Molecular Dynamics Modelling of Biomolecular Interactions with Lipid Membranes

and

Novel Coarse Grain Lipid Model Development

A thesis presented in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy
In
Biochemistry

at Massey University, Albany,
New Zealand

Elisey Kobzev

2017

When I went to school, they asked me what I wanted to be when I grew up.
I wrote down 'happy'.

They told me I didn't understand the assignment
And I told them they didn't understand life.

- Unattributed

Abstract

Lipids comprise a key component of the cellular membrane and are essential to many biological processes. *In silico* investigations provide valuable opportunities to study the dynamics and structure of biological molecules, such as lipid membranes and the molecules that interact with them, at near atomic resolutions. In the context of this thesis three research projects were undertaken with a focus on lipid membrane simulations.

The structure and dynamics of the novel antibacterial battacin analogue peptides and their interactions with model membranes of the common pathogenic gram positive and gram negative bacterial species *Staphylococcus aureus* and *Escherichia coli* were studied. Antibacterial peptides are a key area of research due to their potential medicinal applications in overcoming the current antibiotic resistance crisis. However, detailed knowledge of their mode of action is often lacking. The peptides were to found to insert into the bacterial membranes, facilitated by the insertion of the fatty acid moiety, and showed strong affinity for all three types of membranes studied.

Antifreeze protein 1 (AFP1) is critical to cell survival at near freezing temperatures. Structural analysis of the behavior of AFP1 is presented, including a study of its possible aggregation. Interactions of AFP1 were studied in conjunction with a model of a typical cell membrane. AFP1 units were found to be flexible in solution, adopting a variety of non α-helical structures. In certain cases, two AFP1 proteins aggregated together and interacted with each other. Furthermore, AFP1 interacted with the unsaturated lipid membrane, coming to rest on its surface, providing insight into the freezing damage prevention mechanism.

Finally, in order to facilitate simulation of larger biological membrane systems, a novel supra-atomic phospholipid model was proposed, and model parameters developed for the common lipid 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC). The model is based on and ultimately compatible with the GROMOS 54a8 atomic-level force field[103] including the GROMOS coarse-grained water model[111]. It is also polarisable, unlike many popular supra-atomic models. The DPPC model was developed following a bottom-up approach, and is intended to pave a way for stepwise parameterisation of other lipids, to build a library of “plug and play” lipid parameters.

Thesis Outline

The scope of the thesis covers a variety of molecular dynamics simulation projects, centred on lipid membrane simulations, including novel coarse-graining methodology development, as well as utilising current atomistic modelling methods for biochemical investigations. The thesis is split into four chapters, each relating to a separate topic. The first chapter gives a brief overview of concepts involved in molecular dynamics simulations. The subsequent three chapters cover three separate projects. Each project chapter is split into several parts, including a specific literature review, results, methods, discussion and conclusion. Chapter 2 describes the investigation of novel anti-bacterial peptides and explores their behaviour in the context of lipid membranes. Chapter 3 is about investigation of winter flounder Antifreeze Protein 1 structure and aggregation in the context of lipid membranes. Chapter 4 deals with creation of novel coarse-grained modelling methodology for the lipid DPPC. A comprehensive list of combined and numbered references follows.

Acknowledgements

The author would like to acknowledge the following parties, without whom the outcome of this thesis would not have been possible:

Dr Jane Allison, for supervision and invaluable insight, advice and technical assistance.

Dist. Prof. Peter Schwerdtfeger, for co-supervision and advice.

Ivan Welsh, for advice and technical assistance with GROMOS and GROMACS tools, as well as in house programming.

William Irvine for advice and assistance with GROMACS lipid bilayer simulation.

Dr Thomas Collier for assistance with further development of the coarse-grained project and advice.

Dr Lukas Wirz for help with programming and technical assistance with GROMOS.

Ashar Malik and Shamim Shadfar for advice and assistance as part of the Allison Group.

Assist. Prof. Sereina Riniker, for invaluable help and insight into the GROMOS coarse grain models, as well as technical assistance in enabling GROMOS code updates.

Emeritus Prof. Wilfred van Gunsteren, for invaluable help and insight into the GROMOS coarse grain models.

Prof. Syma Khalid, for provision of bacterial membrane models.

Dr Thomas Piggot, for provision of bacterial membrane models and ongoing technical assistance with supersampling methods and lipid simulation.

Prof. Denis Kobzev and Inga Kobzeva for unwavering support

Brenda Roldan, for being an inspiration

Table of Contents

Abbreviations

AFP1	Antifreeze protein 1
AL	Atomic level
CG	Coarse-grain
Dab	*α,γ*-diaminobutyric acid
DGDG	1,2-digalactosyldiacylglycerol
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DPPE	1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine
DPPC	1,2-dipalmitoyl-sn-glycero-3-phosphocholine
LJ	Lennard-Jones
LPS	Liposaccharide or cardiolipin
MD	Molecular dynamics
NVT	Constant moles, volume, temperature
NVP	Constant moles, volume, pressure
O17	Octapeptide 17
P30	Pentapeptide 30
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
RDF	Radial distribution function
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation

Abstract

In silico simulations pose an attractive opportunity to glean insight into dynamics of molecular processes at near atomic scales. This chapter looks at the basic principles involved in creating *in silico* simulations: representation of particles, interactions, propagation of motion and boundary conditions.

Chapter 1: Introduction to Molecular Dynamics

1.1 Molecular dynamics modelling

Biomolecules are the building blocks of all living matter and their study is crucial to understanding of biochemical processes. The study of intrinsic structural and energetic properties of biomolecules is a daunting prospect, especially at the molecular and atomic levels. Biomolecules comprise a wide class of chemical compounds which vary greatly both in function and size from amino acids to protein assemblies. Biomolecular systems are flexible and mobile at multiple levels of resolution, resulting in a wide array of adopted conformations[1]. Complicated molecules, such as proteins, often do not exist as static rigid structures, but rather adopt a Boltzmann weighted ensemble of conformations.

It is important to be able to study biomolecular dynamics and structures at near atomic level. Knowing how biomolecules move and structure themselves in biological environments is critical in understanding many biochemical mechanisms and problems. Such studies form the backbone of understanding biochemical mechanisms and pathways, and finding out which adopted structures are key to the activity of the chemicals in question.

Many experimental techniques exist for determining molecular structure and dynamics at near atomic level of detail, such as X-ray and neutron diffraction, as well as nuclear magnetic resonance (NMR) and dynamic NMR (dNMR) spectroscopy[2-3]. Diffraction allows for the investigation of structures adopted by molecules in solid (crystal or fibre) form at extremely high (< 0.1 nm) resolution. Information about the homogeneity and local mobility (B factors[4]) of the crystal is also available, giving some indication about which regions of the structure are more mobile than others, with the most mobile parts identified by absence from the diffraction pattern[5]. However, the full nature of such dynamics is not explored, and the structures accessed are only those present in the solid state, which is not always indicative of the native environment of a biological molecule.

NMR spectroscopy provides observables that are time- and ensemble-averages, from which both atomic-level structures and information about dynamics, such as the time-scale of the motion, can be derived. The limitations of NMR are that the averaging of information over a number of structures must be taken into account, and chemicals typically have to be solvated in single component solvent - water, heavy water or other non-biological solvents, which again is not bio-mimetic.

In silico simulations pose an attractive opportunity to glean insight into the structures and dynamics of molecular processes from sub-atomic to molecular scales. Molecular dynamics (MD) simulations allow us to realistically model biomolecules and follow specific molecular processes at near atomic scales explicitly[6]. Since the first simulation by Rahman and Stillinger in 1971, these methods have been extensively used as a powerful tool for mechanistic and structural investigation of biomolecules[7]. Computer simulations bridge the gap in knowledge left by the limitations of classical structure determination techniques and experimental methods that report on thermodynamics. *In silico* methods allow us to simulate increasingly sophisticated models, aiming to reproduce *in vivo* conditions as closely

as possible. It is possible to follow the dynamics of individual biomolecules and study the structures adopted by them, giving close insight into biomolecular mechanisms.

1.2. Simulation concepts

The key principles of molecular dynamics simulations are outlined below. A set of interrelated decisions must be made when designing a molecular simulation. The method of representation of chemical particles and their movement in the simulation must be chosen, as well as what degrees of freedom the particles will possess. Simulation boundaries also have to be established.

1.2.1 Representation

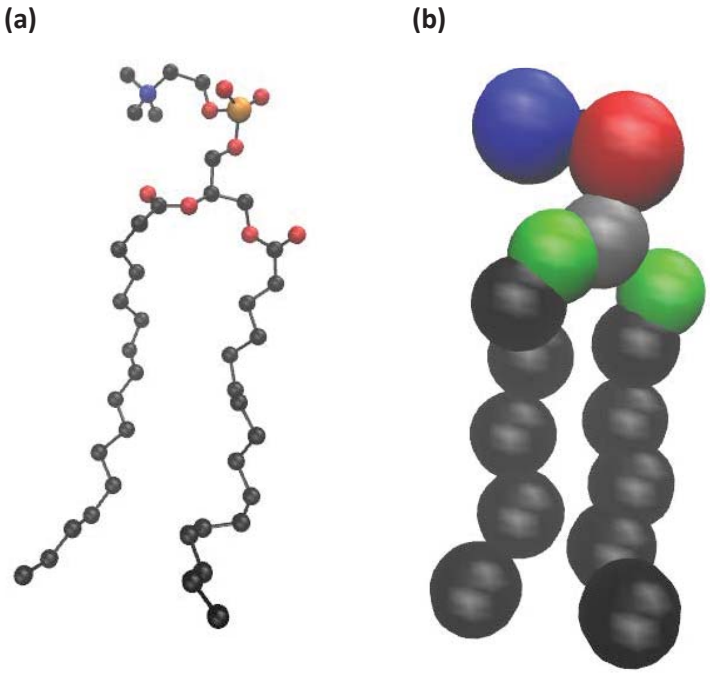


Figure 1.1: Two modelling approaches for the phospholipid dipalmitoylphosphatidylcholine (DPPC): (a) atomic level model, where each particle represents an atom; and (b) CG model, where each particle represents a cluster of atoms, such as a functional group.

It is possible to model a chemical system at many levels, ranging from sub-atomic to supra-atomic[8]. The level of modelling selected primarily depends on the application and on what physical processes are of interest. Typically, the motion of biologically relevant molecules, such as the folding of peptides, functional motion of proteins and formation of biomolecular complexes, membranes and micelles, depends on weak non-bonded interactions, which are best modelled at near-atomic level[9]. Chemical

reactions, in which bonds are broken and/or formed, require a sub-atomic quantum mechanical description. At the time of writing, it is computationally unfeasible to perform quantum mechanics dynamical calculations to determine sub-atomic electronic behaviour for large systems containing more than hundreds of atoms. Instead, the behaviour of atoms can be followed by modelling them as hard spheres via the Born-Oppenheimer approximation, which treats atomic nuclei as stationary objects relative to their electron density. Such models are known as atomic level (AL) models, where each atom is explicitly represented as single particle in the simulation, emulating typical "ball and stick" chemistry models. In this thesis, united atom models in which aliphatic hydrogen atoms are subsumed by the carbon atom to which they are bonded are considered to be a subset of AL models[10-11]. United atom approaches are typically used to reduce the computational cost of simulations, as aliphatic hydrogen behaviour is typically considered to be not mechanistically important and can be accounted for by slight alteration of the carbon parameters. Furthermore, a cluster of multiple atoms can also be represented by a single simulation object, resulting in a supra-atomic or coarse-grained (CG) model. CG models may also be supra-molecular, if the CG representation encompasses multiple molecules. Generally, this thesis will deal with three types of modelling – AL, CG and a combination of the two (*Figure 1.1*).

1.2.2 Interactions

The behaviour of the particles making up the simulated system is governed by a "forcefield", which describes the properties of the particles and the interactions between them. A forcefield consists of a parametric potential energy function for all particles in the system, and its specification requires the choice of a functional form and of the associated parameter set[12]. Many different forcefields exist – their form and parameterisation, including the resolution level at which the molecules are represented, depends heavily on the use intended at the time of generation. Forcefields consist of several component functions, which are described below. Forcefields are typically parametrised to match experimental data at near room temperatures.

Firstly, there are parameter values that are specific to individual particles and do not possess any associated mathematical function, such as the mass of an atom or cluster of atoms. Such parameters are only used as constants in the calculation of forces during the main MD simulation.

The second class of parameters are assigned to individual atoms and are used in the calculation of "non-bonded" interactions between particles that are in different molecules, or are separated by four or more bonds. Such parameters include partial charges and Lennard-Jones (L-J) sigma and epsilon values (*Equation 1.1)* The effects of such parameters are twofold – they establish the effective size of the particles and, also describe the favourability of the "non-bonded" interactions between particles. Typically, inter-particle interactions are only calculated for pairs of particles within a certain distance (e.g. 1.0 - 1.4 nm for most AL models) to minimise calculation time.

The effective radius of a simulated particle, dispersion (van der Waals) interactions and short range attraction between pairs of particles are described by the Lennard-Jones function (*Equation 1.1.1, Figure*

1.2). LJ interactions are calculated for a pair of particles as a function of the inter-particle distance in a pairwise manner, and multi-body terms are not considered. At the LJ radius *σ*, the interactions between particles are neutral, if the particles come closer, the interaction becomes strongly repulsive, if they move further away – attractive (up to the magnitude negative *ε*), and eventually zero. LJ can be written in terms of two coefficients derived from *ε* and σ values – C6 and C12, and this form is typically used by MD engines for ease of calculation (*Equation 1.1.2*). A lookup table frequently exists for all combinations of interacting particles where these values can be looked up. Such look up tables are sometimes generated by multiplying individual parameters associated with particles (√C6 and √C12) together to give and approximation of the real C6 and C12 values. The LJ interaction energy and resulting force can be truncated at the long-range cut-off distance or, alternatively, the LJ function can be shifted so that they approach zero. The LJ function approaches zero at large inter-particle distances and so these interactions may be considered negligible.

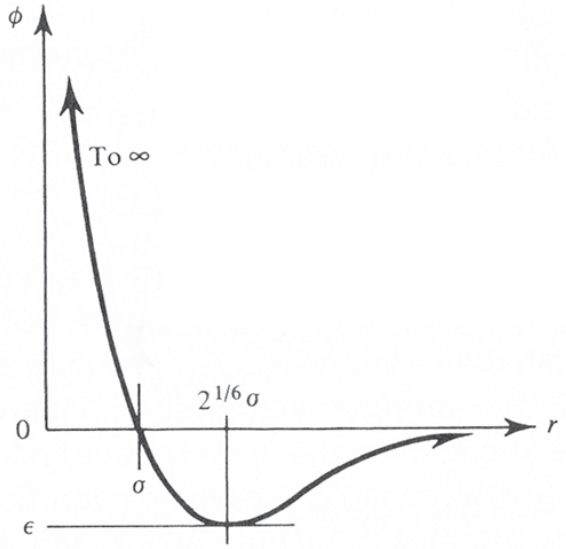


Figure 1.2: Lennard-Jones potential, where Φ is the potential energy, r is the distance between two objects, σ is the LJ radius and ε is the depth of the potential. Adapted from Tien and Lienhard, Statistical Thermodynamics[13].

(1.1.1) $$U_{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Where:
$U_{LJ}(r)$ is the Lennard-Jones potential energy
r is the inter-particle distance
ε is the well depth parameter
σ is the Lennard-Jones radius

(1.1.2) $$U_{LJ}(r) = \frac{C12}{r} - \frac{C6}{r}$$

Where:
$U_{LJ}(r)$ is the Lennard-Jones potential energy
r is the inter-particle distance
C12 is the C12 parameter
C6 is the C6 parameter

Equation 1.1: Two forms of the Lennard-Jones potential energy function.1.1.1: General form. 1.1.2: Typical MD format expressed in terms of tabulated constants.

The other type of non-bonded interaction is the electrostatic interaction between particles, which is described by the Coulomb potential (*Equation 1.2*). Each simulated particle typically has a "partial charge" assigned to it. The charge is assigned to the centre of mass of the simulated particle. Partial charges approximate the average positions of electrons in the molecule. Several ways of allocating charge exist and vary greatly between methods and forcefields. Similarly to the LJ potential, the Coulomb potential depends on the inter-particle distance. It is increasingly attractive at short distances for a pair of particles of opposite charge, and the repulsive for particles of like charge. Unlike dispersion interactions, electrostatic interactions operate over large inter-particle distances, and thus are still significant outside the long-range cut-off distance. Long range electrostatic interactions are therefore either accounted for using simplified algorithms, such as a continuum representation in the case of reaction field electrostatics, or using reciprocal space methods such as particle mesh Ewald.

(1.2) $$U^C(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

Where:
$U^C(r)$ is the Coulomb potential energy
q_1 *and* q_2 are the partial charges on particles *i* and *j*
ϵ_0 is the dielectric permittivity of free space
r is the inter-particle distance

Equation 1.2: Coulomb potential energy function for electrostatic interactions[14].

The final case of parameters contains those responsible for interactions between groups of particles that are separated by 1 to 4 bonds: the bond, angle, torsional dihedral and, in some forcefields, improper dihedral terms. Such interactions are typically handled by harmonic or quartic potential functions and sine/cosine functions with parameters individually and explicitly outlined for most common

combinations. Improper dihedral functions exist to help maintain geometry around tetrahedral centres where some of the atoms may not be explicitly represented. Examples of bonded potential energy functions are given below.

(1.3.1) $$U^{Bond} = k_{ij}^{h}(r - r_0)^2$$

(1.3.2) $$U^{\theta} = k_{ijk}^{h}(\theta - \theta_0)^2$$

(1.3.3) $$U^{\phi} = k_{ijkl}^{\phi}(1 + \cos(m\phi - \gamma))$$

Where:
U^{Bond} is the harmonic bond potential energy
k_{ij}^{h} is the bond harmonic force constant
r is the distance between particles *i* and *j*
r_0 is the ideal bond distance

U^{θ} is the harmonic angle potential energy
k_{ijk}^{h} is the angle harmonic force constant
θ is the bond angle between particles *i*, *j* and *k*
θ_0 is the ideal bond angle

U^{ϕ} is the torsional dihedral potential energy
k_{ijkl}^{ϕ} is the dihedral force constant
m is the multiplicity
ϕ is the dihedral angle between particles *i*, *j*, *k* and *l*
γ is the phase shift

Equation 1.3: Bond, angle and torsional dihedral potential energy functions.

1.2.3 Motion

Methods to generate Boltzmann-weighted ensembles of conformations can broadly be divided into two categories: Monte-Carlo (MC) and Molecular Dynamics (MD). For the purposes of this thesis, only MD will be considered as all the simulation packages used in the work are MD based. In MD simulations numerical integration of Newton's equations of motion allows determination of particle coordinates and velocities at discrete time points (*Equation 1.3*).

(1.3) $$F_i = m\ddot{r}_i = -\frac{\partial}{\partial r_i}U(r)$$

Where:
F_i is the force on particle *i*
$\ddot{r}_i$ is the second derivative of the position variable of particle *i* with respect to time (acceleration of particle *i*)
m is the mass of particle *i*
U(r) is the total potential energy function

Equation 1.3: MD forces are given by the differential of the potential energy fucntion.

Energies of each conformation are evaluated at discrete timesteps. Computing the energy of the current conformation allows for the computation of the forces (hence the term - "forcefield") as the derivative of the energy with respect to the coordinates (*Equation 1.3*). These Newtonian forces give the acceleration on each particle and are combined with the current positions and velocities to calculate the positions and velocities at the next simulation step. The integration timesteps must be short enough (1 fs) to approximate the true trajectory without significant accuracy loss. Using bond constraint algorithms such as SHAKE[15], the time can be increased to 2 fs, as the fastest vibrations from the bonds will be removed. Several algorithms for time integration exist such as leap-frog[16], Verlet , velocity Verlet and others[17-18]. The quality of the forcefield parameters used in the calculation determines how well the simulation matches the generated sets of coordinates to a real Boltzmann weighted ensemble of configurations.

MD simulations can be run under different types of thermodynamic ensembles to follow processes and calculate system properties under different conditions. The commonly simulated ensembles are:

- NVE, microcanonical ensemble – number of moles (particles), volume and energy are kept constant;
- NVT, canonical ensemble – moles, volume and temperature are kept constant;
- NPT, isothermal-isobaric – moles, pressure and temperature are kept constant.

Temperature and pressure can be kept constant via the use of thermostat and barostat algorithms respectively. The full details of all the simulations and ensembles used in the following chapters can be found in the corresponding *Methods* sections.

1.2.4 Boundaries

Biomolecular simulations typically contain one to a few large biomolecules (e.g. a protein), or multiple smaller molecules (e.g. a lipid bilayer) solvated in water. Unfortunately, the availability of current

computational resources imposes a practical feasibility limit on the amount of particles and length of the simulations. A trade-off exists between the size and length of the simulation against the amount of computational resources available to process the information. An MD simulation of a large system (more than 100, 000 particles) using a typical timestep of 2 fs might take more than several CPU years to run for the length of time needed to obtain statistically good sampling. For this reason, most simulations of biomolecules are limited to hundreds of nanoseconds or, more recently, using dedicated computing resources, parallel processing and multithreading, micro- to milliseconds[1].

To minimise the amount of simulated particles whilst simultaneously negating any possible boundary effects, periodic boundary conditions are used to mimic the presence of an infinite bulk of solvent molecule[19]. Periodic boundary conditions translate the environment from one edge of the simulation space to the next, giving an approximation of an infinite medium that interacts with itself and is free from boundary effects.

Typically, the distance from the outer edge of the molecule (or group of molecules) to the edge of the box must be equal to at least half the cut-off distance for long-range interactions in all directions, so that the molecule does not interact with itself. The exceptions to the rule are simulations of continuous media, such as bilayers which only need satisfy the condition in the direction of the bilayer normal only. In practice, this distance that separates the molecule of interest from the edge of the simulation is often created to be more than half the cut-off to allow for conformational changes that change the shape and size of the molecule, as well as other rotational movement to occur. This condition, therefore imparts a minimum size constraint on most simulations.

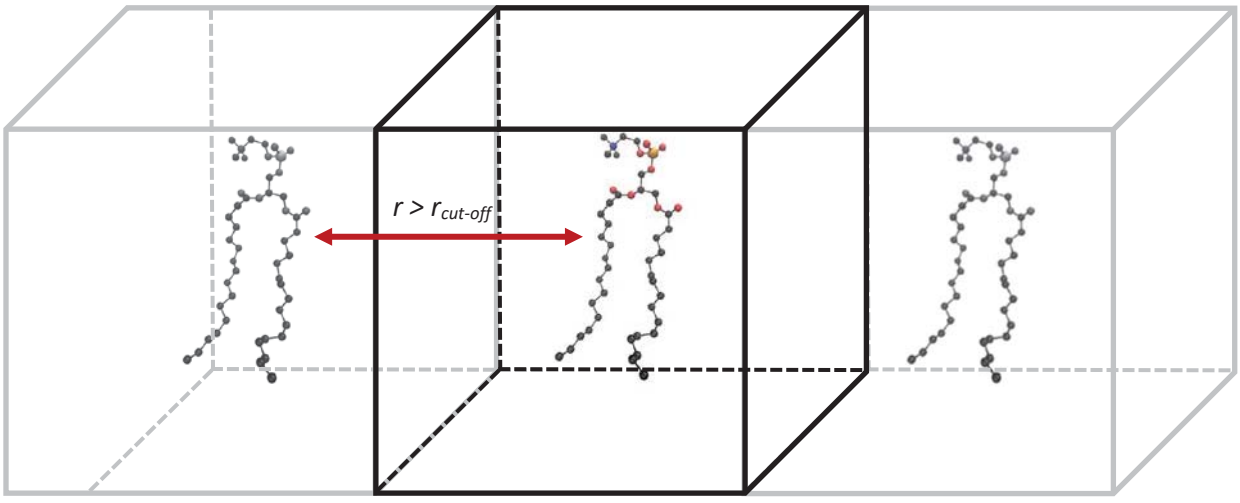


Figure 1.3: Typical simulation set-up, a 3D box is established with periodic boundary conditions. Superimposed periodicity in x direction is shown in greyscale, illustrating importance of situating a box with dimensions of at least half the cut off distance from the molecule of interest to each boundary to avoid self-interaction.

1.3 Summary

AL forcefields for biomolecules are well established, as are the software packages for running MD simulations. Such packages typically contain both a full complement of MD engine and analysis software components, incorporating all of the concepts discussed above into the code. Many differing software exist, to name a few examples: GROMACS, GROMOS, LAMMPS, CHARMM, AMBER and many more. Typically, most programs come with an associated forcefield; e.g. the GROMOS simulation software package comes with a series of GROMOS forcefields, such as GROMOS 54a7. Many simulation software packages, such as GROMACS, can work with a variety of forcefields. It is also possible to edit the forcefield files manually to create new types of simulation objects or to change and improve existing parameters. Development of novel CG methods has resulted in the introduction of new CG forcefields which are growing in popularity.

The following chapters cover simulations of proteins and peptides in the context of lipid membranes using existing AL methodologies, as well as development of novel CG models. Chapter 2 concerns the investigation of novel antibacterial peptides in the context of bacterial membranes and their mechanism of action. In chapter 3, MD investigation of the antifreeze protein AFP1 will be performed to investigate the possible low temperature induced dimerisation and dynamics in the presence of a cell membrane. Finally, chapter 4 will cover the creation and parametrisation of a novel polarisable CG model of the lipid DPPC for use in bilayer simulations. The model will be designed for use with the GROMOS MD engine, utilizing and adapting the GROMOS 54a8 forcefield as a basis for the modelling environment. Together, the work described in this thesis covers the methodology utilised in creation of novel models, as well as using established methods to perform biologically and medically relevant simulation experiments.

The author would like to note that this overview of MD principles is not comprehensive and focuses on aspects that have the most relevance in the subsequent chapters. For further comprehensive technical information, the reader can consult “Understanding Molecular Simulation” by Frenkel, D. and Smitt, B.[19] or specific MD software manuals for the software used in this thesis[11, 20].

Abstract

Antibacterial peptides are a key area of research due to their potential medicinal applications. Recently novel unnatural analogues of the antibacterial peptide battacin have been synthesised by Dr Sarojini's group at the university of Auckland and have been shown to have broad multi-species antibacterial action. Using *in silico* methods the structure and dynamics of the novel peptides were studied, as well as their interaction with bacterial membrane models of common Gram-positive and negative species *Staphylococcus aureus* and *Escherichia coli*. The peptides were found to insert themselves into the bacterial membranes, showing strong hydrogen bonding affinity for all three types of membranes studied. Simulations confirmed the interaction mechanism is facilitated by the insertion of the fatty acid into the bacterial membranes.

Chapter 2: Structural characterisation and investigation of the mode of action of novel antibacterial peptides

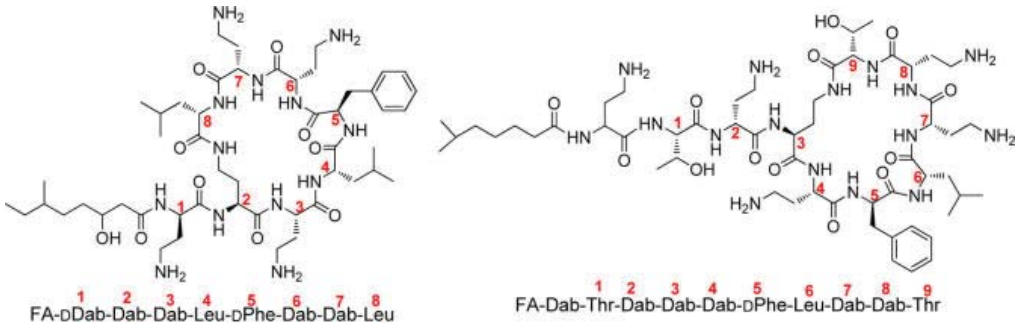
2.1 Antibacterial peptides

Research into new antibiotics is critical due the global rise in number of resistant bacteria[21-23]. Many families of antibacterial agents exist and the search continues for natural product derivatives and their analogues. One approach is to study short peptides which are often used by the bacteria themselves as hostile anti-bacterial agents. Such agents are released in "bacterial wars" when the host bacteria need to compete with other bacterial species for space and resources[24]. Such alternative natural products pose attractive opportunities for research into new broad spectrum antibiotics. This chapter focuses on the study two small novel antibacterial peptides with proven antibacterial activity, derived from non-ribosomal lipopeptides battacin and polymyxin B.

Naturally occurring non-ribosomal lipopeptides belong to a diverse class of non-ribosomal antibiotics. Typically, lipopeptides have a long chain fatty acid conjugated to the N-terminus[25], but the length can vary in individual peptides. Lipopeptides are diverse, consisting of 11 to 13 amino acids, both natural and non-proteinogenic and also typically have cyclic systems[25].

Broad spectrum lipopeptide antibiotics daptomycin and polymyxin B, which are found in soil bacteria are used to treat antibiotic resistant diseases such as MRSA caused by *S. aureus*[26-29]. However, such treatments frequently pose nephrotoxicity, neurotoxicity issues and other negative side effects, such as renal failure[30-32]. Furthermore, as polymyxin was discovered over 50 years ago, there is also an issue of growing bacterial resistance, so these drugs are typically used as a last resort treatment only. Therefore, it would be advantageous to explore alternative similar short peptides, such as battacin and its analogues, for anti-microbial applications to create a broader range of active anti-bacterial agents.

2.1.1 Battacin lipopeptide and analogues



Battacin **Polymyxin B**

Figure 2.1: Battacin and polymyxin B peptide structures and sequences. FA denotes the fatty acid attached to the N-terminus. Taken un-adapted from De Zoysa G.H. et al., 2015[33]

Battacin is a novel naturally occurring cyclic lipopeptide, belonging to the octapeptin group of peptide antibiotics with strong structural similarities to polymyxin B (*Figure 2.1*). Octapeptins are typically characterised by branched termini and a high percentage of the non-proteinogenic amino acid *α,γ*-diaminobutyric acid (Dab)[26, 34], including both the D- and L- isomers. Both polymyxin B and battacin have similarly sized fatty acid chains and contain five Dab residues, along with one Leucine (Leu) and D-Phenylalanine (D-Phe)[26, 35]. Antimicrobial activity of both peptides has been linked to the combination of hydrophobic regions comprised of Phe, Leu and the fatty acid with the hydrophilic five positively charged Dab residues[36-37]. The positive charge on the Dab residues is critical in forming attractive interactions with negatively charged phosphate groups present in the bacterial membranes, especially in lipopolysaccharide (LPS) or cardiolipin[38]. Battacin has been shown to have comparable antimicrobial activity to polymyxin B but is far less toxic[26]. At the time of writing, the synthesis of battacin has not been published. However, peptide analogues of battacin are currently being researched and can offer an alternative.

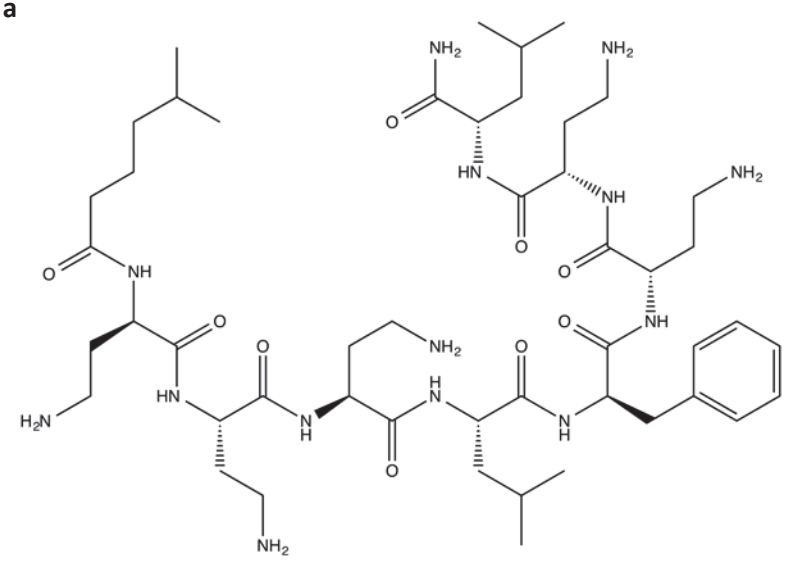
Several mechanisms for the action of polymyxin type antibacterial agents have been proposed[39]. There is strong evidence to suggest that polymyxin type antibiotics have surface interactions with the membrane itself[40]. The initial suggested target of polymyxin in gram-negative bacteria is the LPS present in the outer leaflet of the outer membrane. It is suggested that the amphipathic nature of the peptide enables the permeation of the agent through the cell membrane[41]. A computational study of polymyxin B also suggests interaction with the LPS part of the bacterial membranes via displacement of cations which stabilise the lipopolysaccharide interactions[42]. The fatty acid tail part is reported to be inserted first into the membrane followed by the rest of the molecule, creating disruption in the lipid packing. Additional mechanisms suggest that polymyxin B has the ability to form pore like aggregates which facilitate its insertion into the cell. *In vitro* studies suggest that both of the two proposed mechanisms could be valid and unrelated to each other[43]. Other methods of action have also been hypothesised, such as inhibiting enzyme function[44] or action via creating hydroxyl radicals[45], depending on the species and type of bacteria present.

Dr. Sarojini's research group at the University of Auckland created a range of novel battacin analogues which have shown broad antibiotic activity[33]. Their research focused on synthesising smaller peptides that would mimic the structure and activity of battacin. Two compounds were of particular interest – octapeptide 17 (O17), due to its superior antibacterial activity, and a shorter analogous pentapetide 30 (P30) (*Figure 2.2*). Both peptides exhibited rapid broad range antibacterial activity against both Gram-positive and Gram-negative bacteria of the following six species: *E. coli*, *S. aureus*, *Pseudomonas aeruginosa*, *Erwinia amylovora* 1501, streptomycin resistant *Erwinia amylovora* and *Pseudomonas syringe pv. actinidiae* 16207.

Linear battacin analogues such as O17 and P30 were found to exhibit greater antimicrobial activity than cyclic analogues[33]. The nature of the fatty acid group was also found to affect activity – the highest antimicrobial activity was achieved when the acid carbon chain length numbered between six and twelve carbons[46]. Furthermore, there are inherent advantages in utilising short peptides from the

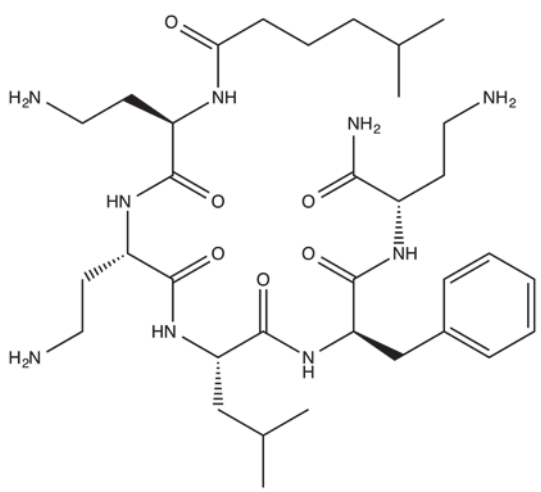
synthetic perspective, as the involved synthesis is typically much simpler and cheaper than that of more complex cyclic natural products comprised of more residues.

a



*Octapeptide O17: 5-methoxyhexanoyl-(**D**)**Dab**-**Dab**-**Dab**-**Leu**-(**D**)**Phe**-**Dab**-**Dab**-**Leu**-NH2*

b



*Pentapeptide P30: 5-methoxyhexanoyl-(**D**)**Dab**-**Dab**-**Leu**-(**D**)**Phe**-**Dab**-NH_2*

Figure 2.2: (a) Octapeptide 17 and (b) Pentapeptide 30 structures and sequences.

The application of these antimicrobial agents has resulted in rapid cell death of Gram-positive and Gram-negative bacteria. The peptide was found to permeate both the outer and inner membranes of

bacteria, confirming similarities between the novel peptides and polymyxin B, which was found to behave similarly[47-48]. Scanning electron microscopy showed cell deformation and severely perturbed cell membranes after treatment with octapeptide O17, often leading to cell lysis. There appeared to be different mechanisms of action for the outer and inner membranes. The outer membrane of Gram-negative bacteria acts as a permeability barrier[49-50]. Using propidium iodide assays[51] Dr Sarojini's group found that O17 had similar activity to *Polymyxin B* in permeating inner bacterial membranes[47]. O17 was also found to be an effective agent against bacterial films, which are typically resistant to antibacterial agents. P30 was originally designed as a smaller active version of O17 and showed similar experimental antibacterial activity.

As the exact mechanisms of antibacterial peptide action and the nature of their interaction with the bacterial membranes are not fully understood, it is important to simulate antibacterial agents in biological context to gain insight. It would be of value to simulate these novel antibacterial peptides, especially in the presence of bacterial membranes, to better understand their dynamics and the nature of interactions with the bacterial membrane.

2.1.2 Bacterial Membrane Composition

Bacterial membranes often have complex compositions containing many different types of lipids and other chemicals[52]. Membrane composition varies greatly species to species, but most have a high density of negatively charged phospholipids[53]. Gram-positive bacteria have just one membrane, primarily comprised of phospholipids. Gram-negative bacteria have two separate membranes (*Figure 2.3*). The inner membrane is comprised of a phospholipid bilayer. The outer, asymmetric membrane contains an inner leaflet comprised of phospholipids and an outer leaflet of LPS[54-56], which leaves the membrane resistant to permeation by polar or hydrophobic large molecules[57-59]. The two membranes are separated by a periplasmic space. For the scope of the research we have selected two types of bacterial species to model – one from the Gram-positive bacteria *S. aureus* and one from the Gram-negative bacteria *E. coli*. These two species can be considered good model organisms, well studied in the literature, representative of both Gram-positive and Gram-negative bacteria and, therefore provide the best research opportunity for investigation due to the possible broadness of anti-microbial applications.

It should be noted that *in vivo* bacterial membrane composition varies significantly depending on the species of bacteria, as well as the current stage of cell cycle[60]. However, due to the limited size of MD simulations and lack of available parameter sets for the multiple molecules involved in the realistic membranes, simplified modelling schemes are often used. Fortunately, existing previously-tested models, were available to use for MD studies. Models kindly provided by Drs. Thomas Piggot and Syma Khalid[61] allow for simulation of both inner and outer membranes of both of the species without the need to create and test new types of cell membrane models, which would be out of scope of this PhD's work. These models have also been used in similar studies of Polymyxin B to show novel insights into the mechanistic interactions of lipopeptides and bacterial membranes[42]. Polymyxin B was found to insert

into the core of these membranes and to interact with the lipopolysaccharide in the asymmetric outer membrane of Gram-negative bacteria and insert into the core of the membranes generally. Therefore, with demonstrable results in similar studies, three types of membrane will be used in this project: *E. coli* inner, *E. coli* outer and *S. aureus*. The lipid composition of the membranes can be found below in *Table 2.1* and typical lipid structures in *Appendix 6*.

Table 2.1: Bacterial membrane model[61] composition by lipid type. Most lipids have one saturated palmitoyl tail and one mono-unsaturated oleoyl tail. Percentages are given by lipid number divided over total number of lipids in both leaflets of the bilayer.

E. Coli Inner membrane

Lipid	*Percentage*
Phosphatidylethanolamine (PE)	75.4%
Phosphatidylglycerol (PG)	19.7%
Diphosphatidylglycerol (DPG)	4.9%

E. coli outer membrane

Inner leaflet lipid	*Percentage*
Phosphatidylethanolamine (PE)	64.3%
Phosphatidylglycerol (PG)	3.6%
Diphosphatidylglycerol (DPG)	3.6%
Outer leaflet lipid	
LPS	28.5%

S. aureus membrane

Lipid	*Percentage*
Phosphatidylglycerol (PG)	56.8%
Phosphoglycero-lysine (PG-lys)	37.9%
Diphosphatidylglycerol (DPG)	5.3%

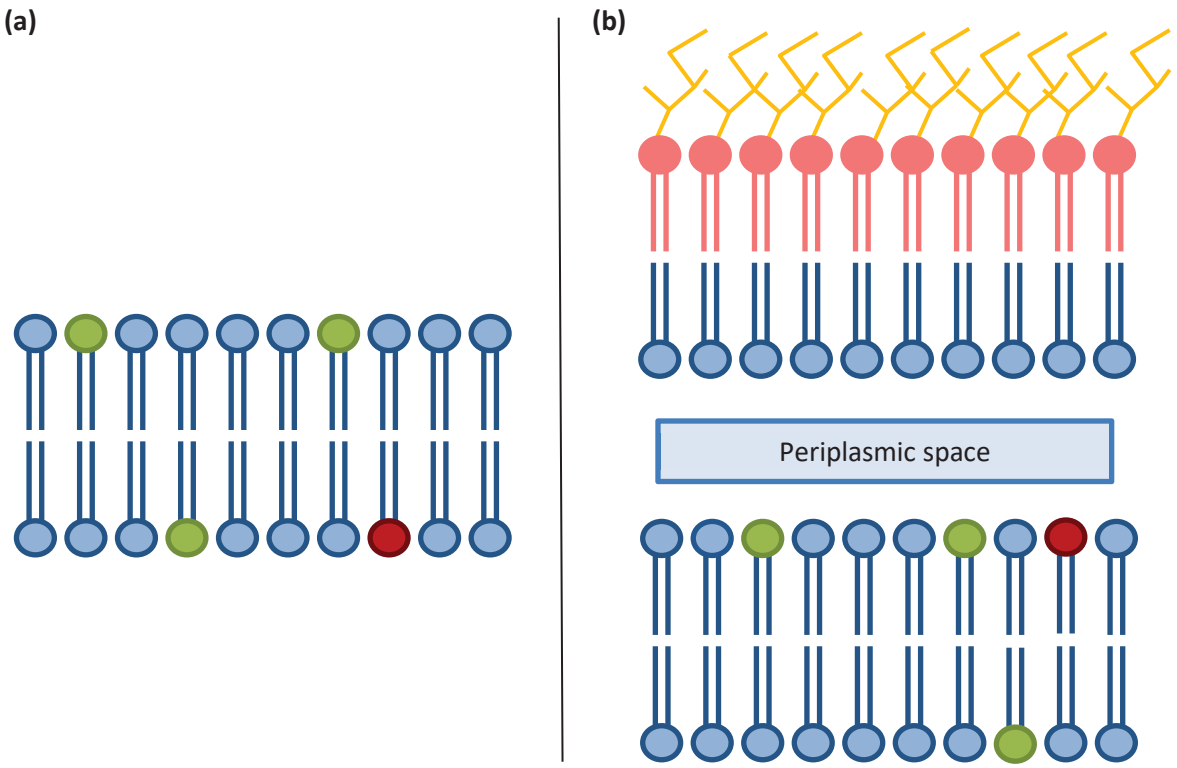


Figure 2.3: Schematic representation of (a) Gram-positive and (b) Gram-negative bacterial cell membranes. Gram-positive bacteria contain just one membrane comprised of a mixed phospholipid bilayer (blue, green and red lipids). Gram-negative bacteria contain two membranes, inner and outer, separated by the periplasmic space. The inner membrane is comprised of a mixed phospholipid bilayer (blue, green and red lipids) and the outer membrane is an asymmetric bilayer, containing a lower phospholipid leaflet (blue) and an outfacing upper leaflet with LPS (pink and yellow).

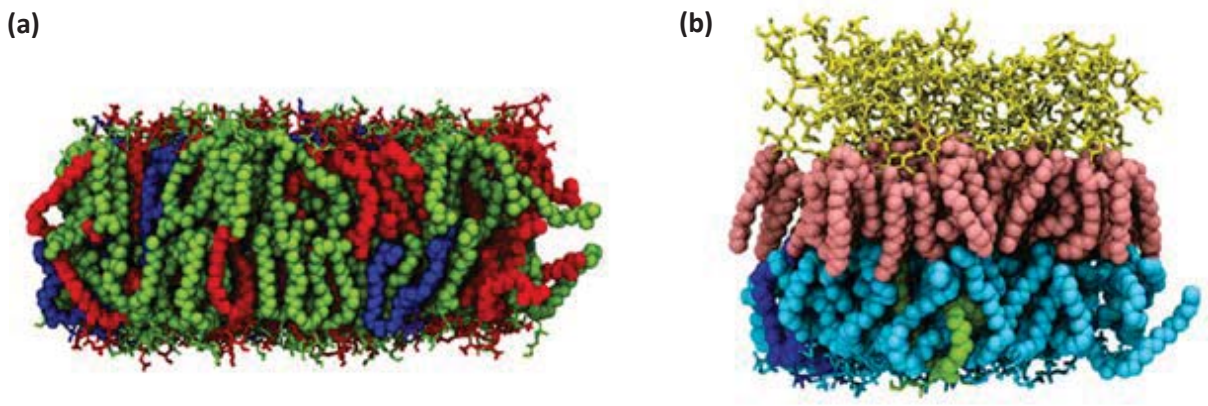


Figure 2.4: Structural snapshots of the membrane models, adapted from Piggot, T.J. et al. 2011[61]; (a) the typical structure of E. coli inner and S. aureus membranes; (b) the structure of E. coli outer membrane, which is asymmetrical and comprised of two leaflets (cardiolipin on top and phospholipid on the bottom).

2.2 Structural investigation of antimicrobial peptides via MD

The first goal of this work was to create models of P30 and O17 battacin peptide analogues and carry out MD simulations of each peptide alone, solvated in water to gain structural and dynamics information. Simulations were carried out at two different temperatures: 301 K (28 °C) and 310 K (37 °C). 301 K was used as it was the experimental temperature reported in the original publication and the higher 310 K temperature would approximate the temperature of the human body where bacterial infections occur. Differences in structure and dynamics between the two peptides and differences across the two different simulation temperatures would be investigated. Simulation data would be compared to experimental NMR data, gathered from peptides in solution.

The second goal was to simulate the peptides with accompanying bacterial membranes of *E. coli* (both inner and outer) and *S. aureus* to investigate their mutual interactions and gain insight into the mechanism of anti-microbial action. These simulations were also carried out at 301 and 310 K. The bacterial membranes were also simulated separately, without the presence of the peptides, to provide a reference.

2.2.1 System design

The MD simulations and subsequent analysis of O17 and P30 were carried out using the GROMACS 5.1.2 software. Two simulations were performed for each peptide at 301 K and 310 K. Further technical details of how the simulations were conducted are given in *Section 2.3: Methods*. The complete table of simulations performed is found below in *Table 2.2*.

Parameters for both peptides were constructed manually based on the existing GROMOS 54a7 force field parameters, due to their novel nature, according to the sequences presented above in *Fig. 2.2*. Firstly, the parameters for the unnatural Dab amino acid was created by adaptation of Leu parameters from the standard GROMOS 54a7 forcefield, due to its structural similarity. The two carbons and their associate bonded parameters in the middle of the Leu side chain were removed to create the Dab residue. No changes to charge distributions were required as the removed carbons both had zero charge. Secondly, to create D-Dab and D-Phe residues the improper dihedral term for the α-Carbon was reversed to create a mirror image geometry to the L- counterparts. Full parameter files and starting configurations for both peptides can be found in *Appendices 1* and *2*.

Initial peptide coordinates for both molecules were generated by Gaussview (v. 5.0.9). GROMACS compatible input files (parameters and equilibrated coordinates) for the bacterial membrane models were kindly provided by Prof. S. Khalid and Dr T. Piggot. The membrane lipid composition can be found in *Table 2.1*.

Table 2.2: List of simulations performed arranged by peptide, presence of membranes and temperature.

Peptide	*Membrane*	*Temperature (K)*
O17	–	301
O17	–	310
O17	*E. coli* inner	301
O17	*E. coli* inner	310
O17	*E. coli* outer	301
O17	*E. coli* outer	310
O17	*S. aureus*	301
O17	*S. aureus*	310
P30	–	301
P30	–	310
P30	*E. coli* inner	301
P30	*E. coli* inner	310
P30	*E. coli* outer	301
P30	*E. coli* outer	310
P30	*S. aureus*	301
P30	*S. aureus*	310

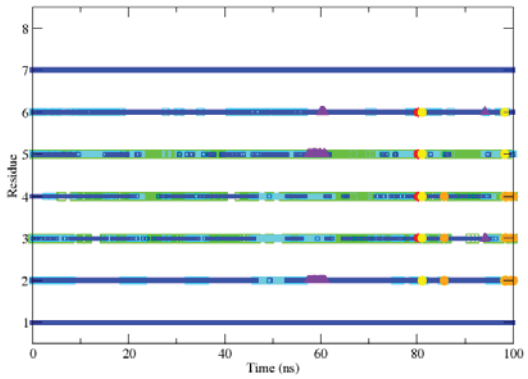
2.2.2 Internal structure

The internal structure of the O17 and P30 antibacterial peptides during the simulation was assessed using a variety of tools. Firstly, the database of secondary structure assignments (DSSP) program was used to analyse the types of secondary structures adopted by the peptide residues. Secondly, the root mean square deviation (RMSD) from the initial structure after equilibration and the per-residue root mean square fluctuation (RMSF) were calculated to assess the internal mobility.

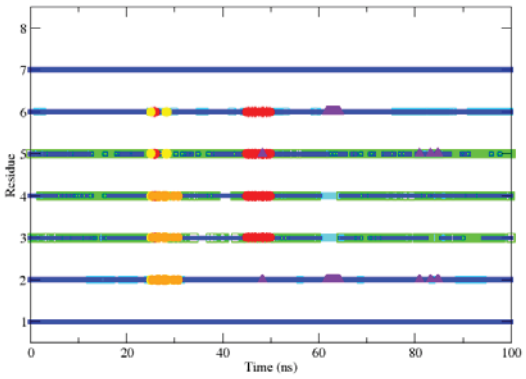
2.2.2.1 DSSP secondary structure

DSSP secondary structure analysis was performed for all peptides in all simulations to assess which types of secondary structures the peptides exhibited under different simulation conditions (*Figures 2.5 and 2.6*). The dihedral angles between members of a chain of residues are computed and based on the results several structure elements can be assigned: α-helix, π-helix, 3_{10}-helix, β-sheet, β-turn, turn, bend and coil. Coil structure indicates that the secondary structure could not be fitted into any of the aforementioned classes.

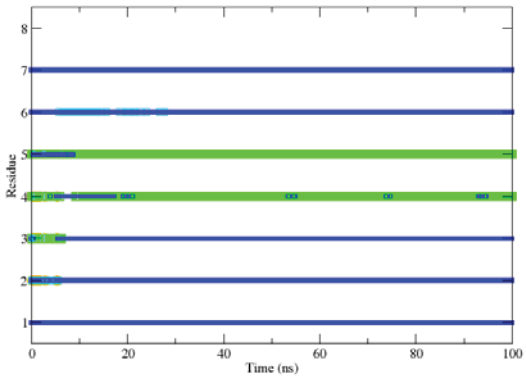
Octopeptide with water 301 K



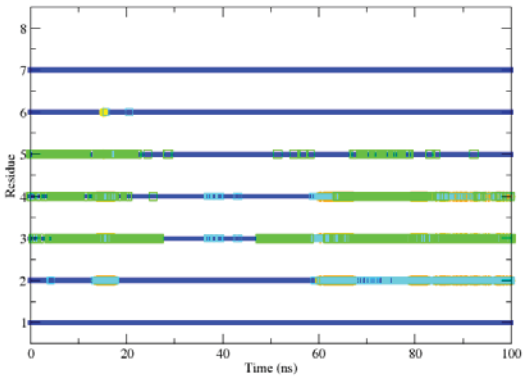
Octopeptide with water 310 K



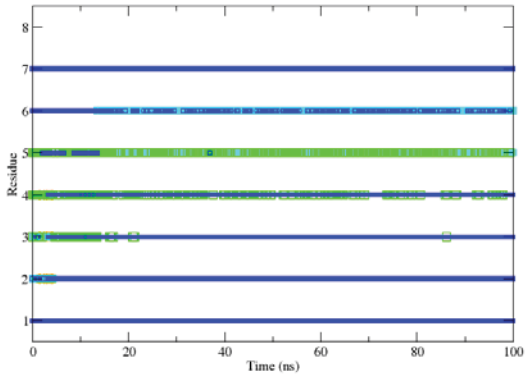
Octopeptide with E. coli inner membrane 301 K



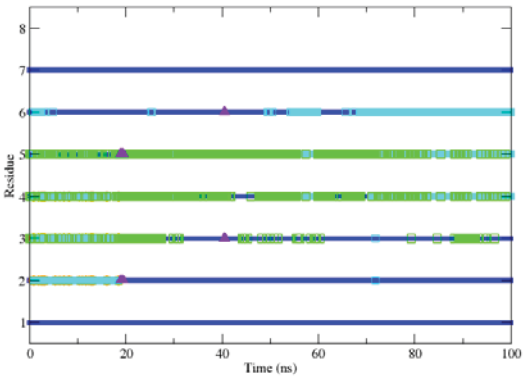
Octopeptide with E. coli inner membrane 310 K



Octopeptide with E. coli outer membrane 301 K



Octopeptide with E. coli outer membrane 310 K



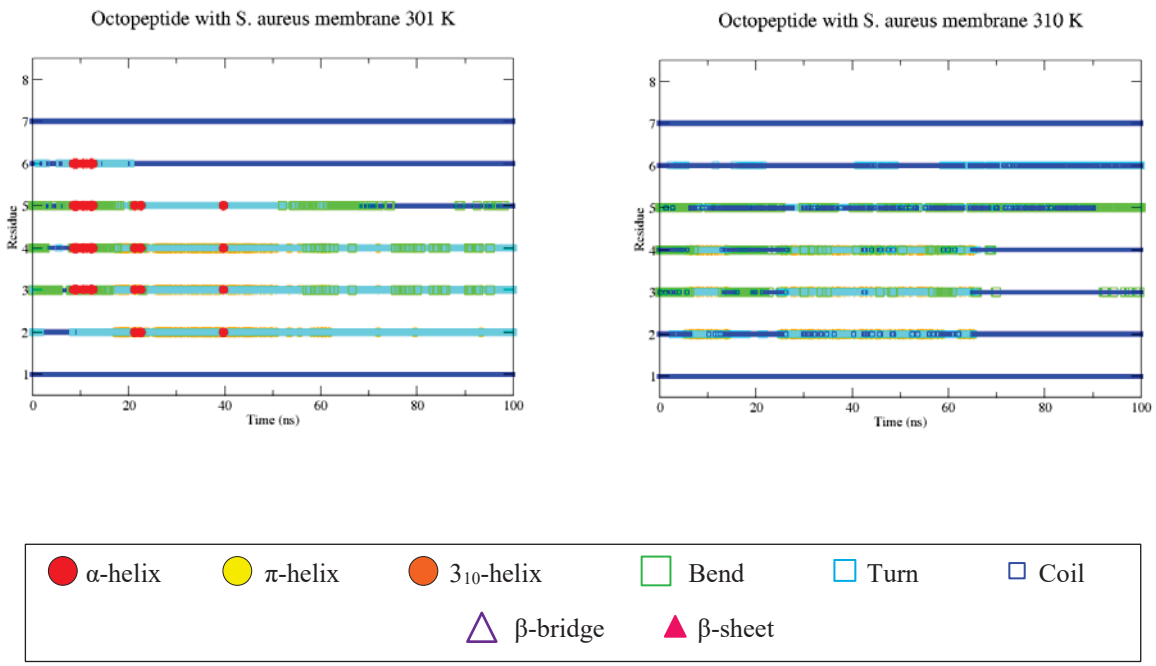
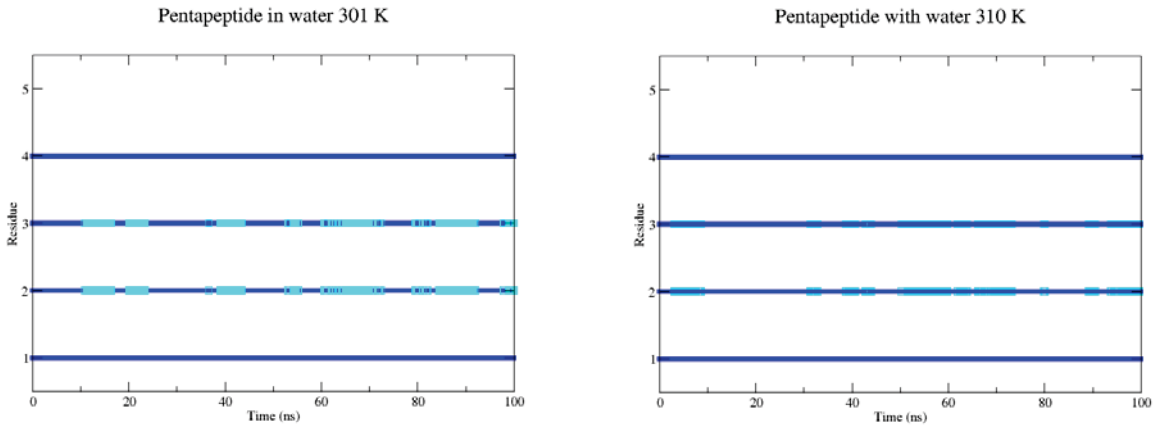


Figure 2.5: DSSP secondary structure analysis by residue of octapeptide O17 during simulations at 301 K and 310 K.



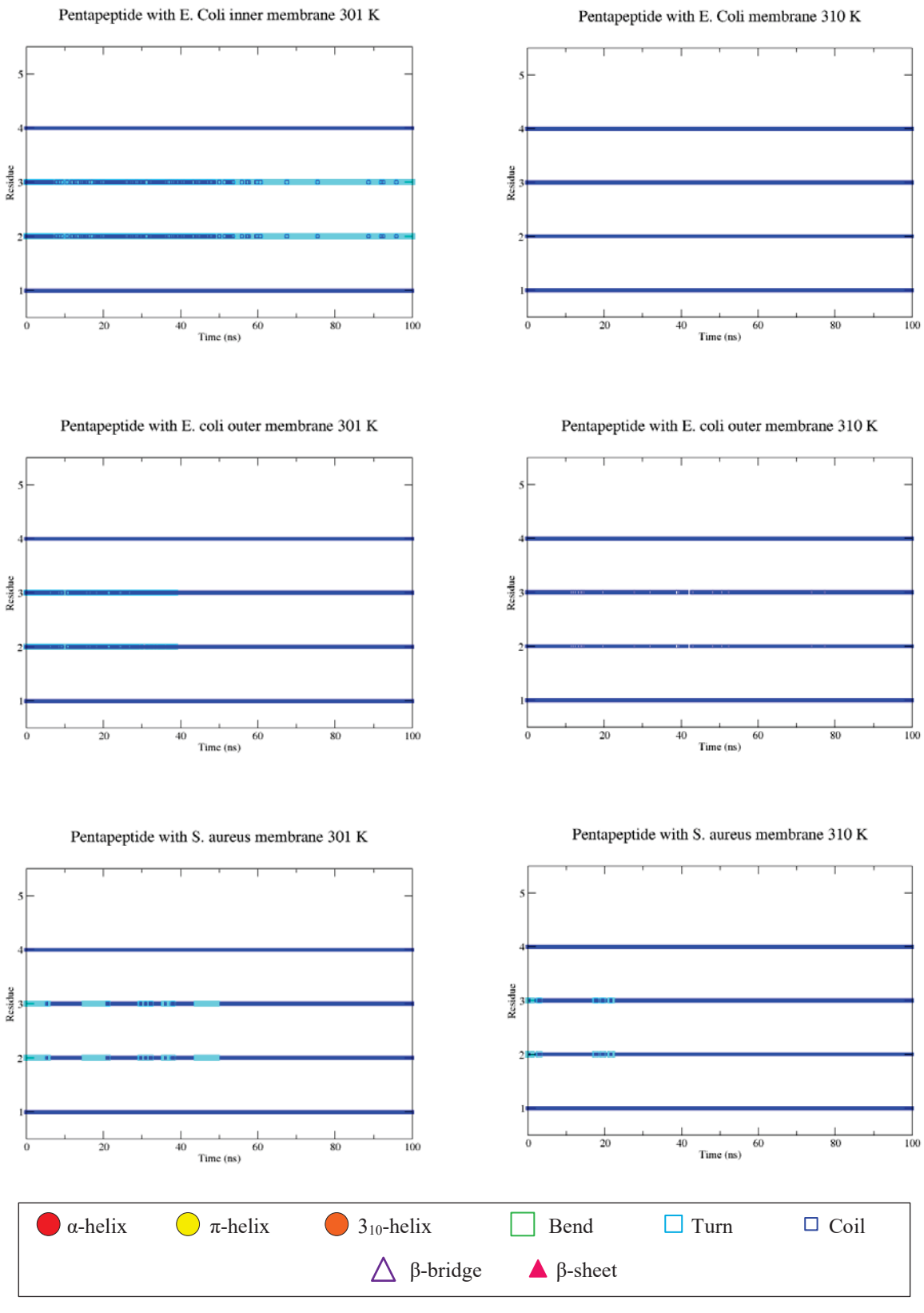


Figure 2.6: DSSP secondary structure analysis by residue of P30 during simulations at 301 K and 310 K.

Both peptides were found to adopt little secondary structure elements, predominantly falling into the DSSP category "coil", for nearly all the residues. Coil structure is assigned to a residue if none of the other structure classifications could be assigned. The "turn" type structure was also present in the middle of O17 and P30 to a lesser extent. This is unsurprising as the peptides are very short and it would be unlikely for them to form regular defined secondary structures, as these are stabilised by hydrogen bonding between non-adjacent residues. Variations of helical and β-sheet type structures were found in snippets of the simulations, but it is difficult to ascertain whether these can be attributed to sampling artefacts, where the local angles between units temporarily align in such a way as to give the detected structure result. The forcefield will typically favour the backbone dihedral angles that correspond to regular secondary structure elements, but without internal hydrogen bonding, regular secondary structure types like α-helices and β-sheets are unlikely to be stable for long periods of time. Furthermore, DSSP classification mechanism is computed based on dihedral angles and matching to Ramachandran plots. Typically, database matching occurs for all the normal encountered L- amino acid angle types, however for D- residues, significantly less matches will be found.

The incidence of different secondary structures was a lot higher for peptides in simulation without the membrane, indicating a degree of extra internal structural flexibility. Unfortunately, the last residue structure was un-diagnosable using the built in GROMACS DSSP tool, as neighbouring groups are required for dihedral angle calculations involved in the program.

2.2.2.2 RMSD (internal)

The atom-positional root mean square deviation from the initial structure was computed separately for each peptide to give a measure of their flexibility in water, as well as enabling us to determine the influence of simulation temperature and membrane presence on peptide flexibility (*Table 2.3* and *Figure 2.7*).

Table 2.3: Whole simulation averages of atom-positional RMSD of all atoms of each peptide unit mapped on to itself at 301 K and 310 K in nm with standard deviation for each graph included as σ.

Temperature	*in water*	*with E. coli inner*	*with E. coli outer*	*with S. aureus*	*σ*
Octapeptide 301 K	0.503	0.473	0.487	0.556	0.070
Octapeptide 310 K	0.484	0.432	0.546	0.490	0.073
Pentapeptide 301 K	0.386	0.160	0.323	0.468	0.073
Pentapeptide 310 K	0.415	0.251	0.496	0.475	0.060

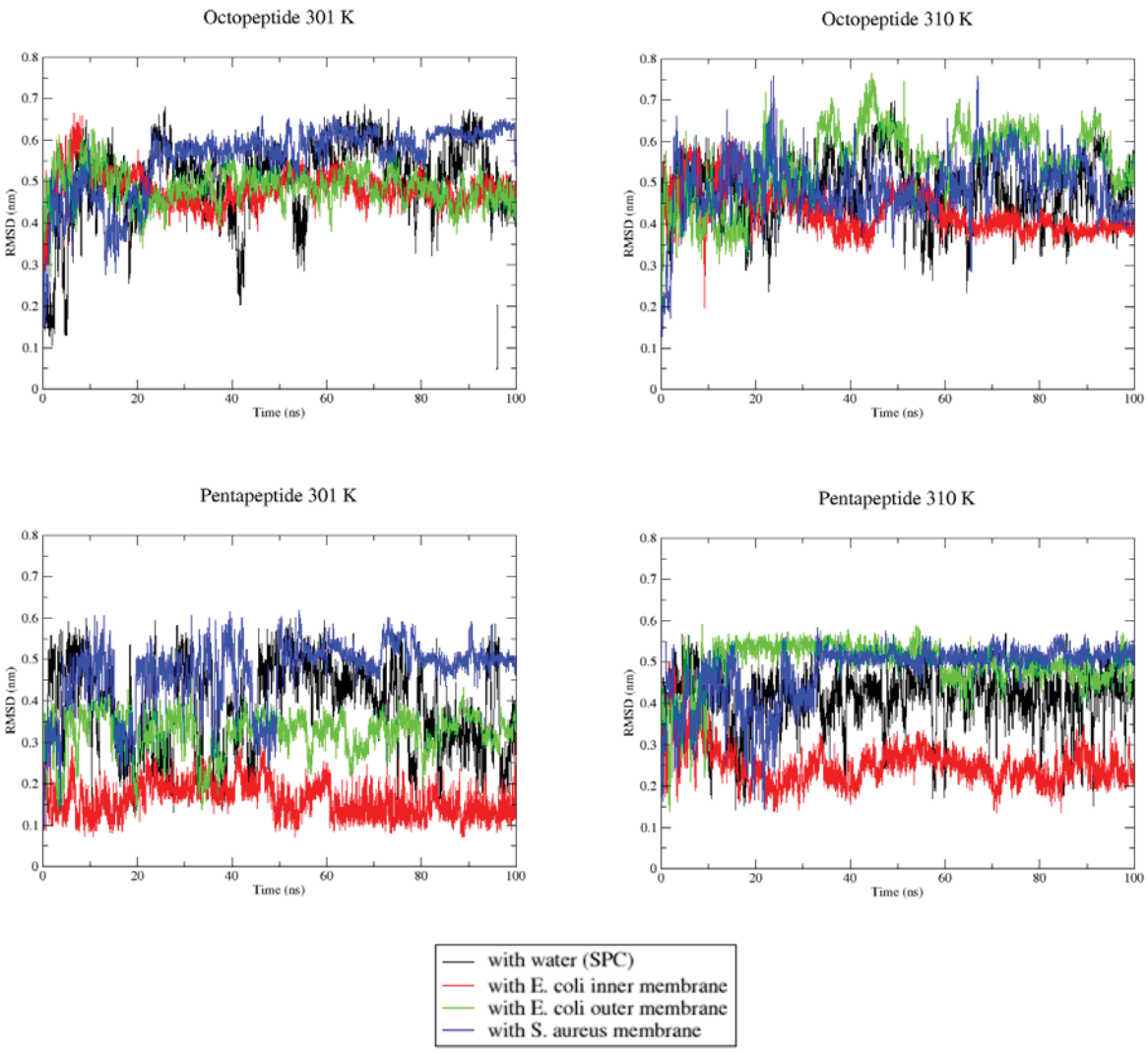


Figure 2.7: Atom-positional RMSD of all atoms of each peptide unit mapped on to itself at 301 K and 310 K.

O17 simulated with SPC water solvent only had RMSD values between 0.3 and 0.6 nm, generally indicating a small degree of internal flexibility. When simulated with the range of bacterial membranes, O17 RMSD values remained similar throughout most of the simulations, with water and *S. aureus* simulations having slightly higher values then that of peptides with both of the *E. coli* membranes. Peptide internal mobility was lower in simulations with the *E. coli* inner membrane, and *E. coli* outer membrane, indicating a reduction in internal flexibility. The differences between the simulations were more pronounced in the P30 simulations, which have a much larger spread of RMSD values from 0.1 to 0.6 nm. P30 RMSD results were affected to a greater extent by the presence of differing membranes, than the O17 results. Simulations of *E. coli* inner membrane with P30 resulted in the lowest RMSD

averages exhibited across all experiments. There was little difference between the octapeptides simulated at different temperatures. P30 at 310 K exhibited higher RMSD values than P30 at 301 K.

2.2.2.3 RMSF

The per-residue root mean square fluctuation (RMSF) has been calculated for all the peptide simulations to localise and quantify internal flexibility (*Figure 2.8*).

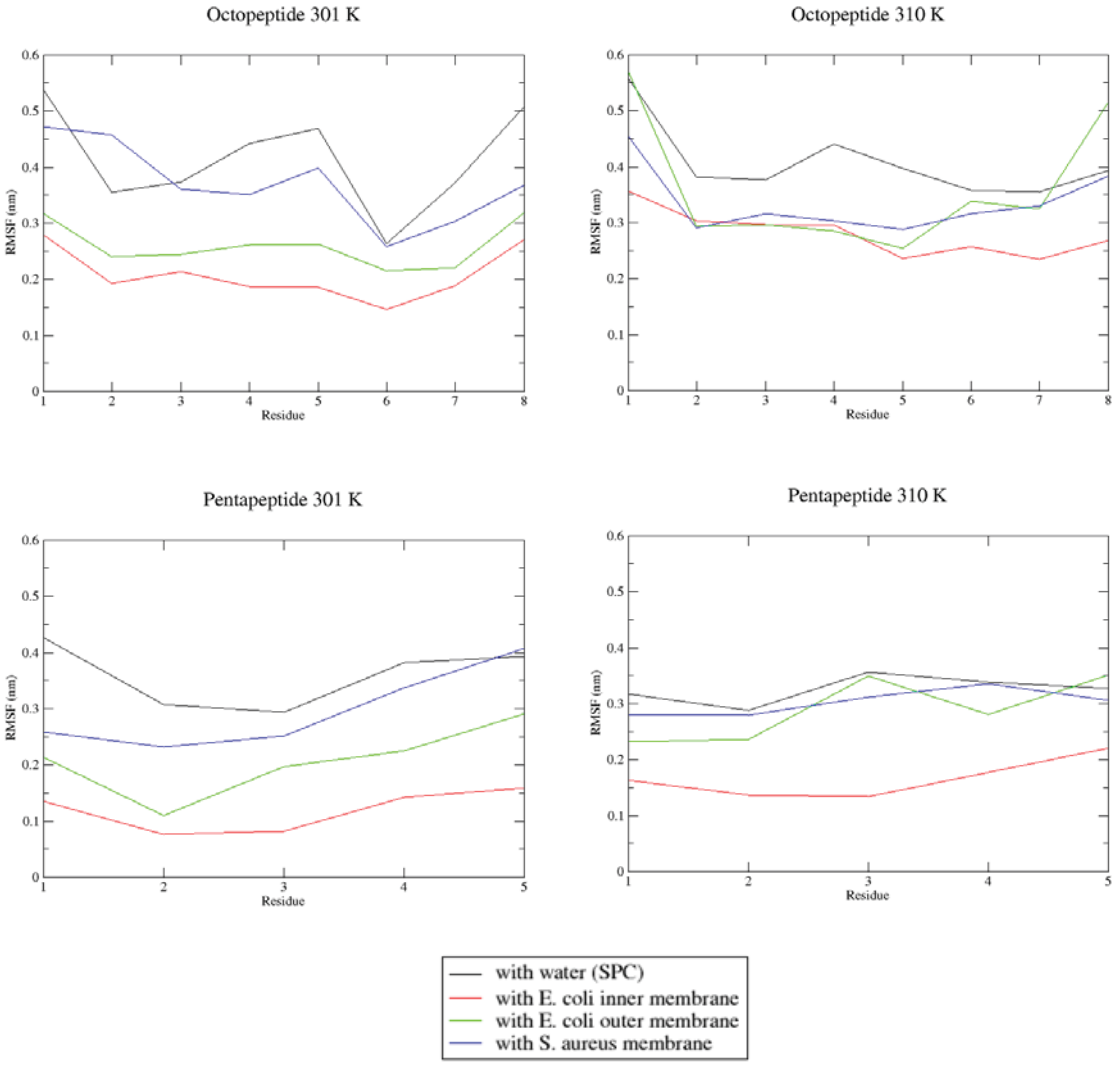


Figure 2.8: RMSF of peptide simulations at 301 K and 310 K.

RMSF values varied across the simulation series depending on the simulation environment of the peptides. RMSF values were highest for peptides simulated in solvent only, indicating greater mobility.

RMSF values were lowest for the simulations with the *E. coli* inner membrane, indicating reduced mobility, possibly as a result of interactions. The simulations with the other two membrane types were found to be in the middle of the previous two results, suggesting that the presence of a membrane in general restricted the internal mobility of both peptides, potentially due to peptide-membrane interactions. Peptide residues were found have generally similar RMSF values for a given simulation, indicating relatively uniform mobility throughout the peptide, matching previous DSSP results indication the lack of regular secondary structure. Higher RMSF values were found toward the ends of the peptides, indicating a slight increase of mobility towards the end residues. The exception being P30 at 310 K, where the N-terminal residues have similar RSMF values to the remainder of the peptide. Possible explanations include increased interaction at the ends of the peptide, thus restricting movement. A narrower range of RMSF values was observed for P30 simulations at 310 than at 301 K, indicating similar behaviour irrespective of environments at the higher temperature. Otherwise, no significant differences between the different simulation temperatures were observed.

2.2.2.4 NMR J-value estimation

Estimations of $^{3}J_{NH\text{-}C^{\alpha}H}$ coupling values were performed for the O17 in water and could be directly compared to experimental data provided, providing a measure of model quality (*Table 2.4*). $^{3}J_{NH\text{-}C^{\alpha}H}$-coupling values are calculated from the φ backbone torsional dihedral angle values using the Karplus relation and give a direct comparison to NMR experimental data, serving as a useful tool in model validation.

Table 2.4: $^{3}J_{NH\text{-}C^{\alpha}H}$-coupling comparison between O17 experimental data (278 K in 70% H_2O and 30% TFE, or 100% CD_3OH) and values calculated from simulations at 301 and 310 K. Error = +/- 0.05 nm

Residue	*Experimental 70% H_2O, 30% TFE 278 K*	*Experimental 100% CD_3OH 278 K*	*Estimate 301 K*	*Percentile difference 301 K**	*Estimate 310 K*	*Percentile difference 310 K**
1 D-Dab	6.6	*6.1*	*n/a*	*n/a*	*n/a*	*n/a*
2 Dab	7.0	7.1	5.7	- 18.6%	5.8	- 17.1%
3 Dab	7.6	7.5	6.0	- 21.1%	5.9	- 22.4%
4 Leu	7.0	7.0	6.8	- 2.9%	6.8	- 2.9%
5 D-Phe	7.0	6.9	5.0	- 28.6%	4.8	- 28.6%
6 Dab	7.6	8.0	6.0	- 21.1%	6.2	- 18.4%
7 Dab	7.3	7.9	6.2	- 15.1%	6.3	- 13.7%
8 Dab	7.2	7.4	5.9	- 18.1%	5.8	- 19.4%

**Differences between estimated and experimental data are expressed as a percentage of the 70% H_2O, 30% TFE experimental NMR data for convenience. Experimental data sourced from De Zoysa, G. H., et al 2015 paper and supplementary information[33].*

The estimated $^{3}J_{NH\text{-}C^{\alpha}H}$ values for O17 were of similar magnitude compared to experimental values. Deviations are typically in the range of 1.5 Hz (or 20% difference), which is reasonable for simulation estimates, especially considering the NMR was taken at a different temperature and in a different solvent system, containing 30% TFE. Results in CD_3OH were also included in the table above for comparison and show a greater degree of minor deviation from the simulated values. Best fit values were found to be for Leu, which is the only natural amino acid in the sequence. Deviations of Dab residues were in the range of 13.7% to 28.6%. The largest deviation from experimental values can be observed for the D-Phe residue of 28.6%. This can be explained by the unnatural nature of the Dab acids and for the D-Phe case the D- nature of the acid. The program used to compute dihedral angles, sought to compute torsional dihedral angles and approximate J-values by the Karplus equation. The nature of the D- amino acids reverses the dihedral around the α-Carbon, making it impossible for the program to match the expected result to the L-acid angle.

Unfortunately, due to the nature of the program, it was not possible to calculate the J-values for the first residue, as the program could not find the initial reference point for calculating the residue one dihedral angle, as a result of peptide numbering in the trajectories.

2.2.3 Interactions

The relative mobility of the peptides with respect to the membrane was given by the RMSD of the peptides fitted to all non-water atoms in the starting configuration of the main simulation run. The minimum distance between the peptides and the bacterial membranes was calculated to assess whether the peptides came into contact with the bilayer. Hydrogen bonding analysis was also performed to assess the nature of any inter-molecular interactions that formed. The interactions of the peptides with the bilayer were also visualised.

2.2.3.1 RMSD (relative mobility)

Table 2.5: Averages of atom-positional RMSD of all atoms of each peptide unit mapped on to all non-water atoms in the simulations at 301 K and 310 K in nm with standard deviation for each graph included as σ.

Temperature	*with E. coli inner*	*with E. coli outer*	*with S. aureus*	*σ*
Octapeptide 301 K	5.521	10.691	7.747	1.20
Octapeptide 310 K	6.020	0.546	8.932	1.40
Pentapeptide 301 K	4.103	7.751	8.967	1.29
Pentapeptide 310 K	5.604	10.829	7.802	1.21

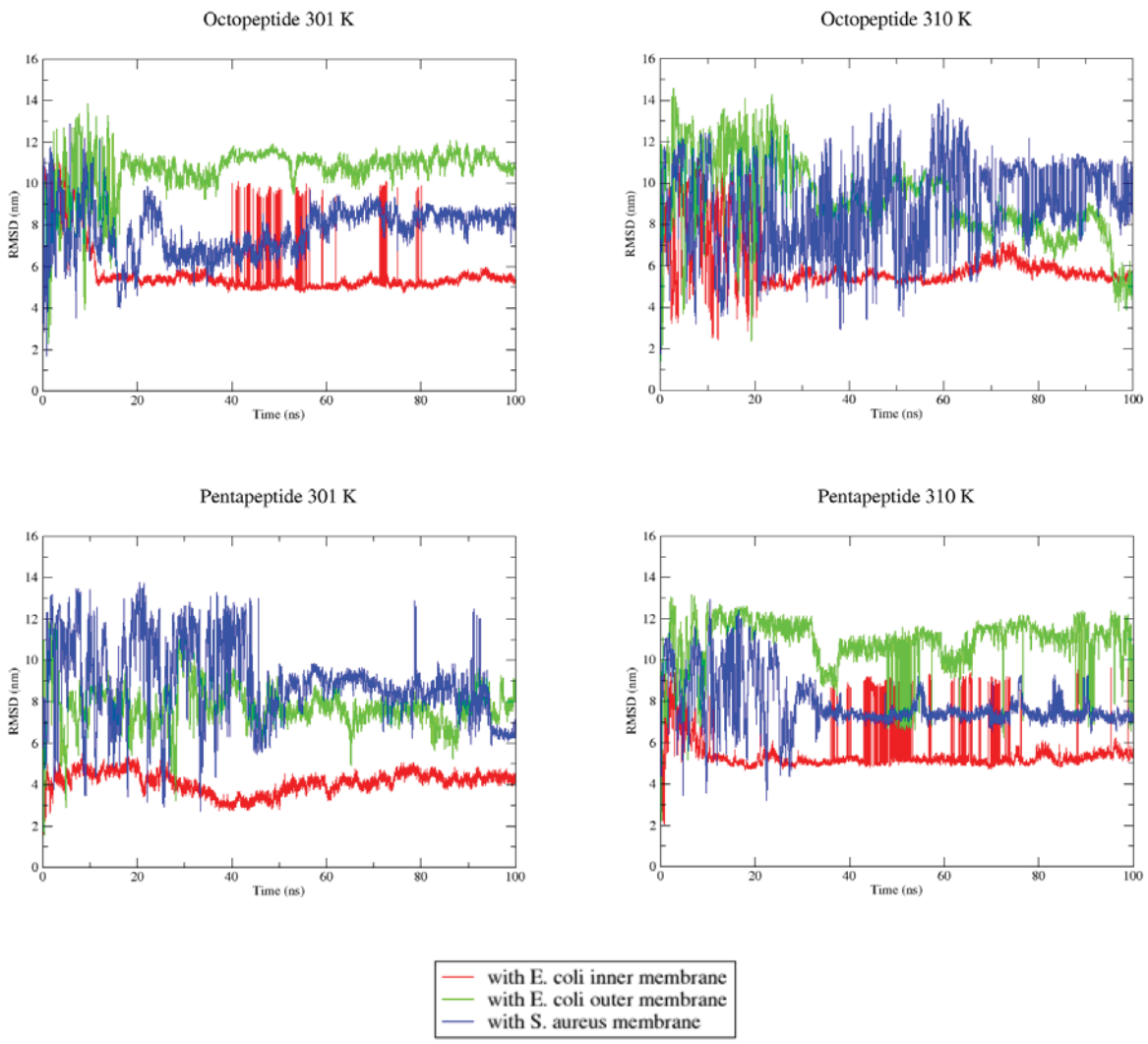


Figure 2.9: Atom-positional RMSD of all atoms of each peptide unit mapped on to all non-water atoms in the simulations at 301 K and 310 K.

Peptides were found to be highly mobile relative to the membrane (*Figure 2.9* and *Table 2.5*). As expected, the RMSD values were found to be generally higher at the higher simulation temperature, but the differences was not too great. The peptides were found to be less mobile when simulated with *E. coli* inner membrane. The RMSD averages for the other two membranes were of similar magnitudes, with the exception of two simulations of peptides with the *E. coli* outer membrane where the RMSD values were over 10 nm. It should be noted that in some cases RMSD values remained consistent, while in others they fluctuated a lot, which could be associated with periodicity issues inherent in simulating multi-component systems. The spike in that case would be indicative of the peptide molecule crossing a periodic simulation boundary.

2.2.3.2 Minimum distance

The minimum distance between the peptides and their respective bacterial membranes was calculated throughout each simulation to monitor any approach of the peptide to the bilayer surface (*Figure 2.10*). For simulations with the *E. coli* outer membrane, distances from the peptide were calculated both for the LPS and the phospholipid leaflets of the membrane separately to determine which leaflet the peptide approached, as both options were possible due to the periodic boundary conditions.

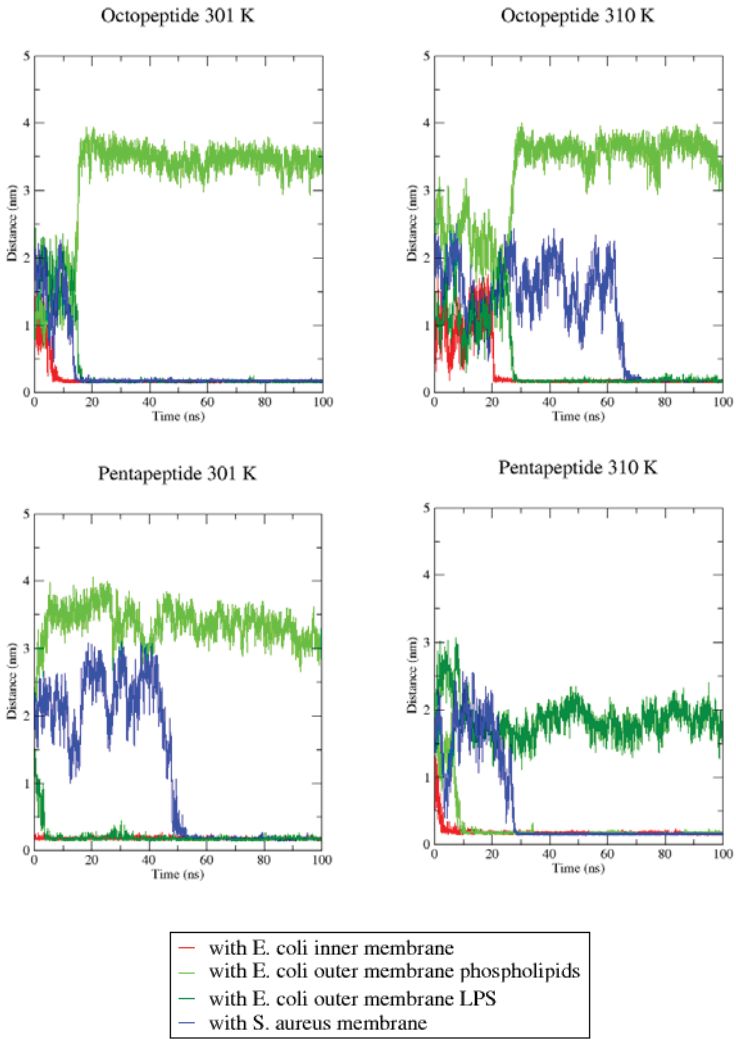


Figure 2.10: Graphs of minimum distance between peptides and the respective lipid bilayers in simulations at 301 K and 310 K.

In all membrane simulations, the peptides were found to approach the membrane and remained in close proximity thereafter. The peptides were found to approach the *E. coli* membranes relatively quickly, within 20 ns for all simulations. In contrast, approach to *S. aureus* membrane was the slowest,

as in three out of four simulation the peptide-membrane approach does not occur until after 30, 50 and 65 ns.

In the simulations with *E. coli* outer membrane, the peptide-membrane distances for to the two leaflets were monitored independently. In three out of the four simulations, the peptides approached the LPS part of the membrane and only in the pentapeptide simulation at 310 K did the peptide interact with the phospholipid part of the membrane, indicating strong preference of peptides for the LPS leaflet. This evidence supports earlier literature claim that LPS is the initial target of the peptide.

2.2.3.3 Hydrogen bonding with solvent

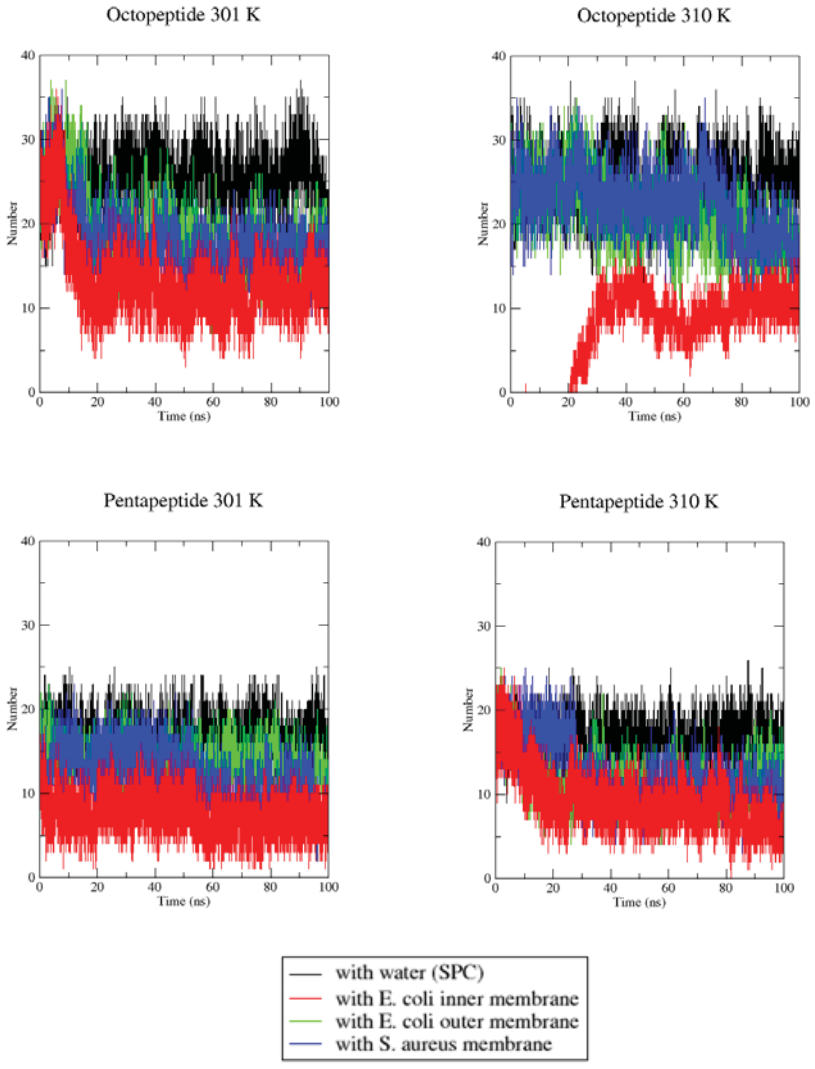


Figure 2.11: Number of hydrogen bonds formed between the antibacterial peptides and the water solvent during simulations at 301 K and 310 K.

Table 2.6: Average number of hydrogen bonds formed between the antibacterial peptides and the water solvent during simulations at 301 K and 310 K.*

Simulation	*Average*	*Pre interaction average*	*Post interaction average*
O17 with SPC 301 K	25.36	n/a	n/a
O17 with SPC 310 K	25.22	n/a	n/a
O17 with E coli inner 301 K	13.46	25.23	12.36
O17 with E coli inner 310 K	7.50	0.00	9.49
O17 with E coli outer 301 K	19.09	25.65	17.65
O17 with E coli outer 3101 K	21.09	24.54	20.00
O17 with S aureus 301 K	18.17	24.13	16.95
O17 with S aureus 3101 K	22.28	23.60	19.94
P30 with SPC 301 K	16.93	n/a	n/a
P30 with SPC 310 K	16.67	n/a	n/a
P30 with E coli inner 301 K	7.84	n/a	n/a
P30 with E coli inner 310 K	9.34	17.92	9.03
P30 with E coli outer 301 K	13.90	16.65	13.80
P30 with E coli outer 3101 K	11.84	16.85	11.29
P30 with S aureus 301 K	12.34	14.00	11.68
P30 with S aureus 3101 K	12.69	1.69	10.87

**Averages are given for the whole simulation length, as well as pre- and post- interaction with the bacterial membrane where interaction was based on minimum distance analysis.*

The number of hydrogen bonds between the peptides and the surrounding solvent was calculated throughout each simulation (*Figure 2.11)*. Both peptides formed a large number (7 to 25 on average) of hydrogen bonds with surrounding solvent molecules. The maximum number of hydrogen bonds was approximately 35 for O17 and 25 for P30, which roughly equated to 4.4 and 5 bonds per residue, respectively.

It was observed that after interaction with bacterial membranes (assessed by minimum distance analysis), the average number of solvent bonds dropped significantly (*Figure 2.10 and Table 2.6*). It can be inferred that this occurred when the peptides approached the membrane and began to interact with the lipid bilayer, switching out hydrogen bonds with the solvent for hydrogen bonds with the bilayer, showing preferential hydrogen bonding affinity for the bilayer.

The extent of reduction of solvent hydrogen bonding varied between the different membranes simulated. The number of solvent hydrogen bonds was lowest for the *E. coli* inner membrane simulations and approximatly equal for the other two membrane types. A notable exception was found

in the case of O17 with *E. coli* inner membrane at 310 K, where no hydrogen bonds with the solvent were observed in the first stages of the simulation, which was anomalous.

2.2.3.4 Hydrogen bonding with bilayer

Analysis was also carried out to establish the extent of hydrogen bonding between the peptides and the bacterial membranes (*Figure 2.12* and *Table 2.7*).

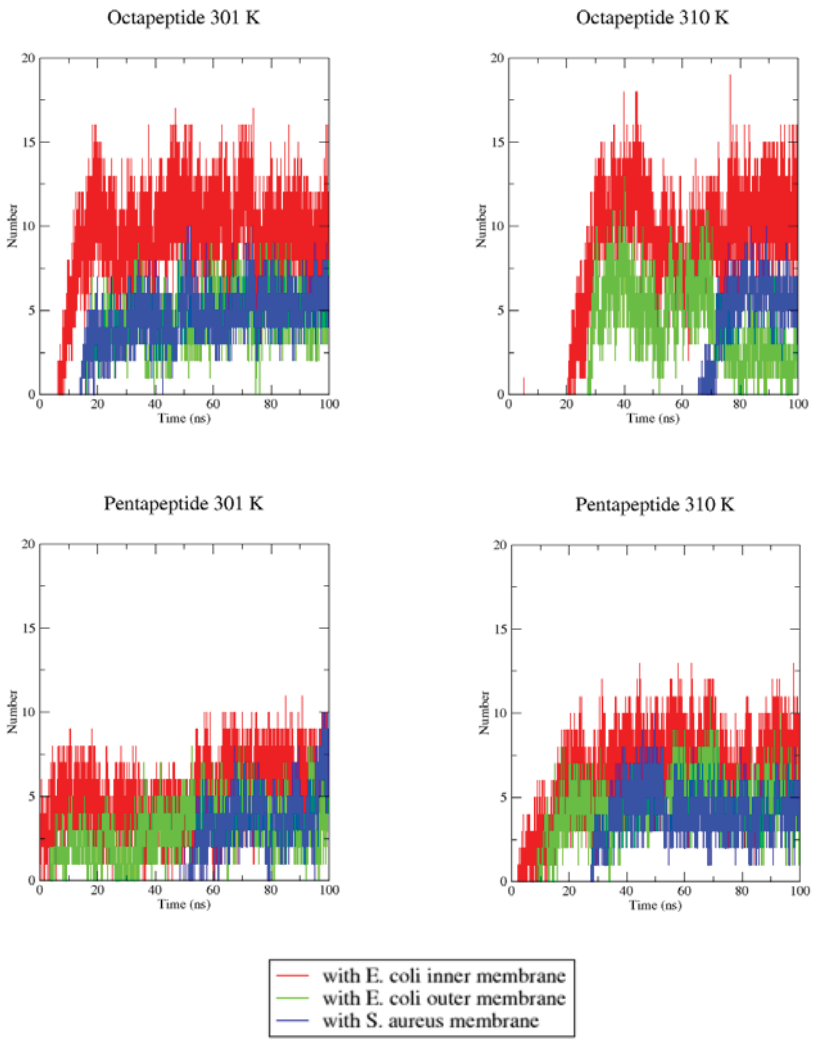


Figure 2.12: Number of hydrogen bonds formed between the peptides and the bacterial membranes at 301 K and 310 K.

Table 2.7: Averages numbers of hydrogen bonds formed between the antibacterial peptides and the lipid membrane during simulations at 301 K and 310 K.*

Simulation	*Average*	*Pre interaction average*	*Post interaction average*
O17 with E. coli inner 301 K	8.88	0.30	9.68
O17 with E. coli inner 310 K	7.50	0.02	9.49
O17 with E. coli outer 301 K	3.88	0.26	4.67
O17 with E. coli outer 310 K	3.19	0.00	4.20
O17 with S. aureus 301 K	3.96	0.33	4.71
O17 with S. aureus 3101 K	1.68	0.00	4.67
P30 with E. coli inner 301 K	4.97	n/a	n/a
P30 with E. coli inner 310 K	5.96	0.13	6.18
P30 with E. coli outer 301 K	2.89	0.00	2.99
P30 with E. coli outer 3101 K	4.08	0.11	4.52
P30 with S. aureus 301 K	1.80	0.00	2.52
P30 with S. aureus 3101 K	2.93	0.03	4.09

**Averages are given for the whole simulation length, as well as pre- and post- interaction with the bacterial membranes, where interaction was based on minimum distance analysis.*

Corroborating the peptide-membrane minimum distance results, it was found that peptides formed hydrogen bonds with membranes approximately at the time when the peptide-membrane distance was reduced. It should be noted that in the simulation of P30 with *E. coli* outer membrane at 310 K, the bonding occured in between P30 and the phospholipid leaflet of the membrane. In all other simulations of the *E. coli* outer membrane, the peptides have preferentially interacted with the LPS leaflet of the membrane instead.

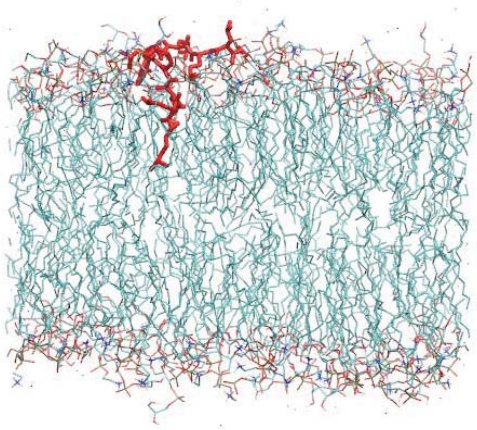
The number of peptide-membrane hydrogen bonds was found to be highest for the simulations with the *E. coli* inner membrane, with the maximum number reaching 19 for the O17 and 13 for the P30 simulations. For the O17 simulations the number of hydrogen bonds to the *S. aureus* membrane was much smaller, reaching a maximum of 10. For the P30 simulation at 301 K for both *E. coli* membranes were similar, but at 310 K the number of peptide-membrane hydrogen bonds formed during the simulation with the *S. aureus* membrane was smaller again. This could be related to the positively charged nature of the *S. aureus* membrane, causing a greater degree of repulsion between the peptides and the membrane, compared to the *E. coli* membranes. The number of hydrogen bonds formed between P30 and the phospholipid component of the *E. coli* outer membrane was found to be similar to the number of hydrogen bonds with the *S. aureus* membrane. All residues were observed to form hydrogen bonding pairs with the membrane lipids (see *Appendix 3* for Hydrogen bond pairs).

These results corroborate the previous analysis of loss of hydrogen bonding with the solvent, indicating that the peptides preferentially bind the membrane over the solvent. This interaction was strongest with the *E. coli* inner membrane by number of hydrogen bonds formed. The drops in peptide-solvent hydrogen bonds corresponded approximately to the number of new hydrogen bonds formed between the solvent and the peptide.

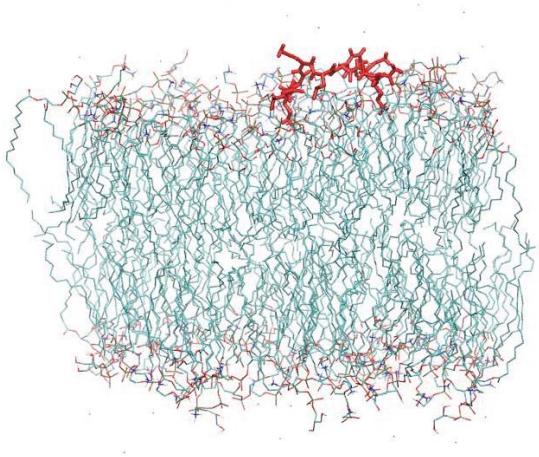
2.2.3.5 Visual analysis

A visualisation of peptides at the points of interaction was performed using VMD to assess the structures and conformations adopted during the simulation, and in particular once the peptides begun interacting with the membranes.

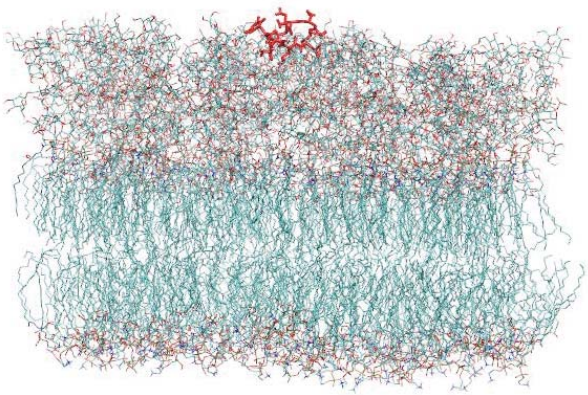
Octapeptide with E. coli inner membrane 301 K



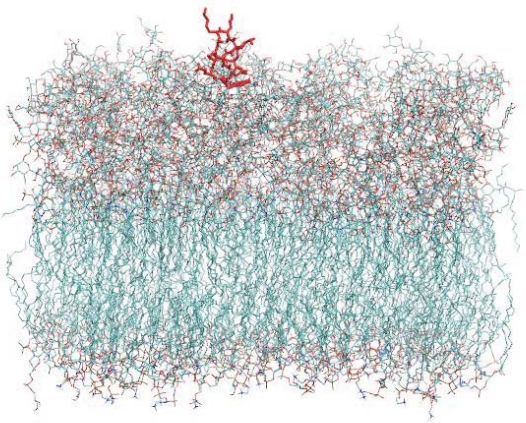
Octapeptide with E. coli inner membrane 310



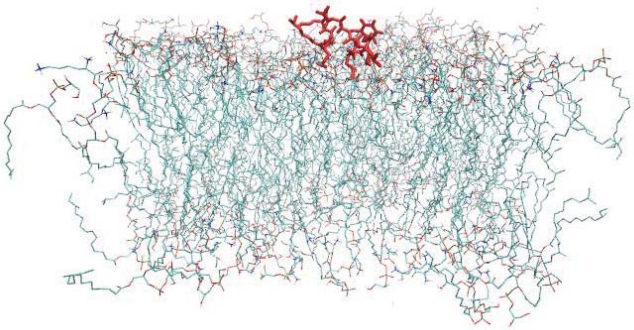
Octapeptide with E. coli outer membrane 301 K



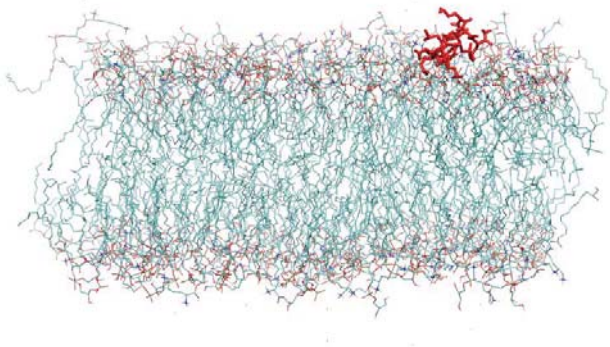
Octapeptide with E. coli outer membrane 310 K



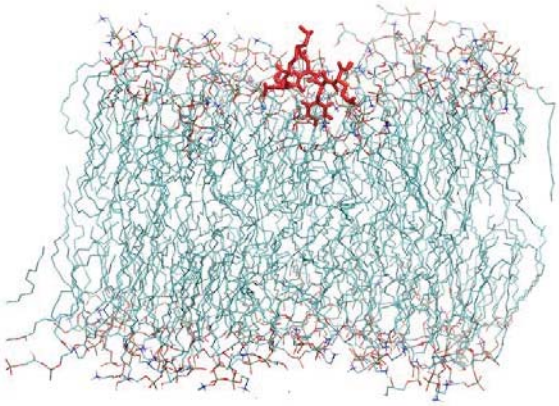
Octapeptide with S. aureus membrane 301 K



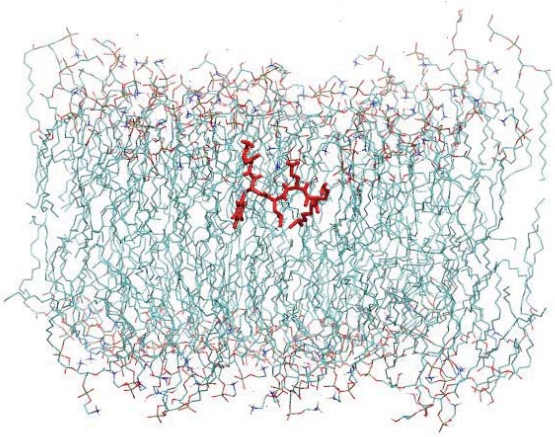
Octapeptide with S. aureus membrane 310 K



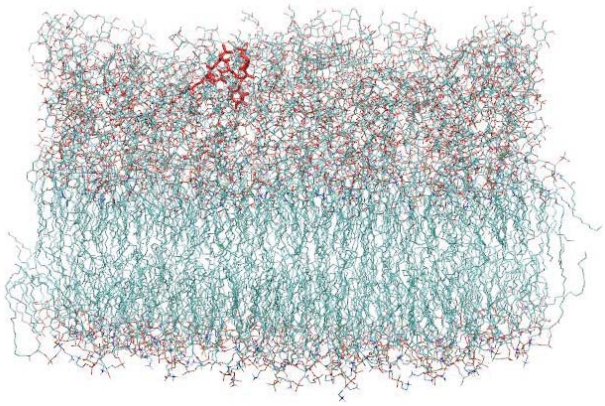
Pentapeptide with E. coli inner membrane 301 K



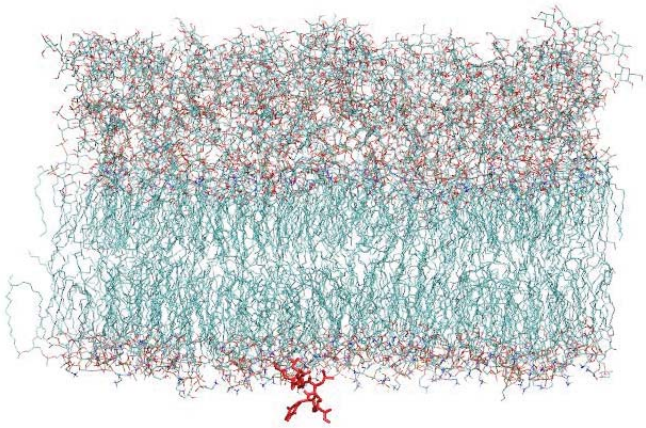
Pentapeptide with E. coli inner membrane 310 K



Pentapeptide with E. coli outer membrane 301 K



Pentapeptide with E. coli outer membrane 310 K



Octapeptide with S. aureus membrane 301 K

Octapeptide with S. aureus membrane 310 K

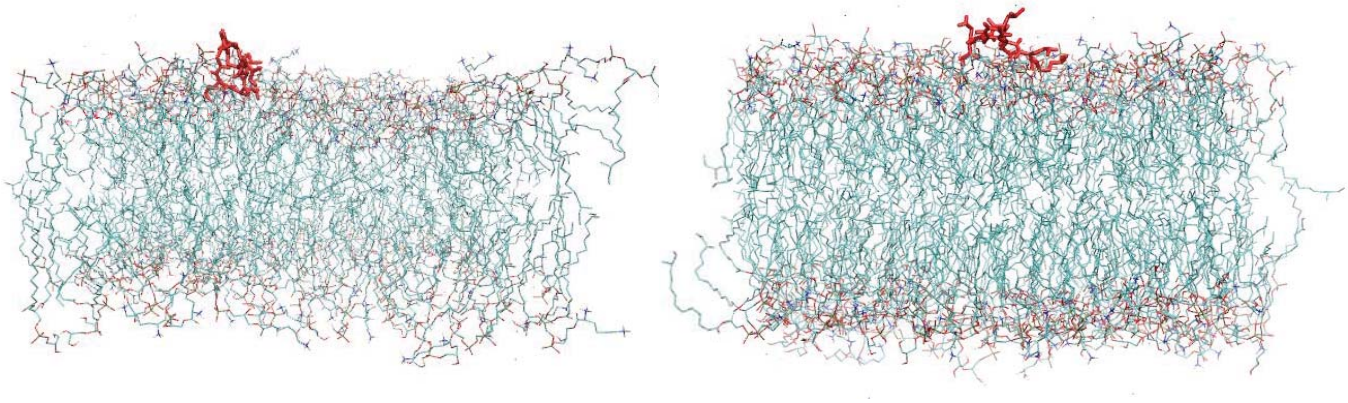


Figure 2.13: Snapshots of each peptide and the respective bacterial membranes at the end of the 100 ns simulations. Peptides are drawn in red, and the lipids are coloured according to atom type (carbon: cyan; oxygen: red; nitrogen: blue, phosphorus: orange). Water molecules are omitted for clarity.

From the visual analysis of the peptides at the end of the simulation runs, it was observed that all of the peptides were resting on the surface or permeating the bacterial membranes (*Figure 2.13*). The visual data confirmed interactions suggested by previous minimum distance and hydrogen bonding analysis. A number of interesting cases were found and are outlined below.

In the case of the O17 simulation with the *E. coli* inner membrane at 301 K, it was found that fatty acid carbon chain has partially permeated the bilayer, lending evidence to the insertion mechanism outlined in the in *Section 2.1.1*. P30 in simulation with the *E. coli* inner membrane at 310 K has become completely inserted into the membrane. P30 in simulation with the *E. coli* outer membrane at 301 K was the only peptide to have become inserted into the LPS leaflet of the outer membrane, indicating that a process of permeation has begun. P30 in simulation with the *E. coli* outer membrane at 310 K was the only peptide to approach the membrane from the other side through periodic boundary conditions. Peptides interacted with the membranes at both of the simulated temperatures, indicating that the peptides would be effective anti-bacterial agents at both of the temperatures.

Further snapshots of the simulation trajectories are outlined below in *Figures 2.14, 2.15, 2.16 and 2.17*. Outlining stepwise insertion of the peptides into the membrane. In the cases of P30 and O17 interaction with *E. coli* inner membrane it can be observed that at first the peptide rests on the surface of the membrane and hydrogen bonds between Dab2 along with several other functionalised amino acids and the membrane become established. Then the fatty acid carbon chain begun to permeate the lipid bilayer, seeking hydrophobic inner region with the lipid tails, eventually leading to the rest of the molecule permeating the bilayer. The mechanism for the other two types of membrane appeared to be similar in its initial stages, albeit it proceeded at a slower rate and did not reach conclusion. O17 became embedded in the LPS leaflet of the *E. coli* outer membrane but did not permeate as deeply as in the case of the inner membrane. The fatty acid tail has worked its way into the lipid tails of the LPS layer, possibly

guided by hydrophobic interactions, while the rest of the peptide stayed amidst the glycan units of the upper part of the LPS leaflet. With *S. aureus,* the interaction was even less advanced, as the peptide just coiled on the membrane surface, forming hydrogen bonds to the lipids.

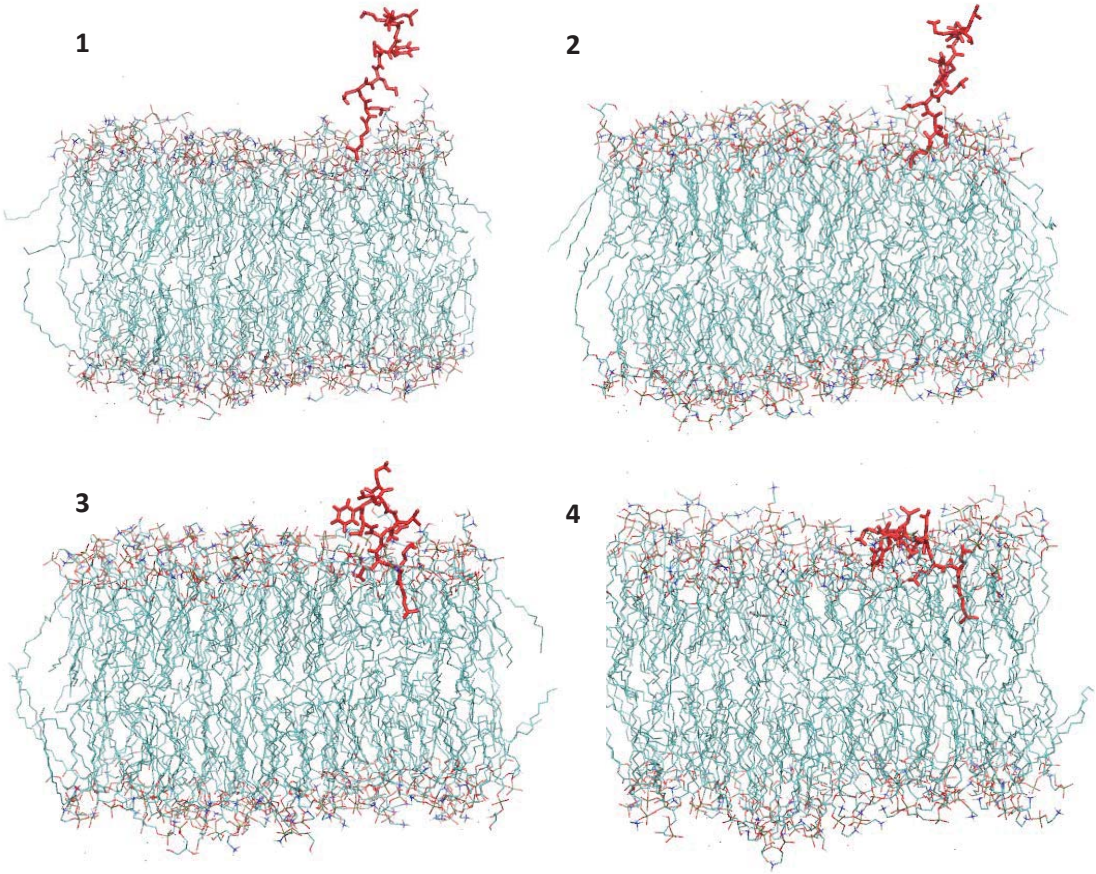
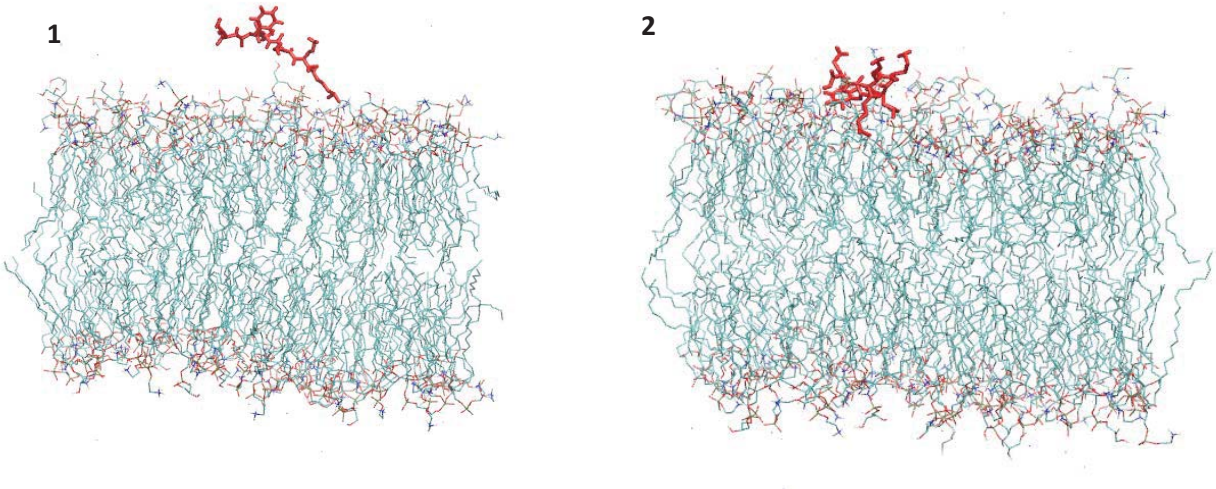


Figure 2.14: Sequential insertion of O17 into the E. coli inner membrane at 301 K. Mechanism starts with insertion of hydrophobic fatty acid into the lipid bilayer, followed by the rest of the peptide.



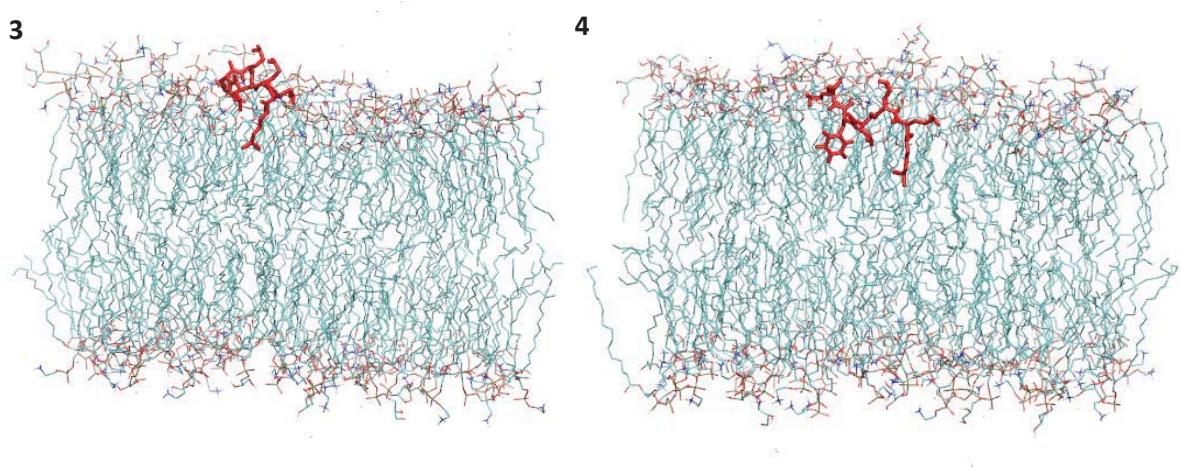


Figure 2.15: Sequential insertion of P30 into the E. coli inner membrane at 310 K. Mechanism starts with insertion of hydrophobic fatty acid into the lipid bilayer, followed by the rest of the peptide.

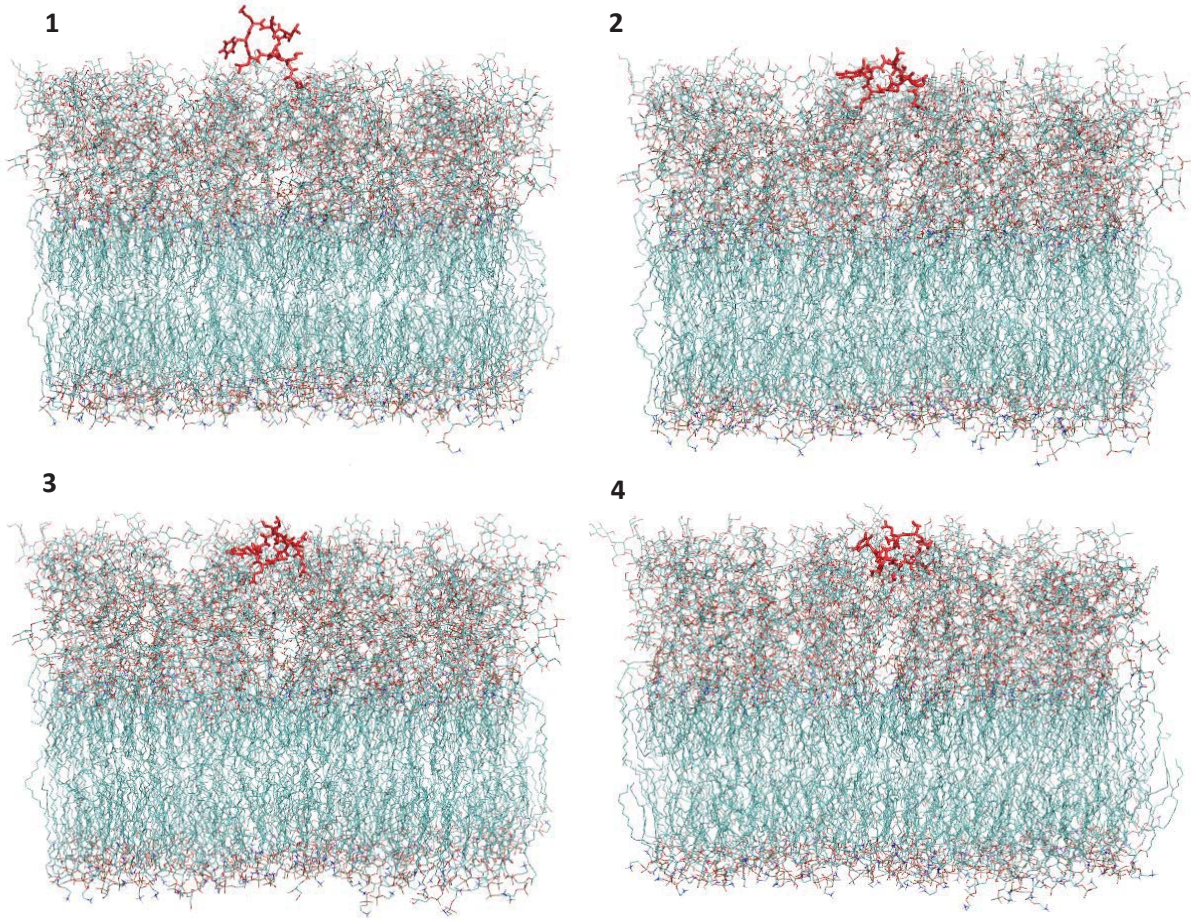


Figure 2.16: Sequential insertion of O17 into the E. coli outer membrane at 301 K. Mechanism starts with insertion of hydrophobic fatty acid into the lipid bilayer, followed by the rest of the peptide.

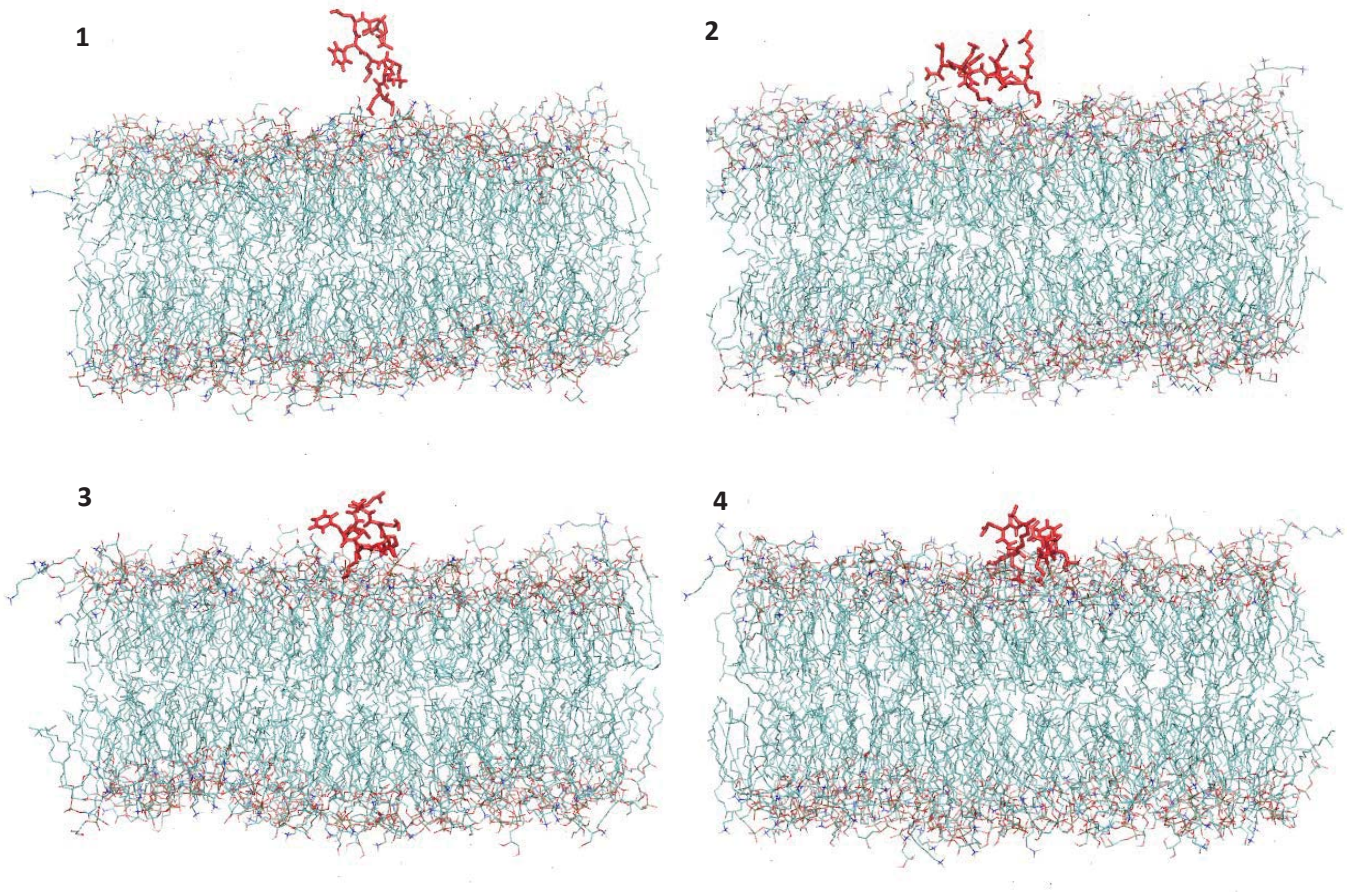


Figure 2.17: Sequential interaction of O17 into the S. aureus membrane at 301 K. Mechanism starts with insertion of hydrophobic fatty acid into the lipid bilayer, followed by the rest of the peptide.

2.2.3.6 Area Per Lipid

Area per lipid can be considered a good indicator of bilayer packing and structure, which in turn affects bilayer thickness and reflects the phase of the bilayer (e.g. gel or liquid crystal). Area per lipid was found to vary greatly across the membrane models, so for the purposes of the analysis, the bacterial membrane area per lipids were only compared to their own reference simulations (*Tables 2.8 and 2.9*). Little difference in the average area per lipid between the reference bilayer simulations and the peptide/membrane simulations were found for the *E. coli* inner and *S. aureus* membranes, indicating little structural rearrangement in the bilayer as a consequence of its interaction with the peptide. A reduction of 6.7-7.6% in area per lipid was observed in the *E. coli* outer membrane simulations containing the peptides, compared to the reference. The biggest discrepancy was found between reference simulations of *E. coli* outer membrane and the simulations with the peptides, indicating that the bilayer packing became looser as a result of peptide interaction.

Table 2.8: Average areas per lipid in nm^2 before and after interaction with anti-bacterial peptides for phospholipids for membrane simulations at 301 K and 310 K. Reference simulation values are included.

E. coli inner membrane

Simulation key	*Before*	*After*
Reference 301 K	0.611 +/- 0.005	
Octapeptide 301 K	0.620 +/- 0.002	0.629 +/- 0.003
Pentapeptide 301 K	0.621+/-0.004	n/a*
Reference 310 K	0.620 +/- 0.003	
Octapeptide 310 K	0.625+/- 0.001	0.623 +/- 0.003
Pentapeptide 310 K	0.618+/- 0.002	0.627 +/- 0.002

E. coli outer membrane

Simulation key	*Before*	*After*
Reference 301 K	0.639 +/- 0.003	
Octapeptide 301 K	0.584 +/- 0.000	0.592 +/- 0.002
Pentapeptide 301 K	0.584 +/- 0.001	0.591 +/- 0.002
Reference 310 K	0.640 +/- 0.000	
Octapeptide 310 K	0.587 +/- 0.001	0.597 +/- 0.001
Pentapeptide 310 K	0.588 +/- 0.001	0.598 +/- 0.002

S. aureus membrane

Simulation key	*Before*	*After*
Reference 301 K	0.646 +/- 0.001	
Octapeptide 301 K	0.642 +/- 0.001	0.646 +/- 0.001
Pentapeptide 301 K	0.634 +/- 0.004	0.655 +/- 0.001
Reference 310 K	0.660 +/- 0.004	
Octapeptide 310 K	0.656 +/- 0.003	0.656 +/- 0.002
Pentapeptide 310 K	0.655 +/- 0.002	0.649 +/- 0.002

2.3 Discussion

Internal structure analysis has demonstrated that both simulated peptides remained generally disordered, adopting coil (undefined) and turn secondary structures predominantly throughout the simulation. Little incidence of regular secondary structures such as α-helices or β-sheets was found, which is unsurprising considering the very short sequence length of the peptides.

The J-coupling value estimation for simulated peptides was found to be in within 10-20% agreement of experimental data, which confirmed that the model replicated real-world conditions reasonably well.

Some discrepancies were found in the J-coupling values of unnatural amino acids, but this was not unexpected as the analysis was seeking to compare dihedral angles of residues with known natural examples.

Both peptides were found to be generally highly mobile relative to the membrane. In all the simulations of the peptides in the presence of the different bacterial membranes, the peptides approached the membrane rapidly and spent most of the simulation time resting on the surface of the membranes after initial contact. Such mobility could also suggest that while the peptides are in contact with the membrane surface, they retained high degree of internal flexibility. The extent of such mobility varies depending on the type of bacterial membrane that the peptides were simulated with. Results were lower for simulations with the *E. coli* inner membrane, indicating the strongest reduction of mobility, which was likely caused by peptide insertion confirmed by later analysis.

All the peptides had approached the bacterial membrane surface, showing high affinity for the membrane surface for all simulated bacterial membranes. A high degree of hydrogen bonding affinity was observed – bonding to the membranes was found to be preferential than bonding with the solvent. Notable cases of interaction included O17 fatty acid chain inserting into the *E. coli* inner membrane at 301 K, indicating that the mechanism of peptide action is similar to that of Polymyxin B, as outlined in the introductory part of this chapter. Two further cases of P30 inserting itself partially into the lipopolysaccharide part of the *E. coli* outer membrane at 301 K and fully into the *E. coli* inner membrane at 310 K were observed via a similar method of action. The mechanism confirmed the previously reported of insertion via fatty acid interaction. The fatty acid component inserted into the membranes first for both O17 and P30 and the rest of the peptide followed. The interaction was likely assisted by the initial formation of hydrogen bonds between Dab, D-Phe and Leu peptide residues in the membrane, which fixed the peptide to the membrane surface.

It is difficult to compare the affinity and the speed of approach of different peptides to their corresponding membranes directly, as the simulations spaces were all of a different size due to differing membrane compositions. All of the peptides approached the membranes very quickly, in under 60 ns. Evidence of hydrogen bonding between the peptides and membranes was observed for all simulations. However, the degree of hydrogen bond interaction with the *S. aureus* membrane was lesser compared to the other membranes. Furthermore, no peptide insertion was observed with the *S. aureus* membrane.

No structural differences in the membranes were observed as the result of their interaction, indicating that peptides did not compromise membrane stability during the simulation time. It is possible that with longer simulation times the peptides could work their way into the membranes. However, extending the simulation runtime even further is very computationally expensive and currently not feasible within sensible timeframes utilising currently available resources. It is also possible that the mechanism of action required further factors, such as multiple peptide molecules interacting together or the presence of other cofactors. In future research, it would also be of interest to simulate peptides in their native ionised states to assess the effect of native ionisation on interaction with negatively charged membranes and in the destabilisation via cation replacement in LPS.

2.4 Methods

Preparation, running and analysis of the simulations were performed using GROMACS 5.1.1[17, 62-67] utilizing the GROMOS 54a7 force field[68] with parameters for the unnatural amino acids added as described below.

2.4.1 Peptide parameter and coordinate generation

Initial coordinates were generated using GaussView version 5.0.9., from literature reported sequences[33]. Topologies were assembled manually from existing and amended amino acid parameters based on the GROMOS 54a7 forcefield. Dab residue parameters were created by deleting two carbon atoms, as well as the associated bond, angle and dihedral parameters, from the side chain of the existing 54a7 Leu residue parameters. The partial charge on the deleted atoms was zero, so there was no need to recalculate partial charges. D-Phe and D-Dab topologies were built from the corresponding L- residues by reversing the improper dihedral term that maintains the tetrahedral geometry of the α-carbon (due to the lack of the aliphatic α-hydrogen atom in the GROMOS united-atom force fields). The fatty acid chain parameters were constructed manually based on existing lipid and branched-chain aliphatic amino acid side chain parameters. Peptides were created in a charge neutral state (non-ionised) according to existing building block parameters. Peptides were then solvated in SPC water[69-70]. A copy of the forcefield files and starting configurations can be found in *Appendices 1* and *2*.

2.4.2 Bacterial membrane simulation coordinates and parameters

Bacterial membrane parameters and pre-equilibrated (200 ns) starting coordinates, including counterions, were obtained from Drs Thomas Piggot and Syma Khalid[61]. The starting coordinates for the membrane reference simulations were the final coordinate sets from the 200 ns pre-equilibrated *E. coli* inner membrane, *E. coli* outer membrane and *S. aureus* membrane simulations. Lipid composition by number, including counterion counts, can be found in *Table 2.9*.

For the simulations of the peptides with the bacterial membranes combined coordinates were generated. Coordinates of each peptide were taken from the end of the 100 ns simulations in water. Similarly, coordinates of each bilayer were taken from the end of the 100 ns reference simulations. Solvent coordinates were deleted from both systems. The simulation box space of the bilayer was expanded by 1.1 nm in the z axis using the gmx editconf program and the peptide coordinates were added to the simulation space at a minimum distance of 1.4 nm away from the bilayer. The combined coordinates were then resolvated with SPC water.

Table 2.9: Membrane lipids by numerical composition

Lipid	*E. coli inner*	*E. coli outer*	*S. aureus*
POPE	92	144	
POPG	24	8	108
DPG	6	8	10
PG-Lys			72
LPS		64	
Na^+	36		56
Mg^{2+}		332	

2.4.3 Energy minimisation and equilibration

The simulations were energy minimised over 50000 steps using the steepest descent algorithm with an energy minimization step of 0.01 ps. Equilibration under NVT conditions was performed for 100 ps, and the systems were heated from 50 K to 301 K during that period to the relevant simulation temperature (301 K or 310 K) prior to equilibration under NPT conditions for 1 ns. Further details of the equilibration and heating simulations were as described below.

2.4.4 MD simulation

Each of the peptides, with and without the bacterial membranes, were simulated for 100 ns at 301 K and 310 K. The cut-off for calculating non-bonded Coulomb and van der Waals interactions was 1.4 nm. Long-range electrostatic interactions were approximated using the reaction-field[20]. Bonds were constrained using the LINCS[71] constraint algorithm with an order parameter of 4, allowing for an integration time step of 2 fs. The temperature was maintained using the weak coupling Berendsen thermostat[72], with temperature coupling time τ_T = 0.001 ps^{-1}. For the NPT simulations, the pressure was maintained at a pressure of 1 atmosphere using the Berendsen barostat with pressure coupling time τ_P = 0.5 ps^{-1} and isothermal compressibility = 4.5 × 10^{-5} atm^{-1}. For simulations involving the bacterial membranes, the pressure coupling was changed to semi-isotropic. Coordinates and energies were written out every 50 ps.

2.4.5 Analysis

The GROMACS 5.1.1 software was used to analyse the simulation trajectories. The analyses carried out are listed in alphabetical order below.

Area per lipid

Area per lipid was calculated by obtaining the simulation box *x* and *y* dimension averages and dividing by the number of lipids per leaflet appropriate to the membrane in particular. The average box dimensions were calculated with the GROMACS energy program. Errors were also calculated by GROMACS energy program and maximum values including the errors for the *x* and *y* dimensions were calculated, multiplied together and the difference between maximum possible error area and the average area was given as the maximum error.

DSSP

DSSP analysis[73-74] was performed using the GROMACS do_dssp program. Outputs were converted by a python script into an ASCII format for plotting.

Hydrogen bonds

Numbers of hydrogen bonds formed was determined using the GROMACS hbond program. Hydrogen bond pair analysis for peptide-bilayer interactions can be found in Appendix 3.

J-value

$^{3}J_{NH\text{-}C^{\alpha}H}$ coupling values parameters were calculated by the GROMACS chi program.

Minimum distance

The minimum distances between a peptide and the lipid bilayer were calculated using the GROMACS mindist program.

RMSD

The atom-positional root-mean-square deviations were calculated using the GROMACS rms program.

When used to report on the internal structure, all atoms of one peptide were fitted to its starting coordinates (of the production run, i.e. the coordinates from the end of the NPT equilibration), and the RMSD was then calculated for all atoms of the peptide only.

When used to report on the general mobility of the peptide, the coordinates of both the peptide and the membrane were fitted to their initial coordinates, but the RMSD was again calculated for all atoms of the peptide only.

RMSF

The atom-positional root-mean-square fluctuations were calculated using the GROMACS rmsf program and both printed per-atom, and summed over all atoms to give the per-residue values.

Visualisation

VMD software package version 1.9.2. was used to visualise the GROMACS coordinate trajectory files[75]. VMD images were screen captured for use in this thesis.

2.5 Conclusion

The structures and dynamics of two novel anti-bacterial peptides, O17 and P30, were studied and analysed. Both peptides were simulated in solution and then simulated with established models of the *E. coli* inner and outer membrane and the *S. aureus* membrane, whose parameters and equilibrated coordinates were kindly provided by Prof. S. Khalid and Dr T. Piggot[61].

The investigations described here have revealed that these small linear antimicrobial peptides were very mobile and flexible in solution, exhibiting a lack of defined secondary structure for both peptides. Both peptides are amphiphilic and have both hydrophobic and hydrophilic residues. Hydrogen bonding analysis has showed that both peptides form up to 35 (O17) and 25 (P30) hydrogen bonds with the surrounding water solvent.

In all the simulations with the bacterial membranes the peptides were found to approach the bacterial membranes rapidly and preferentially hydrogen bond to the respective membranes, forming multiple hydrogen bonding interactions with the membranes. Both peptides were found to rest on the membrane surface by the end of the simulations, with the exceptions of several that have begun their insertion into the membranes themselves. In the cases of O17 with the *E. coli* inner membrane at 301 K, P30 with the *E. coli* inner membrane at 310 K and with the *E. coli* outer membrane at 301 K, parts of or the entirety of the peptide have permeated the lipid bilayer. Thus, indicating that these novel peptides might possess similar antimicrobial mechanisms as that of proposed mechanism of polymyxin B. The peptides were found to affix themselves to the membrane surface via hydrogen bonding interactions and the fatty acid subsequently inserted itself into the lipid membrane, followed by the peptide. Little evidence of membrane destabilisation was found and only *E. coli* outer membranes were found to experience and increase in the average area per lipid as a result of the interaction. However, it is possible that 100 ns simulation time was not sufficient to track these processes from start to finish in most cases and only the beginning of membrane destabilisation could be observed. Furthermore, it is possible that the action of multiple peptide molecules is required to form pores or otherwise destabilise the membrane.

In conclusion, the simulations confirm the interaction and high affinity of novel antibacterial peptides with both gram-positive and gram-negative bacterial membranes, confirming the hypothetical mechanism fatty acid insertion reported in the literature.

Abstract

Antifreeze proteins are found in a variety of species and serve to protect cells from freezing damage. A structural analysis of the behavior of antifreeze protein 1 (AFP1) from winter flounder during molecular dynamics (MD) simulations is presented, as well as a study of possible temperature induced aggregation. Interactions of AFP1 proteins were studied in conjunction with a model of a typical unsaturated cell membrane. AFP1 units were found to show greater flexibility in solution compared to their crystal structures, adopting a variety of non α-helical structures. In certain cases, two AFP1 proteins aggregated together in solution and formed hydrogen bonds between each other. It was also found that AFP1 proteins interacted with an unsaturated lipid membrane, resting on its surface, providing insight into the mechanism of preventing cell damage from freezing conditions.

Chapter 3: Investigation of winter flounder Antifreeze Protein 1 structure and aggregation in the context of lipid membranes.

3.1 Antifreeze protein 1 (AFP1)

Antifreeze proteins (AFPs) play an essential part in preservation of living cells in cold environments. AFPs have a broad range of different functionalities across multiple species and several different structural groups exist. The scope of this research aims to investigate antifreeze protein 1 (AFP1). AFP1 is found a variety of species including bacteria[76-77], psychrophilic yeasts[78], lichens[79], winter plant species[80-81] and marine animals[82].

1				*5*					
ASP	**THR**	ALA	**SER**	**ASP**	ALA	ALA	ALA	ALA	ALA
11				*15*					
ALA	LEU	**THR**	ALA	ALA	**ASN**	ALA	**LYS**	ALA	ALA
21				*25*					
ALA	**GLU**	LEU	**THR**	ALA	ALA	**ASN**	ALA	ALA	ALA
31				*35*					
ALA	ALA	ALA	ALA	**THR**	ALA	**ARG**			

Figure 3.1: Winter flounder AFP1 amino acid sequence with residue numbers[83]. C-terminus is capped by an amide group. Residues with functionalised, non-aliphatic carbon chains are highlighted in bold. Amino acids reported to exhibit ice binding activity are Asp-1/Thr-2/Asp-5, Thr-24/Asn-27, Thr-13/Asn-16 and Thr-35/Arg-37

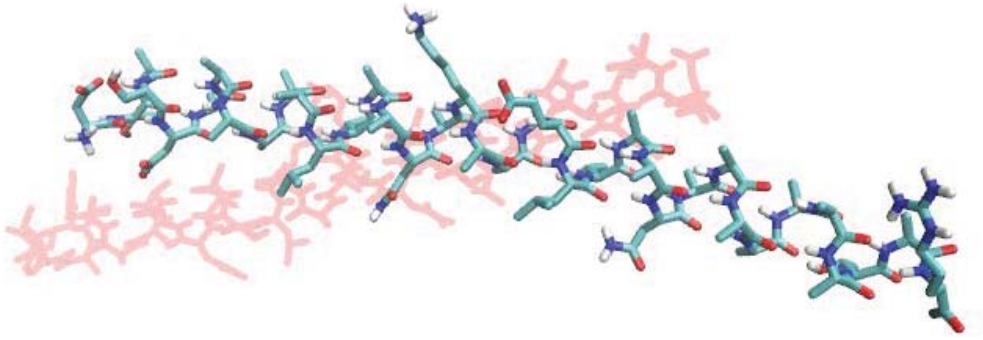


Figure 3.2: Winter flounder AFP1 crystal structure unit cell, containing two AFP1 proteins. Highlighted in red is the second protein unit (B).

Although the mechanism of antifreeze action is controversial, AFP1 protects cells from death at near freezing temperaures[84]. The mechanisms of ice damage are multifold and typically involve cell lysis via mechanical damage and dehydration[85]. Winter flounder AFP1 has predominantly α-helical structure in the solid phase and the crystal structure produced at 277 K contained two AFP1 units made up of 37 amino residues each. Complexed crystallised water molecules were also present at the C- and N- termini of the AFP1 units. Experimental studies utilizing X-ray crystallography have found that the α-helical winter flounder AFP1 formed four ice-binding regions per single protein which could bind ice crystals as a single flat surface[83] (*Figure 3.1*). A simulation of AFP1's ice-binding pattern carried out by Cheng and Merz in 1997 indicated the position of four regularly spaced ice binding regions, two situated at the ends of the molecule (residues Asp-1/Thr-2/Asp-5 and Thr-35/Arg-37) and two middle regions (Thr-13/Asn-16 and Thr-24/Asn-27)[86]. Crystallography studies gave further evidence, where threonine residues were shown to be critical to the ice binding functionality[87]. The ice binding regions are separated by roughly 11 residues along the length of the molecule, the spacing of which corresponds to the periodicity exhibited in certain ice lattices, leading to enhanced hydrogen bonding with the ice lattice. NMR studies of AFP1 confirmed the α-helical nature of the protein[88]. Simulation studies of the protein by Hong *et al* at different temperatures also suggest that the method of action involves adsorption of the molecule onto the surface of the formed ice crystals and further indicating that hydrogen bond formation between the amino acids and water Oxygens was important in the interactions[89].

More recently it has been reported that a "hyperactive" form of the AFP1 protein exists. This protein could be a dimer or higher-order polymer of AFP1 helical protein units[90-92]. Marshall *et al.* report a hyperactive form of AFP1 from winter flounder which they characterise as a long molecule comprised of α-helical structures[92]. This protein has been isolated and found to be twice the mass (34 Da) of two AFP1 monomer units (17 Da each)[90], possibly indicating a dimer precursor structure. It is possible that previous studies have failed to detect this dimer or aggregate due to the thermal instability of the protein at non-water freezing temperatures. Such oligomerisation or complexation behaviour is common in antifreeze proteins in general[93]. The crystal structure of AFP1 suggested that complexation or oligomerization is also possible as the unit cell contains two monomer AFP1 units[86].

It was also demonstrated that AFP1 proteins could interact with unsaturated lipid membranes experimentally[94]. Studies have shown that AFP1 addition stopped freezing damage to unsaturated lipid membranes through a variety of practical experimentation studies[95-98]. Typically, psychrophilic organism cell membranes are comprised of unsaturated phospholipids[99-101]. Further adaptations of such membranes include high degree of acyl branching and presence of large lipid head groups. However, such alterations are not universal and exact composition varies across species and types of cell[102]. However, the mechanism of AFP1 interaction with assorted unsaturated lipid bilayers was not clear.

In the research of Tomczak, *el al*. a mixed membrane, containing phosphocholine and digalacto- lipids, was used as a model to test AFP1 interactions with lipid membranes[94]. Digalacotolipid was chosen for its high degree of unsaturation, as it has three carbon double bonds in each of the lipid tails. AFP1 presence was found to affect the lipid bilayer packing order in experiments with unsaturated lipids, but not in those containing purely saturated phospholipids. The presence of unsaturated lipids also affected the

temperature dependent folding of AFP1 protein. AFP1 folded into an α-helical structure upon lowering the temperature in solution without the membrane present. However, in the presence of the mixed saturated/unsaturated liposome the chilling induced-folding of AFP1 protein was inhibited. Contrasting studies by Kar *et al*. show that AFP's are dynamic and adopt a variety of structures when interacting with the membranes, but still suggest a high degree of interaction[98].

Further evidence for interaction of AFPs with unsaturated membranes was found experimentally by Hays *et al*. when AFP presence was found to stop liposomes leaking their contents as they went through a phase transition[103-105]. Therefore, it was suggested that a direct interaction of AFP's with the bilayer could occur, although this was not tested explicitly.

There are clearly many gaps in the understanding of AFP1 behaviour, both in terms of its potential oligomerisation and its interactions with the cell membrane. In this work, pairs of AFP1 protein units were simulated together at standard (room) and near-freezing temperatures to investigate dynamics, mutual interactions and any possible oligomers precursor formation. Simulations of the protein with a lipid bilayer mimicking a typical cell membrane were carried out to characterise whether and how such dimer structures might interact with lipid bilayers.

3.2 MD investigation of AFP1

The primary goal of this project was to investigate the structural dynamics and possible dimerization modes of a pair of AFP1 monomers using MD simulations. Firstly, the structural dynamics of pairs of AFP1 proteins were investigated in different relative orientations in water at near freezing and close to room temperatures. It was expected that each AFP1 monomer would be more mobile and less rigid at 300 K than at 277 K. Secondly, the interaction between the two protein units was closely monitored to see if they would form a complex in solution at near freezing temperatures, indicating possible further dimerization, and how this could be affected by their initial relative orientation with respect to each other. The second goal of this chapter was to investigate AFP1 behaviour in the context of a cell membrane model, comprised of unsaturated phospholipids. AFP1 simulations with an unsaturated lipid membrane are covered in *Section 3.3*.

3.2.1 System design

Five different configurations of pairs of AFP1 were set up to investigate the structural dynamics and dimer formation of AFP1. All simulations and subsequent analyses were performed using the GROMACS 5.1.2 software package. Each system contained two AFP1 units in different relative orientations (*Figure 3.3 and Table 3.1*), and was simulated at 277 K and 300 K. By simulating multiple different arrangements of pairs of protein units, not only could the effect of initial orientation on subsequent dimer formation

be assessed, additionally, should dimerization not occur, better statistical sampling of the range of conformations accessible to the monomeric protein would be obtained.

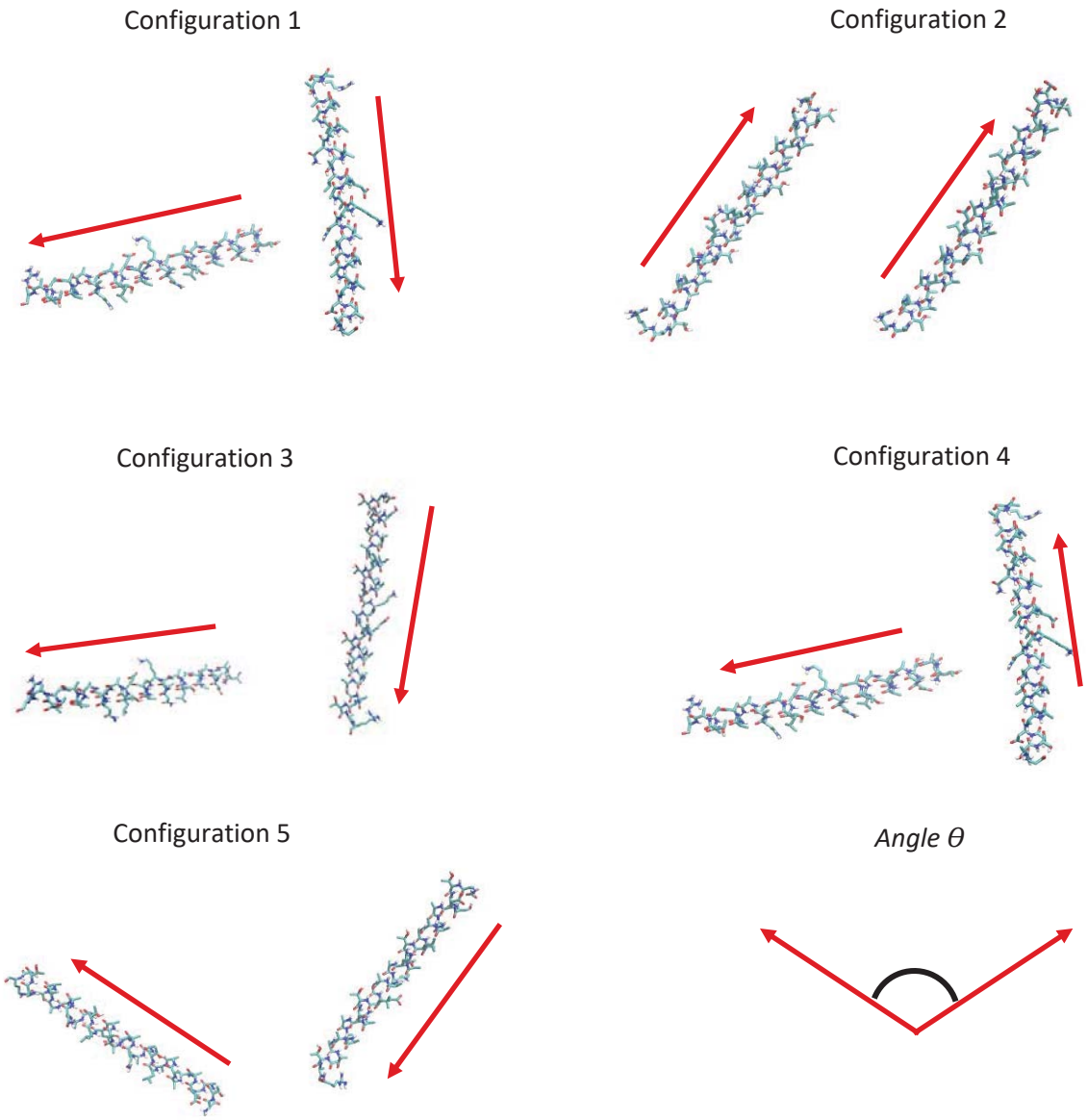


Figure 3.3: Diagram of starting coordinates for all configurations. The arrows represent the direction from N-terminal residue 1 to amide-capped C-terminal residue 37. Both peptides lie in the same plane, so that their orientations differ only in terms of the angle θ.

The initial coordinates of the AFP1 protein were obtained from the Protein Data Bank[106] X-ray crystal structure (PDB ID: 1WFA)[83, 107] and edited to create the five different orientations. The proteins were

solvated in single point charge (SPC) water[70] and the complete system energy minimized, heated to the required temperature and equilibrated for 1 ns before carrying out a 100 ns MD simulation. Further details of how the simulations were conducted are given in section *3.5 Methods*.

The PDB crystal coordinates contained a pair of complexed protein subunits with some complexed waters that were later replaced during the solvation process. Different configurations of the starting relative orientations were created to test whether overall orientation affected the structure and interactions of the AFP1 units. The five configurations were created by separating the protein units and translating and rotating them with respect to each other to vary the relative angle and overall relative orientation. The SPC[108] water model with which the GROMOS force fields are compatible was used in all of the simulations.

Table 3.1: Angle θ between the N to C vectors of the AFP1 units A and B in the equilibrated configurations from which the simulations at 277 K and 300 K were initiated. The peptides lie in the same plane, so that their orientations differ only in terms of θ. Note that the θ values vary slightly from the initial setup during the heating and equilibration procedure.

	Config. 1	*Config. 2*	*Config. 3*	*Config. 4*	*Config. 5*
277 K	76.9°	3.0°	103.6°	66.9°	85.5°
300 K	75.8°	1.9°	103.6°	66.3°	85.9°

Overall the five starting configurations were broadly categorised as follows:

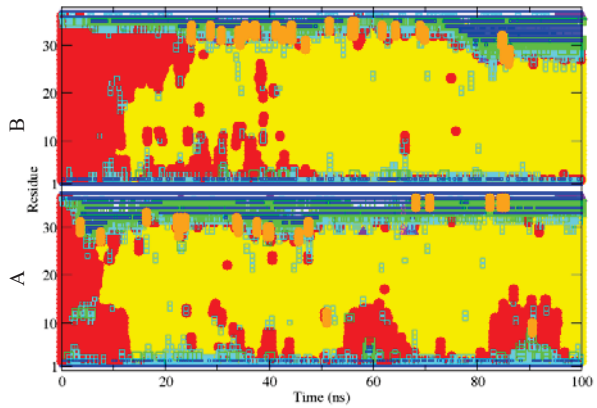
- Configurations 1 and 3 were “T” shaped; molecules were perpendicular to each other
- Configuration 2 has the proteins parallel to each other
- Configurations 4 and 5 were “V” shaped; in 4 the ends of the molecules both point outwards and in 5 one end points towards the start of the other molecule

3.2.2 Internal structure

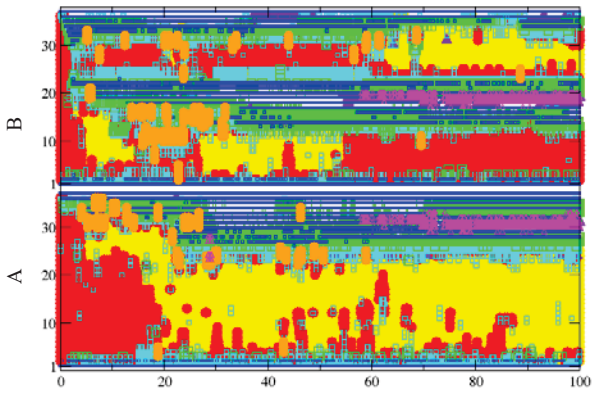
Internal structure of the AFP1 protein units was assessed using a variety of tools. Firstly, a DSSP (database of secondary structure assignments) secondary structure analysis was performed. Secondly, the root mean square deviation (RMSD) from the initial structure and the per-residue root mean square fluctuation (RMSF) were calculated to assess the internal mobility of the individual units.

3.2.2.1 DSSP

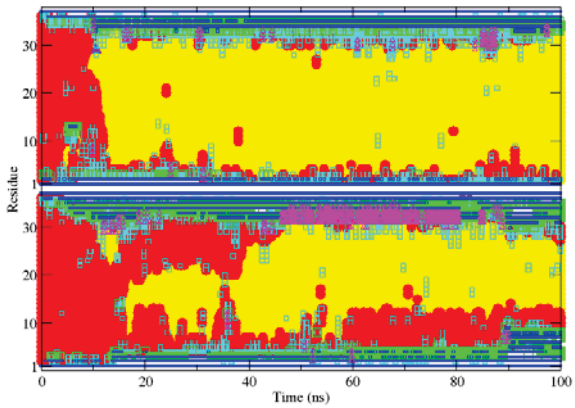
Config. 1 277 K



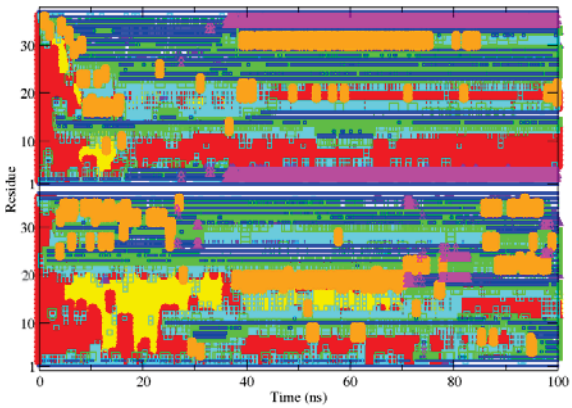
Config. 1 300 K



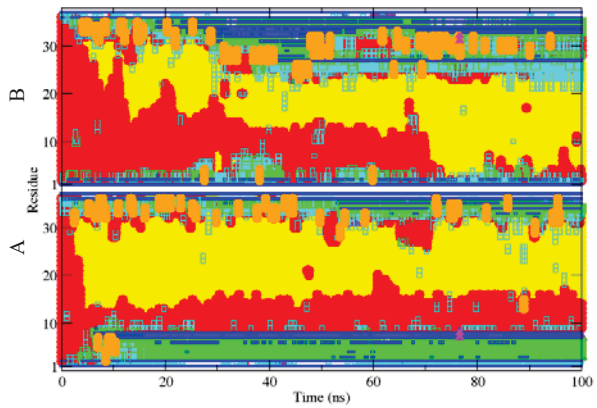
Config. 2 277 K



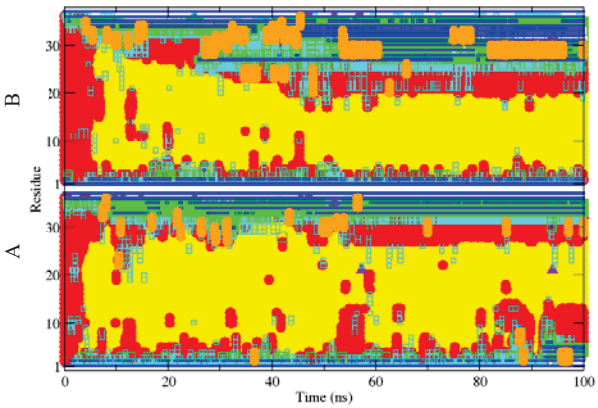
Config. 2 300 K



Config. 3 277 K



Config. 3 300 K



B

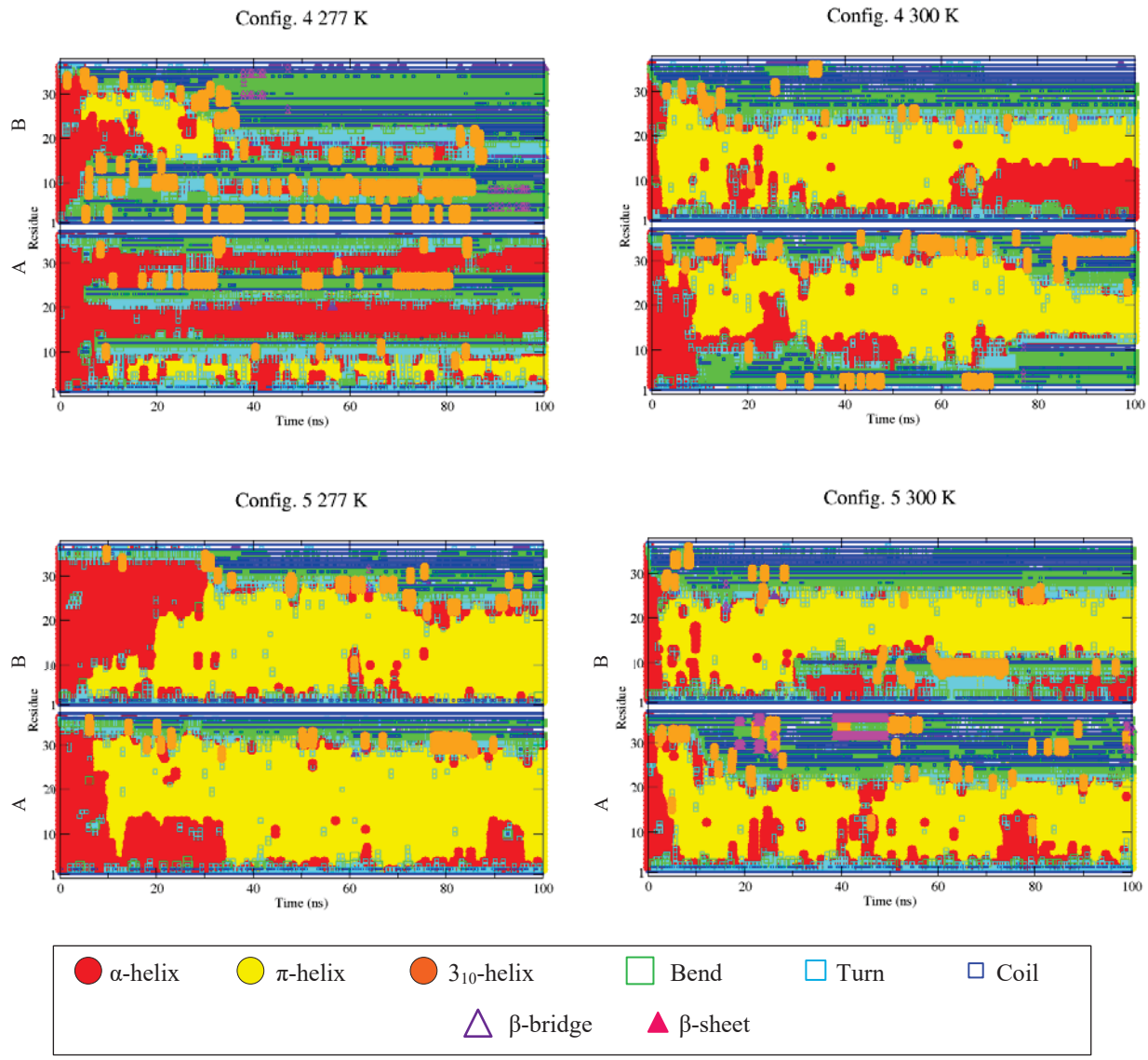


Figure 3.4: Secondary structure content of AFP1 protein units A and B during simulations initiated from each of the five starting configurations and run at 277 K and 300 K.

Both AFP1 subunits started all simulations with a predominantly α-helical structure for the middle residues, in line with experimental crystallography findings. In most simulations the initial α-helical structure gave way to a π-helical one before 10 ns. The residues close to the N- and C- termini of the protein exhibited a number of non-helical secondary structures throughout the entire simulation, predominantly coil, bend and turn, indicating higher flexibility and less rigidity at the ends of the rod like protein. Coil structure is assigned when no other secondary structure elements could be fit to the residue. Other types of structure such as a 3_{10}-Helix and β-sheets and β-bridges occurred infrequently. It is possible that the α-helical structure attributed to AFP1 in the literature is a result of crystal packing or solid phase behaviour, as the MD simulations in solvent all show a clear shift away from strict α-helicity.

A further possibility is that the forcefield is biased against α-helical structure; however, although the earlier GROMOS 53a6 forcefield was known to have such a bias, the 54a7 forcefield used here was developed specifically to overcome this bias[109].

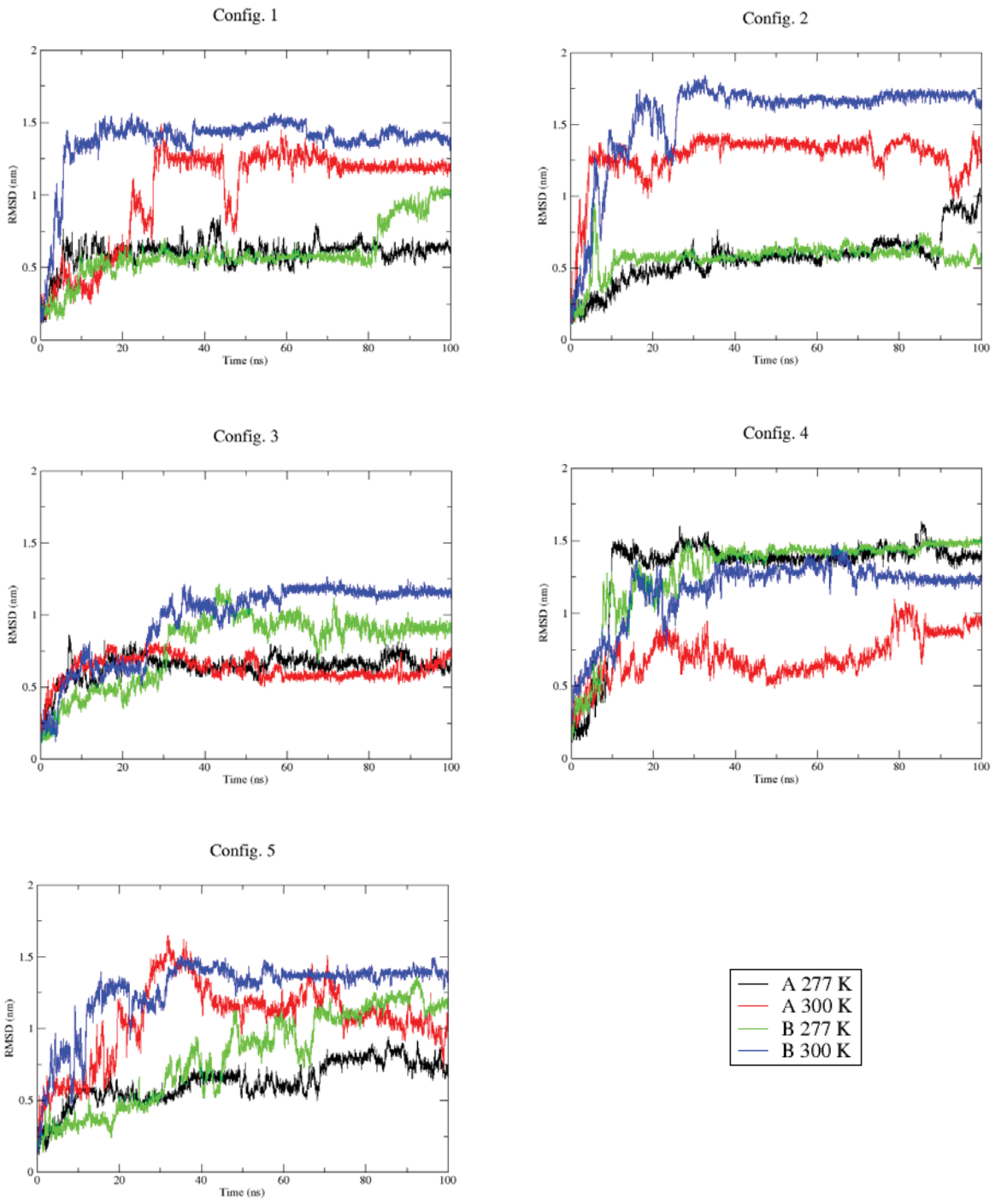


Figure 3.5: Atom-positional RMSD of all atoms of each individual protein unit mapped on to itself (A on A and B on B) at 277 K and 300 K.

Configurations 1, 2 and 5 retained more helical character at 277 K, while structures at 300 K generally exhibited simpler secondary structures such as turns and coils. This trend followed previous experimental results, which found AFP1 proteins to fold into an α-helical structure at near freezing temperatures. However, the temperature trend was reversed in configuration 4, and in configuration 3 there was little difference between the two temperatures, with the protein exhibiting predominantly helical structure in both simulations.

It was hard to assess whether secondary structure variance between the experiments were caused by the difference in starting relative positions. Of special interest were configurations 3 and 4 where the predicted temperature trend was reversed. From *Figure 3.4* and *Table 3.1* it can be seen that configurations 3 and 4 are T shaped and V shaped coordinates, respectively, with inter-unit angles similar to those of configurations 1 and 5 (see *Section 3.2.3.3: Angles*), so it was difficult to establish a definitive correlation between the secondary structure during the simulation and the initial orientations or the simulation temperature.

3.2.2.2 RMSD (internal structure)

The RMSD from the initial conformation is a good indication for the degree of structural change of each unit during a simulation (*Figure 3.5*). The atom-positional RMSD of all atoms from the initial structure was computed separately for each protein unit to remove any contributions from changes in the relative orientation of each subunit (see subsequent *Section 3.2.3.1*).

Generally, all protein units exhibited a good degree of relative internal mobility. RMSD values increased in the first 20 – 30 ns of the simulations and subsequently settled to a steady average. In agreement with the DSSP results, the RMSD time series indicated that protein subunits were generally more mobile in the higher temperature simulations in the cases of configurations 1, 2, and 5. In configuration 3, variance in temperature did not affect the RMSD, particularly for subunit A. This again shown some agreement with the secondary structure content, which was similar for both subunits at 277 K and 300 K (*Figure 3.5*). In the case of configuration 4, the trend was reversed; unit A demonstrated higher internal flexibility at the lower temperature and unit B similar RMSD for both temperatures. Again, there was some correspondence between this behaviour and the secondary structure content, which was mostly helical at higher temperature (*Figure 3.4*).

3.2.3 Interactions of AFP1 units

It was important to analyse the interactions of AFP1 units with each other and with the solvent environment to determine whether dimerisation could occurr spontaneously in solution, and whether this was affected by the initial relative orientation of the subunits. The RMSD of each individual protein

unit was calculated from its initial conformation after fitting to the starting equilibrated coordinates of both units to assess the mobility of the units with respect to each other. The minimum distance between the two AFP1 units during the simulation was calculated to show any possible aggregation (*Section 3.2.3.3: Separation*), and hydrogen bonding analysis was also performed to assess the nature of any inter-molecular interactions that formed (*Section 3.2.3.5: Hydrogen Bonding*).

3.2.3.1 RMSD (relative mobility)

To assess the degree of intermolecular mobility of each protein unit, the atom-positional RMSD of all atoms of each individual protein unit was calculated after fitting both subunits simultaneously to their coordinates at the start of the simulation (*Figure 3.6*) and averaged over the entire simulation (*Table 3.2*). The RMSD values for relative mobility were found to be much higher than the ones indicative of internal flexibility (*Figure 3.5*), spanning the ranges of 2 to 14 nm for configurations 1 and 2; and 2 to 12 nm for configurations 3, 4 and 5. High values were related to high relative mobility of the protein units with respect to each other.

All configurations showed a high variability, as well as a high magnitude, of RMSD values in the initial stages of the simulation. In configurations 1 and 4 at 300 K, the RMSD values settled to approximately 4 nm after 55 and 50 ns, respectively, but at 277 K they continued to fluctuate throughout the simulation. For configuration 3, all four simulations settled to values of approximately 4 to 5 nm within the first 20 ns, but the simulations at 277 K, in particular, exhibit frequent sudden excursions to higher RMSD values. Similarly, there appeared to be little correlation between simulation temperatures and the RMSD values. Averages of RMSD values from all of the simulations, found in *Table 3.2*, indicated that AFP1 units all configurations had an RMSD of approximately 4 to 6 nm.

The sudden spikes in the RMSD time series, particularly those observed for configuration 3, were caused by a single unit jumping across the periodic simulation boundary. It was difficult to obtain trajectories where both the protein units would never cross the simulation boundary due to their high mobility. The standard approach to accounting for periodicity, which involved centering the protein in the simulation box, was not applicable for two highly mobile peptides.

Table 3.2: Atom-positional RMSD values (in nm) averaged over the entire simulation at 277 K and 300 K for individual AFP1 units A and B after fitting both subunits simultaneously to their coordinates at the start of the simulation, standard deviation σ is included as a measure of error.

	Config. 1	*Config. 2*	*Config. 3*	*Config. 4*	*Config. 5*	*σ*
A 277 K	5.472	9.024	4.993	6.000	6.907	0.925
A 300 K	5.074	5.833	4.238	4.614	6.533	0.238
B 277 K	5.510	6.577	4.634	6.232	5.892	0.232
B 300 K	5.217	6.131	4.704	4.811	6.263	0.274

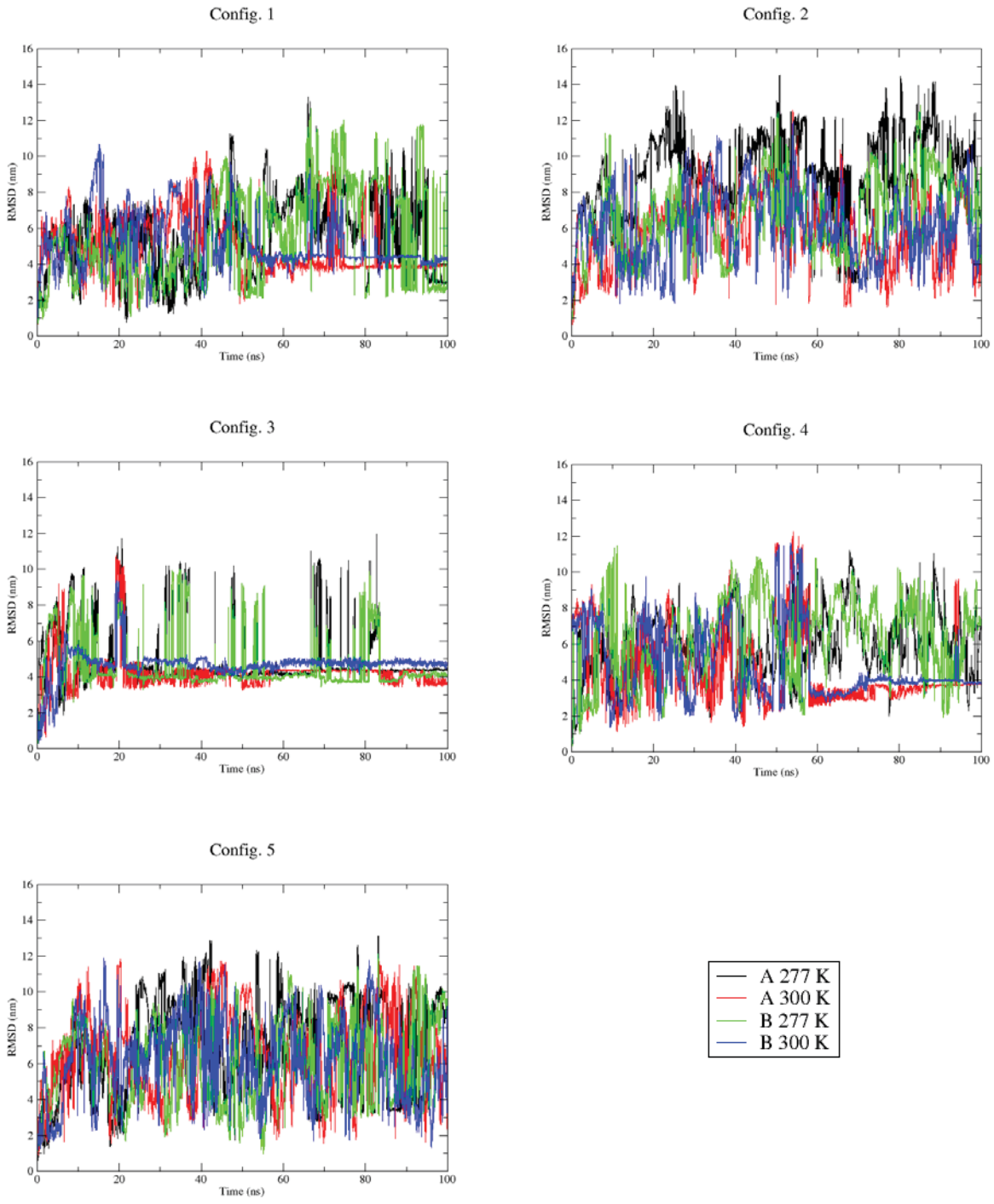


Figure 3.6: Atom-positional RMSD values at 277 K and 300 K for individual AFP1 units A and B after fitting both units simultaneously to their coordinates at the start of the simulation.

3.2.3.2 RMSF

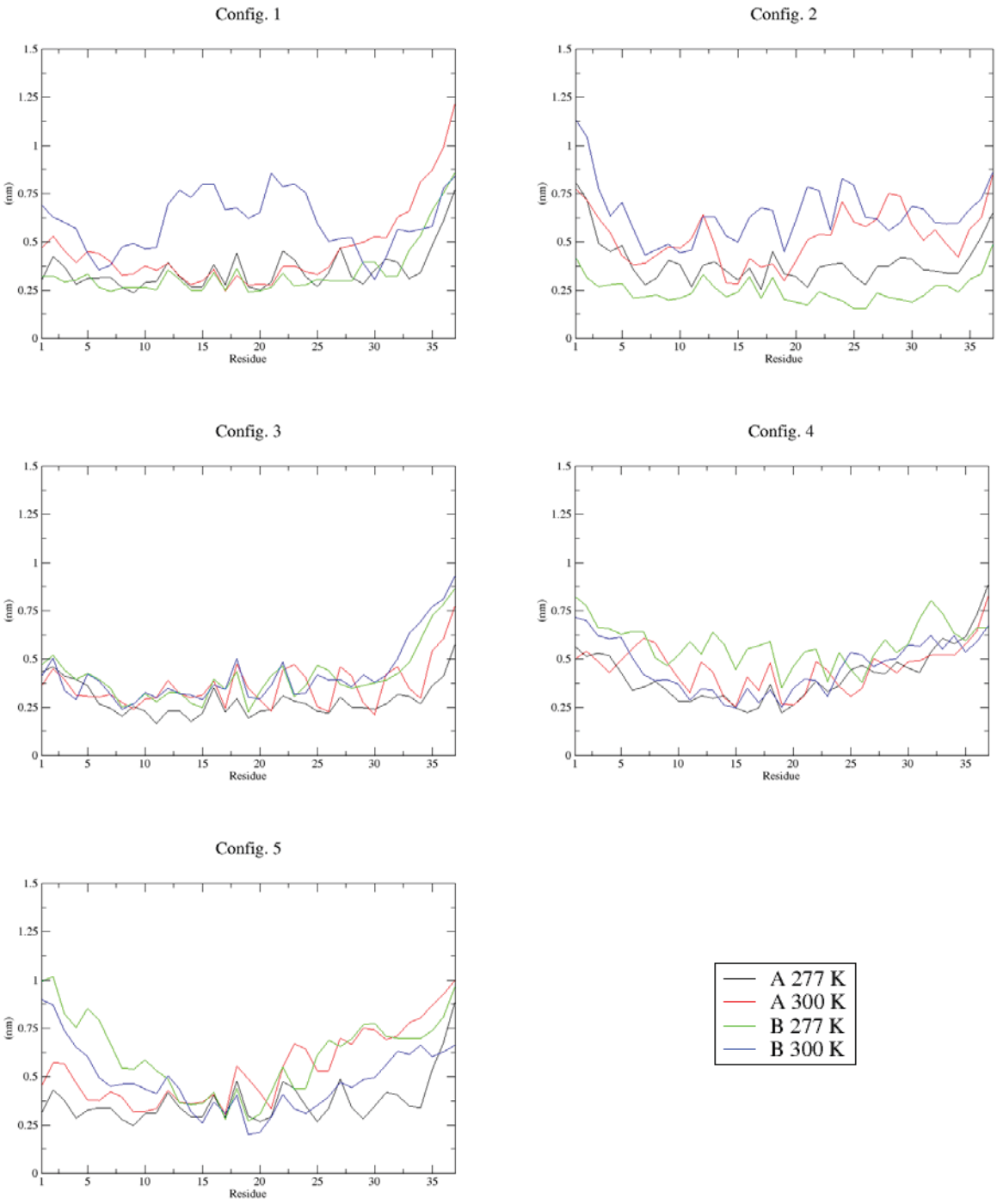


Figure 3.7: RMSF of each protein unit A and B in the simulations at 277 K and 300 K per residue.

The RMSF was calculated for each residue (sum of atoms in that residue) in *Figure 3.7*. The fluctuation in RMSF of individual atoms varied across a wide range of values from 0.25 to 1.5 nm. Baseline RMSF for most residues was around 0.3 nm. RMSF of residues situated at the ends of the molecules was greater than that of middle ones, indicating higher molecular movement and flexibility of the ends of each protein units, agreeing with previous DSSP results. RMSF values were found to be similar for both temperature ranges, with 300 K series largely exceeding the 277 K values only in configurations 1, 2 and 5 and varying little in 2 or 3. Localised maxima were found in the RMSF results, consistent across both temperatures, units and most configurations. The peaks typically corresponded to residues with functionalised side chains, such as THR, ASN (*Figure 1.1*), which were linked to ice-binding functionality. This is not surprising as the functionalised residues typically contain highly mobile functionalised side chain groups.

3.2.3.3 Separation

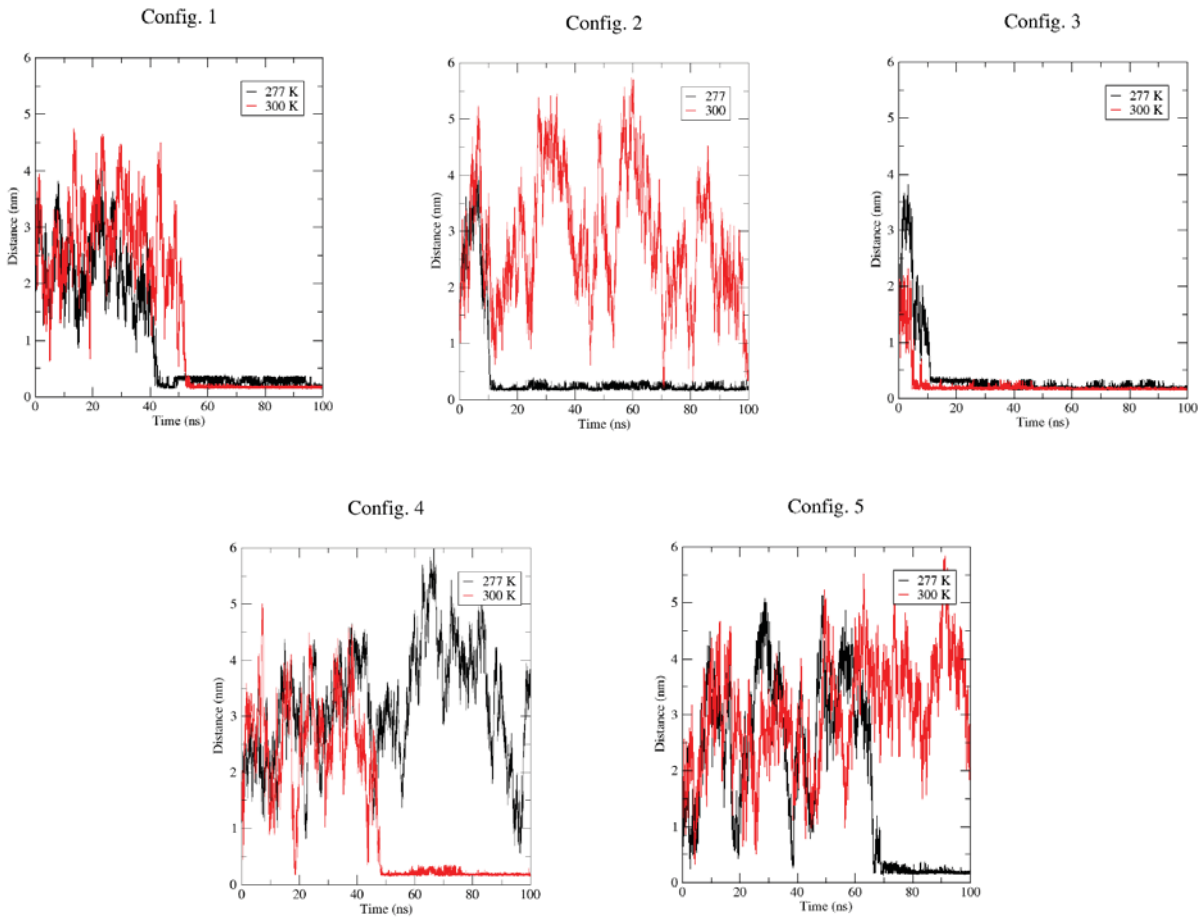


Figure 3.8: Minimum distance between AFP1 monomer units A and B for all five starting configurations for both 277 K and 300 K simulations.

To determine whether the monomers converged positions during the simulation, the minimum distance between the two protein units was calculated (*Figure 3.8*). Seven out of ten simulations have resulted in the two AFP1 protein units approaching each other. A minimum distance of less than 0.25 nm was taken to indicate association of the monomer units, as it is the closest possible distance A and B due to the effective atomic radius generated by the LJ potential. Generally, a distance of up to 0.6 nm was also be indicative of possible interactions.

In configurations 1 and 3 at both temperatures the monomer units aggregated at similar times, whereas for configurations 2 and 5 at 300 K the monomers remained separated for the duration of the simulation and furthermore, in configuration 4, the monomers remained also separated at 277 K.

The minimum distance results can be correlated with RMSD results (*Figure 3.6*). In config. 3 the RMSD values remained low and settled quickly, which correlated to the time at which both proteins approached each other. The correlation was repeated for configuration 1 at 300 K at approximately 50 ns and for configuration 4, 300 K at approximately 55 ns. It was harder to see the correlation in configurations 2 and 5 due the large range in RMSD values adopted, which did not seem to be affected by the A-B unit distance. The process of monomers approaching each other did not correlate with the temperature across the spread of the simulations. Two simulations at 300 K resulted in no aggregation as did one at 277 K.

3.2.3.4 Relative orientations

In most of the simulations, the AFP1 protein units were found to change relative directional orientation frequently with respect to one another, indicative of a high degree of mobility (*Figure 3.9*). A few exceptions to this general trend were present – in several of the simulations a smaller spread of angle values or less frequent fluctuations were observed. In the simulations of configuration 3 at both 277 and 300 K and config 4 at 300 K the relative angle was consistent in the region of 120-150° after the 40 ns mark, as is configuration 2 at 277 K in the latter part of the simulation, with angle values restricted to approximately 150°. In contrast, configurations 1 at 300 K and 5 at 277 K settled on angles of approximately 50° with more fluctuation.

It was possible to correlate several angle preferences with the minimum distance separation data presented above in *Figure 4.8*. For configuration 3 both AFP1 units aggregated at a very early stage and remained together for the rest of the simulation. Their relative angle values fluctuated less rapidly, and settled into a stable and limited range of values after approximately 40 ns. Configuration 2 at 277 K exhibited a similar correlation. After the 10 ns mark the two units aggregated and, correspondingly, the angle between them was restricted to a value of roughly 150°. A similar trend was seen in configuration 1 at 300 K, which settled to angle values less than 60° after 50 ns, and configuration 4 at 300 K and

configuration 5 at 277 K, which settled at angle values around 120° and less than 80°, respectively, towards the end of the simulations.

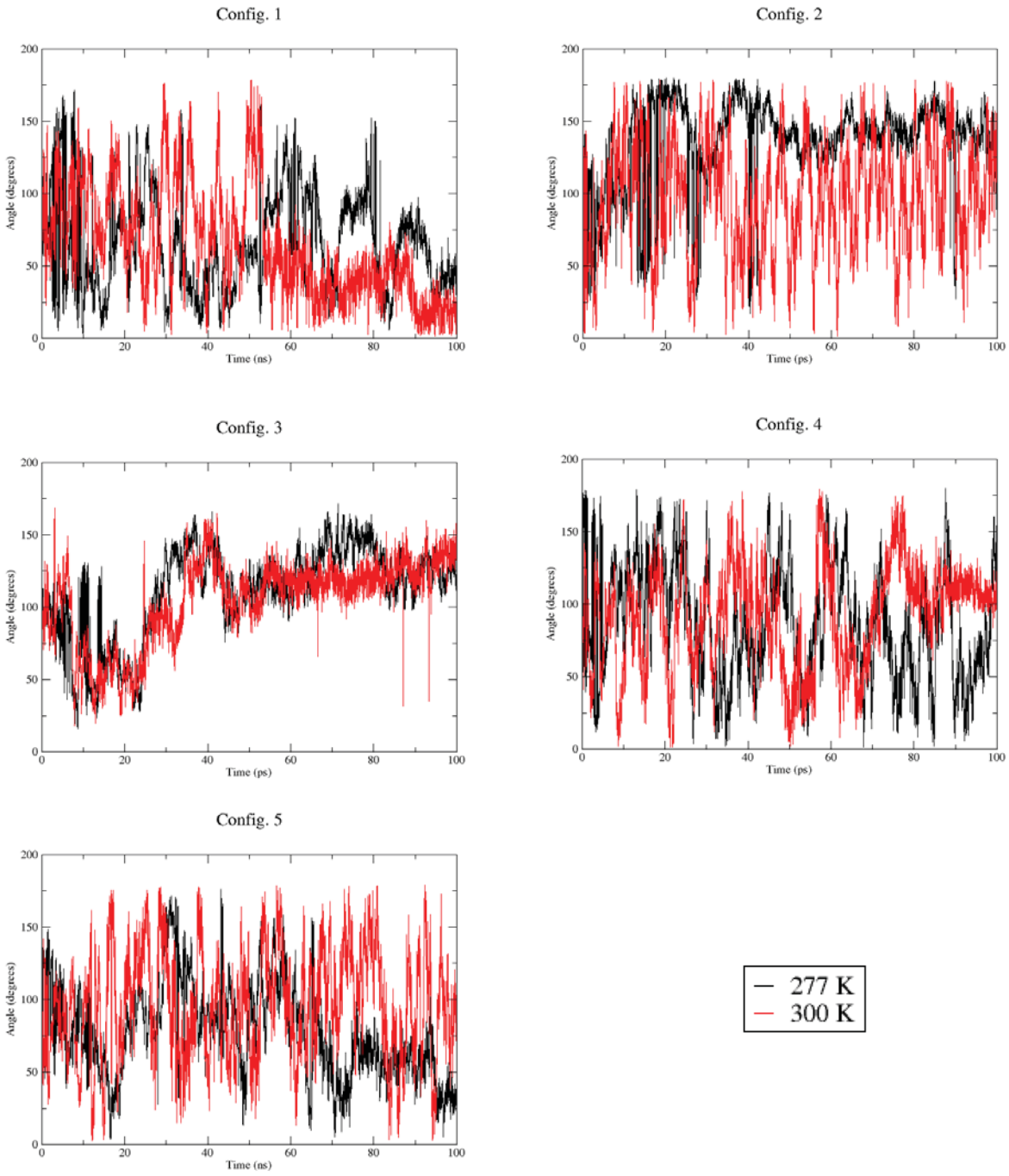


Figure 3.9: Time series of angles between AFP1 units A and B for all five configurations during simulations at both 277 K and 300 K.

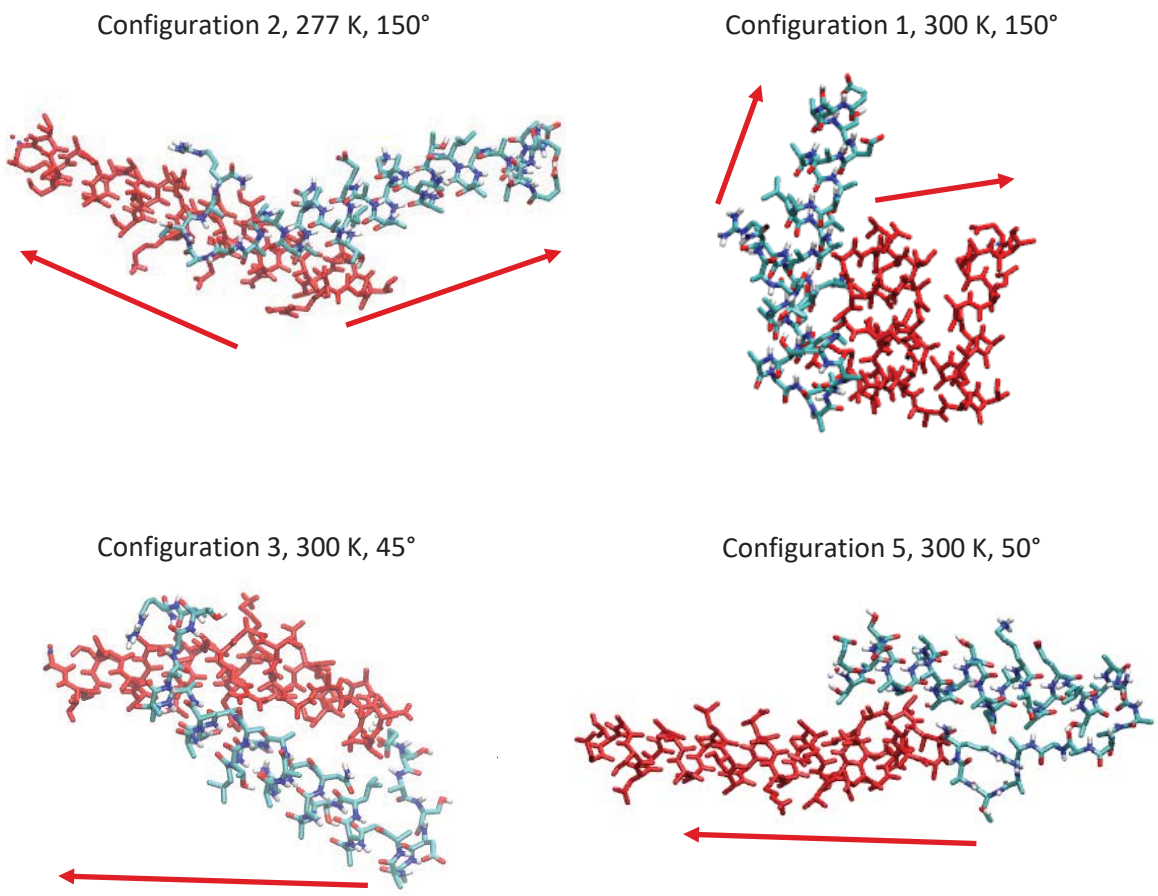


Figure 3.10: Snapshots of orientations of units A and B at 150° and 50° relative angles. Arrows outlining vectors from first to last atoms of the molecule are included for convenience.

These correlations may suggest that when two AFP1 units converged positions, their relative orientation angles became restricted. Typically, this "locking down" of the angle values occurred at a delayed time interval following the convergence, suggesting a delayed effect between initial contact and restriction of orientation. However, not all proteins that aggregated obeyed this trend. In configurations 1 at 277 K and 5 at 300 K small regions of preference for a particular angle could be found, but overall a wide degree of values was exhibited.

Simulation snapshots have revealed that in the majority of cases the proteins wrapped around each other and did not retain rod like structures comparable to those found the crystal structure (*Figure 3.10*). The angle program calculated angles of two vectors, defined by the first and last atoms of the protein subunits. When the rod like structure became folded or bent, the angle between vectors becomes dependent on relative positions of the first and last residues and may not have been indicative of the overall relative orientation of the entire unit. Therefore, loss of rod like structure and increasing flexibility in the molecule explained the high variation of angles in some of the simulations.

3.2.3.5 Hydrogen bonding

To assess the nature of any interactions that form between the two protein units as well as the interactions between the proteins and their environment the hydrogen bonding during the simulations was analysed (*Figure 3.11 and 3.12*).

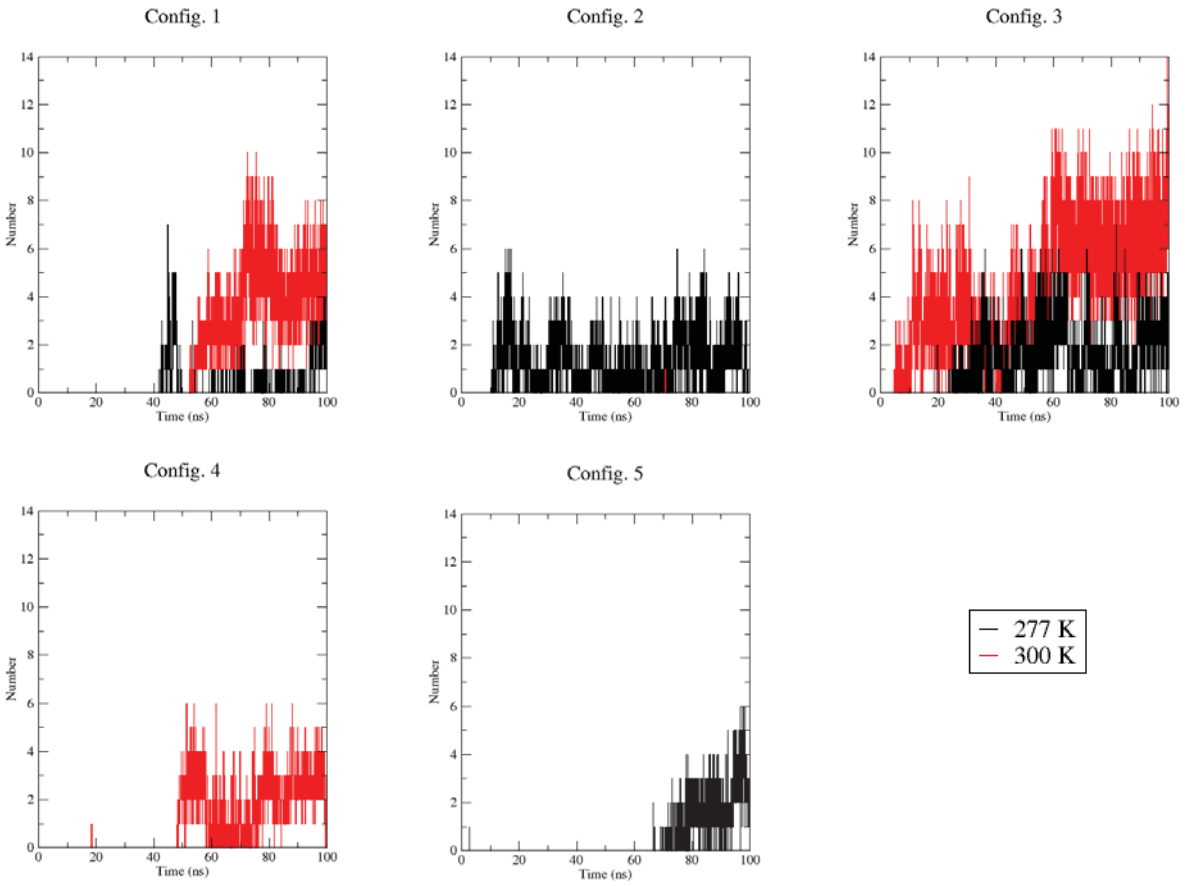


Figure 3.11: Number of hydrogen bonds formed between the two AFP1 protein units A and B during the simulations at 277 K and 300 K plotted as a time series.

The timing of hydrogen bond formation between the AFP1 proteins during the simulation was largely correlated with the minimum distance between the two protein subunits. In configuration 2 at 300 K almost no hydrogen bonds were formed between the units as the monomers did not aggregate. Similarly, in the simulations of configuration 4 at 277 K and configuration 5 at 300 K there were no hydrogen bonds formed due to lack of contact between the protein units. The number of inter-subunit hydrogen bonds varied with differing initial configurations and with simulation temperature significantly. Configurations 1, 2, 3 and 5 at 277 K and 4 at 300 K formed a maximum of 5 to 6 bonds. Configurations 1 and 3 (T-shaped) at 300 K formed up to 10 to 12 bonds.

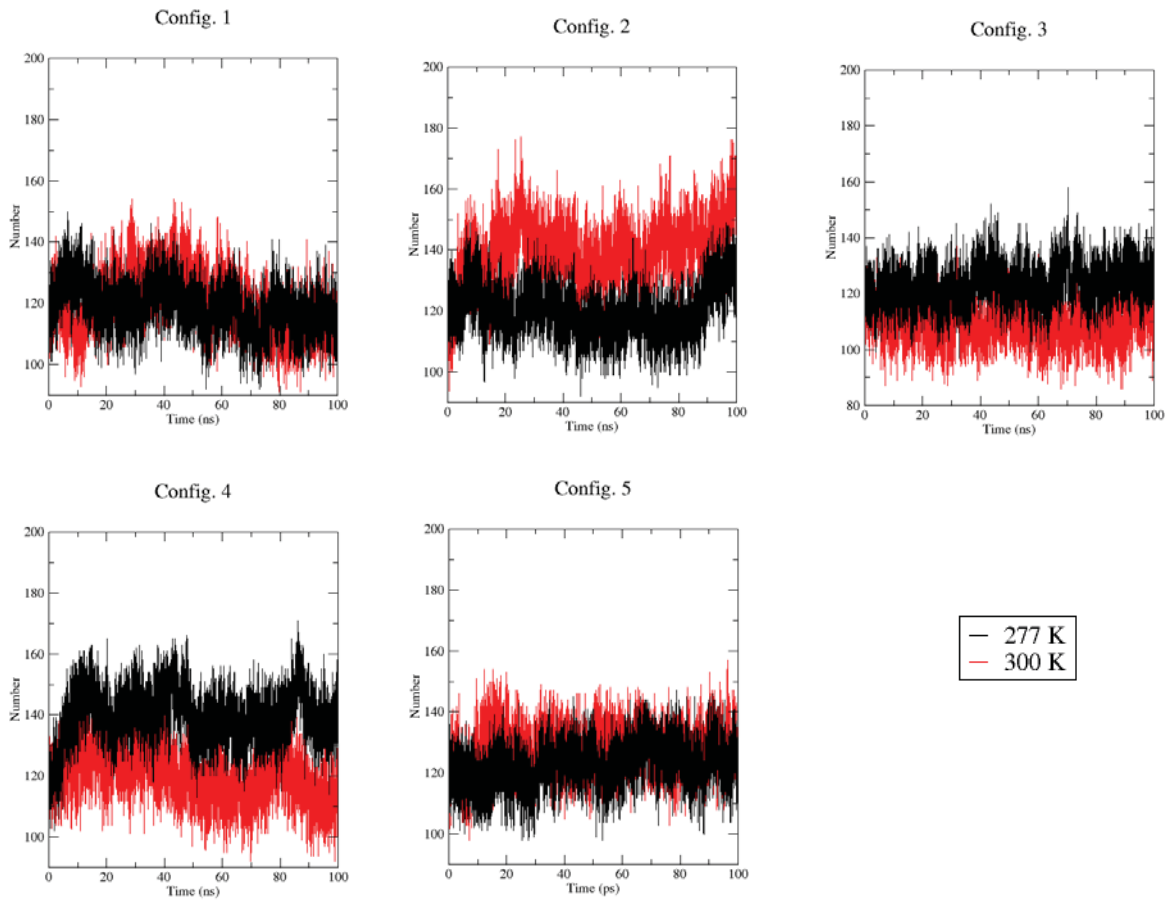


Figure 3.12: Number of hydrogen bonds between both AFP1 protein units and solvent during the simulations at 277 K and 300 K plotted as a time series.

It was found that in all of the simulations, both of the AFP1 protein units together formed 100-160 hydrogen bonds with the SPC water solvent (*Figure 3.12*). There was no general correlation between temperature and the number of bonds formed. In configuration 1 and 5 the number of bonds formed for either temperature was approximately equivalent. In configuration 2 the number of bonds was generally higher for the higher temperature simulation and in 3 and 4 the number was lower for the higher temperature.

A correlation could be drawn between the number of bonds formed between the proteins and the number formed with solvent. In simulations where the AFP1 units have not aggregated (configurations 2 at 300 K and 4 at 277 K) a higher average number of bonds with the solvent occurred, compared to other simulations. The difference was quantified to be approximately 15 bonds. However, for configuration 5 at 300 K, where the units did not approach each other, the difference in bonding was much less (approx. 5-10). The hydrogen bonds that formed between the protein units in the simulations where aggregation did occur could partially account for the difference. The number of inter-protein unit

hydrogen bonds is effectively half the number of bonds that could be formed to solvent if both units involved hydrogen bonded to solvent instead. The hydrogen bonding ability of several groups would be taken up with inter-protein bonding and neighbouring hydrogen bond donor or acceptor groups could be unfavourably positioned to bind with the solvent. From this an assumption was made that AFP1 units that did not converge positions have the maximum number of hydrogen bonds possible for the AFP1 protein, all of which were hydrogen bonds between AFP1 and water.

Analysis of hydrogen bonding pairs by residue between units A and B has revealed that typically residues in the second half of the molecule (residue number 16 and more) formed the most pairings between units A and B (Appendix 4). Prevalent bonding residues were ARG37, THR35 as well as the NH_2 terminal cap. However, a range of residues, including ALA backbone were also involved in cross unit hydrogen bond formation.

3.3 MD investigation of AFP1 with an unsaturated lipid membrane

The second research goal of the project was to investigate dynamics and interactions of AFP1 proteins with an unsaturated phospholipid membrane, representative of a psychrophilic organism cell membrane. To investigate the presence and nature of any such interactions, the AFP1 dimer (configuration 1) was simulated in the presence of a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer as an analogue of a cell membrane. This also allowed comparison of the behavior of AFP1 in the presence and absence of a lipid membrane.

3.3.1 System design

A POPC bilayer was chosen to serve as a basic model of an unsaturated cell membrane. Tomczak *et al*. reported possible AFP1 interaction with a 50/50 mixed DGDG (Digalactosyldiacylglycerol) and DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) lipid bilayer[96]. DGDG unsaturated membranes were of particular interest due to their stability at near freezing temperatures. However, while models exist for polyunsaturated monogalactolipids[110], not many other galactolipid parameters were available at the time of writing, especially for the digalactolipid family. Creating and parametrising new types of lipids would be outside of the scope of the project. Instead, the research focused on mimicing the unsaturated aspects of DGDG membranes, which was one of the main criteria for its original selection[94]. POPC was deemed to be the best analogue for the mixed DGDG/DMPC, as it contained both a saturated carbon tail and an unsaturated one, echoing the mixed bilayer properties of the experimental research. The POPC head group also matched that of one of the lipid types (DMPC) used in that research. Furthermore, the gel phase transition temperature of POPC bilayers is reported to be 271 K, ensuring that the bilayer remained in the liquid crystal phase at the necessary temperatures[111-112]. The chemical structure of POPC can be found in *Appendix 6, Figure A6.1*.

The initial protein coordinates used to generate the new membrane simulation were the same as for configuration 1, for convenient comparison with the previous simulations in water. The two AFP1 monomer protein units were added to a simulation box with the pre-equilibrated POPC bilayer[113-114]. The combined coordinates were then solvated with SPC water, and the entire system energy minimized, heated and equilibrated prior to a 100 ns simulation. The size of the system was limited by available computational resources and due to the periodic boundary conditions, the concentration of AFP1 could be considered higher than it might be under physiological conditions. However, this remains a reasonable representation of the local environment for a pair of AFP1 units interacting with a small portion of the membrane. Reference simulations of POPC lipid bilayer at 277 K and 300 K were also carried out. Further details of the simulation procedures are given in *Section 3.5: Methods*.

3.3.2 Internal structure

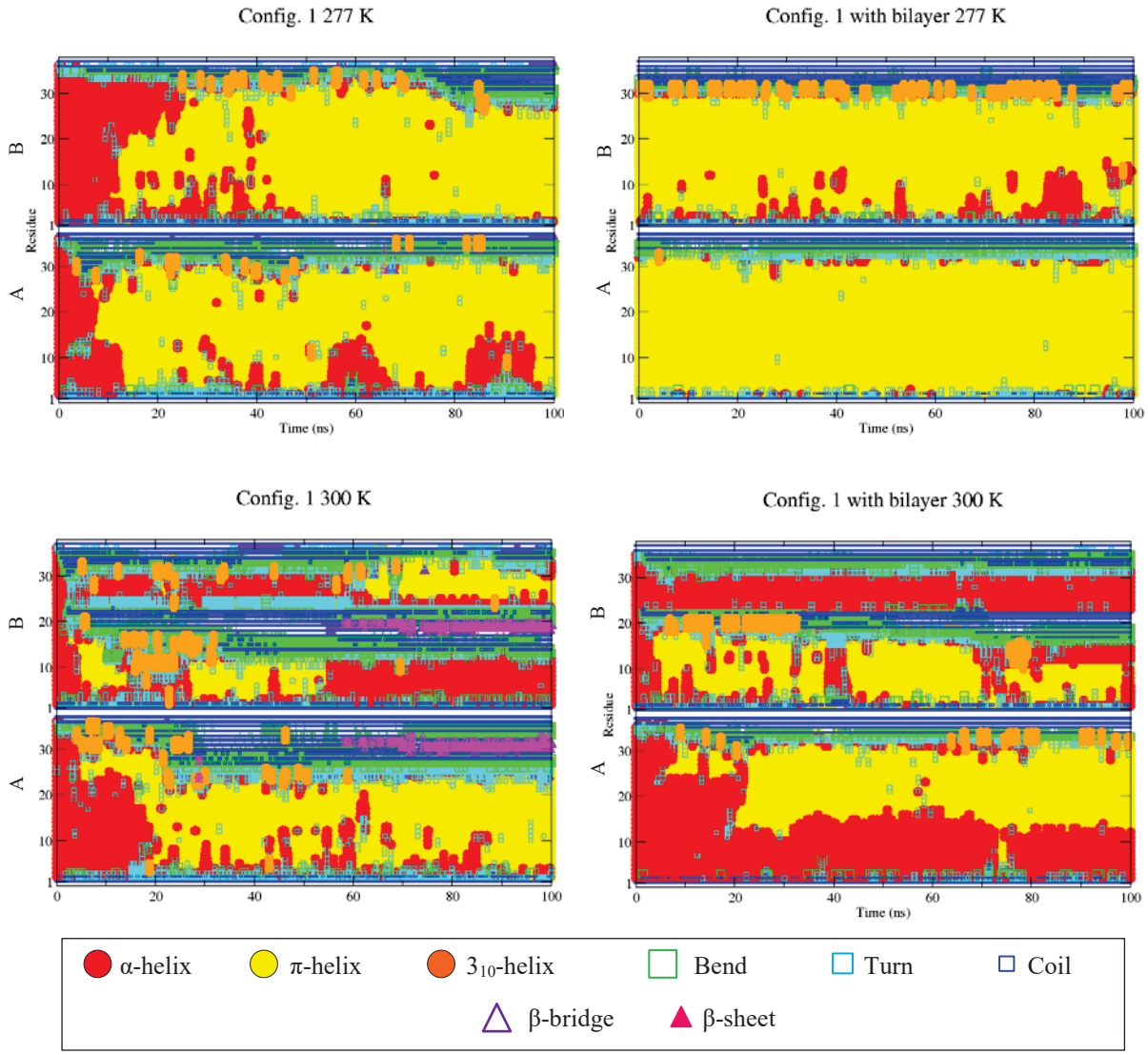


Figure 3.13: Secondary structure of AFP1 units A and B in the presence and absence of a POPC bilayer during simulation at both 277 K and 300 K.

The internal structure analysis of AFP1 proteins in the simulation with the lipid bilayer was carried out in the same manner as in *Section 4.2.2*. DSSP analysis was used to obtain information about secondary structure and the per-subunit atom-positional RMSD was computed to quantify the internal mobility of the AFP1 molecules. The results from the membrane simulations were compared with the results from the water only simulations.

3.3.2.1 DSSP

Clear differences in internal structure were observed via comparison of simulations of AFP1 with and without the bilayer (*Figure 3.13*). At 300 K the α-helical character of the crystal structure was preserved nearly all the way through the simulation for around half of the residues (residues 1 to 15 for unit A and residues 22-38 for unit B, approximately). In contrast, in the 277 K simulations, no α-helical character was observed at all in the bilayer simulation. Both proteins exposed to the POPC bilayer were predominantly π-helical, as in the simulations without a bilayer present. The structures of AFP1 units with the bilayer present were more ordered then their counterparts without the bilayer, in that they exhibited a lower degree of turn, bend and coil structures near the ends of the proteins.

3.3.2.2 RMSD (internal)

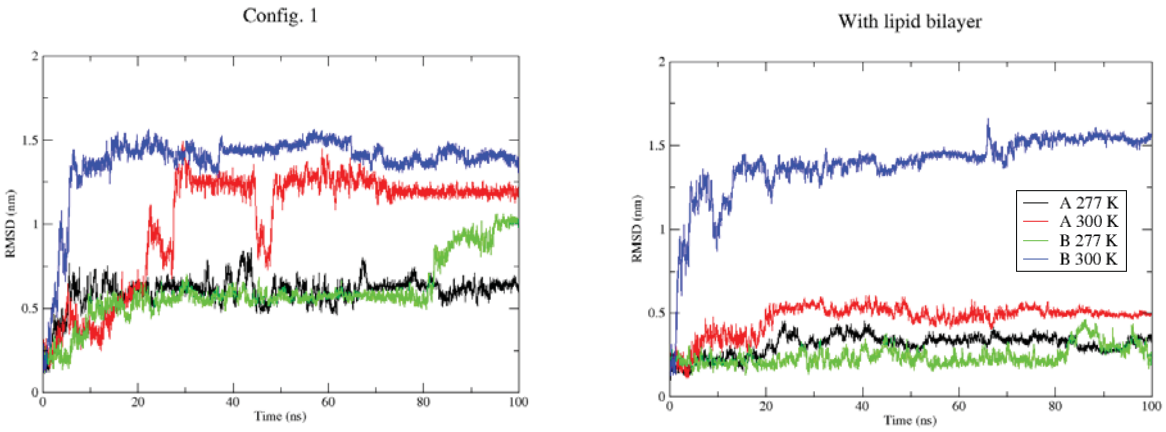


Figure 3.14: Atom-positional RMSD of all atoms of AFP1 subunits A and B mapped on to itself (A on A and B on B) at 277 K and 300 K from the simulations at both 277 K and 300 K (a) without a bilayer (configuration 1) and (b) with a POPC bilayer.

The RMSD values of monomer units A and B fitted onto themselves were similar to the results shown above for the simulations of AFP1 without the bilayer (*Figure 3.14*). Both monomers appeared to show greater internal flexibility at the higher temperatures as expected, however this effect was more pronounced for monomer B, especially in the presence of the POPC bilayer, despite the elements of α-

helical structure of unit B. RMSD averages for the AFP1 units at 277 K were lower in the simulation with the membrane, indicating more restricted internal movement.

3.3.3 Interactions

3.3.3.1 RMSD (relative mobility)

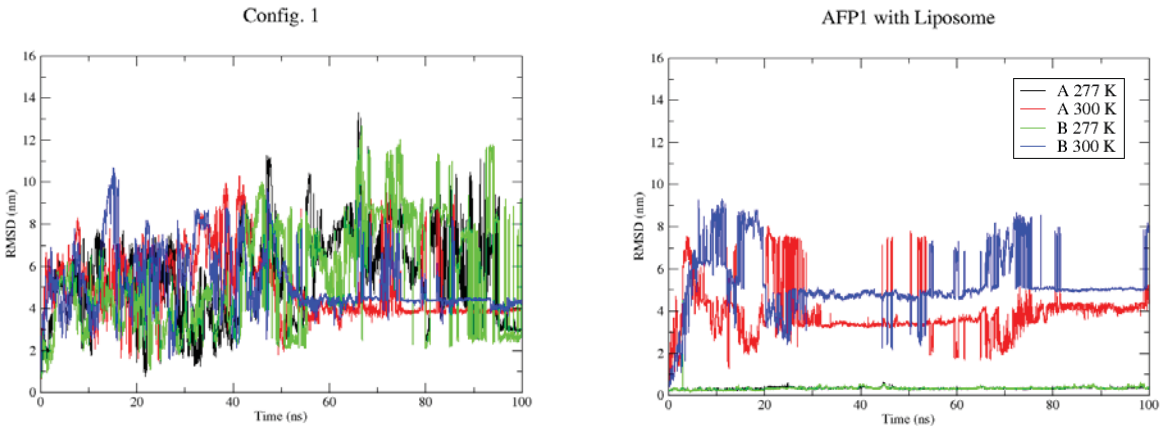


Figure 3.15: Comparison of atom-positional RMSD values at 277 K and 300 K for individual AFP1 units A and B after fitting both subunits simultaneously to their coordinates at the start of the simulation. (a) AFP1 in solution; (b) AFP1 with the POPC bilayer.

Table 3.3: Atom-positional RMSD values (in nm) averaged over the entire simulation at 277 K and 300 K with and without a POPC bilayer present for individual AFP1 units A and B after fitting both subunits simultaneously to their coordinates at the start of the simulation, standard deviation σ is included as a measure of error.

	Config. 1	*Config 1 with POPC bilayer*	*σ*
A 277 K	5.472	0.330	0.082
A 300 K	5.074	3.912	0.086
B 277 K	5.510	0.335	0.268
B 300 K	5.217	5.222	0.220

Clear differences were observed in the relative mobility of the AFP1 subunits in the presence of a POPC bilayer compared to in the reference simulation. The RMSD values did not fluctuate as rapidly in the presence of the bilayer. In the simulations at 277 K, the RMSD (after fitting to both subunits

simultaneously) was very low (approximately 0.33 nm), which indicated a lack of relative mobility, suggesting a stable complexation of the AFP1 subunits with each other or the bilayer. The simulations of AFP1 with the POPC membrane at 300 K showed higher RMSD values than the 277 K counterparts, but were still lower on average then in the simulation without a bilayer for unit A (*Table 3.3*). The results were strongly suggestive of restricted relative mobility of the AFP1 units with respect to one another in the presence of the bilayer.

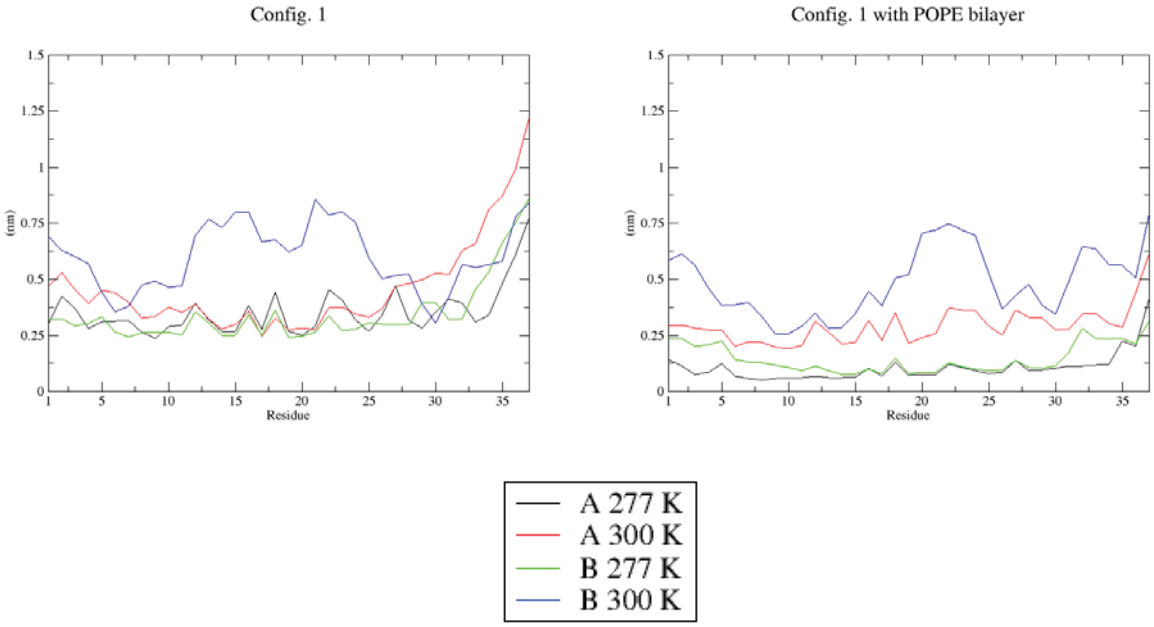


Figure 3.16: Comparison of the per- residue RMSF of AFP1 units during simulations at 277 K and 300 K with and without a POPC bilayer present.

Generally, the AFP1 RMSF values were found to be lower for the simulations in the presence of a POPC bilayer, further adding to the evidence suggesting that AFP1 proteins become less mobile as result of their interaction with the bilayer. The RMSF values were higher for the 300 K simulations, especially for unit B, which exhibited high RMSF values in the middle of the molecule regardless of the presence of the bilayer.

3.3.3.2 Distance between AFP1 subunits

Calculation of the minimum distance between each subunit and the POPC bilayer showed that both AFP1 subunits rapidly approached the POPC bilayer, making lasting contact almost immediately at 277 K and within 20 ns at 300 K (*Figure 3.17*). Both AFP1 proteins spent the rest of the simulation time in contact with the membrane. Additionally, in contrast with the original simulations without the bilayer

(*Figure 3.8, Config. 1*), the AFP1 subunits approached each other at the very start of the simulation, rather than at the 40 to 50 ns mark.

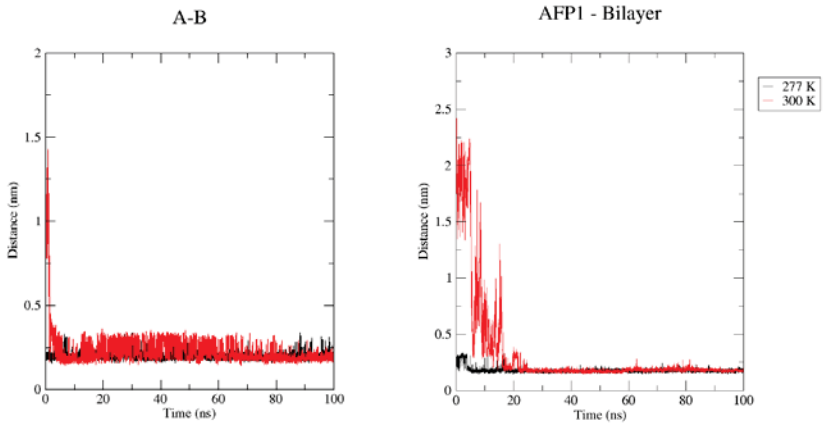


Figure 3.17: Minimum distance between AFP1 monomer units A and B and to the bilayer during the 277 K and 300 K simulations in the presence of the POPC bilayer: (a) A – B, (b) A – bilayer and (c) B – bilayer.

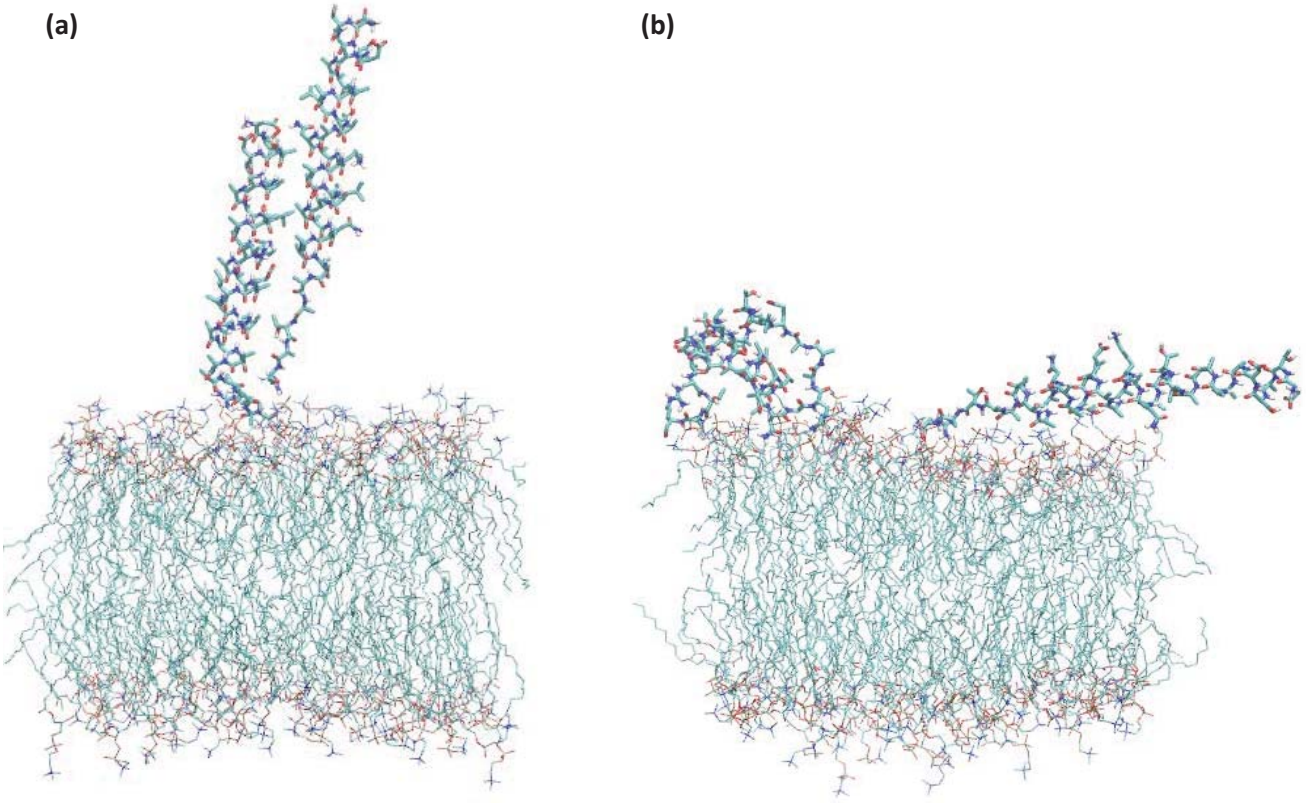


Figure 3.18: AFP1 protein interaction at 100 ns of simulation time with the POPC lipid membrane at (a) 277 and (b) 301 K.

The AFP1 units remained together after making contact, as they did in the original simulation, further indicating mutual interaction. It appears that both the protein units and the bilayer experience mutual attraction and all the objects aggregated together in one space. The close proximity of the AFP1 units to each other and to the membrane at the start of the 277 K simulations may be caused by convergence during the heating and equilibration stages, despite initial starting distances between the protein units being in excess of 1.4 nm (above cut-off distance).

In the bilayer simulations it was found that the N-terminus of unit B was situated next to the bilayer and the C-terminus was found at a distance away, indicating that at both temperatures unit A acted as a rod situated perpendicular to the bilayer surface (*Figure 3.18*). Unit B residues were found to uniformly closely situated to the bilayer surface, indicating that the rod like structure was aligned with the x y axis of the simulation. The protein unit to unit distance as generally close, indicating interaction. In the case of Unit A residues at 300 K the N-terminus was situated further away. And in the case of unit B, the C-terminus was found at a greater distance away then the rest of the molecule at both temperatures.

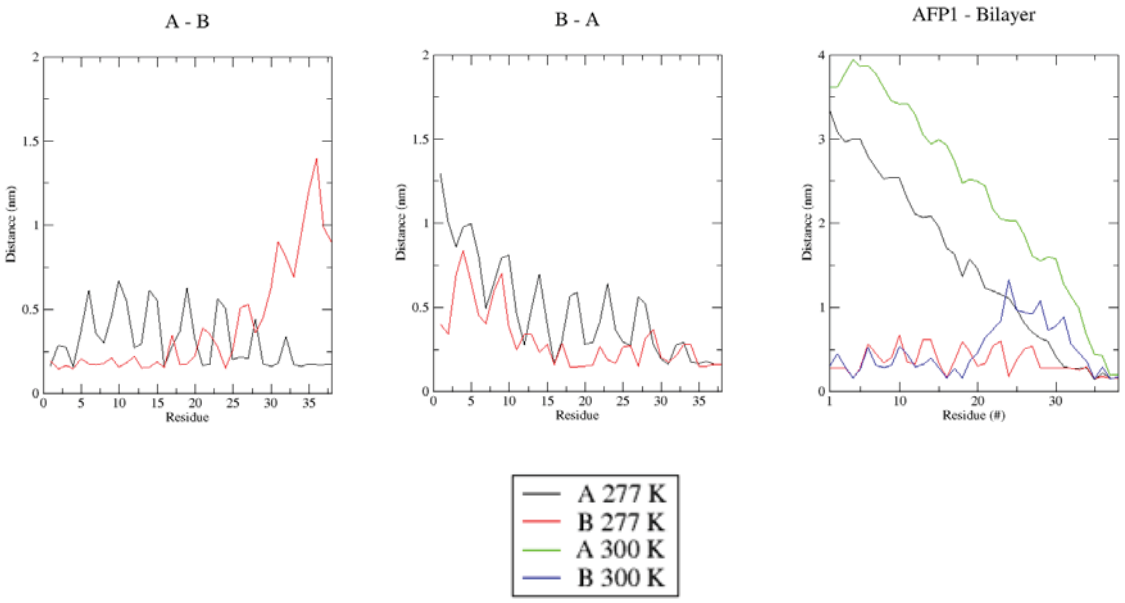


Figure 3.19: Minimum distance from each residue of the specified (first) subunit of AFP1 to any atom of the other (second) subunit during the 277 K and 300 K simulations in the presence of the POPC bilayer: subunit A – subunit B, subunit B – subunit A and subunit A or subunit B – bilayer

3.3.3.3 Relative angles

The relative AFP1 inter-unit angle distribution was different in the presence of the lipid bilayer. At 277 K AFP1 were locked into a small range of values, at approximately 20°, which indicated a locking of orientation and conformation due to possible interactions. The angle value is much smaller than that of

any of the non-bilayer experiments (*Figure 3.9*). At 300 K the effect was less pronounced, but nonetheless the angle of AFP1 with the membrane had far less variation than in the reference simulations, again demonstrating restriction of orientation as a result of membrane interactions.

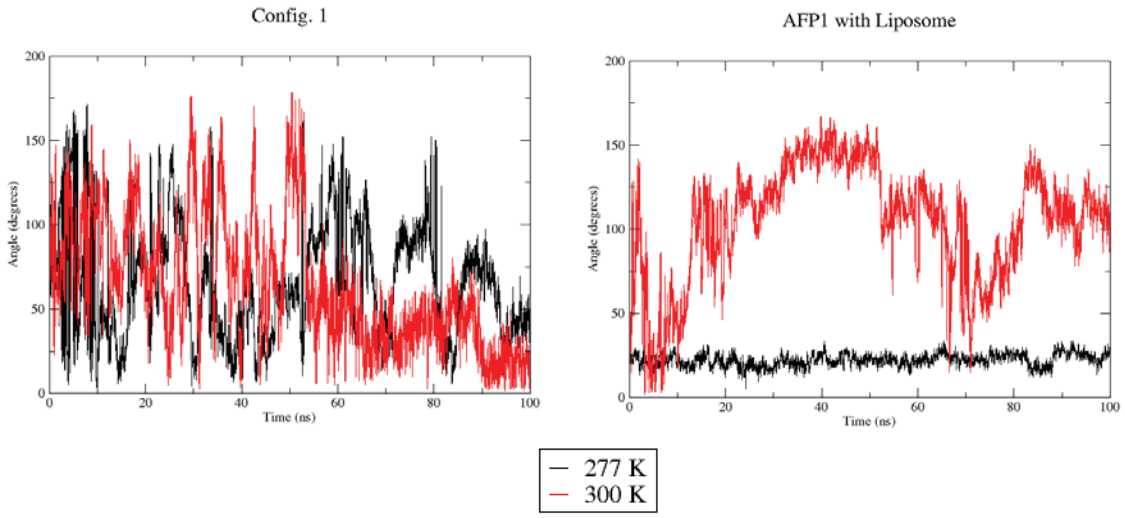


Figure 3.20: Relative angles between AFP1 units in simulations with and without the liposome, at 277 K and 300 K.

3.3.3.4 Hydrogen Bonding

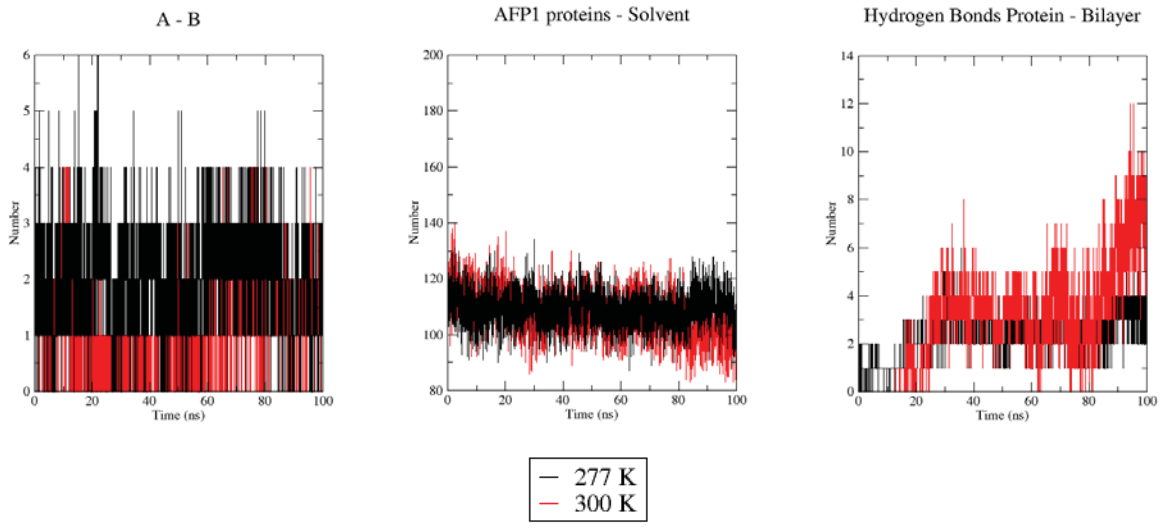


Figure 3.21: Time series of number of hydrogen bonds at 277 K and 300 K: (a) AFP1 units A and B to POPC bilayer, (b) Monomer A – Monomer B, (c) AFP1 units A and B – solvent.

Up to four hydrogen bonds were detected between the two AFP units, increasing up to a maximum of 6 in the 277 K simulation. This number was much smaller compared to the original simulations of AFP1 in water, despite the units converging positions very early in the simulation. The number of bonds between the AFP1 units together and the solvent was again lower than in the water simulations. If the number of hydrogen bonds that AFP1 units could form in water would be taken as the maximum possible amount of hydrogen bonds formed by the molecule (approximately 145, see *Figure 3.12*, configuration. 1), it can be deduced that some of the hydrogen bonding ability of AFP1 proteins was being utilised elsewhere. Furthermore, the number of hydrogen bonds between AFP1 and the solvent decreased throughout the 100 ns simulation. Protein bilayer bonds started appearing early on in the simulations and the number increased up to 5 in the 277 K simulation and 12 in the 300 K simulation. This lent more evidence that AFP1 hydrogen bonding groups stopped forming bonds with solvent and instead formed bonds with each other and the liposome surface.

Hydrogen bond forming residues in the interaction of units A and B in the presence of the lipid bilayer were found to be more varied than in previous simulations, covering the whole range of the molecule rather than residues 16 or above and again included both the ALA and the functionalised residues alike. Key residues for hydrogen bond formation between the AFP1 units and the POPC membrane found to be THR35, ARG37 and the NH2 terminal cap.

3.3.4 Bilayer properties

The analysis of lipid membrane behaviour was compared between the reference membrane simulations (without AFP1) and the simulations of AFP1 with the membrane. Calculation of the area per lipid gives information about any possible changes in the dimensions of the membrane during either simulation, or due to the presence of AFP1. Deuterium order parameters for the lipid tails were also calculated to assess whether AFP1 presence had any effects on the bilayer structure.

3.3.4.1 Area per lipid

Area per lipid (A_l) can be considered a good indicator of bilayer packing, which in turn affects bilayer thickness and reflects the phase of the bilayer (e.g. gel or liquid crystal). Area per lipid was found to be similar throughout all the POPC simulations (*Table 3.4*). However, A_l was found to be marginally higher in the higher temperature simulations, as expected. There was also a slight decrease in Area per lipid in the presence of AFP1 at 277 K, but not at 300 K.

Table 3.4: Area per lipid for the POPC bilayer simulated with and without AFP1 present at 277 K and 300 K in nm^2.

Temperature	*POPC*	*POPC with AFP1*
277 K	0.597 +/- 0.005	0.585 +/- 0.001
300 K	0.615 +/- 0.005	0.617 +/- 0,002

3.3.4.2 Deuterium order parameter estimation

As an indicator of bilayer lipid ordering and packing order parameters were calculated for the aliphatic tails of the POPC lipids in the bilayer (*Figure 4.22*). Order parameters are highest near the lipid head group (atoms with low numbers), and decrease towards the tail, reflecting the expected higher mobility of the lipid tails. It was found that order parameters for both temperature ranges were marginally higher in the presence of AFP1 proteins, indicating a slight increase in the degree of order for the membrane lipid tails. This type of behavior has been observed in other situations where a peptide or protein bind to a membrane. Order parameters were higher for the lower temperature simulations, indicating greater ordering and more structure at near freezing temperatures as expected.

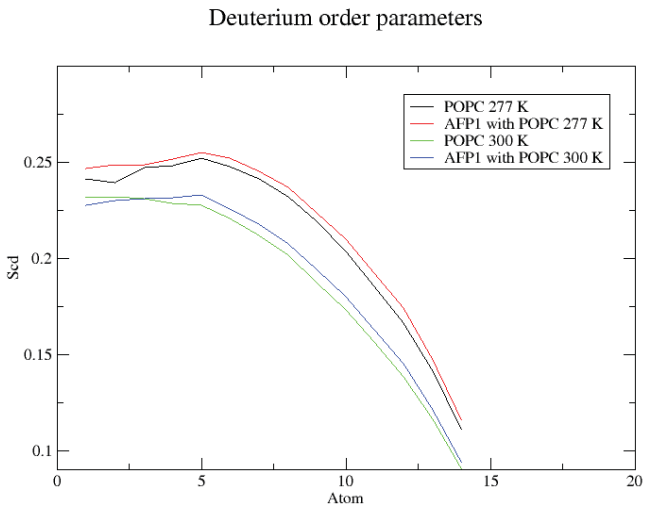


Figure 3.22: Order parameters for the aliphatic carbon tails of the POPC bilayer simulated with and without AFP1 present at 277 K and 300 K.

3.4 Discussion

The primary goal of this project was to investigate the structural dynamics and possible dimerisation modes of a pair of AFP1 monomers using MD simulations. It was of particular interest to find out whether the AFP1 units converged to form a precursor complex, which could subsequently lead to dimer or higher order oligomer formation. AFP1 units were simulated in differing relative orientations to investigate the correlation between orientation and their ability to aggregate. The secondary goal was to simulate and investigate the interactions of the two AFP1 protein units with a model of a psychrophilic cell membrane. All simulations were performed at near freezing and room temperatures to analyse whether dynamics were affected by temperature as they were in experimental studies.
During the simulations of AFP1 the monomer units adopted a variety of structures, rapidly losing their rigid α-helical nature present in the initial (X-ray crystal) structure, especially towards the ends of the molecules. It may be that the helicity of the crystal structure is in part due to crystal packing, and is not stable for isolated monomers in solution. The α-helices rapidly turned into π-helices in the middle of each unit. A range of bend, turn, and coil structures was observed around either end of the protein indicating flexibility in the C- and N- termini regions. Helix fraying is not unusual for isolated α-helical structures in solution. This behaviour nonetheless indicated that AFP1 proteins could be less rigid in solution than previously believed. Surprisingly little correlation between secondary structure and simulation temperature has been found, going against the initial hypothesis of predicting temperature dependent structural differences between the experiments.

The large-scale changes in secondary structure are matched by relatively high internal atom-positional RMSD values. The atom-positional RMSF revealed that the majority of these motions are due to flexibility of the amino acid side chains, particularly of the functionalized side chains around the residue regions that are linked to the AFP1 ice binding ability. Thus, overall, although the structure of each AFP1 subunit changes from its helical crystal structure, the structural changes are likely to be related to its antifreeze function.

The crystal structure of AFP1 winter flounder at 277 K shows two monomer units situated in proximity to each other as rigid α-helical rods (*Figure 3.2*), supporting the possibility of aggregation and dimer formation at low temperatures. The simulations here started from a range of different initial orientations, with the proteins units separated, to test the determinants of dimer formation. In the majority of the simulations, at least some residues of the two AFP1 units came close together and remained close, as evidenced by a low minimum distance. There was no clear link, however, between the final configuration, or even the formation of the aggregate, and the initial relative orientations. The AFP1 units coiled around each other, rarely resulting in stable relative orientations, party as the relative orientation became hard to define as a vector when the secondary structure of the units became flexible.

The formation of multiple hydrogen bonds between the units indicated formation of a stable complex. Such complexes may be a precursor to dimer formation and perhaps, ultimately, higher-order

aggregation, *in vivo*. The hydrogen bond interactions only occurred after a time delay compared to the initial (stable) contact between the subunits, showing that it took a certain amount of time to develop complex bonding patterns. Hydrogen bonding evidence suggests that functionalised groups towards the C-terminus, such as ARG, THR, NH2 terminal group as well as some ALA amino acids played a greater role in AFP1-AFP1 bonding.

The presence of an unsaturated membrane changed the AFP1 behaviour compared to the original simulation in water. The two AFP1 protein subunits rapidly approached each other and the membrane in both membrane simulations. In the 277 K simulation, the proteins came to rest lying flat on the surface of the membrane, whereas in the 300 K simulation they approached the membrane as α-helical rods and became positioned on the membrane surface end on. It is possible therefore, that the orientation of the AFP1 relative to the membrane and retention of helicity may be important to the mechanism of AFP1 action, as the orientation of the rod-like units may be key in binding developing ice-lattices and preventing ice crystal formation near the membrane surface. In that respect, at 277 K the orientation of the AFP1 relative to the membrane could cover and therefore protect more of the membrane surface.

The structures exhibited by the AFP1 protein units in the presence in the lipid bilayer showed reduced mobility and a restriction of structure compared to in solution. In some cases, some degree of helical structure was retained, which appeared to be more rigid when the protein units interacted with the membrane. Higher degrees of hydrogen bonding developed between the AFP1 units and the membrane as the simulations progressed, indicating preferential binding affinity between AFP1 units and the membrane. The primary residues involved at 277 K were THR35, ARG37 and the NH2 terminal group. At 300 K, more interactions were seen involving these residues with the addition of a further ASN16 and LYS18. These residues were the same ones that were involved in AFP1-AFP1 unit bonding earlier and therefore the functionality of these particular residues is switched from inter-unit bonding to unit-membrane bonding due to preferences in bonding.

The membrane parameters were not greatly affected by the presence of the AFP1 proteins, with the deuterium order parameters and area per lipid remaining largely similar between the reference and the AFP1 simulations. Membrane order parameters were found to be higher in the presence of AFP1 by a small margin. The effect of temperature on order parameters was found to be greater than that of AFP1 presence.

Temperature, overall, has not been found to correlate greatly with AFP1 behaviour in these simulations, which was contrary to literature findings. In the literature AFP1 was found to change structure and lose functionality at higher temperatures. However, a potential caveat of *in silico* simulations could be that temperature variation could not be very physical. Altering simulation temperature results in molecules moving faster or slower and is dependent heavily on the reliability of thermostat and the forcefield, which are both parametrised to work best at near room temperatures, where the forces due to the forcefield will be in balance with those due to the kinetic energy.

On an additional note, initially five sets of different starting configurations of the AFP1 molecule were utilised to give a broader indication of the structures adopted by the proteins as well as investigating whether initial positions had any effect on behaviour. While the five different configurations cannot be constituted as exact replicas, having multiple simulations did allow access to a greater amount of structural sampling, at least prior to aggregation. Simulating five sets of coordinates was computationally expensive, however, we can clearly see that the results of each simulation are quite different from each other, which can be argued was beneficial in establishing an extended spectrum of results. Three of the simulations followed the predicted trends in terms of temperature dependency while the other two showed contrary evidence. For similar simulation projects, it could be beneficial to consider the use of enhanced sampling methods such as Hamiltonian replica exchange (HREX)[115-116] to access more available structures at a reduced computational cost. Furthermore, as it was found that as the effect of simulation temperature did not correlate greatly with dynamics of AFP1 proteins, it might be of benefit to further research to run simulations at higher temperatures to increase conformational sampling [117].

In future projects it would be of great interest to extend simulation types and use more membrane models involving a wider range of lipids. It would also be beneficial investigate other similar antifreeze proteins and some of the reported hyperactive polymer structures to discern any trends in their dynamics and interactions. Finally, different forcefields that allow use of alternative water models should also be explored, as water models that are flexible or include an extra pseudo-atom to mimic the lone pairs on the water oxygen might provide different results, particularly given that the reproduction of experimentally-measured properties of water different between models.

3.5 Methods

3.5.1 Simulation

Preparation, running and analysis of the simulations were performed using GROMACS 5.1.1[62] utilising the GROMOS 54a7 forcefield[68] and the SPC water model[69-70].

3.5.1.1 Coordinates and parameters

Winter flounder AFP1 X-ray crystal structure coordinates[83,107] were obtained from the Protein Data Bank[106]. AFP1 parameters were built from amino acid building blocks from the 54a7 forcefield according to the PDB sequence. The ionization state of the amino acid side chains and N- and C-termini of the peptides were assigned as per a typical physiological pH of 7.0, giving an overall charge on the peptide of –1. The termini were not capped. Coordinates for the initial arrangements of the two monomers (configurations 1-5) were built using the GROMACS edit_conf program to translate and rotate one monomer with respect to the other (*Figure 3.3*). POPC parameters compatible with the 54a7 forcefield

and coordinates of a pre-equilibrated (250 ns) POPC bilayer section comprised of 128 lipids were obtained from the ATB database[6], submitted by David Poger[118]. The membrane was re-equilibrated under NPT conditions for 1 ns and simulation of the membrane was carried out for 100 ns to provide reference simulations for membrane analysis.

3.5.1.2 Preparation

Each arrangement of two AFP1 units was solvated in a cubic box of SPC water with a 1.4 nm minimum distance to simulation box edge. Each system was energy minimised using steepest descent algorithm for 50,000 steps, heated under NVT conditions in 6 increments from 60 K to the relevant simulation temperature and then equilibrated for 1 ns under NPT conditions (277 K or 300 K; see below for simulation details).

The POPC bilayer was energy minimised over 50,000 simulation steps using the steepest descent algorithm with an energy minimization step of 0.01. Equilibration under NVT conditions was performed for 100 ps, and the system was heated from 50 K to the relevant simulation temperature (277 K or 300 K) prior to equilibration under NPT conditions for 1 ns.

For the simulations of AFP1 with a POPC bilayer the configuration 1 AFP1 coordinates were added to an extended simulation box containing the equilibrated POPC bilayer and the entire system solvated with SPC water. The system was energy minimized and re-equilibrated in the same manner as described above.

3.5.1.3 Production

Each of the five AFP1 configurations, the POPC bilayer alone, and the POPC bilayer with the two AFP1 subunits in configuration 1 were each simulated for 100 ns at 277 K and 300 K. The cut-off for calculating non-bonded Coulomb and van der Waals interactions was 1.4 nm. Long-range electrostatic interactions were approximated using the reaction-field[20]. Bonds were constrained using the LINCS[71] constraint algorithm with an order parameter of 4, allowing for an integration time step of 2 fs. The temperature was maintained using the Berendsen thermostat[72], with temperature coupling time τ_T = 0.001 ps^{-1}. Berendsen isotropic pressure coupling was used For the NPT simulations, the pressure was maintained at a pressure of 1 atmosphere using the Berendsen barostat with pressure coupling time τ_P = 0.5 ps^{-1} and isothermal compressibility = 4.5×10^{-5} atm^{-1}. In the case of simulations involving the POPC lipid bilayer, the pressure coupling was changed to semi-isotropic. Coordinates and energies were written out every 50 ps.

3.5.2 Analysis

GROMACS 5.1.2 software was used to analysis the trajectories. The alphabetical list of analyses is listed below.

Angles
Angles between AFP1 units A and B were calculated using the GROMACS gangle program, which calculated the angle between two vectors. Vectors were defined as running between the first and last atom of each unit.

Area per lipid
Area per lipid was calculated by obtaining the simulation box x and y dimension averages and dividing by the number of lipids per leaflet (64). The average box dimensions were calculated with the gmx energy program.

DSSP
DSSP analysis[73-74] was performed using the GROMACS do_dssp programs. Outputs were converted by a python script into an ASCII format for plotting.

Hydrogen bonds
Hydrogen bond formation was determined using the GROMACS hbond program. Results were post-processed using an in-house python script to label the most populated hydrogen bonds with the residues and atoms involved.

Minimum distance
The minimum distance between pairs of AFP1 units or between an AFP1 unit and the lipid bilayer were calculated using the GROMACS mindist program.
Order parameters
Order parameters were calculated using the GROMACS g_order program. Special index files were created containing the atoms at the start and at the end of each lipid tail to be used in conjunction with the program.

RMSD
The atom-positional root-mean-square deviations were calculated using the GROMACS rms program.

When used to report on the internal structure, all atoms of one AFP1 subunit were fitted to the starting coordinates (of the production run, i.e. the coordinates from the end of the NPT equilibration) of the same unit (atoms 1-275 for unit A, 276-570 for unit B), and the RMSD was then calculated for all atoms of that unit only.

When used to report on the general mobility of one subunit with respect to the other, the coordinates of both of AFP1 units (atoms 1-570) were fitted to their initial coordinates, but the RMSD was again calculated for all atoms of each unit individually.

RMSF

The atom-positional root-mean-square fluctuations were calculated using the GROMACS rmsf program and both printed per-atom, and summed over all atoms to give the per-residue values.

Secondary structure

Secondary structure content was analysed using the DSSP program called using the GROMACS do_dssp program. The outputs were converted using an in-house python script into an ASCII format for flexible plotting.

Diffusion coefficient

The diffusion coefficient was also calculated for each AFP1 unit using the GROMACS msd program, which carries out a linear regression fit to the mean-squared displacement. However due to the limited sampling of conformational space, the uncertainties were of a similar magnitude to the coefficient values, therefore the diffusion coefficient data was not considered further.

Visualisation

VMD software package version 1.9.2. was used to visualise the GROMACS coordinate trajectory files[75]. VMD images were screen captured and edited for use in this thesis.

3.6 Conclusion

Antifreeze proteins are integral in preventing ice induced damage to the cells and have long been thought to interact with cell membranes at low temperatures. The research focused on MD simulation and analysis of AFP1 proteins at near freezing and room temperatures, as well as simulating the AFP1 proteins in the context of a lipid bilayer, as a model for the cell membrane.

Simulations of several AFP1 configurations have demonstrated the loss of strict α-helical general structure, replaced with π-helical structure in the majority of configurations. It was found that the most mobile parts of the structure were the functionalised side chains of residues key to the protein's ice binding activity. This suggested that AFP1 was less rigid in solution than previously thought and that the structure differed to that exhibited in the crystal structure. Despite previous literature findings, no strong correlation between AFP1 dynamics and temperature was found.

The two AFP1 monomers were found to converge positions in the simulation in the majority of cases, forming hydrogen bonds between the units. Initial variance in relative orientation did not correlate to whether AFP1 units aggregated. When AFP1 units aggregated, their relative orientations became more restricted. Commonly the "second" half of a protein unit would bind to the other, followed by proteins coiling around each other. While interaction between the two units certainly occurred, it is still unclear whether any of these conformations could lead to a stable complex or a dimerisation precursor.

AFP1 behaviour changed in the presence of the POPC unsaturated phospholipid membrane, but had little impact on the physical membrane characteristics. AFP1 monomer units and the membrane very quickly aggregated together. AFP1 retained more of its original α-helical structure in the higher temperature simulation. In both simulations, AFP1 units interacted and formed hydrogen bonds with the lipid bilayer. In the lower temperature simulations, the AFP1 units approached the membrane as rods and pointed away from the surface. In the higher temperature experiment, only one of the proteins exhibited this behaviour, while the other unit coiled on the surface. The orientation and direction α-helical rods might prove physiologically important if they remain in a stable position on the lipid membrane and potentially stop ice crystal formation near the membranes.

In conclusion, this study demonstrated that AFP1 proteins exhibited a wide range of secondary structure motifs in solution, and were capable of forming varying complex structures together stabilised by hydrogen bonds. In the presence of an unsaturated POPC bilayer representative of a psychrophilic organism's cell membrane, the helical secondary structure was stabilised on the membrane surface, and the AFP1 monomers appeared more likely to form a stable complex, and also interacted with the membrane surface, forming extensive protein-membrane contacts.

Abstract

Lipids comprise a key component of the cellular membrane and play an essential part in many biological processes. *In silico* investigations provide valuable opportunities to study dynamics and structure of biologically relevant constructs at near atomic resolutions. An attractive opportunity is posed by supra-atomic approaches to save computational resources when simulating large numbers of identical units, such as lipids in a bilayer. We propose a series of novel supra-atomic phospholipid models that are based on and are compatible with the GROMOS 54a8 atomic-level forcefield[109] including the GROMOS CG water model[119]. These models are intended for use in multi-scale simulations, and are also polarisable, unlike many popular supra-atomic models. The lipid headgroup is split into five beads mapped by group functionality and geometric properties. This facilitates a "plug and play" approach for building models of multiple types of lipids, as different functional group beads are available and interchangeable. The polarisability of the lipid headgroup arises from each supra-atomic bead comprising two oppositely charged particles utilising a charge-on-a-spring model. The lipid tails are modelled by non-polarisable beads that map two to four carbon atoms[120]. These models are intended to be compatible with both AL and CG models in the GROMOS forcefield and used in multiscale simulations of lipid bilayers in multicomponent systems.

Chapter 4: Novel GROMOS supra-atomic polarisable lipid models

As discussed in Chapter 1, it is possible to create coarse-grained (CG) supra-atomic or supra-molecular models by representing a group of several atoms or molecules as a single bead. AL simulations are limited in terms of the size of the system that can be simulated, and the length of time for which the simulation can be run, due to the large amount of calculations required. Timescale of biological process, such as protein folding and protein membranes vary greatly, often requiring microseconds to complete. Currently available computational resources therefore impose an efficiency/practicality barrier to completing simulations of large systems within a reasonable timeframe. While recent advances in parallel computing and especially the use of GPUs help to advance the technological barrier, the overall problem can be greatly simplified by decreasing the amount of performed calculations and, concomitantly, the resolution via the use of CG models. CG simulations lack a high proportion of non-essential degrees of freedom and so provide a substantial increase in the simulation speed and/or the size of the system that can be simulated[121].

The use of CG models needs to be tailored to specific applications due to a range of limitations. The loss of resolution needs to be carefully considered, as it will be impossible to track fine atomistic mechanisms and structural dynamics with CG models. Furthermore, as only clusters of atoms or molecules are represented a range of non-bonded interactions, such as hydrogen bonding, cannot be directly modelled. Another consequence of this is the loss of entropy in the simulation. This can lead to difficulties if CG and AL models are mixed together and also could lead to inaccuracies when dealing with intermolecular non-bonded interactions, so special multiscale implementations are required. CG methods find common application in solvent modelling, as solvent typically comprises over 90% of the simulated atoms/particles in a typical AL simulation and usually does not require the same level of detail as the solute. The loss of resolution in this component of the simulation is significantly offset by the large gains in computational speed. Nonetheless, CG methodologies can be applied to more complex molecules with careful consideration of application. The inherent loss in degrees of freedom and simulation resolution needs to be weighed up versus the need to increase the simulation size or timescale. There have been multiple applications of CG modelling approaches in literature from solvent to multi-protein assembly modelling. Lipids pose a great opportunity for CG modelling due to their presence in large numbers in membranes, highly repetitive nature and relative structural simplicity.

4.1 Existing CG modelling approaches

The contents of a typical simulation can be broadly split into two groups: solvent and solute. Many CG models for both solvent and solute exist. Different types of models exist with differing applications and functionalities. Models can be polarisable and non-polarisable, differ in degree of coarse-graining (number of atoms per CG bead) and the nature of non-bonded interactions, as well as bead shapes (e.g. spherical vs ellipses). Methods of parametrisation also differ – force matching, Boltzmann inversion and manual methodologies are all common. The most well-known solvent and solute CG models are described in the sections below.

4.1.1 Solvents

Solvent modelling is perhaps the most widespread area of application for CG methods. Since the initial attempts[122] in 1971, there have been many approaches to solvent CG, with water playing a focal point in the research due to its biological relevance and its consequent prevalence in biomolecular simulations. Ingolfsson H. I *et al*. 2014[123], Hadley K. R. and McCabe C 2012[124], Darré L., *et al*. 2012[125] and Izvekov S. and Voth G. A. 2005[126] have previously reviewed the most popular CG water models. Typically, these aim to accurately replicate experimental thermodynamic and dielectric screening properties. It should also be noted that it is possible to treat the bulk solvent implicitly with no physical solvent present[127-128], but for the purposes of our applications these models will not be considered, as explicit solvent molecules can play an important role in biomolecular processes.

A typical CG model "bead", representing the equivalent of several AL particles is initially set up as a single particle that has a certain volume and LJ or similar parameters that establish the size. Additional particles can then be added inside the existing "volume" of this bead to give rise to polarisability. The balance of charges between the central and any additional particles will affect the strength and directionality of the polarisation and the overall charge. The splitting of the masses between the beads will affect the inertia – typically the central bead is the heaviest to replicate the atoms moving as one group. Outlined below in *Figure 4.1* are the more successful particle based CG models for water.

MARTINI

One of the most well-known CG forcefields, the MARTINI (*Figure 4.1(a)*), includes a water model that was created by Marrink S. J., *et al*. in 2007 to facilitate MD simulations of CG lipids. The use of the model speed up the simulation by a factor of three compared to AL methods[129]. The model utilised a 4:1 CG bead mapping (four H_2O molecules are represented by one CG bead) and a LJ interaction potential for VdW non-bonded interactions. The downside of the model was that it did not account for electrostatic interactions. A number of variations exist that focus on further refining the interaction parameters, such as using different potentials (e.g. Morse) to describe the non-bonded interactions[130]. The MARTINI water model has also been updated to include polarisation by introducing two extra particles that were bonded to the central particle and have opposing charges, creating the MARTINIp model[131] (*Figure 4.1(b)*).

SSD

Soft sticky dipole (SSD) is a 1:1 CG water model (*Figure 4.1(c)*)[132-133]. The non-bonded interactions were again modelled by a LJ potential and it had a dipole term included to model electrostatic interactions, as well as a tetrahedral "sticky" term for hydrogen bond formation modelling. The inclusion of extra interaction terms combined with the 1:1 mapping lead to this model being computationally intensive compared to coarser alternatives and the simple point charge (SPC) water model[108] often used in AL simulations.

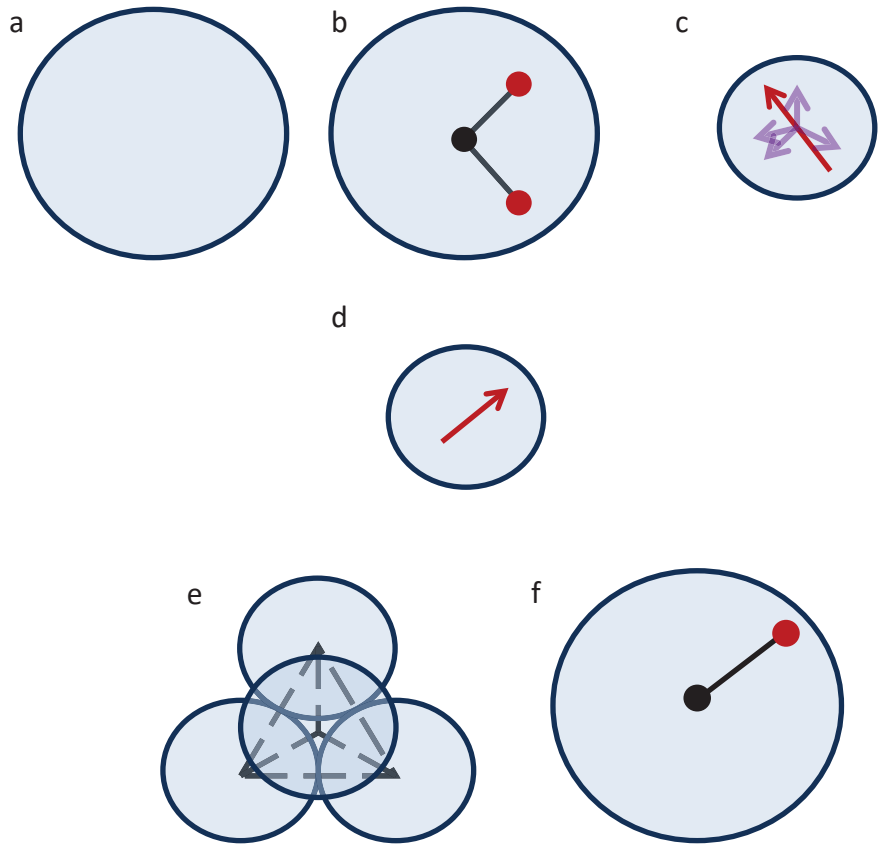


Figure 4.1 CG water models: (a) Regular MARTINI, (b) Polarisable MARTINIp, (c) Soft Sticky Dipole (SSD), (d) ELBA, (e) WatFour, (f) GROMOS. Adapted from H. I. Ingolfsson et al. (2014)[3]. Key: blue areas correspond to the LJ radius of the molecules, proportionally across the range of models; black circles represent the centre of the particle, red circles represent possible charged or dipole particles; arrows indicate directional electrostatic vectors.

SPC

While not strictly a CG water model, the SPC water model[69], frequently used in GROMOS AL simulations, is simplified in that the three water atoms are held in a rigid conformation by use of constraints on the bond lengths and bond angle. It is the water model against which the GROMOS forcefields are parameterised, and can also be used with other atomic-level forcefields as well as with the GROMOS CG alkane[120] and solvent[119, 134-135].

ELBA

Orsi M. and Essex J. W. 2011 have created the "ELBA" (electrostatics based) CG force-field for biomolecular simulations[136], with a primary focus on investigating phospholipid bilayers[137]. Based on the Stockmayer potential for polar fluids[138] and the SPC water model[69], each CG water bead, representing one H_2O molecule, was modelled as a soft LJ sphere with a point dipole (*Figure 4.1(e)*).

WatFour

The WatFour model maximises the coarse-graining factor, utilising 1:11 mapping, further reducing the number of required solvent computations. In a somewhat different approach to the typical single bead design, a cluster of 11 water molecules is modelled as four spheres locked into in a tetrahedral formation[139] (*Figure 4.1(e)*). The WatFour model is a great example of CG methodology as it shows that non-intuitive methods can replicate experimental parameters well.

GROMOS

Riniker S. *et al*. developed a polarisable 5:1 CG water model[119] that has since been shown to be compatible with the GROMOS AL force fields (*Figure 4.1(f)*). The CG bead contains two particles, one central and one additional "dipole particle" to add polarisability. The central and dipole particles have opposing charges and are linked by a bond described by a half-harmonic attractive potential, resulting in a changing dipole moment. Fluctuation in the length of this bond, and the rotational motion of the dipole particle around the central particle, result in fluctuation of the dipole moment of the CG bead, making the CG models polarisable.

While in the original model, only the central particle had LJ parameters, in the updated version, the mixing of CG water and AL solvent or other molecules was facilitated by introducing an extra repulsive term for the dipole particle. This avoided a "polarisation catastrophe", which could occur as a result of the AL water molecules being able to approach the CG beads too closely. The GROMOS CG water accurately reproduced the thermodynamic and electrostatic properties of water. The "movable dipole" approach has been adapted for other simple solvent molecules, such as dimethyl sulphoxide, chloroform and methanol by Allison J., *et al*. 2012[135].

Multiple models exist for simulating solvents. Generally, for greater computational gains the coarser the model is, the faster the simulations will run. Polarisability also offers attractive features over other models, in reproducing more realistic behaviour of solvent clusters. The GROMOS MD package will be used for the majority of the work in this chapter, so GROMOS compatible models are of most interest for the current work. The GROMOS polarisable CG water offers a good balance between the degree of coarse graining with 5:1 mapping, residing towards the coarser end of the CG solvent spectrum. Furthermore, GROMOS forcefields are cross compatible with a range of other MD software packages (e.g GROMACS) and therefore offer potential for a high degree of transferability between different programs.

4.1.2 CG lipids

Cell membranes are comprised primarily from a lipid bilayer, which an essential part in many biological processes[41]. Phospholipids are the main component of lipid bilayers which make up most living cell membranes[140]. Cell membranes are important to many cellular processes, such as transport of chemicals across membranes and many proteins also interact with the membrane surface[141-143]. Proteins

frequently also interact with the membranes, embedding into the bilayers or coming into peripheral contact. Many of these proteins help facilitate the transport of many key chemicals from ions and small molecules such as glucose to peptides and other proteins. The primary focus of the most popular AL biomolecular forcefields has traditionally been proteins, but recently the importance of lipids in such in membrane interactions has been highlighted[144]. Lipids play an important part in recognition and partitioning of proteins to the particular parts[145] of the membrane and in defining the structure and dynamics of the membrane itself. Antimicrobial peptides will preferentially target certain, specific bacterial cell types over other via selective lipid recognition[146-147]. Processes of endocytosis, signalling and vesicle budding can also depend on selective membrane interaction and recognition[148-149].

It is important to get the bulk properties of the lipid bilayer correct, especially as membranes can be complex structures with multiple types of lipids and other chemicals present, such as cholesterol. Typically, a cell membrane will consist of a lipid bilayer where the lipids are orientated with the polar groups facing outwards and the tails facing inwards (*Figure 4.2*). The nature of the lipids and other chemicals present will affect the structural properties of the bilayer. Lipids have simpler structure than many proteins (most phospholipids typically consist of a polar headgroup and a pair of hydrophobic tails) There has been significant work to create specialised or adapted versions of these existing forcefields that include parameters for different types of lipids. These are typically tested by simulating pure or occasionally mixed lipid bilayers, with the goal of reproducing experimental data for electron density profiles, bilayer thickness, deuterium order parameters and temperature dependent transition dynamics [130, 150-151].

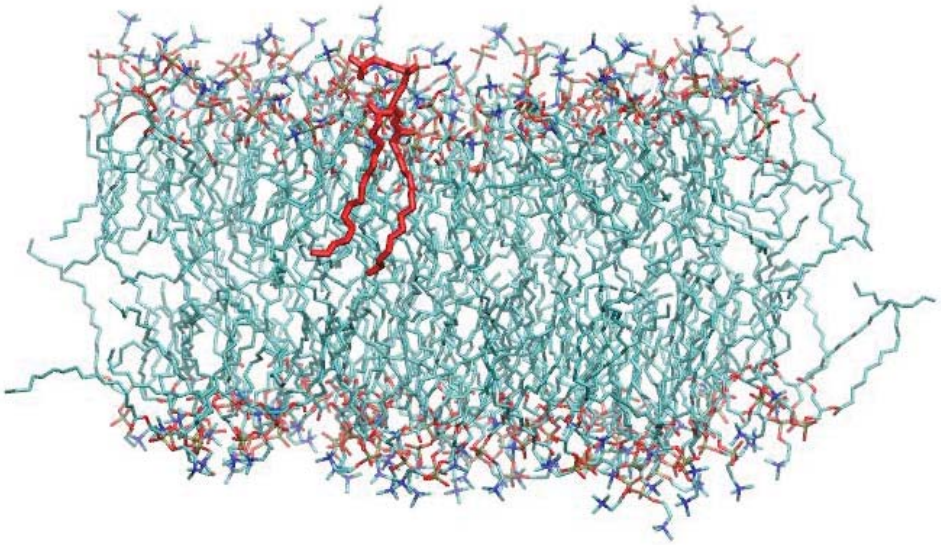


Figure 4.2: Simulation snapshot of a periodic DPPC bilayer simulation comprised of 128 lipids. A single DPPC lipid has been highlighted in red for reference.

CG approaches to modelling lipids present an attractive opportunity to save computational resources because naturally occurring lipid systems, such as bilayers, contain a large number of lipid molecules, are highly repetitive, and seldom form the key mechanistic focus. Unlike many water models, in which the CG model does not necessarily bear physical resemblance to the underlying collection of water molecules, the process for generating CG lipid models typically involves mapping CG beads onto the atomic level structure of the lipid molecule according to functional groups or using force matching techniques to obtain parameters directly from AL simulations. Each of the functional group CG beads can then be parameterised according to the properties of that functional group in the original AL model [152]. CG lipids have been recently used to study bilayer permeability by small molecules, phase behaviour in multi-component systems and self-assembly[153-155]. A selection of common CG lipid models is outlined below.

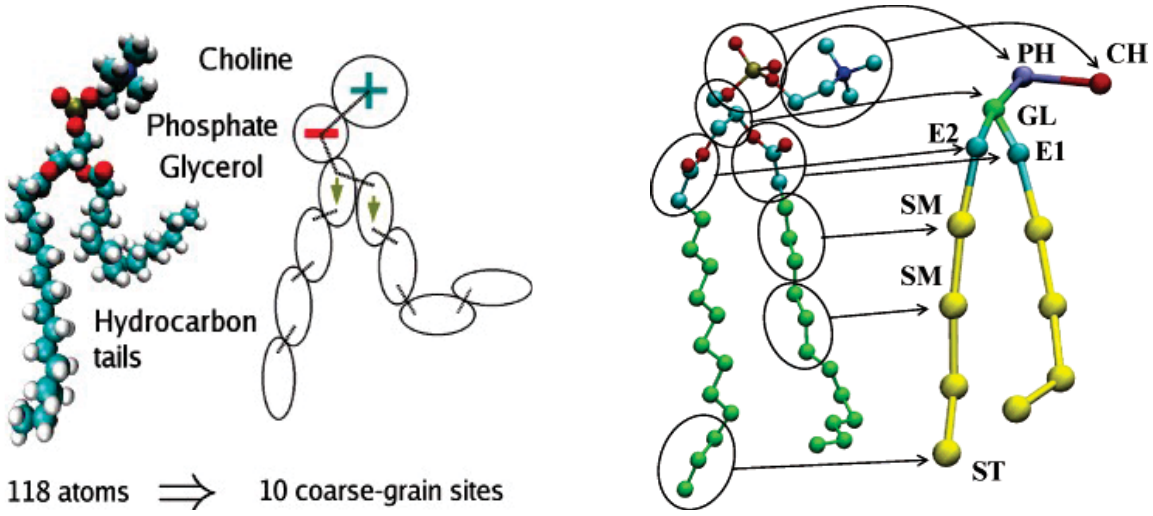


Figure 4.3: Two examples of lipid coarse graining strategies, demonstrating AL lipid model and their corresponding CG counterparts. Shown are mapping of CG groups, as well as linking between CG beads and some charge information. On the left - Orsi, M. et al., 2008[152]. On the right – Izvekov, S. and Voth, G. A., 2005[129]

MARTINI models

MARTINI CG models are very well known and widely used [156]. MARTINI lipid CG model was originally published in 2004[157] and then further improved and extended in 2007[129] for other molecules, including polymers[158], ring systems, sterols and then further to proteins in 2008[159]. They showed good agreement to existing models and offering a five to ten fold increase in simulation speed compared to AL methods[129]. They model was designed for a broad range of applications, including investigations of self-assembly[160] , vesicle fusion[161], lamellar phase transformation and simulations of membranes, including protein interactions and permeability studies[162-166].

MARTINI models on average utilise 4:1 mapped CG particles with single interactions sites and are split into four categories – polar, non-polar, apolar and charged types. 18 further subtypes are also included to further refine polarity and hydrogen bonding characteristics. Non bonded interactions are modelled by a 12-6 LJ potential and bonded interactions are modelled using the typical standard forcefield function, similar to the recent GROMOS forecfield[167] and are used in conjunction with the GROMACS MD engine. CG MARTINI models are typically parametrised to match atomistic simulations.

One of the limitations of the MARTINI is the static nature of the mapping system. Lipid bilayer properties such as transition temperatures, packing parameters and stability depend on the length of the acyl chains in lipid tails. Due to the locked nature of 4:1 mapping, MARTINI model can only approximate the length of the actual carbon tails, which are not always divisible by groups of 4. Furthermore, as it is based on atomistic models, a precedent can be had for multiplying the inaccuracies inherently present in AL models, if they are used as a basis for parametrisation.

Klein models

Klein *et al*. pioneered one of the first force fields for CG lipids based on matching structural and thermodynamic experimental and atomistic simulation data of 1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid[154, 168, 46]. The model has since been refined and seen applications in studying lipoprotein interactions[166] and membrane behaviour. In this CG model, a lipid is comprised of 13 CG beads, constituting roughly 3.5 atoms per bead on average. Bonded terms were modelled by standard harmonic and quartic potentials and 6-12 LJ potential was used to model non-bonded interactions. A "soft" LJ 4-12 potential was used to model dispersion forces in water. This force field has been used extensively in lipid monolayer and bilayer studies and more specialist studies, such as membrane partitioning for fullerenes[155, 169]. However, the model is not polarisable.

Voth models

The Voth group focused on the use of force matching methods to obtain CG model parameters directly from AL simulations. They have developed a large number of models for different lipid types and other systems. Their lipid models typically comprise 15 CG beads. Other lipids and mapping methods were also attempted by the group, for example in the logical extreme of lipid CG, an ultra-CG lipid model, comprised of only one bead was also developed[126].

Initially, a CG forcefield was developed in 2005 for the simulation of a DMPC bilayer[8, 126]. The forcefield was parametrised by force matching AL reference simulations. Bonded terms were described with harmonic potential for CG bonds. Non-bonded interactions were replicated by force matching AL simulations globally for the CG models. However, this approach resulted in combined non-bonded terms that gave the total combined potential and therefore did not consider LJ and Coulombic interactions individually. This meant that the method had transferability drawbacks as each new molecule must be force matched individually. The methodology was then further expanded to other types of lipids[170-171] and the models were then consequently coupled with AL and developed for use in multiscale simulations (S*ection 4.1.3*)[172].

ELBA

As well as the water model mentioned previously, Orsi, M. *et al*. have also developed lipid models, in keeping with their aim of studying bilayer behaviour, including bilayer permeation studies with small molecules[137, 152-153]. ELBA included parameters for a range of CG Phosphatidylcholine-based lipids, and simulations of bilayers accurately replicated both experimental data and the results of simulations using AL models. In the initial version of ELBA, both elliptical (via Gay-Berne potential[173]) and spherical geometry beads were used for the lipids. More recent versions used spherical beads, due to the large number of parameters and computational costs of defining and calculating the ellipse orientations. Non-bonded interactions were modelled by standard 6-12 LJ potential. No angle or torsional dihedral bonded terms were used in the initial lipid model. ELBA models are currently limited to the LAAMPS software due to the use of point dipoles and the need to calculate torques. While this is advantageous, as most other MD software does not allow for such calculations, it limits the user to only one specific software.

4.1.3 Multiscale models

For larger systems that contain complex environments, requiring detailed AL investigation (e.g. a protein embedded in a lipid bilayer) – a combinational approach of AL and CG modelling could be applied. While it is possible to simulate large biomolecular systems at AL atomistic resolutions, the timescales needed to study structural changes exceed what is reasonably feasible using modern computer resources. Multiscale simulations combine combinations of scales (e.g. quantum mechanics and MD or AL and CG) to allow a detailed representation of the areas of interest, while saving computational resources where possible by coarse-graining the remainder of the system. Such approaches to molecular modelling present an excellent opportunity to gain new insights into physiological processes by taking advantages of the computational speed up offered by CG methods, while maintaining high resolution for the molecules of interest[126, 174]. Multiscaling provides an excellent opportunity to give structural insights into heterogeneous bilayers, multi-component systems, protein-membrane interactions and protein assemblies[175-177].

Many multiscale simulation methodologies exist, including back mapping[178-179], adaptive resolution[180], positional restraining[180] and free mixing[181-182]. Back mapping methods involve creating a set of CG models from an AL system, running a relatively faster CG simulation and then mapping the AL coordinates back onto the gathered data, restoring the original atomistic resolution. Positional restraining and free mixing approaches comprise running a simulation containing both CG and AL particles in the same simulation box. Of the fixed-resolution methods, positional restraining restrains the differing types of beads to their respective allocated areas according to their resolution, while free mixing lifts that restriction and allows particles to interact freely. Free mixing requires the implementation of further parameters to accommodate AL-CG interactions, but results in a more adaptable, flexible system. In such simulations, CG methods will usually be used to model the solvents and AL for more complex molecules, such as proteins.

Orsi M. *et al*. (2009) have performed multiscale simulations to investigate permeation of lipid bilayers by small particles[153]. In their system the bulk water solvent and lipids constituting the bilayer were coarse-grained and the small AL permeating molecules. The CG models were compatible with AL ones and therefore a "free mixing" approach could be utilised. The results agreed with those of previous AL simulations, as well as with experimental data, but the multiscale method was faster by a factor of two. Since then continuous refinement of ELBA water and lipids has been carried out in multiscale environments, including direct mixing experiments and testing of GPU parrellisation[183-184].

Ayton G.S and Voth G.A (2007) developed mesoscopic level membrane that can couple to atomistic models for use in membrane simulations and investigation of AL proteins[185-186]. Focus was made on making AL and CG models directly compatible with each other. A range of Voth models described above has been used in direct combination multiscale experiments[187], involving lipid bilayers and proteins[171, 188]. The models are typically fitted for specific applications and therefore vary highly in the degree of coarse-graining as well as application.

Another example of multiscale simulation that illustrates both the position restraining and free mixing techniques is the work of Rinker S. et al. (2012)[189], who continuously developed multiscale simulations of a selection of structurally diverse biomolecules[190-192]. In one case, a protein was modelled using the (AL) 54A7 GROMOS forcefield, and surrounded by a layer of AL SPC water molecules, with the GROMOS CG water model used for the bulk solvent. The AL water molecules were only loosely restrained to remain close to the protein centre of mass, so that at the AL-CG boundary, the two types of water model freely mixed. Analysis of the structural properties of the proteins and comparison of back-calculated NMR data with the experimental measurements showed that the protein structure was not distorted by the multiscale solvent representation. Simulating a AL protein in only CG water also resulted in stable protein structures, but the inclusion of AL water adjacent to the protein allowed correct modelling of the water-protein, and therefore also the intra-protein, hydrogen bonding (CG water cannot be used to model hydrogen bonds). In the position-restraining simulations, the CG water comprised the bulk of the solvent not adjacent to the protein, leading to significant computational speed gains, without the reduced accuracy commonly associated with CG simulations. Unlike other positional restraint multiscale methods, this technique does not completely fix the different levels of model to different regions of the simulation. The model presented in this chapter aims to expand on Riniker and van Gunsteren's methodology and attempts to create a lipid model based on the GROMOS polarisable CG water for multiscale simulation uses.

The aim of this project was to create a new series of lipid CG models, as the current methods all have a range of drawbacks. The new lipid models would be polarisable and compatible with AL models for direct mixing multiscale simulations. The non-bonded interactions would have separate functional potentials for Coulombic and LJ interactions. The bonded terms would be described using standard quartic potentials. Terms for angle and torsional dihedral terms would also be added.

4.2 Design of CG DPPC lipid model

The initial aim of the project was to establish a good CG model for a typical phosphocholine (PC) lipid. DPPC made a good candidate for the development of the initial model, as it is well studied both experimentally and computationally, resulting in many thermodynamic and structural properties (e.g. energy of solvation, area per lipid, bilayer thickness, etc.) being readily available for comparison. Furthermore, DPPC has two saturated lipid tails, so that the presence of carbon-carbon double bonds does not need to be considered in the first instance.

The CG polarisable DPPC model would be designed according to a modular approach, splitting the AL lipid into several CG beads of varying size by functionality (*Figure 2.4*). Such an approach is in line with an overall goal of creating a set of "plug and play" CG parameters for a variety of different interchangeable lipid head groups and tails. All models would be parameterised by matching structural and electrostatic parameters to reference AL simulations of the DPPC bilayer manually. Existing atomic simulations of DPPC bilayer that match experimental data exist, therefore directly matching CG and AL simulations should also create good agreement with experimental data. As it is theoretically challenging to get all model parameters (CG or AL) to match the full extent of *in vivo* experimentally monitored properties, it is necessary to prioritise certain model features for primary matching. In this case the features of importance and interest were the physical dimensions and bilayer behaviours of the lipids and the goal of the new model would be to replicate a full DPPC bilayer, both in terms of its dimensions and interactions with other molecules. The eventual project goal would be to create a range of parameters for interchangeable lipid functional groups so that whole phospholipid models could be assembled from a combination of existing functional group building blocks and simulated without need for further individual parameterisation.

4.2.1 Mapping

DPPC can be broadly split into two parts – the polar functionalised "headgroup" and the non-polar carbon tails. The headgroup part of the lipid contains the following functional groups: choline (C), phosphate (P), glycerol (G) backbone and two ester (E1 and E2) groups. Each functional group is modelled by a separate CG bead, so that each functional group can serve as a potentially interchangeable building block (*Figure 4.4*). Each CG bead present in the headgroup is comprised of two particles: a central particle, which has a full LJ potential, a partial charge, and a mass, and a dipole particle, which has an opposing partial charge and a repulsive-only LJ potential. Repulsive LJ interactions were added to all the dipole particles to prevent a charge catastrophe in the presence of AL water and other molecules, following the recent reparametrisation of the CG water model[189].

The tails would consist of five non-polar carbon beads each. The groups of Emeritus Prof. Wilfred van Gunsteren and Assist. Prof. Sereina Riniker at ETH Zurich, who developed the original GROMOS CG water model, have concurrently also developed a CG alkane model, which should be directly applicable to modelling the lipid tails. Alkanes of any length can be constructed from a set of six different types of

beads, comprising 2, 3 or 4 carbon atoms and their associated hydrogen atoms, and representing either central ($(CH_2)_n$) or end ($(CH_2)_n$-CH_3) portions of the hydrocarbon chain. Unlike the GROMOS CG water model, the carbon beads are not polarisable and are uncharged.

The AL-CG mapping of the headgroup was further rationalised by analysing the distributions of distances, angles and torsional dihedrals between the centres of mass of grouped atoms calculated from AL simulations. For easy parameterisation of an accurate CG model that is representative of the behaviour of the underlying AL model, the distance and angle distributions would ideally be well defined, with a single peak, as at a CG level, these properties are controlled using a simple potential energy functions.

Preliminary simulation of an AL DPPC bilayer for 40 ns was carried out to provide AL reference data. The results were used to determine how the AL molecule should be split best into CG beads. Three approaches for splitting were proposed – layouts 1, 2 and 3 (*Table 4.1* and *Figure 4.4*). In layout 1 the beads were split according to chemically intuitively rational midpoints between the various functional groups, with the exception of Choline and phosphate split, where they are split to make beads of approximately equal size. The aim was to keep the beads as uniform as possible in size. In layout 2 the concept of a larger glycerol bead, containing the joining oxygen atoms was trialled. Layout three represented a split by functional groups fully, so that in theory any group could be substituted out for a similar one without too much issue (e.g. choline for serine). Layout 3 differs to layout one by including carbon 6 with the choline group and leaving phosphate bead carbon free.

Table 4.1: Different ways of splitting the AL DPPC into CG beads. Numbers refer to the atomic numbering system present in Figure 2.4. Atoms 18-31 and 37-50 would be modelled using the CG alkane parameters and so were not considered here.

CG bead	*Layout 1*	*Layout 2*	*Layout 3*
Choline	1-5	1-6	1-6
Phosphate	6-11	7-11	7-11
Glycerol	12, 13, 32	12, 13, 14 32, 33	12, 13, 32
Ester 1	33-36	34-36	33-36
Ester 2	14-17	15-17	14-17

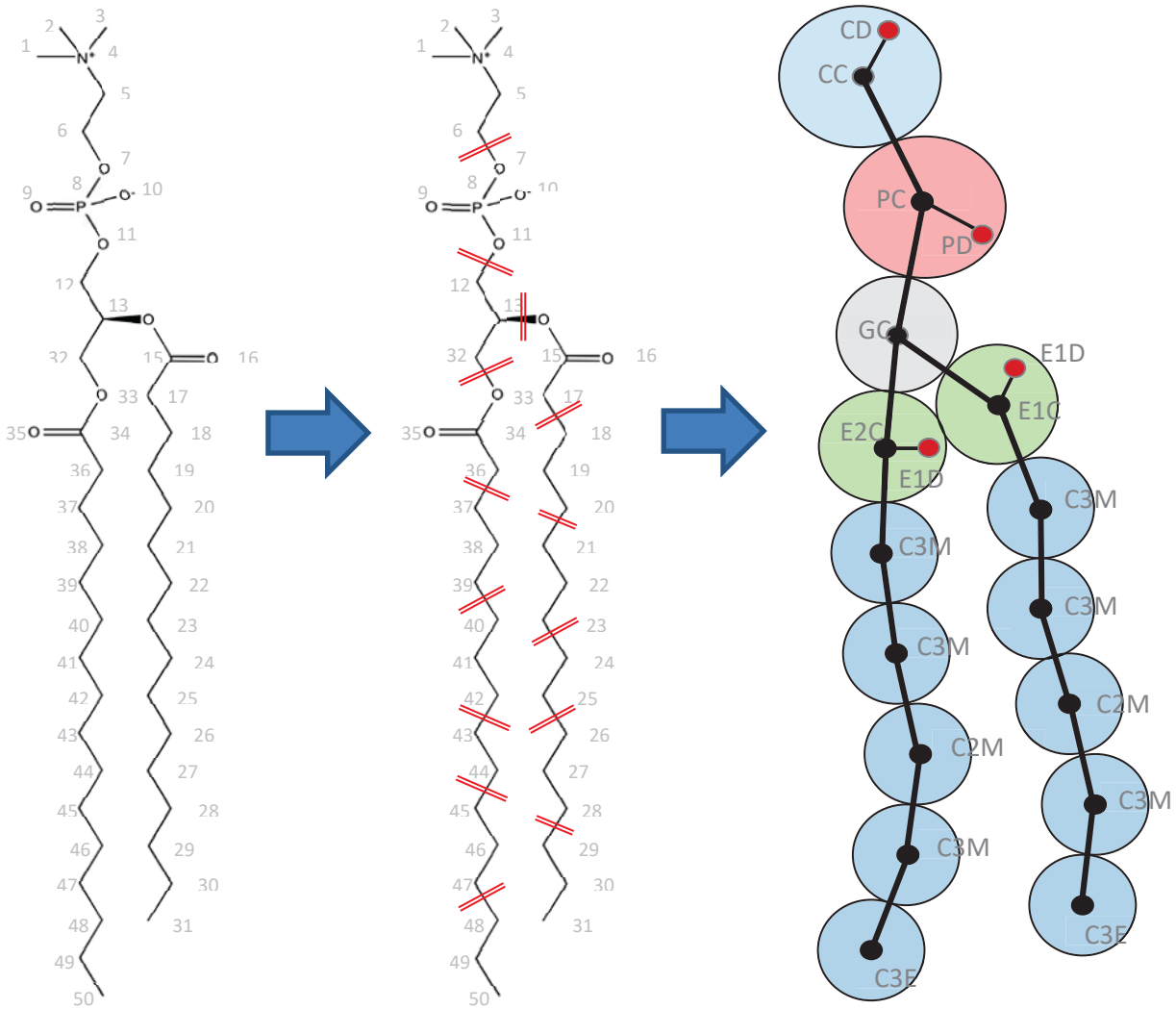


Figure 4.4: Mapping from AL DPPC to CG DPPC, all AL and CG atoms and particles are labelled in grey (Layout 3). Red double lines represent boundary points between groups of atoms that are to be assigned to different CG beads.

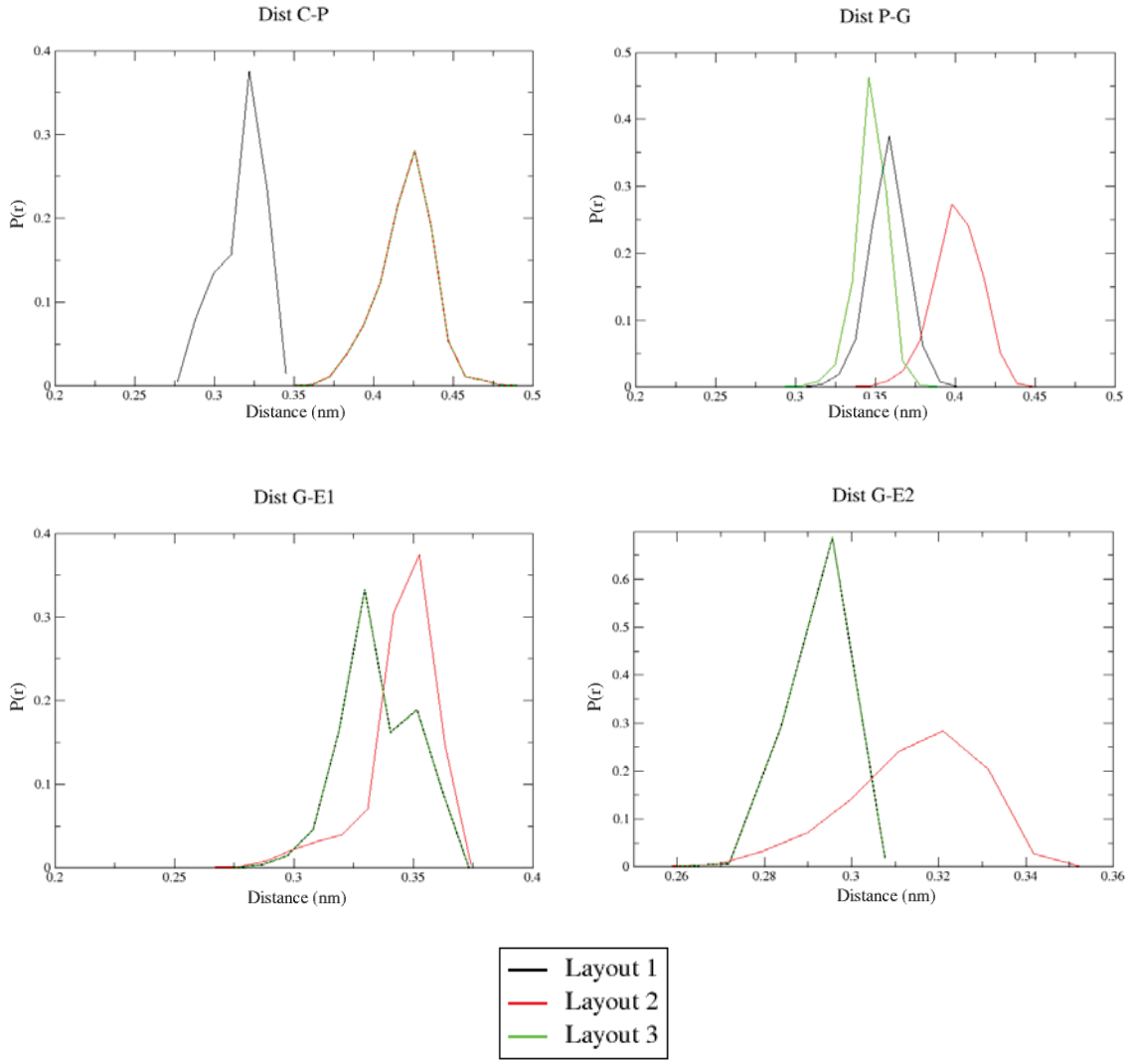


Figure 4.5: Distributions of the distance between the centres of mass of pairs of groups of atoms, corresponding to CG beads during a 40 ns AL simulation of a DPPC bilayer. The three layouts are defined in Table 2.1.

All three layouts produced mono-modal distance distributions, and layouts 2 and 3 also produced essentially mono-modal angle distributions (*Figures 4.5 and 4.6*). Layout 3 was chosen as it offered the best solution for splitting the DPPC molecule due to the flexibility it would offer when applying the methodology to other lipids. Layout 3 provided a cleaner separation between the phosphate and choline beads by chemical intuition, as well as a clean separation of the ester groups from the glycerol backbone. With this approach CG beads could be cleanly swapped out when going between models. For example, the phosphate group only has phosphate atoms associated with the CG bead (i.e. no attached carbons, as in the case of layout 1), meaning that a variety of CG phospholipids could make use of the same phosphate CG bead parameters in the future. Furthermore, no extra incentives were found for the

case of the large Gly bead in layout 2, so it was decided to keep oxygens attached to the polar ester groups.

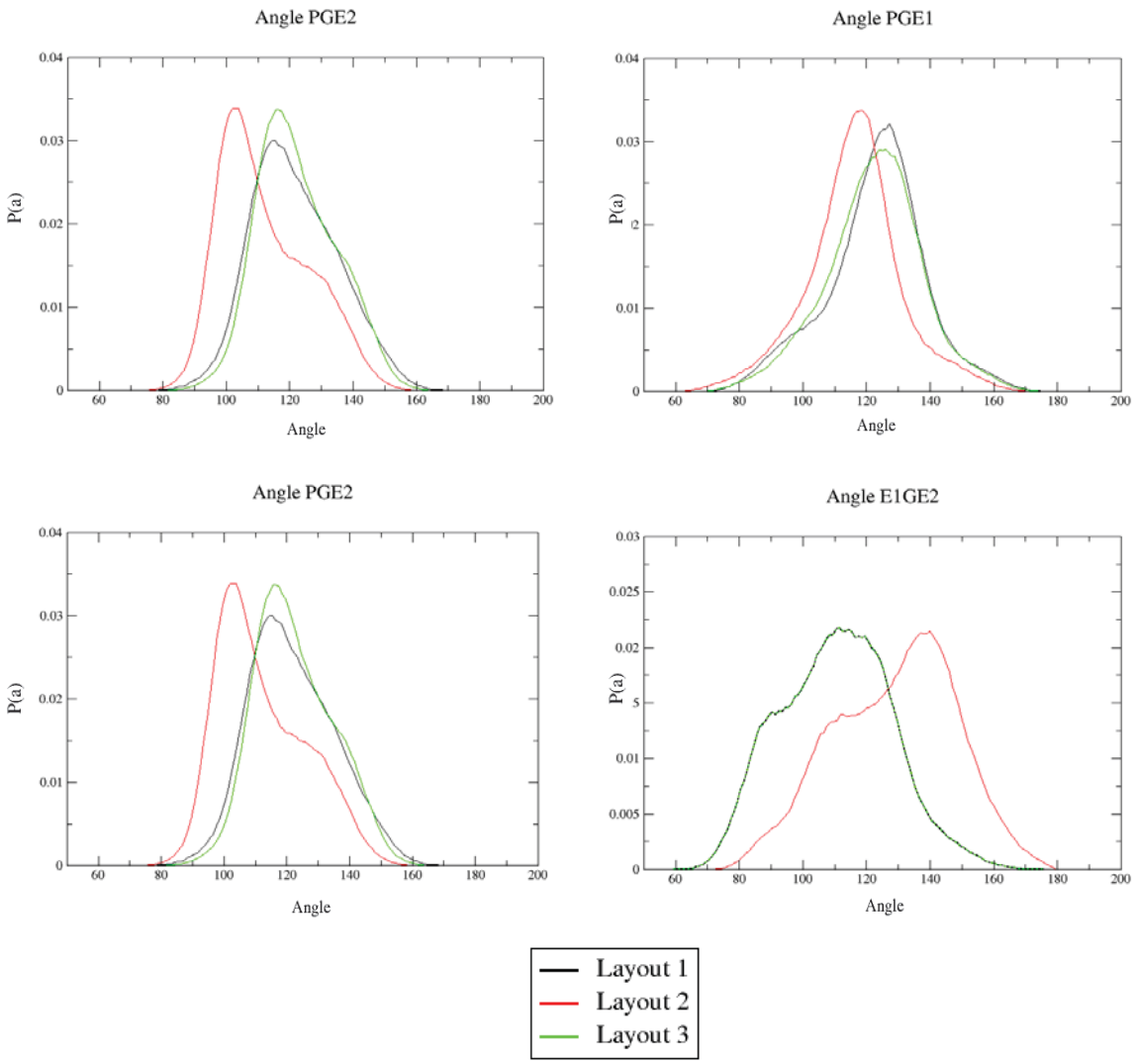


Figure 4.6: Distributions of the angle between the centres of mass of three groups of atoms, corresponding to CG beads during a 40 ns AL simulation of a DPPC bilayer. The three layouts are defined in Table 2.1.

4.2.2 New CG force field parameters

A range of new forcefield parameters needed to be introduced to the existing GROMOS 54A8 to accurately portray and describe the relationship between the new CG beads. New parameters can be split into bonded and non-bonded categories. Non-bonded parameters include CG bead masses, charges on both central and dipole particles, and LJ C6 and C12 parameters for the LJ interactions between beads (see *Figure 1.2 and Equation 1.1.2*). Bonded parameters (*Figure 4.7*) include the force constants and equilibrium values for all bonds, angles and torsional dihedral terms between CG particles, and the centre-dipole particle bonds. A dipole was established between the centre and dipole beads of the particles by ensuring that they were of opposing charges.

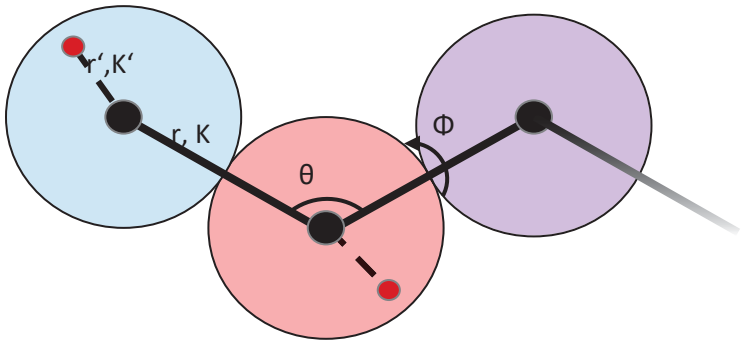


Figure 4.7: A schematic of the required bonded interaction parameters for multiple joined CG beads. Black circles represent the centre of the particles and the corresponding coloured areas are the effective LJ radius; red circles are the dipole particles; black lines correspond to the inter-particle bonds, dotted lines are the particle-dipole bonds. The corresponding bond and angle terms are described below.

Bonds between the central CG particles were described using the standard GROMOS quartic bond stretching potential (*Equation 4.1*). The central and dipole particles were linked by a special case half-harmonic bond implemented separately in the code and described in more detail below (*Equation 2.2*). There were no angle terms attached to the dipole particles, so that they would have free unrestricted movement within the sphere. While all AL solvent bonds were constrained during AL simulations using the SHAKE[15] algorithm, to allow for an integration time step of 2 fs by eliminating fast vibrational dynamics. The bonds between CG central particles were not constrained.

(4.1) $$V^{\mathrm{r}} = \frac{1}{4}k^{\mathrm{r}}(r^2 - (r^0)^2)^2$$

Where:

V^r – quartic bond potential energy function

k^r – quartic bond stretching constants
r – bond length
r^0 – ideal bond length

Equation 4.1: Bonded quartic potential energy function.

(4.2) if $r > r^0$ $V_{hh}^r = \frac{1}{2}k^{\mathrm{r}}(r - r^0)^2$;
else $V_{hh}^r = 0$

Where:
V_{hh}^r – half harmonic bond potential energy function
k^r – quartic bond stretching constants
r – bond length
r^0 – ideal bond length

Equation 4.2: Half harmonic bonded potential energy function, used to model centre-dipole particle bonds.

The angles between triplets and dihedrals between quartets of central CG particles were also treated using the standard GROMOS interaction functions (*Equations 2.3 and 2.4*).

Angles (θ) – angles between three connected beads are described by the harmonic angle bending function below:

(4.3) $$V^{\theta} = \frac{1}{2}k^{\theta}(cos\theta - \cos\theta^0)^2$$

Where:
V^{θ} – angle bending potential energy function
k^{θ} – harmonic angle bending force constant
θ – angle between three connected objects
θ^0 – ideal angle between three connected objects

Equation 4.3: Angle harmonic potential energy function.

Torsional dihedral angles (φ) are described by the periodic torsional dihedral angle rotation function below:

(Eq. 4.4) $$V^{\phi} = k^{\phi}(1 + cos(\mathrm{m}^{\phi}\phi - \phi')) \; with \; \phi \in [0, 2\pi]$$

Where:
V^{Φ} – torsional dihedral rotation potential energy function
k^{Φ} – dihedral rotation force constant
m^{Φ} - multiplicity
Φ – dihedral angle between four connected objects
Φ' – dihedral phase shift

Equations 4.4: Torsional dihedral rotation potential energy function.

Running a CG simulation further necessitated changes both to the GROMOS MD code and inputs. Electrostatic parameters (e.g. cut-off distances and dielectric permittivity) were adjusted to account for the multi-atom encompassing nature of CG beads, as well their extended size. Furthermore, changes to CG bead pressure scaling adjustment had to be included. Initially GROMOS code treated all CG beads as a single special class of molecules, irrespective whether the CG beads were supra-atomic or supra-molecular. New pressure adjustment inputs meant that correct pressure scaling was used for CG beads that were not supra-molecular and that simulations could be run without simulation-breaking errors. The corresponding input parameter code and structure had to be altered by the programming team at ETH Zurich in order to fix the issue after extensive correspondence. See *Section 4.5: Methods* for differences between CG and AL simulation parameters.

Ideal values for the new forcefield parameters were estimated based on the simulations of the AL DPPC bilayer. Key structural properties such as the distance, angles and dihedral angles between the centre-of-mass of each group of AL atoms were used to provide an initial estimate of the equilibrium values of the CG bond, angle and dihedral terms. Initial force constant values were based on standard GROMOS forcefield AL values, as well as the CG alkane parameters[109, 120]. CG bead charges were the sum of the partial charges of their constituent atoms. LJ parameters for new CG particles would be estimated on structural parameter distributions between clusters of atoms. More details on the derivation of these parameters for specific stages of the CG DPPC model are given in subsequent sections.

4.2.3 Reference simulations

Reliable AL forcefield parameters for DPPC were available and have been shown to reproduce the experimentally determined properties of DPPC bilayers[109, 150]. AL simulations were used as a basis for the initial CG model and as a point of reference for evaluating the CG parameter sets (*Table 4.2*). The advantage of using AL simulations as a reference was that a greater quantity of quantifiable, calculated properties were accessible, such as the structural distributions of atoms in each molecule, dipole moments of chemical groups, etc. Simulations of both a DPPC bilayer and a single DPPC molecule in water were performed.

It was also important to ascertain whether isolated parts of the lipid molecule functional groups could be successfully simulated as individual components. A Choline (Cho) and Phosphate (Pho) only headgroup (PCm) was created by truncating the DPPC molecule at atom 12 (*Figure 4.4*) and capping the terminal oxygen atom in the phosphate group with a chemically inert methyl group, to mimic the head group being connected to a larger lipid structure. Technical details of the simulations can be found in *Section 4.5: Methods*.

Table 4.2: List of AL reference simulations with solute, solvent and total run-time details

Solute	*Solvent*	*Run-time (ns)*
DPPC bilayer*	SPC	100
DPPC bilayer initial*	SPC	40
DPPC single	SPC	50
PCm** single	SPC	10

** The DPPC bilayer simulations are a continuation of a pre-equilibrated system, previously simulated for 200 ns[6]. **PCm is the truncated, methyl capped phosphocholine headgroup.*

These simulations allowed comparison of the conformations adopted by DPPC in a bilayer, DPPC by itself in aqueous solution, and the capped phosphocholine by itself in aqueous solution, and surrounded by other head group molecules.

4.3 Model Construction

Several CG models, each increasing in complexity, were built during the course of the work covered in this thesis. Due to the number of variables and complexity of interactions, simpler "proof of principle" models were designed and tested first before moving onto simulations of the full DPPC CG molecule and the DPPC CG bilayer as a whole. This section follows the development of these models sequentially.

Firstly, a simple dimer of two CG type beads was established (CG-PC model). Then the functionalised headgroup of the lipid (CG-HG model), including choline, phosphate, glycerol and ester beads was parameterised separately, before the addition of the carbon tails, which were developed separately. Then, with the parameterisation of the tails by our colleagues at ETH Zurich finally reaching completion, a single full lipid molecule (CG-DPPC model) was parameterised. After a working, parameterised version of CG-DPPC was established, attempts to simulate a CG bilayer system comprised of multiple CG-DPPC molecules was attempted. Each of these three models is discussed in a separate section below, with all associated parameterisation and code development.

4.3.1 Two-bead test system: CG-PC

Initially a simplistic two-bead CG model was created to test feasibility of extending the CGW style methodology for biomolecules. The two-bead molecule aimed to create the phosphate (P) and choline (C) units linked together, and to test whether it was feasible to have GROMOS solute CG beads linked together in a simulation. Both novel bonded and non-bonded parameters for choline and phosphate beads were created based on the methodology described above. The primary aim for this particular model was to prove the principle that two CG beads containing both centre and dipole particles could be bonded together. As soon as a stable CG-PC simulation could be carried out, the next set of subsequent models, focusing on the lipid headgroup could be parameterised.

4.3.1 Design of CG-PC model and initial parameter estimates

The test model comprised two beads linked by a bond (*Figure 4.8*). Each bead contained a centre (suffix -C) and a dipole (suffix -D) particle inside it. Centre and dipole particles were linked by the centre-dipole special case half harmonic bond. In this simplified version of the model, no angle or torsional dihedral terms were required. Each bead had an associated mass equivalent to the cumulative masses of the atoms comprising the choline and phosphate groups, respectively. The mass was split in a 2:1 ratio between the centre and the dipole particles, following the GROMOS CGW example. Such a split was meant to ensure that most of the mass of the CG bead was concentrated in the centre of the particle so that the dipole particle did not gain too much inertia. This approach worked well in the previous CGW model and was left as such due to the large number of other parameters in need of refinement. The distribution of weight between the centre and dipole beads could be addressed again at a later point after the main CG-DPPC bilayer model has become functional. To imitate the AL simulation models, the choline bead needed to have an overall charge of +1 and phosphate of -1. Initial estimates of the LJ σ parameters, which determined the effective bead sizes, were estimated from the radial distribution function (RDF) of the centres of mass calculated for atoms mapped to each CG particle, calculated from the AL DPPC bilayer simulation (*Figure 4.10*).

The exact nature of the bonding, charge and LJ parameters was determined in the iterative parameterisation step outlined below. Initial estimates for parameters were taken from AL simulations of a DPPC bilayer (*Section 4.2.3: Parameterisation*). An initial estimate of the length of the bond between the two CG beads comprising the head group was obtained from the distance between virtual atoms representing the centres of mass of the two functional groups, Cho and Pho. The equilibrium length and force constant of the CP-DP bond were set to be the same as in the GROMOS CG water model (0.2 nm and 2×10^6 kJ.mol^{-1}) in the first instance. Parameters for the LJ function applied to the central particle were estimated from the reference distribution of distances between centre of masses of atoms comprising choline and phosphate functional groups. The dipole particles initially had the LJ parameters set to zero, so that the particles would be volumeless.

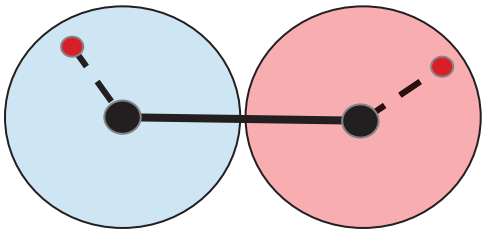


Figure 4.8: Two bead CG-PC test system design. The choline bead is coloured blue and the phosphate red. Centre particles are represented by black circles and the edge of the coloured circle represent the effective LJ radius. Dipole particles are represented by red circles. The bond between centre particles is in solid black and the centre-dipole bonds are dashed black lines.

The CG-PC model was then simulated in a box of CGW. It would have been advantageous to also simulate the model in SPC water, however GROMOS tools for coordinate generation were very limited when this work was carried out. SPC water model is the standard AL water model for the GROMOS package – it contains all three atoms of the solvent molecule in a locked orientation. For the CG-PC in CGW simulation the starting coordinates were generated by taking a box of 2560 CGW molecules and manually editing two nearby water molecules to create the two bead CG-PC coordinates. All initial attempts to generate two bead CG-PC coordinates in AL SPC water using the available GROMOS tools (sim_box, ran_box and build_box) have failed due to constraint algorithm errors in the first few time steps, implying physically unrealistic initial sets of coordinates. Further details on simulations and coordinate generation can be located in *Section 4.5: Methods*.

4.3.2 Simulation failure and software modifications

First attempts to run MD simulations of the CG-PC model using the parameter estimates generated as described above revealed multiple unforeseen problems with the simulation software and resulted in a series of simulation failures. Severe problems occurred due to the implementation of CG features in the GROMOS MD++ software package being specific to solvents or pure liquids (the CG alkanes were parameterised in pure liquid form), many of which were difficult to diagnose.

Initially it was found that CG simulations of CG-PC terminated after a small number of integration steps due to failure in the SHAKE algorithm (3000-55000 steps, machine dependent). While this could have resulted from almost any problem with initial coordinates or model parameters, it was eventually ascertained that there was an issue with the treatment of the central-dipole particle bond. The central-dipole bond potential was meant to be calculated using a half-harmonic bonding potential function (*Equation 4.2*).

It was found that the bonds between the central particles were improperly treated at the initial simulations with a half-harmonic potential, due to non-differentiation of bonds between all CG particles. After diagnosis and extended correspondence with collaborators at ETH Zurich, CG particle bonds were now split into two categories in the topology into normal and “dipole” bonds. Normal category included centre-centre particle bonds and is described by the normal GROMOS quartic harmonic potential. Dipole bonds are between centre and dipole particles, use the half harmonic bonded potential instead of the normal quartic and are exempt from bond constraints. SHAKE constraints on the simulation were still necessitated in the presence of SPC water, as the solvent bonds had to be constrained.

Further difficulties arose due to the implementation of facets particular to CG solvent models that were not appropriate for CG lipid models, such as the assignation of different pressure scaling factors to account for the contribution of supra-molecular CG beads to the system, and the adjusted CG intra-bead electrostatic screening. It was found that pressure correction was not necessary for lipid models as they were not supra-molecular and therefore could be subjected to pressure scaling like normal AL particles.

An additional problem that was also observed for the GROMOS CG water model was that when one or more dipole particles experienced a strong attractive force from the central particle of another CG bead they were forced to move too far away from its own central particle, effectively “breaking” the bond between these two particles, or conversely superimpose themselves on one another, resulting in simulation failure due to extremely high bond and LJ energies, as result of overstretched bonds and overlap of LJ spheres.

Identifying and addressing these issues required extensive collaboration with our colleagues at ETH Zurich, particularly as the GROMOS molecular dynamics software code was being continuously updated as our testing proceeded. This ultimately resulted in several changes to the internal structure of the GROMOS MD software, as well as to the user inputs. New options were added to the MD input file to facilitate correct pressure scaling and treatment of electrostatic screening for supra-atomic CG models. A small repulsive LJ term was also added to the dipole particle, following the recent update of the GROMOS CG water model. A repulsive LJ on the dipole CG atoms prevented the over-attraction arising between the dipole and a neighbouring central particle.

4.3.3 Bonded parameters

After these issues were addressed, optimisation of the CG-PC parameters could occur. During the parametrisation process, parameters were fitted sequentially. Firstly, a range of bond parameters were scaled to establish a reasonable set for bonding to CG beads together. For this CG water bead non-bonded parameters were used, as two CG beads were linked in a box of CG water for a simplistic test system. After bonded parameters were established, non-bonded LJ and charge parameters were refined to match the reference data to the point of making the model functional.

For the bonded parameters, the distances between the centres of mass of functional groups corresponding to CG beads were calculated from each CG simulation and plotted as a histogram and compared to reference AL bilayer data. A range of bond force constant values was tested for the C-P bond, ranging from $1.0x10^5$ kJ.mol^{-1} through the typical $2.00x10^6$ kJ.mol^{-1} to $8.00x10^6$ kJ.mol^{-1}. An extended range of values was tested, as it was found difficult to accurately fit CG results to reference data. The lower threshold for the range of values was based on the CG alkane parameters reported in A.P Eichenberger's PhD thesis, staring at $1x10^3$ kJ.mol^{-1}. The upper end was extended a power of ten above the common top values found for AL-AL bonds ($1x10^7$ kJ.mol^{-1}) The C-P bond distance distribution, with a special focus on peak values arising from CG test simulations (energy minimised and equilibrated for 1 ns) was compared to those computed for the DPPC bilayer to assess whether the CG-PC parameters were suitable. Simulations had to be kept short, due to the large number of values scanned and the relatively slow nature of GROMOS MD code, which wasn't well parallelised for multiple processing systems above 16 CPU cores. Therefore, focus was made on getting approximate initial parameter values correct to be refined at a later date with longer simulations.

Table 4.3: Final CG-PC bond stretching parameters relating to Equations 4.1 and 4.2

Bond	*K_b (final)*	*r (final)*
CC-PC	2.00000e+08	4.31e-01
CC-CD	2.00000e+06	2.00e-01
PC-PD	2.00000e+06	2.00e-01

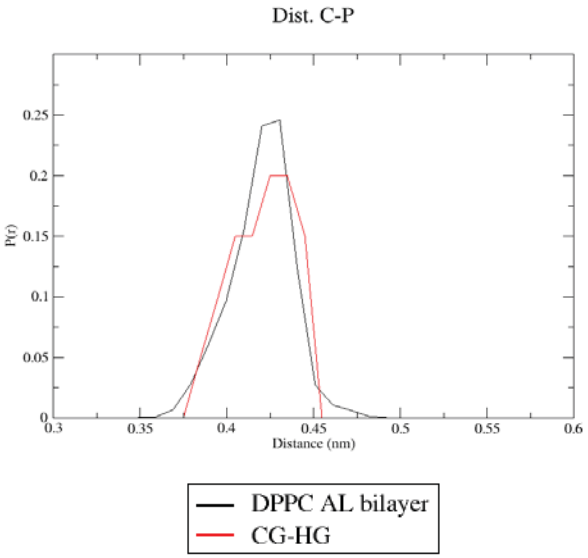


Figure 4.9: Bond distance distribution comparison between reference simulations and the CG-PC model.

After scanning through the proposed bond force constant values, the CG-PC distance distribution that most closely matched the average of the AL DPPC bilayer distribution was identified. Increasing the bond force constant resulted in a narrower distribution, as expected. Reducing the force constant not

only caused the distributions to become flatter, but also resulted in the peak value shifting from that of the reference simulation, which was also the equilibrium bond distance in the quartic potential, indicating some competition between the bond term and the other forcefield terms. Note that there was no reference point for the centre-dipole bond; while this could be parameterised by reference to the dipole moment, which, in turn, was influenced by the other charged particles in the simulation and their non-bonded parameters.

4.3.4 Non-bonded parameters

The non-bonded parameters were also refined. First, partial charges for the C and P beads had to be established. To mimic the functional groups, the C bead was given a +1 overall charge and the P bead an overall charge of -1, also following the AL model by summing all the constituent atom charges. To original CG water had an approximately +0.5/-0.5 charge split across the central and dipole beads, roughly equalling to a total charge difference of magnitude 1. It was decided for the initial model to have a similar distribution for the charged CG lipid beads, but arrange it in a way that the total sums added up to positive and negative one for Cho and Pho respectively. (*Table 4.4*). However, it was found that the dipoles produced by charges of such magnitude were far too large, compared to original distributions. Ultimately, the charge split had to be scaled down by one order of magnitude such that the both the dipole moment distribution was of the same magnitude the AL reference data (*Figure 4.9*). The charges were not parametrised further in the "proof-of-concept" model.

Table 4.4: CG-PC initial and final particle charges

CG particle	*Initial Charge*	*Final Charge*
CC	+ 1.5	+ 1.05
CD	- 0.5	- 0.05
PC	- 1.5	- 1.05
PC	+ 0.5	+ 0.05

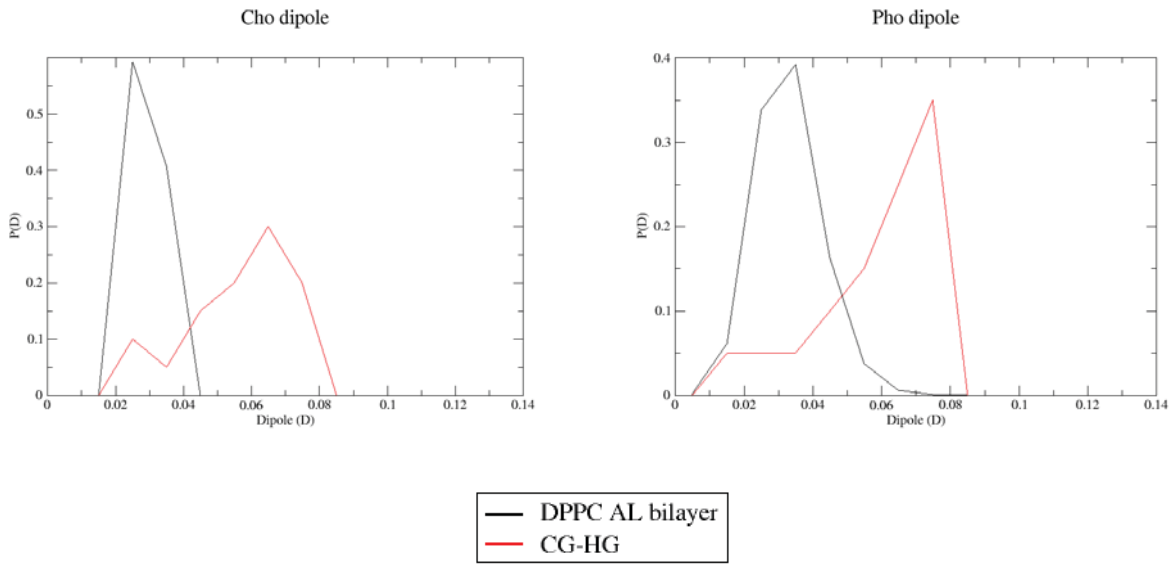


Figure 4.10: Dipole moment distributions for the two bead CG-PC model.

To parameterise LJ σ, the radial distribution function (RDF) analysis of the AL bilayer was performed to give better indication of intramolecular distances between mapped beads (*Figure 4.11*). Unfortunately, the software that computes the RDF was not able to do so for the centre of mass of groups of atoms, so the nitrogen-nitrogen (from the choline moiety) and phosphorus-phosphorus (from the phosphate moiety) pairings were used as a best estimate of the inter-molecular distances between the CC-CC and CP-CP particles in a lipid bilayer. The RDF was used rather than the intra-molecular distance to parameterise the bond equilibrium value because atoms, or CG beads separated by less than three bonds are excluded from LJ interactions.

Furthermore, in GROMOS forcefields, LJ parameters are assigned by multiplying individual atom/particle type parameters to get pair interactions between pairs. Therefore, intermolecular distances between pairs of like CG beads were required for the fitting. Distances between Nitrogen and Phosphorus pairs in the reference bilayer were calculated to provide data for Cho-Cho and Pho-Pho interactions (*Figure 4.10*). In principle, because we were interested in the radial distances, the periodicity of lipid molecules arranged in the bilayer should have minimised the effect of not using the centre of mass calculations. The first peak of the RDF indicated the most common short-range distance between the centre of each CG bead of the same type and therefore should approximately equate to the LJ radius σ, the inter-bead distance at which the LJ energy switches from being favourable (distances > σ) to unfavourable (distances < σ).

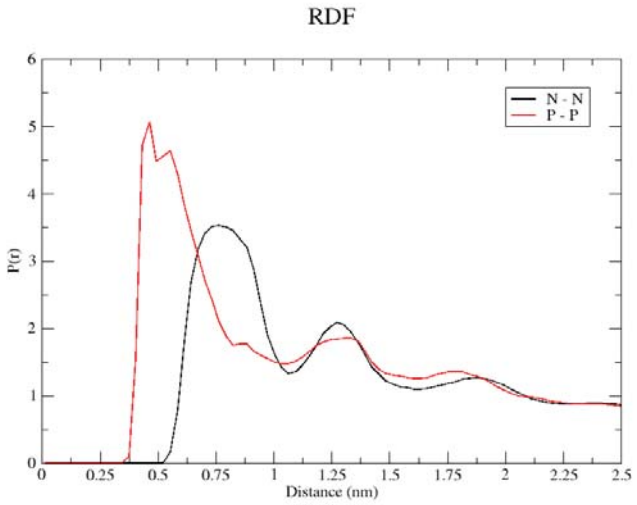


Figure 4.11: Radial distribution function of Nitrogen-Nitrogen and Phosphorus-Phosphorus atom pairs form the initial reference 40 ns DPPC bilayer simulation.

The ε parameter of the LJ interaction function was estimated to be 1.5 kJ.mol^{-1}, based on the GROMOS CGW model. Other GROMOS CG solvents[135] also had similar ε parameters, ranging from 2.1 to 3.5 kJ.mol^{-1}. During parameterisation, a range of values were scanned to establish better, more appropriate LJ parameters for the two beads. A pass/fail test based on the successful completion of a 1 ns simulation was implemented to quickly scan through different combinations of *σ* values for CC and CP. All following LJ parameter value combinations resulting from the combination of the initial *σ* value lists were attempted:

CC σ (nm): 0.3, 0.4, 0.5, 0.6, 0.7
PC σ (nm): 0.3, 0.4, 0.5

These *σ* values were then further refined and ultimately increased to be 0.65 nm (CC) and 0.6 nm (PC) to more closely match N-N and P-P RDF peak distances (*Figure 4.11*). The final values can be seen in *Table 4.5*. LJ parameters for the dipole particles were kept the same as the GROMOS CG water dipole particles.

Finally, the CG-PC model was found to result in stable simulation for at least 1 ns, proving that polarisable CG beads could be joined together by a bond and, therefore a more complex model of the CG lipid headgroup could be attempted. The development of this model had required substantial modification of the GROMOS code (carried out by our collaborators at ETH Zurich). The parameter set generated after the parameterisation procedures described above replicated the structural properties calculated from the reference simulation applicable to this two-bead system used for parameterisation. Further parameterisation at this stage was not carried out, as the parameters were likely to require

further refinement once the remainder of the head group was added. Thus, having carried out this proof of principle test, the model could be expanded to replicate the entire headgroup of the DPPC molecule.

Table 4.5 CG-PC LJ parameters.

CG particle	*σ (nm)*	*√C6*	*√C12*
CC	0.65	0.4525	0.0341
CD	0.15*	0.0000	2.7901×10^{-5}
PC	0.60	0.2799	0.0131
PD	0.15*	0.0000	2.7901×10^{-5}

**As repulsive potential for the dipole particles was refined as pure C12 value, σ was back calculated from the C12 values, assuming the epsilon is equal to 1.5 kJ.mol^{-1}.*

4.2.3 Headgroup model (CG-HG)

Once the CG-PC "proof of principle" two-bead model had been established to work, the model for the entire headgroup could be created. As work was still continuing on the parameterisation of the alkane parameters for the carbon tails, the lipid headgroup CG beads could be parameterised at the same time. The key points of parameterisation of this model were the introduction of novel CG angle and torsional dihedral type terms due to the presence of multiple beads. Several new bonded terms had to be introduced and parameterised also. New non-bonded parameters also had to be fitted for all of the new bead types.

Design of CG-HG model and initial parameter estimates

The CG-HG headgroup molecule consists of five CG beads: C, P, G, E1 and E2 (*Figure 4.10*), which are analogous to the previous model, apart from the non-polar G bead, that only has a central particle. The initial charges and parameters for the LJ function and the CP-DP (*Equation 4.2*) and CP-CP bonds (*Equation 4.1*) were assigned as described for the CG-PC model. In order to describe the behaviour of a multi-bead CG molecule, new bonded interaction terms had to be created and parameterised (*Equations. 4.3* and *4.4*). Parameterisation was carried out by simulating a single CG-HG molecule in SPC AL water due to the difficulties encountered with CGW solvation and for direct comparison with AL simulations (See *Section 4.5: Methods*).

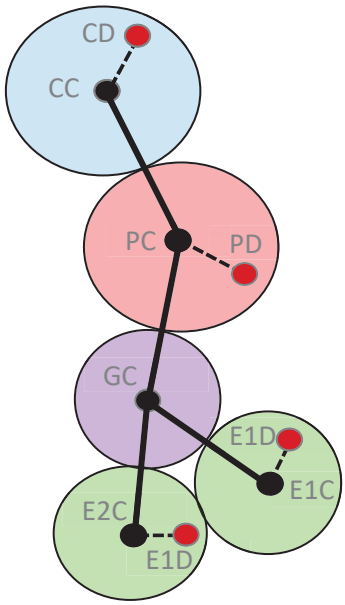


Figure 4.12: CG-HG model. The spheres represent the (blue) choline, (red) phosphate, (purple) glycerol and (green) ester portions of the head group of a phosphocholine lipid. The central particles are shown as black spheres, the dipole particles as red spheres, bonds between particles as black lines, bonds between centre and dipole particles as dashed black lines and the effective LJ radius of each bead is represented by the respective coloured areas.

Bonded term parameterisation

The parameterisation of bonded terms was completed by distribution matching between CG test and AL reference simulations. After starting from initial values parameters were iteratively improved through several versions of the model. Values for the force constant k_x for all bonded functions scanned across multiple values typically from 1.0×10^4 to 1.5×10^6 kJ.mol^{-1}. In some cases the ideal distance or angle value had to be adjusted manually to ensure that the modal value of the CG distribution matched that of the AL reference simulation.

Table 4.6: Final CG-HG bond stretching potential parameters.

Bond	k_b *(kJ.mol^{-1})*	*r (nm)*
CC-PC	$5.50 \times 10^{+04}$	0.410
PC-GC	$4.50 \times 10^{+05}$	0.347
GC-E1C	$3.50 \times 10^{+05}$	0.336
GC-E2C	$2.00 \times 10^{+06}$	0.292

Table 4.7: Final CG-HG half harmonic dipole bond stretching potential parameters.

Bond	*k_b (kJ.mol^{-1})*	*r (nm)*
CC-CD	2.00 x10^{+06}	0.100
PC-PD	2.00 x10^{+06}	0.100
E1C-E1D	2.00 x10^{+06}	0.100
E2C-E2D	2.00 x10^{+06}	0.100

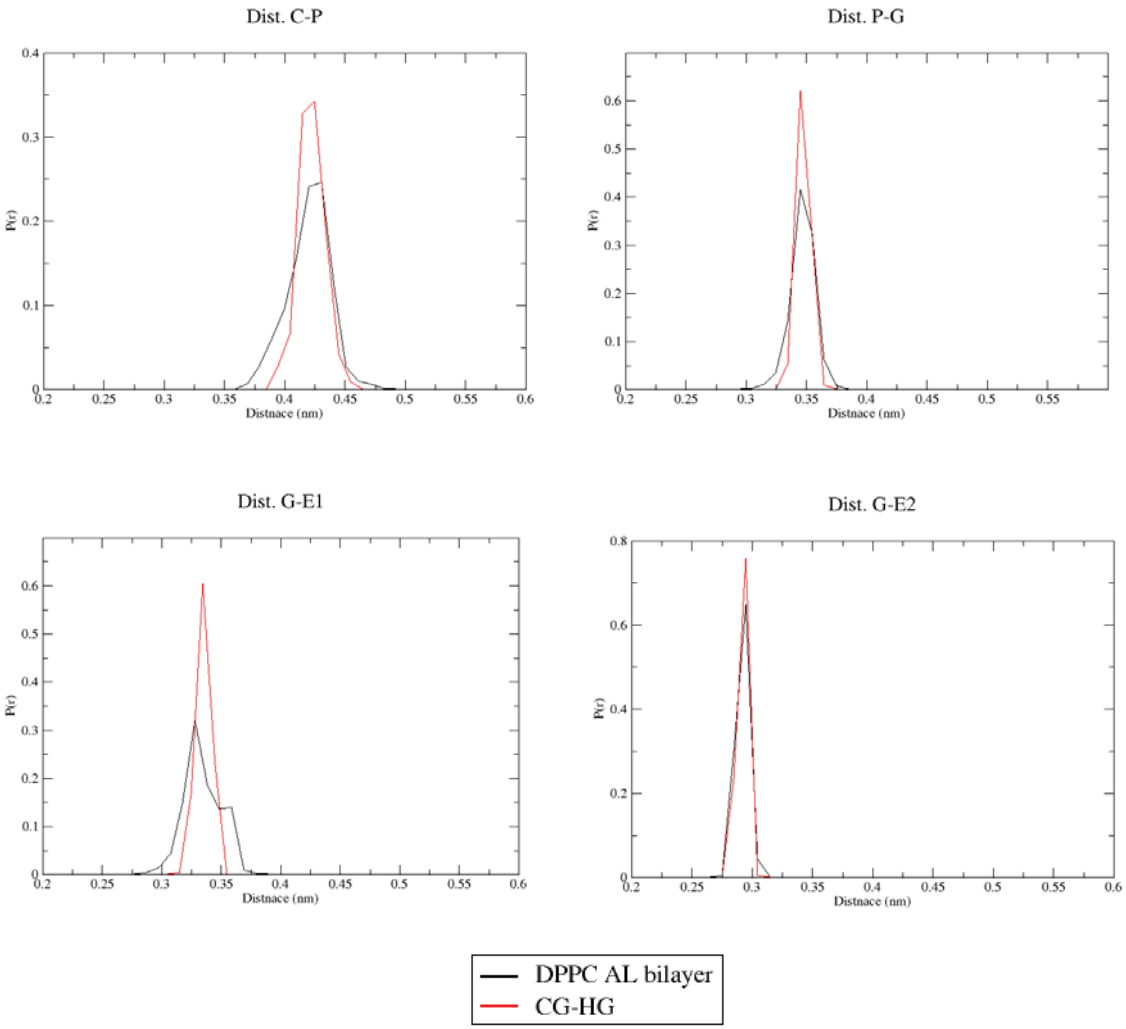


Figure 4.13: Distributions of the inter-bead distances calculated from the CG-HG simulation carried out with the bonded parameters listed in Tables 4.6 and 4.7 and from the reference AL DPPC bilayer simulation.

Table 4.8: Final CG-HG angle bending potential parameters.

Angle	k_θ *(kJ.mol^{-1})*	θ *(degrees)*
CC-PC-GC	4.50 x10^{+02}	86.9
PC-GC-E1C	1.20 x10^{+02}	128.0
PC-GC-E2C	2.00 x10^{+02}	123.0
E1C-GC-E2C	1.70 x10^{+02}	250.0

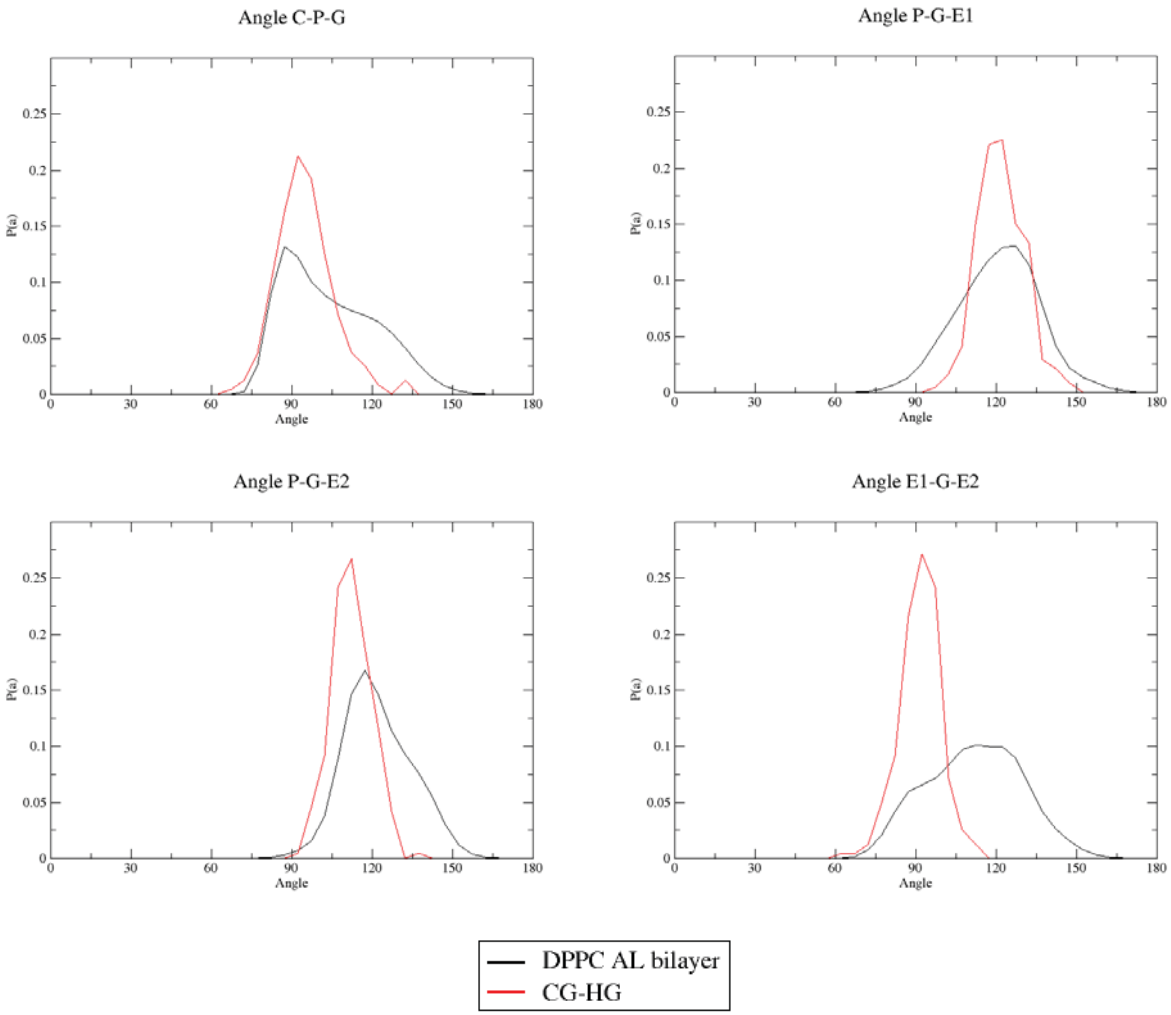


Figure 4.14: Distributions of the angles calculated from the CG-HG simulation carried out with the bonded parameters listed in Tables 4.8 and from the reference AL DPPC bilayer simulation.

From the final parameterisation results (*Figures 4.13 and 4.14*) of the distance and angle terms, it was observed that most distributions and modal values of structural properties of the CG-HG molecule were similar to those calculated for the AL DPPC bilayer.

The exceptions were apparent in the case of P-G-E2 and E1-G-E2 angles where it proved not possible to match the exact AL DPPC distributions. A range of k_a and θ^0 values was tried for P-G-E2 and E1-G-E2 but it was found that the peak value was not strongly affected by changes in parameter magnitudes. At higher k_a values the distribution would become too narrow, but the modal value did not shift to match the ideal θ_0 value. At lower k_a values, the peak of the distribution would deviate strongly from the ideal values. Changing the ideal angle value had little effect on the resultant distribution peak angle. The biggest discrepancy was observed in the E1-G-E2 case. One possible explanation is that the behaviour of the angles for the single CG-HG molecule in solution might differ from that observed for the AL DPPC bilayer due to the absence of lipid tails and general bilayer structure and packing, which provided higher-level organisation. It is also possible that there was strong competition between the bonded and non-bonded terms in the CG model that prevented the satisfaction of the angle potential term. Alternatively, inherent changes in structure and loss of asymmetry in this part of the lipid as a result of the CG mapping could also be the cause for this.

Interesting cases were also observed in the G-E1 distance and C-P-G angle distributions, where the original AL simulation distribution had shoulders. Non-symmetrical distribution features such as these are common and arise from centre of mass averaging of atom clusters in distribution analysis. All CG model bonded terms were represented by symmetrical potentials, which do not allow for replication of bimodal peaks or shoulders. It was decided, therefore, to fit the CG behaviour to the primary peak of the AL distribution, which represented the most probable incidence of values for those particular structural properties.

Non-bonded term parameterisation

LJ parameters were also refined at this stage. Initial estimates for the LJ radius σ were taken from both the initial RDF peaks and the distances taken between centres of mass of AL atoms that represented CG beads, as it was difficult to assess if RDF data would provide a good basis for Gly, E1 and E2 parameterisation, as the RDF estimates were perceived as being too high for LJ radii. Distances between atomic groups representing the CG beads were the next best starting point for initial LJ σ values. The ϵ parameter was estimated to be 1.5 kJ.mol^{-1} based on CGW GROMOS water, as for CG-PC. A range of values of σ was tested for each bead in combination with each other (*Tables 4.9 and 4.10*). Two tests were performed – first for the various combinations of CC and PC σ values, keeping the GC and EC σ (√C6 and √C12) parameters constant, and second for the GC and EC combination, while keeping new CC and PC σ parameter values constant. In the first instance, the same 'pass/fail' test was used as for the CG-PC, as before the parameters could be refined so that the simulation results matched the reference data, parameter combinations needed to be found for which the simulations did not end prematurely due to extremely high energies or SHAKE constraint algorithm failure. It was found that certain

combinations of σ values for the CG beads would result in simulation failure, while other similar values caused no issue. No correlation explaining this behaviour could be established.

Final parameters were chosen based on the combination of matching AL DPPC bilayer RDF distributions (Figure 4.15), as well as the bonded inter bead distances, to provide a mix of values to parametrise to, as the RDF distributions results indicated that LJ σ should have been approximately 0.5 nm. However, a lot of the trialled values of that magnitude have failed, so smaller σ values were chosen (*Table 4.11*).

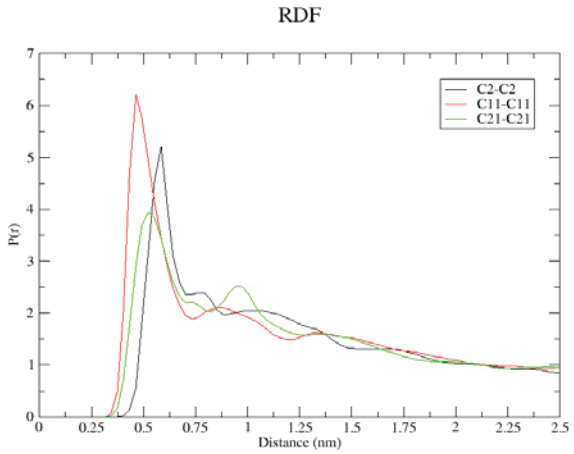


Figure 4.15: Radial distribution function of C2-C2 (from glycerol), C11-C11 (from ester 1) and C21-C21 (ester2) atom pairs form the initial reference 40 ns DPPC bilayer simulation.

Table 4.9: Results of a simulation pass/fail test for an array of trial CC and PC LJ σ values (nm). σ_{GC} = 0.425, $\sigma_{E1C/E2C}$ = 0.45. "✓" marks denote successful simulations, "X" marks failed simulations.

σ CC/σ PC -	*0.35*	*0.40*	*0.45*	*0.50*
0.45	X	✓	✓	X
0.475	X	✓	X	✓
0.485	X	X	X	X
0.5	X	✓	✓	✓
0.55	X	X	X	X

Table 4.10: Results of a simulation pass/fail test for an array of trial GC and EC LJ σ values (nm). σ_{CC} = 0.6, σ_{PC} = 0.65. "✓" marks denote successful simulations, "X" marks failed simulations.

σ GC/σ EC -	*0.35*	*0.36*	*0.37*	*0.38*	*0.39*	*0.40*	*0.425*
0.35	X	X	X	X	X	X	X
0.36	✓	X	✓	X	X	✓	X
0.37	✓	X	X	X	X	X	X
0.38	✓	✓	X	X	✓	X	✓
0.39	X	X	✓	X	✓	X	✓
0.40	✓	✓	X	✓	✓	✓	✓
0.425	✓	✓	✓	✓	✓	✓	✓
0.45	X	X	X	X	X	X	X

Table 4.11: CG-HG final LJ parameters.

CG particle	*σ (nm)*	*√C6*	*√C12*
CC	0.480	0.2790	9.681×10^{-4}
CD	0.150	0.0000	4.797×10^{-7}
PC	0.400	0.1555	9.884×10^{-3}
PD	0.150	0.0000	4.797×10^{-7}
GC	0.400	0.1555	9.884×10^{-3}
E1C	0.400	0.1555	0.009884
E1D	0.150	0.0000	4.797×10^{-7}
E2C	0.400	0.1555	9.884×10^{-3}
E2C	0.150	0.0000	4.797×10^{-7}

Table 4.12: Initial and final charge parameters for the CG-HG model beads.

CG particle	*Initial Charge*	*Final Charge*
CC	+ 1.10	+ 1.05
CD	- 0.10	- 0.05
PC	- 1.10	- 1.48
PD	+ 0.10	+ 0.48
GC	+ 1.20	+ 1.20
E1C	- 1.00	- 1.02
E1D	+ 0.50	+ 0.52
E2C	- 1.00	- 1.05
E2D	+ 0.30	+ 0.35

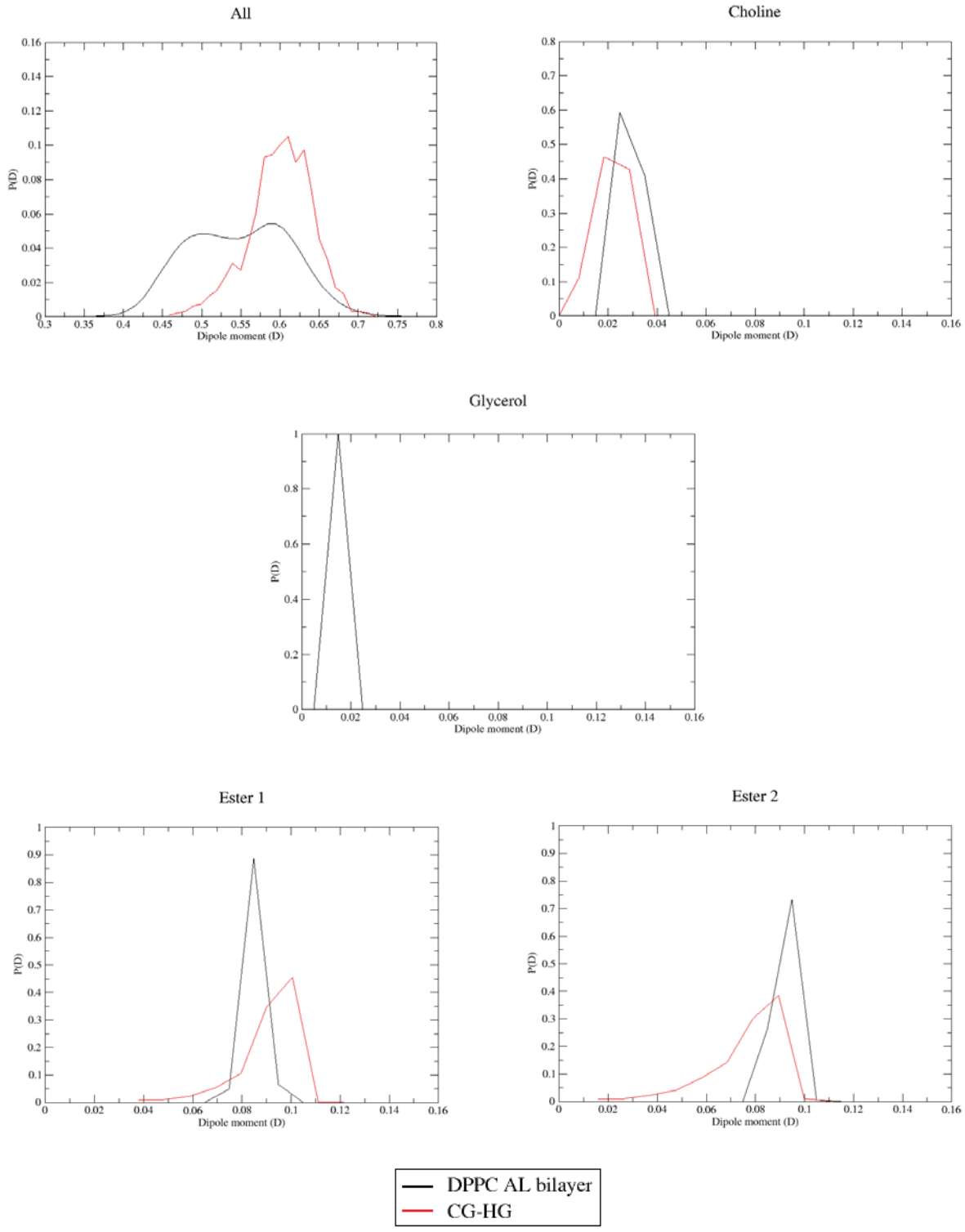


Figure 4.16: Matched dipole moment distributions for the CG-HG model. Reference graph for glycerol is included as justification for not having a polarisable CG bead – the distribution is only 1 bin wide, as the spread of dipole moments for Gly group was very narrow and small in magnitude.

The values of the partial charges, including how these are distributed between the centre and dipole particles on the PC-HG, were optimised by matching test simulation dipole distributions to those calculated from the reference bilayer simulation (*Figure 4.16*). The centre particle carried the bigger charge and its sign set the overall positive/negative nature of the CG bead. It was found that to replicate the magnitudes of the AL dipoles, the dipole charge needed to be much smaller than the central charge, in a 1:10 - 1:20 ratio (*Table 4.12)*.

The dipole distributions calculated from the reference AL DPPC simulation for the phosphate and ester groups were well replicated in the CG-HG model, both in terms of peak alignment and distribution width. It was more difficult to parameterise the choline bead because of the small magnitude of the dipole involved. The positive nitrogen atom in the AL equivalent is situated in the middle of the three negatively charged methyl carbons and one negatively charged CH_2 atom, making the top of the headgroup almost symmetrical. Instead of removing the dipole, as was done in the case of non-polar G bead, it was decided to keep the choline dipole intact, as there is still a certain degree of polarisability present in the molecule. Furthermore, if choline was substituted for a different, more polar group (e.g. the serine moiety in phosphoserine), the methodology still needed to be generally applicable, making the polarisable choline a useful, general test case.

Despite the good reproduction of the individual dipole moment distributions of each CG bead, the overall dipole moment of the CG headgroup was not well matched to that calculated from the reference AL DPPC simulation. It is possible that this could be fixed in the further iterations of the CG models, for instance with attachment of the carbon tails and a simulation of an entire bilayer (rather than of individual molecules), which would change the local environment experienced by the lipid. Other possible future tactics could include increasing overall magnitude of charges on the opposing ends of the headgroup. As the focal point of the CG-HG model was introduction of several new bonded terms, and the environmental context would highly likely to affect the overall dipole moment, further fine tuning of dipole interactions was left to subsequent models.

In conclusion, a basic model of the DPPC lipid headgroup has been established. Working bonded and non-bonded parameters were created for the newly introduced G, E1 and E2 beads and their associated interactions. Several features, such as the P-G-E2 and E1-G-E2 angles, as well as torsional dihedrals could not be matched well to the reference AL data, despite scanning a wide range of parameter values. It was decided to proceed to the full DPPC lipid CG model regardless, adding the hydrocarbon tails, as it would be inefficient to over-refine the parameters for a preliminary model, only to have to re-refine them if the completion of the molecule and ability to simulate a bilayer had a large effect on the molecular properties.

4.3.4 Complete DPPC model (CG-DPPC)

Now that a functional headgroup model that reproduced the majority of the mapped AL data had been obtained, the next step was to combine it with hydrocarbon lipid tails, based on the recently parameterised GROMOS CG alkane models[193] to create a complete CG-DPPC model (*Figure 4.4*). The addition of the lipid tails required incorporating all the associated CG alkane and CG-HG parameters into a single set of parameter files, as well as establishing several new bonded, angle and torsional dihedral interactions related to the addition of the lipid tails.

Tails would consist of five alkane bead particles each - C3M, C3M, C2M, C3M, C3E (*Figure 4.4*). The C- denotes that the bead is an alkane chain carbon, the number is associated with number of equivalent AL carbon atoms and -M and -E denote whether the bead is the end or middle bead for its particular chain. All carbon beads contain no dipole particle and are uncharged (just like their AL analogues). All bond, angle and torsional dihedral parameters for the CG alkane beads were taken directly from Eichenberger et al[193].

During the parameterisation of the CG-DPPC molecule, the LJ parameters were taken directly from the CG-HG model for the headgroup beads and from A. P. Eichenberger *et. al*.[193] for the alkane tails. CG-DPPC parameterisation focused on optimisation of the electrostatic parameters and ensuring that the existing bonded parameters were still suitable, as well as adding the alkane tail related bonded parameters. Optimisation of the LJ parameters was reserved for the next stage of the parameterisation, when a CG-DPPC bilayer would be simulated, and so data on lipid-lipid intermolecular interaction could be gathered and compared to the reference AL DPPC bilayer simulation.

Solvent model use during parameterisation

There were two water models that could be used during the parameterisation of the CG lipid models: SPC water (standard GROMOS forcefield AL solvent) and the GROMOS CG water. Using the AL SPC water model allowed straightforward comparison with the reference AL DPPC simulations, which were also carried out in SPC water. It also guaranteed that the CG lipid models were capable, from the outset, of interacting with AL molecules in a multigrain simulation scenario. On the other hand, simulations in CG water would run faster, and would avoid the problems with mixing model resolutions that required reparameterisation of the CG water model when first used with AL proteins, for instance.

Initially the goal was to use both water models during the later stages of the parameterisation process. Unfortunately, however, this was not feasible, despite several different methods for solvating the CG-HG, CG-DPPC and CG-DPPC bilayers in CG water being tested. Thus the parameterisation of CG-HG, CG-DPPC and the re-parameterisation of CG-DPPC once in a bilayer all took place only in SPC water.

Bonded term parameterisation

New bonded parameters associated with headgroup-tail interactions had to be introduced in this version of the model. Bonds between ester and tail carbon beads E1-C3, E2-C3; angles in the same area G-E1-C3, G-E2-C3; and torsional dihedrals P-G-E1-C3, P-G-E2-C3, G-E1-C3-C3, G-E2-C3-C3. The starting basis for these parameters was taken from the published alkane parameters[193], treating G, E1 and E2 groups as comparable C3 and C4 beads. A series of 20 simulations were performed, iteratively, where parameters were changed manually to increase or decrease the charges on the headgroup particles, to better fit the reference dipole moment data. Furthermore, centre dipole particle bond lengths were shortened in cases where charge differences got minute, as an alternative to minimising the dipole moment. The distributions of the structural properties corresponding to the bonded interaction terms for the final set of parameters can be found below in *Figures 4.17 to 4.19* and the parameters themselves in *Tables 4.13 to 4.16.*

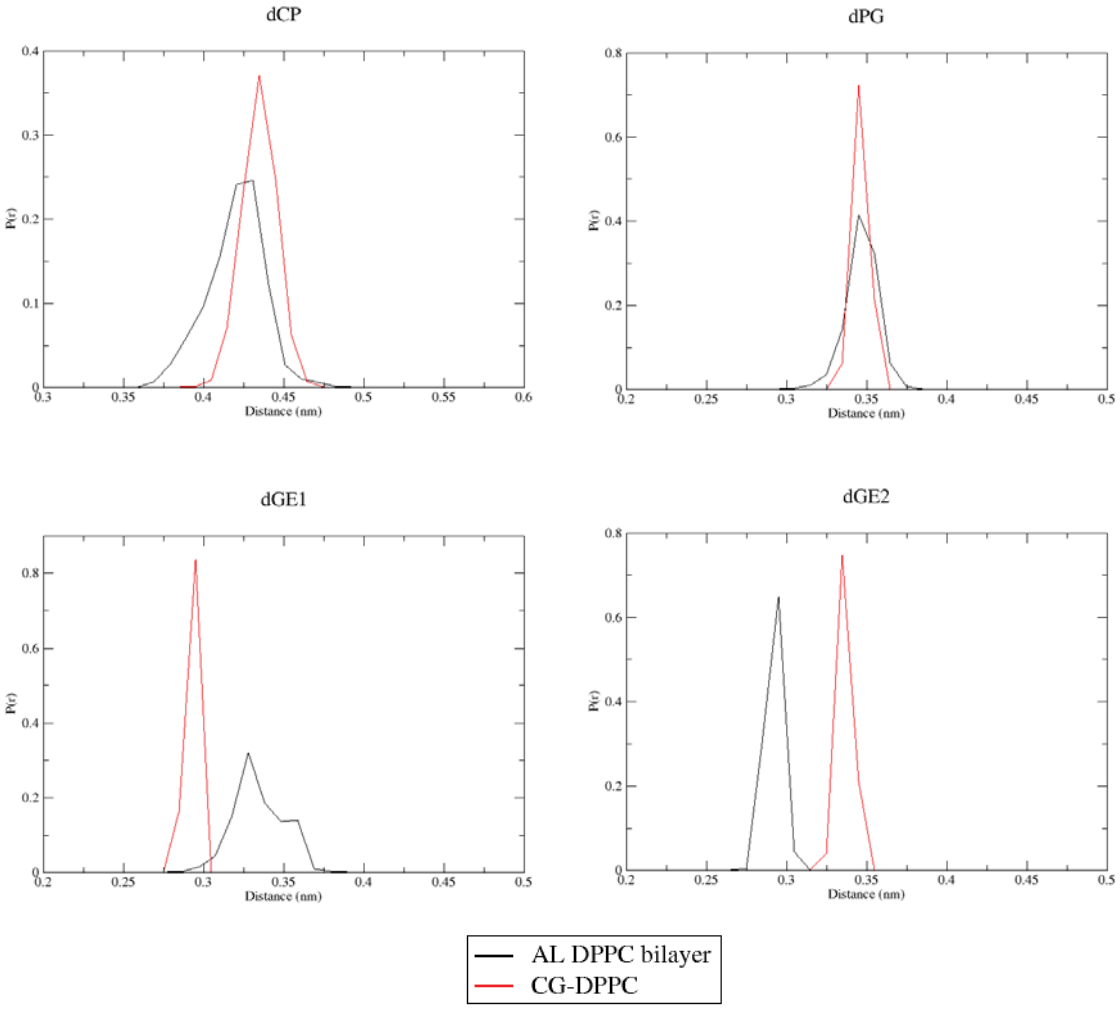


Figure 4.17: Distributions of the inter-bead distances calculated from the CG-DPPC simulation carried out with the final set of optimised parameters listed in Tables 4.11-4.12 and from the reference AL DPPC bilayer simulations.

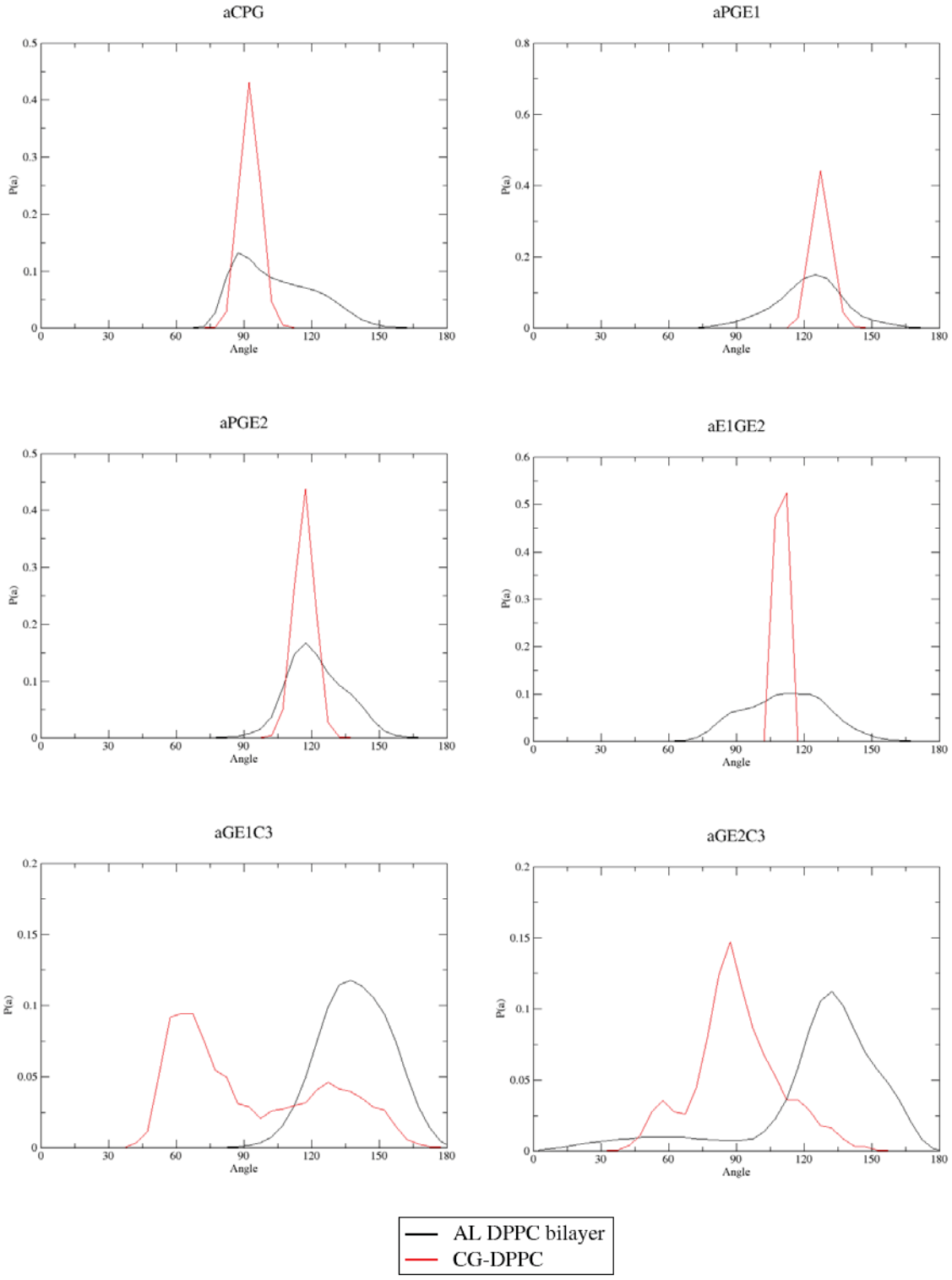


Figure 4.18: CG-DPPC parameter matching for angle terms with DPPC AL bilayer and CG-HG are included for reference.

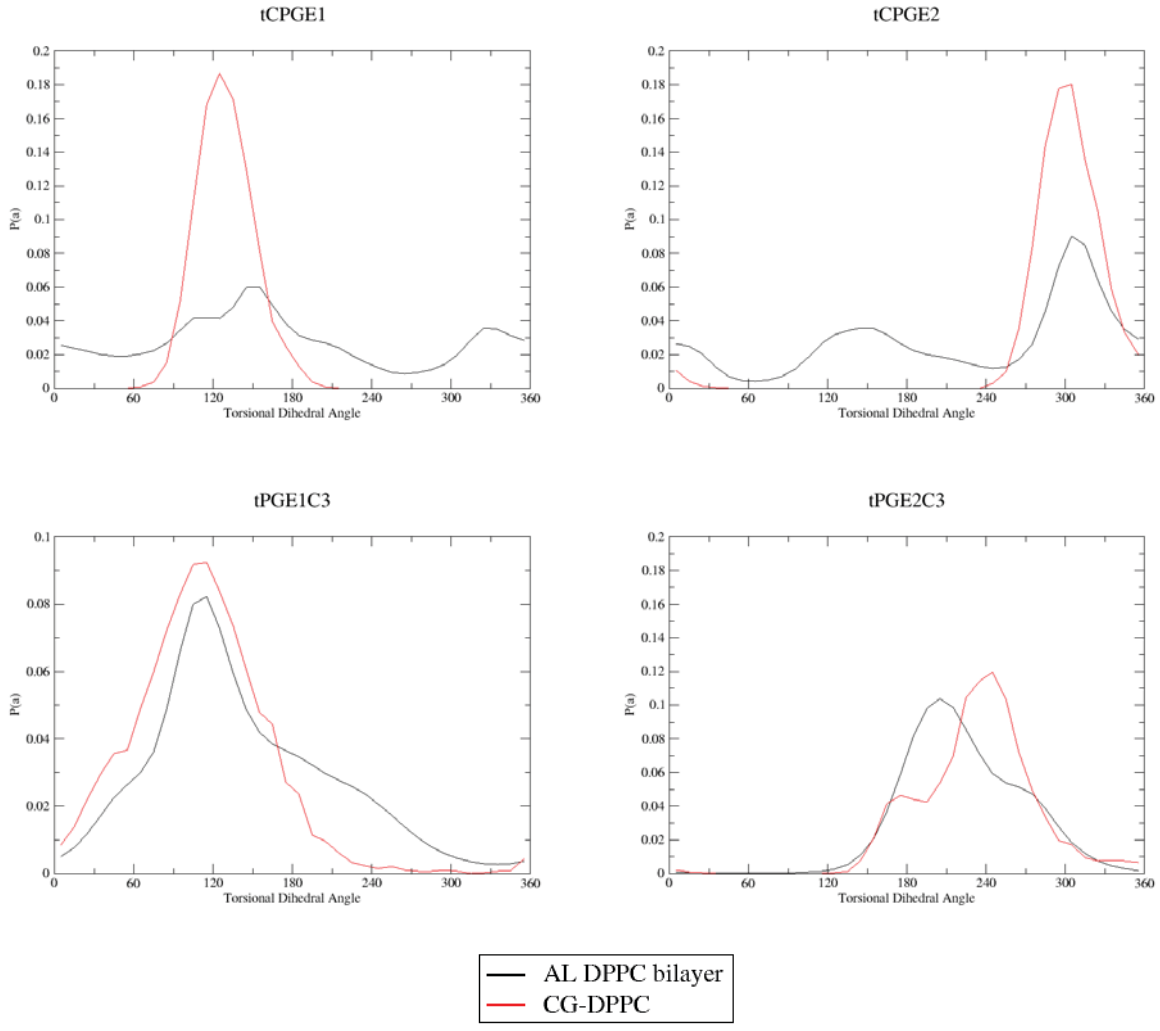


Figure 4.19: CG-DPPC parameter matching for torsional dihedral terms with DPPC AL bilayer and CG-HG are included for reference.

Table 4.13: CG-DPPC bond stretching potential parameters.

Bond	*k_b (kJ.mol^{-1})*	*r (nm)*
CC-PC	4.50 x10^{02}	0.430
PC-GC	4.50 x10^{02}	0.346
GC-E1C	3.50 x10^{03}	0.336
GC-E2C	4.00 x10^{03}	0.292
E1C-C3M	1.80 x10^{04}	0.338
E2C-C3M	1.80 x10^{04}	0.338
C3M-C2M	6.00 x10^{04}	0.303
C3M-C3E	2.50 x10^{04}	0.358

Table 4.14: CG-DPPC dipole bond stretching potential parameters

Bond	k_b *(kJ.mol^{-1})*	*r (nm)*
CC-CD	2.00 x10^{08}	0.06
PC-PD	2.20 x10^{08}	0.06
E1C-E1D	2.00 x10^{08}	0.10
E2C-E2D	2.00 x10^{08}	0.10

Table 4.15: CG-DPPC angle potential parameters

Angle	k_θ *(kJ.mol^{-1})*	θ *(degrees)*
CPG	4.50 x10^{02}	101.0
PGE1	4.50 x10^{02}	154.0
PGE2	3.50 x10^{03}	146.0
E1GE2	4.00 x10^{03}	112.0
#GE1C3	2.00 x10^{02}	142.5
#GE2C3	2.00 x10^{02}	133.0
E1C3C3	2.00 x10^{02}	174.3
E2C3C3	2.00 x10^{02}	154.7
C3C3C2	5.89 x10^{01}	157.1
C3C2C3	5.33 x10^{01}	147.2

Table 4.16: CG-DPPC torsional dihedral angle potential parameters

Angle	k_Φ *(kJ.mol^{-1})*	ϕ*' (degrees)*	*Multiplicity*
CPGE1	6.000	335	1
CPGE2	6.000	120	1
PGE1C3	8.000	300	1
PGE2C3	8.000	25	1
CC in tails	0.737	180	1

The distance distributions calculated from the CG-DPPC model mostly matched those from the reference AL DPPC simulation, and therefore did not require much adjustment to the parameters inherited from the previous CG-HG model. The exceptions were the G-E1 and G-E2 distances. In these two cases the addition of the hydrocarbon tails shifted the distance distributions such that the CG G-E1 distance distribution matches the AL G-E2 distribution and vice versa. A possible cause was the loss of asymmetry around the glycerol – ester links as result of coarse-graining combined with the steric effects of adding the hydrocarbon tails.

The agreement of the E1-G-E2 and P-G-E2 angle distributions with their AL reference counterparts improved compared to the previous CG-HG model, while retaining agreement of the C-P-G and P-G-E1 angle distributions. However, it was not possible to obtain a good match to the reference data for the G-E1-C3 and G-E2-C3 angle distributions, despite testing a range of parameter values. It was decided to leave further refinement of these properties until the bilayer simulations.

Non-bonded term parameterisation

The charges were fitted by manual iterative improvement through the same set of 20 MD simulations, as it was found that changing the charge distributions did not have any effect on the reproduction of the structural properties described by the bonded terms, so to save time, these parameterisation procedures could be carried out simultaneously. The charges were optimised based on reproduction of the dipole moment distributions and average values calculated from the reference AL DPPC simulation. Primarily, the magnitude of the partial charge on each particle and the difference between the magnitude of the centre and dipole particle changes were altered, without changing the total charge of each CG bead. Furthermore, the lengths of the centre- dipole bond were varied and decreased in order to decrease the dipole magnitude without further decreasing the difference in magnitude of the charge on each particle and still retaining a degree of polarisability for the bead. The centre-dipole bond force constant was increased compared to the CG-HG model to ensure that the dipoles particles could not move outside the effective LJ radius of the bead due to attraction from a neighbouring opposite charge. The final charge distribution for CG-DPPC can be found below in *Table 4.17* and the matching graphs in *Figure 4.20*.

Table 4.17 CG-DPPC final charges.

CG particle	*Final Charge*
CC	+ 1.080
CD	- 0.080
PC	- 1.100
PD	+ 0.100
GC	+ 1.200
E1C	- 1.085
E1D	+ 0.585
E2C	- 1.230
E2D	+ 0.530

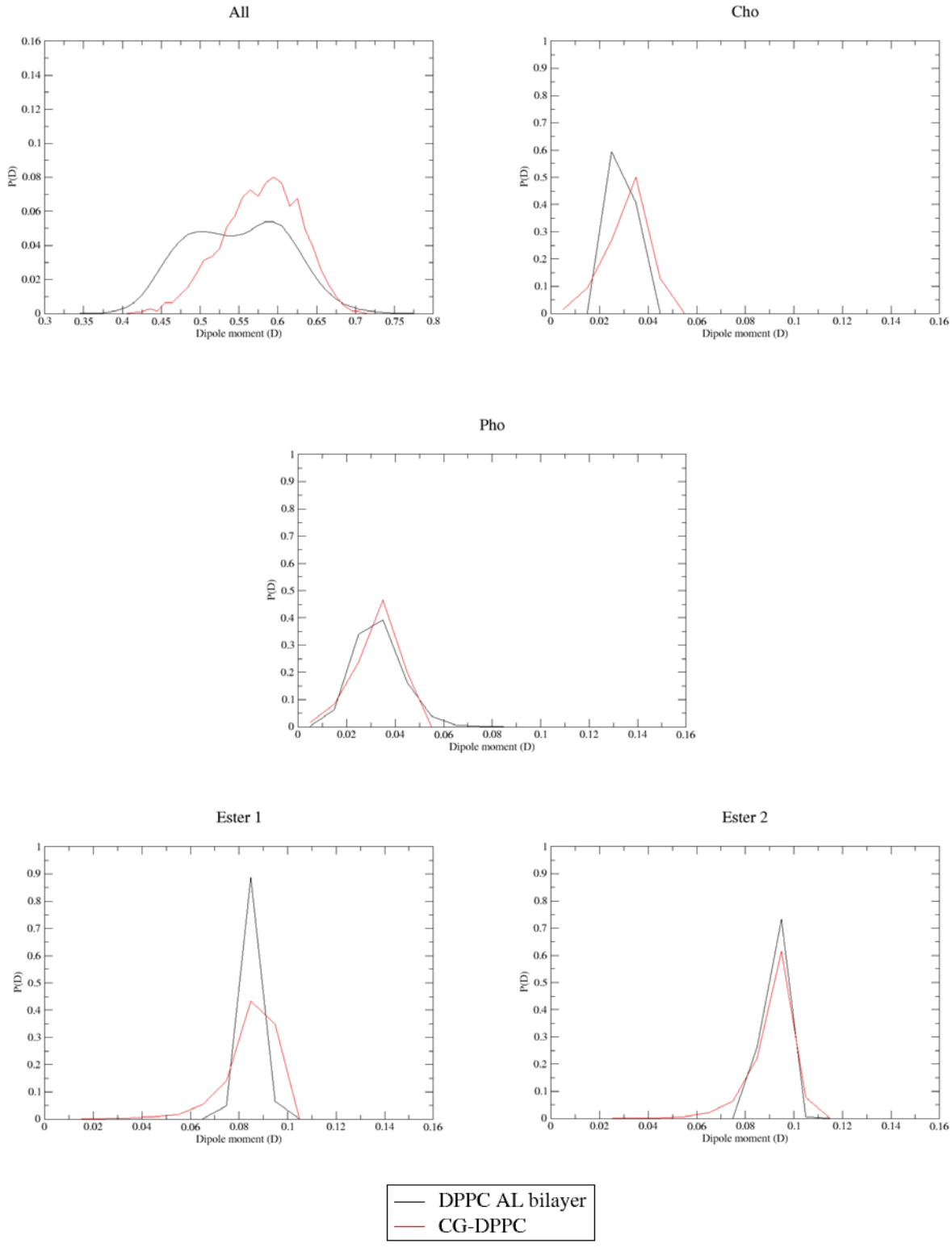


Figure 4.20: Matched dipole moment distributions for the CG-DPPC model.

*Table 4.18: CG-DPPC LJ parameters.**

CG particle	*σ (nm)*	*√C6*	*√C12*
CC	0.600	0.5291	0.1142
CD	0.150	0.0000	4.7970×10^{-7}
PC	0.650	0.6727	0.1847
PD	0.150	0.0000	4.7970×10^{-7}
GC	0.425	0.1880	0.0144
E1C	0.425	0.1880	0.0144
E1D	0.150	0.0000	4.7970×10^{-7}
E2C	0.425	0.1880	0.0144
E2D	0.150	0.0000	4.797×10^{-7}
C3M	0.488	0.2849	2.7112×10^{-2}
C2M	0.420	0.1811	1.5283×10^{-2}
C3E	0.507	0.3201	3.2986×10^{-2}

**ϵ for the alkane tails and for the dipole particles was assumed to be 1.5 kJ.mol^{-1}to allow back-calculation of the corresponding σ values.*

CG-DPPC model was parameterised and found to produce stable simulations of a single CG molecule in AL water solvent. AL reference data was matched well for bonded and dipole terms. The next stage of modelling would involve setting up a simulation of a CG-DPPC lipid bilayer to test the model further. Further refinement of LJ parameters could occur in the bilayer simulation, as it would be possible to get intermolecular lipid-lipid interaction data.

4.3.5 CG-DPPC bilayer

After establishing a CG-DPPC model that worked in SPC solvent, the next step was to try and recreate a DPPC bilayer, containing 128 CG-DPPC molecules, and solvated in SPC water. All simulations were run according to the bilayer simulation methods described in *Section 4.5: Methods*.

Unfortunately, serious problems were encountered with all bilayer simulations of the polarisable CG-DPPC model. While polarisability itself did not seem to be an issue, a range of related parameter and implementation problems persisted. Each attempted simulation of the CG-DPPC bilayer failed after less than 0.1 ns due to SHAKE or “not a number” energy related problems linked to the bonded terms. The bilayer disintegrated, losing all structure and forming pockets of vacuum before the actual errors were reported. While the errors that ultimately cause the simulations to fail were present in the bonded terms, it was more likely that the causes initiated elsewhere. For instance, the CG models do not use

SHAKE, thus to instigate SHAKE errors in the SPC water, there must be inappropriate non-bonded interactions between the CG particles and the water atoms.
Possible proposed problem sources were inappropriate starting co-ordinates (e.g. particles/atoms located too close to each other), or incorrect non-bonded parameters (as interactions between CG molecules had not yet been tested), in particular, over-attraction between dipole and centre particles of CG beads in different molecules. In support of the latter theory, it was found upon analysis of the few simulations that did run for more than a few thousand integration steps that the dipole moments of the CG-DPPC beads were far greater than in the single CG-DPPC simulation or the reference simulations by a factor of five. Each of these possible problems was investigated in turn, as outlined below.

To test whether inappropriate initial coordinates were at fault, several sets of starting coordinates with different inter-molecular separations were tested. The initial coordinates were generated using an in-house bilayer-building program (Ivan Welsh, Allison Group, Massey University). The program took a single configuration of a CG-DPPC molecule, extracted from the previous simulations of CG-DPPC in SPC water, and replicated it in the xy plane at a user-specified spacing (ranging from 0.2 to 1.0 nm) to make one leaflet, which is then reflected and duplicated in the z-direction to form the complete bilayer. An alternative program – VOTCA[194-195] was also used to generate CG-DPPC bilayer coordinates by CG-AL mapping based on the reference AL DPPC bilayer coordinates. None of the bilayers built using these methods resulted in stable simulations. While the use of a single initial structure to build the bilayer, selected from a simulation of a single CG-DPPC, might be expected to be highly dependent on the choice of structure, and require substantial equilibration, the bilayer mapped directly from the AL DPPC bilayer using VOTCA should have provided a reasonable set of initial coordinates. A range of input single lipid configurations was also trialled and no built bilayer configurations were found to produce stable results Therefore, it was assumed that the issues were not caused by the coordinate generation methods.

To test whether the dipole particle and associated parameters were causing problems within the simulation, simulations were carried out without the presence of the dipole particles. The dipole particles and associated bonded parameters were removed from the model topology. The entire partial charge of each bead was therefore situated on the central particle, while all other parameters remained constant. Such an approach did not resolve the problems of the simulations ending prematurely due to SHAKE errors or “not a number” bonded energies.

While it remains possible that the LJ parameters were not suitable for representing the inter-molecular interactions between CG-DPPC molecules, it was difficult to refine these and diagnose the issues further without the ability to run simulations for more than a few integration timesteps.

Non-polarisable model of CG-DPPE with new LJ parameters

A new approach to parameterise CG lipid models using the GROMACS simulation software currently under development by Dr Thomas Collier (Allison Group, Massey University), who is focusing on DPPE (di-palmitoyl-sn-glycero-3-phosphoethanolamine). For these CG-DPPC models, the partial charge on

each bead was the sum of the partial charges of its constituent atoms in the AL forcefield, but the bonded terms were derived by Boltzmann Inversion using the VOTCA program[194-195] from two 500 ns reference simulations of AL DPPE. LJ parameters derived in the same way were not reasonable (resulting in simulation failure), so initial estimates were based on the ELBA CG lipid model[137]. A semi-CG water model based on a one-bead mapping (also using Boltzmann Inversion) from SPC water was used, as the GROMOS CG water model required changes to the GROMACS software and was not fully implemented at the time of writing. Using this non-polarisable model, a DPPE bilayer could be simulated for over 500 ns without dipole particles in GROMACS.

A GROMOS model based on the above principles and LJ parameters remains stable for at least 5 ns. Addition of the dipole particles to the lipid headgroup, however, still causes problems and is the subject of further development. Therefore, it can be hypothesised that the dipole particles, possibly in combination with inappropriate LJ parameters, were causing the problems. Specifically, it is likely to be the novel combination of properties pioneered by these polarisable CG lipid models, namely the bonding (proven to work for the non-polarisable CG alkanes) between polarisable beads containing dipole particles (proven to work for CG solvents), and the subsequent close packing of these molecules into a bilayer structure.

4.3.6 Further models and future work

In addition to the CG-DPPC models and the reference simulations, a range of other work pertaining to similar lipids was carried out. The initial scope of the PhD work anticipated that after the completion of the CG-DPPC model, the methodology could be applied to a range of similar phospholipids (see *Appendix 6* for examples of similar lipid structures). Initial reference simulations of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine) and POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine) lipids were carried out (up to 10 ns). A truncated methyl capped phospho-ethanolamide was also simulated and development of an enthanolamide CG bead was begun. However, due to the difficulties encountered in the creation of a functioning CG-DPPC bilayer, the work on other lipids was put on hold. Furthermore, anomalous motion of the amide hydrogen atoms during the reference simulations of AL POPE required substantial investigative work that ultimately revealed severe problems with the implementation of distance cut-offs and time-step algorithms in the GROMACS simulation software[196].

Further changes to CG-DPPC model will be required to create a functional bilayer. The current hypothesis is that the current LJ parameters were inappropriate for bilayer simulations. The issue was further compounded by the fact that even when the LJ parameters were adjusted with a similar DPPE model, the presence of dipole particles still caused the simulation to fail. Therefore, the next tests should focus on parameterising both LJ values for bilayer simulations along with establishing appropriate dipole particle parameters, perhaps re-evaluating centre-dipole bonded interactions and reparameterising repulsive LJ parameters on the dipole particles. Once a working CG-DPPC model has been developed, these reference simulations should allow for rapid development of analogous CG-POPE

and CG-POPS models, ultimately resulting in a library of GROMOS-compatible CG lipid models. The end goal would be to create a repository of CG lipids compatible with AL proteins, existing CG solvent and other molecules.

One major disadvantage of the GROMOS simulation software was the lack of efficient parallelisation for more than 16 processers, resulting in the inability to make best use of the increasingly parallel cluster computing resources available and therefore relatively slow simulation speeds. Additionally, in part due to their speed (although the technical issues mentioned above should serve as a caveat to maximising speed alone), other MD simulation codes such as GROMACS are much more widely used among the biomolecular simulation community. Furthermore, GROMACS and GROMOS programs share common ancestry and were both written at the University of Groningen at inception. Since then GROMOS has been developed at ETH Zurich and GROMACS has development teams all over the world, lead by a development team based in Stockholm. As the GROMOS forcefields are compatible with the GROMACS software, it would therefore be advantageous to adapt these CG lipid models to GROMACS, which is already capable of utilising the GROMOS AL forcefields. While the GROMOS CG water model has been used in GROMACS[197], at present the pressure correction has not been implemented[197] and it requires the use of tabulated potentials, which are not trivial to develop or use. This will therefore be a substantial undertaking, but would greatly improve the uptake and use of these CG models.

4.4 Discussion

Reference simulations of the DPPC bilayer, as well as a single molecule of DPPC in water were successfully performed. The internal structural properties of DPPC were found to be conserved between the bilayer and single molecule simulations, suggesting that these properties of the CG-DPPC could be optimised for a single molecule in solution. Similarly, the distribution of bond lengths was conserved when a truncated methyl-capped phosphocholine moiety was simulated alone in solution, legitimising the building block approach to parameterising CG-DPPC.

The approach taken to parameterise the CG-DPPC model involved several iterations of CG models, starting with a small part of the whole molecule and adding more CG beads with each iteration. An initial proof of principle two-bead CG methyl-capped phosphocholine model (CG-PC) was established to show that two polarisable CG beads could be bonded together. This then led to the parameterisation of the polar CG headgroup part of the lipid (CG-HG), involving the addition of angle and dihedral terms. Finally, the entire CG-DPPC model was created by the addition of non-polarisable CG alkane tails.

This piecemeal parameterisation procedure greatly simplified the process of adding and parameterising new bonded terms in a multi-variable system. For the most part, the existing bonded terms did not have to be changed with addition of further beads. However, the non-bonded terms proved more problematic, as these could not be fully optimised until the later editions of the model, due to the effects of other molecules, not added until the bilayer stage or beads situated more than 3 bonds away

within the same molecule (only one such case present for CG-HG). Here, each addition of new beads changed the behaviour of the model slightly, and, at the point of packing the CG-DPPC molecules into a bilayer, enormously. Thus, unfortunately, the success of models representing a sub-part of the molecule did not necessarily ensure the success of the final model in all environments.

Manual parameter matching as a methodology proved to be a useful, if laborious method for CG parameterisation. It is worth noting here that attempts by Dr Collier to use Iterative Boltzmann Inversion, implemented in VOTKA[194] as an automated alternative, did not provide appropriate estimates of the LJ parameters, which are the most difficult to derive manually. Generally, the distributions of bonded properties calculated from the CG simulations were always sharper than their AL equivalents, and due to the use of a single harmonic function, between two or three beads, to encompass the behaviour described by several such functions, (between multiple atoms at AL, the CG bond and angle distributions always had a single peak. This is simply a consequence of moving from a more detailed to less detailed description.

This non-correspondence derailed attempts to automate the parameterisation of the bonded terms via quantitative comparison of the histograms. If the AL reference simulation produced multi-modal distributions, or a substantial shoulder on an otherwise unimodal distribution, this would result in poor agreement between the CG and reference data and thus driving the parameters away from those values, which could be the best available. While human intuition can easily account for these issues, it proved difficult to automate.

The project has encountered a set of technical challenges, which became apparent from the very onset, but took some time to diagnose due to uninformative error messages and the incomplete CG implementation documentation still under development. When this work was initiated, two types of CG simulations had been run successfully. Firstly, pure CG solvent systems, where groups of molecules were comprised of a single polarisable CG bead, did not have any CG bonded terms other than the centre-dipole particle bond. Secondly, CG alkanes, which were still under development, involved non-polarisable CG beads with a full complement of bonded terms, but were only tested as pure solutions or in the presence of CG water, meaning that bond constraints (SHAKE) could simply be switched off globally and the normal AL bonded terms used. Thus, simulations of polarisable individual CG beads, and bonded non-polarisable CG beads using the standard bonded terms were known to work.

After moving beyond the two-bead CG-PC model, problems with the CG water model were undergoing diagnosis by our collaborators at ETH Zurich, making it necessary to run the CG-HG simulations in SPC water while those were fixed, and later, solvation with CG water also proved problematic. To run the simulations in AL SPC water, it is necessary to use bond constraints on the water molecules, meaning that a completely new set of CG bonded terms had to be added to the code by our collaborators at ETH Zurich.

Furthermore, issues were found with the pressure scaling of CG particles with the GROMOS MD inputs. CG solvents are supra-molecular CG models – e.g. one CG water bead represented five AL water

molecules. This necessitated an adjustment to the pressure calculation during the simulation, which in turn affected the pressure coupling needed to run under NPT conditions. This was initially implemented in a way that affected all CG molecules in the system, so changes to the code were required to allow partitioning of groups of CG molecules between the adjusted and non-adjusted pressure calculations. Again, this had to be carried out by our collaborators at ETH Zurich, who were responsible for the code developments.

In general, when scanning ranges of values of different parameters in parallel, there was no clear correspondence between combinations that resulted in stable simulations (of at least 1 ns) and those that did not. This made further refinement of these parameters, which included the LJ parameters, very difficult to systemise or automate. While this may simply be the result of working with a complex system in which all parameter values are interdependent, very recently, a potential issue associated specifically with the LJ parameters has become apparent. In the first instance, a LJ ϵ value of 1.5 kJ.mol^{-1} was used, the same value as in the GROMOS CG water models. The intention was to refine this value in the later stages of the parameterisation procedure. However, other published CG lipid models, including the polarisable ELBA CG lipids[137], use significantly higher values (2 to 6 kJ.mol^{-1}, in the case of ELBA). While CG models are not directly comparable due to the different design principles (for instance, ELBA uses point dipoles, and other CG lipid models are not polarisable) and parameterisation procedures, the preliminary work on a GROMOS CG-DPPE model showed that larger LJ ϵ look promising, thus in the future it would be useful to further test the GROMOS CG-DPPC model with a wider range of increased LJ ϵ values.

4.5 Methods

4.5.1 AL simulations of reference systems

The AL simulations used the 54A8 atomistic GROMOS forcefield[109] and the SPC water model[69]. All simulations were performed using the GROMOS MD++ program[67,68] and prepared and analysed using the GROMOS++ suite of programs[198].

Initial coordinates for the pre-equilibrated (190 ns) DPPC bilayer were obtained from the ATB[6] online resource. Coordinates for the single DPPC and PCm molecules were extracted from the DPPC bilayer, and the system comprising multiple PCm molecules was built using the GROMOS++ build_box program to distribute exact copies of the input molecule uniformly in a box of size 5.56 × 5.56 × 5.56 nm^3.

The DPPC bilayer and single DPPC and PCm solute molecules were solvated in cubic boxes with a minimum distance to the wall of 1.4 nm. This resulted in the DPPC bilayer becoming surrounded by 5841 water molecules (6.40 × 6.33 × 8.38 nm^3), the single DPPC molecule by 3767 water molecules (4.89 × 4.89 × 4.89 nm^3) and the single PCm molecule by 1361 water molecules (3.48 × 3.48 × 3.48 nm^3 box). The solvent molecules were energy-minimized while the solute molecule structures were positionally

restrained to their starting structure via a harmonic potential energy term with a force constant k_b of 2.5 × 10^4 Nm^{-1}. Initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 60 K. The temperature was then increased stepwise from 60 to 300 K in 2 ps steps of 60 K under NVT conditions. During this heating phase, the solute position restraints were simultaneously relaxed by lowering the harmonic restraint force constant by a factor of 10 at each step. Finally, each system was equilibrated at 298 K under NPT conditions for 0.12 ns. The final coordinates resulting from the energy minimisation, heating and equilibration were the starting point for the MD simulations from which the reference data were calculated.

Simulations were energy minimised for 10,000 steps using steepest descent algorithm. In the NVT and NPT simulations, the temperature was maintained at an approximate temperature of 298 K by weak coupling to a temperature bath with a relaxation time of 0.1 ps[69]. Analogously, for the NPT simulations, the pressure was maintained close to 1.013 bar (1 atm) by weak coupling to a pressure bath with a relaxation time of 0.5 ps and an isothermal compressibility κ_T of 4.575 × 10^{-4} $kJmol^{-1}nm^{-3}$. The leap-frog scheme[70] was used to integrate Newton's equations of motion with a time step of 2 fs. Solute and solvent bond lengths were constrained by the SHAKE algorithm to their ideal lengths[12]. During MD, a twin-range cut-off method was used for the non-bonded interactions, with a short-range cut-off radius of 0.8 nm and an intermediate-range cut-off radius of 1.4 nm. The non-bonded interactions were calculated and updated every 5 integration steps. Solvation effects outside the cut-off sphere were approximated with a reaction field force[71], which was applied using the experimental relative dielectric permittivity of water ε_{rf} of 78.5[199]. The dielectric permittivity within the cut-off sphere ε_{CS} was 1.0 for AL-AL interactions, and was initially estimated to be for 1.5 for CG-CG interactions.

4.5.2 CG simulations

Coordinate generation

An equilibrated box of 2560 CG water beads (equivalent to 12,800 water molecules) was provided by Assist. Prof. S. Riniker. The system was equilibrated at 298 K under NPT conditions for 1 ns. The initial CG bead configuration (coordinates) for CG-PC was obtained by taking the last frame of a simulation of 2560 CG water beads (7.29 × 7.29 × 7.29 nm^3 box) and transforming two adjacent CG water beads into the two CG PCm head group beads.

Initial coordinates for the CG headgroup and single CG DPPC molecule were obtained by taking a randomly selected DPPC molecule from the equilibrated DPPC AL bilayer coordinates and computing the centre of geometry of the group of atoms corresponding to each CG bead. Initial coordinates for the dipole particle were generated by manually adding a dipole particle 0.2 nm away from the central molecule in the y axis manually. CG-DPPC and CG-HG were solvated in SPC water using GROMOS sim_box program.

CG DPPC bilayer coordinates were generated using an in-house bilayer building tool built by Ivan Welsh. The tool took the coordinates of the aforementioned DPPC CG molecule and translated them at regular specified intervals in the xy plane to produce the top leaflet of the bilayer. The leaflet was then mirror in the z direction to produce the complete bilayer, which was then solvated with SPC water.

Molecular dynamics

The CG simulations used the 54A8 atomistic GROMOS forcefield[18] that was edited to include the experimental parameters relating to CG beads. The multiscale simulations used the SPC water model[35], and the pure CG simulations used the GROMOS CG water model37.

Simulations were energy minimised for 10000 steps using steepest descent algorithm. In the NVT and NPT simulations, the temperature was maintained at an approximate temperature of 298 K by weak coupling to a temperature bath with a relaxation time of 0.1 ps69. Analogously, for the NPT simulations, the pressure was maintained close to 1.013 bar (1 atm) by weak coupling to a pressure bath with a relaxation time of 0.5 ps and an isothermal compressibility κT of 4.575 × 10-4 kJmol-1nm-3. The leap-frog scheme70 was used to integrate Newton's equations of motion with a time step of 2 fs. Solvent bond lengths were constrained by the SHAKE algorithm to their ideal lengths12. During MD, a twin-range cut-off method was used for the non-bonded interactions, with an increased (compared to AL) short-range cut-off radius of 1.4 nm and an intermediate-range cut-off radius of 2.0 nm. Extra pressure scaling parameters had to be introduced into the system to compensate for the supra-molecular nature of the CG water beads. Unlike in the case of the CG water model, calculation of the virial and therefore the pressure is unaffected by the inclusion of CG molecules in the simulation (i.e. the scaling factor of 5 for water, 1 for lipids), therefore no need for CG pressure correction was required for the CG lipids themselves.

Initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 60 K. The temperature was then increased stepwise from 60 to 300 K in 2 ps steps of 60 K under NVT conditions. During this heating phase, the solute position restraints were simultaneously relaxed by lowering the harmonic restraint force constant by a factor of 10 at each step. Finally, each system was equilibrated at 298 K under NPT conditions for 0.12 ns. The final coordinates resulting from the energy minimisation, heating and equilibration were the starting point for the MD simulations from which key properties were calculated for comparison to the AL reference simulations and, ultimately to experimental data.

4.5.3 Analysis tools

Distance, angle and torsional dihedral distributions

The GROMOS tser program was used to generate time series of special property values extracted from the molecular dynamics trajectories. The time series data values were then processed into histograms, and the mean values calculated.

Dipole moments

The dipole moments of a group of atoms at each step in the simulation was computed using the GROMOS dipole program and plotted as histograms. The dipole moment can be worked out relative to the origin using the following general equation:

$$p = \sum_{i}^{n} q_i r_i$$

Where:
p is the dipole moment,
n is the number of particles,
q_i is the charge of particle *i*,
r_i is the distance to the origin in a vector format.

To avoid overall molecular motion being incorporated into the calculated dipole moment, each frame of the trajectory was first superimposed onto the initial coordinates.

Radial distribution function

The radial distribution function between groups of atoms corresponding to CG beads, or the CG beads themselves, was calculated by the GROMOS program rdf according to the following function:

$$g(r) = \frac{N_i(k)}{4\pi r^2 dr \rho_i}$$

Where:
r is the distance between the particles or groups of particles
$N_i(k)$ is the number of particles or groups of particles of type *i* found at a distance between *r(k) – 0.5dr* and *r(k) + 0.5dr*
P_j is the number density of particles or groups of particles *i*

4.6 Conclusion

A large extent of parameterisation groundwork was laid down for the CG-DPPC lipid modelling project over the 3 years of this PhD. A working version of CG-DPPC lipid was created for a simulation of a single CG lipid molecule in AL water. The joining of multiple polarisable CG beads containing dipole particles worked for isolated CG molecules in solution, but not for multiple CG molecules packed into a bilayer. The parameterisation of the bonded terms can be considered largely complete, as these were fitted against structural data for the reference bilayer simulation throughout and largely unaffected by each expansion of the model. However, the non-bonded terms required further work. This pertained to both the interactions between CG lipid molecules in the bilayer, and between the CG lipids and CG water. The success of simulations without dipole particles suggested that these are the main cause of the problems, yet dipole particles wer essential to the polarisability of the model, a key point of difference along with the model's compatibility with the GROMOS AL forcefield. The large degree of similarity between DPPC and other common phospholipids meant that once these remaining problems have been solved, generation of other lipid CG models, such as DPPE/POPE and DPPS/POPS should be straightforward, utilising the plug-and-play philosophy that underlies the design of the CG-DPPC model developed here.

References

1. Zwier, M. C.; Chong, L. T., Reaching biological timescales with all-atom molecular dynamics simulations. *Current opinion in pharmacology* **2010,** *10* (6), 745-52.
2. Drenth, J., *Principles of Protein X-Ray Crystallography*. 3 ed.; Springer-Verlag: New York, 2007.
3. Wüthrich, K., *NMR of Proteins and Nucleic Acids*. 1986.
4. K. N. Trueblood, H.-B. B., H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Shmueli, S. C. Abrahams, Atomic displacement parameter nomenclature. Report of a Subcommittee on atomic displacement parameter nomenclature. *Acta Cryst.* **1996,** (A52), 770-781.
5. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M., Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS journal* **2008,** *275* (1), 1-21.
6. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E., An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of chemical theory and computation* **2011,** *7* (12), 4026-37.
7. Kollman, P., Perspective on "Molecular dynamics study of liquid water". In *Theoretical Chemistry Accounts: New Century Issue*, Cramer, C. J.; Truhlar, D. G., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2001; pp 306-307.
8. Noid, W. G., Perspective: Coarse-grained models for biomolecular systems. *J Chem Phys* **2013,** *139* (9), 090901.
9. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B., Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie (International ed. in English)* **2006,** *45* (25), 4064-92.
10. Levitt, M.; Lifson, S., Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of molecular biology* **1969,** *46* (2), 269-79.
11. van Gunsteren, W. F., Billeter, S. R.; Eising, A. A., et al., Biomolecular Simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zurich: Zurich, Groningen, 1996.
12. Horta, B. A.; Fuchs, P. F.; van Gunsteren, W. F.; Hunenberger, P. H., New Interaction Parameters for Oxygen Compounds in the GROMOS Force Field: Improved Pure-Liquid and Solvation Properties for Alcohols, Ethers, Aldehydes, Ketones, Carboxylic Acids, and Esters. *Journal of chemical theory and computation* **2011,** *7* (4), 1016-31.
13. Tien, C. L. a. L., J. H., *Statistical Thermodynamics*. CRC Press: 1979; Vol. Revised Ed.
14. Allen, P. M., Introduction to Molecular Dynamics Simulation. In *Computational Soft Matter: From Synthetic Polymers to Proteins*, John von Neumann Institute for Computing, Julich: 2004; Vol. 23, pp 1-28.
15. Ryckaert, J.-P. C., G.; Berendsen, H. J. C., Numerical-integration of Cartesian equations of motion of a system with contraints: molecular dynamics of n-alkanes. *J. Comput. Phys* **1977,** (23), 327-341.
16. Cuendet, M. A.; Gunsteren, W. F. v., On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of Chemical Physics* **2007,** *127* (18), 184102.
17. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1–2*, 19-25.
18. Tsai, S.-H.; Krech, M.; Landau, D. P., Symplectic integration methods in molecular and spin dynamics simulations. *Brazilian Journal of Physics* **2004,** *34*, 384-391.

19. Frenkel, D. S., B., *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications*. 2nd ed.; Academic Press: 2002.
20. M.J. Abraham, D. v. d. S., E. Lindahl, B. Hess and the GROMACS; team, d., GROMACS User Manual version 5.1.2. www.gromacs.org, 2016.
21. Spellberg, B., Guidos, Robert, Gilbert, David, Bradley, John, Boucher, Helen W., Scheld, W. Michael, Bartlett, John G., Edwards, Jr ,John, The Epidemic of Antibiotic-Resistant Infections: A Call to Action for the Medical Community from the Infectious Diseases Society of America. *Clinical Infectious Diseases* **2008,** *46* (2).
22. Levy, S. B., Factors impacting on the problem of antibiotic resistance. *Journal of Antimicrobial Chemotherapy* **2002,** *49* (1).
23. Ventola, C. L., The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics* **2015,** *40* (4), 277-283.
24. Cornforth, D. M.; Foster, K. R., Antibiotics and the art of bacterial war. *Proceedings of the National Academy of Sciences* **2015,** *112* (35), 10827-10828.
25. Strieker, M.; Marahiel, M. A., The structural diversity of acidic lipopeptide antibiotics. *Chembiochem : a European journal of chemical biology* **2009,** *10* (4), 607-16.
26. Qian, C. D.; Wu, X. C.; Teng, Y.; Zhao, W. P.; Li, O.; Fang, S. G.; Huang, Z. H.; Gao, H. C., Battacin (Octapeptin B5), a new cyclic lipopeptide antibiotic from Paenibacillus tianmuensis active against multidrug-resistant Gram-negative bacteria. *Antimicrob Agents Chemother* **2012,** *56* (3), 1458-65.
27. Danner, R. L.; Joiner, K. A.; Rubin, M.; Patterson, W. H.; Johnson, N.; Ayers, K. M.; Parrillo, J. E., Purification, toxicity, and antiendotoxin activity of polymyxin B nonapeptide. *Antimicrobial Agents and Chemotherapy* **1989,** *33* (9), 1428-1434.
28. Benedict, R. G.; Langlykke, A. F., Antibiotic activity of Bacillus polymyxa. *Journal of bacteriology* **1947,** *54* (1), 24.
29. Kirkpatrick, P.; Raja, A.; LaBonte, J.; Lebbos, J., Daptomycin. *Nature reviews. Drug discovery* **2003,** *2* (12), 943-4.
30. Pablo S. Corona Pérez-Cardona, V. B. O., Dolors Rodriguez Pardo, Carlos Pigrau Serrallach, Ernesto Guerra Farfán, Carles Amat Mateu, Xavier Flores Sanchez, Clinical experience with daptomycin for the treatment of patients with knee and hip periprosthetic joint infections. *J Antimicrob Chemother* **2012,** *67* (7), 1749-1754.
31. Zavascki, A. P.; Goldani, L. Z.; Li, J.; Nation, R. L., Polymyxin B for the treatment of multidrug-resistant pathogens: a critical review. *J Antimicrob Chemother* **2007,** *60* (6), 1206-15.
32. Falagas, M. E.; Kasiakou, S. K., Toxicity of polymyxins: a systematic review of the evidence from old and recent studies. *Critical care (London, England)* **2006,** *10* (1), R27.
33. De Zoysa, G. H.; Cameron, A. J.; Hegde, V. V.; Raghothama, S.; Sarojini, V., Antimicrobial peptides with potential for biofilm eradication: synthesis and structure activity relationship studies of battacin peptides. *Journal of medicinal chemistry* **2015,** *58* (2), 625-39.
34. Meyers, E. P., W. L.; Brown, W. E.; Linnett, P.; Stroming, J. L., Em49 - new polypeptide antibiotic active against cell-membranes. *Ann. N. Y. Acad. Sci.* **1974,** *235*, 493−501.
35. Storm, D. R.; Rosenthal, K. S.; Swanson, P. E., Polymyxin and related peptide antibiotics. *Annual review of biochemistry* **1977,** *46*, 723-63.
36. Tsubery, H.; Ofek, I.; Cohen, S.; Fridkin, M., N-terminal modifications of Polymyxin B nonapeptide and their effect on antibacterial activity. *Peptides* **2001,** *22* (10), 1675-81.
37. Tsubery, H.; Ofek, I.; Cohen, S.; Fridkin, M., Structure-function studies of polymyxin B nonapeptide: implications to sensitization of gram-negative bacteria. *Journal of medicinal chemistry* **2000,** *43* (16), 3085-92.
38. Nation, R. L.; Velkov, T.; Li, J., Colistin and Polymyxin B: Peas in a Pod, or Chalk and Cheese? *Clinical Infectious Diseases* **2014,** *59* (1), 88-94.

39. Velkov, T.; Thompson, P. E.; Nation, R. L.; Li, J., Structure--activity relationships of polymyxin antibiotics. *Journal of medicinal chemistry* **2010,** *53* (5), 1898-916.
40. Velkov, T.; Roberts, K. D.; Nation, R. L.; Thompson, P. E.; Li, J., Pharmacology of polymyxins: new insights into an 'old' class of antibiotics. *Future microbiology* **2013,** *8* (6), 10.2217/fmb.13.39.
41. Hancock, R. E.; Chapple, D. S., Peptide antibiotics. *Antimicrob Agents Chemother* **1999,** *43* (6), 1317-23.
42. Berglund, N. A.; Piggot, T. J.; Jefferies, D.; Sessions, R. B.; Bond, P. J.; Khalid, S., Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of E. coli: A Molecular Dynamics Study. *PLoS computational biology* **2015,** *11* (4), e1004180.
43. Soon, R. L.; Velkov, T.; Chiu, F.; Thompson, P. E.; Kancharla, R.; Roberts, K.; Larson, I.; Nation, R. L.; Li, J., Design, synthesis, and evaluation of a new fluorescent probe for measuring polymyxin-lipopolysaccharide binding interactions. *Analytical biochemistry* **2011,** *409* (2), 273-83.
44. Mogi, T.; Murase, Y.; Mori, M.; Shiomi, K.; Omura, S.; Paranagama, M. P.; Kita, K., Polymyxin B identified as an inhibitor of alternative NADH dehydrogenase and malate: quinone oxidoreductase from the Gram-positive bacterium Mycobacterium smegmatis. *Journal of biochemistry* **2009,** *146* (4), 491-9.
45. Sampson, T. R.; Liu, X.; Schroeder, M. R.; Kraft, C. S.; Burd, E. M.; Weiss, D. S., Rapid killing of Acinetobacter baumannii by polymyxins is mediated by a hydroxyl radical death pathway. *Antimicrob Agents Chemother* **2012,** *56* (11), 5642-9.
46. Cochrane, S. A.; Lohans, C. T.; Brandelli, J. R.; Mulvey, G.; Armstrong, G. D.; Vederas, J. C., Synthesis and structure-activity relationship studies of N-terminal analogues of the antimicrobial peptide tridecaptin A(1). *Journal of medicinal chemistry* **2014,** *57* (3), 1127-31.
47. Zhang, L.; Dhillon, P.; Yan, H.; Farmer, S.; Hancock, R. E., Interactions of bacterial cationic peptide antibiotics with outer and cytoplasmic membranes of Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **2000,** *44* (12), 3317-21.
48. Clausell, A.; Garcia-Subirats, M.; Pujol, M.; Busquets, M. A.; Rabanal, F.; Cajal, Y., Gram-Negative Outer and Inner Membrane Models: Insertion of Cyclic Cationic Lipopeptides. *The Journal of Physical Chemistry B* **2007,** *111* (3), 551-563.
49. Scortichini, M.; Marcelletti, S.; Ferrante, P.; Petriccione, M.; Firrao, G., Pseudomonas syringae pv. actinidiae: a re-emerging, multi-faceted, pandemic pathogen. *Molecular plant pathology* **2012,** *13* (7), 631-40.
50. Renzi, M.; Copini, P.; Taddei, A. R.; Rossetti, A.; Gallipoli, L.; Mazzaglia, A.; Balestra, G. M., Bacterial Canker on Kiwifruit in Italy: Anatomical Changes in the Wood and in the Primary Infection Sites. *Phytopathology* **2012,** *102* (9), 827-840.
51. Nicoletti, I.; Migliorati, G.; Pagliacci, M. C.; Grignani, F.; Riccardi, C., A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry. *Journal of immunological methods* **1991,** *139* (2), 271-9.
52. Cronan, J. E., Bacterial membrane lipids: where do we stand? *Annual review of microbiology* **2003,** *57*, 203-24.
53. Sohlenkamp, C. a. G., O., Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiology Reviews* **2015**.
54. Silhavy, T. J.; Kahne, D.; Walker, S., The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology* **2010,** *2* (5), a000414.
55. Raetz, C. R., Enzymology, genetics, and regulation of membrane phospholipid synthesis in Escherichia coli. *Microbiological Reviews* **1978,** *42* (3), 614-659.
56. J., O. M., Structure and Biosynthesis of the Bacterial Cell Wall. *Annual review of biochemistry* **1969,** *38*, 1-1034.

57. Qureshi, N.; Honovich, J. P.; Hara, H.; Cotter, R. J.; Takayama, K., Location of fatty acids in lipid A obtained from lipopolysaccharide of Rhodopseudomonas sphaeroides ATCC 17023. *The Journal of biological chemistry* **1988,** *263* (12), 5502-4.
58. Raetz, C. R., Biochemistry of endotoxins. *Annual review of biochemistry* **1990,** *59*, 129-70.
59. Osborn, M. J.; Rick, P. D.; Lehmann, V.; Rupprecht, E.; Singh, M., Structure and biogenesis of the cell envelope of gram-negative bacteria. *Annals of the New York Academy of Sciences* **1974,** *235* (0), 52-65.
60. Barák, I.; Muchová, K., The Role of Lipid Domains in Bacterial Cell Processes. *International Journal of Molecular Sciences* **2013,** *14* (2), 4050-4065.
61. Piggot, T. J.; Holdbrook, D. A.; Khalid, S., Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. *The journal of physical chemistry. B* **2011,** *115* (45), 13381-8.
62. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *Journal of computational chemistry* **2005,** *26* (16), 1701-18.
63. S. Páll, M. J. A., C. Kutzner, B. Hess, E. Lindahl, Tackling Exascale Software Challenges in Molecular Dynamics Simulations with

GROMACS. *S. Markidis & E. Laure (Eds.), Solving Software Challenges for Exascale* **2015,** (8759), 3-27.
64. S. Pronk, S. P., R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29*, 845-54.
65. 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* **2008,** *4* (3), 435-447.
66. Lindahl, E.; Hess, B.; van der Spoel, D., GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual* **2001,** *7* (8), 306-317.
67. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995,** *91* (1), 43-56.
68. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal : EBJ* **2011,** *40* (7), 843-56.
69. Berendsen, H. J. C., Postma, J. P., von Gunsteren W. P. and Hermans J., *Intermolecular Forces*. Reidel: 1981.
70. Berweger, C. D. v. G., W.F and Müller-Plathe, F, Force field parametrization by weak coupling. Re-engineering SPC water. *Chemical Physics Letters* **1995,** *232* (5-6), 429-436.
71. Hess, B. B., H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J., LINCS: A Linear Constraint Solver for Molecular Simulations. *Comput. Chem.* **1997,** *18* (12), 1463-1472.
72. H. J. C. Berendsen, J. P. M. P., W. F. van Gunsteren, A. DiNola, and and J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (3684).
73. Frishman, D.; Argos, P., Knowledge-based protein secondary structure assignment. *Proteins* **1995,** *23* (4), 566-79.
74. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22* (12), 2577-637.
75. Humphrey, W., Dalke, A. and Schulten, K.,, VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996,** *14*, 33-38.
76. Gilbert, J. A.; Davies, P. L.; Laybourn-Parry, J., A hyperactive, Ca2+-dependent antifreeze protein in an Antarctic bacterium. *FEMS microbiology letters* **2005,** *245* (1), 67-72.

77. Gilbert, J. A.; Hill, P. J.; Dodd, C. E.; Laybourn-Parry, J., Demonstration of antifreeze protein activity in Antarctic lake bacteria. *Microbiology (Reading, England)* **2004,** *150* (Pt 1), 171-80.
78. Hashim, N. H. F.; Bharudin, I.; Nguong, D. L. S.; Higa, S.; Bakar, F. D. A.; Nathan, S.; Rabu, A.; Kawahara, H.; Illias, R. M.; Najimudin, N.; Mahadi, N. M.; Murad, A. M. A., Characterization of Afp1, an antifreeze protein from the psychrophilic yeast Glaciozyma antarctica PI12. *Extremophiles* **2013,** *17* (1), 63-73.
79. Doucet, C. J.; Byass, L.; Elias, L.; Worrall, D.; Smallwood, M.; Bowles, D. J., Distribution and characterization of recrystallization inhibitor activity in plant and lichen species from the UK and maritime Antarctic. *Cryobiology* **2000,** *40* (3), 218-27.
80. Duman, J. G., Purification and characterization of a thermal hysteresis protein from a plant, the bittersweet nightshade Solanum dulcamara. *Biochim Biophys Acta* **1994,** *1206* (1), 129-35.
81. Antikainen, M.; Griffith, M.; Zhang, J.; Hon, W. C.; Yang, D. S. C.; Pihakaski-Maunsbach, K., Immunolocalization of Antifreeze Proteins in Winter Rye Leaves, Crowns, and Roots by Tissue Printing. *Plant Physiology* **1996,** *110* (3), 845-857.
82. Harding, M. M.; Anderberg, P. I.; Haymet, A. D., 'Antifreeze' glycoproteins from polar fish. *European journal of biochemistry* **2003,** *270* (7), 1381-92.
83. Sicheri, F.; Yang, D. S., Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature* **1995,** *375* (6530), 427-31.
84. Yeh, Y.; Feeney, R. E., Antifreeze Proteins: Structures and Mechanisms of Function. *Chemical reviews* **1996,** *96* (2), 601-618.
85. Yamazaki, T.; Kawamura, Y.; Uemura, M., Extracellular freezing-induced mechanical stress and surface area regulation on the plasma membrane in cold-acclimated plant cells. *Plant Signaling & Behavior* **2009,** *4* (3), 231-233.
86. Cheng, A.; Merz, K. M., Ice-binding mechanism of winter flounder antifreeze proteins. *Biophysical Journal* **1997,** *73* (6), 2851-2873.
87. Zhang, W.; Laursen, R. A., Structure-function relationships in a type I antifreeze polypeptide. The role of threonine methyl and hydroxyl groups in antifreeze activity. *The Journal of biological chemistry* **1998,** *273* (52), 34806-12.
88. Graether, S. P.; Slupsky, C. M.; Davies, P. L.; Sykes, B. D., Structure of type I antifreeze protein and mutants in supercooled water. *Biophys J* **2001,** *81* (3), 1677-83.
89. Hong, J. S.; Jung, D. H.; Jhon, M. S., Molecular Dynamics Study on Winter Flounder Antifreeze Protein and its Binding Mechanism. *Molecular simulation* **1998,** *20* (5), 303-314.
90. Gauthier, S. Y.; Marshall, C. B.; Fletcher, G. L.; Davies, P. L., Hyperactive antifreeze protein in flounder species. The sole freeze protectant in American plaice. *The FEBS journal* **2005,** *272* (17), 4439-49.
91. Marshall, C. B.; Fletcher, G. L.; Davies, P. L., Hyperactive antifreeze protein in a fish. *Nature* **2004,** *429* (6988), 153.
92. Marshall, C. B.; Chakrabartty, A.; Davies, P. L., Hyperactive antifreeze protein from winter flounder is a very long rod-like dimer of alpha-helices. *The Journal of biological chemistry* **2005,** *280* (18), 17920-9.
93. Yu, X.-M.; Griffith, M., Antifreeze Proteins in Winter Rye Leaves Form Oligomeric Complexes. *Plant Physiology* **1999,** *119* (4), 1361-1370.
94. Tomczak, M. M.; Vigh, L.; Meyer, J. D.; Manning, M. C.; Hincha, D. K.; Crowe, J. H., Lipid unsaturation determines the interaction of AFP type I with model membranes during thermotropic phase transitions. *Cryobiology* **2002,** *45* (2), 135-42.
95. Tomczak, M. M.; Hincha, D. K.; Estrada, S. D.; Feeney, R. E.; Crowe, J. H., Antifreeze proteins differentially affect model membranes during freezing. *Biochim Biophys Acta* **2001,** *1511* (2), 255-63.

96. Tomczak, M. M.; Hincha, D. K.; Estrada, S. D.; Wolkers, W. F.; Crowe, L. M.; Feeney, R. E.; Tablin, F.; Crowe, J. H., A mechanism for stabilization of membranes at low temperatures by an antifreeze protein. *Biophysical Journal* **2002,** *82* (2), 874-881.
97. Tomczak, M. M.; Hincha, D. K.; Crowe, J. H.; Harding, M. M.; Haymet, A. D., The effect of hydrophobic analogues of the type I winter flounder antifreeze protein on lipid bilayers. *FEBS letters* **2003,** *551* (1-3), 13-9.
98. Kar, R. K.; Mroue, K. H.; Kumar, D.; Tejo, B. A.; Bhunia, A., Structure and Dynamics of Antifreeze Protein–Model Membrane Interactions: A Combined Spectroscopic and Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2016,** *120* (5), 902-914.
99. D'Amico, S.; Collins, T.; Marx, J.-C.; Feller, G.; Gerday, C., Psychrophilic microorganisms: challenges for life. *EMBO Reports* **2006,** *7* (4), 385-389.
100. Chintalapati, S.; Kiran, M. D.; Shivaji, S., Role of membrane lipid fatty acids in cold adaptation. *Cellular and molecular biology (Noisy-le-Grand, France)* **2004,** *50* (5), 631-42.
101. Russell, N. J., Psychrophilic bacteria--molecular adaptations of membrane lipids. *Comparative biochemistry and physiology. Part A, Physiology* **1997,** *118* (3), 489-93.
102. Viarengo A., A. R., Roma G. and Orunesu M., Differences in lipid composition of cell membranes from Antarctic and Mediterranean scallops. *Comparative biochemistry and physiology. B, Comparative biochemistry* **1994,** *109(4)*, 576-584.
103. Hays, L. M. Antifreeze glycoproteins from antarctic fish: their structure and function in regards to lipid thermotropic phase transitions. University of California, 1998.
104. Hays, L. M.; Crowe, J. H.; Wolkers, W.; Rudenko, S., Factors affecting leakage of trapped solutes from phospholipid vesicles during thermotropic phase transitions. *Cryobiology* **2001,** *42* (2), 88-102.
105. Hays, L. M.; Feeney, R. E.; Crowe, L. M.; Crowe, J. H.; Oliver, A. E., Antifreeze glycoproteins inhibit leakage from liposomes during thermotropic phase transitions. *Proceedings of the National Academy of Sciences of the United States of America* **1996,** *93* (13), 6835-6840.
106. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* (1), 235-242.
107. Yang, D. S. C., Sicheri, F., WINTER FLOUNDER ANTIFREEZE PROTEIN ISOFORM HPLC6 AT 4 DEGREES C. RSCB PDB, 1995.
108. Zielkiewicz, J., Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *J Chem Phys* **2005,** *123* (10), 104501.
109. Reif, M. M.; Winger, M.; Oostenbrink, C., Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. *Journal of chemical theory and computation* **2013,** *9* (2), 1247-1264.
110. Baczynski, K.; Markiewicz, M.; Pasenkiewicz-Gierula, M., A computer model of a polyunsaturated monogalactolipid bilayer. *Biochimie* **2015,** *118*, 129-40.
111. Silvius, J. R., *Thermotropic Phase Transitions of Pure Lipids in Model Membranes and Their Modifications by Membrane Proteins*. John Wiley & Sons: New York, 1982.
112. Avanti Polar Lipids, I. Phase Transition Temperatures for Glycerophospholipids. https://avantilipids.com/tech-support/physical-properties/phase-transition-temps/ (accessed 26th Dec 2016).
113. Poger, D.; Mark, A. E., On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *Journal of chemical theory and computation* **2010,** *6* (1), 325-336.
114. Domański, J.; Stansfeld, P. J.; Sansom, M. S. P.; Beckstein, O., Lipidbook: A Public Repository for Force-Field Parameters Used in Membrane Simulations. *The Journal of Membrane Biology* **2010,** *236* (3), 255-258.

115. Bussi, G., Hamiltonian replica exchange in GROMACS: a flexible implementation. *Molecular Physics* **2014,** *112* (3-4), 379-384.
116. Terakawa, T.; Kameda, T.; Takada, S., On easy implementation of a variant of the replica exchange with solute tempering in GROMACS. *Journal of computational chemistry* **2011,** *32* (7), 1228-1234.
117. Wang, Y.; Chen, C. H.; Hu, D.; Ulmschneider, M. B.; Ulmschneider, J. P., Spontaneous formation of structurally diverse membrane channel architectures from a single antimicrobial peptide. **2016,** *7*, 13535.
118. Poger, D.; Van Gunsteren, W. F.; Mark, A. E., A new force field for simulating phosphatidylcholine bilayers. *Journal of computational chemistry* **2010,** *31* (6), 1117-25.
119. Riniker, S.; van Gunsteren, W. F., A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. *The Journal of Chemical Physics* **2011,** *134* (8), 084110.
120. Eichenberger, A. P.; Huang, W.; Riniker, S.; van Gunsteren, W. F., Supra-Atomic Coarse-Grained GROMOS Force Field for Aliphatic Hydrocarbons in the Liquid Phase. *Journal of chemical theory and computation* **2015,** *11* (7), 2925-37.
121. Riniker, S.; Allison, J. R.; van Gunsteren, W. F., On developing coarse-grained models for biomolecular simulation: a review. *Physical chemistry chemical physics : PCCP* **2012,** *14* (36), 12423-30.
122. Levitt, M.; Warshel, A., Computer simulation of protein folding. *Nature* **1975,** *253* (5494), 694-8.
123. Ingolfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J., The power of coarse graining in biomolecular simulations. *Wiley interdisciplinary reviews. Computational molecular science* **2014,** *4* (3), 225-248.
124. Hadley, K. R.; McCabe, C., Coarse-Grained Molecular Models of Water: A Review. *Molecular simulation* **2012,** *38* (8-9), 671-681.
125. Darré, L.; Machado, M. R.; Pantano, S., Coarse-grained models of water. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012,** *2* (6), 921-930.
126. Izvekov, S.; Voth, G. A., A multiscale coarse-graining method for biomolecular systems. *The journal of physical chemistry. B* **2005,** *109* (7), 2469-73.
127. Król, M., Comparison of various implicit solvent models in molecular dynamics simulations of immunoglobulin G light chain dimer. *Journal of computational chemistry* **2003,** *24* (5), 531-546.
128. Kleinjung, J.; Fraternali, F., Design and application of implicit solvent models in biomolecular simulations. *Current opinion in structural biology* **2014,** *25* (100), 126-134.
129. 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. *The journal of physical chemistry. B* **2007,** *111* (27), 7812-24.
130. Chiu, S. W.; Scott, H. L.; Jakobsson, E., A Coarse-Grained Model Based on Morse Potential for Water and n-Alkanes. *Journal of chemical theory and computation* **2010,** *6* (3), 851-63.
131. Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J., Polarizable water model for the coarse-grained MARTINI force field. *PLoS computational biology* **2010,** *6* (6), e1000810.
132. Liu, Y.; Ichiye, T., Soft Sticky Dipole Potential for Liquid Water: A New Model. *The Journal of Physical Chemistry* **1996,** *100* (7), 2723-2730.
133. Tan, M.-L.; Fischer, J. T.; Chandra, A.; Brooks, B. R.; Ichiye, T., A temperature of maximum density in soft sticky dipole water. *Chemical Physics Letters* **2003,** *376* (5–6), 646-652.
134. Renevey, A.; Riniker, S., Improved accuracy of hybrid atomistic/coarse-grained simulations using reparametrised interactions. *J Chem Phys* **2017,** *146* (12), 124131.
135. Allison, J. R.; Riniker, S.; van Gunsteren, W. F., Coarse-grained models for the solvents dimethyl sulfoxide, chloroform, and methanol. *J Chem Phys* **2012,** *136* (5), 054505.
136. Orsi, M., Comparative assessment of the ELBA coarse-grained model for water. *Molecular Physics* **2014,** *112* (11), 1566-1576.

137. Orsi, M.; Essex, J. W., The ELBA force field for coarse-grain modeling of lipid membranes. *PloS one* **2011,** *6* (12), e28637.
138. Mourits F. M. and Rummens, F. H. A., A critical evaluation of Lennard-Jones and Stockmayer potential parameters and of some correlation methods. *Can. J. Chem.* **1976,** *55*.
139. Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S., Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *Journal of chemical theory and computation* **2010,** *6* (12), 3793-3807.
140. Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D., Enzyme dynamics during catalysis. *Science (New York, N.Y.)* **2002,** *295* (5559), 1520-3.
141. Lodish H, B. A., Zipursky SL, et al., *Molecular Cell Biology. 4th edition.* New York, 2000.
142. Lee, A. G., Lipid–protein interactions in biological membranes: a structural perspective. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2003,** *1612* (1), 1-40.
143. Alberts B, J. A., Lewis J, et al., *Molecular Biology of the Cell. 4th edition.* Garland Science: New York, 2002.
144. Hirtz, M.; Kumar, N.; Chi, L., Simulation modeling of supported lipid membranes - a review. *Current topics in medicinal chemistry* **2014,** *14* (5), 617-23.
145. Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V., Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014,** *510* (7503), 172-175.
146. Bandyopadhyay, A.; McCarthy, K. A.; Kelly, M. A.; Gao, J., Targeting bacteria via iminoboronate chemistry of amine-presenting lipids. *Nature Communications* **2015,** *6*, 6561.
147. Cochrane, S. A.; Findlay, B.; Bakhtiary, A.; Acedo, J. Z.; Rodriguez-Lopez, E. M.; Mercier, P.; Vederas, J. C., Antimicrobial lipopeptide tridecaptin A1 selectively binds to Gram-negative lipid II. *Proceedings of the National Academy of Sciences* **2016,** *113* (41), 11561-11566.
148. Mazhab-Jafari, M. T.; Rohou, A.; Schmidt, C.; Bueler, S. A.; Benlekbir, S.; Robinson, C. V.; Rubinstein, J. L., Atomic model for the membrane-embedded VO motor of a eukaryotic V-ATPase. *Nature* **2016,** *539* (7627), 118-122.
149. Barrera, N. P.; Zhou, M.; Robinson, C. V., The role of lipids in defining membrane protein interactions: insights from mass spectrometry. *Trends in Cell Biology* **2013,** *23* (1), 1-8.
150. Poger, D.; Mark, A. E., Lipid Bilayers: The Effect of Force Field on Ordering and Dynamics. *Journal of chemical theory and computation* **2012,** *8* (11), 4807-17.
151. Horta, B. A. C.; de Vries, A. H.; Hünenberger, P. H., Simulating the Transition between Gel and Liquid-Crystal Phases of Lipid Bilayers: Dependence of the Transition Temperature on the Hydration Level. *Journal of chemical theory and computation* **2010,** *6* (8), 2488-2500.
152. Orsi, M.; Haubertin, D. Y.; Sanderson, W. E.; Essex, J. W., A quantitative coarse-grain model for lipid bilayers. *The journal of physical chemistry. B* **2008,** *112* (3), 802-15.
153. Orsi, M.; Sanderson, W. E.; Essex, J. W., Permeability of small molecules through a lipid bilayer: a multiscale simulation study. *The journal of physical chemistry. B* **2009,** *113* (35), 12019-29.
154. Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L., A Coarse Grain Model for Phospholipid Simulations. *The Journal of Physical Chemistry B* **2001,** *105* (19), 4464-4470.
155. Shinoda, W.; DeVane, R.; Klein, M. L., Zwitterionic lipid assemblies: molecular dynamics studies of monolayers, bilayers, and vesicles using a new coarse grain force field. *The journal of physical chemistry. B* **2010,** *114* (20), 6836-49.
156. Marrink, S. J.; Tieleman, D. P., Perspective on the Martini model. *Chemical Society Reviews* **2013,** *42* (16), 6801-6822.
157. Marrink, S. J.; de Vries, A. H.; Mark, A. E., Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B* **2004,** *108* (2), 750-760.

158. Lee, H.; de Vries, A. H.; Marrink, S. J.; Pastor, R. W., A coarse-grained model for polyethylene oxide and polyethylene glycol: conformation and hydrodynamics. *The journal of physical chemistry. B* **2009,** *113* (40), 13186-94.
159. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of chemical theory and computation* **2008,** *4* (5), 819-34.
160. Marrink, S. J.; Mark, A. E., Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *Journal of the American Chemical Society* **2003,** *125* (49), 15233-42.
161. Knecht, V.; Marrink, S.-J., Molecular Dynamics Simulations of Lipid Vesicle Fusion in Atomic Detail. *Biophysical Journal* **2007,** *92* (12), 4254-4261.
162. Wong-Ekkabut, J.; Baoukina, S.; Triampo, W.; Tang, I. M.; Tieleman, D. P.; Monticelli, L., Computer simulation study of fullerene translocation through lipid membranes. *Nature nanotechnology* **2008,** *3* (6), 363-8.
163. Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S., Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *Journal of structural biology* **2007,** *157* (3), 593-605.
164. Bond, P. J.; Sansom, M. S., Insertion and assembly of membrane proteins via simulation. *Journal of the American Chemical Society* **2006,** *128* (8), 2697-704.
165. Shih, A. Y.; Freddolino, P. L.; Arkhipov, A.; Schulten, K., Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *Journal of structural biology* **2007,** *157* (3), 579-92.
166. Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K., A coarse grained protein-lipid model with application to lipoprotein particles. *The journal of physical chemistry. B* **2006,** *110* (8), 3674-3684.
167. Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **2004,** *25* (13), 1656-76.
168. Bradley, R.; Radhakrishnan, R., Coarse-Grained Models for Protein-Cell Membrane Interactions. *Polymers* **2013,** *5* (3), 890.
169. Jusufi, A.; DeVane, R. H.; Shinoda, W.; Klein, M. L., Nanoscale carbon particles and the stability of lipid bilayers. *Soft Matter* **2011,** *7* (3), 1139-1146.
170. Ayton, G. S.; Voth, G. A., Hybrid Coarse-Graining Approach for Lipid Bilayers at Large Length and Time Scales. *The Journal of Physical Chemistry B* **2009,** *113* (13), 4413-4424.
171. Srivastava, A.; Voth, G. A., Hybrid Approach for Highly Coarse-Grained Lipid Bilayer Models. *Journal of chemical theory and computation* **2013,** *9* (1), 750-765.
172. Ayton, G. S.; McWhirter, J. L.; Voth, G. A., A second generation mesoscopic lipid bilayer model: connections to field-theory descriptions of membranes and nonlocal hydrodynamics. *J Chem Phys* **2006,** *124* (6), 64906.
173. Berne, J. G. G. a. B. J., Modification of the overlap potential to mimic a linear site–site potential. *The Journal of Chemical Physics* **1981,** *74* (6), 3316-3319.
174. Peter, C., Kremer K., Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back. *Soft Matter* **2009,** *5* (22), 4357-4366.
175. Pandit, S. A.; Scott, H. L., Multiscale simulations of heterogeneous model membranes. *Biochimica et biophysica acta* **2009,** *1788* (1), 136-148.
176. Saunders, M. G.; Voth, G. A., Coarse-graining of multiprotein assemblies. *Current opinion in structural biology* **2012,** *22* (2), 144-50.
177. Davtyan, A.; Simunovic, M.; Voth, G. A., Multiscale simulations of protein-facilitated membrane remodeling. *Journal of structural biology* **2016,** *196* (1), 57-63.

178. Rzepiela, A. J.; Schafer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J., Reconstruction of atomistic details from coarse-grained structures. *Journal of computational chemistry* **2010,** *31* (6), 1333-43.
179. Samiotakis, A.; Homouz, D.; Cheung, M. S., Multiscale investigation of chemical interference in proteins. *J Chem Phys* **2010,** *132* (17), 175101.
180. Heyden, A.; Truhlar, D. G., Conservative Algorithm for an Adaptive Change of Resolution in Mixed Atomistic/Coarse-Grained Multiscale Simulations. *Journal of chemical theory and computation* **2008,** *4* (2), 217-21.
181. Masella, M.; Borgis, D.; Cuniasse, P., Combining a polarizable force-field and a coarse-grained polarizable solvent model. II. Accounting for hydrophobic effects. *Journal of computational chemistry* **2011,** *32* (12), 2664-78.
182. Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J., Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Physical chemistry chemical physics : PCCP* **2011,** *13* (22), 10437-48.
183. Orsi, M.; Ding, W.; Palaiokostas, M., Direct Mixing of Atomistic Solutes and Coarse-Grained Water. *Journal of chemical theory and computation* **2014,** *10* (10), 4684-4693.
184. Shkurti, A., Orsi, Mario, Macii, Enrico, Ficarra, Elisa, Acquaviva, Andrea, Acceleration of coarse grain molecular dynamics on GPU architectures. *J. Comput. Chem.* **2012,** *34* (10).
185. Ayton, G. S.; Voth, G. A., Multiscale simulation of transmembrane proteins. *Journal of structural biology* **2007,** *157* (3), 570-578.
186. Ayton, G. S.; Noid, W. G.; Voth, G. A., Multiscale modeling of biomolecular systems: in serial and in parallel. *Current opinion in structural biology* **2007,** *17* (2), 192-198.
187. Ayton, G. S.; Voth, G. A., Systematic multiscale simulation of membrane protein systems. *Current opinion in structural biology* **2009,** *19* (2), 138-44.
188. Lyman, E.; Pfaendtner, J.; Voth, G. A., Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys J* **2008,** *95* (9), 4183-92.
189. Riniker, S.; van Gunsteren, W. F., Mixing coarse-grained and fine-grained water in molecular dynamics simulations of a single system. *J Chem Phys* **2012,** *137* (4), 044120.
190. Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F., Solvating atomic level fine-grained proteins in supra-molecular level coarse-grained water for molecular dynamics simulations. *European biophysics journal : EBJ* **2012,** *41* (8), 647-61.
191. Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F., Structural effects of an atomic-level layer of water molecules around proteins solvated in supra-molecular coarse-grained water. *The journal of physical chemistry. B* **2012,** *116* (30), 8873-9.
192. Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H., A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *Journal of chemical theory and computation* **2016,** *12* (8), 3825-3850.
193. Eichenberger, A. P. Molecular dynamics simulation of alkanes and proteins: methodology, prediction of properties and comparison to experimental data. ETH Zurich, 2012.
194. Rühle V, J. C., Lukyanov A, Kremer K, Andrienko D., Versatile object-oriented toolkit for coarse-graining applications. *Journal of chemical theory and computation* **2009,** *5*.
195. Rühle V, J. C., Hybrid Approaches to Coarse-Graining using the VOTCA Package: Liquid Hexane. . *Macromolecular Theory and Simulations* **2011,** *20*.
196. Sabine Reißer, M. S., David Poger and Alan E. Mark, The Real Cost of Speed: The Effect of a Time-Saving Multiple-Time-Stepping Algorithm on the Accuracy of Molecular Dynamics Simulations. *Submitted* **2017**.

197. Kuhn, A. B.; Gopal, S. M.; Schäfer, L. V., On Using Atomistic Solvent Layers in Hybrid All-Atom/Coarse-Grained Molecular Dynamics Simulations. *Journal of chemical theory and computation* **2015,** *11* (9), 4460-4472.
198. Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D.; van Gunsteren, W. F., GROMOS++ Software for the Analysis of Biomolecular Simulation Trajectories. *Journal of chemical theory and computation* **2011,** *7* (10), 3379-90.
199. Lide, D. R., *Handbook of chemistry and physics*. CRC Press/Taylor and Francis, Boca Raton: 2007-2008; Vol. 88th ed.

Appendix 1: Octapeptide O17 – GROMOS 54a7 forcefield parameters and starting configuration

2.1 GROMOS 54a7 forcefield parameters

```
[moleculetype ]
; Name   nrexcl
octopep    3
[ atoms ]
;  nr  type  resnr  resid  atom  cgnr  charge   mass         total_charge
    1   CH3    1    DDAB     C6    1    0.00000  15.0350
    2   CH3    1    DDAB     C7    1    0.00000  15.0350
    3   CH1    1    DDAB     C5    1    0.00000  14.0270
    4   CH2    1    DDAB     C4    1    0.00000  14.0270
    5   CH2    1    DDAB     C3    1    0.00000  14.0270
    6   CH2    1    DDAB     C2    1    0.00000  14.0270     ;  0.00000
    7     C    1    DDAB     C1    2    0.45000  12.0110
    8    OM    1    DDAB     O1    2   -0.45000  15.9994     ;  0.00000
    9    NL    1    DDAB      N    3   -0.31000  14.0067
   10     H    1    DDAB     H3    3    0.31000   1.0080
   11   CH1    1    DDAB     CA    3    0.00000  13.0190     ;  0.00000
   12   CH2    1    DDAB     CB    4    0.00000  14.0270     ;  0.00000
   13   CH2    1    DDAB     CE    5   -0.24000  14.0270
   14    NT    1    DDAB     NZ    5   -0.64000  14.0067
   15     H    1    DDAB    HZ1    5    0.44000   1.0080
   16     H    1    DDAB    HZ2    5    0.44000   1.0080     ;  0.00000
   17     C    1    DDAB      C    6    0.45000  12.0110
   18     O    1    DDAB      O    6   -0.45000  15.9994     ;  0.00000
   19     N    2     DAB      N    7   -0.31000  14.0067
   20     H    2     DAB      H    7    0.31000   1.0080     ;  0.00000
   21   CH1    2     DAB     CA    8    0.00000  13.0190
   22   CH2    2     DAB     CB    8    0.00000  14.0270     ;  0.00000
   23   CH2    2     DAB     CE    9   -0.24000  14.0270
   24    NT    2     DAB     NZ    9   -0.64000  14.0067
```

```
25     H    2     DAB    HZ1    9    0.44000   1.0080
26     H    2     DAB    HZ2    9    0.44000   1.0080     ;  0.00000
27     C    2     DAB     C    10    0.45000  12.0110
28     O    2     DAB     O    10   -0.45000  15.9994     ;  0.00000
29     N    3     DAB     N    11   -0.31000  14.0067
30     H    3     DAB     H    11    0.31000   1.0080     ;  0.00000
31   CH1    3     DAB    CA    12    0.00000  13.0190
32   CH2    3     DAB    CB    12    0.00000  14.0270     ;  0.00000
33   CH2    3     DAB    CE    13   -0.24000  14.0270
34    NT    3     DAB    NZ    13   -0.64000  14.0067
35     H    3     DAB    HZ1   13    0.44000   1.0080
36     H    3     DAB    HZ2   13    0.44000   1.0080     ;  0.00000
37     C    3     DAB     C    14    0.45000  12.0110
38     O    3     DAB     O    14   -0.45000  15.9994     ;  0.00000
39     N    4     LEU     N    15   -0.31000  14.0067
40     H    4     LEU     H    15    0.31000   1.0080     ;  0.00000
41   CH1    4     LEU    CA    16    0.00000  13.0190
42   CH2    4     LEU    CB    16    0.00000  14.0270     ;  0.00000
43   CH1    4     LEU    CG    17    0.00000  13.0190
44   CH3    4     LEU   CD1    17    0.00000  15.0350
45   CH3    4     LEU   CD2    17    0.00000  15.0350     ;  0.00000
46     C    4     LEU     C    18    0.45000  12.0110
47     O    4     LEU     O    18   -0.45000  15.9994     ;  0.00000
48     N    5    DPHE     N    19   -0.31000  14.0067
49     H    5    DPHE     H    19    0.31000   1.0080     ;  0.00000
50   CH1    5    DPHE    CA    20    0.00000  13.0190
51   CH2    5    DPHE    CB    20    0.00000  14.0270
52     C    5    DPHE    CG    20    0.00000  12.0110     ;  0.00000
53     C    5    DPHE   CD1    21   -0.14000  12.0110
54    HC    5    DPHE   HD1    21    0.14000   1.0080     ;  0.00000
55     C    5    DPHE   CD2    22   -0.14000  12.0110
56    HC    5    DPHE   HD2    22    0.14000   1.0080     ;  0.00000
```

```
57     C    5    DPHE    CE1   23   -0.14000  12.0110
58    HC    5    DPHE    HE1   23    0.14000   1.0080     ;  0.00000
59     C    5    DPHE    CE2   24   -0.14000  12.0110
60    HC    5    DPHE    HE2   24    0.14000   1.0080     ;  0.00000
61     C    5    DPHE     CZ   25   -0.14000  12.0110
62    HC    5    DPHE     HZ   25    0.14000   1.0080     ;  0.00000
63     C    5    DPHE      C   26    0.45000  12.0110
64     O    5    DPHE      O   26   -0.45000  15.9994     ;  0.00000
65     N    6     DAB      N   27   -0.31000  14.0067
66     H    6     DAB      H   27    0.31000   1.0080     ;  0.00000
67   CH1    6     DAB     CA   28    0.00000  13.0190
68   CH2    6     DAB     CB   28    0.00000  14.0270     ;  0.00000
69   CH2    6     DAB     CE   29   -0.24000  14.0270
70    NT    6     DAB     NZ   29   -0.64000  14.0067
71     H    6     DAB    HZ1   29    0.44000   1.0080
72     H    6     DAB    HZ2   29    0.44000   1.0080     ;  0.00000
73     C    6     DAB      C   30    0.45000  12.0110
74     O    6     DAB      O   30   -0.45000  15.9994     ;  0.00000
75     N    7     DAB      N   31   -0.31000  14.0067
76     H    7     DAB      H   31    0.31000   1.0080     ;  0.00000
77   CH1    7     DAB     CA   32    0.00000  13.0190
78   CH2    7     DAB     CB   32    0.00000  14.0270     ;  0.00000
79   CH2    7     DAB     CE   33   -0.24000  14.0270
80    NT    7     DAB     NZ   33   -0.64000  14.0067
81     H    7     DAB    HZ1   33    0.44000   1.0080
82     H    7     DAB    HZ2   33    0.44000   1.0080     ;  0.00000
83     C    7     DAB      C   34    0.45000  12.0110
84     O    7     DAB      O   34   -0.45000  15.9994     ;  0.00000
85     N    8     LEU      N   35   -0.31000  14.0067
86     H    8     LEU      H   35    0.31000   1.0080     ;  0.00000
87   CH1    8     LEU     CA   36    0.00000  13.0190
88   CH2    8     LEU     CB   36    0.00000  14.0270     ;  0.00000
```

```
   89   CH1    8     LEU     CG   37    0.00000  13.0190
   90   CH3    8     LEU    CD1   37    0.00000  15.0350
   91   CH3    8     LEU    CD2   37    0.00000  15.0350     ;  0.00000
   92     C    8     LEU      C   38    0.29000  12.0110
   93     O    8     LEU     O1   38   -0.45000  15.9994
   94    NT    8     LEU      N   38   -0.72000  14.0067
   95     H    8     LEU     H1   38    0.44000   1.0080
   96     H    8     LEU     H2   38    0.44000   1.0080     ;  0.00000
; total charge of the molecule:   0.00000
[ bonds ]
;  ai   aj  funct   c0         c1
    9   10    2   0.1000     1.8700e+07
   14   15    2   0.1000     1.8700e+07
   14   16    2   0.1000     1.8700e+07
   19   20    2   0.1000     1.8700e+07
   24   25    2   0.1000     1.8700e+07
   24   26    2   0.1000     1.8700e+07
   29   30    2   0.1000     1.8700e+07
   34   35    2   0.1000     1.8700e+07
   34   36    2   0.1000     1.8700e+07
   39   40    2   0.1000     1.8700e+07
   48   49    2   0.1000     1.8700e+07
   53   54    2   0.1090     1.2300e+07
   55   56    2   0.1090     1.2300e+07
   57   58    2   0.1090     1.2300e+07
   59   60    2   0.1090     1.2300e+07
   61   62    2   0.1090     1.2300e+07
   65   66    2   0.1000     1.8700e+07
   70   71    2   0.1000     1.8700e+07
   70   72    2   0.1000     1.8700e+07
   75   76    2   0.1000     1.8700e+07
   80   81    2   0.1000     1.8700e+07
```

80	82	2	0.1000	1.8700e+07
85	86	2	0.1000	1.8700e+07
94	95	2	0.1000	1.8700e+07
94	96	2	0.1000	1.8700e+07
1	3	2	0.1530	7.1500e+06
2	3	2	0.1530	7.1500e+06
3	4	2	0.1530	7.1500e+06
4	5	2	0.1530	7.1500e+06
5	6	2	0.1530	7.1500e+06
6	7	2	0.1530	7.1500e+06
7	8	2	0.1230	1.6600e+07
7	9	2	0.1330	1.1800e+07
9	11	2	0.1470	8.7100e+06
11	12	2	0.1530	7.1500e+06
11	17	2	0.1530	7.1500e+06
12	13	2	0.1530	7.1500e+06
13	14	2	0.1470	8.7100e+06
17	18	2	0.1230	1.6600e+07
17	19	2	0.1330	1.1800e+07
19	21	2	0.1470	8.7100e+06
21	22	2	0.1530	7.1500e+06
21	27	2	0.1530	7.1500e+06
22	23	2	0.1530	7.1500e+06
23	24	2	0.1470	8.7100e+06
27	28	2	0.1230	1.6600e+07
27	29	2	0.1330	1.1800e+07
29	31	2	0.1470	8.7100e+06
31	32	2	0.1530	7.1500e+06
31	37	2	0.1530	7.1500e+06
32	33	2	0.1530	7.1500e+06
33	34	2	0.1470	8.7100e+06
37	38	2	0.1230	1.6600e+07

```
37   39    2   0.1330     1.1800e+07
39   41    2   0.1470     8.7100e+06
41   42    2   0.1530     7.1500e+06
41   46    2   0.1530     7.1500e+06
42   43    2   0.1530     7.1500e+06
43   44    2   0.1530     7.1500e+06
43   45    2   0.1530     7.1500e+06
46   47    2   0.1230     1.6600e+07
46   48    2   0.1330     1.1800e+07
48   50    2   0.1470     8.7100e+06
50   51    2   0.1530     7.1500e+06
50   63    2   0.1530     7.1500e+06
51   52    2   0.1530     7.1500e+06
52   53    2   0.1390     1.0800e+07
52   55    2   0.1390     1.0800e+07
53   57    2   0.1390     1.0800e+07
55   59    2   0.1390     1.0800e+07
57   61    2   0.1390     1.0800e+07
59   61    2   0.1390     1.0800e+07
63   64    2   0.1230     1.6600e+07
63   65    2   0.1330     1.1800e+07
65   67    2   0.1470     8.7100e+06
67   68    2   0.1530     7.1500e+06
67   73    2   0.1530     7.1500e+06
68   69    2   0.1530     7.1500e+06
69   70    2   0.1470     8.7100e+06
73   74    2   0.1230     1.6600e+07
73   75    2   0.1330     1.1800e+07
75   77    2   0.1470     8.7100e+06
77   78    2   0.1530     7.1500e+06
77   83    2   0.1530     7.1500e+06
78   79    2   0.1530     7.1500e+06
```

```
   79   80    2   0.1470     8.7100e+06
   83   84    2   0.1230     1.6600e+07
   83   85    2   0.1330     1.1800e+07
   85   87    2   0.1470     8.7100e+06
   87   88    2   0.1530     7.1500e+06
   87   92    2   0.1530     7.1500e+06
   88   89    2   0.1530     7.1500e+06
   89   90    2   0.1530     7.1500e+06
   89   91    2   0.1530     7.1500e+06
   92   93    2   0.1230     1.6600e+07
   92   94    2   0.1330     1.0600e+07
[ pairs ]
;  ai   aj  funct  ;  all 1-4 pairs but the ones excluded in GROMOS itp
    1    5    1
    2    5    1
    3    6    1
    4    7    1
    5    8    1
    5    9    1
    6   10    1
    6   11    1
    7   12    1
    7   17    1
    9   13    1
    9   18    1
    9   19    1
   10   12    1
   10   17    1
   11   14    1
   11   20    1
   11   21    1
   12   15    1
```

```
12   16    1
12   18    1
12   19    1
13   17    1
17   22    1
17   27    1
18   20    1
18   21    1
19   23    1
19   28    1
19   29    1
20   22    1
20   27    1
21   24    1
21   30    1
21   31    1
22   25    1
22   26    1
22   28    1
22   29    1
23   27    1
27   32    1
27   37    1
28   30    1
28   31    1
29   33    1
29   38    1
29   39    1
30   32    1
30   37    1
31   34    1
31   40    1
```

```
31   41    1
32   35    1
32   36    1
32   38    1
32   39    1
33   37    1
37   42    1
37   46    1
38   40    1
38   41    1
39   43    1
39   47    1
39   48    1
40   42    1
40   46    1
41   44    1
41   45    1
41   49    1
41   50    1
42   47    1
42   48    1
43   46    1
46   51    1
46   63    1
47   49    1
47   50    1
48   52    1
48   64    1
48   65    1
49   51    1
49   63    1
50   53    1
```

```
50   55    1
50   66    1
50   67    1
51   64    1
51   65    1
52   63    1
63   68    1
63   73    1
64   66    1
64   67    1
65   69    1
65   74    1
65   75    1
66   68    1
66   73    1
67   70    1
67   76    1
67   77    1
68   71    1
68   72    1
68   74    1
68   75    1
69   73    1
73   78    1
73   83    1
74   76    1
74   77    1
75   79    1
75   84    1
75   85    1
76   78    1
76   83    1
```

```
   77   80    1
   77   86    1
   77   87    1
   78   81    1
   78   82    1
   78   84    1
   78   85    1
   79   83    1
   83   88    1
   83   92    1
   84   86    1
   84   87    1
   85   89    1
   85   93    1
   85   94    1
   86   88    1
   86   92    1
   87   90    1
   87   91    1
   87   95    1
   87   96    1
   88   93    1
   88   94    1
   89   92    1
[ angles ]
;  ai   aj   ak  funct   angle     fc
    7    9   10    2    123.00    415.00
   10    9   11    2    123.00    415.00
   13   14   15    2    109.50    425.00
   13   14   16    2    109.50    425.00
   15   14   16    2    109.50    380.00
   17   19   20    2    123.00    415.00
```

20	19	21	2	115.00	460.00
23	24	25	2	109.50	425.00
23	24	26	2	109.50	425.00
25	24	26	2	109.50	380.00
27	29	30	2	123.00	415.00
30	29	31	2	115.00	460.00
33	34	35	2	109.50	425.00
33	34	36	2	109.50	425.00
35	34	36	2	109.50	380.00
37	39	40	2	123.00	415.00
40	39	41	2	115.00	460.00
46	48	49	2	123.00	415.00
49	48	50	2	115.00	460.00
52	53	54	2	120.00	505.00
54	53	57	2	120.00	505.00
52	55	56	2	120.00	505.00
56	55	59	2	120.00	505.00
53	57	58	2	120.00	505.00
58	57	61	2	120.00	505.00
55	59	60	2	120.00	505.00
60	59	61	2	120.00	505.00
57	61	62	2	120.00	505.00
59	61	62	2	120.00	505.00
63	65	66	2	123.00	415.00
66	65	67	2	115.00	460.00
69	70	71	2	109.50	425.00
69	70	72	2	109.50	425.00
71	70	72	2	109.50	380.00
73	75	76	2	123.00	415.00
76	75	77	2	115.00	460.00
79	80	81	2	109.50	425.00
79	80	82	2	109.50	425.00

81	80	82	2	109.50	380.00
83	85	86	2	123.00	415.00
86	85	87	2	115.00	460.00
92	94	95	2	109.50	425.00
92	94	96	2	109.50	425.00
95	94	96	2	109.50	380.00
1	3	4	2	111.00	530.00
2	3	4	2	111.00	530.00
1	3	2	2	111.00	530.00
3	4	5	2	111.00	530.00
4	5	6	2	111.00	530.00
5	6	7	2	111.00	530.00
6	7	8	2	121.00	685.00
6	7	9	2	115.00	610.00
9	11	12	2	109.50	520.00
9	11	17	2	109.50	520.00
12	11	17	2	109.50	520.00
11	12	13	2	111.00	530.00
12	13	14	2	111.00	530.00
11	17	18	2	121.00	685.00
11	17	19	2	115.00	610.00
18	17	19	2	124.00	730.00
17	19	21	2	122.00	700.00
19	21	22	2	109.50	520.00
19	21	27	2	109.50	520.00
22	21	27	2	109.50	520.00
21	22	23	2	111.00	530.00
22	23	24	2	111.00	530.00
21	27	28	2	121.00	685.00
21	27	29	2	115.00	610.00
28	27	29	2	124.00	730.00
27	29	31	2	122.00	700.00

29	31	32	2	109.50	520.00
29	31	37	2	109.50	520.00
32	31	37	2	109.50	520.00
31	32	33	2	111.00	530.00
32	33	34	2	111.00	530.00
31	37	38	2	121.00	685.00
31	37	39	2	115.00	610.00
38	37	39	2	124.00	730.00
37	39	41	2	122.00	700.00
39	41	42	2	109.50	520.00
39	41	46	2	109.50	520.00
42	41	46	2	109.50	520.00
41	42	43	2	111.00	530.00
42	43	44	2	111.00	530.00
42	43	45	2	111.00	530.00
44	43	45	2	111.00	530.00
41	46	47	2	121.00	685.00
41	46	48	2	115.00	610.00
47	46	48	2	124.00	730.00
46	48	50	2	122.00	700.00
48	50	51	2	109.50	520.00
48	50	63	2	109.50	520.00
51	50	63	2	109.50	520.00
50	51	52	2	111.00	530.00
51	52	53	2	120.00	560.00
51	52	55	2	120.00	560.00
53	52	55	2	120.00	560.00
52	53	57	2	120.00	560.00
52	55	59	2	120.00	560.00
53	57	61	2	120.00	560.00
55	59	61	2	120.00	560.00
57	61	59	2	120.00	560.00

```
50   63   64    2    121.00    685.00
50   63   65    2    115.00    610.00
64   63   65    2    124.00    730.00
63   65   67    2    122.00    700.00
65   67   68    2    109.50    520.00
65   67   73    2    109.50    520.00
68   67   73    2    109.50    520.00
67   68   69    2    111.00    530.00
68   69   70    2    111.00    530.00
67   73   74    2    121.00    685.00
67   73   75    2    115.00    610.00
74   73   75    2    124.00    730.00
73   75   77    2    122.00    700.00
75   77   78    2    109.50    520.00
75   77   83    2    109.50    520.00
78   77   83    2    109.50    520.00
77   78   79    2    111.00    530.00
78   79   80    2    111.00    530.00
77   83   84    2    121.00    685.00
77   83   85    2    115.00    610.00
84   83   85    2    124.00    730.00
83   85   87    2    122.00    700.00
85   87   88    2    109.50    520.00
85   87   92    2    109.50    520.00
88   87   92    2    109.50    520.00
87   88   89    2    111.00    530.00
88   89   90    2    111.00    530.00
88   89   91    2    111.00    530.00
90   89   91    2    111.00    530.00
87   92   93    2    121.00    685.00
87   92   94    2    115.00    610.00
93   92   94    2    124.00    730.00
```

```
[ dihedrals ]
; GROMOS improper dihedrals
;  ai   aj   ak   al  funct   angle     fc
   19   17   21   20    2      0.00    167.42
   29   27   31   30    2      0.00    167.42
   39   37   41   40    2      0.00    167.42
   48   46   50   49    2      0.00    167.42
   53   52   57   54    2      0.00    167.42
   55   52   59   56    2      0.00    167.42
   58   53   61   57    2      0.00    167.42
   60   55   61   59    2      0.00    167.42
   61   57   59   62    2      0.00    167.42
   65   63   67   66    2      0.00    167.42
   75   73   77   76    2      0.00    167.42
   85   83   87   86    2      0.00    167.42
   11    9   17   12    2    -35.26    334.85
   17   11   19   18    2      0.00    167.42
   21   19   27   22    2     35.26    334.85
   27   21   29   28    2      0.00    167.42
   31   29   37   32    2     35.26    334.85
   37   31   39   38    2      0.00    167.42
   41   39   46   42    2     35.26    334.85
   42   44   45   43    2     35.26    334.85
   46   41   48   47    2      0.00    167.42
   50   48   63   51    2    -35.26    334.85
   52   53   55   51    2      0.00    167.42
   52   53   57   61    2      0.00    167.42
   52   55   59   61    2      0.00    167.42
   53   52   55   59    2      0.00    167.42
   53   57   61   59    2      0.00    167.42
   55   52   53   57    2      0.00    167.42
   55   59   61   57    2      0.00    167.42
```

```
   63   50   65   64    2      0.00    167.42
   67   65   73   68    2     35.26    334.85
   73   67   75   74    2      0.00    167.42
   77   75   83   78    2     35.26    334.85
   83   77   85   84    2      0.00    167.42
   87   85   92   88    2     35.26    334.85
   88   90   91   89    2     35.26    334.85
   92   87   94   93    2      0.00    167.42
[ dihedrals ]
;  ai   aj   ak   al  funct    ph0      cp     mult
   12   13   14   15    1      0.00      3.77    3
   22   23   24   25    1      0.00      3.77    3
   32   33   34   35    1      0.00      3.77    3
   68   69   70   71    1      0.00      3.77    3
   78   79   80   81    1      0.00      3.77    3
   87   92   94   95    1    180.00     33.50    2
    1    3    4    5    1      0.00      5.92    3
    3    4    5    6    1      0.00      5.92    3
    4    5    6    7    1      0.00      5.92    3
    5    6    7    9    1    180.00      3.50    2
    6    7    9   11    1    180.00     33.50    2
    7    9   11   17    1      0.00      3.77    3
    9   11   12   13    1      0.00      5.92    3
    9   11   17   19    1    180.00      3.50    2
   11   12   13   14    1      0.00      5.92    3
   11   17   19   21    1    180.00     33.50    2
   17   19   21   27    1      0.00      2.80    3
   19   21   22   23    1      0.00      5.92    3
   19   21   27   29    1    180.00      3.50    2
   21   22   23   24    1      0.00      5.92    3
   21   27   29   31    1    180.00     33.50    2
   27   29   31   37    1      0.00      2.80    3
```

```
   29   31   32   33    1      0.00      5.92    3
   29   31   37   39    1    180.00      3.50    2
   31   32   33   34    1      0.00      5.92    3
   31   37   39   41    1    180.00     33.50    2
   37   39   41   46    1      0.00      2.80    3
   39   41   42   43    1      0.00      5.92    3
   39   41   46   48    1    180.00      3.50    2
   41   42   43   44    1      0.00      5.92    3
   41   46   48   50    1    180.00     33.50    2
   46   48   50   63    1      0.00      2.80    3
   48   50   51   52    1      0.00      5.92    3
   48   50   63   65    1    180.00      3.50    2
   50   51   52   53    1      0.00      1.00    6
   50   63   65   67    1    180.00     33.50    2
   63   65   67   73    1      0.00      2.80    3
   65   67   68   69    1      0.00      5.92    3
   65   67   73   75    1    180.00      3.50    2
   67   68   69   70    1      0.00      5.92    3
   67   73   75   77    1    180.00     33.50    2
   73   75   77   83    1      0.00      2.80    3
   75   77   78   79    1      0.00      5.92    3
   75   77   83   85    1    180.00      3.50    2
   77   78   79   80    1      0.00      5.92    3
   77   83   85   87    1    180.00     33.50    2
   83   85   87   92    1      0.00      2.80    3
   85   87   88   89    1      0.00      5.92    3
   85   87   92   94    1    180.00      3.50    2
   87   88   89   90    1      0.00      5.92    3
[ exclusions ]
;  ai   aj  funct  ;  GROMOS 1-4 exclusions
    1    5
    2    5
```

```
 3    6
 4    7
 5    8
 5    9
 6   10
 6   11
 8   10
 8   11
51   54
51   56
51   57
51   59
52   58
52   60
52   61
53   56
53   59
53   62
54   55
54   58
54   61
55   57
55   62
56   60
56   61
57   60
58   59
58   62
60   62
93   95
93   96
```

1.2 Starting coordinates (pdb format)

```
TITLE     octapep_o17
CRYST1   30.000   36.203   30.000  90.00  90.00  90.00 P 1           1
MODEL        1
ATOM      1  C6  DDAB    1       0.760  32.520   0.870  1.00  0.00
ATOM      2  C7  DDAB    1      -0.510  31.190  -1.360  1.00  0.00
ATOM      3  C5  DDAB    1      -0.230  32.200  -0.250  1.00  0.00
ATOM      4  C4  DDAB    1      -1.210  33.360  -0.300  1.00  0.00
ATOM      5  C3  DDAB    1      -0.330  34.590  -0.370  1.00  0.00
ATOM      6  C2  DDAB    1      -1.140  35.870  -0.480  1.00  0.00
ATOM      7  C1  DDAB    1      -0.100  36.980  -0.610  1.00  0.00
ATOM      8  O1  DDAB    1       1.000  36.750  -1.110  1.00  0.00
ATOM      9  N   DDAB    1      -0.550  38.140  -0.170  1.00  0.00
ATOM     10  H3  DDAB    1      -1.540  38.240  -0.010  1.00  0.00
ATOM     11  CA  DDAB    1       0.230  39.390  -0.070  1.00  0.00
ATOM     12  CB  DDAB    1       1.230  39.220   1.080  1.00  0.00
ATOM     13  CE  DDAB    1       2.610  39.810   0.830  1.00  0.00
ATOM     14  NZ  DDAB    1       3.350  40.110   2.060  1.00  0.00
ATOM     15  HZ1 DDAB    1       4.000  40.870   1.900  1.00  0.00
ATOM     16  HZ2 DDAB    1       3.950  39.350   2.370  1.00  0.00
ATOM     17  C   DDAB    1      -0.780  40.500   0.250  1.00  0.00
ATOM     18  O   DDAB    1      -1.980  40.220   0.180  1.00  0.00
ATOM     19  N   DAB     2      -0.330  41.740   0.280  1.00  0.00
ATOM     20  H   DAB     2       0.620  41.960   0.040  1.00  0.00
ATOM     21  CA  DAB     2      -1.140  42.940   0.610  1.00  0.00
ATOM     22  CB  DAB     2      -1.480  42.880   2.110  1.00  0.00
ATOM     23  CE  DAB     2      -2.320  44.000   2.730  1.00  0.00
ATOM     24  NZ  DAB     2      -2.730  43.620   4.080  1.00  0.00
ATOM     25  HZ1 DAB     2      -3.140  42.710   4.030  1.00  0.00
ATOM     26  HZ2 DAB     2      -3.480  44.220   4.390  1.00  0.00
ATOM     27  C   DAB     2      -0.330  44.190   0.200  1.00  0.00
ATOM     28  O   DAB     2       0.680  44.070  -0.490  1.00  0.00
```

```
ATOM     29  N   DAB     3      -0.930  45.360   0.400  1.00  0.00
ATOM     30  H   DAB     3      -1.810  45.380   0.880  1.00  0.00
ATOM     31  CA  DAB     3      -0.300  46.670   0.210  1.00  0.00
ATOM     32  CB  DAB     3      -0.250  47.060  -1.280  1.00  0.00
ATOM     33  CE  DAB     3       0.820  48.120  -1.560  1.00  0.00
ATOM     34  NZ  DAB     3       0.680  48.650  -2.920  1.00  0.00
ATOM     35  HZ1 DAB     3       0.660  47.910  -3.610  1.00  0.00
ATOM     36  HZ2 DAB     3       1.410  49.310  -3.140  1.00  0.00
ATOM     37  C   DAB     3      -1.110  47.720   0.990  1.00  0.00
ATOM     38  O   DAB     3      -2.280  47.510   1.330  1.00  0.00
ATOM     39  N   LEU     4      -0.480  48.860   1.200  1.00  0.00
ATOM     40  H   LEU     4       0.370  49.050   0.710  1.00  0.00
ATOM     41  CA  LEU     4      -1.000  49.960   2.020  1.00  0.00
ATOM     42  CB  LEU     4      -0.780  49.590   3.480  1.00  0.00
ATOM     43  CG  LEU     4      -1.910  50.130   4.340  1.00  0.00
ATOM     44  CD1 LEU     4      -3.230  49.390   4.140  1.00  0.00
ATOM     45  CD2 LEU     4      -1.470  50.150   5.790  1.00  0.00
ATOM     46  C   LEU     4      -0.290  51.270   1.660  1.00  0.00
ATOM     47  O   LEU     4       0.930  51.300   1.740  1.00  0.00
ATOM     48  N   DPHE    5      -1.120  52.270   1.340  1.00  0.00
ATOM     49  H   DPHE    5      -2.080  52.180   1.630  1.00  0.00
ATOM     50  CA  DPHE    5      -0.850  53.580   0.690  1.00  0.00
ATOM     51  CB  DPHE    5       0.590  53.810   0.190  1.00  0.00
ATOM     52  CG  DPHE    5       1.010  52.910  -0.970  1.00  0.00
ATOM     53  CD1 DPHE    5       0.210  52.740  -2.100  1.00  0.00
ATOM     54  HD1 DPHE    5      -0.840  53.030  -2.100  1.00  0.00
ATOM     55  CD2 DPHE    5       2.300  52.410  -0.950  1.00  0.00
ATOM     56  HD2 DPHE    5       2.920  52.550  -0.060  1.00  0.00
ATOM     57  CE1 DPHE    5       0.740  52.110  -3.220  1.00  0.00
ATOM     58  HE1 DPHE    5       0.130  52.000  -4.120  1.00  0.00
ATOM     59  CE2 DPHE    5       2.820  51.770  -2.060  1.00  0.00
ATOM     60  HE2 DPHE    5       3.840  51.390  -2.060  1.00  0.00
```

```
ATOM     61  CZ  DPHE    5       2.050  51.650  -3.210  1.00  0.00
ATOM     62  HZ  DPHE    5       2.560  51.440  -4.140  1.00  0.00
ATOM     63  C   DPHE    5      -1.410  54.810   1.460  1.00  0.00
ATOM     64  O   DPHE    5      -2.600  54.820   1.770  1.00  0.00
ATOM     65  N   DAB     6      -0.710  55.940   1.390  1.00  0.00
ATOM     66  H   DAB     6       0.050  56.060   0.750  1.00  0.00
ATOM     67  CA  DAB     6      -1.030  57.140   2.180  1.00  0.00
ATOM     68  CB  DAB     6      -0.290  57.020   3.520  1.00  0.00
ATOM     69  CE  DAB     6      -0.850  58.010   4.550  1.00  0.00
ATOM     70  NZ  DAB     6      -0.220  57.920   5.880  1.00  0.00
ATOM     71  HZ1 DAB     6      -0.260  56.970   6.220  1.00  0.00
ATOM     72  HZ2 DAB     6      -0.770  58.460   6.520  1.00  0.00
ATOM     73  C   DAB     6      -0.570  58.390   1.420  1.00  0.00
ATOM     74  O   DAB     6       0.370  58.320   0.630  1.00  0.00
ATOM     75  N   DAB     7      -1.280  59.500   1.630  1.00  0.00
ATOM     76  H   DAB     7      -2.160  59.470   2.100  1.00  0.00
ATOM     77  CA  DAB     7      -0.910  60.820   1.090  1.00  0.00
ATOM     78  CB  DAB     7      -1.540  61.020  -0.290  1.00  0.00
ATOM     79  CE  DAB     7      -1.120  62.280  -1.050  1.00  0.00
ATOM     80  NZ  DAB     7      -1.870  62.400  -2.300  1.00  0.00
ATOM     81  HZ1 DAB     7      -2.800  62.740  -2.140  1.00  0.00
ATOM     82  HZ2 DAB     7      -1.370  62.980  -2.940  1.00  0.00
ATOM     83  C   DAB     7      -1.340  61.920   2.070  1.00  0.00
ATOM     84  O   DAB     7      -2.130  61.700   2.990  1.00  0.00
ATOM     85  N   LEU     8      -0.660  63.050   1.930  1.00  0.00
ATOM     86  H   LEU     8      -0.010  63.170   1.180  1.00  0.00
ATOM     87  CA  LEU     8      -0.830  64.220   2.800  1.00  0.00
ATOM     88  CB  LEU     8       0.090  64.010   4.000  1.00  0.00
ATOM     89  CG  LEU     8      -0.220  64.980   5.150  1.00  0.00
ATOM     90  CD1 LEU     8      -1.630  64.760   5.680  1.00  0.00
ATOM     91  CD2 LEU     8       0.780  64.720   6.260  1.00  0.00
ATOM     92  C   LEU     8      -0.440  65.490   2.050  1.00  0.00
```

```
ATOM     93  O1  LEU     8       0.520  65.510   1.270  1.00  0.00
ATOM     94  N   LEU     8      -1.350  66.450   2.150  1.00  0.00
ATOM     95  H1  LEU     8      -1.060  67.310   1.740  1.00  0.00
ATOM     96  H2  LEU     8      -1.650  66.590   3.090  1.00  0.00
TER
ENDMDL
```

Appendix 2: Pentapeptide P30 – GROMOS 54a7 forcefield parameters and starting configuration

2.1 GROMOS 54a7 forcefield parameters

```
[ moleculetype ]

; Name   nrexcl

pentapep 3

[ atoms ]

;  nr  type  resnr  resid  atom  cgnr  charge    mass    total_charge
added_charge

    1   CH3    1    DDAB     C6    1    0.00000  15.0350

    2   CH3    1    DDAB     C7    1    0.00000  15.0350

    3   CH1    1    DDAB     C5    1    0.00000  14.0270

    4   CH2    1    DDAB     C4    1    0.00000  14.0270

    5   CH2    1    DDAB     C3    1    0.00000  14.0270

    6   CH2    1    DDAB     C2    1    0.00000  14.0270     ;  0.00000

    7     C    1    DDAB     C1    2    0.45000  12.0110

    8    OM    1    DDAB     O1    2   -0.45000  15.9994     ;  0.00000

    9    NL    1    DDAB      N    3   -0.31000  14.0067

   10     H    1    DDAB     H3    3    0.31000   1.0080

   11   CH1    1    DDAB     CA    3    0.00000  13.0190     ;  0.00000

   12   CH2    1    DDAB     CB    4    0.00000  14.0270     ;  0.00000

   13   CH2    1    DDAB     CE    5   -0.24000  14.0270

   14    NT    1    DDAB     NZ    5   -0.64000  14.0067

   15     H    1    DDAB    HZ1    5    0.44000   1.0080

   16     H    1    DDAB    HZ2    5    0.44000   1.0080     ;  0.00000

   17     C    1    DDAB      C    6    0.45000  12.0110

   18     O    1    DDAB      O    6   -0.45000  15.9994     ;  0.00000

   19     N    2     DAB      N    7   -0.31000  14.0067

   20     H    2     DAB      H    7    0.31000   1.0080     ;  0.00000

   21   CH1    2     DAB     CA    8    0.00000  13.0190

   22   CH2    2     DAB     CB    8    0.00000  14.0270     ;  0.00000

   23   CH2    2     DAB     CE    9   -0.24000  14.0270
```

```
24    NT    2     DAB     NZ    9   -0.64000  14.0067
25     H    2     DAB    HZ1    9    0.44000   1.0080
26     H    2     DAB    HZ2    9    0.44000   1.0080     ;  0.00000
27     C    2     DAB      C   10    0.45000  12.0110
28     O    2     DAB      O   10   -0.45000  15.9994     ;  0.00000
29     N    3     LEU      N   11   -0.31000  14.0067
30     H    3     LEU      H   11    0.31000   1.0080     ;  0.00000
31   CH1    3     LEU     CA   12    0.00000  13.0190
32   CH2    3     LEU     CB   12    0.00000  14.0270     ;  0.00000
33   CH1    3     LEU     CG   13    0.00000  13.0190
34   CH3    3     LEU    CD1   13    0.00000  15.0350
35   CH3    3     LEU    CD2   13    0.00000  15.0350     ;  0.00000
36     C    3     LEU      C   14    0.45000  12.0110
37     O    3     LEU      O   14   -0.45000  15.9994     ;  0.00000
38     N    4    DPHE      N   15   -0.31000  14.0067
39     H    4    DPHE      H   15    0.31000   1.0080     ;  0.00000
40   CH1    4    DPHE     CA   16    0.00000  13.0190
41   CH2    4    DPHE     CB   16    0.00000  14.0270
42     C    4    DPHE     CG   16    0.00000  12.0110     ;  0.00000
43     C    4    DPHE    CD1   17   -0.14000  12.0110
44    HC    4    DPHE    HD1   17    0.14000   1.0080     ;  0.00000
45     C    4    DPHE    CD2   18   -0.14000  12.0110
46    HC    4    DPHE    HD2   18    0.14000   1.0080     ;  0.00000
47     C    4    DPHE    CE1   19   -0.14000  12.0110
48    HC    4    DPHE    HE1   19    0.14000   1.0080     ;  0.00000
49     C    4    DPHE    CE2   20   -0.14000  12.0110
50    HC    4    DPHE    HE2   20    0.14000   1.0080     ;  0.00000
51     C    4    DPHE     CZ   21   -0.14000  12.0110
52    HC    4    DPHE     HZ   21    0.14000   1.0080     ;  0.00000
53     C    4    DPHE      C   22    0.45000  12.0110
54     O    4    DPHE      O   22   -0.45000  15.9994     ;  0.00000
55     N    5     DAB      N   23   -0.31000  14.0067
```

```
    56     H    5     DAB      H   23    0.31000   1.0080     ;  0.00000
    57   CH1    5     DAB     CA   24    0.00000  13.0190
    58   CH2    5     DAB     CB   24    0.00000  14.0270     ;  0.00000
    59   CH2    5     DAB     CE   25   -0.24000  14.0270
    60    NT    5     DAB     NZ   25   -0.64000  14.0067
    61     H    5     DAB    HZ1   25    0.44000   1.0080
    62     H    5     DAB    HZ2   25    0.44000   1.0080     ;  0.00000
    63     C    5     DAB      C   26    0.29000  12.0110
    64     O    5     DAB     O1   26   -0.45000  15.9994
    65    NT    5     DAB      N   26   -0.72000  14.0067
    66     H    5     DAB     H1   26    0.44000   1.0080
    67     H    5     DAB     H2   26    0.44000   1.0080     ;  0.00000
; total charge of the molecule:   0.00000
[ bonds ]
;  ai   aj  funct   c0         c1
    9   10    2   0.1000     1.8700e+07
   14   15    2   0.1000     1.8700e+07
   14   16    2   0.1000     1.8700e+07
   19   20    2   0.1000     1.8700e+07
   24   25    2   0.1000     1.8700e+07
   24   26    2   0.1000     1.8700e+07
   29   30    2   0.1000     1.8700e+07
   38   39    2   0.1000     1.8700e+07
   43   44    2   0.1090     1.2300e+07
   45   46    2   0.1090     1.2300e+07
   47   48    2   0.1090     1.2300e+07
   49   50    2   0.1090     1.2300e+07
   51   52    2   0.1090     1.2300e+07
   55   56    2   0.1000     1.8700e+07
   60   61    2   0.1000     1.8700e+07
   60   62    2   0.1000     1.8700e+07
   65   66    2   0.1000     1.8700e+07
```

65	67	2	0.1000	1.8700e+07
1	3	2	0.1530	7.1500e+06
2	3	2	0.1530	7.1500e+06
3	4	2	0.1530	7.1500e+06
4	5	2	0.1530	7.1500e+06
5	6	2	0.1530	7.1500e+06
6	7	2	0.1530	7.1500e+06
7	8	2	0.1230	1.6600e+07
7	9	2	0.1330	1.1800e+07
9	11	2	0.1470	8.7100e+06
11	12	2	0.1530	7.1500e+06
11	17	2	0.1530	7.1500e+06
12	13	2	0.1530	7.1500e+06
13	14	2	0.1470	8.7100e+06
17	18	2	0.1230	1.6600e+07
17	19	2	0.1330	1.1800e+07
19	21	2	0.1470	8.7100e+06
21	22	2	0.1530	7.1500e+06
21	27	2	0.1530	7.1500e+06
22	23	2	0.1530	7.1500e+06
23	24	2	0.1470	8.7100e+06
27	28	2	0.1230	1.6600e+07
27	29	2	0.1330	1.1800e+07
29	31	2	0.1470	8.7100e+06
31	32	2	0.1530	7.1500e+06
31	36	2	0.1530	7.1500e+06
32	33	2	0.1530	7.1500e+06
33	34	2	0.1530	7.1500e+06
33	35	2	0.1530	7.1500e+06
36	37	2	0.1230	1.6600e+07
36	38	2	0.1330	1.1800e+07
38	40	2	0.1470	8.7100e+06

```
   40   41    2   0.1530     7.1500e+06
   40   53    2   0.1530     7.1500e+06
   41   42    2   0.1530     7.1500e+06
   42   43    2   0.1390     1.0800e+07
   42   45    2   0.1390     1.0800e+07
   43   47    2   0.1390     1.0800e+07
   45   49    2   0.1390     1.0800e+07
   47   51    2   0.1390     1.0800e+07
   49   51    2   0.1390     1.0800e+07
   53   54    2   0.1230     1.6600e+07
   53   55    2   0.1330     1.1800e+07
   55   57    2   0.1470     8.7100e+06
   57   58    2   0.1530     7.1500e+06
   57   63    2   0.1530     7.1500e+06
   58   59    2   0.1530     7.1500e+06
   59   60    2   0.1470     8.7100e+06
   63   64    2   0.1230     1.6600e+07
   63   65    2   0.1330     1.0600e+07
[ pairs ]
;  ai   aj  funct  ;  all 1-4 pairs but the ones excluded in GROMOS itp
    1    5    1
    2    5    1
    3    6    1
    4    7    1
    5    8    1
    5    9    1
    6   10    1
    6   11    1
    7   12    1
    7   17    1
    9   13    1
    9   18    1
```

```
 9   19    1
10   12    1
10   17    1
11   14    1
11   20    1
11   21    1
12   15    1
12   16    1
12   18    1
12   19    1
13   17    1
17   22    1
17   27    1
18   20    1
18   21    1
19   23    1
19   28    1
19   29    1
20   22    1
20   27    1
21   24    1
21   30    1
21   31    1
22   25    1
22   26    1
22   28    1
22   29    1
23   27    1
27   32    1
27   36    1
28   30    1
28   31    1
```

```
29   33    1
29   37    1
29   38    1
30   32    1
30   36    1
31   34    1
31   35    1
31   39    1
31   40    1
32   37    1
32   38    1
33   36    1
36   41    1
36   53    1
37   39    1
37   40    1
38   42    1
38   54    1
38   55    1
39   41    1
39   53    1
40   43    1
40   45    1
40   56    1
40   57    1
41   54    1
41   55    1
42   53    1
53   58    1
53   63    1
54   56    1
54   57    1
```

```
   55   59    1
   55   64    1
   55   65    1
   56   58    1
   56   63    1
   57   60    1
   57   66    1
   57   67    1
   58   61    1
   58   62    1
   58   64    1
   58   65    1
   59   63    1
[ angles ]
;  ai   aj   ak  funct   angle     fc
    7    9   10    2    123.00    415.00
   10    9   11    2    123.00    415.00
   13   14   15    2    109.50    425.00
   13   14   16    2    109.50    425.00
   15   14   16    2    109.50    380.00
   17   19   20    2    123.00    415.00
   20   19   21    2    115.00    460.00
   23   24   25    2    109.50    425.00
   23   24   26    2    109.50    425.00
   25   24   26    2    109.50    380.00
   27   29   30    2    123.00    415.00
   30   29   31    2    115.00    460.00
   36   38   39    2    123.00    415.00
   39   38   40    2    115.00    460.00
   42   43   44    2    120.00    505.00
   44   43   47    2    120.00    505.00
   42   45   46    2    120.00    505.00
```

46	45	49	2	120.00	505.00
43	47	48	2	120.00	505.00
48	47	51	2	120.00	505.00
45	49	50	2	120.00	505.00
50	49	51	2	120.00	505.00
47	51	52	2	120.00	505.00
49	51	52	2	120.00	505.00
53	55	56	2	123.00	415.00
56	55	57	2	115.00	460.00
59	60	61	2	109.50	425.00
59	60	62	2	109.50	425.00
61	60	62	2	109.50	380.00
63	65	66	2	109.50	425.00
63	65	67	2	109.50	425.00
1	3	2	2	111.00	530.00
1	3	4	2	111.00	530.00
2	3	4	2	111.00	530.00
3	4	5	2	111.00	530.00
4	5	6	2	111.00	530.00
5	6	7	2	111.00	530.00
6	7	8	2	121.00	685.00
6	7	9	2	115.00	610.00
9	11	12	2	109.50	520.00
9	11	17	2	109.50	520.00
12	11	17	2	109.50	520.00
11	12	13	2	111.00	530.00
12	13	14	2	111.00	530.00
11	17	18	2	121.00	685.00
11	17	19	2	115.00	610.00
18	17	19	2	124.00	730.00
17	19	21	2	122.00	700.00
19	21	22	2	109.50	520.00

19	21	27	2	109.50	520.00
22	21	27	2	109.50	520.00
21	22	23	2	111.00	530.00
22	23	24	2	111.00	530.00
21	27	28	2	121.00	685.00
21	27	29	2	115.00	610.00
28	27	29	2	124.00	730.00
27	29	31	2	122.00	700.00
29	31	32	2	109.50	520.00
29	31	36	2	109.50	520.00
32	31	36	2	109.50	520.00
31	32	33	2	111.00	530.00
32	33	34	2	111.00	530.00
32	33	35	2	111.00	530.00
34	33	35	2	111.00	530.00
31	36	37	2	121.00	685.00
31	36	38	2	115.00	610.00
37	36	38	2	124.00	730.00
36	38	40	2	122.00	700.00
38	40	41	2	109.50	520.00
38	40	53	2	109.50	520.00
41	40	53	2	109.50	520.00
40	41	42	2	111.00	530.00
41	42	43	2	120.00	560.00
41	42	45	2	120.00	560.00
43	42	45	2	120.00	560.00
42	43	47	2	120.00	560.00
42	45	49	2	120.00	560.00
43	47	51	2	120.00	560.00
45	49	51	2	120.00	560.00
47	51	49	2	120.00	560.00
40	53	54	2	121.00	685.00

```
   40   53   55    2    115.00    610.00
   54   53   55    2    124.00    730.00
   53   55   57    2    122.00    700.00
   55   57   58    2    109.50    520.00
   55   57   63    2    109.50    520.00
   58   57   63    2    109.50    520.00
   57   58   59    2    111.00    530.00
   58   59   60    2    111.00    530.00
   57   63   64    2    121.00    685.00
   57   63   65    2    115.00    610.00
   64   63   65    2    124.00    730.00
[ dihedrals ]
; GROMOS improper dihedrals
;  ai   aj   ak   al  funct   angle     fc
   19   17   21   20    2      0.00    167.42
   29   27   31   30    2      0.00    167.42
   38   36   40   39    2      0.00    167.42
   43   42   47   44    2      0.00    167.42
   45   42   49   46    2      0.00    167.42
   48   43   51   47    2      0.00    167.42
   50   45   51   49    2      0.00    167.42
   51   47   49   52    2      0.00    167.42
   55   53   57   56    2      0.00    167.42
   11    9   17   12    2    -35.26    334.85
   17   11   19   18    2      0.00    167.42
   21   19   27   22    2     35.26    334.85
   27   21   29   28    2      0.00    167.42
   31   29   36   32    2     35.26    334.85
   32   34   35   33    2     35.26    334.85
   36   31   38   37    2      0.00    167.42
   40   38   53   41    2    -35.26    334.85
   42   43   45   41    2      0.00    167.42
```

```
   42   43   47   51    2      0.00    167.42
   42   45   49   51    2      0.00    167.42
   43   42   45   49    2      0.00    167.42
   43   47   51   49    2      0.00    167.42
   45   42   43   47    2      0.00    167.42
   45   49   51   47    2      0.00    167.42
   53   40   55   54    2      0.00    167.42
   57   55   63   58    2     35.26    334.85
   63   57   65   64    2      0.00    167.42
[ dihedrals ]
;  ai   aj   ak   al  funct    ph0      cp     mult
   12   13   14   15    1      0.00      3.77    3
   22   23   24   25    1      0.00      3.77    3
   58   59   60   61    1      0.00      3.77    3
   57   63   65   66    1    180.00     33.50    2
    1    3    4    5    1      0.00      5.92    3
    3    4    5    6    1      0.00      5.92    3
    4    5    6    7    1      0.00      5.92    3
    5    6    7    9    1    180.00      3.50    2
    6    7    9   11    1    180.00     33.50    2
    7    9   11   17    1      0.00      3.77    3
    9   11   12   13    1      0.00      5.92    3
    9   11   17   19    1    180.00      3.50    2
   11   12   13   14    1      0.00      5.92    3
   11   17   19   21    1    180.00     33.50    2
   17   19   21   27    1      0.00      2.80    3
   19   21   22   23    1      0.00      5.92    3
   19   21   27   29    1    180.00      3.50    2
   21   22   23   24    1      0.00      5.92    3
   21   27   29   31    1    180.00     33.50    2
   27   29   31   36    1      0.00      2.80    3
   29   31   32   33    1      0.00      5.92    3
```

```
   29   31   36   38    1    180.00      3.50    2
   31   32   33   34    1      0.00      5.92    3
   31   36   38   40    1    180.00     33.50    2
   36   38   40   53    1      0.00      2.80    3
   38   40   41   42    1      0.00      5.92    3
   38   40   53   55    1    180.00      3.50    2
   40   41   42   43    1      0.00      1.00    6
   40   53   55   57    1    180.00     33.50    2
   53   55   57   63    1      0.00      2.80    3
   55   57   58   59    1      0.00      5.92    3
   55   57   63   65    1    180.00      3.50    2
   57   58   59   60    1      0.00      5.92    3
[ exclusions ]
;  ai   aj  funct  ;  GROMOS 1-4 exclusions
    1    5
    2    5
    3    6
    4    7
    5    8
    5    9
    6   10
    6   11
    8   10
    8   11
   41   44
   41   46
   41   47
   41   49
   42   48
   42   50
   42   51
   43   46
```

```
  43   49
  43   52
  44   45
  44   48
  44   51
  45   47
  45   52
  46   50
  46   51
  47   50
  48   49
  48   52
  50   52
  64   66
  64   67
```

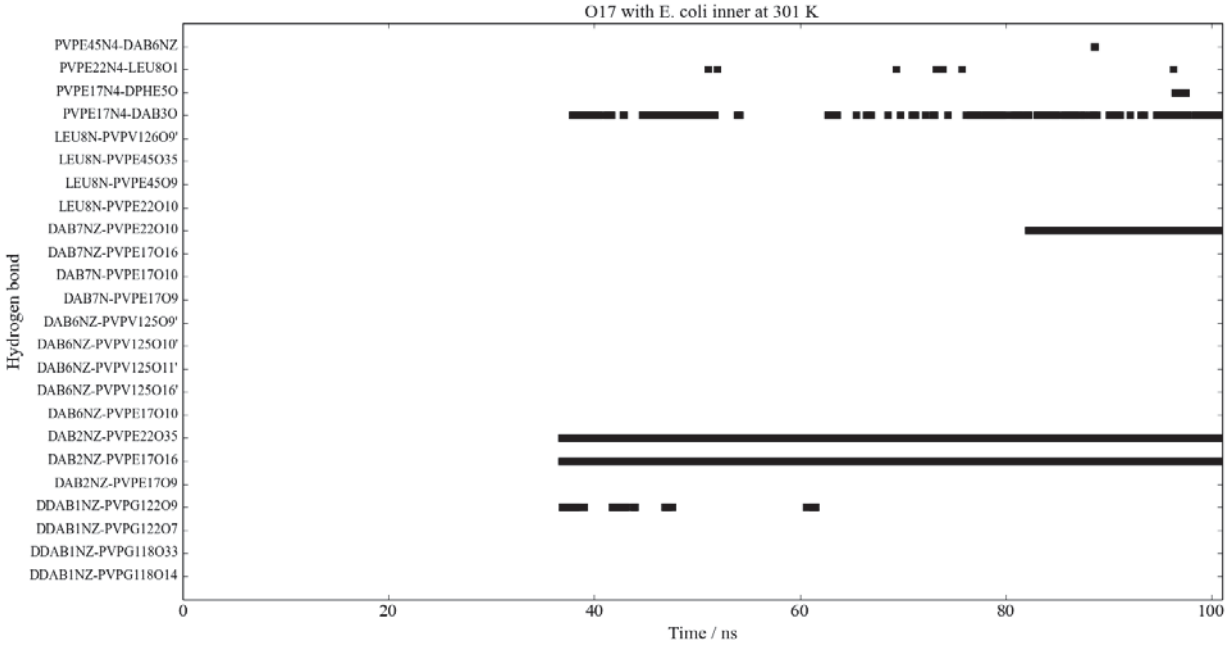
2.2 Starting coordinates (pdb format)

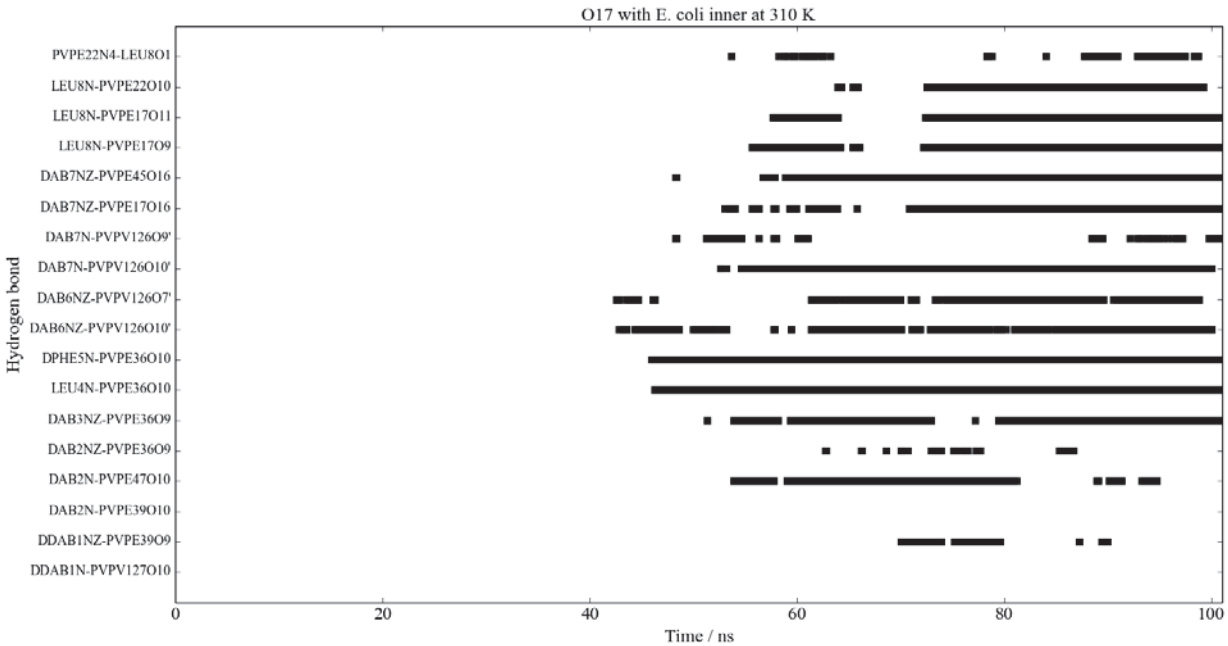
```
TITLE     pentapep_p30
CRYST1   30.000   30.000   30.000  90.00  90.00  90.00 P 1           1
MODEL        1
ATOM      1  C6  DDAB    1       0.440  24.680  27.790  1.00  0.00
ATOM      2  C7  DDAB    1      -1.400  25.190  28.970  1.00  0.00
ATOM      3  C5  DDAB    1       0.060  25.550  28.970  1.00  0.00
ATOM      4  C4  DDAB    1       0.140  27.050  28.650  1.00  0.00
ATOM      5  C3  DDAB    1       0.810  27.990  29.650  1.00  0.00
ATOM      6  C2  DDAB    1       0.690  29.380  29.030  1.00  0.00
ATOM      7  C1  DDAB    1       1.280  30.450  29.930  1.00  0.00
ATOM      8  O1  DDAB    1       2.500  30.680  29.890  1.00  0.00
ATOM      9  N   DDAB    1       0.520  31.540  29.920  1.00  0.00
ATOM     10  H3  DDAB    1      -0.310  31.560  29.350  1.00  0.00
ATOM     11  CA  DDAB    1       0.840  32.770  30.660  1.00  0.00
ATOM     12  CB  DDAB    1       0.720  32.500  32.160  1.00  0.00
```

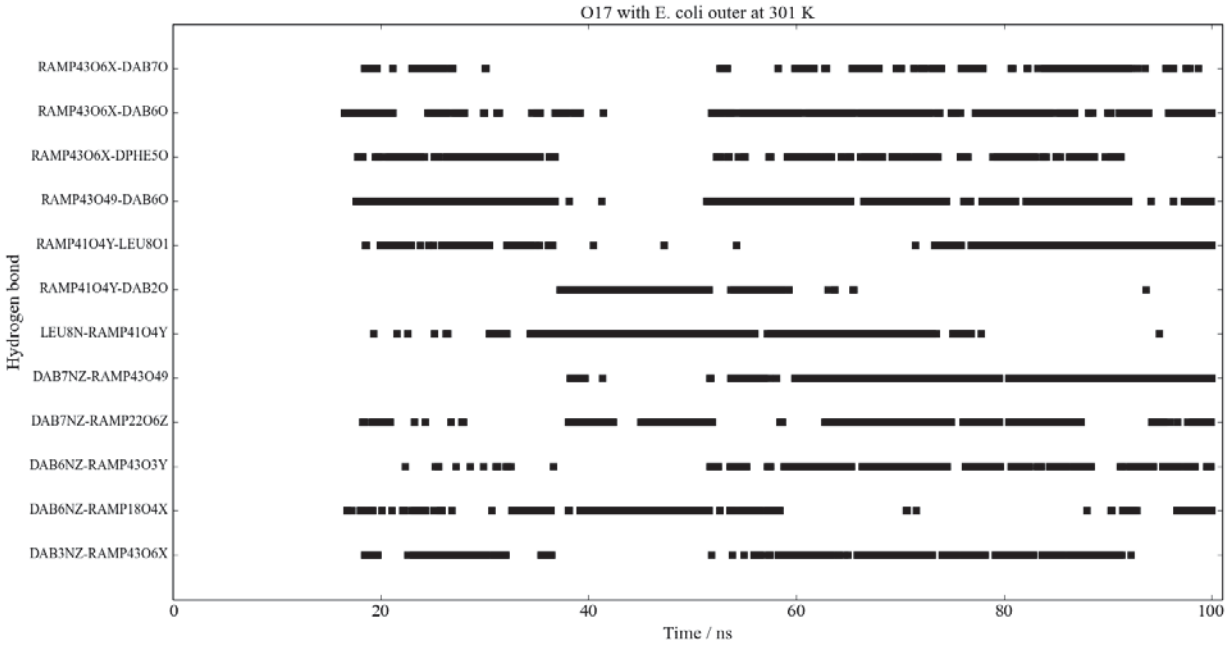
```
ATOM     13  CE   DDAB    1       1.410  33.580  32.980  1.00  0.00
ATOM     14  NZ   DDAB    1       1.330  33.210  34.390  1.00  0.00
ATOM     15  HZ1  DDAB    1       1.690  33.970  34.950  1.00  0.00
ATOM     16  HZ2  DDAB    1       1.890  32.400  34.560  1.00  0.00
ATOM     17  C    DDAB    1      -0.100  33.900  30.200  1.00  0.00
ATOM     18  O    DDAB    1      -1.050  33.620  29.450  1.00  0.00
ATOM     19  N    DAB     2       0.420  35.110  30.290  1.00  0.00
ATOM     20  H    DAB     2       1.390  35.260  30.470  1.00  0.00
ATOM     21  CA   DAB     2      -0.380  36.350  30.170  1.00  0.00
ATOM     22  CB   DAB     2      -1.240  36.420  31.440  1.00  0.00
ATOM     23  CE   DAB     2      -2.270  37.540  31.540  1.00  0.00
ATOM     24  NZ   DAB     2      -2.940  37.470  32.850  1.00  0.00
ATOM     25  HZ1  DAB     2      -3.810  36.950  32.770  1.00  0.00
ATOM     26  HZ2  DAB     2      -3.160  38.390  33.170  1.00  0.00
ATOM     27  C    DAB     2       0.490  37.610  30.020  1.00  0.00
ATOM     28  O    DAB     2       1.650  37.660  30.420  1.00  0.00
ATOM     29  N    LEU     3      -0.200  38.650  29.560  1.00  0.00
ATOM     30  H    LEU     3      -1.200  38.600  29.620  1.00  0.00
ATOM     31  CA   LEU     3       0.290  40.040  29.380  1.00  0.00
ATOM     32  CB   LEU     3       0.360  40.370  27.890  1.00  0.00
ATOM     33  CG   LEU     3       1.670  39.940  27.240  1.00  0.00
ATOM     34  CD1  LEU     3       1.490  39.540  25.780  1.00  0.00
ATOM     35  CD2  LEU     3       2.740  41.040  27.380  1.00  0.00
ATOM     36  C    LEU     3      -0.660  41.010  30.100  1.00  0.00
ATOM     37  O    LEU     3      -1.830  40.640  30.250  1.00  0.00
ATOM     38  N    DPHE    4      -0.280  42.290  30.080  1.00  0.00
ATOM     39  H    DPHE    4       0.270  42.590  29.300  1.00  0.00
ATOM     40  CA   DPHE    4      -0.680  43.410  30.970  1.00  0.00
ATOM     41  CB   DPHE    4      -2.150  43.390  31.470  1.00  0.00
ATOM     42  CG   DPHE    4      -2.450  42.440  32.640  1.00  0.00
ATOM     43  CD1  DPHE    4      -3.710  41.850  32.690  1.00  0.00
ATOM     44  HD1  DPHE    4      -4.430  42.080  31.910  1.00  0.00
```

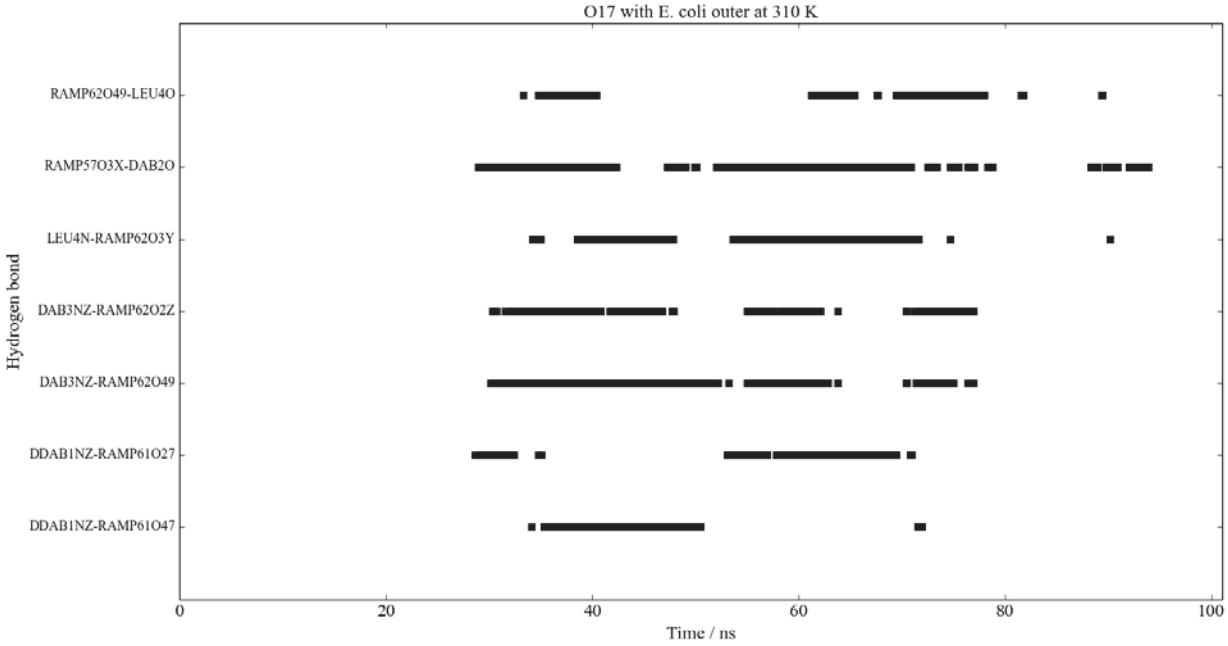
```
ATOM     45  CD2 DPHE    4      -1.570  42.190  33.680  1.00  0.00
ATOM     46  HD2 DPHE    4      -0.600  42.690  33.720  1.00  0.00
ATOM     47  CE1 DPHE    4      -4.070  41.030  33.760  1.00  0.00
ATOM     48  HE1 DPHE    4      -5.110  40.710  33.880  1.00  0.00
ATOM     49  CE2 DPHE    4      -1.910  41.360  34.740  1.00  0.00
ATOM     50  HE2 DPHE    4      -1.190  41.160  35.530  1.00  0.00
ATOM     51  CZ  DPHE    4      -3.170  40.770  34.780  1.00  0.00
ATOM     52  HZ  DPHE    4      -3.450  40.170  35.650  1.00  0.00
ATOM     53  C   DPHE    4      -0.340  44.820  30.440  1.00  0.00
ATOM     54  O   DPHE    4       0.330  44.930  29.400  1.00  0.00
ATOM     55  N   DAB     5      -0.530  45.820  31.280  1.00  0.00
ATOM     56  H   DAB     5      -0.760  45.630  32.250  1.00  0.00
ATOM     57  CA  DAB     5      -0.360  47.270  31.010  1.00  0.00
ATOM     58  CB  DAB     5      -1.620  47.840  30.350  1.00  0.00
ATOM     59  CE  DAB     5      -1.520  49.300  29.890  1.00  0.00
ATOM     60  NZ  DAB     5      -2.490  49.600  28.820  1.00  0.00
ATOM     61  HZ1 DAB     5      -3.450  49.490  29.070  1.00  0.00
ATOM     62  HZ2 DAB     5      -2.330  50.420  28.250  1.00  0.00
ATOM     63  C   DAB     5      -0.100  48.030  32.320  1.00  0.00
ATOM     64  O1  DAB     5      -0.500  47.590  33.400  1.00  0.00
ATOM     65  N   DAB     5       0.620  49.150  32.210  1.00  0.00
ATOM     66  H1  DAB     5       0.720  49.630  33.070  1.00  0.00
ATOM     67  H2  DAB     5       0.480  49.580  31.300  1.00  0.00
TER
ENDMDL
```

Appendix 3: Antibacterial peptide-membrane lipid hydrogen bonds






O17 with E. coli outer at 301 K
RAMP43O6X-DAB7O
RAMP43O6X-DAB6O
RAMP43O6X-DPHE5O
RAMP43O49-DAB6O
RAMP41O4Y-LEU8O1
RAMP41O4Y-DAB2O
LEU8N-RAMP41O4Y
DAB7NZ-RAMP43O49
DAB7NZ-RAMP22O6Z
DAB6NZ-RAMP43O3Y
DAB6NZ-RAMP18O4X
DAB3NZ-RAMP43O6X
Hydrogen bond
0
20
40
60
80
100
Time / ns


O17 with E. coli outer at 310 K
RAMP62O49-LEU4O
RAMP57O3X-DAB2O
LEU4N-RAMP62O3Y
DAB3NZ-RAMP62O2Z
DAB3NZ-RAMP62O49
DDAB1NZ-RAMP61O27
DDAB1NZ-RAMP61O47
Hydrogen bond
0
20
40
60
80
100
Time / ns

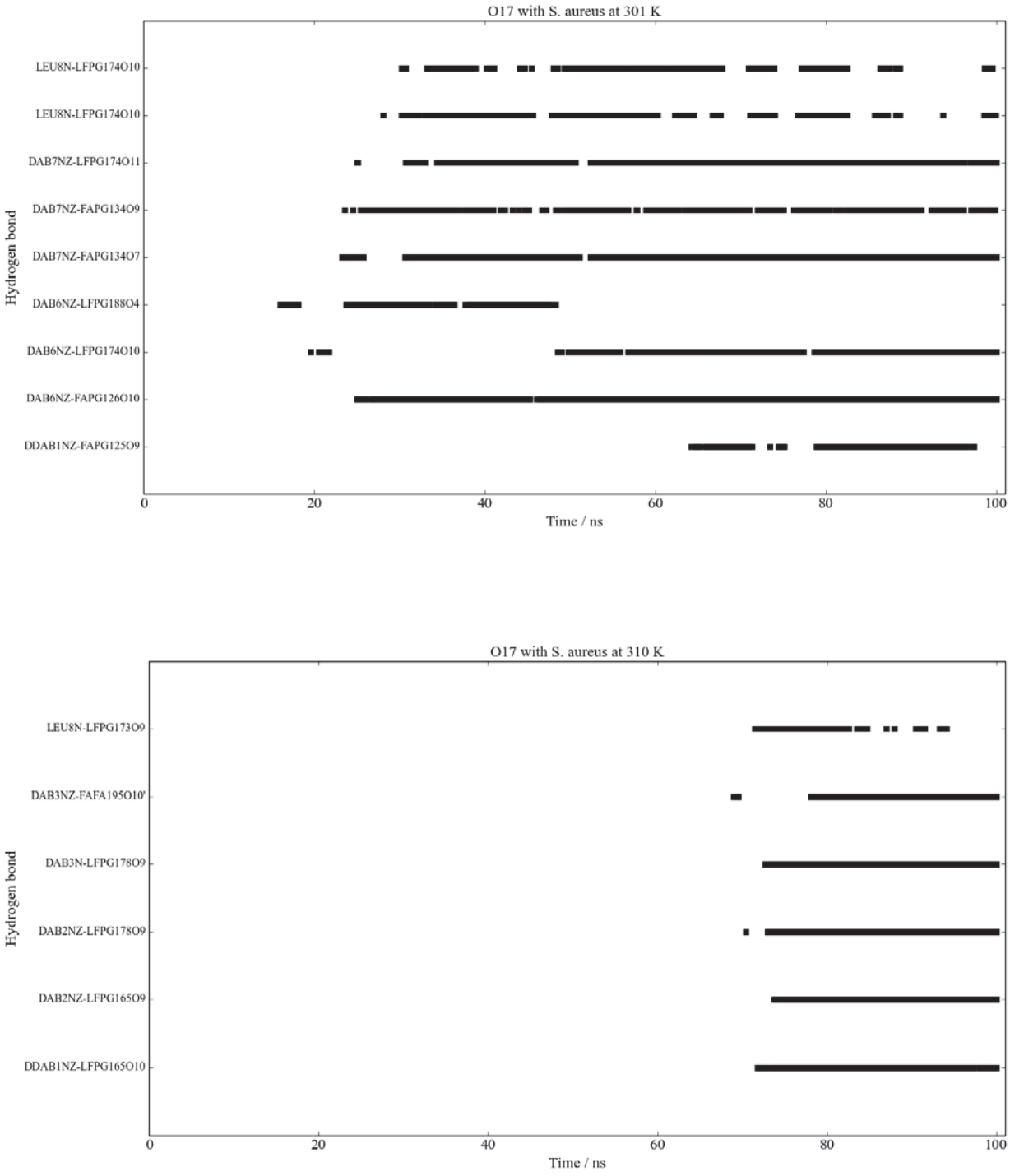
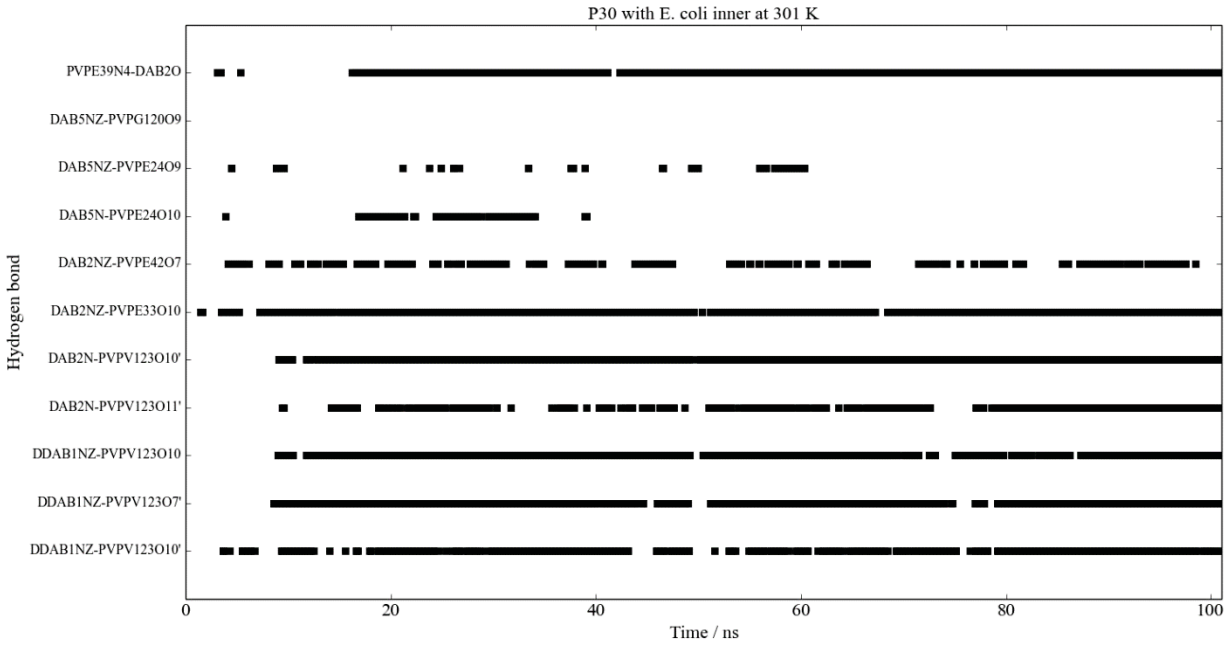
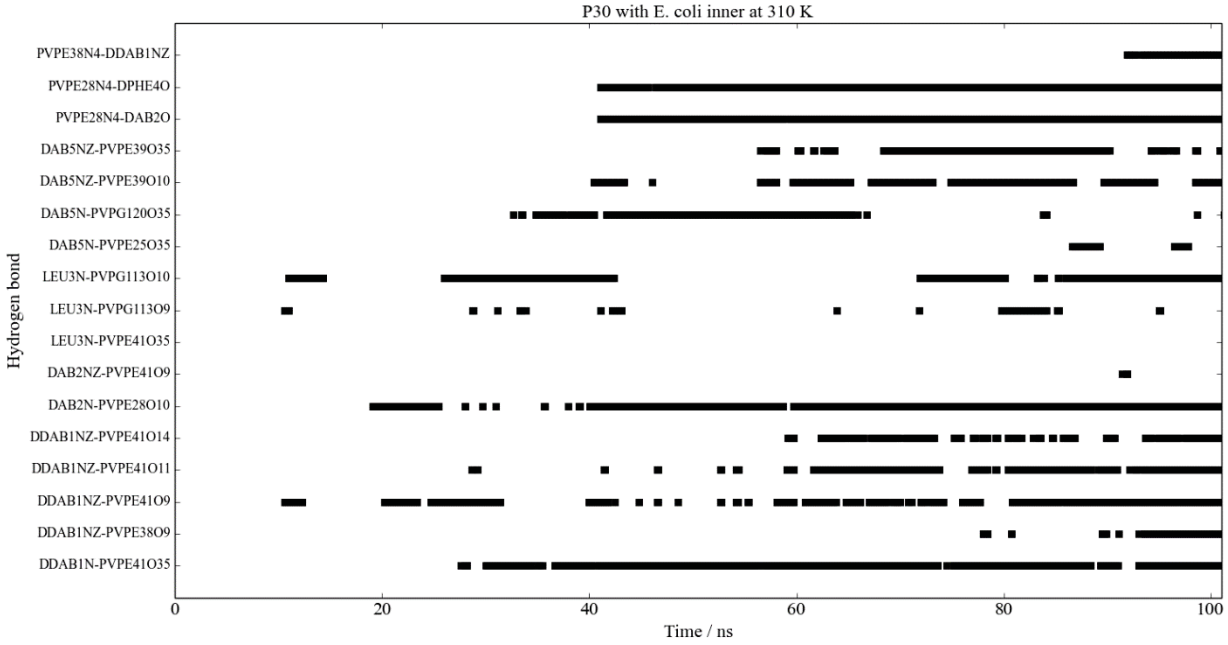
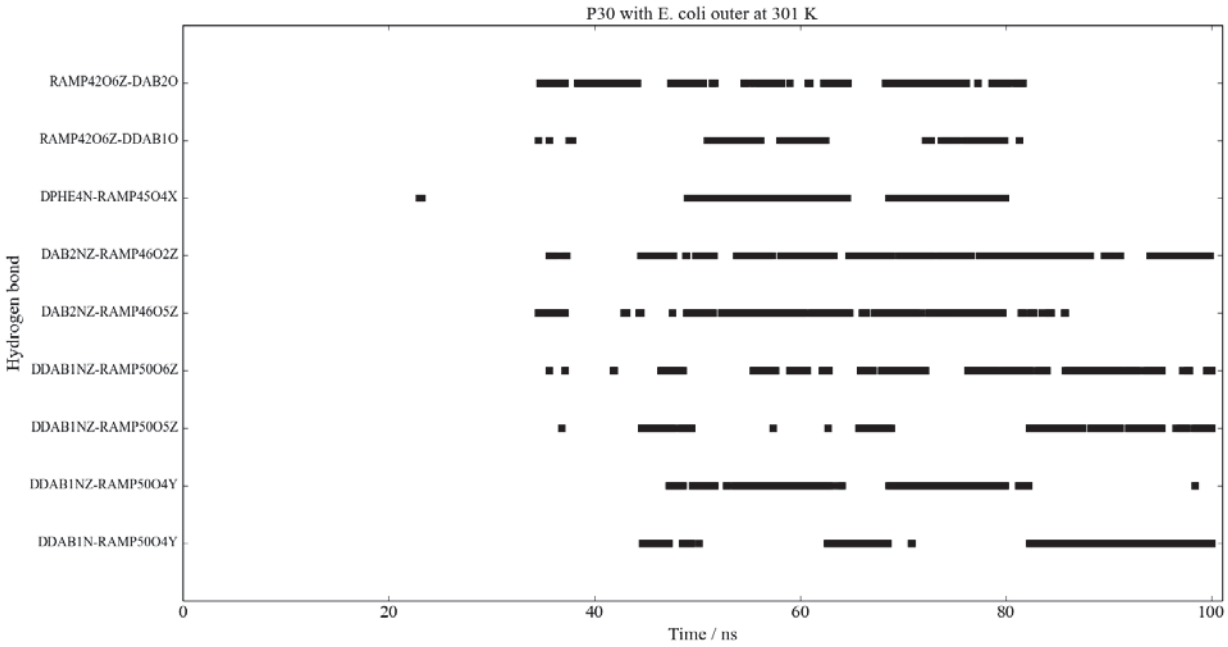
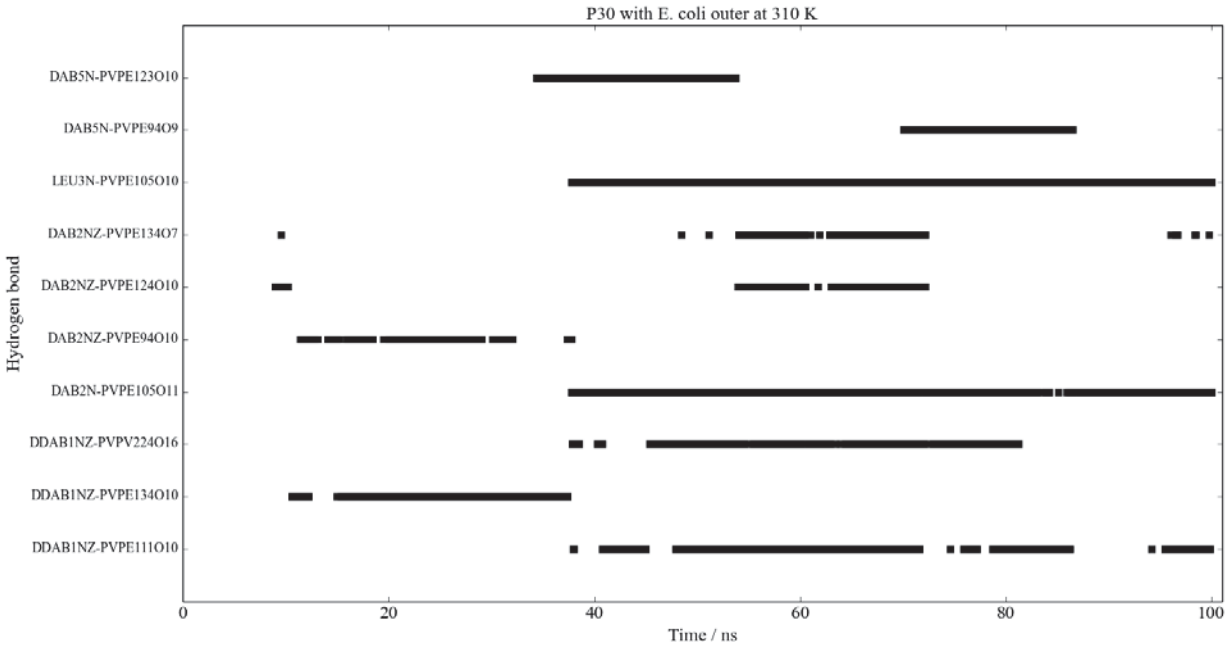


Figure A3.2: Hydrogen bonds between O17 and bacterial membrane lipids for simulations at 301 and 310 K.


P30 with E. coli inner at 301 K
PVPE39N4-DAB2O
DAB5NZ-PVPG120O9
DAB5NZ-PVPE24O9
DAB5N-PVPE24O10
DAB2NZ-PVPE42O7
DAB2NZ-PVPE33O10
DAB2N-PVPV123O10'
DAB2N-PVPV123O11'
DDAB1NZ-PVPV123O10
DDAB1NZ-PVPV123O7'
DDAB1NZ-PVPV123O10'
Hydrogen bond
0
20
40
60
80
100
Time / ns


P30 with E. coli inner at 310 K
PVPE38N4-DDAB1NZ
PVPE28N4-DPHE4O
PVPE28N4-DAB2O
DAB5NZ-PVPE39O35
DAB5NZ-PVPE39O10
DAB5N-PVPG120O35
DAB5N-PVPE25O35
LEU3N-PVPG113O10
LEU3N-PVPG113O9
LEU3N-PVPE41O35
DAB2NZ-PVPE41O9
DAB2N-PVPE28O10
DDAB1NZ-PVPE41O14
DDAB1NZ-PVPE41O11
DDAB1NZ-PVPE41O9
DDAB1NZ-PVPE38O9
DDAB1N-PVPE41O35
Hydrogen bond
0
20
40
60
80
100
Time / ns


P30 with E. coli outer at 301 K
RAMP42O6Z-DAB2O
RAMP42O6Z-DDAB1O
DPHE4N-RAMP45O4X
DAB2NZ-RAMP46O2Z
DAB2NZ-RAMP46O5Z
DDAB1NZ-RAMP50O6Z
DDAB1NZ-RAMP50O5Z
DDAB1NZ-RAMP50O4Y
DDAB1N-RAMP50O4Y
Hydrogen bond
0
20
40
60
80
100
Time / ns


P30 with E. coli outer at 310 K
DAB5N-PVPE123O10
DAB5N-PVPE94O9
LEU3N-PVPE105O10
DAB2NZ-PVPE134O7
DAB2NZ-PVPE124O10
DAB2NZ-PVPE94O10
DAB2N-PVPE105O11
DDAB1NZ-PVPV224O16
DDAB1NZ-PVPE134O10
DDAB1NZ-PVPE111O10
Hydrogen bond
0
20
40
60
80
100
Time / ns

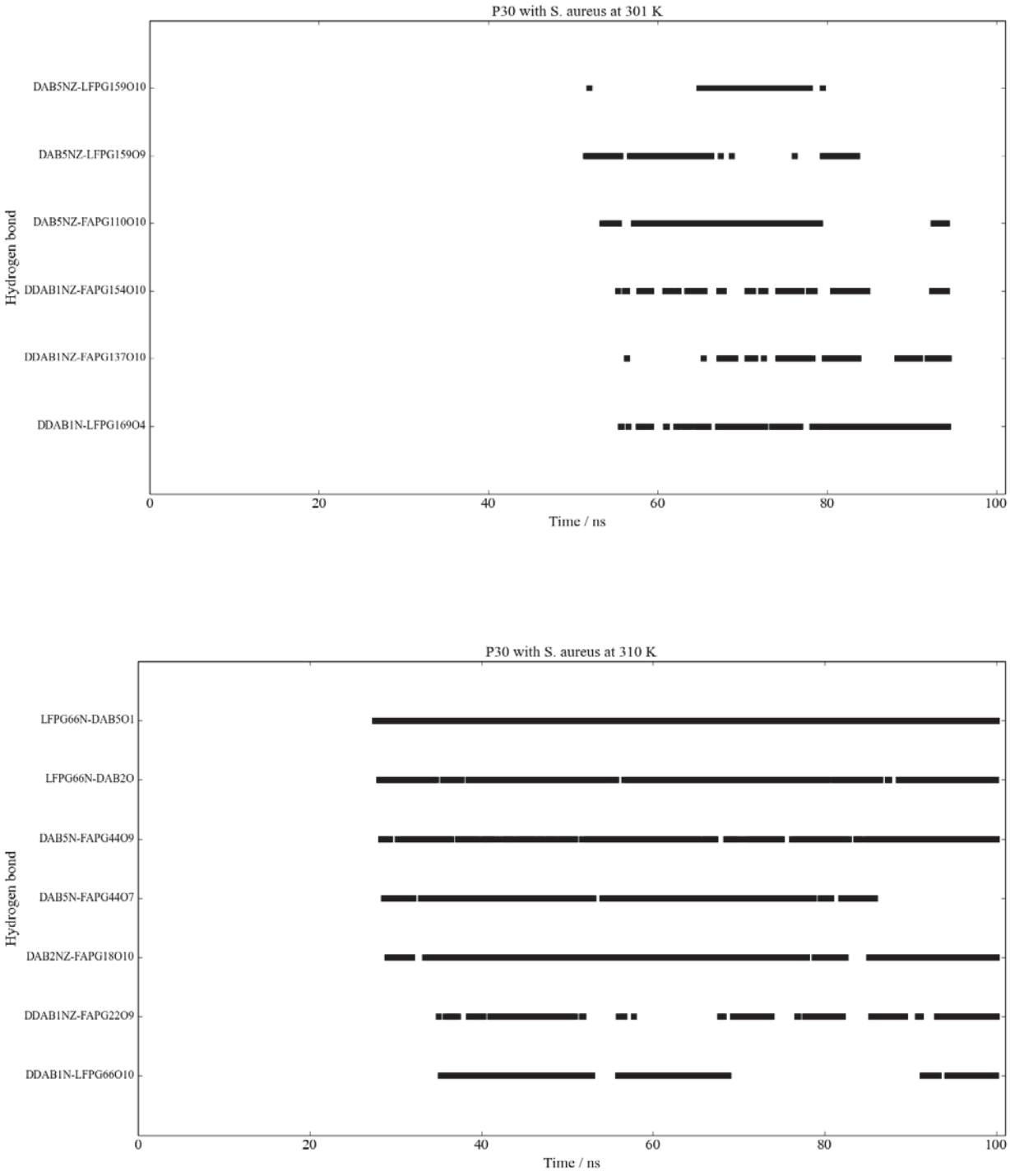
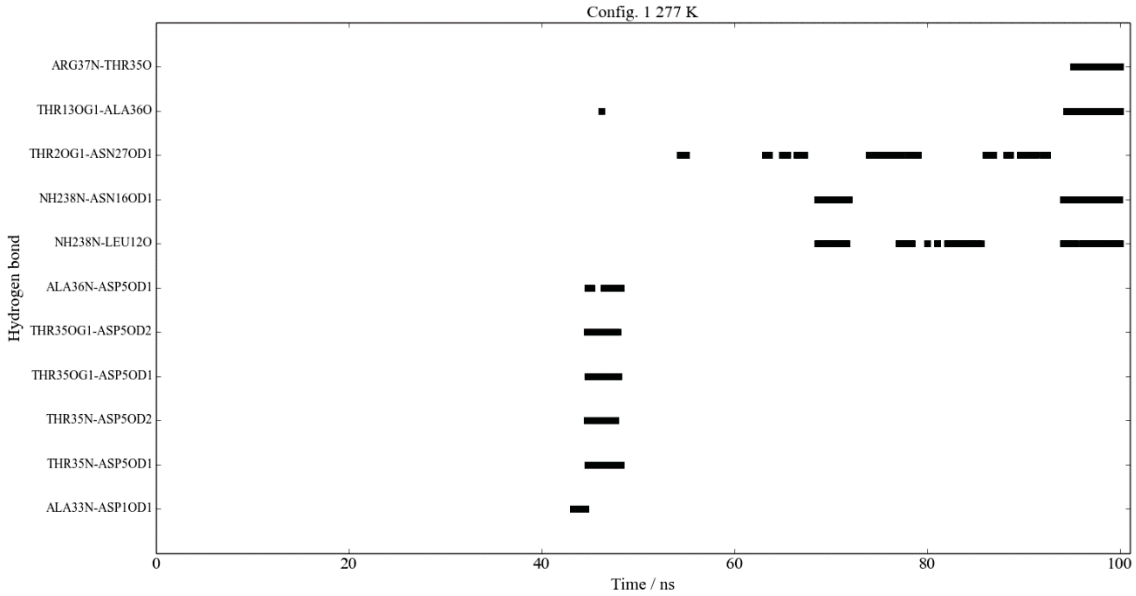
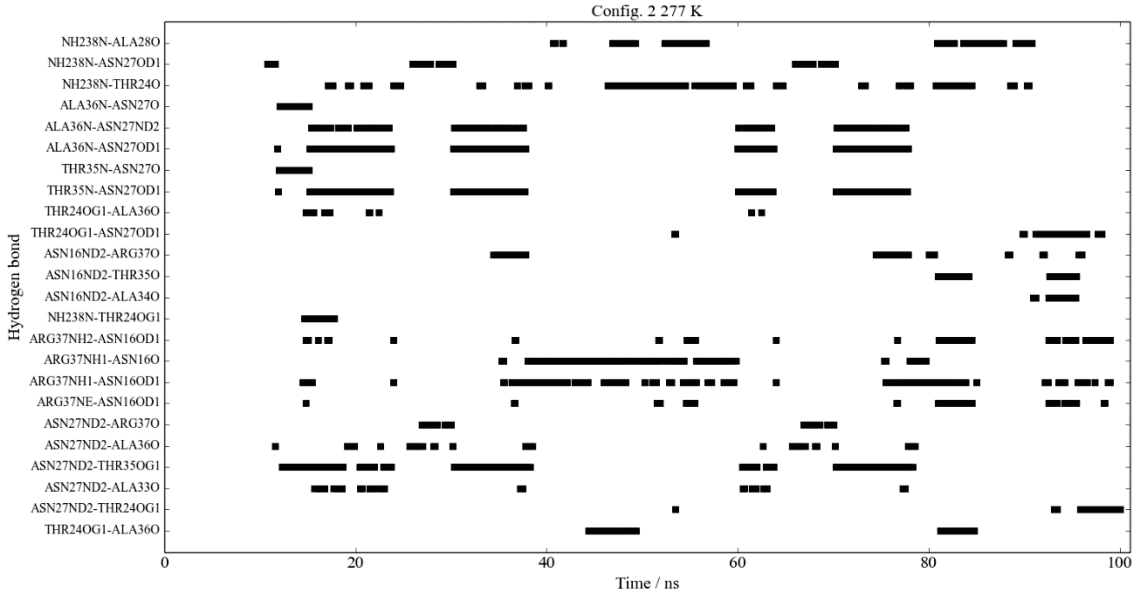
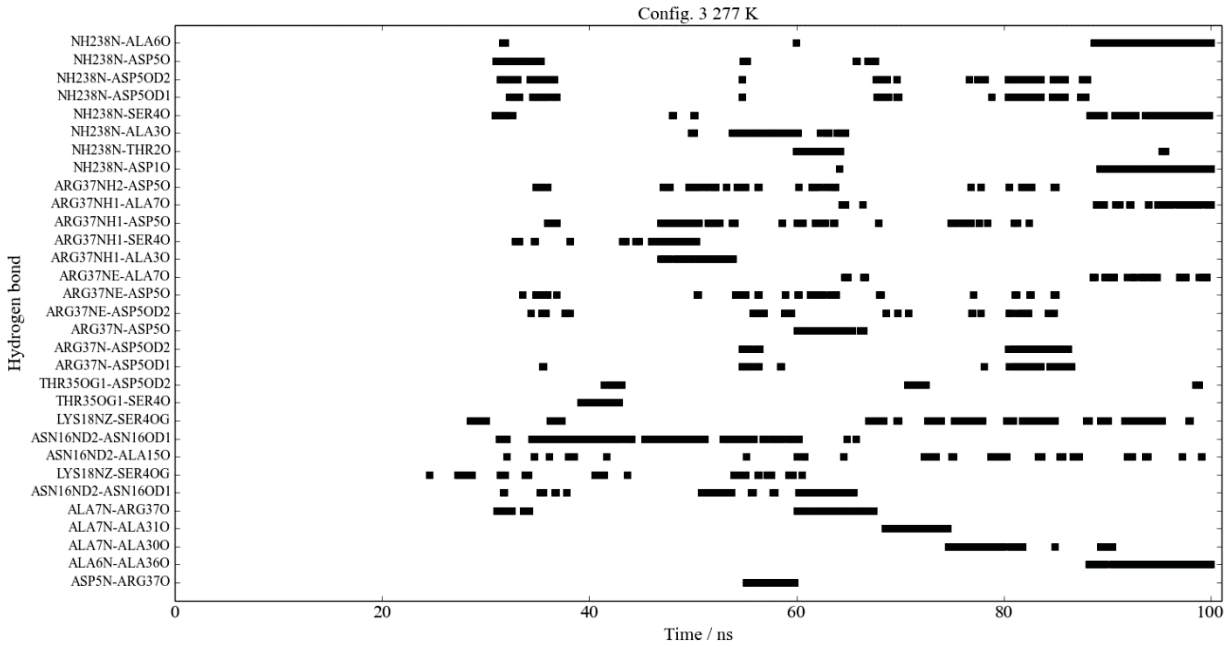


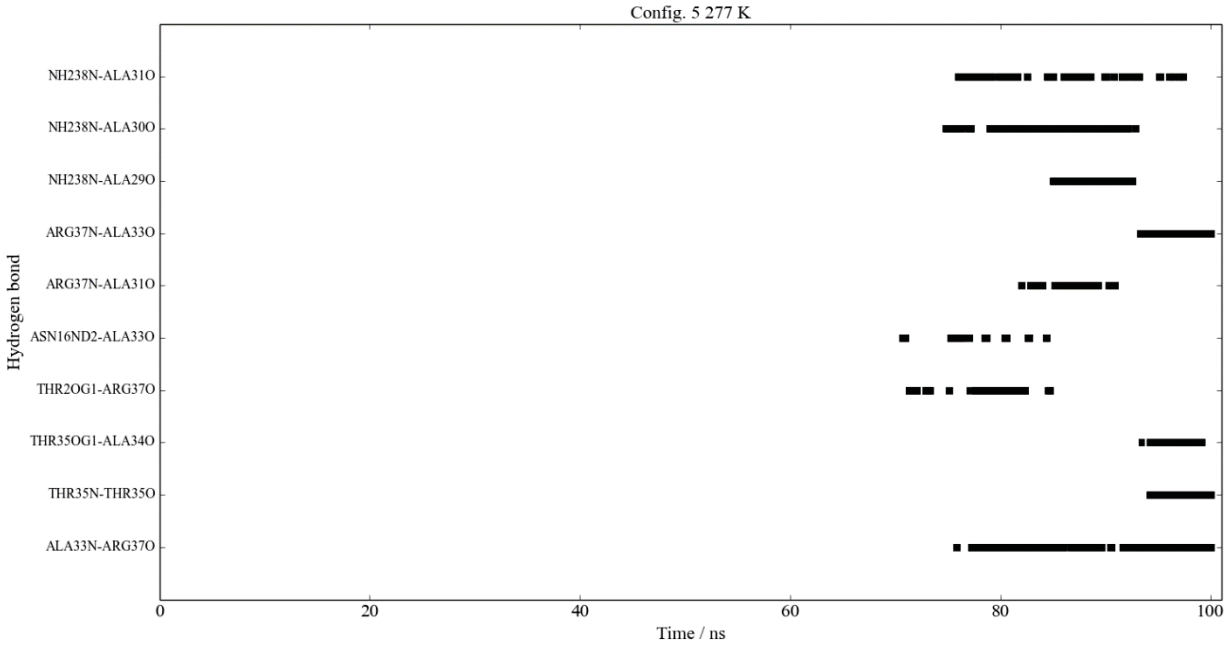
Figure A3.2: Hydrogen bonds between P30 and bacterial membrane lipids for simulations at 301 and 310 K.

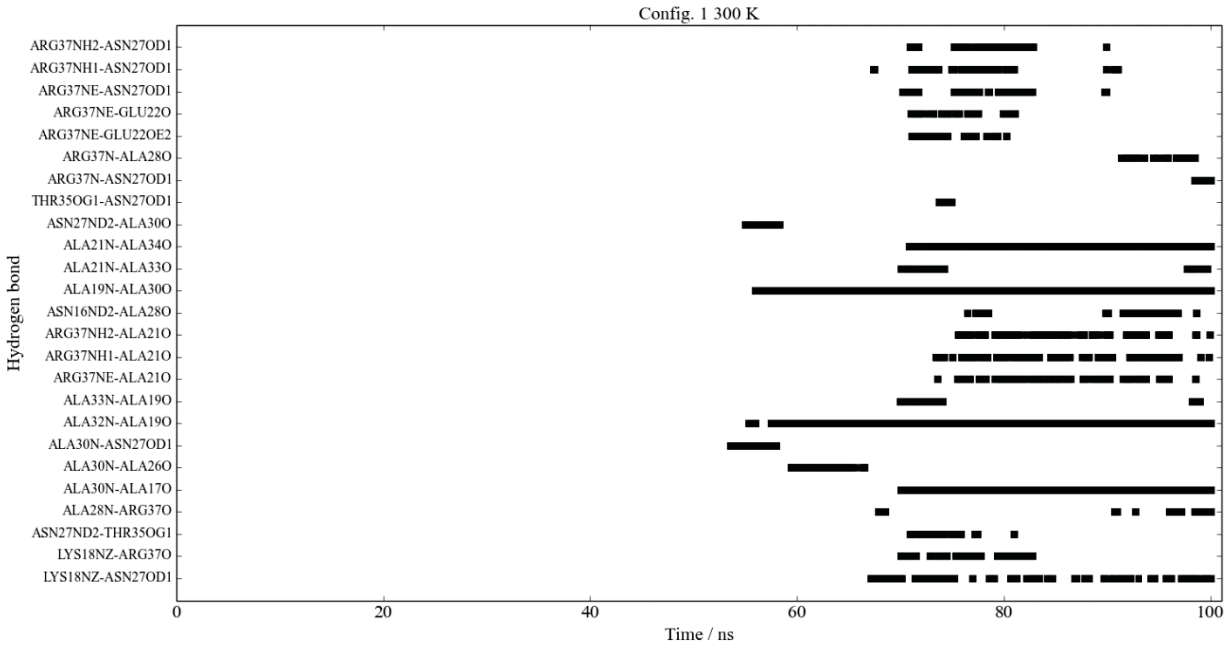
Appendix 4: AFP1 protein inter-subunit hydrogen bonds

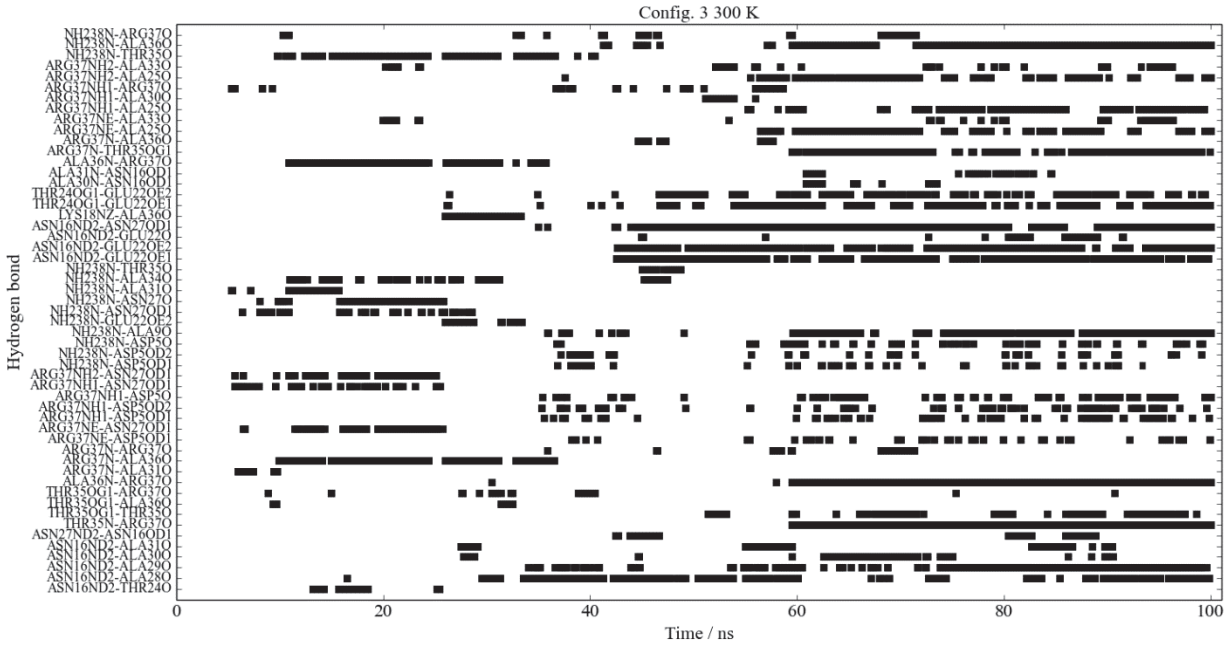





Config. 3 277 K
NH238N-ALA6O
NH238N-ASP5O
NH238N-ASP5OD2
NH238N-ASP5OD1
NH238N-SER4O
NH238N-ALA3O
NH238N-THR2O
NH238N-ASP1O
ARG37NH2-ASP5O
ARG37NH1-ALA7O
ARG37NH1-ASP5O
ARG37NH1-SER4O
ARG37NH1-ALA3O
ARG37NE-ALA7O
ARG37NE-ASP5O
ARG37NE-ASP5OD2
ARG37N-ASP5O
ARG37N-ASP5OD2
ARG37N-ASP5OD1
THR35OG1-ASP5OD2
THR35OG1-SER4O
LYS18NZ-SER4OG
ASN16ND2-ASN16OD1
ASN16ND2-ALA15O
LYS18NZ-SER4OG
ASN16ND2-ASN16OD1
ALA7N-ARG37O
ALA7N-ALA31O
ALA7N-ALA30O
ALA6N-ALA36O
ASP5N-ARG37O
Hydrogen bond
0
20
40
60
80
100
Time / ns


Config. 5 277 K
NH238N-ALA31O
NH238N-ALA30O
NH238N-ALA29O
ARG37N-ALA33O
ARG37N-ALA31O
ASN16ND2-ALA33O
THR2OG1-ARG37O
THR35OG1-ALA34O
THR35N-THR35O
ALA33N-ARG37O
Hydrogen bond
0
20
40
60
80
100
Time / ns


Config. 1 300 K
ARG37NH2-ASN27OD1
ARG37NH1-ASN27OD1
ARG37NE-ASN27OD1
ARG37NE-GLU22O
ARG37NE-GLU22OE2
ARG37N-ALA28O
ARG37N-ASN27OD1
THR35OG1-ASN27OD1
ASN27ND2-ALA30O
ALA21N-ALA34O
ALA21N-ALA33O
ALA19N-ALA30O
ASN16ND2-ALA28O
ARG37NH2-ALA21O
ARG37NH1-ALA21O
ARG37NE-ALA21O
ALA33N-ALA19O
ALA32N-ALA19O
ALA30N-ASN27OD1
ALA30N-ALA26O
ALA30N-ALA17O
ALA28N-ARG37O
ASN27ND2-THR35OG1
LYS18NZ-ARG37O
LYS18NZ-ASN27OD1
Hydrogen bond
0
20
40
60
80
100
Time / ns


Config. 3 300 K
NH238N-ARG37O
NH238N-ALA36O
NH238N-THR35O
ARG37NH2-ALA33O
ARG37NH2-ALA25O
ARG37NH1-ARG37O
ARG37NH1-ALA30O
ARG37NH1-ALA25O
ARG37NE-ALA33O
ARG37NE-ALA25O
ARG37N-ALA36O
ARG37N-THR35OG1
ALA36N-ARG37O
ALA31N-ASN16OD1
ALA30N-ASN16OD1
THR24OG1-GLU22OE2
THR24OG1-GLU22OE1
LYS18NZ-ALA36O
ASN16ND2-ASN27OD1
ASN16ND2-GLU22O
ASN16ND2-GLU22OE2
ASN16ND2-GLU22OE1
NH238N-THR35O
NH238N-ALA34O
NH238N-ALA31O
NH238N-ASN27O
NH238N-ASN27OD1
NH238N-GLU22OE2
NH238N-ALA9O
NH238N-ASP5O
NH238N-ASP5OD2
NH238N-ASP5OD1
ARG37NH2-ASN27OD1
ARG37NH1-ASN27OD1
ARG37NH1-ASP5O
ARG37NH1-ASP5OD2
ARG37NH1-ASP5OD1
ARG37NE-ASN27OD1
ARG37NE-ASP5OD1
ARG37N-ARG37O
ARG37N-ALA36O
ARG37N-ALA31O
ALA36N-ARG37O
THR35OG1-ARG37O
THR35OG1-ALA36O
THR35OG1-THR35O
THR35N-ARG37O
ASN27ND2-ASN16OD1
ASN16ND2-ALA31O
ASN16ND2-ALA30O
ASN16ND2-ALA29O
ASN16ND2-ALA28O
ASN16ND2-THR24O
Hydrogen bond
0
20
40
60
80
100
Time / ns

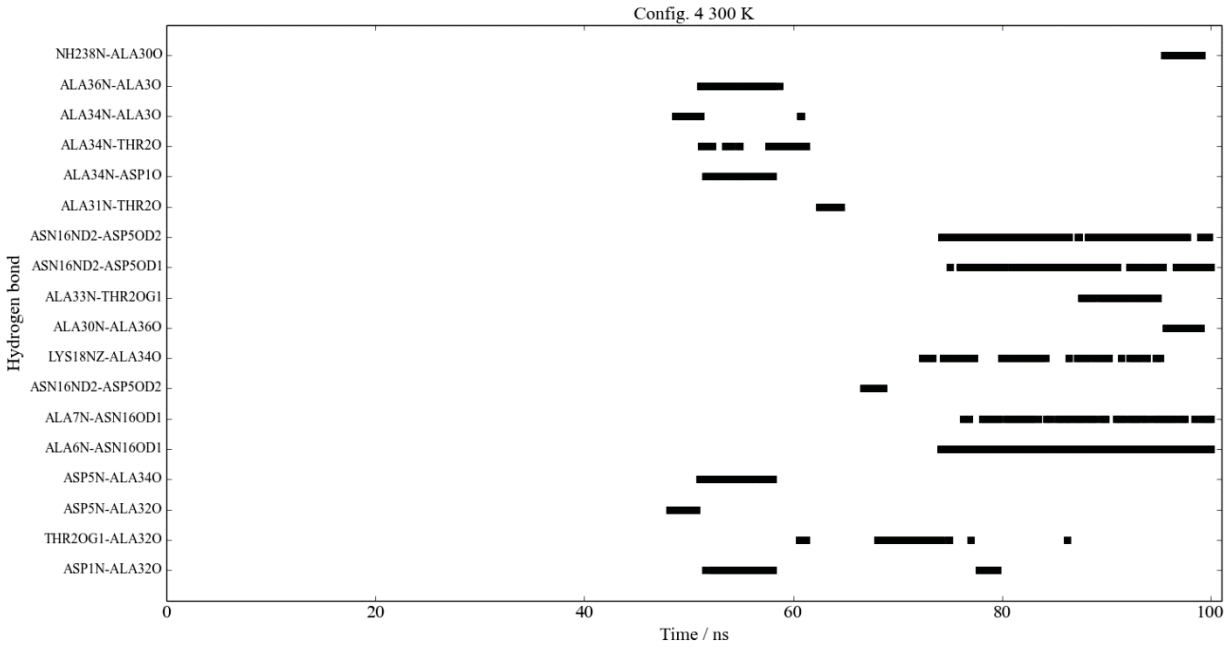
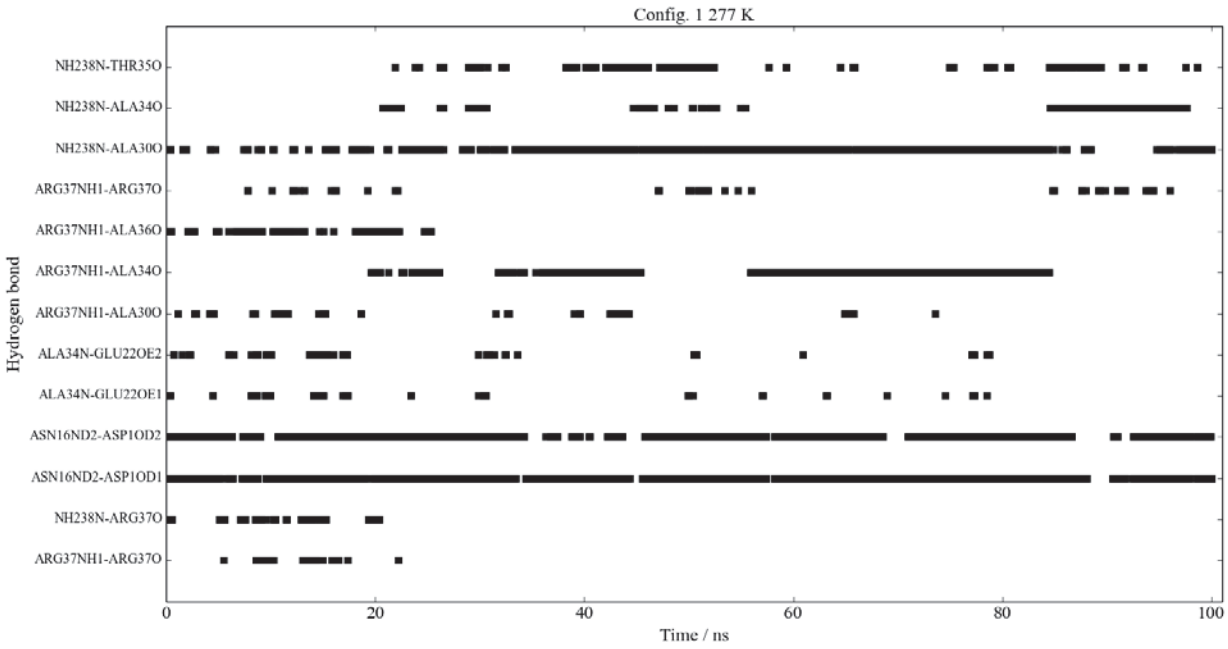


Figure A4.1: Hydrogen bonds between subunits A and B of AFP1 by residue as a time series for simulations at 277 K and 300 K. Graphs for configuration 4, 277 K and configurations 2 and 5 at 300 K have been omitted as no hydrogen bond formation occurred.



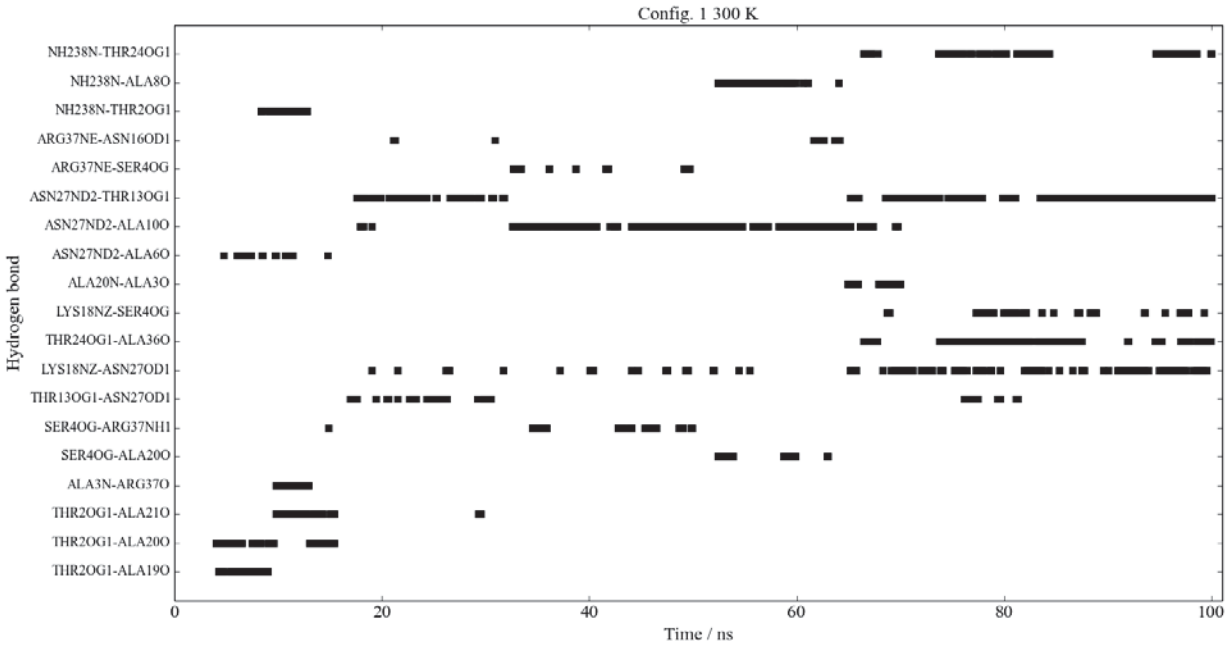
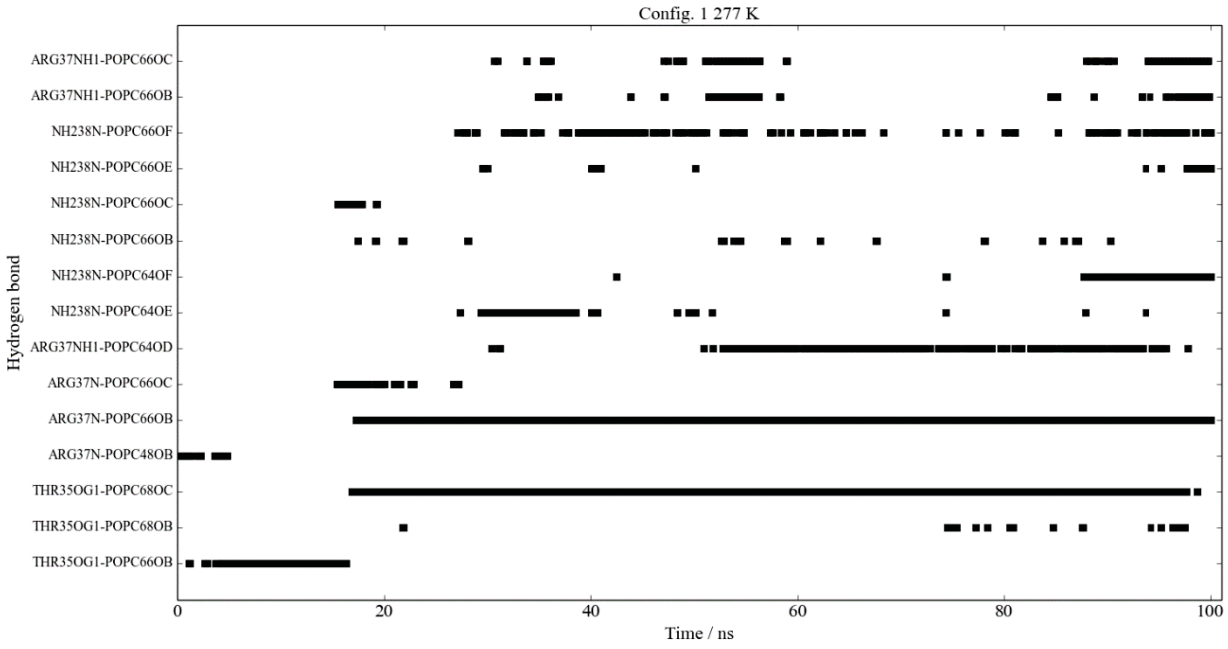


Figure A4.2: Hydrogen bonds between subunits A and B of AFP1 by residue as a time series for simulations with POPC bilayer at 277 K and 300 K.



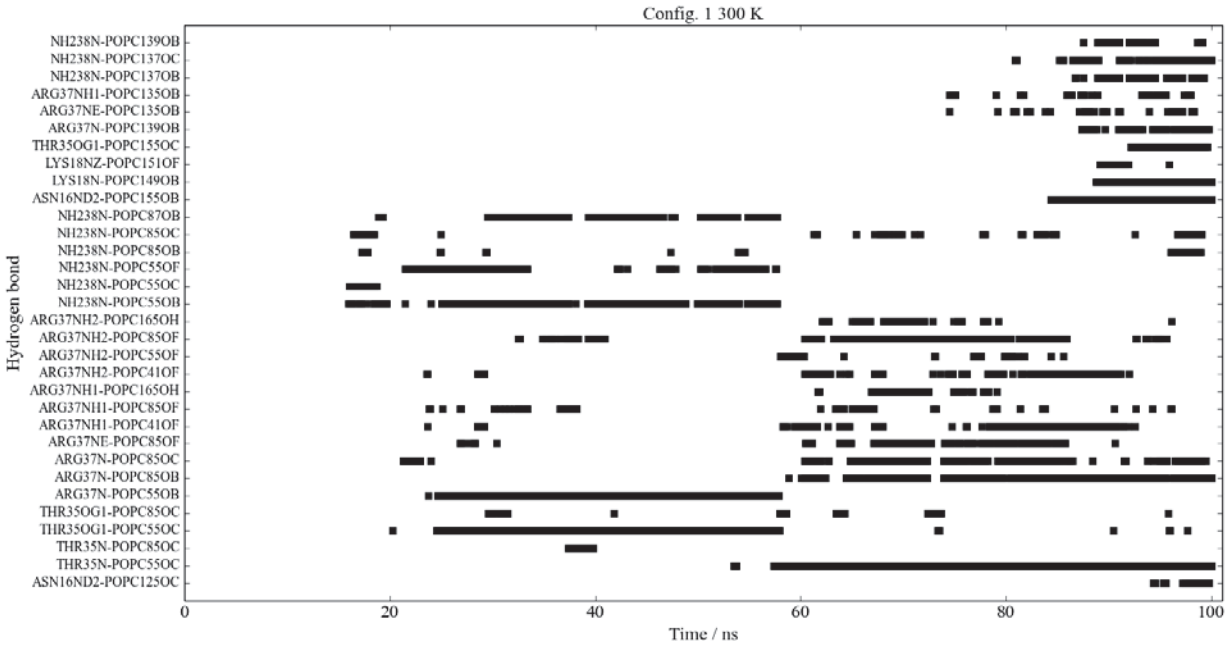


Figure A4.3: Hydrogen bonds between AFP1 and POPC bilayer by residue as a time series for simulations with POPC bilayer at 277 K and 300 K.

Appendix 5: CG-DPPC GROMOS parameters

GROMOS 54a8 building block for CG-DPPC. All CG-DPPC forcefield parameters can be found in Chapter 4.

```
MTBUILDBLCGSOLUTE
# building block (residue, nucleotide, etc.)
# RNME (CG-DPPC)
DPCG
# number of atoms, number of preceding exclusions
# NMAT NLIN
    19    0
# preceding exclusions
#ATOM                                MAE MSAE
# atoms
#ATOM ANM  IACM MASS        CGMICGM MAE MSAE
    1 CC      57   49    1.0800   0   4     2 3 4 5
    2 CD      58   25   -0.0800   1   3     3 4 5
    3 PC      59   82   -1.1000   0   6     4 5 6 7 8 9
    4 PD      60   41    0.1000   1   5     5 6 7 8 9
    5 GC      61   83    1.2000   0   6     6 7 8 9 10 15
    6 E1C     63   85   -1.0850   0   5     7 8 9 10 11
    7 E1D     64   86    0.5850   0   4     8 9 10 11
    8 E2C     65   87   -1.2300   0   3     9 15 16
    9 E2D     66   88    0.5300   1   2    15 16
   10 C3M     71   103   0.0000   0   2    11 12       #tail1
   11 C3M     71   103   0.0000   0   2    12 13
   12 C2M     69   101   0.0000   0   2    13 14
   13 C3M     71   103   0.0000   0   1    14
   14 C3E     72   104   0.0000   1   0
   15 C3M     71   103   0.0000   0   2    16 17        #tail2
   16 C3M     71   103   0.0000   0   2    17 18
```

```
   17 C2M     69   101   0.0000   0   2    18 19
   18 C3M     71   103   0.0000   0   1    19
   19 C3E     72   104   0.0000   1   0
# CG regions
# NCG
    1
# start end massscale
    1    19    1
# bonds
#  NB
    14
#  IB   JB  MCB
    1    3   54
    3    5   57
    5    6   61
    5    8   59
    6    10  71  #e1 -t1 c4-c3?
    8    15  71 #e2 -t2
    10   11  68
    11   12  70
    12   13  70
    13   14  68
    15   16  68
    16   17  70
    17   18  70
    18   19  68
# cg-dp bonds?
#  NCGB
   4
#  IB  JB   MCB
    1   2    55
```

```
    3   4    56
    6   7    60
    8   9    62
# bond angles
# NBA
  12
#  IB   JB   KB   MCB
    1   3    5    55  #CPG
    3   5    6    56  #PGE1
    3   5    8    57  #PGE2
    6   5    8    58  #E1GE2
    5   6    10   69  #GE1C3
    5   8    15   70  #GE2C3
    6  10    11   71  #E1C3C3
    8  15    16   72  #E2C3C3
   10  11    12   62  #tail 1
   11  12    13   64
   12  13    14   62
   15  16    17   62  #tail2
   16  17    18   64
   17  18    19   62
# improper dihedrals
# NIDA
    0
#  IB   JB   KB   LB  MCB
# dihedrals
# NDA
   13
   1    3   5   6   46
   1    3   5   6   48 #cpge1
   1    3   5   8   47
```

```
   3    5   6   10  54  #PGE1C3
   3    5   8   15  55  #PGE2C3
   5    6   10  11  56  #GE1C3C3
   5    8   15  16  57  #GE2C3C3
   6    10  11  12  49  #E1C3C3C2
   8    15  16  17  49  #E2C3C3C2
  10    11  12  13  49  #t1
  11    12  13  14  49  #t1
  15    16  17  18  49  #t2
  16    17  18  19  49  #t2
#  IB   JB   KB   LB  MCB
# LJ exceptions
# NEX
    0
END
```

Appendix 6: Common lipid structures

Biological cell membranes typically comprise phospholipid bilayers. Phosphatidylcholine (e.g. 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), *Fig. A6.1,* and 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), *Fig. A6.2*) and phosphatidylethanolamine (e.g. 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE), *Fig. A6.3*) lipids typically form the main component of the model membranes used in this thesis. Note the difference in lipid tails of POPC and DPPC – despite both lipids having the same headgroup, DPPC has two saturated carbon tails, while POPC contains one mono-unsaturated tail and one saturated tail. Other structures of biologically relevant lipids included in the *E. coli* and *S. aureaus* membranes are provided: the phosphatidylglycerol 1,2-dipalmitoyl-*sn*-glycero-3-phosphoglycerol (DPPG, *Fig. A6.4*), lysyl-DPPG (*Fig. A6.5*), the phosphatidylserine 1,2-dipalmitoyl-*sn*-glycero-3-phosphoserine (DPPS, *Fig. A6.6*), and lipopolysaccharide (LPS, *Fig. A6.7*). Images are adapted from the Avanti Lipids Catalogue[108].

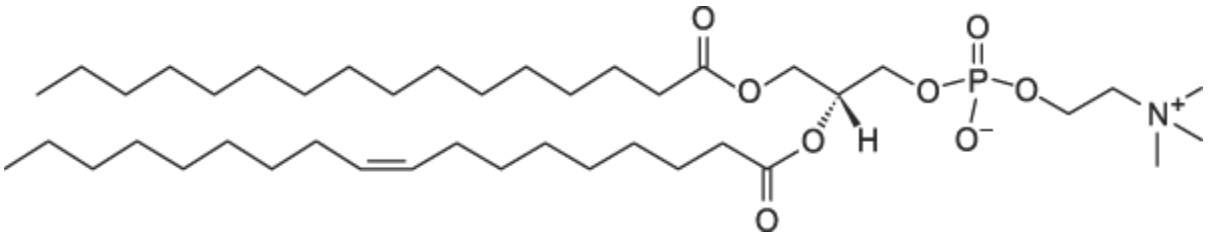


*Figure A6.1: Unsaturated phosphatidylcholine POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine).*

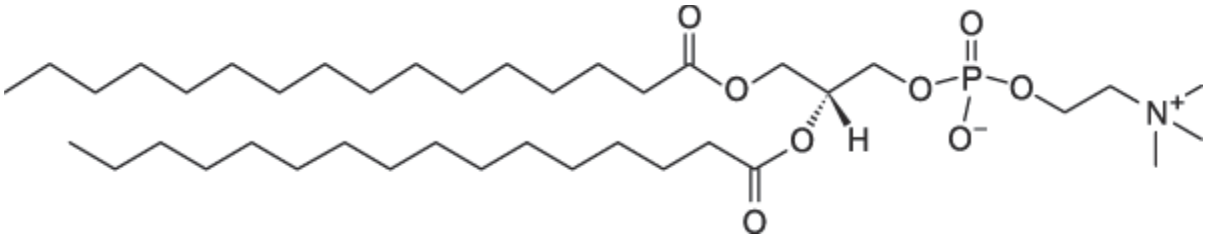


*Figure A6.2: Saturated phosphatidylcholine DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine).*

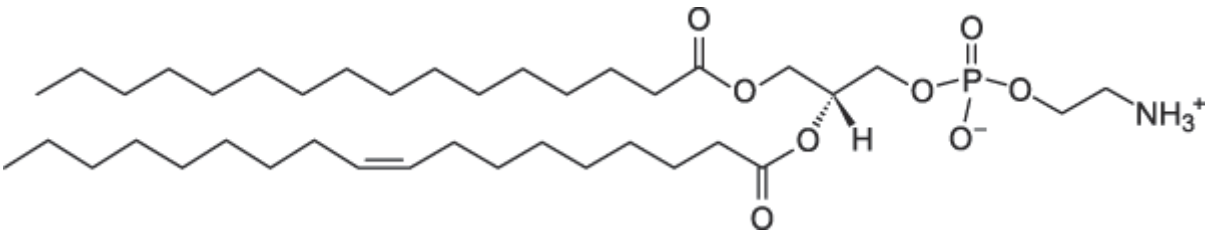


*Figure A6.3: Unsaturated phosphatidylethanolamine POPE (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine).*

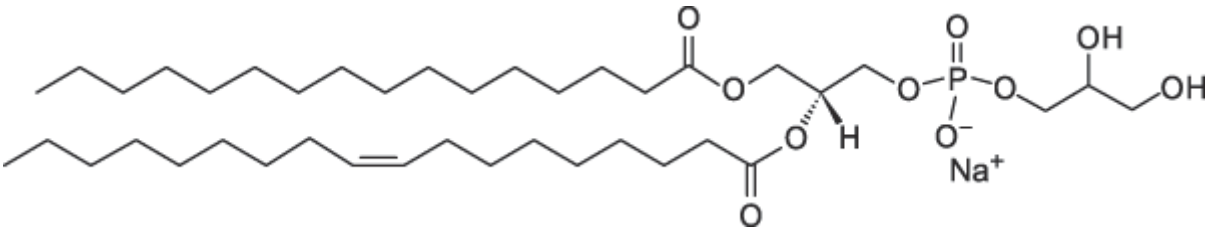


*Figure A6.4: Unsaturated phosphatidylglycerol POPG (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) sodium salt).*

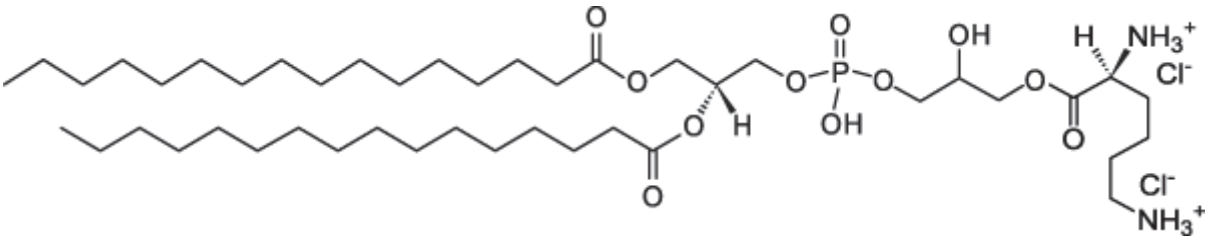


*Figure A6.5: Saturated phosphatidylglycero-lysine DPPG-Lys (1,2-dipalmitoyl-*sn*-glycero-3-[phospho-rac-(3-lysyl(1-glycerol))] chloride salt).*

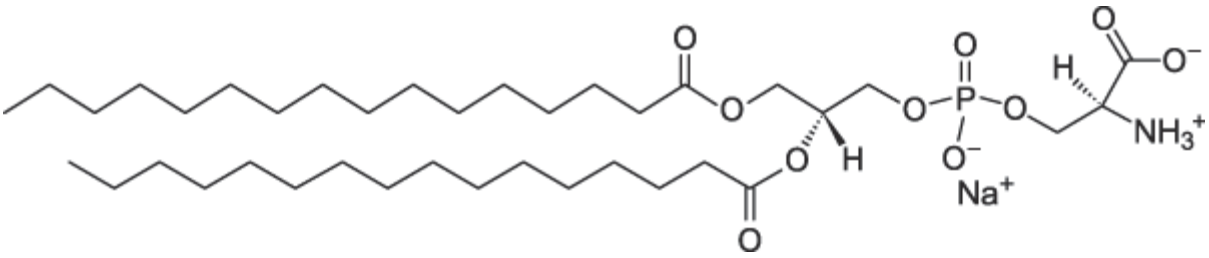


*Figure A6.6: Saturated phosphatidylserine DPPS 1,2-dipalmitoyl-*sn*-glycero-3-phospho-L-serine (sodium salt).*

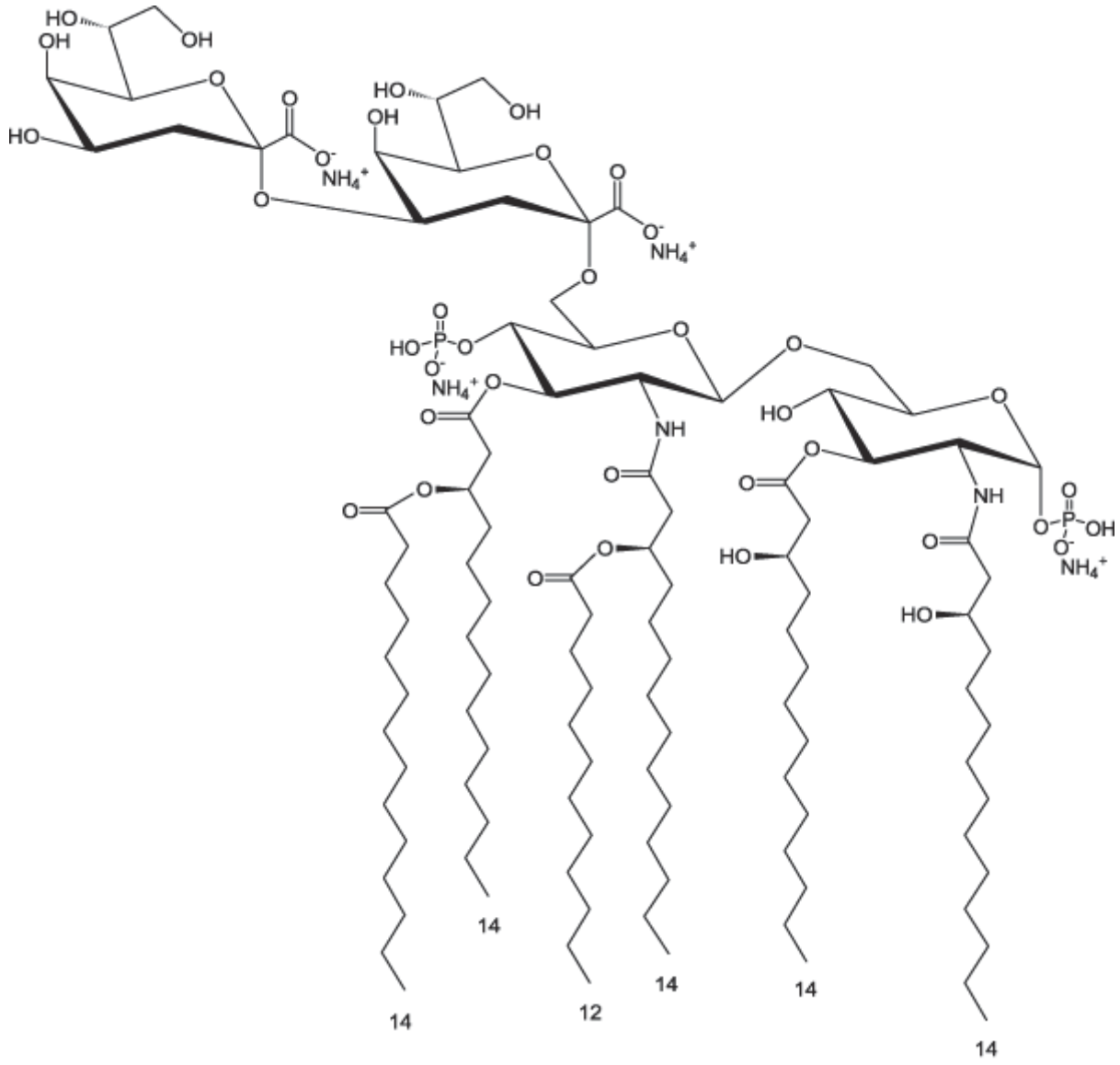


Figure A6.7: Example structure of Lipid A LPS. Structure is one of the many possible arrangements of a typical lipopolysaccharide in the bacterial outer membrane.