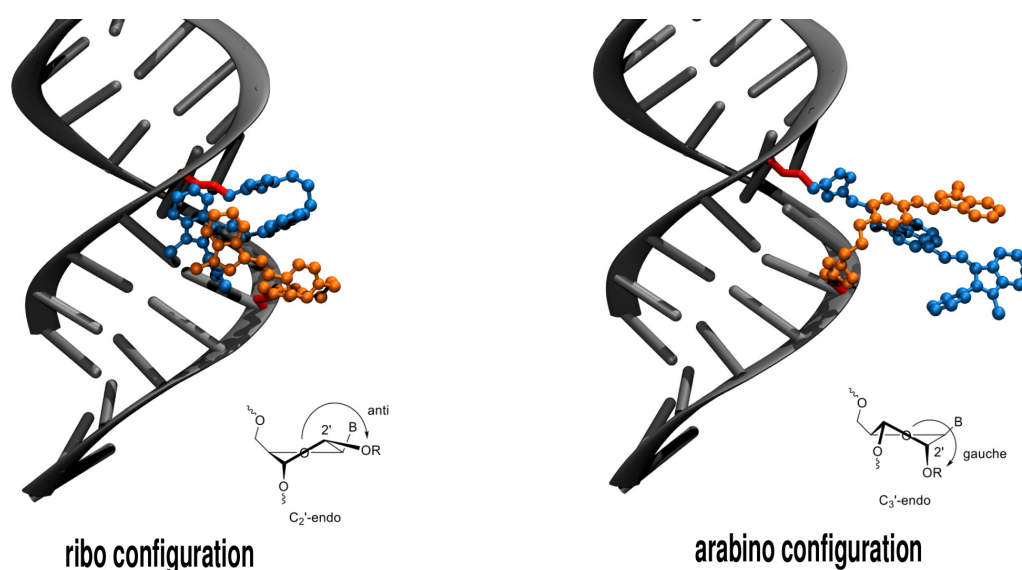Figure 5.9: Starting structure of the double stranded siRNA with the donor and the acceptor dye attached to the uracil building blocks via linker.

5.5.4 MD Simulations

The double stranded RNA species with the fluorescent dyes (two systems, one for arabino and one for the ribo-configuration) were placed in the middle of a rhombic dodecahedron box with a box size that ensures a distance of at

least 1.3 nm between the RNA molecule and any box face. Then, the box was filled with 24,650 TIP3P [113] water molecules. To compensate the negative charge of RNA, 40 sodium ions (Cheatham and Young parameter set [118]) were added. Therefore, 40 water molecules were excluded. In total, the number of atoms was 75,505 atoms. All hydrogen atoms in the RNA-species were considered as virtual sites, for which no equations of motion were solved, rather their new positions were obtained with a lever rule, and the forces obtained on these atoms were distributed on neighbouring atoms. The bond lengths were constrained to their respective equilibrium values. Electrostatics were treated with the particle-mesh-Ewald algorithm [77] with a direct-space cut-off of 1.0 nm and the group cut-off scheme. The Lennard-Jones interactions were cut off at 1.0 nm. The equations of motion were integrated by means of the leap-frog algorithm [119] with a time step of 4 fs. All bond lengths were constrained with p-LINCS [67]. The simulations were performed unrestrained, both for the ribo-configured and the arabino configured fluorescent FRET dyes.

5.5.4.1 Simulation protocol

First, an energy minimisation using the steepest-descent algorithm was performed. Afterwards, the system was equilibrated at a constant volume in the NVT (N: number of atoms, V: volume and T: temperature) ensemble. Random initial velocities from the Maxwell-Boltzmann distribution at 10 K were assigned, followed by 20 ns ($5 \cdot 10^6$ steps) of constant-volume heating to 300 K using the Nosé-Hoover thermostat [87] with a coupling time of 2 ps. Two coupling groups were considered, containing RNA with the dyes, and water together with ions. Subsequently, 20 ns equilibration at constant pressure in the NPT (N: number of atoms, P: pressure and T: temperature) ensemble at 300 K and 1 bar was executed. Like in the NVT equilibration two separate Nosé-Hoover thermostats with a coupling time of 2 ps were considered. For regulation of the pressure the Parinello-Rahman barostat [89] was applied with a coupling time of 2 ps. Lastly, for the production simulation 4000 ns of NPT simulation ($1 \cdot 10^9$ steps) were performed at 300 K using the Nosé-Hoover thermostat with a coupling time of 2 ps and 1 bar using the Parinello-Rahman barostat with a coupling time of 2 ps.

5.5.5 Analysis

The visualisation and graphical analysis was performed with the tools *PyMol 1.8.x* [176], *Grace* [122] and *VMD 1.9.2* [123].

For visualisation the trajectory was processed in several steps with the GROMACS tool *trjconv* to account for periodic boundary conditions (PBC). The GROMACS tool *rmsf* was used to calculate the root mean square fluctuations RMSF (for all heavy atoms) of the RNA structure with the added dyes to obtain information about the flexibility of the constructs. In this case, the starting structure of the production simulation obtained by the equilibration (energy minimisation and equilibration of NVT and NPT) was selected as reference.

To shed light on the different conformational flexibility of the arabino and ribo-configured dyes, the *b-factor* or *temperature factor* was calculated from the RMSFs. The *b-factor* is a measure of thermal-dependent atomic displacement and thus of the flexibility of the structure.

For the calculation of the root mean square deviation RMSD the GROMACS tool *rms* was used. The RMSD describes the structural stability in comparison with a reference structure over the simulation time. For the arabino and ribo-configured dyes this was calculated by computing the root mean square deviation of these in comparison to the starting structure. A least square fit to the complete RNA with dyes was performed before.

Furthermore, the distance between several parts of the constructs over time was computed with the GROMACS tool *distance*, and the fluctuation of dihedrals over time were calculated with the GROMACS tool *angle*.

5.6 Results

Regarding the time-consuming 4 µs long free classical MD-simulations of the arabino- and ribo-configured dyes attached to RNA a clear difference between the simulations can be detected.

5.6.1 Ribo-configured Dye Simulation

The ribo-configured dyes retain in a stable π-stacked orientation similar to the initial structure. This complex remains located near the minor groove. Even over the whole simulation time the RNA strands were more or less *rigid* with perfectly stacked base pairs. Only the overlapping ends of the RNA strands fluctuate.

The stacked dye complex itself fluctuates during the simulation merely by conveniently displacing the dyes in relation to each other. The dyes interact at the beginning, while during the simulation this reduces to only parts of the dyes interacting. Importantly, the dyes never completely lose the contact.

Representative patterns of interaction are shown in figure 5.10.

5.6.2 Arabino-configured Dye Simulation

The arabino-configured dyes undergo large structural rearrangements over the time of a few hundred nanoseconds in the simulation. A loose complex forms for brief periods of time on several occasions. Instead of stacking they interact with nucleobase pairs. The donor is more flexible than the acceptor and interacts with a broader part of the minor groove in the 5'-direction.

The acceptor dye interacts during the first 400 ns in a loose complex with the donor, and both dyes were in close contact to the RNA interface. For the following 2 µs the acceptor interacts with out-flipping bases from the RNA strand which unwinds at this distinct positions. Importantly, the acceptor also interacts with the nucleotide connected to the donor. At the same time, the donor interacts with the minor groove.

Interestingly, several time points can be detected where the acceptor dye unwinds the RNA helix and intercalates into the RNA and undergoes π-stacking.

Figure 5.10: Representative patterns of the ribo-configured dyes at several points in time of the 4 μs classical MD simulation. The donor is coloured in orange and the acceptor in blue. RNA strands are more or less perfectly stacked. The dyes retain in a stable π-stacked orientation and remain located near the minor groove. Fluctuations occur mainly in the stacking interface where parts of the dye shift against each other.
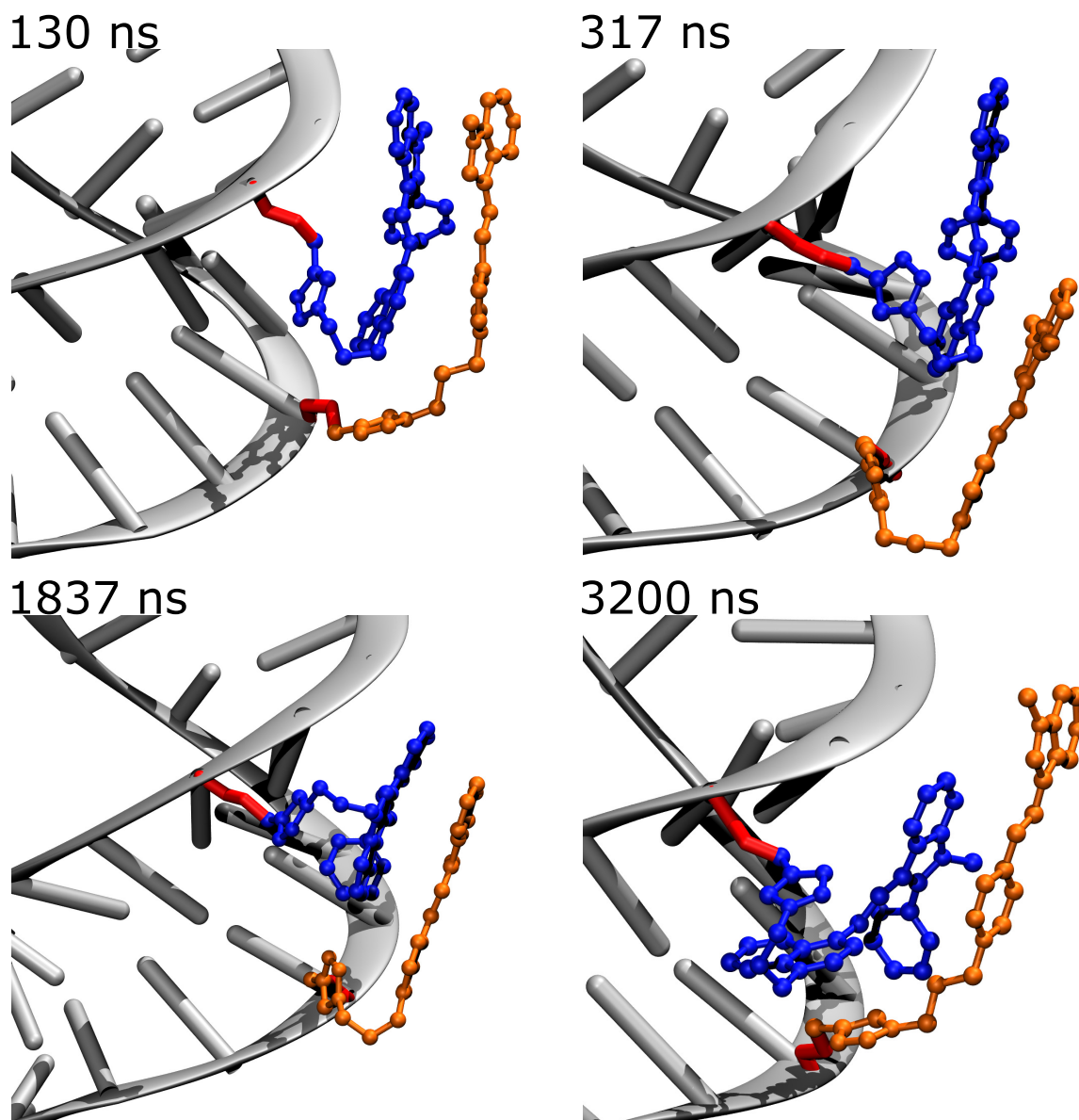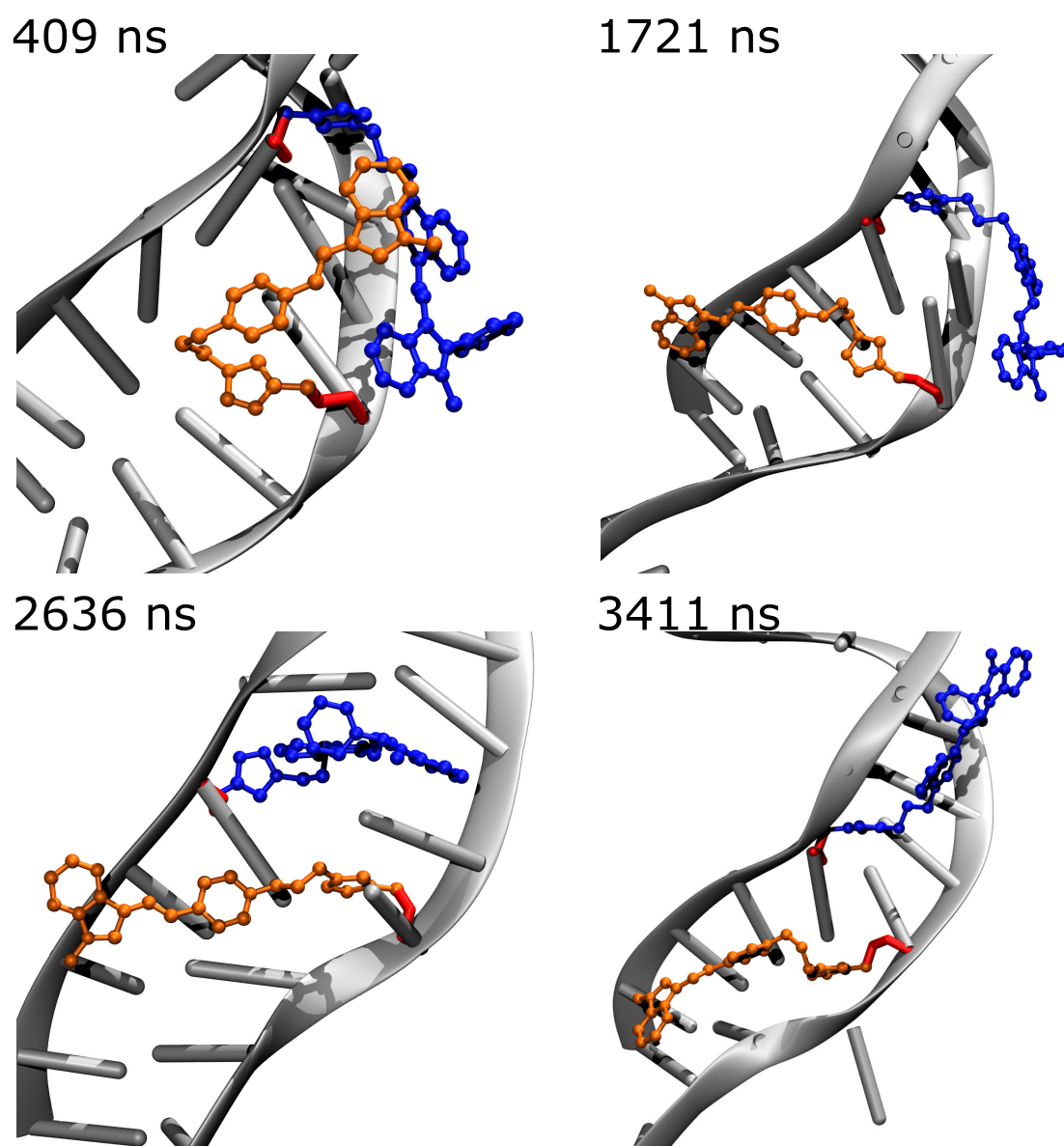
Figure 5.11: Representative patterns of the arabino-configured dyes at several points of time from the 4 μs classical MD simulation. Overall, the dyes were more flexible and moved to more different interaction patterns. The snapshots represent four main types of such interactions. Donor and acceptor loosely form a complex near the minor groove (409 ns). Also, they rotate in opposite directions and interact with a broader range of the minor groove (3411 ns). Furthermore, the RNA strands are more flexible with reversible unwinding and out-flipping of nucleobases (1721 ns). The acceptor dye interacts with such rotated nucleobases. Interestingly, one of these nucleobases is the donor-base. Additionally, the acceptor dye can intercalate into the α-helical structure (2636 ns). It has to be noted that the acceptor dye undergoes new interaction modes, but the donor has a broader flexibility interacting with the minor groove downwards. The donor is coloured in orange and the acceptor in blue.

At the end of the simulation the dyes are rotated apart from each other and interact with the minor groove.

All in all, the RNA strands were highly flexible, not only the overlapping ends. Reversible unwinding and flipping-out of nucleobases was observed. Structural differences became apparent not only in the range of motion, but also in the character of interaction motifs. In figure 5.11 representative patterns for the distinct interactions of the arabino-configured dyes over time are illustrated.

To quantify the visual observations, distinct specifications were chosen to describe the different structural ensemble dynamics of the ribo- and arabino-configured dyes.

5.6.3 Calculation of the Root Mean Square Deviation

One standard tool is the calculation of the root mean square deviation (RMSD). This is the average distance of the heavy atoms (except hydrogens) and a quantitative measure of similarity. In this case the RMSD of the dyes in respect to their starting structures were calculated (shown in figure 5.12).

Ribo-configured dyes

It can be seen that the ribo-configured dyes underwent two stable binding patterns with respect to the starting complex. The first one is stable for several nanoseconds, whereas the second is stable until the simulation ends. The RMSD for the donor is less fluctuating than for the acceptor. To conclude, the RMSD results agree with the observation that a stable π-stacked complex is formed.

Arabino-configured dyes

By contrast, the RMSD calculations for the arabino-configured dyes show that during the entire simulation the positions of the dyes fluctuate. Also the acceptor dye is much more floppy then the donor. Moreover, the RMSD values are wider spread with larger structural displacement of the dyes during the simulation. The RMSD of the arabino-configured dyes corresponds with the proposal of a loosely stacked complex.

5.6.4 Calculation of the Root Mean Square Fluctuation

Another tool is the calculation of the root mean square fluctuations (RMSF). Thereby, the flexibility of structures can be shown.
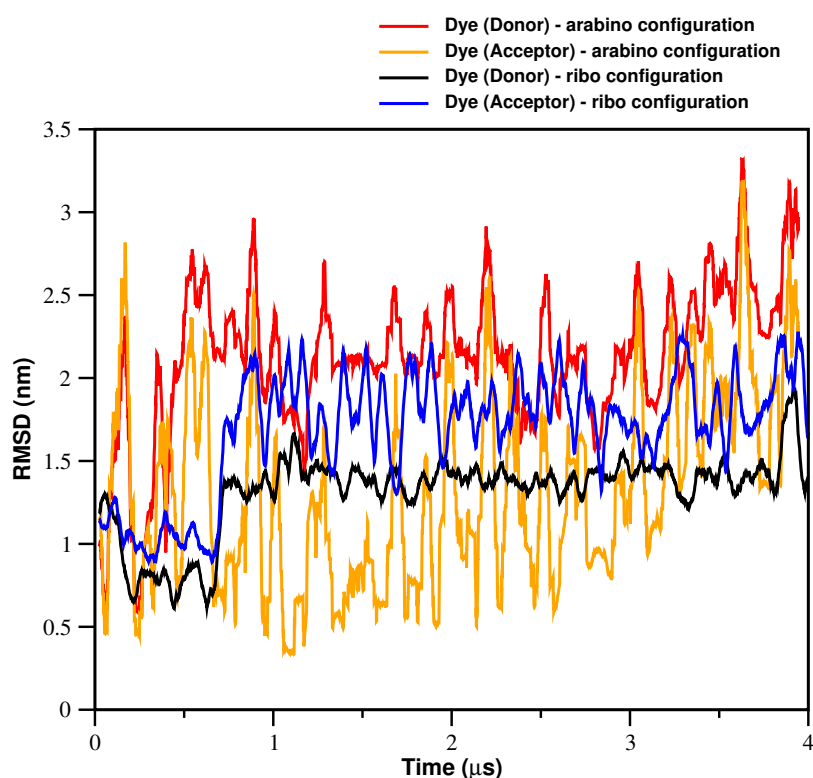
Figure 5.12: Root mean square deviation (RMSD) of the donor and acceptor dyes covalently attached to the ribo- and arabino-configured RNA. The RMSD is calculated with respect to the starting structure of the dyes.

The calculated RMSFs values can be converted into *b-factor* values which are known from crystallography as a displacement parameter. These can be used as a measure of mobility.

The calculation of the RMSF and the conversion to *b-factors* afterwards were used to shed light on the flexibility of the dyes themselves. In figure 5.13 the ribo- and arabino-configured dyes in their shape at the last point in simulation time are shown. In the upper half of the figure the RMSF are colour-coded: blue for the lowest value over green and orange to red for the highest value.

Ribo-configured dyes

The ribo-configured dyes are very rigid mainly for the linker-part of the acceptor dye and small parts of the rings, while there is hardly any fluctuation. This affirms the assumption that the dyes themselves were very rigid.

Arabino-configured dyes

The picture is different for the arabino dyes. The donor dye is very flexible compared to the acceptor, in particular the linker. Especially, the acceptor dye itself is more or less rigid, and only the linker-part is flexible.

In the upper half of figure 5.13 RMSF values of arabino- and ribo-configured dyes are compared and visualised. The arabino-configured dye is significantly more flexible than the ribo dye. Additionally, the arabino donor and the acceptor can be distinguished: the donor is more flexible than the acceptor dye, but the linker of the donor is nearly as flexible as the donor.

5.6.5 RMSF calculation of the RNA

The RMSF for the RNA strand was calculated for each residue (base of the RNA) to test the assumption of the *rigid* RNA structure in the ribo-configuration and the very flexible behaviour of the arabino-configured double-stranded RNA.

The residue number stands for a RNA base, only the dye-attached bases are composed of three residues, one for the RNA base, one for the linker and one for the dye. The RMSF was only calculated for the base. It should be noted that only the dye-attached base in the arabino-configuration is converted, all others are ribo-configured (naturally). The results are shown in figure 5.14.

For better understanding, it should be noticed that residue numbers 1, 2, 25 and 26 are the overlapping non-stacking base pairs. The dyes have residue numbers 6-8 (donor) and 42-44 (acceptor). So the double-strand helix starts from residue 3 interacting with residue 48, residue 24 with residue 27 and so on.
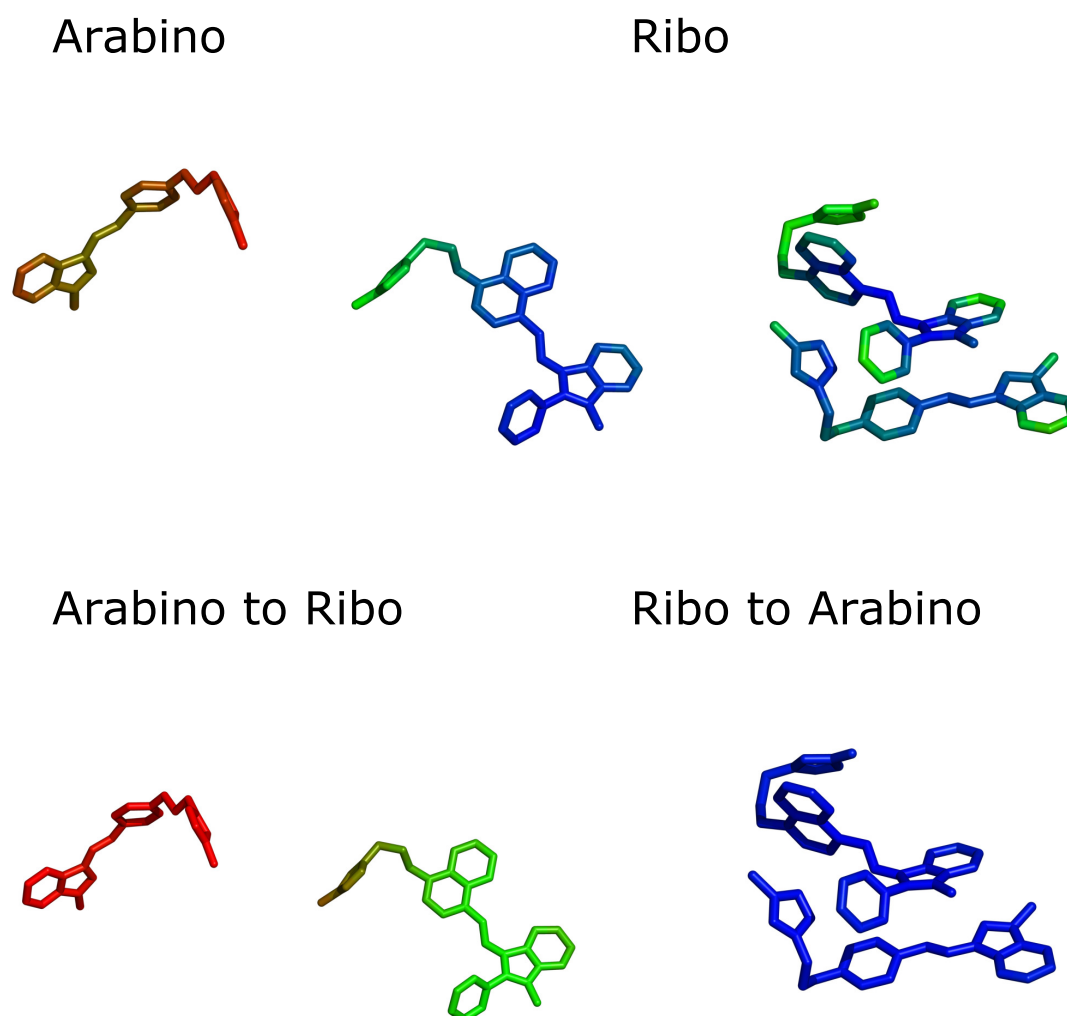
Figure 5.13: Representation of the structural flexibility of the dyes. For this purpose the RMSF values for the ribo- and arabino dyes with linker were calculated, converted to *b-factors* represented by a colour range. This colour range varies from blue for the lowest values to red for the highest. In the upper half of the figure the segmentation of the colour range is limited to the dye itself, whereas in the lower half the colour range is assigned to the values of arabino and ribo RMSF. However, the colour spectrum only serves as a tool to show the relative contrast between the flexibility. From this analysis it can be concluded that the ribo-dyes are less flexible than the arabino-dyes. Additionally, the more flexible arabino-donor and the less flexible arabino-acceptor can be distinguished. The observation of the flexibility of the dyes themselves indicates that the dye structure is *rigid* and more or less flat, whereas the linker is the flexible part.

Ribo-configured dyes

The RNA helix is very stable for the ribo-configuration. Only the overlapping ends and of them the part to the dye-attached ones are more flexible.

Arabino-configured dyes

By contrast, the arabino-configured representatives are significantly more flexible, not only at the ends. Interestingly, the strand with the donor-dye is more flexible compared to the acceptor. Possibly, the structural reorganisation leads to the unwinding of the double helix which is originating from the movement of the donor dye.
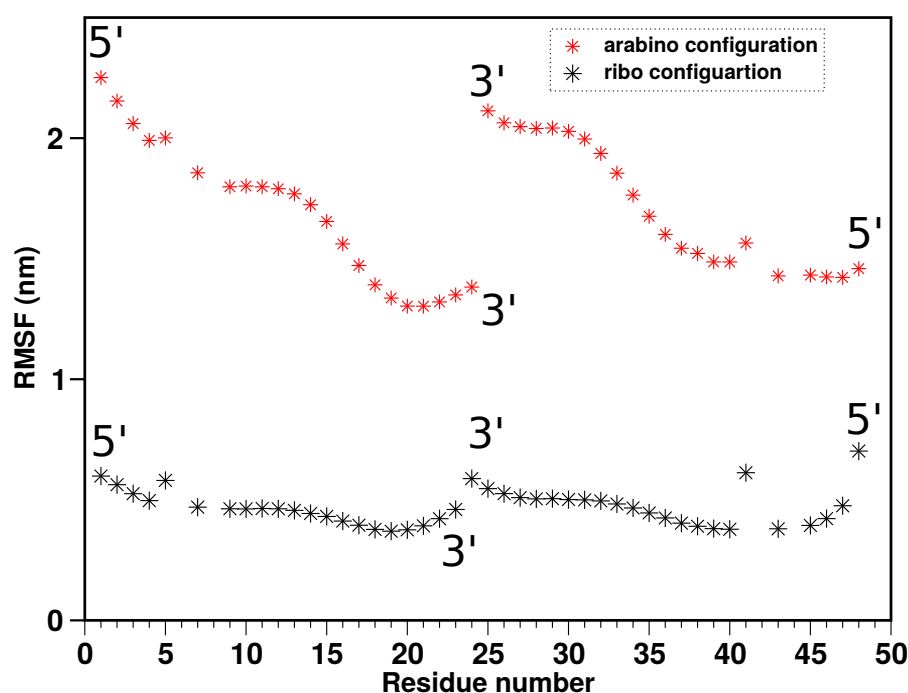
Figure 5.14: Representation of the calculated RMSF values for the RNA strands yielding values for each residue number. Residue numbers 6-8 (donor) and 42-44 (acceptor) represent the dye attached bases. Residue numbers 1, 2, 25 and 26 are the 3'-overlapping ends. The interacting residues are residue numbers 3 with 48, 24 with 27 and so on. Only the dye-attached base in the arabino-configuration is converted, all other bases are natural-like ribo-configured. The ribo-configured dye is very stable and rigid, whereas the arabino-configured dye is highly flexible. Also, the part with the attached donor is more flexible than the acceptor strand in both cases.

5.6.6 Calculation of the Distance between the Dyes

To shed light on the complex formation and the structural rearrangements, the centre-of-mass distance between the two dyes was calculated (shown in figure 5.15).

Ribo-configured dyes

The ribo-configured dyes remain in a stable complex with a short centre-of-mass distance of 0.5 nm, which is in a range for a π-stacked complex.

Arabino-configured dyes

On the other hand, the arabino-configured dyes underwent large structural rearrangements after the first 400 ns of stacking, partly with distances over 2.5 nm. Still, they were briefly approaching from time to time.
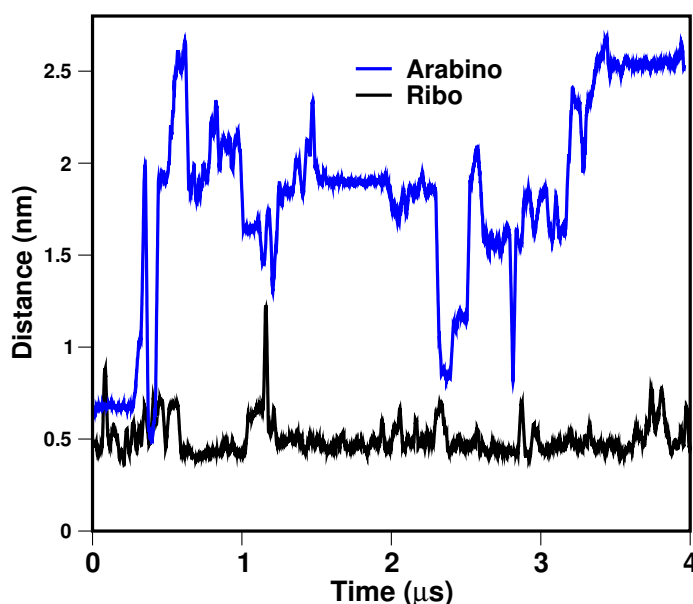
Figure 5.15: Centre-of-mass distance of the conjugated dye pairs. The short centre-of-mass distance of 0.5 nm for the ribo-conjugated dye indicates a very stable complex. In contrast, the arabino-configured dye undergoes large structural rearrangements after ca. 400 ns.

5.6.7 Torsional Angle Calculation

As depicted in figure 5.13 the most flexible part of the attached dyes is the linker. To elucidate the flexibility and to characterise the dynamics of the dyes, the torsion angle C1'-C2'-O2'-CA was monitored where CA is the first carbon atom of the dye linker.

Ribo-configured dyes

The angle of the ribo-configured dyes remains at ~82° with little fluctuation. In agreement with the RMSF calculation, the donor dye fluctuates slightly more than the acceptor dye.

Arabino-configured dyes

By contrast, the torsional angle of the arabino-configured dyes oscillates strongly, and the mean value differs largely from the ribo. In the interval from 1.5 to 2.5 µs the donor dye rotates into opposite directions. This concurs with the observation that the dyes are separated by a larger distance and that they are rotating closer to the RNA surface.

Finally, the torsional angle exhibits more fluctuations for the donor dyes than the acceptor dyes. This concurs with the results from the RMSD, RMSF and the inspection of the trajectory.

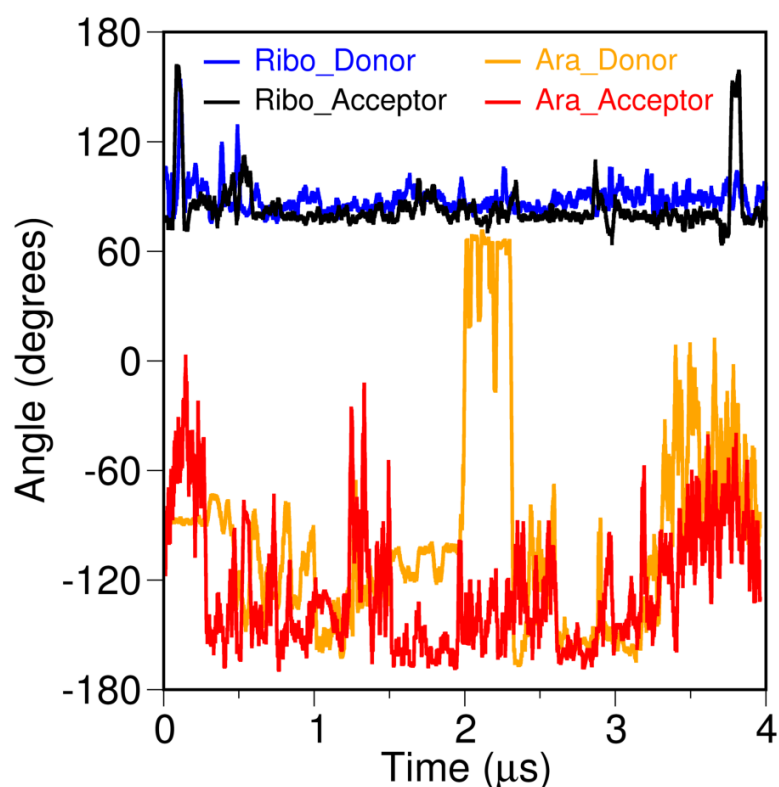Figure 5.16: Torsional angle C1'-C2'-O2'-CA, which describes the rotation of the fluorescent dyes conjugated to the ribose/arabinose ring of RNA. For the ribo-configuration the angle remains at 82° with little fluctuation, whereas in the arabino-configuration the torsional angle oscillates enormously while in the same time the donor dye rotates to opposite directions.

5.7 Discussion

Extended classical MD simulations on atomistic level provided an explanation of the striking difference between the FRET efficiencies of the ribo- and arabino-configured dyes.

Ribo- and arabino-configured dyes differ in their connection of the 2'-OH group to the sugar ring. The natural ribo-configuration is *down*, and the arabino-configuration is in epimeric form and *up*. This small change results in a large structural transition of the secondary helical structure to a more A-DNA like appearance.

The 4 µs simulations indicate that the ribo-configured dyes are in a stable π-stacked orientation over the whole simulation, whereas the arabino-configured dyes undergo large structural rearrangements.

Furthermore, for the ribo-configured dyes the RNA strands in the simulations were quite *rigid* and the RNA bases were perfectly stacked, in contrast to the RNA strands in the arabino-configured case. These strands were structural flexible. In the proximity of the dyes, nucleobases were flipping out interacting strongly with the acceptor dye, and the RNA unwound. However, the unravelling of the RNA strands in the arabino-configured dye simulation was a reversible process.

An interesting experimental result was the melting point difference of 2 °C between the RNA-dye constructs. The 2 °C higher melting temperature of the ribo-configuration could be explained by the close distance and the robust interaction between the π-stacked dyes. In contrast, the loose interactions of the arabino-configured dyes, the minor groove and distinct RNA bases may lead to a reduction of the stability and the resulting decreased melting temperature.

Why is the FRET efficiency of the two RNA constructs different?

The question of FRET efficiency difference between the ribo- with low efficiency and the arabino-configured dyes with high efficiency could be confirmed by the elasticity.

FRET needs unstacked dyes to arise. The stacked dyes in the ribo-configuration fail to undergo an efficient energy transfer. These ground state interactions interfere with energy transfer of an uncoupled dye for energy acceptance. Due to charge separation by excitation the energy transfer

is quenched. For this purpose the loosely interacting arabino-configured construct is compatible with the mentioned requirement. The unstacked uncoupled dye can serve as donor and subsequently transfer the energy to an unstacked unexcited dye, the acceptor.

To conclude, it was possible to interpret the striking difference between the FRET-efficiencies in terms of structural differences and mobility. In addition, excited state issues could be explained, in this case, with ground state structural ensemble dynamics using classical MD simulations which are less time-consuming than quantum mechanical methods.

Furthermore, the best construct for the siRNA approach and the cell imaging, namely the arabino-configured dye, was identified.

Bibliography

1. Kukol, A. *Molecular modeling of proteins: Second edition* (2014).

2. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology* **19,** 120–127 (Apr. 2009).

3. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys* **41,** 429–52 (2012).

4. Christen, M. & van Gunsteren, W. F. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *Journal of Computational Chemistry* **29,** 157–166 (Jan. 2008).

5. Van Gunsteren, W. F. *et al.* Biomolecular Modeling: Goals, Problems, Perspectives. *Angewandte Chemie International Edition* **45,** 4064–4092 (June 2006).

6. Wallin, E. & Heijne, G. V. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **7,** 1029–1038 (Dec. 2008).

7. Terstappen, G. C. & Reggiani, A. In silico research in drug discovery. *Trends in pharmacological sciences* **22,** 23–6 (Jan. 2001).

8. Goñi, F. M. The basic structure and dynamics of cell membranes: An update of the Singer–Nicolson model. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838,** 1467–1476 (June 2014).

9. Lindahl, E. & Sansom, M. S. Membrane proteins: molecular dynamics simulations. *Current Opinion in Structural Biology* **18,** 425–431 (Aug. 2008).

10. Thiel, W. & Hummer, G. Nobel 2013 Chemistry: Methods for computational chemistry. *Nature* **504,** 96–97 (Dec. 2013).

11. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9,** 646–652 (Sept. 2002).

12. Centers for Disease Control and Prevention. United States of America. *Antibiotic / Antimicrobial Resistance | CDC* `https://www.cdc.gov/drugresistance/index.html`.

13. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. *Bulletin of the World Health Organization* **79,** 780–90 (2001).

14. Tan, S. Y. & Tatsumura, Y. Alexander Fleming (1881-1955): Discoverer of penicillin. *Singapore medical journal* **56,** 366–7 (July 2015).

15. Bigger, J. Treatment of Staphylococcal Infections with penicillin by intermittent sterilisation. *The Lancet* **244,** 497–500 (Jan. 1944).

16. Lewis, K. Persister Cells and the Paradox of Chronic Infections. *Microbe Magazine* (2010).

17. Del Pozo, J. L. & Patel, R. *The challenge of treating biofilm-associated bacterial infections* 2007.

18. Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial Biofilms: A Common Cause of Persistent Infections. *Science* **284,** 1318 (May 1999).

19. Spoering, A. L. & Lewis, K. Biofilms and planktonic cells of Pseudomonas aeruginosa have similar resistance to killing by antimicrobials. *Journal of Bacteriology* (2001).

20. Berghoff, B. A., Hoekzema, M., Aulbach, L. & Wagner, E. G. H. Two regulatory RNA elements affect TisB-dependent depolarization and persister formation. *Molecular Microbiology* (2017).

21. Duerr, T. & Lewis, K. Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS Biology* (2010).

22. Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature Chemical Biology* **12,** 208–214 (Mar. 2016).

23. Wagner, E. G. H. & Unoson, C. *The toxin-antitoxin system tisB-istR1: Expression, regulation, and biological role in persister phenotypes.* 2012.

24. Ramage, H. R., Connolly, L. E. & Cox, J. S. Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. *PLoS Genetics* **5** (ed Rosenberg, S. M.) e1000767 (Dec. 2009).

25. Gerdes, K. *et al.* Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon. *The EMBO journal* **5,** 2023–9 (Aug. 1986).

26. Brantl, S. & Jahn, N. sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiology Reviews* **39,** 413–427 (May 2015).

27. Yamaguchi, Y., Park, J.-H. & Inouye, M. Toxin-Antitoxin Systems in Bacteria and Archaea. *Annual Review of Genetics* **45,** 61–79 (Dec. 2011).

28. Goeders, N., Chai, R., Chen, B., Day, A. & Salmond, G. P. C. Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems. *Toxins* **8** (Sept. 2016).

29. Wang, X. *et al.* A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nature chemical biology* **8,** 855–61 (Oct. 2012).

30. Lee, K.-Y. & Lee, B.-J. Structure, Biology, and Therapeutic Application of Toxin-Antitoxin Systems in Pathogenic Bacteria. *Toxins* **8** (2016).

31. Vogel, J., Argaman, L., Wagner, E. G. H. & Altuvia, S. The small RNA istR inhibits synthesis of an SOS-induced toxic peptide. *Current Biology* (2004).

32. Hartmut Heinzmann Johannes Peters, S. R. & Ulrich, A. S. *Protein Origami* `http://www.ibg.kit.edu/protein_origami`.

33. Unoson, C. & Wagner, E. G. H. A small SOS-induced toxin is targeted against the inner membrane in Escherichia coli. *Molecular Microbiology* (2008).

34. Gurnev, P. A., Ortenberg, R., D??rr, T., Lewis, K. & Bezrukov, S. M. Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers. *FEBS Letters* (2012).

35. Steinbrecher, T. *et al.* Peptide-lipid interactions of the stress-response peptide tisb that induces bacterial persistence. *Biophysical Journal* (2012).

36. Walther, T. H. *et al.* Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism. *Cell* (2013).

37. Zimpfer, B. *Struktur- und Funktionsanalyse des membranaktiven bakterientoxischen Peptids TisB aus Escherichia coli* PhD thesis (Karlsruhe Institute of Technology, 2015).

38. Brantl, S. & Brueckner, R. *Small regulatory RNAs from low-GC Gram-positive bacteria* 2014.

39. Jahn, N., Preis, H., Wiedemann, C. & Brantl, S. BsrG/SR4 from Bacillus subtilis- the first temperature-dependent type I toxin-antitoxin system. *Molecular Microbiology* **83,** 579–598 (Feb. 2012).

40. Brantl, S. & Jahn, N. Heat-shock-induced refolding entails rapid degradation of bsrG toxin mRNA by RNases Y and J1. *Microbiology* **162,** 590–599 (Mar. 2016).

41. Jahn, N. & Brantl, S. One antitoxin—two functions: SR4 controls toxin mRNA decay and translation. *Nucleic Acids Research* **41,** 9870–9880 (Nov. 2013).

42. Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics (Oxford, England)* **14,** 378–9 (1998).

43. White, S. H. & Wimley, W. C. Membrane Protein Folding and Stability: Physical Principles. *Annual Review of Biophysics and Biomolecular Structure* **28,** 319–365 (June 1999).

44. Lear, J. D., DeGrado, W. F., Choma, C. & Gratkowski, H. Asparagine-mediated self-association of a model transmembrane helix. *Nature Structural Biology* **7,** 161–166 (Feb. 2000).

45. Jahn, N., Brantl, S. & Strahl, H. Against the mainstream: the membrane-associated type I toxin BsrG from Bacillus subtilis interferes with cell envelope biosynthesis without increasing membrane permeability. *Molecular Microbiology* **98,** 651–666 (Nov. 2015).

46. Becker, K. *Charakterisierung des bakteriellen Toxins BsrG aus Bacillus subtilis* PhD thesis (Karlsruhe Insitute of Technology, 2017).

47. Hespelt, A. *Darstellung und Charakterisierung des bakteriellen Toxins BsrG, Diploma Thesis* PhD thesis (Karlsruhe Institute of Technology, 2015).

48. Bishop, D. G., Rutberg, L. & Samuelsson, B. The Chemical Composition of the Cytoplasmic Membrane of Bacillus subtilis. *European Journal of Biochemistry* **2,** 448–453 (Nov. 1967).

49. Schlick, T. *Molecular modeling and simulation : an interdisciplinary guide* (Springer Science+Business Media, LLC, 2010).

50. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267,** 585–590 (June 1977).

51. Hansson, T., Oostenbrink, C. & van Gunsteren, W. Molecular dynamics simulations. *Current Opinion in Structural Biology* **12,** 190–196 (Apr. 2002).

52. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **389,** 457–484 (Jan. 1927).

53. Berendsen, H. J. *Simulating the physical world: Hierarchical modeling from quantum mechanics to fluid dynamics* (2007).

54. Dinur, U. & Hagler, A. T. in (John Wiley & Sons, Inc., 2007).

55. Jorgensen, W. L. & Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences* **102,** 6665–6670 (May 2005).

56. Atkins, P. W. & Friedman, R. *Molecular quantum mechanics* (Oxford University Press, 2011).

57. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30,** 1545–1614 (2009).

58. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–8 (June 2010).

59. Zgarbová, M. *et al.* Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation* **7,** 2886–2902 (Sept. 2011).

60. Yildirim, I., Stern, H. A., Kennedy, S. D., Tubbs, J. D. & Turner, D. H. Reparameterization of RNA χ Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *Journal of Chemical Theory and Computation* **6,** 1520–1531 (May 2010).

61. Jämbeck, J. P. M. & Lyubartsev, A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *Journal of Physical Chemistry B* **116,** 3164–3179 (2012).

62. Cramer, C. J. *Essentials of Computational Chemistry Theories and Models, 2nd Ed.* (2004).

63. Jensen, F. *Introduction to computational chemistry 3rd Edition* (Wiley and Sons, 2017).

64. Lennard-Jones, J. E. On the Forces between Atoms and Ions. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **109,** 584–597 (Dec. 1925).

65. Newton, I. Philosophiae Naturalis Principia Mathematica. *Pan* (1687).

66. Leach, A. R. *Molecular modelling : principles and applications* 2nd ed., 744 (Prentice Hall, Apr. 2001).

67. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18,** 1463–1472 (Sept. 1997).

68. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23,** 327–341 (Mar. 1977).

69. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13,** 952–962 (Oct. 1992).

70. Verlet, L. Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159,** 98–103 (July 1967).

71. Hockney, R., Goel, S. & Eastwood, J. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14,** 148–158 (Feb. 1974).

72. Berendsen, H. J. C., Van Gunsteren, W. F. & Fermi, E. Practical Algorithms for Dynamic Simulation, in Molecular-Dynamics Simulations of Statistical-Mechanical Systems. *Enrico Fermi Summer School, Varenna, Italy,* (1986).

73. Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids* (Clarendon Press, 1989).

74. Allen, M. P. Introduction to Molecular Dynamics Simulation. *Computational Soft Matter* **23,** 1–28 (2004).

75. Saito, M. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. *The Journal of Chemical Physics* **101,** 4055–4061 (Sept. 1994).

76. Saito, M. Molecular dynamics/free energy study of a protein in solution with all degrees of freedom and long-range Coulomb interactions. *The Journal of Physical Chemistry* **99,** 17043–17048 (Nov. 1995).

77. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98,** 10089–10092 (June 1993).

78. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103,** 8577–8593 (Nov. 1995).

79. Sagui, C. & Darden, T. A. Molecular Dynamics Simulation of Biomolecules: Long-Range Electrostatic Effects. *Annual Review of Biophysics and Biomolecular Structure* **28,** 155–179 (June 1999).

80. Frenkel, D. & Smit, B. *Understanding molecular simulation : from algorithms to applications* 638 ().

81. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369,** 253–287 (Jan. 1921).

82. Deserno, M. & Holm, C. How to mesh up Ewald sums (I): A theoretical and numerical comparison of various particle mesh routines (July 1998).

83. Hub, J. S., de Groot, B. L., Grubmüller, H. & Groenhof, G. Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge. *Journal of Chemical Theory and Computation* **10,** 381–390 (Jan. 2014).

84. Monticelli, L. & Salonen, E. *Biomolecular Simulations: Methods and Protocols* (Humana Press, 2016).

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

86. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126,** 014101 (Jan. 2007).

87. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52,** 255–268 (June 1984).

88. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31,** 1695–1697 (Mar. 1985).

89. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (Dec. 1981).

90. Marrink, S. J., De Vries, A. H. & Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B,* 7812–7824 (2007).

91. De Jong, D. H. *et al.* Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation* **9,** 687–697 (Jan. 2013).

92. Periole, X. & Marrink, S.-J. The Martini Coarse-Grained Force Field BT - Biomolecular Simulations. *Biomolecular Simulations* (2012).

93. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B* (2007).

94. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26,** 1701–1718 (Dec. 2005).

95. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26,** 1781–1802 (Dec. 2005).

96. Christen, M. *et al.* The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry* **26,** 1719–1751 (Dec. 2005).

97. Chair-Horner-Miller, C. & Barbara. *SC 2006 Conference, Proceedings of the ACM/IEEE.* (2006).

98. Monticelli, L. *et al.* The MARTINI goarse-grained force field: extension to proteins. *Journal of Chemical Theory and Computation* (2008).

99.   Marrink, S. J. & Tieleman, D. P. Perspective on the Martini model. *Chemical Society Reviews* **42,** 6801 (July 2013).

100.   Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Computational Biology* **6** (ed Levitt, M.) e1000810 (June 2010).

101.   Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational Lipidomics with insane : A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation* (2015).

102.   Wassenaar, T. A. *et al.* High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach. *Journal of Chemical Theory and Computation* (2015).

103.   Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J. & Tieleman, D. P. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation* (2014).

104.   Pluhackova, K., Wassenaar, T. A., Kirsch, S. & Böckmann, R. A. Spontaneous Adsorption of Coiled-Coil Model Peptides K and E to a Mixed Lipid Bilayer. *The Journal of Physical Chemistry B* **119,** 4396–4408 (Mar. 2015).

105.   Berendsen, H., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91,** 43–56 (Sept. 1995).

106.   Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **7,** 306–317 (Aug. 2001).

107.   Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4:- Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4,** 435–447 (2008).

108.   Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (Apr. 2013).

109.   Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. in *Lecture Notes in Computer Science, vol 8759. Springer, Cham* 3–27 (Springer, Cham, 2015).

110.   Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (Sept. 2015).

111.   Case, D. A. *et al. Amber 16* 2016.

112.   Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11,** 3696–3713 (Aug. 2015).

113. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79,** 926–935 (July 1983).

114. Jämbeck, J. P. M. & Lyubartsev, A. P. An extension and further validation of an all-atomistic force field for biological membranes. *Journal of Chemical Theory and Computation* **8,** 2938–2948 (2012).

115. Jämbeck, J. P. M. & Lyubartsev, A. P. Another piece of the membrane puzzle: Extending slipids further. *Journal of Chemical Theory and Computation* **9,** 774–784 (2013).

116. Ermilova, I. & Lyubartsev, A. P. Extension of the Slipids Force Field to Polyunsaturated Lipids. *The Journal of Physical Chemistry B* **120,** 12826–12842 (Dec. 2016).

117. Wolf, M. G., Hoefling, M., Aponte-Santamaría, C., Grubmüller, H. & Groenhof, G. g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *Journal of Computational Chemistry* **31,** 2169–2174 (Mar. 2010).

118. Joung, I. S. & Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B* **112,** 9020–9041 (July 2008).

119. Hockney, R. W. & Eastwood, J. W. *Computer simulation using particles* 540 (McGraw-Hill International Book Co, 1981).

120. Páll, S. & Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Computer Physics Communications* **184,** 2641–2650 (Dec. 2013).

121. Williams Thomas, K. C. & many others. *Gnuplot 5.0: an interactive plotting program* 2016. `http://gnuplot.sourceforge.net/`.

122. Stambulchik, E. *Grace - A 2D plotting tool* `http://plasma-gate.weizmann.ac.il/Grace/`.

123. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14,** 33–38 (1996).

124. Gower, J. C. & Ross, G. J. S. Minimum Spanning Trees and Single Linkage Cluster Analysis. *Applied Statistics* **18,** 54 (1969).

125. Kumar, R. *g_distMat -Analysis Tool* 2015. `https://github.com/rjdkmr/g_distMat`.

126. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **12,** 405–413 (Jan. 2016).

127. Wu, E. L. *et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations* 2014.

128. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal* **97,** 50–58 (2009).

129. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29,** 1859–1865 (2008).

130. Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS one* **2** (ed Yuan, A.) e880 (Jan. 2007).

131. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **34,** 2135–2145 (Sept. 2013).

132. Jämbeck, J. P. M. & Lyubartsev, A. P. Implicit inclusion of atomic polarization in modeling of partitioning between water and lipid bilayers. *Physical chemistry chemical physics : PCCP* **15,** 4677–86 (2013).

133. Lewis, B. A. & Engelman, D. M. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *Journal of molecular biology* **166,** 211–7 (May 1983).

134. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **114,** 7830–7843 (June 2010).

135. Ulmschneider, J. P., Smith, J. C., Ulmschneider, M. B., Ulrich, A. S. & Strandberg, E. Reorientation and dimerization of the membrane-bound antimicrobial peptide PGLa from microsecond all-atom MD simulations. *Biophysical journal* **103,** 472–82 (Aug. 2012).

136. Reißer, S., Roesky, P., Ulrich, A. & Elstner, M. Computational studies of membrane-active antimicrobial peptides and comparison with NMR data. `https://d-nb.info/1064504116/34` (2014).

137. Epand, R. M. & Epand, R. F. Lipid domains in bacterial membranes and the action of antimicrobial agents. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1788,** 289–294 (Jan. 2009).

138. Ulmschneider, J. P. Charged Antimicrobial Peptides Can Translocate across Membranes without Forming Channel-like Pores. *Biophysical journal* **113,** 73–81 (July 2017).

139. Wang, T. *et al.* Challenges and opportunities for siRNA-based cancer treatment. *Cancer Letters* **387,** 77–83 (Feb. 2017).

140. Steinmeyer, J., Rönicke, F., Schepers, U. & Wagenknecht, H.-A. Synthesis of Wavelength-Shifting Fluorescent DNA and RNA with Two Photostable Cyanine-Styryl Dyes as the Base Surrogate Pair. *ChemistryOpen* **6,** 514–518 (Aug. 2017).

141. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature* (1998).

142. Elbashir, S. M. *et al.* Duplexes of 21 ± nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* (2001).

143. Hannon, G. & Rossi, J. Unlocking the potential of the human genome with RNA interference. *Nature* (2004).

144. Coburn, G. A. & Cullen, B. R. Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference. *J. Virol* (2002).

145. Kim, D. H. & Rossi, J. J. Strategies for silencing human disease using RNA interference. *Nat Rev Genet* (2007).

146. Davis, M. E. *et al.* Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature* (2010).

147. Wu, S. Y., Lopez-Berestein, G., Calin, G. A. & Sood, A. K. RNAi therapies: drugging the undruggable. *Science translational medicine* **6,** 240ps7 (June 2014).

148. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nature Reviews Molecular Cell Biology* **16,** 95–109 (Feb. 2015).

149. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* (1998).

150. Stryer, L. Fluorescence Energy Transfer as a Spectroscopic Ruler. *Annual Review of Biochemistry* **47,** 819–846 (June 1978).

151. Hirsch, M. & Helm, M. Live cell imaging of duplex siRNA intracellular trafficking. *Nucleic acids research* **43,** 4650–60 (May 2015).

152. Alabi, C. A. *et al.* FRET-Labeled siRNA Probes for Tracking Assembly and Disassembly of siRNA Nanocomplexes. *ACS Nano* **6,** 6133–6141 (July 2012).

153. Wahba, A. S. *et al.* Phenylpyrrolocytosine as an Unobtrusive Base Modification for Monitoring Activity and Cellular Trafficking of siRNA. *ACS Chemical Biology* **6,** 912–919 (Sept. 2011).

154. Kummer, S. *et al.* Fluorescence Imaging of Influenza H1N1 mRNA in Living Infected Cells Using Single-Chromophore FIT-PNA. *Angewandte Chemie International Edition* **50,** 1931–1934 (Feb. 2011).

155. Holzhauser, C., Liebl, R., Goepferich, A., Wagenknecht, H.-A. & Breunig, M. RNA "Traffic Lights": An Analytical Tool to Monitor siRNA Integrity. *ACS Chemical Biology* **8,** 890–894 (May 2013).

156. Holzhauser, C. & Wagenknecht, H.-A. "DNA Traffic Lights": Concept of Wavelength-Shifting DNA Probes and Application in an Aptasensor. *ChemBioChem* **13,** 1136–1138 (May 2012).

157. Bohländer, P. R. & Wagenknecht, H.-A. Bright and photostable cyanine-styryl chromophores with green and red fluorescence colour for DNA staining. *Methods and Applications in Fluorescence* **3,** 044003 (Sept. 2015).

158. Bohländer, P. R. & Wagenknecht, H.-A. Synthesis and evaluation of cyanine–styryl dyes with enhanced photostability for fluorescent DNA staining. *Organic & Biomolecular Chemistry* **11,** 7458 (Oct. 2013).

159. Walter, H. K., Bohländer, P. R. & Wagenknecht, H. A. Development of a wavelength-shifting fluorescent module for the adenosine aptamer using photostable cyanine dyes. *ChemistryOpen* (2015).

160. Bohländer, P. R. & Wagenknecht, H. A. Synthesis of a photostable energy-transfer pair for "dNA traffic lights". *European Journal of Organic Chemistry* (2014).

161. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie (International ed. in English)* **40,** 2004–2021 (June 2001).

162. Walter, H.-K., Olshausen, B., Schepers, U. & Wagenknecht, H.-A. A postsynthetically 2'-"clickable" uridine with arabino configuration and its application for fluorescent labeling and imaging of DNA. *Beilstein Journal of Organic Chemistry* **13,** 127–137 (Jan. 2017).

163. Frisch, M. J. *et al. Gaussian~16 Revision A.03* 2016.

164. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98,** 5648–5652 (Apr. 1993).

165. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **98,** 11623–11627 (1994).

166. Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **54,** 724–728 (Jan. 1971).

167. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics* **23,** 69–89 (Apr. 1951).

168. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25,** 247–260 (Oct. 2006).

169. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **97,** 10269–10280 (Oct. 1993).

170. Sosa, C. *et al.* A local density functional study of the structure and vibrational frequencies of molecular transition-metal compounds. *The Journal of Physical Chemistry* **96,** 6630–6636 (Aug. 1992).

171. Godbout, N., Salahub, D. R., Andzelm, J. & Wimmer, E. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Canadian Journal of Chemistry* **70,** 560–571 (Feb. 1992).

172. Stroud, J. *Web server that constructs nucleic acid helices from sequences* `http://structure.usc.edu/make-na/server.html`.

173. Cheatham, T. E., Brooks, B. R., Kollman, P. A. & III. Molecular modeling of nucleic acid structure. *Current protocols in nucleic acid chemistry* **Chapter 7,** Unit 7.5 (Nov. 2001).

174. Pérez, A. *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical Journal* **92,** 3817–3829 (June 2007).

175. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophysical Journal* **100,** L47–L49 (May 2011).

176. Schrödinger LLC. *The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version˜1.8* Nov. 2015.

"In scientific thinking are always present elements of poetry.
Science and music requires a thought homogeneous."

— Albert Einstein (1879-1955)

Carsten Wiebusch
For the inspiration, the passion for music,
experiences, enabled opportunities..
and all which can't be expressed in words.