Environment Induced Conformational Transition in Peptides

by

Cahit Dalgıçdir

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy

in

Computational Sciences & Engineering

Koç University

August, 2014

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

Cahit Dalgıçdir

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Assist. Prof. Mehmet Sayar

_____

Assist. Prof. Alkan Kabakçıoğlu

_____

Prof. Özlem Keskin

_____

Prof. Burak Erman

_____

Assoc. Prof. Banu Özkan

Date: _____

To my wife and my family

ABSTRACT

Proteins are fascinating molecular machines with their ability to fold into unique 3-dimensional structures identified as their native states. Despite their surprisingly robust folding capability, proteins and peptides exhibit a strong tendency to form ordered aggregates if the environmental conditions are correctly tuned. The amphiphilic nature of peptides plays an important role in enabling aggregation in aqueous environment or at interfaces and surfaces or by allowing peptides to penetrate through or aggregate in membranes. In many cases the aggregation or the interaction of a peptide with a hydrophobic/hydrophilic interface triggers a conformational change in the molecule, which is usually coupled to the partitioning of the hydrophobic/hydrophilic residues of the peptide. Well known examples of the interplay of conformational change and aggregation or partitioning at interfaces are the misfolding of proteins upon amyloid aggregation, or more generally the induction of higher β-sheet content by aggregation or by the presence of an interface. In order to better understand and ultimately control structure formation in peptide aggregates and peptide-based materials, knowledge of the relevant interactions, driving forces, pathways and assembly mechanisms is essential.

In this thesis we utilize molecular dynamics simulations to provide microscopic structural and thermodynamic insight into the interplay of folding, aggregation and partitioning in peptide based systems. In order to illustrate environment driven conformational change, the first model system we have focused on is phenylalanine dipeptide (FF). With its only two aminoacid long sequence, this molecule forms one-dimensional nanotubes, in which the molecules adopt a cis-like conformation unlike their preferred state in water. Here, by analyzing molecular dynamics simulations of FF in bulk water and cyclohexane/water interface, we demonstrate how the hydropho-

bic/hydrophilic interface triggers the trans-to-cis conformational change. Moreover, we demonstrate that even a molecular interface can lead to a similar conformational change, and discuss the similarities and differences between macroscopic and molecular interfaces. Next, in order to overcome the time and length scale barriers in observing aggregation of peptides in molecular simulations, we develop a coarse-grained (CG) model capable of representing the conformational behavior of FF. Our CG model is unique in its ability to capture the correct representation of the target molecule in two different environments. We show that correct representation of a structural change, such as a trans-to-cis conformational switch, relies on thermodynamic driving forces. Hence, a solvation free energy based tuning is required to capture the correct partitioning behavior.

In the second study we switch to the LK peptide which is a designed synthetic molecule. We demonstrate how the interplay of hydrogen bonding, hydrophobic interactions, and electrostatics leads to an intrinsically disordered peptide. When isolated in bulk water it lacks a well defined secondary structure and only in the presence of a macroscopic or molecular interface its targeted α-helical secondary structure can be realized. In the case of LK the presence of an interface leads to a population shift in the conformational phase space of the molecule. We also calculate the potential of mean force as a function of aggregate size and demonstate that in agreement with experimental findings tetramers of LK are the stable form in solution.

Our findings highlight the challenges associated with the coupled nature of aggregation, folding and partitioning for peptides. We show that molecular dynamics simulations provide atomistic resolution analysis of the driving forces for such phenomena, perfectly complementing experimental techniques.

ÖZETÇE

Birer moleküler makine olan proteinler, üç boyutlu doğal yapılarına katlanmadaki yeteneklerine rağmen, çevresel etkiler doğru ayarlandığı takdirde dizili birikimler oluştururlar. Solüsyon içinde, yüzeylerde veya arayüzeylerde görülen birikimler ya da etkileşimlerde peptitlerin amfifilik doğasının rolü önemlidir. Bu rol hücre zarlarından geçişte veya bu zarlarda birikmede de görülebilir. Çoğu durumda, peptitlerin birikmesi veya hidrofobik/hidrofilik arayüzeylerle etkileşimi, moleküllerin şekillerinde değişime sebep olur. Bu durum genelde hidrofobik/hidrofilik rezidülerin ayrılmasıyla birlikte görülür. Arayüzeylerde görülen konformasyon değişimi, birikme ve ayrılmanın en bilinir örneği amiloid birikimiyle görülen proteinlerdeki hatalı katlanmadır. Ya da daha genel tabirle arayüzey veya birikme etkisiyle yüksek miktarda beta plaklarının oluşumudur. Hem peptit birikmelerindeki yapının hem de peptit-tabanlı malzemelerin oluşumunu kontrol edebilmek ve bu süreçleri daha iyi anlamak için ilgili etkileşimlerin, harekete geçirici kuvvetlerin, yolakların ve birleşim mekanizmalarını anlamak şarttır.

Bu tezde, peptit tabanlı sistemlerde katlanma, birikim ve ayrılmanın birbirleriyle etkileşimine mikroskopik yapısal ve termodinamik açıdan bakma amacıyla moleküler dinamik simülasyonlarını kullandık. Çevrenin etkisiyle gerçekleşen yapısal değişiklikleri göstermek amacıyla ilk model sistemimiz olan fenilalanin dipeptidi (FF) üzerine odaklandık. Sadece iki aminoasit uzunluğundaki sekansına sahip olmasına rağmen tek boyutlu nanotüp oluşturabiliyor. Bu nanotüplerdeki moleküllerde su yığını içindeki yapıların aksine cis-benzeri yapılar görülmektedir. Hem su yığınındaki hem de siklohekzan/su arayüzeyindeki FF molekülünü moleküler dinamik simülasyonları ile tahlil ederek hidrofobik/hidrofilik arayüzeylerin transtan cise yapısal değişimi tetiklediklerini gösterdik. Dahası, moleküler arayüzeyin bile

benzer yapısal değişiklikler yaratacağını kanıtladık ve makroskopik ve moleküler arayüzeyler arasındaki farkları ve benzerlikleri tartıştık. Ardından, zaman ve ebat ölçeklerindeki engelleri aşarak moleküler simülasyonlar yardımıyla peptitlerdeki birikimi gözlemlemek için, düşük çözünürlüklü (DÇ) model geliştirdik. Bu DÇ model FFde görülen farklı yapısal davranışları temsil edebilmekte, yani FFnin iki farklı ortamdaki yapısal farklılıklarını doğru bir şekilde betimleyebilmektedir. Transtan cise yapısal değişim gibi süreçleri doğru temsil edebilmenin termodinamik kuvvetlerle yakından alakalı olduğunu gösterdik. Doğru ayrılma davranışını temsil edebilmek için de çözünme serbest enerjilerini kullanmanın gerekliliğini gösterdik.

İkinci çalışmada tasarlanan sentetik bir molekül olan LK peptitlerine geçtik. Burada, hidrojen bağları, hidrofobik ve elektrostatik etkileşimlerin içsel olarak dağınık peptite sebep olduğunu ortaya koyduk. Tek başına su yığını içinde belirli bir ikincil yapıya sahip olmayan bu molekülün makroskopik ve moleküler arayüzeylerde hedeflenen alfa-heliks ikincil yapısını geçtiği görüldü. Böylece arayüzeyin varlığı molekülün yapısal faz uzayında bir popülasyon kayması yaratmış oldu. Ortalama kuvvetin potansiyelini birikim ebatının bir fonksiyonu olarak hesapladık ve deneysel çalışmalarla uyumlu olarak LK peptidinin solüsyon içinde dörtlü bir yapıda stabil kaldığını gözlemledik.

Bulgularımız peptitlerde birikim, katlanma ve ayrılmanın birlikteliğinin doğasına ilişkin zorluklara vurgu yaptı. Moleküler dinamik simülasyonlarının atomistik ölçekte bu fenomenleri harekete geçiren kuvvetlerin tahlilini sağladığı ve bunların deneysel teknikleri tamamladığını göstermiş olduk.

ACKNOWLEDGMENTS

I would like to thank my advisor Mehmet Sayar for providing insight and his patience throughout my PhD studies. He also taught me how to approach the problems and pay attention to the small details which usually turn out to be important.

I am grateful to the thesis committee members, Asisst. Prof. Alkan Kabakçıoğlu, Prof. Özlem Keskin, Prof. Burak Erman and Assoc. Prof. Banu Özkan for spending their valuable time reading this thesis and participating in my defence. I also thank Assoc. Prof. Seda Keskin for her advices during my studies.

I would like to thank Prof. Christine Peter for her collaboration and valuable advices for both the diphenylalanine and the LK projects.

I thank my friends from the research group, Beytullah Özgur for his friendship and contributions on my thesis presentation, Dr. Özge Yılmaz who initiated the diphenylalanine project and helped me get acquainted with the concepts when I started in Koc, Dr. Christoph Globisch for his collaboration in the LK project and Farhad Ramezanghorbani. I also want to thank my friends in Koc, Osman Eryurt, Derya Aydın, Zuhal Taşdemir and many others.

I would like to thank my wife for her ceaseless support and patience during my PhD. I also thank my brother, parents and grandparents and my wife's sisters, parents and grandparents for their support.

The studies in this thesis were supported by TUBITAK with the project number 212T184.

TABLE OF CONTENTS

List of Tables **xi**

List of Figures **xiii**

Nomenclature **xxx**

Chapter 1: Introduction **1**

Chapter 2: Theory **8**

2.1 Molecular Dynamics 8

2.2 Free Energy Calculations 15

2.3 Replica Exchange Method 19

2.4 Coarse-Graining 20

Chapter 3: Phenylalanine dipeptide: A two-state system for investigating conformational transition **22**

3.1 Introduction 22

3.2 Methods 25

3.3 Conformational Transition Upon Environmental Change 27

3.4 Transferable CG Model 32

3.4.1 CG Solvent Model 33

3.4.2 CG FF Model 38

3.4.3 Parametrization of the Dihedral Potential 43

3.4.4 Modifying the CG models 51

3.5 Concluding Remarks 60

Chapter 4: Environment Induced Conformational Transition and Aggregation of LKα14 Peptides **63**

4.1 Introduction 63

4.2 Methods 66

4.3 Environment Driven Conformational Transition of LKα14 Peptide . . 69

4.3.1 Bulk Water Simulations 69

4.3.2 Vacuum/Water Interface Simulations 77

4.4 Aggregation of LKα14 Peptides 83

4.4.1 Bulk Water Simulations 83

4.4.2 Vacuum/Water Interface Simulations 99

4.5 Coarse-grain Model of the LKα14 Peptide 113

4.6 Concluding Remarks 128

Chapter 5: Conclusion **130**

Bibliography **132**

Vita **153**

LIST OF TABLES

3.1 Solvation free energies of water and cyclohexane molecules for the constituent subphases of the interface system measured from atomistic (AA) simulations. Van der Waals and electrostatic contributions, as well as the change in total free energy are listed. Lennard-Jones interactions, where parameters ϵ and σ are tuned to reproduce the measured atomistic free energy, are used for the nonbonded interactions of water (**W**) and cyclohexane (**C**) CG beads. Parametrization of the **W-C** interaction is done in two different ways: 1) CG-FE: solvation free energy of water in cyclohexane (**W/C**) is used as reference. 2) CG-ST: interaction is tuned to reproduce the surface tension of the biphasic system. 34

3.2 Coarse-grain mapping for diphenylalanine showing the position of the superatoms, mass, and the atoms used for calculating the mass of the superatoms. 39

3.3 Equilibrium bond length and spring constant for the harmonic potentials connecting the CG beads. Beads are numbered from C-terminus to N-terminus. 41

3.4 Solvation free energy of FF fragments (that are mapped to the CG beads) in the respective solvents obtained from AA simulations. CG beads' nonbonded interactions are represented via Lennard Jones interactions, where the depth ϵ and radius σ of the interaction are fitted to reproduce the AA free energy. *For the B/W interaction CG bead is tuned to reproduce half of the hydration free energy of zwitterionic diglycine. Results with the modified **P/C** interaction are also shown for the different solvent models used: CG-FE and CG-ST. 44

3.5 Solvation free energy of ethylbenzene in cyclohexane obtained from AA simulations and the corresponding solvation free energies of **P** in solvent of **C** shown along with those of the modified CG-FE and CG-ST models. The nonbonded interactions of the CG models are represented via Lennard Jones interactions, where the depth ϵ and radius σ of the interaction are fitted to reproduce the free energy of the AA model. Note that as the atoms of the cyclohexane molecule does not have any partial charges the electrostatic contribution equals to 0 and is not shown. 54

4.1 Highly populated conformations for a LKα14 dimer at the vacuum/water interface defined in terms of inter-dimer distance and inter-dimer angle obtained from the correlation plot in Fig. 4.30. . . . 103

4.2 Area of hydrophobic leucine sidechains that are exposed to water molecules for LKα14 in nm^2. For the cases involving a dimer, exposure of leucines per peptide is calculated. Comparison of the SASA values with the single peptide in bulk case is listed as percentages. The errors for each case are lower than 0.01. 108

4.3 Definition of the beads of the CG model of LKα14 and their masses . 114

4.4 Equilibrium bond length and spring constant for the harmonic potentials connecting superatoms in the LKα14 CG model. 120

LIST OF FIGURES

1.1 Schematics showing the relation between folding, environment induced folding and aggregation in peptides. Figure is adapted from [17]. . . . 2

1.2 Schematics showing the hierarchy for multiscale modelling via spatial and temporal dimensions. The figure was obtained from [34]. 4

2.1 Schematic showing the downhill motion of energy minimization algorithms. Figure was adapted from [58]. 11

2.2 Schematic flow of replica exchange molecular dynamics simulation. Replicas at different temperatures are run in parallel and at regular intervals exchanges between replicas are attempted. The diagram is taken from Riken Research Website [79] 19

3.1 Molecular packing of diphenylalanine. Taken from [122]. 24

3.2 Diphenylalanine molecule. Taken from [122]. 25

3.3 Diphenylalanine peptide in cis-like conformation at the cyclohexane/water interface (top) and in trans-like conformation in bulk water (bottom). Mapping of the peptide to a 4 bead CG model (**PBBP**, **P** for sidechain and **B** for backbone beads) is also shown. Solvent molecules (shown as continuum for visual clarity) are mapped on to single beads: **W** and **C** for water and cyclohexane, respectively. . . . 28

3.4 Probability distribution for the **PBBP** dihedral angle from atomistic simulations. In bulk water (solid black line), the molecule adopts an extended structure. At the cyclohexane/water interface the molecule adopts a cis-like conformation (dashed blue line), highly distinct from bulk water behavior, but similar to the crystal structure (solid red impulse). In a finite-concentration solution of FF peptides molecules can switch from a trans-like to a cis-like conformation (probability distribution for such a molecule is shown with dashed brown line), when presented with a hydrophobic interface formed by a small cluster of FF molecules. 29

3.5 A diphenylalanine in cis-like conformation (shown in yellow) at the interface of a nano-cluster of diphenylalanine molecules in water. Water molecules are not shown for clarity. The nano-cluster of FF molecules presents a hydrophilic/hydrophobic interface, which enables the transition of the corresponding molecule to a cis-like conformation. Peptide residues within 4 Å of the corresponding peptide are shown in red color. 30

3.6 Probability distribution of the dihedral angle from a finite-concentration solution of FF peptides. Dihedral distribution averaged over all of the peptides, as well as the results for individual distributions for selected FF molecules are shown. 31

3.7 Dihedral angle among **PBBP** superatoms of the molecule in Fig. 3.5 as a function of simulation time. Unlike the peptide in bulk water (where only trans-like conformation is observed), in a solution of peptides in water, formation of nano-clusters with hydrophilic/hydrophobic interfaces enables sampling of the cis-like conformation. 32

3.8 CG non-bonded interaction potentials for the solvent beads (Lennard-Jones parameters are given in Table 3.1). **W-W** and **C-C** interactions are shown with solid black and blue lines, respectively. Two different potentials are used for the **W-C** interaction, where either the solvation free energy of water in cyclohexane (CG-FE, dashed black line) or the surface tension of cyclohexane/water system (CG-ST, long-dashed blue line) is taken as a reference. 36
3.9 Comparison of water-water (W-W) (black color) and cyclohexane-cyclohexane (C-C) (blue color) radial distribution functions in AA (solid lines) and CG (dashed lines) simulations. 37
3.10 Density profile of water (black color) and cyclohexane (blue color) phases in AA (solid lines) and CG (dashed lines) simulations. 38
3.11 Mapping scheme for FF. The phenyl rings are mapped onto the **P** superatoms (brown) and the backbones are represented by the **C** superatoms (blue). 40
3.12 Angle potentials for the P1-B1-B2 and B1-B2-P2 angles obtained via Boltzmann inversion of the corresponding atomistic probability distributions. 41
3.13 Probability distributions for bonds between FF superatoms (BB, PB and BP) from atomistic simulations in bulk water and at the cyclohexane/water interface. 42
3.14 Probability distributions between the BBP and PBB superatom triplets from atomistic simulations in bulk water and at the cyclohexane/water interface. 42
3.15 CG nonbonded interaction potentials between the FF and solvent beads (Lennard-Jones parameters are given in Table 3.4). 43
3.16 Dihedral potential for the P1-B1-B2-P2 obtained via Boltzmann inversion of the corresponding atomistic probability distribution. 45

3.17 Bulk water: Comparison of **PBBP** dihedral angle probability distribution from CG simulations with the atomistic reference (solid-black line). Without an explicit dihedral potential, single-bead representation (see text) of the **P** beads leads to a cis-like conformation (dashed brown line). When the size of the **P** beads is reduced, a flat dihedral distribution is obtained (solid brown line). Thus, dihedral potential obtained via Boltzmann Inversion (BI) of the atomistic reference in bulk water (dashed blue line) reproduces the correct behavior. On the other hand, BI of the interface distribution fails in bulk water (solid-blue line). 46

3.18 Comparison of the dihedral probability distribution of the atomistic simulations at the water/cyclohexane interface for the 0-50 ns and 50-100ns intervals. 46

3.19 Comparison of the dihedral probability distribution of the free energy based coarse grained model (CG-FE) simulations at the water/cyclohexane interface for the first 50ns and 50-100ns intervals. 47

3.20 Cyclohexane/Water interface: Comparison of **PBBP** dihedral angle probability distribution from CG simulations with the atomistic reference (black line). CG model uses reduced **P-W** interaction radius and dihedral potential based on BI of atomistic bulk water dihedral distribution. The CG solvent model with surface tension tuned **W-C** interaction fails to capture the correct dihedral behavior (CG-ST, brown line), whereas the **W-C** interaction tuned to reproduce solvation free energy of a water molecule in cyclohexane solvent (CG-FE, blue line) yields a fairly close result. 48

3.21 Probability distribution for the distance of FF from the interface. Unlike the AA simulation (solid black line), where the molecule is placed right at the interface, CG model yields a submerged FF molecule with both of the tested solvent models (CG-FE and CG-ST shown with blue and brown lines, respectively), with CG-ST model displaying looser adsorption to the interface. 49
3.22 Probability distribution for the orientation angle (γ) of the molecule with respect to the interface for atomistic (solid black line) and CG model with two different solvent models (solid blue line for CG-FE and solid brown line for CG-ST). 50
3.23 Snapshots showing the orientation of CG model of FF peptide with respect to interface. The snapshot on the left corresponds to an orientation of approximately 90° and the one on the right to an angle of30° 50
3.24 CG nonbonded interaction potentials between the FF and solvent beads (Lennard-Jones parameters are given in Table 3.4). For the **P-C** potential the modified CG interactions are also shown for both the CG-FE (solid brown line) and the CG-ST models (dashed brown line), which correct the dihedral distribution at the interface. 52

3.25 Cyclohexane/Water interface: Comparison of **PBBP** dihedral angle probability distribution from CG simulations with the atomistic reference (solid black line). CG model uses reduced **P-W** interaction radius and dihedral potential based on BI of atomistic bulk water dihedral distribution. The CG solvent model with surface tension tuned **W-C** interaction fails to capture the correct dihedral behavior (CG-ST, solid brown line), whereas the **W-C** interaction tuned to reproduce solvation free energy of a water molecule in cyclohexane solvent (CG-FE, solid blue line) yields a fairly close result. By increasing the attraction for the **P-C** interaction both CG models can be enhanced to yield a better match with the AA result (dashed blue and dashed brown line for CG-FE and CG-ST solvent models, respectively). 53

3.26 Probability distribution for the distance of FF from the interface. Modified **P-C** interaction enhances both of the CG models (dashed lines), where modified CG-ST case yields almost a perfect match with atomistic results. 55

3.27 Probability distribution for the orientation angle (γ) of the molecule with respect to the interface for atomistic (solid black line) and CG model with two different solvent models (solid blue line for CG-FE and solid brown line for CG-ST). Upon modification of the **P-C** interaction, both CG-FE and CG-ST case yields a narrower distribution compared to the atomistic results (dashed lines). 56

3.28 Potential of mean force, $V(r)$, for the transfer of FF from bulk water to the cyclohexane/water interface. CG-FE solvent model yields the closest match to the atomistic result in terms of depth (solid blue line), whereas CG-ST solvent model displays a weaker attraction towards the interface. In both cases the minimum is shifted towards the water side of the interface. Modification of the **P-C** interaction pulls down the PMF curve to much deeper values, largely overestimating the adsorption energy (dashed lines). 57

3.29 Potential of mean force, $V(r)$, for the transfer of FF from bulk water to the bulk cyclohexane. The plot is the full form of Fig. 3.28 showing the transfer to the cyclohexane phase. 57

3.30 Snapshots showing the hydration shell when FF is pulled into the cyclohexane phase from the bulk water. The CG-ST solvent model (middle) where the surface tension matches that of the AA, correctly captures the hydration shell of the AA (left) whereas the cg-FE (right) which yields a higher surface tension value compared to AA, fails. 58

4.1 Helical wheel representation of LKα14 showing the separation of hydrophobic leucine residues (green) and hydrophilic lysine residues (blue). 65

4.2 Atomistic view of KL peptide in bulk water (bottom) and at the vacuum/water interface (top). The translucent grey area represents the water slab. At the vacuum/water interface the lysine (blue) residues reside inside the water whereas the leucine (red) sidechains extend into the vacuum. The sidechains are shown with licorice and the backbones are shown with new cartoon representations. 66

4.3 A single LKα14 in bulk water: The snapshots are hand-picked to represent different conformations observed throughout the simulation. The solvent accessible surface area of the hydrophobic leucine sidechains are plotted vs. time (middle). Time evolution of the secondary structure of a single LKα14 in bulk water color coded for each residue (bottom) along with the corresponding snapshots show that full-helical conformation was not stable and the peptide adopted various conformations including random-coil, and half-helix-half-random-coil. 70

4.4 Time evolutions of the secondary structure of the LKα14 peptide in bulk water for the remaining 2 simulations with different initial seeds. For each simulation the initial secondary structure of α-helix was not stable, however for the figure in the middle a refolding into α-helix was observed. In general, the peptide assumed a variety of secondary structures including half-α-helix half-random coil, beta sheet and random coil. 71

4.5 A single LKα14 peptide in bulk water initiated from a random structure. The peptide did not fold into α-helix, but instead formed an unstable beta-sheet. 72

4.6 Replica exchange MD in bulk water: Time evolution of the secondary structure of each replica conformation for a single LKα14 peptide in bulk water. None of the simulations converge to a single secondary structure which indicates that for a single LKα14 in bulk water there is no unique equilibrium secondary structure. The plots are ordered in increasing temperature from left to right and top to bottom. 74

4.7 Most populated 6 conformations (from left to right and top to bottom) for the REMD simulation at 298 K for a single LKα14 in bulk water. The conformations are calculated using Jarvis-Patrick clustering algorithm [168] with a number of nearest neighbor cutoff of 60. The backbone is drawn with NewCartoon representation and colored according to secondary structure. The lysine (blue) and leucine (white) residues are shown as Licorice. 75
4.8 Matrix showing the exchanges between replicas for the REMD simulation in bulk water. The x scale represents the replica ids and the color bar represents the initial replica ids. The amount of mixing of the REMD simulations is correlated by the amount of mixing in this figure. 76
4.9 Mean number of residues with the shown secondary structure vs temperture. 76
4.10 Time evolution of the secondary structure of the LKα14 peptide at the vacuum/water interface color coded per residue. The alpha helical secondary structure of the peptide is not disrupted for the duration of the simulation. 77
4.11 Time evolution of the secondary structure of single LKα14 simulation at the vacuum/water interface which is initially random and folds into α-helix. 78

4.12 Single LKα14 at the vacuum/water interface: mean of the distance between the centers of mass of hydrophobic leucine sidechains and the charged NH_3^+ groups for a single peptide at the vacuum/water interface. The plot shows the amount of partitioning: as the distance between these groups increases partitioning increases. Initially the peptide is submerged inside the water slab, and moves to the interface in 40 ns (see Fig. 4.11). The α-helix conformation at the interface maximizes partitioning by preventing the hydration of leucine sidechains. 79

4.13 Mean of solvent accessible surface area for leucine sidechains for a single LKα14 at the vacuum/water interface. The peptide is initially random coil and submerged in bulk water. The initial drop corresponds to the peptide moving to the interface. The area of hydrophobic leucine sidechains exposed to water is minimised when the peptide adopts full α-helix (see Fig. 4.11). The area of the leucine sidechains facing vacuum is subtracted from the total area of the leucine sidechains accessible from scanning the protein surface to achieve the SASA. 80

4.14 Replica exchange MD at the vacuum/water interface: Time evolution of the secondary structure of a single LKα14 peptide between temperatures 298-366 K. Secondary structure content was evaluated with the DSSP algorithm. The plots are ordered in increasing temperature from left to right and top to bottom. 81

4.15 Time evolution of the secondary structure for a dimer of LKα14 in bulk water for antiparallel (top) and parallel (bottom) orientations. The α-helix secondary structure is stable for both orientations for the duration of the simulation. In the snapshots, the lysine sidechains (blue licorice) extend into bulk water whereas the leucine sidechains (white licorice) form a hydrophobic core. The peptide backbone is colored according to secondary structure with NewCartoon representation of VMD. The ACE cap residues are drawn with orange licorice. 84

4.16 Comparison of parallel and anti-parallel orientation of LKα14 dimer in bulk water: Timeline of inter-dimer distance and inter-dimer angle (top) and the corresponding histograms (bottom). The distance is the distance between the centers of mass distance of the backbone atoms. The angle is the angle between the vectors passing through the backbone of the peptides, where the first and last 3 residues are not taken into account due to occasional unfolding of the end residues. The parallel orientation adopts an average inter-dimer angle of 30° and an inter-dimer distance of 0.97 nm where the anti-parallel one yields 150° and 1.06 nm. 86

4.17 Distance angle correlation for a LKα14 dimer in bulk water for peptides aligned in parallel (left) and anti-parallel (right) 87

4.18 LKα14 dimer in bulk water: Two initially unfolded and separately positioned LKα14 peptides in bulk water first associate and then fold. Snapshots handpicked from the simulation are shown along with the time evolution of the secondary structure. Solvent accessible surface area for the leucine sidechains is plotted (third from above) and the minimum distance between the peptides is presented (bottom). The initial drop in both the minimum distance and SASA plots correspond to the association that takes around 180 ns. 88

4.19 PMFs for separating one peptide from a dimer for parallel and anti-parallel orientations. The depth of the PMFs are the same for both orientations which suggests that they are equally preferable. The convergence for the PMF was checked by block analysis. Entropic correction was applied to the PMFs. 89
4.20 PMFs for separating one peptide from an anti-parallel oriented LKα14 dimer in bulk water obtained by umbrella simulations comparing the peptides where the lysine sidechains are charged and uncharged. The PMF of the charged peptides shows a decreasing slope after 2.5 nm which corresponds to a repulsion whereas the uncharged peptides display a flat PMF at large distances. Therefore the repulsion observed is of electrostatic origin. Entropic correction was applied to the PMFs. 90
4.21 LKα14 dimer in bulk water: Initially, the peptides are full α-helix and separated. The peptides merged after 300 ns and α-helix for one of the peptides the full α-helix was not stable between 180-340 ns. Compared to the previous simulations where two randomly oriented peptides merged in 40 ns (Fig. 4.18) here the formation of the dimer takes more time. The reason behind this increased time for aggregation is the electrostatic repulsion observed at large distance as seen from the PMF curves in Figs. 4.19 and 4.20. 91
4.22 Frequency histograms for the umbrella simulations for pulling one peptide from the anti-parallel dimer. The histograms should intersect if one wants to achieve a reliable PMF curve. 92
4.23 Time evolution of the secondary structure for each residue for trimer, tetramer, pentamer and octamer in bulk water color coded for each secondary structure. 94
4.24 Snapshots for the trimer, tetramer, pentamer and octamer simulations in bulk water. 95

4.25 PMF curves obtained by umbrella sampling simulations for pulling away a single LKα14 peptide from the LKα14 aggregate. The tetramer has the highest depth among the four aggregates with the trimer and the pentamer having the next highest depth. Error bars are smaller than the point sizes, therefore may not be clearly visible. 96
4.26 Umbrella PMF curves for pulling away a single LKα14 from the dimer, trimer, tetramer and pentamer compared to dividing the octamer into two tetramers. 97
4.27 Difference in solvent accessible surface area between the aggregate and the sum of the individual chains, compared for different aggregate sizes of trimer, tetramer, pentamer and octamer. That is the sum of the SASA for individual chains is subtracted from the SASA of the whole aggregate and plotted according to time. 98
4.28 Time evolution of the secondary structure for a dimer at the vacuum/water interface. 100
4.29 LKα14 Dimer in bulk water (red) and vacuum/water (green) interface. Timelines: Inter-dimer distance (top), inter-dimer angle (2nd from above). Histograms: inter-dimer distance (bottom, left), inter-dimer angle (bottom,right). Unlike bulk water the aggregate is loosely positioned with a weaker drive to form a stable aggregate. 101
4.30 Inter-dimer distance vs inter-dimer angle correlation for a LKα14 dimer at the vacuum/water interface. The anti-parallel orientation is more populated compared to the parallel one and the dimer does not dissociate but instead prefers to position with larger distances. The three distinct peaks are can be seen in Fig. 4.31. 102

4.31 Representative snapshots of high probability conformations for a LKα14 dimer at the vacuum/water interface. The top three figures are the anti-parallel orientations and correspond to the peaks observed in distance angle correlation plot in Fig. 4.30. The ones at the bottom are parallel orientations. Lysine and leucine residues are colored blue and red, respectively, and the ACE cap is colored gold with CPK representation. 104

4.32 Comparison of PMF curves for separating one LKα14 peptide from a dimer in bulk water and vacuum/water interface obtained with umbrella sampling method. The bulk water PMF is deeper than those for the dimer at the vacuum/water interface. The anti-parallel orientation has a higher depth compared to that of the parallel in vacuum/water interface. Note that the at distances larger than 2.4 same simulations were used as windows for the vacuum/water simulations because at these distances the anti-parallel and parallel orientations are not meaningful and only the separation distance matters. 105

4.33 Histogram of the minimum inter-peptide distance between charged N atoms at the end of lysine sidechains for bulk water and basin 1 and 3 of vacuum/water interface. 106

4.34 Mean number of inter-peptide leucine-leucine sidechain contacts for each highly populated conformation defined in Table 4.1. The snapshots for these conformations are shown in Fig. 4.31. The threshold distance was determined to include the first peak of the RDF (shown in Fig. 4.35) for the tetramer which was the most stable aggregate size in bulk water. 107

4.35 RDF of leucine-leucine sidechains 107

4.36 Difference between solvent accessible surface area of dimer and sum of individual peptides in bulk water for anti-parallel (top) and parallel (middle) orientations and for dimer at vacuum/water interface (bottom).109
4.37 The initial configurations for eight LKα14 peptides at the vacuum/water interface 110
4.38 The final conformations for eight LKα14 peptides at the vacuum/water interface. All the peptides remain α-helix and are positioned at the vacuum/water interface with hydrophobic and hydrophilic residues partitioned as discussed before. 110
4.39 Time evolution of the secondary structure for eight LKα14 peptides simulated at the vacuum/water interface. Except intermittent unfolding of the end residues, α-helix structure is stable for all peptides. . . 111
4.40 The CG mapping scheme for the atomistic LKα14 peptide. The coloring of superatoms are as follows: **L** are red, **KC** and **KC** are blue, **CA** and **CN** are cyan. The radius of the superatoms were tweaked to contain their atomistic counterparts and are arbitrary. 113
4.41 The coarse-grain mapped bond-stretching distributions for the tetramer (red) and the single peptide in implicit water with the non-bonded interactions excluded (labeled as *vacuum-excl* with green). 115
4.42 The CG mapped angle-bending distributions for the tetramer (red) and a single peptide in implicit water with non-bonded interactions excluded (green labeled as *vacuum-excl*). 116
4.43 The coarse-grain mapped dihedral distributions for the tetramer and the single peptide in vacuum with non-bonded interactions excluded. 117
4.44 The coarse-grain mapped improper dihedral distributions for the tetramer (red) and a single peptide in implicit water with non-bonded interactions excluded (green labeled as *vacuum-excl*). 118

4.45 Distributions from tetramer and vacuum atomistic simulations for backbone dihedrals (top): CA-CN-CA-CN (left) and CN-CA-CN-CA (right) and CN_i-CN_{i+4} distance distribution (bottom). The difference between a random coil and α-helix is highlighted in CN-CA-CN-CA dihedral angle and the CN_i-CN_{i+4} distance distributions. 120
4.46 Tabulated potentials for angle interactions used in the LKα14 CG model obtained via Boltzmann-inverting the related angular distributions of the atomistic tetramer simulations in bulk water. 121
4.47 Tabulated potentials for the proper backbone dihedral interactions. These potentials are the Boltzmann-inverted forms of the corresponding distributions of atomistic tetramer simulations in bulk water. . . . 121
4.48 Tabulated potentials for implicit dihedral interactions. 122
4.49 Dihedral distribution of CA-CN-CA-L for the CG without the dummy improper dihedral potential (red) and with it (green), compared to that of the atomistic tetramer. The improper fix potential is displayed at the right. 122
4.50 Different tabulated potentials used for the CN_i-CN_{i+4} interaction (right) and the corresponding CN_i-CN_{i+4} distance distributions (left). Without any defined pair interactions the CG model fails to reproduce the α-helix (*cg-p0*) and the Boltzmann inverted form of the CN_i-CN_{i+4} distance distribution for the atomistic tetramer reproduces the correct distribution of the atomistic tetramer in bulk water (*cg-p1*). 123
4.51 Distributions comparing the CG models with the tetramer derived backbone dihedrals and the implicit water derived ones. No pair potential between the CN_i-CN_{i+4} beads were used. The CG models display similar distributions to to those of their atomistic counterparts. . . . 124

4.52 Backbone dihedral angle (top) and CN_i-CN_{i+4} distance distributions (bottom) comparing the CG models with and without the CN_i-CN_{i+4} pair potentials for the CG model with the backbone dihedral potentials derived from the atomistic implicit water simulations. 125

4.53 Distribution of the distance between CN_i-CN_{i+4} compared to different depths of wall potentials without the CN_i-CN_{i+4} defined (*cg-w0.1, cg-w.5, cg-21* where *cg-w0.1* means the wall potential obtained from that is a counterpart for the interaction for the atomistic *leucine-interface* is multiplied by 0.1). The CN_i-CN_{i+4} pair potential (*cg-ii4p*) is defined as the inverse of the atomistic tetramer distributions. 126

4.54 Snapshot of CG LKα14 on the wall where the wall is shown with a black line. 126

4.55 Histograms of the CN_i-CN_{i+4} distances (left) obtained by the corresponding tabulated potentials of used to represent the CN_i-CN_{i+4} pair interactions. Here the wall potential is multiplied by 0.1 and the CN_i-CN_{i+4} potentials are scaled. 127

4.56 Tabulated interaction potentials of superatoms (right) with the wall except the L bead(i.e. CA-wall, CN-wall, KC-wall, KN-wall) and the distribution of their distances from the wall (left). 127

NOMENCLATURE

MD	Molecular Dynamics
CG	Coarse-grained
PMF	Potential of Mean Force
FF	Phenylalanine Dipeptide
LK	Lysine-Leucine Peptide
LKα14	$(LKKLLKL)_2$ Sequenced Lysine-Leucine Peptide
AA	All Atomistic

Chapter 1

INTRODUCTION

Proteins are the building blocks of life that function as enzymes for the chemical reactions in living organisms. [1] Proteins perform tasks ranging from transport of molecules to DNA replication, from signalling messages inside the cell to acting as agents that transform molecules. [2] All proteins consist of different combinations of 20 different aminoacids and adopt a collection of structures. Nuclear magnetic resonance (NMR) and X-ray diffraction techniques along with several spectroscopic methods such as circular dichroism and Raman spectroscopy are used to determine the structures of proteins. The link between structure and function of proteins is a hot topic being actively researched. Therefore, predicting the native structure of a protein is one of the notable goals in biological studies. [3] Initial theory was that the sequence determined the function of the protein. Although it plays a major role, studies showed that the sequence is not the only factor. In fact, identical sequences were observed to adopt different structures in different environments. Thus the environmental effects were identified as one of the key factors. [4] In 1984, Kabsch et al. demonstrated that a pentapeptide assumed different structures according to their context in the protein. [5] Later, Waterhous et. al along with Zhang et al., showed that the choice of solvent as one of the factors that define the final structure. [6, 7] Mutter et al. reported environment dependent and pH-dependent mechanisms for conformational transition mechanisms based on peptides. [8, 9]

Peptides are shorter versions of proteins containing up to 50 amino acid residues. [10] Amphiphilic nature of peptides enable ordered structures at the interfaces and surfaces. Such conformational changes upon change of media have uses in biomed-

ical applications such as implant devices [11] or in anti-microbial uses [12]. Currently, switch peptides are used to study conformational transitions which are regarded as key events in neurodegenerative diseases. [13, 14] Spontaneous aggregation of switched peptides resulting in amyloid plaque formation is thought to be the source of Alzheimer's disease. [15, 16]

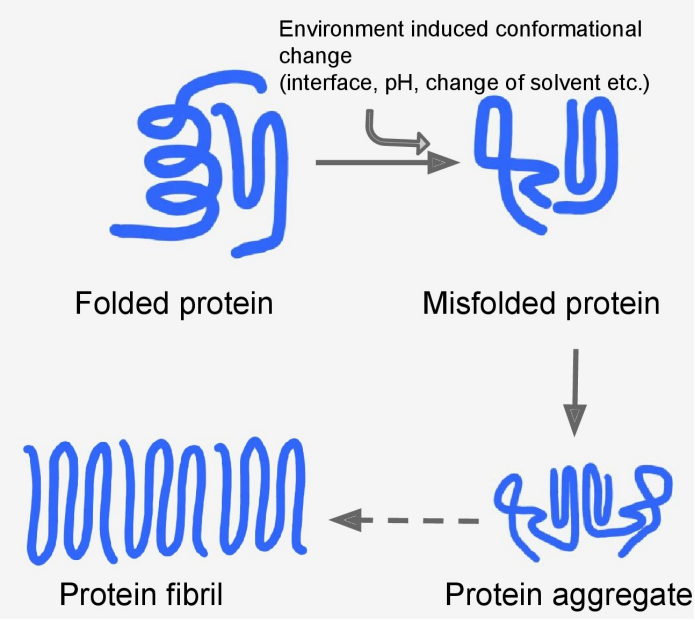


Figure 1.1: Schematics showing the relation between folding, environment induced folding and aggregation in peptides. Figure is adapted from [17].

Moreover, the amphiphilicity promotes self-assembly on the cell membranes [18–20] which can also be initiated manually by alterations in environmental factors, such as pH, temperature, insertion of solute, change of solvent and such physical or chemical changes. [21] There exist a range of diseases that are associated with aggregation or conformational transition of proteins such as Alzheimer's, Huntington's, Type II diabetes. [22] For example, aggregation of misfolded proteins leads to amyloid plaque which plays a role in neurodegenerative diseases. [23]. Other than diseases, the self-assembly and environment dependent conformational transition in proteins and peptides have wide variety of uses in designing biological hybrid materials [24], smart materials that respond to environmental changes [25, 26] such as the introduction of

an interface [27].

The ultimate aim is to control these processes and to control them one has to understand the underlying mechanisms and lay out the interactions and driving forces responsible in those mechanisms. To this end, molecular simulation is a valuable tool that provides access to a range of information at an atomistic detail, which may be inaccessible by current experimental techniques. Molecular simulation can be useful in interpreting spectroscopic information by providing both a structural interpretation of the data and also giving access to the thermodynamic driving forces. Therefore, molecular dynamics (MD) simulations are increasingly used to study biomolecular processes such as protein folding, lipid layer formation, and aggregate formation as a result of advances in computational hardware and improvements in numerical algorithms and force-fields. [28–30]

Challenges in collecting experimental data for the switch processes deems the computational studies more important. [13] Furthermore, the environment dependent complex arrangement of peptides complicates multiscale modeling. Furthermore, simulating structural changes of such large-scale biological processes requires enormous computational resources. [31] In terms of aggregation processes, MD simulations are generally limited to either small molecules or simulating pre-aggregated forms. Sometimes the level of detail in a simulation may be too much to identify significant parameters. [32] For such cases, a coarse-grained (CG) model can be used which decreases the computational demand and speeds up the simulations by alleviating the unwanted details. [33] Therefore CG models act as bridges between different scales where the processes are unattainable by standard MD simulations. A rough outline for the temporal and spatial hierarchy for multiscale modeling can be seen in Fig. 1.2.

In coarse-graining methods, atoms are grouped together and each group is represented as beads, or spherical group of atoms which are also called superatoms. The missing interaction parameters can be generated using available systematic methods: iterative Boltzmann inversion [35], inverse Monte Carlo [36] or force-matching [37]. These methods yield potential functions that act as a force-field to reproduce the

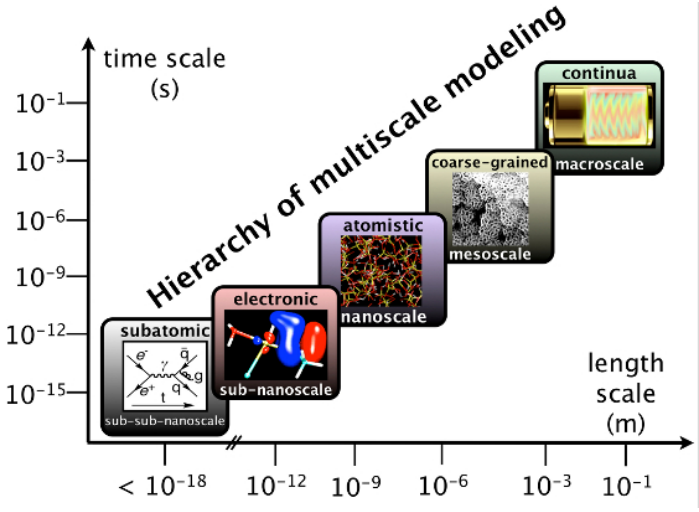


Figure 1.2: Schematics showing the hierarchy for multiscale modelling via spatial and temporal dimensions. The figure was obtained from [34].

desired properties of the target systems. The target systems are higher resolution systems which can be experimental values, full-atomistic models or any other system with a higher degree of freedom compared to the CG model. [38] When choosing the mapping of the CG model for the system, it is important to clearly define and focus on the properties to be reproduced. Because, the degree of freedom of the coarse-grained model is lower than the target system, evidently some of the information will not be represented.

As the current computational resources demand CG models for replicating large scale biological processes, a meaningful representation of these processes is only possible if the model is able to reproduce the behavior of peptides/proteins in different environments. For example, dynamic processes such as the conformational switch constitute at least two distinct conformations. However standard CG models are state-dependent, i.e. they mimic only the system of interest they are modeled after. Meaning that the state-dependent CG models fail to reproduce the behaviors of their higher degree counterparts in different conditions such as different solvent, concentration or temperature. Therefore, the state-dependent models fail to represent the processes where environment induced conformational transition takes place, because

they are not transferable. To this end, transferability is one of the key questions in current coarse-grain studies.

Overall there are two approaches to coarse-graining: 1) the top-down approach where the properties to be reproduced are obtained from experimental observations that are large scale and not accessible by atomistic models and 2) the bottom-up approach where the properties are obtained via higher degree models, such as all atomistic models. [39] Although the top-down approach has seen some success in CG models, the interaction potentials do not have a physical basis and are defined to imitate certain large-scale phenomena. CG models based on MARTINI force fields represent the top-down approach where large scale molecular arrangements such as lipid bilayers as well as the partitioning of proteins at those bilayers can be successfully reproduced. [31,40] However the predetermined interaction potentials that do not have a basis from the atomistic standpoint makes this approach not suitable for modeling systems that display environment induced conformational transition.

The bottom-up approach makes use of several strategies to generate interaction potentials. One of those strategies is to make use of the solvation free energies to generate the interaction parameters for the CG model. [41] Free energy gives information on the probability that a system adopts a state which can be used to construct the interaction potentials. [42, 43] An alternative method is to define interactions to reproduce the radial distribution functions (RDF) where Boltzmann inversion can be utilized. [44] The scheme is made effective by iteratively updating the potentials. Using iterative boltzmann inversion (IBI) [35], the RDF of pure water system was reproduced by mapping the atomistic water molecules into single CG beads. [45]

Villa et. al. developed a solvent-free coarse-grained model for peptides which reproduced the aggregation of a dipeptide in solution [33] and then extended the model to include the solvent parameters and mimicked the chirality of the atomistic simulation. [29] Engin et. al. devised solvent-free CG model for diphenylalanine that reproduced structural and thermodynamic properties qualitatively and showed that these potentials were transferable for two other hydrophobic dipeptides, namely,

valinephenylalanine and isoleucinephenylalanine. [46] In another study, the IBI was modified to reproduce the Kirkwood-Buff integrals instead of the RDF and the CG model for urea-water in was produced where the molecules were mapped onto single beads. The model was transferable up to concentrations of 2 M. [47] The study was expanded to reproduce the free energy for forming benzene clusters in urea-water. [48] An informed selection of the initial potentials along with suppressing many-body interactions in the RDF allowed representing the inhomogenous mixtures where a system of liquid perfluoropolyether film coated on a carbon surface was coarse-grained. [49] Although the coarse-graining is increasingly studied there are few studies concerning modelling the environment dependent conformational transition of peptides or proteins.

The purpose of this thesis is to understand the mechanisms of environment induced conformational switch processes through computational simulations. The studies focus on structural and thermodynamic properties of the systems under inspection. This thesis contributes to the literature by 1) displaying environment induced conformational transition via atomistic molecular dynamics simulations and 2) presents a coarse-grained model that can capture the environment dependent behavior. Moreover, it is shown that the aggregates behave as interfaces and produce the ordered structures observed in interfaces.

Chapter 2 gives an outline of the concepts used in this thesis such as molecular dynamics and algorithms and approaches applied in simulations. Furthermore, the reader is briefly introduced to the free energy calculations used in this thesis. Then the theory behind coarse-graining and different approaches to creating CG models are mentioned in a concise manner.

In chapter 3 using atomistic MD simulations the hydrophilic/hydrophobic interface was shown to trigger a structural change in phenylalanine dipeptide. Moreover aggregates of these dipeptides were identified as generating micro-interfaces where similar conformational switches observed in interfaces were detected. The second part of the chapter presents coarse-grain models of diphenylalanine (FF) that are

able to mimic the environment dependent conformational switches. The different approaches to determine the nonbonded interactions for the CG models are compared where it is shown that the success of the CG model for systems that exhibit conformational changes, also depends on the interactions that do not involve the peptide, i.e. solvent-solvent interactions. This study was published in the Journal of Chemical Physics in 2013. [50]

The focus of Chapter 4 is an amphiphilic 14-residue peptide LKα14 that adopts different secondary structures in bulk water and hydrophilic/hydrophobic interface. The secondary structure of a single peptide is tested in bulk water and at the air/water interface which is the simplest of hydrophilic/hydrophobic interfaces. Then the folding and aggregation of a dimer in bulk water is shown as an evidence of aggregation induced change in secondary structure is demonstrated. Later, self-assembly of LKα14 is investigated by comparing the structural and thermodynamic properties as well as the stability of various aggregates with different sizes. In the last section of the LKα14 studies, the coarse-grain model of a single LKα14 peptide is shown where the model mimics the full-atomistic resolution model of a helical structure of a peptide in a four-helix bundle.

As a concluding remark Chapter 5 presents an overview of the studies and summarizes the findings.

Chapter 2

THEORY

2.1 Molecular Dynamics

Molecular Dynamics [1] (MD) is a simulation technique where the Newton's equations of motion for hard-sphere atoms are solved numerically for a number of steps. MD is widely used in simulating biological processes such as protein folding, lipid layer formation, aggregation, mutation etc. [51–54] The simulations allow access to atomistic detail properties that are not easily accessible by experimental techniques. [55]

The potential energy, U consists of bonded and nonbonded parts.

$$U = U_{bonded} + U_{nonbonded} \tag{2.1}$$

$$U_{bonded} = U_{bond} + U_{angle} + U_{dihedral} \tag{2.2}$$

$$U_{nonbonded} = U_{vdW} + U_{electrostatic} \tag{2.3}$$

U_{bonded} consists of bond stretching, angle bending and dihedral angle energies and the $U_{nonbonded}$ includes van der Waals and electrostatic contributions where the former is represented as a typical 12-6 Lennard Jones potential and the latter a Coulombic one (Eq. 2.4 and 2.5).

$$U_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^6 - \left(\frac{\sigma}{r}\right)^{12}\right] \tag{2.4}$$

$$U_{electrostatic}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r} \tag{2.5}$$

[1] The scope of this thesis is limited by classical models and therefore does not consider quantum physics based methods such as ab initio MD. Thus when MD is mentioned, classical MD should be understood.

where σ corresponds to the radius of the van der Waals interaction and ϵ is the depth of the interaction potential. Q_1 and Q_2 are charges, and ϵ_0 is the permittivity of vacuum. Note that the nonbonded interactions are composed of pair potentials. To simplify the calculations multi-body effects are not taken into account.

The bonded interactions are defined as mass-spring systems.

$$U_{bond} = k_r(r - r_0)^2 \tag{2.6}$$

$$U_{angle} = k_\theta(\theta - \theta_0)^2 \tag{2.7}$$

$$U_{dihedral} = k_\phi(1 + cos(n\phi - \phi_0)) \tag{2.8}$$

where k_r, k_θ and k_ϕ are spring constants, r is the distance, θ is the angle, ϕ is the dihedral angle and n represents periodicity. r_0, θ_0 and ϕ_0 are the equilibrium values of corresponding properties. Bond stretching has the fastest degree of freedom among the bonded interactions and therefore limits the time step of the simulations. Therefore bonds are usually constrained to a fixed length to be able to user larger time steps and thus speed up the simulation [56].

$$Q = \frac{1}{h^{3N}N!}\sum\sum exp(-H(p,r)dpdr \tag{2.9}$$

where N is the number of atoms, is the inverse temperature, $H(p,r)$ is the Hamiltonian.

The pair interactions for each atom pair are defined by force-fields. Most widely used force fields in biological simulations are the versions of GROMOS, AMBER, OPLS. These interactions determine the energy of the system. The most time consuming part of an MD simulation is the calculation of energies and forces. Force can be calculated by taking the derivative of the potential energy with respect to position. (Eq. 2.10)

$$F = -\frac{\partial U}{\partial r} \tag{2.10}$$

Using MD simulation the property of interest, X can be calculated via Eq. 2.11.

$$\bar{X} = \frac{1}{M}\sum_{i}^{M} X(t_i) \tag{2.11}$$

where $\bar{X}$ is the time average of X and M is the number of time steps. For ergodic systems the time average is equal to the ensemble average, that is $\bar{X} = \langle X \rangle$.

Long range interactions Considering that the number of atoms in a typical MD simulation is on the order of 10^4 to 10^5, calculation of interactions between all atoms makes long time scales inaccessible. The workaround for this problem is exploiting the range of nonbonded interactions.

When reading Eqs. 2.4 and 2.5 one should remember that the energy functions that decay faster than $1/r^3$ diverge [57] which reveals that Lennard Jones interactions are short ranged and electrostatic ones are long ranged. To this end, the fast decay of van der Waals energy function is utilized by truncating at a cutoff distance. However as Coulombic interactions decay slowly, truncation introduces serious artifacts. As a workaround, Coulombic interactions are divided into short and long range contributions by a cutoff value and the former are calculated in real space whereas the latter are calculated in Fourier space.

$$U_{nonbonded} = U_{short} + U_{long} \tag{2.12}$$

and the short and long range potential energy functions can be defined

$$U_{short} = \overbrace{4\epsilon\left[\left(\frac{\sigma}{r_S}\right)^{12} - \left(\frac{\sigma}{r_S}\right)^{6}\right]}^{U_{vdw-short}} + \overbrace{Q_i Q_j \frac{erfc(\alpha r_S)}{r_S}}^{U_{electrostatic-short}} \tag{2.13}$$

$$U_{long} = \overbrace{Q_i Q_j \frac{erf(\alpha r_S)}{r_S}}^{U_{electrostatic-long}} \tag{2.14}$$

where $r_S = |r_i - r_j + S|$ is the distance between the ions considering the periodic images of the ion and S refers to the lattice vectors. $erf(x)$ is the error function and $erfc(x)$ is the complement of the error function such that $erf(x) + erfc(x) = 1$. This

approach is known as *Ewald summation*. The technique has been extended to various related methods such as *particle mesh Ewald*.

Energy minimization MD simulations are dependent on initial positions which should be a reasonable configurations close to the equilibrium value. However the initial simulation box is usually not at the equilibrium but positioned close to it. To this end energy minimization methods are performed to push the configuration to an equilibrium. Note that the "equilibrium" mentioned here does not necessarily correspond to the "chemical equilibrium" but to the "energy minimum" which is usually close to the chemical equilibrium.

Steepest descent is a widely used energy minimization method where at each step the system is moved to a lower energy configuration. The direction of the move is defined by [58]

$$s_k = -\frac{g_k}{|g_k|} \tag{2.15}$$

and the next direction will be perpendicular to the previous direction such that $g_k \cdot g_{k-1} = 0$. Then the new coordinates are updated as

$$x_{k+1} = x_k + \lambda_k s_k \tag{2.16}$$

where λ_k is the step size.

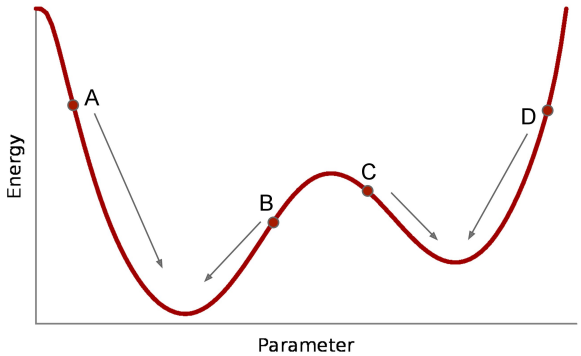


Figure 2.1: Schematic showing the downhill motion of energy minimization algorithms. Figure was adapted from [58].

Energy minimization methods do not guarantee to find the global minimum from an arbitrary position. That is, a conformation which is far away from the global minimum may never reach equilibrium due to being stuck in a local minimum. Energy minimization only carries the system downhill to the lowest energy, which could be a local minimum position if the initial configuration is chosen properly (Fig. 2.1).

Integrator To integrate the equations of motion of a system one of the simplest and most commonly used algorithm is the *Verlet* algorithm. The algorithm takes the Taylor expansion of the position of the particle at previous and forward time steps of Δt and sums them up (Eq. 2.18-2.19).

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}r + \dots \quad (2.17)$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}r + \dots \quad (2.18)$$

where $r(t)$ defines position, $v(t) = \dot{r}(t)$ velocity and $f(t) = \ddot{r}(t)$ force and m is the mass. Summing these two equations and truncating the higher order terms yield the position at the next time step $r(t + \Delta t)$.

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + \frac{f(t)}{m}t^2 \quad (2.19)$$

The *Verlet* algorithm has been extended to different schemes such as *velocity Verlet* and *leapfrog* where the former uses velocity to calculate the equations of motion instead of position. *Leapfrog* method was employed as an integrator for simulations in this thesis. The idea is to calculate the new positions by using the velocities at half-integer time steps (Eq. 2.20-2.21).

$$v(t-\Delta t/2) = \frac{r(t) - r(t-\Delta t)}{\Delta t} \quad (2.20)$$

$$v(t+\Delta t/2) = \frac{-r(t) + r(t+\Delta t)}{\Delta t} \quad (2.21)$$

Rewriting Eq. 2.21 yields the updated positions.

$$r(t + \Delta t) = r(t) + \Delta t v(t + \Delta t/2) \tag{2.22}$$

where the forward half-integer velocity can be obtained from the *Verlet* algorithm.

$$v(t + \Delta t/2) = v(t - \Delta t/2) + \frac{\Delta t}{m} F(t) \tag{2.23}$$

Thermostats For constant temperature simulations, the system is coupled to an external heat bath. Periodically the velocities are rescaled to keep the temperature at a desired value. There are a number of thermostats such as *Berendsen*, *velocity-rescaling*, *Nose-Hoover*, *Andersen* and *Langevin* thermostats that use different approaches to rescale the velocities. The simulations in this thesis utilise velocity-rescale algorithm. Velocity-rescale is a variant of Berendsen thermostat that retains its speed and, unlike Berendsen thermostat, allows proper sampling the canonical ensemble.

The general approach is that all velocities are multiplied by a factor to match the kinetic energy of the canonical ensemble.

$$\alpha = \sqrt{\frac{E_{kin}^t}{E_{kin}}} \tag{2.24}$$

where α is the scaling factor, E_{kin} is the kinetic energy and E_{kin}^t is the kinetic energy drawn from the canonical distribution. This method ensures canonical distribution while at the same time decreases fluctuations by spreading the change in kinetic energy into multiple time steps.

$$dE_{kin} = (\bar{E}_{kin} - E_{kin})\frac{dt}{\tau} + 2\sqrt{\frac{\bar{E}_{kin} E_{kin}}{N_f}}\frac{dW}{\sqrt{\tau}} \tag{2.25}$$

where E_{kin} is the kinetic energy, N_f is the number of degrees of freedom, τ determines the time scale of the thermostat. [59] Here dE_{kin} is the difference in kinetic energy that evolves the system to the desired temperature while keeping the canonical distribution invariant.

Barostats To keep the pressure at a constant value, one needs to use a barostat which scales dimensions similar to thermostats scaling velocities. *Berendsen* barostat was employed for the simulations in this thesis, the other available algorithms are *Parrinello-Rahman* and *Andersen*. In isothermal-isobaric simulations (NPT), the number of molecules, temperature and pressure are kept constant but the box dimensions, therefore the volume, and energy are allowed to change.

The pressure, P can be calculated via

$$P = \frac{2}{3V}(E_{kin} - \Xi) \tag{2.26}$$

where V is the volume, E_{kin} the kinetic energy and Ξ is the internal virial for pair-additive potentials (Eq. 2.27).

$$\Xi = -\frac{1}{2}\sum_{i<j}(r_i - r_j)F_{ij} \tag{2.27}$$

where r_i is the position of ith particle and F_{ij} is the force between the particles i and j.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \tag{2.28}$$

where P is the pressure, P_0 the target pressure and τ_P is the pressure coupling constant. [60]

The scaling factor μ can be calculated via Eq. 2.29.

$$\mu = 1 - \frac{\beta\Delta t}{3\tau_P}(P_0 - P) \tag{2.29}$$

where β is the compressibility factor which can be chosen to match experimental values of the system, where for water at 1 atm and 300 K $\beta = 4.6 \times 10^{-5}\, bar^{-1}$.

The bulk water simulations in this thesis use isotropic pressure coupling is, that is, the box dimensions are multiplied with the same scaling factor, μ. However for interfacial simulations the treatment of x and y dimensions is different from that of the z dimension, therefore a semi-isotropic coupling is applied. The purpose for semi-

isotropic coupling in interfacial simulations is to alleviate possible vacuum pockets while keeping the area of the interface constant.

2.2 Free Energy Calculations

Free energy is an important measure that gives insight into the preferences of a system. Free energy calculations are widely used in biological and chemical processes such as ligand binding [61], conformational changes [62], solvation [63]. [64] Computer simulations facilitate these calculations and are therefore increasingly utilized. Latest improvements in computational power and algorithms paved the way for reliable free energy estimations. [65]

The Helmholtz free energy A and Gibbs free energy G can be calculated by

$$A = -kT\, lnQ(N,V,T) \tag{2.30}$$

$$G = -kT\, lnQ(N,P,T) \tag{2.31}$$

where k is the Boltzmann constant, T is temperature and Q is the partition function. However Q is dependent on the volume of the phase and therefore cannot be determined directly. Furthermore, free energy of a system cannot be expressed as an ensemble average. [66,67] One approach is to calculate the free energy differences ΔA or ΔG between two states A and B, which is usually the parameter of interest. [68]

$$\Delta A = -kT\, ln\, \frac{Q_A}{Q_B} \tag{2.32}$$

If the masses of states A and B are the same, which corresponds to different states of the same molecule, then ratio of configurational integrals could be used instead of partition functions which then gives

$$\Delta A = -kT\, ln \left\langle exp\left(-\frac{1}{kT}[\Delta U]\right)\right\rangle \tag{2.33}$$

where ΔU is the difference between the potential energies of the A and B states. As free energy is a state function that does not depend on the path taken, one could obtain ΔU irrespective of the path taken. This makes the calculation of free energy

differences between two states easier by allowing a different reversible path that is more convenient to obtain via simulations. [69]

Note that in computer simulations these reversible paths do not need to be natural paths that exists physically. Considering 2.33 absolute free energy can be obtained through constructing a reversible path by employing non-interacting ghost particles in the thermodynamic cycle. [64]

Several methods have been put forward to calculate free energy difference along a trajectory such as umbrella sampling, thermodynamic integration, free energy perturbation, slow growth, non-equilibrium work and constraint force pulling. [70] The first two of these methods are used in this thesis.

Thermodynamic Integration Thermodynamic integration modifies the Hamiltonian by introducing a coupling parameter such that the Hamiltonian becomes $H(p, q, \lambda)$. Here the coupling parameter defines the two states with $\lambda = 1$ and $\lambda = 0$, and a number of simulations are run for different λ values between 1 and 0. The free energy difference between the states can be obtained via

$$A(\lambda = 1) - A(\lambda = 0) = \sum_{\lambda=0}^{\lambda=1} d\lambda \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \tag{2.34}$$

which uses the ensemble average of the difference in potential energy with respect to λ, and performs Gaussian quadrature along λ. The process should satisfy the inequality $\partial^2 A/\partial \lambda^2 \leq 0$ which can be used to test validity of the simulations. [71]

Thermodynamic Integration is used in calculating solvation and ligand binding free energies. For particle insertion methods, $\lambda = 0$ corresponds to the solute behaving as a ghost particle that does not interact with its environment, and $\lambda = 1$ is the state where solute-environment interactions are fully represented.

Umbrella Sampling Umbrella sampling is a method designed to extract the free energy difference of two states by applying a bias potential to the parameter of interest along the reaction coordinate. [72] The reaction coordinate is divided into n overlapping windows wherein the parameter of interest is biased towards a target value. [73]

Umbrella sampling is generally applied to calculate ligand binding energy [61], and solvation of a particle.

As the system contains a bias potential, one must remove the effects of the bias to calculate PMF. To this end, weighted histogram analysis (WHAM) method can be utilised. [74] One could get the unbiased probability distribution $P(\xi)$ using the histograms of each window

$$P(\xi) = \frac{\sum_{i=1}^{N_w} g_i^{-1} h_i(\xi)}{\sum_{j=1}^{N_w} n_j g_j^{-1} exp[-\beta(w_j(\xi) - f_j]} \tag{2.35}$$

where $h_i(\xi)$ is the histogram of the ith window, β is the inverse temperature $1/k_B T$, with k_B being the Boltzmann constant, T temperature. n_j is the number of data points in jth window, f_j is the free energy constant, N_w is the number of windows. g_i is given by the equation $g_i = 1 + 2\tau_i$ where τ_i is the integrated autocorrelation time of the ith window. [75, 76]

The integrated autocorrelation time

$$\tau_i = \sum_{\Delta t}^{\infty} R_i(\Delta t) \tag{2.36}$$

where $R_i(\Delta t)$ is the normalized autocorrelation function for the ith window. Using the unbiased probability distribution, the PMF could be obtained

$$W(\xi) = -\beta^{-1} \ln\left(\frac{P(\xi)}{P(\xi_0)}\right) \tag{2.37}$$

where ξ_0 is a reference point along the reaction coordinate that makes the PMF zero, $W(\xi_0) = 0$.

Using the trajectories, ξ_i, of each window in umbrella sampling simulations, new trajectories, ξ_i^b, are generated by randomly selecting the data points where the probability of choosing each data is the same. Here, the superscript b denotes *bootstrap*, referring to the fact that the actual trajectories are resampled to obtain a new set of trajectories, namely the bootstrap trajectories. One can think of the bootstrap data as hypothetical observations. While forming these new trajectories, each data is allowed to be selected multiple times.

Histograms, h_i^b, of the new set of trajectories are calculated which are then used to generate the bootstrap PMFs, W_k^b where k denotes the index of bootstrap sets, from 1 to N_b, N_b being the total number of the bootstrap sets. The uncertainty of the PMF $W(\xi)$ is equal to the unbiased standard deviation of N_b bootstrap PMFs $W_b(\xi_i^b)$.

$$\sigma^2(\xi) = \frac{\sum_{k=1}^{N_b} \left(W_k^b - \langle W^b \rangle\right)^2}{N_b - 1} \tag{2.38}$$

where $\sigma(\xi)$ is the unbiased standard deviation and $\langle W^b \rangle$ is the average of the bootstrap PMFs. [75]

The GROMACS toolkit g_wham allows three approaches to calculate the error estimate. One is that, the errors can be calculated from the bootstrap analysis as described above. This approach quantifies the uncertainties correctly if each umbrella window samples the phase space effectively (the more technical definition is the equality of the autocorrelation time of each window). However, for simulations of inhomogenous systems such as those involving large biomolecules, e.g. membranes, the assumption that each histogram h_i is a representation of the underlying distribution, underestimates the errors and therefore needs to be modified.

If the umbrella windows cannot sample the phase space effectively, the advice is to increase the same windows with different initial configurations. Histograms of all windows are then pooled together to form a group of "complete histograms". Each histogram (including the multiple runs for each window) is treated as an "individual observation" and a new set of histograms are selected from the "complete histograms" group, where multiple selections are allowed. Compared to the above method, no new trajectories or histograms are generated. Instead available histograms are resampled and the error estimate is calculated using the bootstrapped PMFs. This approach ensures a more accurate error estimate for the simulations where the phase space is hard to sample.

In the "complete histograms" method, one should pay attention to the overlap of the bootstrapped histograms to obtain a correct PMF. If the randomly chosen his-

tograms leave a gap along the reaction coordinate then PMF could not be calculated correctly. Instead of choosing the histograms randomly, histograms can be grouped along the reaction coordinate and each group can be separately bootstrapped to avoid any possible gaps in the reaction coordinate. [77]

The third method, called "Bayesian bootstrapping of complete histograms" ensures the overlap of the histograms by assigning random weights, w_i, to all histograms to avoid a non-intersecting set of histograms. The probability of choosing a histogram becomes w_i instead of $1/n$. [78]

2.3 Replica Exchange Method

Complex molecules such as proteins, membranes, polymers etc. have rugged free energy landscapes with many local minima separated by high energy barriers. Adequately sampling the phase space for such systems requires A number of copies of the system with different temperatures are run in parallel and at regular intervals an exchange between neighboring replicas is attempted. The exchange is accepted based on a Metropolis criterion shown in Eq. 2.39. The idea is to excite the system with the help of high temperatures and to scan for other low energy conformations that were otherwise unavailable due to high energy barriers.

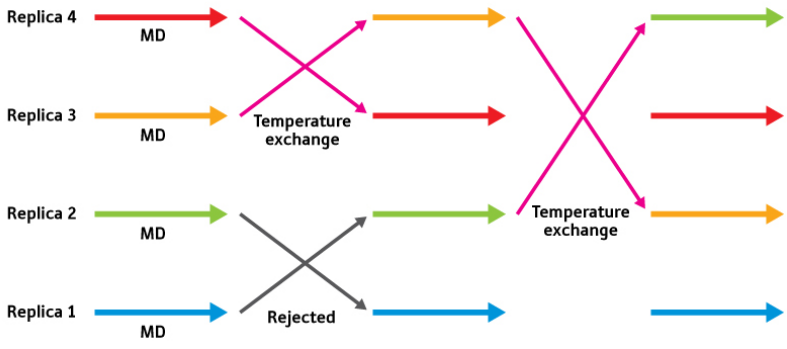


Figure 2.2: Schematic flow of replica exchange molecular dynamics simulation. Replicas at different temperatures are run in parallel and at regular intervals exchanges between replicas are attempted. The diagram is taken from Riken Research Website [79]

Figure 2.2 shows a schematic flow of REMD, where replicas of the same structure at different temperatures are run in parallel, independent of each other. At regular intervals, the configurations of the neighboring replicas are exchanged based on a probability criteria shown in Equation 2.39. The exchanges should be done often to sample the phase space adequately.

$$P(1 \leftrightarrow 2) = min(1, exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right)(U_1 - U_2)\right] \tag{2.39}$$

where $P(1 \leftrightarrow 2)$ is the probability of an exchange between replicas 1 and 2. k_B is the Boltzmann's constant, T_1 and T_2 are the temperatures and U_1 and U_2 are the instantaneous potential energies of replicas 1 and 2. [80] As a result, high energy conformations including the ones that are kinetically trapped are replaced by conformations with lower energies.

The minimum temperature is chosen to be the temperature of the system and maximum temperature should be chosen such that the conformations can escape out of local minima. The temperatures are distributed to obtain an exchange probability. As the number of degrees of freedom increases the temperature difference between the neighboring replicas should be smaller. Therefore for large systems, if one wants to set a high temperature, then the number of replicas should be high. One alternative is to keep the number of replicas low and decrease the exchange probability. However, reducing the exchange probability leads to lower number of successful exchanges and therefore delays sampling of the phase space. [81]

The idea is extended to various parallel tempering methods such as *Hamiltonian replica exchange* which allows the use of a Hamiltonian $H(q, p)$ as a variable between replicas instead of temperature. *Gibbs sampling* extends exchanges to non-neighboring replicas, speeding convergence.

2.4 Coarse-Graining

To alleviate the high demand of computational power demanded for simulating large systems or for longer time steps, alternative techniques such as coarse-graining may

be used. CG decreases the number of degrees of freedom of a system by representing a group of atoms as a single superatom. As a side hand rule, the masses of the superatoms should be similar, and the regions that have different properties (such as backbone vs. sidechain) should be mapped to different superatoms. However these are flexible and act as mere suggestions instead of rules. As the CG model has less number of degrees of freedom some information regarding the atomic model will be lost. Therefore the important part is to clearly define and focus on the properties to be reproduced. [39]

Although there are several approaches for determining the interactions between these superatoms, there is no single recipe that guarantees reproducing the parameter of interest. One approach is to define the interactions between the superatoms by matching the solvation free energies of their atomistic counterparts [41]. Another example is Boltzmann inversion (BI) method that aims to match the atomistic radial distribution function (RDF). [44] The BI method has been extended to an iterative scheme that updates the pair potentials gradually. [35, 45] Force matching aims to mimic the atomistic forces. [37]

Chapter 3

PHENYLALANINE DIPEPTIDE: A TWO-STATE SYSTEM FOR INVESTIGATING CONFORMATIONAL TRANSITION

3.1 Introduction

Peptide aggregation may be triggered by environmental changes such as change of solvent, pH or concentration. Such cases are visible especially for the amphiphilic peptides at hydrophilic/hydrophobic interfaces where the hydrophobic and the hydrophilic aminoacid residues of the peptides interact selectively with the solvents. [82–84]

Well known examples of the interplay of conformational change and aggregation or partitioning at interfaces are the misfolding of proteins upon amyloid aggregation [85, 86], or more generally the induction of higher β-sheet content by aggregation [87] or by the presence of an interface [88, 89]. Such examples are important for a large variety of fields, such as biomedical research, food science, or material science [90–95] where folding, aggregation, interaction with interfaces, and partitioning of hydrophobic and hydrophilic residues influence each other.

In order to better understand the structure formation in peptide aggregates knowledge of the relevant interactions, driving forces, pathways and assembly mechanisms is essential. Here, molecular simulation can be a valuable tool to provide microscopic structural and thermodynamic insight into such systems, complementing experimental data and analytical models.

However, the processes of aggregation and conformational change generally occur at time scales on the order of hundreds of microseconds to days, where the molecular

simulations of all-atomistic models are not able to simulate due to computational restrictions. To this end, coarse-grain models are employed where the atomistic degrees of freedom are reduced to enable access to higher time scales. The CG models are engineered to mimic the desired properties of their higher resolution counterparts.

A common problem associated with the CG models is that the interaction parameters are generated to represent one thermodynamic state. However, the process of conformational switch contains at least two distinct states: e.g. one state may be the configuration of the peptide in bulk water, and the other state may be the configuration at the hydrophobic/hydrophilic interface. Therefore representing the two different environments with one set of interaction potentials presents problems in terms of reproducing the structural or thermodynamic properties of target systems which are more detailed.

Available methods for parametrization of nonbonded interaction potentials can be divided into two general categories: *(i) thermodynamics-based* methods where the CG parameters are refined so that the system displays a certain thermodynamic behavior [96–100] or *(ii) structure-based* methods where the CG system aims at reproducing the configurational phase space sampled by an atomistic reference system [101–113]. Inherent to the process of coarse graining, the resulting CG models typically suffer from *representability* problems [114]: i.e. a structure-based approach does not necessarily yield correct thermodynamic properties such as solvation free energies or partitioning data while thermodynamics-based potentials may not reproduce microscopic structural data such as the local packing or the structure of solvation shells. Thus, a careful assessment of the system and problem of interest needs to be made before choosing an appropriate coarse graining strategy.

Closely related problems arise from the limited *transferability* of CG models: all CG models (in fact also all classical atomistic forcefields) are state-point dependent and cannot necessarily be transferred to different thermodynamic conditions such as temperature, density, concentration, system composition, phase, etc. or a different chemical or molecular environment such as a certain chemical unit being part of

different macromolecular chains. Structural and thermodynamic representability and state-point transferability questions are often intimately linked, since the response to a change in state point corresponds to representing certain thermodynamic properties, and intensive research is currently devoted to this problem [109, 110, 115–121].

To study biomolecular processes such as peptide aggregation and biomaterials structure formation, CG models are often developed based on smaller less complex reference systems (individual peptides, subsystems, etc.). Consequently, it is essential to achieve transferability among different concentrations and environments, to maintain correct partitioning and solvation properties, and to reproduce coupling between conformational behavior and environment.

Diphenylalanine (FF) repeat unit has been suggested by several experimental as well as computational studies as an important motif in amyloid formation. FF is a minimal molecule which exhibits the conformational transition seen in proteins upon formation of amyloid fibrils. And it is important to understand why and how proteins/peptides display such a conformational switch.

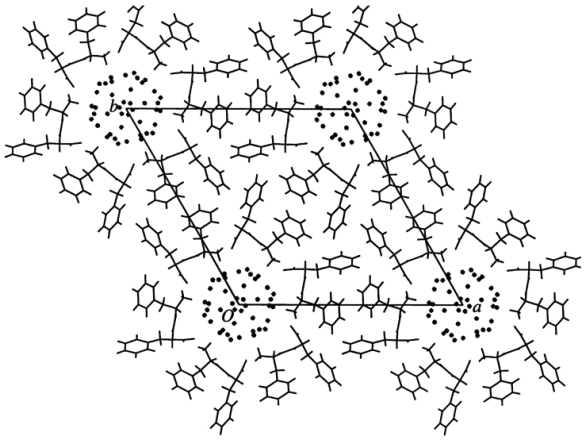


Figure 3.1: Molecular packing of diphenylalanine. Taken from [122].

Diphenylalanine (FF) assumes an extended conformation when immersed in bulk water and when placed at hydrophilic/hydrophobic interfaces and in aggregates (Fig. 3.1), FF adopts a compact conformation (Fig. 3.2). The conformational switch of FF coupled by its small size makes FF an ideal molecule to study the effect of

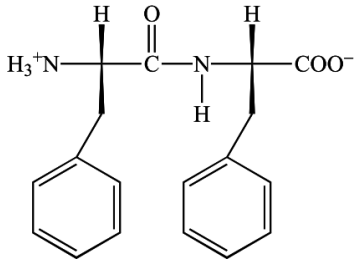


Figure 3.2: Diphenylalanine molecule. Taken from [122].

non-bonded interactions on its conformational behavior. Here we first demonstrate this environment-induced conformational switch using all-atomistic MD simulations. Our aim is to generate a transferable CG model of FF that is able to mimic this conformational switch.

This study focuses on the interplay between bonded and nonbonded interactions, in particular upon the transition from a bulk water environment to a hydrophilic/hydrophobic interface, i.e. cyclohexane/water. The all-atomistic resolution of an FF molecule is used as the model system, and the study demonstrates how one can parametrize a CG model capable of representing the conformational properties of the molecule in two different environments.

3.2 Methods

Atomistic MD simulations were performed for a single FF molecule in bulk water and in a cyclohexane/water biphasic system representing the behavior at a hydrophobic/hydrophilic interface by using GROMACS 4.5 simulation package [123]. For the bulk water simulation a cubic box with a dimension of 5 nm is used and it contains 4123 water molecules. The cyclohexane/water biphasic system for the interface simulations is performed in a box of size 4x4x6.5 nm's and contains 1601 water and 267 cyclohexane molecules. Box sizes are chosen to be larger than twice the largest cut-off value to prevent interactions with periodic images.

Atomistic simulations were obtained by using the leap-frog integrator with a time-step of 2 fs. GROMOS 53a6 force-field [124] is used for the FF molecule and SPC/E

model [125] for water. Hydrogen bonds were constrained with Lincs algorithm [126]. Temperature is set to 298 K via velocity-rescaling algorithm [59] with a coupling time constant of 1 ps. For bulk water simulations isotropic Berendsen pressure coupling at a reference pressure of 1 atm with a coupling time constant of 1 ps and a compressibility of 4.5e-5 bar^{-1} is used. For the cyclohexane/water interface simulations semi-isotropic pressure coupling is used such that the pressure along the z direction is maintained at 1 atm and dimensions of the box in x and y directions are held constant. Electrostatic interactions are handled by particle mesh Ewald (PME) [127] with a Coulomb cutoff of 1 nm.

The conformational distributions of the atomistic simulations of FF in both environments are used as the references for devising the CG model. The atomistic reference systems were mapped on to the CG model representations of the molecules by using VOTCA package [128], while maintaining the same number of molecules. CG simulations were performed in canonical ensemble (NVT) to avoid large volume fluctuations. The volume of the CG box was reduced, to match the reference pressure of 1 atm: for bulk water simulations a cubic box of length 4.74 nm and for the interface simulations a rectangular box of size 4x4x6 nm's . VDW interactions were cut off at 1.4 nm. For the CG simulations a larger timestep (5 fs) was used, which accelerates the CG simulations by a factor of 2.5, while maintaining correct sampling for the stiffest bond in our CG model. All remaining simulation parameters were kept identical to the atomistic case.

Surface tension of the atomistic and CG interface systems were calculated from the difference between the normal and the lateral pressure as described in detail in the Gromacs manual [129] Eqns. 3.61 and 3.62. All molecular visualizations were performed by VMD [130].

Block analysis method was used to check the convergence of the probability distributions for all systems. Analysis results for atomistic and CG interface systems are given in Supplementary information Figs. S1 and S2.

Free energy calculations were performed by thermodynamic integration using Ben-

nett acceptance ratio (BAR) method. [66,131] The free energy calculation simulations are performed under isothermal-isobaric ensembles (NPT) with stochastic dynamic integrator [132], using Berendsen pressure coupling [60].

Umbrella sampling [133] was used for PMF calculations where a harmonic constraint potential was applied between the centers of mass of the water slab and the FF dipeptide. Cosine weighting [134] was applied for the center of mass of the water slab. FF dipeptide is pulled along the direction perpendicular to the interface (z-direction) with a pull rate of 0.01 nm/ps with a harmonic constant of 1000 kJ mol^{-1} nm^{-2} and snapshots of the system are taken at regular intervals. Energy minimization using the steepest-descent algorithm and an NVT equilibration for 50000 steps were subsequently applied to each of these snapshots which were then used as the initial configurations for the umbrella sampling windows. Each constraint distance for the full-atomistic and CG simulations were sampled for 10 ns and 50 ns, respectively. Weighted histogram analysis method [74] was used to obtain the PMF curves.

3.3 Conformational Transition Upon Environmental Change

In its crystal structure, FF adopts a cis-like conformation, where the two phenyl rings point in the same direction (see Fig. 3.1) [122]. Here, it will be shown that a single FF molecule in water displays a trans-like conformation, and that the corresponding structural change can be triggered by exposure to a hydrophobic interface or aggregation. As such, the FF molecule is an ideal model system for our study, and in order to understand and characterize the conformational preference of the FF molecule we will first present the results of three different atomistic simulations.

When a single FF molecule is simulated in bulk water, the molecule adopts a trans-like conformation as shown in Fig. 3.3 bottom. Such a conformation relaxes the backbone dihedral angles, but the hydrophobic phenylalanine side chains are separated from each other and fully hydrated. The different conformations of FF can best be characterized via a dihedral angle between four superatoms (also indicated in Fig. 3.3). Note that, later these superatoms are also going to be used as the

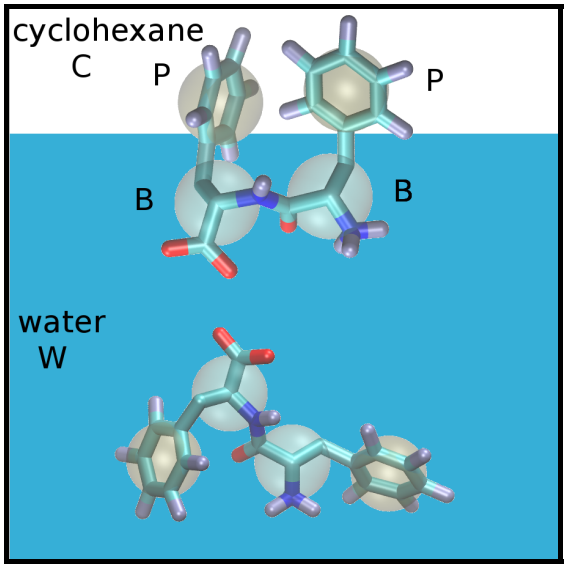


Figure 3.3: Diphenylalanine peptide in cis-like conformation at the cyclohexane/water interface (top) and in trans-like conformation in bulk water (bottom). Mapping of the peptide to a 4 bead CG model (**PBBP**, **P** for sidechain and **B** for backbone beads) is also shown. Solvent molecules (shown as continuum for visual clarity) are mapped on to single beads: **W** and **C** for water and cyclohexane, respectively.

interaction sites in the CG model, i.e. they reflect the mapping scheme that relates the CG beads with the atomistic coordinates of FF. Two superatoms which are positioned in the respective centers of mass of the phenylalanine rings will be referred to as **P**. The backbone and charged end-groups of the FF molecule will be represented by two identical superatoms, which are located at the positions of the two C_α atoms and which will be referred to as **B**. The probability distribution for the dihedral angle among **PBBP** superatoms in bulk water is peaked at an extended conformation with a maximum around -150 degrees (Fig. 3.4 solid black line).

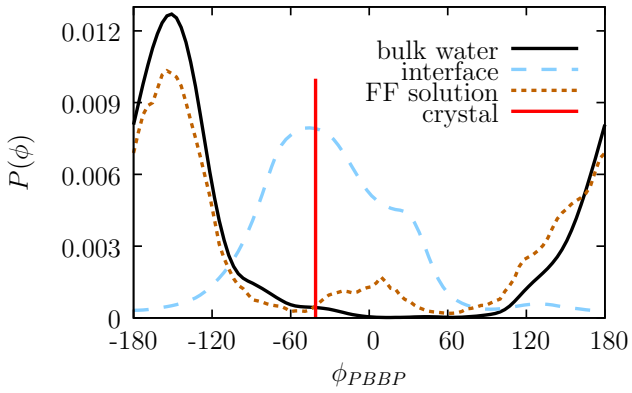


Figure 3.4: Probability distribution for the **PBBP** dihedral angle from atomistic simulations. In bulk water (solid black line), the molecule adopts an extended structure. At the cyclohexane/water interface the molecule adopts a cis-like conformation (dashed blue line), highly distinct from bulk water behavior, but similar to the crystal structure (solid red impulse). In a finite-concentration solution of FF peptides molecules can switch from a trans-like to a cis-like conformation (probability distribution for such a molecule is shown with dashed brown line), when presented with a hydrophobic interface formed by a small cluster of FF molecules.

In atomistic simulations of a biphasic cyclohexane/water environment, the FF molecule (initially placed in the water layer) quickly adsorbs to the interface, and adopts a cis-like conformation as indicated in the representative snapshot in Fig. 3.3 top. The dihedral probability distribution at the interface (Fig. 3.4 dashed blue line) is very distinct from the bulk water distribution and its maximum matches perfectly with the dihedral adopted in the crystal structure [122] (shown with the red impulse line in Fig. 3.4). Phenyl side chains are driven into the cyclohexane phase via

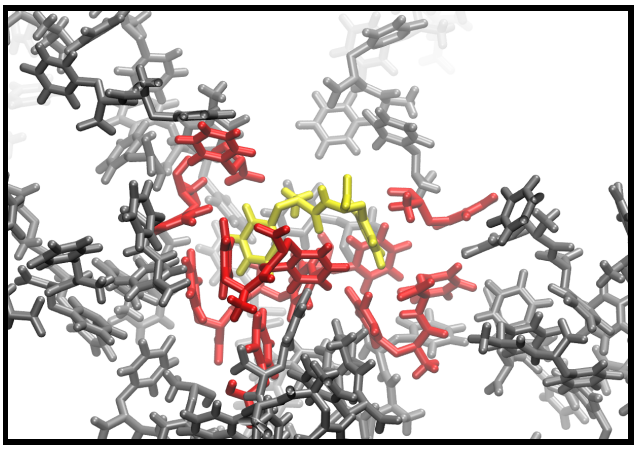


Figure 3.5: A diphenylalanine in cis-like conformation (shown in yellow) at the interface of a nano-cluster of diphenylalanine molecules in water. Water molecules are not shown for clarity. The nano-cluster of FF molecules presents a hydrophilic/hydrophobic interface, which enables the transition of the corresponding molecule to a cis-like conformation. Peptide residues within 4 Å of the corresponding peptide are shown in red color.

hydrophobic forces and this force triggers an environment mediated conformational transition in the molecule.

The conformational transition observed at the interface can also take place in bulk water with a finite concentration of FF molecules as observed in both the simulations presented here and previously by Jeon et. al. [135] As the FF molecules start forming an aggregate the hydrophobic environment, provided by the surrounding phenylalanine side chains, lower the energy barrier for the transition to a cis-like conformation. As a result, a certain fraction of the FF molecules sample the cis-like conformation during the simulation. In the presence of multiple FF molecules, the hydrophobic environment provided by the neighboring FF molecules could lower the energy barrier for transition to the cis conformation. We have performed an atomistic simulation with 101 FF molecules and 15274 water molecules, where all of the FF molecules are initially in trans conformation and randomly distributed in the box. Simulations are performed in the NPT ensemble with identical settings to the single FF molecule in bulk water simulations. In Fig. 3.6 both the probability distribution averaged over

all of the FF molecules, and the distributions for selected FF molecules are given. Molecule 6 represents the behavior for the majority of the FF molecules, it remains in trans conformation through out the 200 ns simulation. On the other hand molecules 9, 26, 28, 42 and 69 switch to the cis-conformation during the simulation as seen from the probability distributions similar to the behavior of molecule 24 for which the timeline is given in Fig. 3.7 and histogram of the dihedral probability distribution is given in Fig. 3.4.

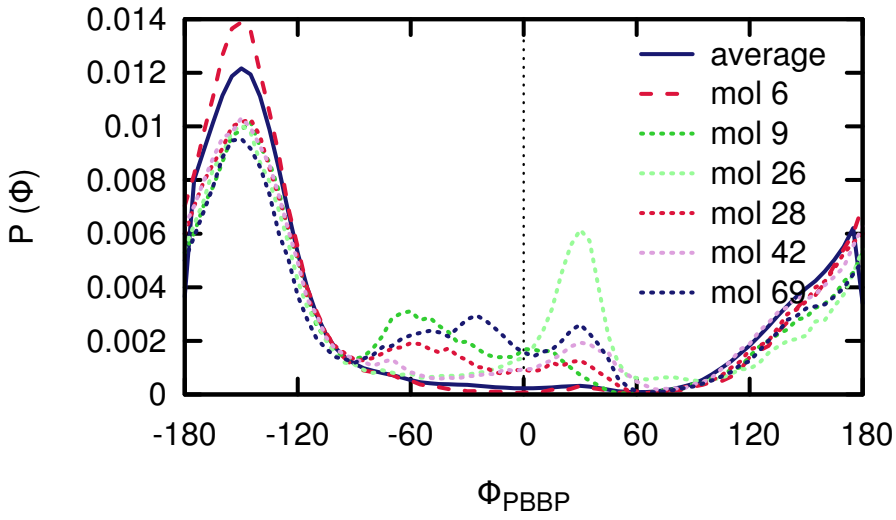


Figure 3.6: Probability distribution of the dihedral angle from a finite-concentration solution of FF peptides. Dihedral distribution averaged over all of the peptides, as well as the results for individual distributions for selected FF molecules are shown.

A representative snapshot for the conformation of such a molecule is shown in Fig. 3.5. The yellow colored molecule displays a cis-like conformation when its sidechains are in contact with a hydrophobic/hydrophilic interface of a nano-cluster of neighboring molecules. The dihedral distribution of the corresponding molecule clearly shows (Fig. 3.4 dashed brown line) two peaks, one corresponding to the trans-like (characteristic for bulk water environment) and one to the cis-like (characteristic for the interface behavior) conformation. This behavior is not specific to a single molecule, but can be seen for many such molecules (Fig. 3.6). Note that this con-

formational transition is only temporary as seen in the time line evolution of the dihedral angle (Fig. 3.7). During the 200 ns simulation, FF molecules form small clusters, however no stable nucleus for crystallization is observed.

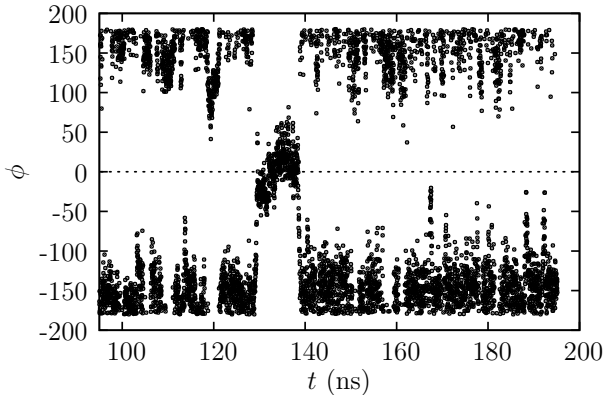


Figure 3.7: Dihedral angle among **PBBP** superatoms of the molecule in Fig. 3.5 as a function of simulation time. Unlike the peptide in bulk water (where only trans-like conformation is observed), in a solution of peptides in water, formation of nano-clusters with hydrophilic/hydrophobic interfaces enables sampling of the cis-like conformation.

Based on these observations on the atomistic model of the FF molecule, in the next section we will develop a coarse-grained model which can capture the behavior in these two distinct states.

3.4 Transferable CG Model

A common approach in development of coarse-grained models is to adopt a Hamiltonian with separate terms for bonded and nonbonded interactions:

$$U^{cg} = U^{cg}_{bonded} + U^{cg}_{nonboned} \tag{3.1}$$

where U^{cg} is the total coarse-grained potential. Similarly, bond-length, angle-bending and dihedral interactions are included into the Hamiltonian as separate terms.

$$U^{cg}_{bonded} = U^{cg}_{r} + U^{cg}_{\theta} + U^{cg}_{\phi} \tag{3.2}$$

Often the parametrization of the different terms is done independently, implying that the relevant degrees of freedom are decoupled, or even if the resulting bonded and nonbonded distributions are weakly correlated, the coarse-grained model will successfully mimic the all atom case and display an identical or similar coupling. Hence, the strategy to devise nonbonded interactions plays no role for parametrization of the bonded potentials. A discussion on these assumptions, where they fail, and how the problems can be overcome has already been provided elsewhere [136, 137].

Conformational probability distributions, obtained by mapping the atomistic configurations to the respective coarse grained coordinates, can be used to obtain potentials of mean force (PMF) for each interaction [101, 138, 139].

$$U^{cg}(r,T) = -k_B T \; ln(P^{cg}(r,T)/r^2) + C_r \tag{3.3}$$

$$U^{cg}(\theta,T) = -k_B T \; ln(P^{cg}(\theta,T)/sin(\theta)) + C_\theta \tag{3.4}$$

$$U^{cg}(\phi,T) = -k_B T \; ln(P^{cg}(\phi,T)) + C_\phi \tag{3.5}$$

where $P^{cg}(r,T)$ is the bonded, $P^{cg}(\theta,T)$ is the angle-bending, and $P^{cg}(\phi,T)$ is the dihedral distribution. k_B is the Boltzmann constant, T is the temperature, and C's are constant values.

The next section will introduce the CG solvent models for water and cyclohexane and present the performance of the models by comparing with the AA solvent models.

3.4.1 CG Solvent Model

To analyze the conformation of the FF molecule both in bulk water and at the cyclohexane/water interface, an explicit solvent CG model is used. Both the water and cyclohexane solvent molecules are mapped onto one bead per molecule (**W** for water and **C** for cyclohexane), which is positioned at the center of mass of the corresponding solvent molecule. For the parametrization of the nonbonded solvent-solvent interactions a Lennard-Jones 9-6 potential is used. The radii (σ) for the Lennard-Jones interaction are matched to the first peak in the radial distribution functions in the AA simulations. The depth of the Lennard-Jones interaction (ϵ) is tuned to reproduce

AA solvation free energies of the respective groups, which are measured via thermodynamic integration [140]. All free energy values (both from AA and CG simulations) that are discussed in the following, as well as the ϵ and σ values for all solvent parameters are listed in Table 3.1. The respective potentials are also graphically depicted in Fig. 3.8.

Ref./Sol.	System	VdW kJ/mol	Elec. kJ/mol	Total kJ/mol	ϵ kJ/mol	σ nm
W/W	AA	9.02	-39.24	-30.22	-	-
	CG	-28.36	-	-28.36	4.64	0.288
C/C	AA	-17.60	-	-17.60	-	-
	CG	-16.95	-	-16.95	4.0	0.547
W/C	AA	2.86	-	2.86	-	-
	CG-FE	2.71	-	2.71	1.33	0.455
	CG-ST	-10.62	-	-10.62	3.19	0.455
C/W	AA	9.17	-	9.17	-	-
	CG-FE	114.4	-	114.4	1.33	0.455
	CG-ST	41.30	-	41.30	3.19	0.455

Table 3.1: Solvation free energies of water and cyclohexane molecules for the constituent subphases of the interface system measured from atomistic (AA) simulations. Van der Waals and electrostatic contributions, as well as the change in total free energy are listed. Lennard-Jones interactions, where parameters ϵ and σ are tuned to reproduce the measured atomistic free energy, are used for the nonbonded interactions of water (**W**) and cyclohexane (**C**) CG beads. Parametrization of the **W-C** interaction is done in two different ways: 1) CG-FE: solvation free energy of water in cyclohexane (**W/C**) is used as reference. 2) CG-ST: interaction is tuned to reproduce the surface tension of the biphasic system.

The pure solvent (i.e. **W-W** and **C-C**) interactions are parametrized to match the AA thermodynamic integration results from the self-solvation of the corresponding molecules. For the parametrization of the (mixed) **W-C** interaction the case is a little more ambiguous: in principle one has two reference values from the solvation

of cyclohexane in water **C/W** and from the solvation of water in cyclohexane **W/C**, and ideally the mixed interaction should reproduce both values.

As a first approach the CG **W-C** interaction was tuned to reproduce the AA solvation free energy of water in cyclohexane (**W/C**). The CG free energy changes linearly with the depth of the potential ϵ, and after a few iterations the depth was set as ϵ=1.33 kJ/mol, which yields the CG solvation free energy of **W/C** to 2.71 kJ/mol – fairly close to the AA result of 2.86 kJ/mol. Using this interaction function, the CG solvation free energy of cyclohexane in water (**C/W**) is measured as 114.4 kJ/mol, which is an overestimation compared to the AA result (9.17 kJ/mol), i.e. cyclohexane beads are too strongly repelled from the aqueous phase. If, alternatively, one chooses to tune the CG interaction potential for **W-C** to match the **C/W** AA solvation free energy, one obtains a very deep **W-C** interaction (data not shown). This model results in a negative **W/C** solvation free energy, i.e. the solvation of water in cyclohexane becomes highly favorable, destabilizing the interface. Hence, this model clearly leads to unphysical results. It turns out that with the approach chosen one cannot simultaneously reproduce both the **W/C** and the **C/W** solvation free energy in the CG simulations. The first CG model – tuned based on the **W/C** solvation free energy – will be referred to as CG-FE solvent model in the remainder of the text.

The **W-C** interaction determines the interfacial properties of the system such as surface tension which gives information about how the two slabs interact, e.g. the amount of mixing at the interface. To be able to successfully mimic the AA system, interfacial properties should also be taken into account in the CG model. To this end, we compared the surface tension of the CG-FE model with that of the AA model and observed that the CG-FE model yields a surface tension of 157 ± 0.3 mN/m, a value much higher than the surface tensions of AA cyclohexane/water simulation which is 51 ± 0.1 mN/m. The surface tension results show that the amount of repulsion at the interface between the cyclohexane and water molecules is higher for the CG-FE model compared to that of the AA model. Moreover, given the ambiguity discussed

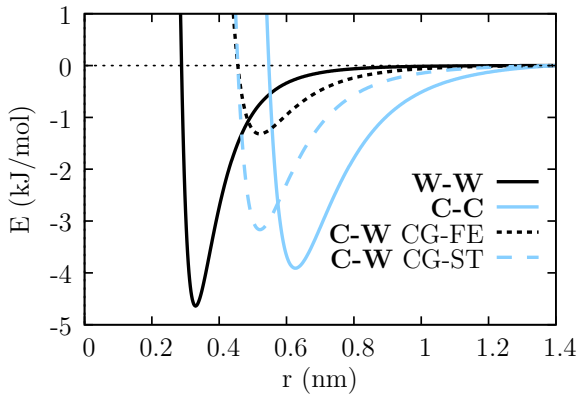


Figure 3.8: CG non-bonded interaction potentials for the solvent beads (Lennard-Jones parameters are given in Table 3.1). **W-W** and **C-C** interactions are shown with solid black and blue lines, respectively. Two different potentials are used for the **W-C** interaction, where either the solvation free energy of water in cyclohexane (CG-FE, dashed black line) or the surface tension of cyclohexane/water system (CG-ST, long-dashed blue line) is taken as a reference.

above the **W-C** interaction potential can be tuned to yield a correct surface tension.

To this end, the **W-C** interaction potential was alternatively tuned to reproduce the surface tension of the cyclohexane/water system with the target surface tension of that of the AA simulations where the surface tension of the new CG model was obtained as 50.4 ± 0.75 mN/m. The resulting model will be referred to as CG-ST solvent model where ST refers to surface tension. Figure 3.8 shows that the **W-C** potential is significantly deeper than that of the CG-FE but not as deep as the discarded free-energy model tuned from the **C/W** solvation which led to mixing of the solvents and destabilized the interface. Therefore CG-ST by construction better represents the interface properties. However, this is achieved at the cost of the solvation free energy of **W/C** calculated as -10.62 kJ/mol, i.e. water is now more favorable inside the cyclohexane phase. However the solvation free energy for **C/W** is found to be 41.30 kJ/mol, a value that is closer to that of the AA model of 9.17 kJ/mol than the CG-FE model which yielded a far off value of 114.4 kJ/mol (see Table 3.1). These results show that unlike the CG-FE model, the CG-ST model does not take sides between reproducing the solvation free energies of **W/C** and **C/W** and instead

makes a sacrifice, improving the latter at the expense of the former.

As a further comparison of the CG solvent models with the AA model, radial distribution function (RDF) of water-water and cyclohexane-cyclohexane superatoms are obtained for the biphasic system with two cyclohexane/water interfaces (i.e. the interfaces at the top and bottom of the simulation box). AA simulation trajectory is mapped to the CG model and g(r) is calculated for the mapped system. Figure 3.9 shows the comparison of the RDF results for the CG-FE and AA models. Considering that the pair interaction potentials were parametrized to reproduce solvation free energy of the corresponding molecules, the match between AA and CG g(r) curves is surprisingly good. The pair potential used for the cyclohexane-cyclohexane interaction among CG beads yields a good representation of the cyclohexane phase. However, pair potential used for water-water interactions yield an over structured distribution compared to the atomistic case, which is typical for Lennard-Jones representation of water molecules.

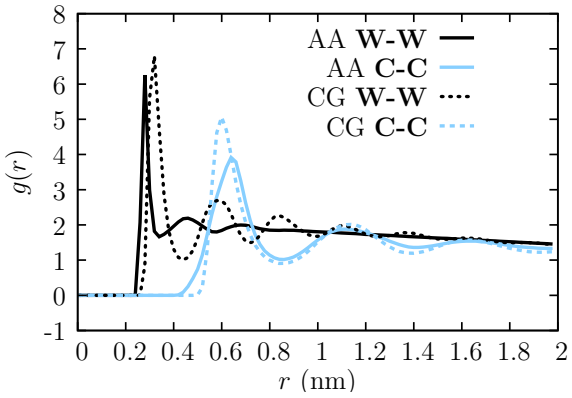


Figure 3.9: Comparison of water-water (W-W) (black color) and cyclohexane-cyclohexane (C-C) (blue color) radial distribution functions in AA (solid lines) and CG (dashed lines) simulations.

Figure 3.10 depicts the density profile of the cyclohexane/water system for AA (solid lines) and CG-FE (dashed lines) models. CG-ST yields a slightly broader interface (not shown), due to the lower surface tension compared to CG-FE. Water and cyclohexane molecules' profiles are shown with black and blue lines, respectively. CG

water beads display a higher density, since the model is tuned to capture the solvation free-energy via spherically symmetric Lennard-Jones potential. The amplified undulations observed in the CG representation of cyclohexane is a finite size artifact. Small number of molecules and presence of two sharp interfaces leads to an over structured density profile for cyclohexane and formation of a diffuse layered-structure is observed with 7 layers on the average. As verified by the g(r) curves the CG cyclohexane beads are still in liquid phase and molecules diffuse freely in between the diffuse layers. Hence, we conclude that the solvent model we use is adequate for the CG model development for FF molecule.

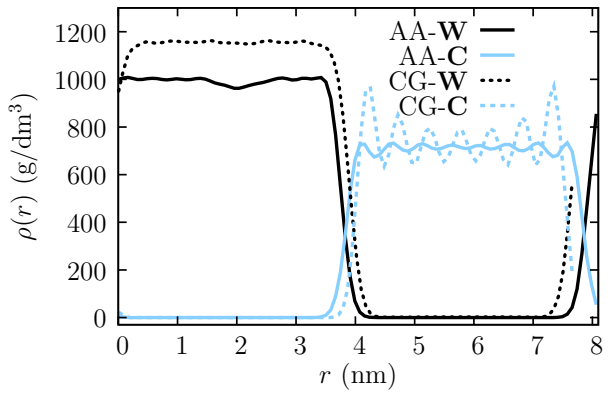


Figure 3.10: Density profile of water (black color) and cyclohexane (blue color) phases in AA (solid lines) and CG (dashed lines) simulations.

In the next section, we present our approach to developing a transferable coarse-grained model for FF, and demonstrate its success and drawbacks compared to the all-atomistic simulations.

3.4.2 CG FF Model

The CG model for FF used here is composed of four beads depicted in Fig. 3.11, with masses of the sidechains assigned to the **P** beads and the mass of the backbone atoms and the end groups distributed equally to the two **B** beads. The **P** beads are mapped to the center of mass of the phenylalanine groups and **B** beads are mapped to the C_α

atoms. Despite mapping onto a single atom, the mass of the **B** beads are the sum of the masses of the backbone atoms as shown in Table 4.3.

beads	position	mass	atoms used for calculating mass
P	CoM of phenyl ring	77.106	C_γ $C_{\delta1}$ $C_{\delta2}$ $C_{\eta1}$ $C_{\eta2}$ C_ζ $H_{\delta1}$ $H_{\delta2}$ $H_{\eta1}$ $H_{\eta2}$ H_ζ
B	C_α	72.0871	NH_3 C_α C_β C O

Table 3.2: Coarse-grain mapping for diphenylalanine showing the position of the superatoms, mass, and the atoms used for calculating the mass of the superatoms.

One alternative mapping could be the inclusion of the C_β atom to the **P** bead. The reason behind the choice in this study was to be able to represent in the CG model the distance between the phenylalanine sidechain and the backbone observed for the AA model. When the **P** beads were mapped onto the C_β atoms, the resulting distribution of the pseudo-dihedral and the angle motions failed to capture the difference in the orientation of FF between the bulk water and the cyclohexane/water interface. Other alternatives where the **B** beads are positioned at the center of mass of the backbone atoms, yield complex dihedral distributions where the contrast between bulk water and the cyclohexane/water interface is not as pronounced as the mapping used in this study.

This small molecule represents an interesting challenge for coarse-graining methodology. In order to design a coarse-grained model which can capture the behavior of the molecule both in bulk water and at the cyclohexane/water interface, one has to reconsider the assumption of the decoupling of bonded and nonbonded interactions where the latter encompass solvent-solvent as well as solute-solvent interactions. The dihedral angle **PBBP** displays an environment driven transition, from a trans-like conformation in bulk water to a cis-like conformation at the interface. It cannot be taken for granted, that the combination of a dihedral potential based on the bulk dihedral angle distribution (which strongly favors the trans state) with the CG nonbonded interactions of P and B beads with the solvent molecules will result in an

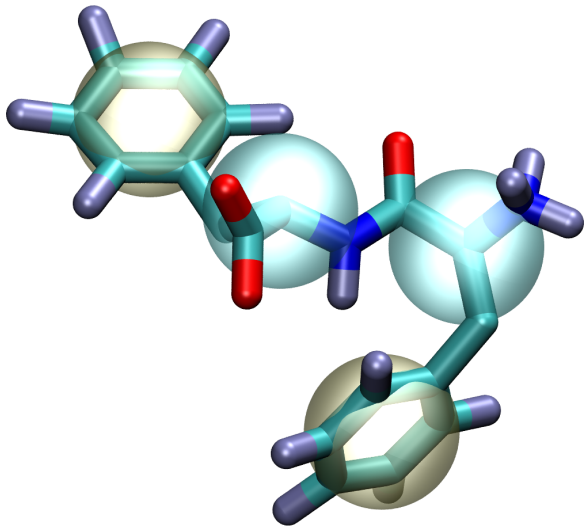


Figure 3.11: Mapping scheme for FF. The phenyl rings are mapped onto the **P** superatoms (brown) and the backbones are represented by the **C** superatoms (blue).

identical change in conformation when the molecule is placed at the interface. One would rather expect that the transition behavior should be strongly dependent on the type of parametrization of the nonbonded CG model, in particular one would suspect the nonbonded interaction of P and B beads with the solvent molecules to be relevant.

As discussed above, the environment-driven conformational transition of FF can be successfully represented in this mapping scheme via the **PBBP** dihedral angle. Since the solvent superatoms are not polarizable and in the present simulations no peptide-peptide interactions are considered, the **B** beads on the peptides do not carry any charge.

With this mapping the CG FF model requires 3 bond, 2 angle, and 1 dihedral angle as bonded interactions, and nonbonded interaction potentials between the peptide and the solvent molecules. As a further simplification, we ignored the small differences in bond and angle distributions between bulk water and at the interface, and focused primarily on the dihedral angle. The interaction potentials for bonds (**PB**, **BP**, **BB**) and angles (**PBB**, **BPP**) are obtained via Boltzmann inversion (Eqns 3.3, and 3.4) of the corresponding probability distributions in bulk water (see Figs. 3.13 and 3.14).

The bonds are represented via harmonic potentials (Table 3.3), whereas the angle

Bond	x_0 (nm)	k_x (kJ/(mol × nm^2))
P1-B1	0.382	16.9×10^3
B1-B2	0.384	82.0×10^3
B2-P2	0.386	21.6×10^3

Table 3.3: Equilibrium bond length and spring constant for the harmonic potentials connecting the CG beads. Beads are numbered from C-terminus to N-terminus.

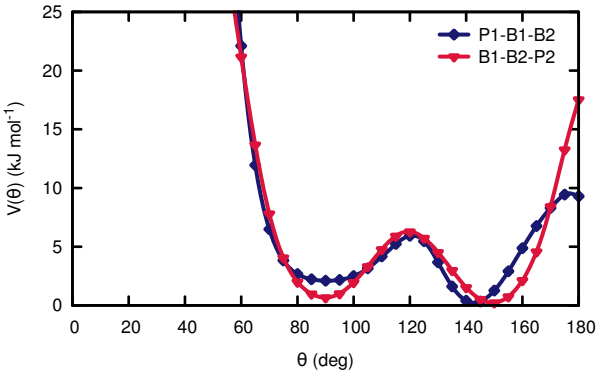


Figure 3.12: Angle potentials for the P1-B1-B2 and B1-B2-P2 angles obtained via Boltzmann inversion of the corresponding atomistic probability distributions.

and dihedral potentials are represented in tabular form (Figs. 3.12-3.16) where all are fitted to the Boltzmann-invert of atomistic distributions. The resulting distributions from the CG model and comparison with the atomistic bulk and interface simulations are given in Figs. 3.14 and 3.4. For the bond and angle distributions the CG model reproduces the AA system well. Hence the focus was on the dihedral angle **PBBP**.

Despite the dramatic change in the dihedral angle upon transfer of the molecule from bulk water to the interface, the bond length and angle distributions show only modest changes. For each bond and angle distribution the bulk and interface results are shown in Figures 3.13 and 3.14, with solid and dashed lines, respectively. Except for a small shift in the BB bond length, in general bond lengths remain identical in bulk water and interface simulations. For the bond angle distributions, a bimodal distribution is observed for both BBP and PBB angles. Even though one can see a

slight shift in peak location and change in height, compared to the dihedral angle these can be considered as secondary minor changes.

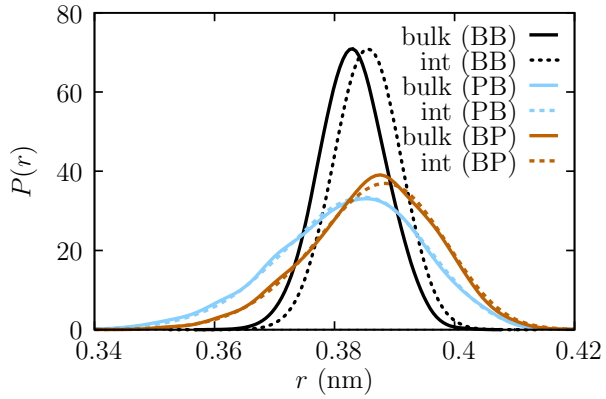


Figure 3.13: Probability distributions for bonds between FF superatoms (BB, PB and BP) from atomistic simulations in bulk water and at the cyclohexane/water interface.

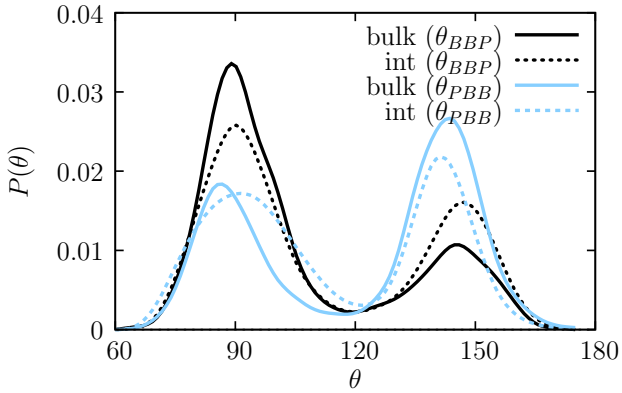


Figure 3.14: Probability distributions between the BBP and PBB superatom triplets from atomistic simulations in bulk water and at the cyclohexane/water interface.

Since the **PBBP** dihedral angle is coupled to the nonbonded interactions, firstly the initial parametrization of the nonbonded interactions between the peptide and the solvent molecules will be shown. Similar to the solvent parametrization, these are also derived based on the solvation free energy of fragments representing the respective groups obtained via thermodynamic integration. Note that the derivation of parameters based on molecule fragments is potentially problematic since free energy

is not per se additive, but these fragment-based parameters serve as a good initial guess. The free energy change upon transfer of the relevant group from vacuum into the bulk solvent medium is obtained from AA simulations. The relevant solvation free energy values measured from AA simulations, and the Lennard-Jones parameters for the matched CG interaction and the resulting CG free energy values are shown in Table 3.4. Ethyl-benzene is used in atomistic free energy calculations to parametrize the CG **P** beads. For the **B** beads, the solvation free energy of zwitterionic diglycine is computed since a smaller charge-neutral atomistic reference would have been difficult to construct. Half of its transfer free energy is used for the initial parameterization of the CG model. Peptide-peptide **B-B** and **B-P** interactions are not included, since we are only interested in a single FF molecule. The **P-P** interaction is set as a purely repulsive Lennard-Jones potential (short ranged interaction with a cut-off at the minimum), representing only the excluded volume effect in the CG simulations. The nonbonded interaction potentials which play an important role in the FF molecule's conformational behavior are shown in Fig. 3.15.

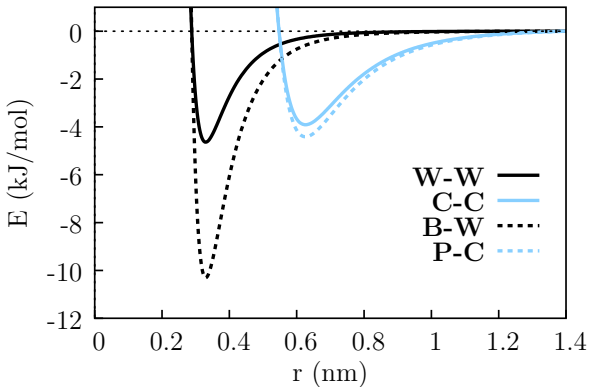


Figure 3.15: CG nonbonded interaction potentials between the FF and solvent beads (Lennard-Jones parameters are given in Table 3.4).

3.4.3 Parametrization of the Dihedral Potential

In this section, the parametrization of the CG dihedral potential **PBBP** and the performance of the FF CG model in the CG-FE and CG-ST solvent models will be

Ref./Sol.	Sys.	VdW kJ/mol	Elec. kJ/mol	Total kJ/mol	ϵ kJ/mol	σ nm
P/W	AA	9.08	-9.19	-0.11	-	-
	CG	-0.21	-	-0.21	2.81	0.288
P/C	AA	-23.36	-	-23.36	-	-
	CG	-23.21	-	-23.21	4.515	0.547
***B/W**	AA	28.81	-283.6	-254.8	-	-
	CG	-119.0	-	-119.0	10.31	0.288
B/C	AA	-18.62	-	-18.62	-	-
	CG	-9.78	-	-9.78	4.0	0.547

Table 3.4: Solvation free energy of FF fragments (that are mapped to the CG beads) in the respective solvents obtained from AA simulations. CG beads' nonbonded interactions are represented via Lennard Jones interactions, where the depth ϵ and radius σ of the interaction are fitted to reproduce the AA free energy. *For the B/W interaction CG bead is tuned to reproduce half of the hydration free energy of zwitterionic diglycine. Results with the modified **P/C** interaction are also shown for the different solvent models used: CG-FE and CG-ST.

discussed. Firstly, the behavior of the CG model in bulk water in the absence of an explicit dihedral potential will be investigated. In bulk water both CG solvent models are identical since the **W-C** interaction only concerns the systems with the cyclohexane/water interface.

The interaction potential for the **PBBP** dihedral angle is shown in Fig. 3.16. This dihedral potential is obtained by Boltzmann-inverting the **PBBP** dihedral angle distribution of the atomistic bulk water simulation.

The first problem arises from the spherical bead representation of the phenylalanine side chain, which has a planar geometry. The single-bead representation of **P** prevents penetration of CG water beads into the space between the two sidechains, and as a result, unlike the AA bulk water simulation, the CG simulation yields a cis-like conformation (Fig. 3.17 dashed brown line). This can be overcome if the diameter

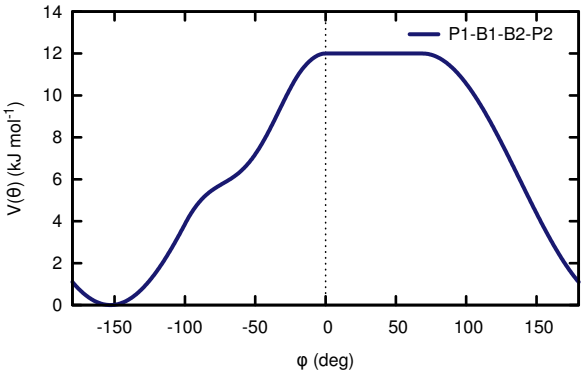


Figure 3.16: Dihedral potential for the P1-B1-B2-P2 obtained via Boltzmann inversion of the corresponding atomistic probability distribution.

of the **P-W** interaction is reduced from (initial) 0.5 nm to 0.28 nm which corresponds to the σ of the **W-W** interaction. With these "reduced size" **P** beads, the CG model yields a uniform dihedral angle probability distribution (Fig. 3.17 solid brown line). In other words, with all other bonded and nonbonded interactions defined, this CG FF model has no dihedral preference on its own in bulk water. Therefore, to capture the atomistic behavior in bulk water, one can simply Boltzmann invert the atomistic dihedral probability distribution in bulk water to obtain an effective dihedral potential. After adding this dihedral potential, the CG model is perfectly capable of representing the bulk water behavior, as shown in Fig. 3.17 (dashed blue line). Alternatively, if one obtains a dihedral potential by Boltzmann inverting the AA simulation's dihedral distribution at the interface, a cis-like conformation is observed also in bulk water as expected (solid blue line in Fig. 3.17), i.e. not the appropriate bulk behavior.

For all the atomistic and coarse-grained systems the convergence of the simulations is checked by block analysis. The probability distributions for the dihedral angle of FF molecule in atomistic and coarse-grained simulations is shown in Figs. 3.18 and 3.19, respectively. The 100 ns simulations are divided into two equal halves, and probability distribution is compared for these two blocks.

Next, we use this CG model which reproduces the bulk water behavior (with reduced size **P-W** interaction and dihedral potential obtained from BI of the bulk water

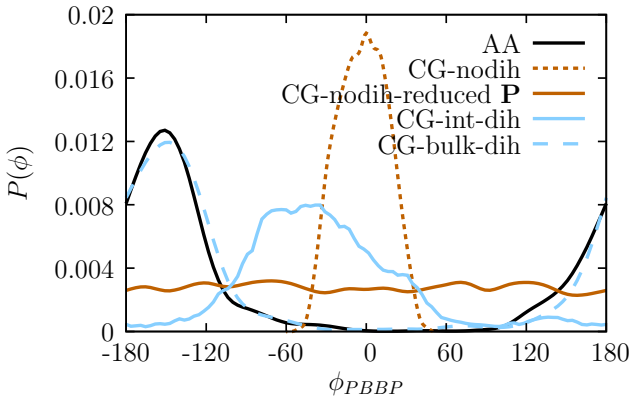


Figure 3.17: Bulk water: Comparison of **PBBP** dihedral angle probability distribution from CG simulations with the atomistic reference (solid-black line). Without an explicit dihedral potential, single-bead representation (see text) of the **P** beads leads to a cis-like conformation (dashed brown line). When the size of the **P** beads is reduced, a flat dihedral distribution is obtained (solid brown line). Thus, dihedral potential obtained via Boltzmann Inversion (BI) of the atomistic reference in bulk water (dashed blue line) reproduces the correct behavior. On the other hand, BI of the interface distribution fails in bulk water (solid-blue line).

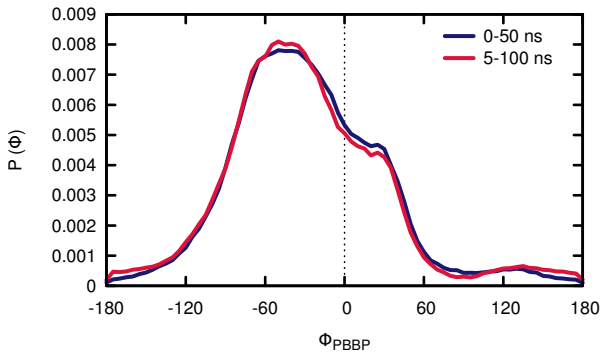


Figure 3.18: Comparison of the dihedral probability distribution of the atomistic simulations at the water/cyclohexane interface for the 0-50 ns and 50-100ns intervals.

distribution) for the biphasic system with a cyclohexane/water interface. Fig. 3.20 shows that in the CG-ST solvent model the dihedral distribution at the interface is not correct: the FF molecule remains in a trans state (Fig. 3.20 brown line). In the CG-FE solvent model (Fig. 3.20 blue line) however, the FF molecule undergoes a transition towards a cis-like state, with a dihedral distribution that is shifted compared to

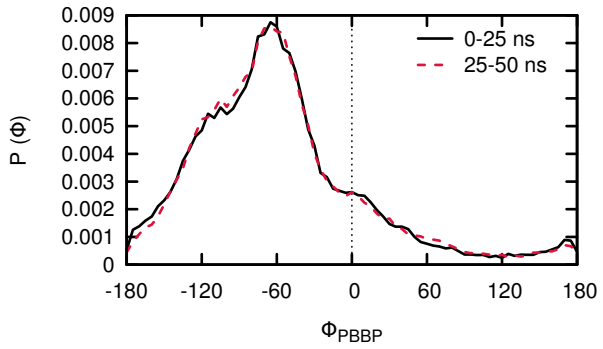


Figure 3.19: Comparison of the dihedral probability distribution of the free energy based coarse grained model (CG-FE) simulations at the water/cyclohexane interface for the first 50ns and 50-100ns intervals.

the AA result, but that nevertheless qualitatively displays the correct conformational transition. Therefore it could be argued that unlike the CG-ST model, the CG-FE model mimics a conformational transition upon changing the environment from bulk water to cyclohexane/water interface. This is in a sense quite a remarkable difference since the two solvent models only differ in the interaction between cyclohexane and water, i.e. the interface properties, there is no difference in the direct interaction between the peptide and the solvents. Nevertheless, one of the two models mimics the environment-driven conformational transition, and the other does not. Remarkably, the model that displays the transition is *not* the model that had been parametrized according to the surface tension, i.e. the interface behavior but the one where the solvation free energy of water in cyclohexane is reproduced. This supports the view that although the conformational transition is structural the underlying forces that exhibit this transition are thermodynamic.

The behavior of FF at the interface is analyzed further, to understand the transition, as well as to investigate what would have to be done to reproduce the conformational change at the interface. To this end, we first compared the relative positions in the dimension normal to the interface plane (namely the z-dimension) of the FF molecule with respect to the interface among AA and CG models. The positioning

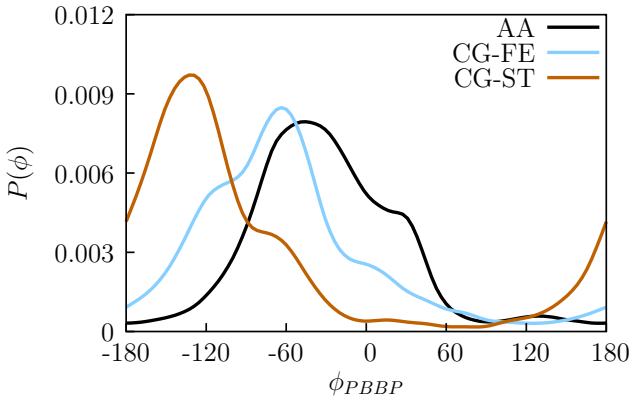


Figure 3.20: Cyclohexane/Water interface: Comparison of **PBBP** dihedral angle probability distribution from CG simulations with the atomistic reference (black line). CG model uses reduced **P-W** interaction radius and dihedral potential based on BI of atomistic bulk water dihedral distribution. The CG solvent model with surface tension tuned **W-C** interaction fails to capture the correct dihedral behavior (CG-ST, brown line), whereas the **W-C** interaction tuned to reproduce solvation free energy of a water molecule in cyclohexane solvent (CG-FE, blue line) yields a fairly close result.

at the cyclohexane/water interface is calculated as follows: the distance between the center of mass of the FF molecule and water is calculated and the half height of the water slab is subtracted from this distance. The height of the water slab is determined by the point where the density of water is reduced to its half. As far as the positioning of the FF molecule at the interface is concerned, in both CG solvent models the FF molecule is submerged somewhat towards the water side of the interface compared to the AA reference, with the CG-FE model being closer to the AA result (see Fig. 3.21 blue and brown lines).

Next, we calculated the orientation of the molecule with respect to the interface to observe the orientation of the FF molecule and test the performances of the CG models in reproducing this structural behavior. Figure 3.22 shows the orientation of the FF molecule with respect to the interface, where the CG model does not display the broken symmetry observed in the atomistic simulation. The distribution of the orientation angle γ has a maximum at $\gamma = 90$ degrees in the CG model. Here, γ denotes the angle between the interface normal and the normal of a plane fitted to

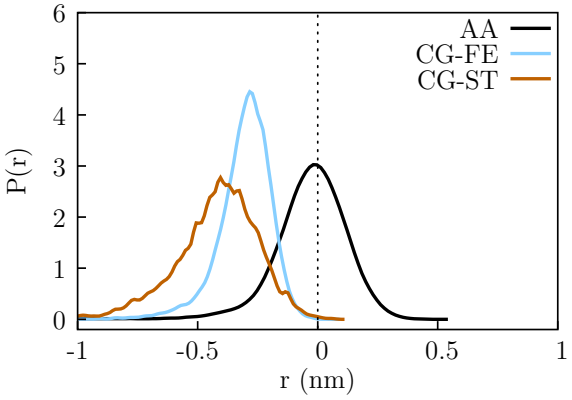


Figure 3.21: Probability distribution for the distance of FF from the interface. Unlike the AA simulation (solid black line), where the molecule is placed right at the interface, CG model yields a submerged FF molecule with both of the tested solvent models (CG-FE and CG-ST shown with blue and brown lines, respectively), with CG-ST model displaying looser adsorption to the interface.

the four beads of the CG peptide where the former corresponds to the unit vector in z direction and the latter is calculated by using the vectors from the center of the **B** beads to each **P** bead (normal of the $\mathbf{B}_c\mathbf{P}_1$ and $\mathbf{B}_c\mathbf{P}_2$ vectors where the c subscript indicates *center*). The corresponding snapshots in Fig. 3.23 illustrate the γ angle. With the lower resolution of the CG model, failure of the model to capture this broken symmetry is not surprising.

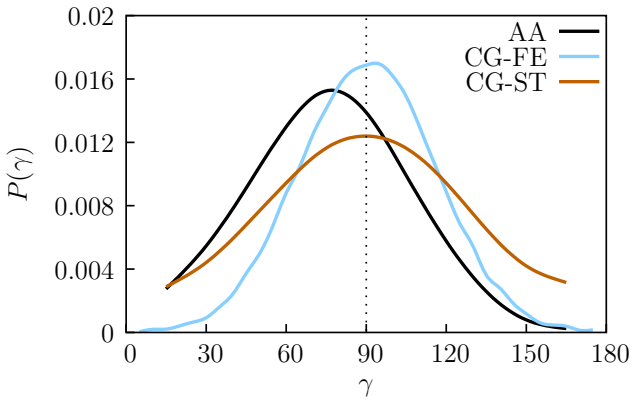


Figure 3.22: Probability distribution for the orientation angle (γ) of the molecule with respect to the interface for atomistic (solid black line) and CG model with two different solvent models (solid blue line for CG-FE and solid brown line for CG-ST).

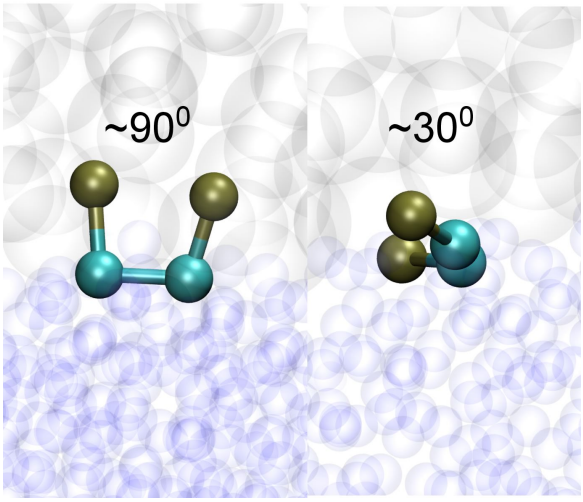


Figure 3.23: Snapshots showing the orientation of CG model of FF peptide with respect to interface. The snapshot on the left corresponds to an orientation of approximately 90° and the one on the right to an angle of30°

3.4.4 Modifying the CG models

Considering the **PBBP** dihedral angle distributions in Fig. 3.20 the CG-ST model failed to capture the conformational transition at the interface and CG-FE still shows a finite probability for the *trans*. This section will answer the questions of 1) how should the nonbonded interactions be arranged to achieve a structural change upon moving from bulk water to the cyclohexane/water interface and 2) how does the CG-FE model respond to similar changes introduced to the CG-ST model. Both CG models are modified to reproduce such a behavior and the modified model is compared to the previous findings.

Note that the CG-ST model captures the atomistic dihedral distribution in bulk water. Therefore any modifications to the nonbonded interactions that involve water will also harm the correct trans behavior of the CG FF in bulk water. Keeping this in mind, the safest approach is to modify the FF-cyclohexane interactions, after all, it is the cyclohexane/water interface properties we are trying to reproduce. The two options are changing the **B-C** (the backbone-cyclohexane) or the **P-C** (the sidechain-cyclohexane) interactions. The effect of the interactions of the backbone of the peptide on the cis-trans conformations is lower than that of the phenyl sidechains.

To this end, we chose to modify the **P-C** interaction which affects the positioning of the sidechains the most and at the same time does not disturb the bulk water behavior. Upon several trials, it is found that if the depth of interaction is increased then the dihedral distribution mimics the distribution observed in the AA model. Remember that although the CG-FE model displayed conformational transition, the **PBBP** dihedral distribution at the cyclohexane/water interface was shifted and the FF molecule spent more time at the trans conformation compared to those of the AA model. To this end, we increased the depth of the **P-C** interaction for the CG-FE model and we found that a potential with a lower value compared to that of the CG-ST model was necessary for reproducing a dihedral distribution closer to that of the AA. Such a difference is expected as the unmodified CG-FE model already displayed a near cis behavior in cyclohexane/water interface whereas the unmodified CG-ST

model exhibited trans behavior.

The modified **P-C** potentials along with the unmodified ones can be seen in Figure 3.24 and the necessary parameters needed to generate the potentials are presented in Table 3.5. The **PBBP** dihedral distributions of the modified CG models can be seen in Fig. 3.25 where the first peak (around $-50°$) of the modified CG-ST model as well as the height of the distributions are imitated exactly and there is a slight shift both in the position and the height for the second peak (around 45°). The position of the first peak of the modified CG-FE model is also moved to the correct position observed in AA but the probability of observing this peak is a little higher than the one in AA. The position of the second peak is the same as the position of the peak for the CG-ST, which suggests that this slight inaccuracy cannot be fixed with tuning the **P-C** interaction alone. Overall it could be argued that the modified CG models perform better in terms of dihedral distribution compared to the unmodified ones.

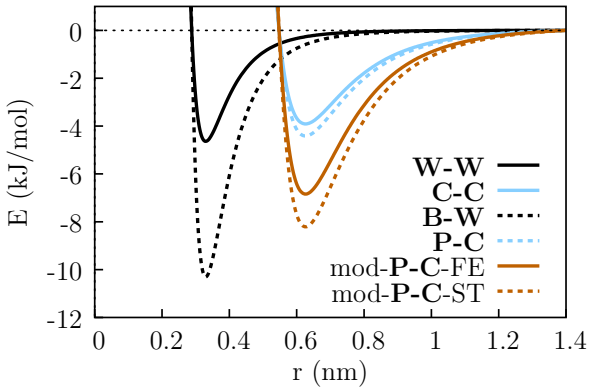


Figure 3.24: CG nonbonded interaction potentials between the FF and solvent beads (Lennard-Jones parameters are given in Table 3.4). For the **P-C** potential the modified CG interactions are also shown for both the CG-FE (solid brown line) and the CG-ST models (dashed brown line), which correct the dihedral distribution at the interface.

Table 3.5 presents the solvation free energy of ethylbenzene in cyclohexane for the AA model and the solvation free energy of the corresponding superatoms, namely **P** in a solvent of **C** for the CG models. Remember that the **P-C** interaction potential of the unmodified CG models were generated to match the solvation free energy of ethyl-

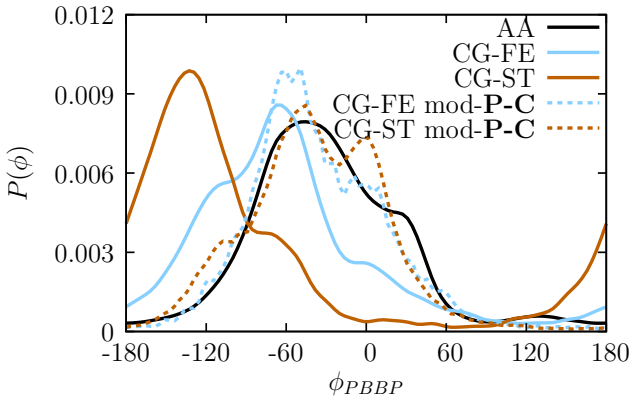


Figure 3.25: Cyclohexane/Water interface: Comparison of **PBBP** dihedral angle probability distribution from CG simulations with the atomistic reference (solid black line). CG model uses reduced **P-W** interaction radius and dihedral potential based on BI of atomistic bulk water dihedral distribution. The CG solvent model with surface tension tuned **W-C** interaction fails to capture the correct dihedral behavior (CG-ST, solid brown line), whereas the **W-C** interaction tuned to reproduce solvation free energy of a water molecule in cyclohexane solvent (CG-FE, solid blue line) yields a fairly close result. By increasing the attraction for the **P-C** interaction both CG models can be enhanced to yield a better match with the AA result (dashed blue and dashed brown line for CG-FE and CG-ST solvent models, respectively).

benzene in cyclohexane in AA. The solvation free energies of the CG models with the modified **P-C** interactions steer away from the solvation free energy of ethylbenzene in cyclohexane for the AA. As the depth of the **P-C** interaction for the CG-ST was increased more than that of the CG-FE, we observe a larger difference in solvation free energy for the CG-ST. Therefore, if one wants to reproduce the exact distribution of the AA model, one has to sacrifice from the free energies.

Figure 3.26 compares the positioning with respect to the interface of modified CG models with those of the unmodified ones and the AA. The unmodified CG models yielded a submerged FF in bulk water whereas in the AA the peptide was positioned exactly at the interface. Increasing the depth of the **P-C** potential made the interaction of the peptide with the cyclohexane more favorable. Therefore the position of FF is shifted to the interface for the modified CG models compared to the unmodified ones. Here, the modified CG-ST presents an almost exact match with

Ref./Sol.	Sys.	VdW kJ/mol	Total kJ/mol	ϵ kJ/mol	σ nm
P/C	AA	-23.36	-23.36	-	-
	CG	-23.21	-23.21	4.515	0.547
mod-**P-C** CG-FE	CG	-54.30	-54.30	7.00	0.547
mod-**P-C** CG-ST	CG	-72.59	-72.59	8.40	0.547

Table 3.5: Solvation free energy of ethylbenzene in cyclohexane obtained from AA simulations and the corresponding solvation free energies of **P** in solvent of **C** shown along with those of the modified CG-FE and CG-ST models. The nonbonded interactions of the CG models are represented via Lennard Jones interactions, where the depth ϵ and radius σ of the interaction are fitted to reproduce the free energy of the AA model. Note that as the atoms of the cyclohexane molecule does not have any partial charges the electrostatic contribution equals to 0 and is not shown.

the AA. However these improvements come at another cost. Looking at the results for the orientation with respect to interface, the modified CG models show an almost identical distribution which is narrower compared to the unmodified CG and the AA models. As the **P** beads have a more favorable interaction with the **C** beads the FF peptides spend less time with the orientation where the **P** beads are positioned closer to **W** beads, which mimic water molecules.

Since we assumed that the conformational transition of the molecule is driven by the segregation forces that are only present at the interface, we calculated the potential of mean force for pulling the FF molecule from the middle of the water layer to the middle of the cyclohexane layer. The PMF curves corresponding to the transfer to the interface for the AA and CG systems are shown in Fig. 3.28. The atomistic system shows a 20.6 kJ/mol change in free energy upon adsorption of the molecule to the interface. The minimum of the potential is positioned right at the interface, in agreement with the free simulation results shown in Fig. 3.26. The CG-FE model (Fig. 3.28 solid blue line) yields a similar adsorption free energy (23.6 kJ/mol). However, unlike the atomistic system, the potential minimum is located

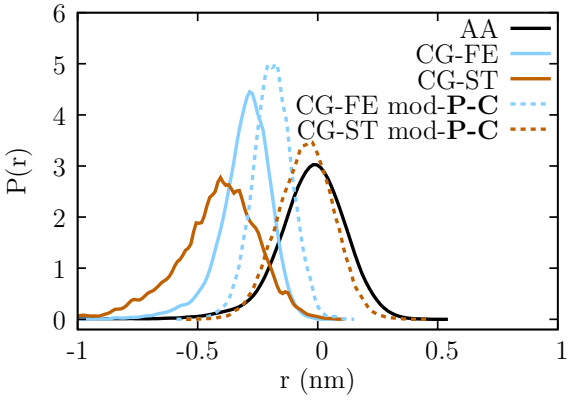


Figure 3.26: Probability distribution for the distance of FF from the interface. Modified **P-C** interaction enhances both of the CG models (dashed lines), where modified CG-ST case yields almost a perfect match with atomistic results.

slightly below the interface (also in agreement with the results shown in Fig. 3.26). The CG-ST solvent model shows a weaker adsorption to the interface (solid brown line), i.e. in this aspect it shows less agreement with the atomistic reference (solid brown line).

Although the behavior of the FF inside the bulk cyclohexane phase is not within the scope of this study we analyzed the performance of the CG models in that regard. Fig. 3.29 shows the PMF curves to transfer the FF from bulk water phase to the cyclohexane phase. The plots are shifted such that $V(r) = 0$ when FF is at the center of the bulk cyclohexane. All of the CG models fail to capture the initial point for the drop in free energy for transferring FF from bulk cyclohexane to the interface. This drop does not begin until FF is 0.9 nm away from the interface whereas the same property for AA is 1.5 nm. This inconsistency might be fixed by increasing the range of the potentials for the peptide-cyclohexane interactions. The CG-ST solvent model predicts the slope of the PMF for the same transfer better than the CG-FE which has a steeper slope compared to the AA. The atomistic PMF for transferring the dipeptide from bulk water to bulk cyclohexane is best reproduced by the CG-FE model whereas the CG-ST model underestimates this free energy difference. Increasing the depth of the **P-C** interaction (dashed lines) decreases the difference in favorability of the

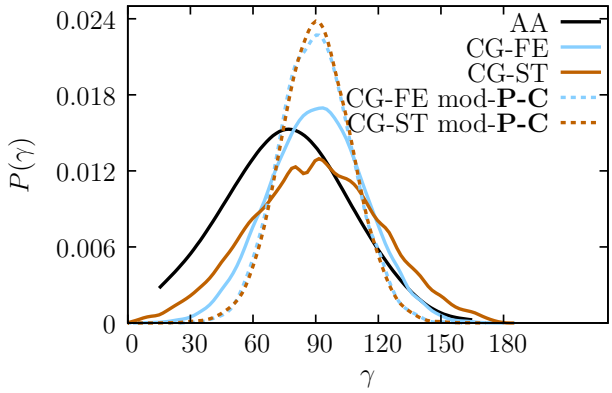


Figure 3.27: Probability distribution for the orientation angle (γ) of the molecule with respect to the interface for atomistic (solid black line) and CG model with two different solvent models (solid blue line for CG-FE and solid brown line for CG-ST). Upon modification of the **P-C** interaction, both CG-FE and CG-ST case yields a narrower distribution compared to the atomistic results (dashed lines).

solvents by the FF and therefore do not yield a good estimate for free energy values compared to those of the AA. Therefore free energy estimates are worse for the CG models with the more favorable **P-C** interaction compared to the unmodified CG models.

An interesting difference between the CG-FE and CG-ST results is that the CG-FE PMF curve rises more steeply compared to the AA PMF when the molecule is dragged into the cyclohexane phase. The analysis of the trajectory shows that this is a direct consequence of the hydration shell of the FF molecule. For the CG-FE solvent model, the reduced **W-C** interaction potential results in a stripping of all the water molecules in the hydration shell of FF as the molecule is pulled out of the water layer. In contrast, the CG-ST successfully mimics the AA results in this regard and maintains part of the hydration shell of the FF molecule as the molecule enters the cyclohexane phase (see Fig. 3.30).

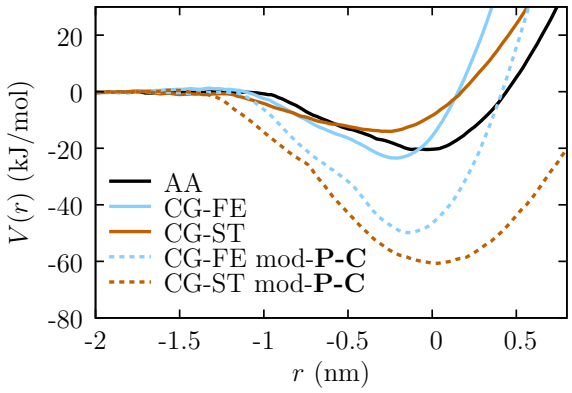


Figure 3.28: Potential of mean force, $V(r)$, for the transfer of FF from bulk water to the cyclohexane/water interface. CG-FE solvent model yields the closest match to the atomistic result in terms of depth (solid blue line), whereas CG-ST solvent model displays a weaker attraction towards the interface. In both cases the minimum is shifted towards the water side of the interface. Modification of the **P-C** interaction pulls down the PMF curve to much deeper values, largely overestimating the adsorption energy (dashed lines).

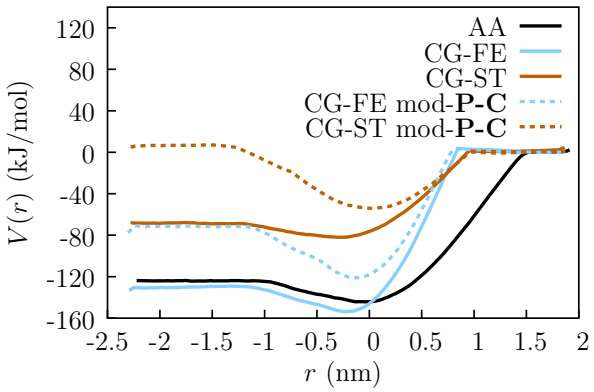


Figure 3.29: Potential of mean force, $V(r)$, for the transfer of FF from bulk water to the bulk cyclohexane. The plot is the full form of Fig. 3.28 showing the transfer to the cyclohexane phase.

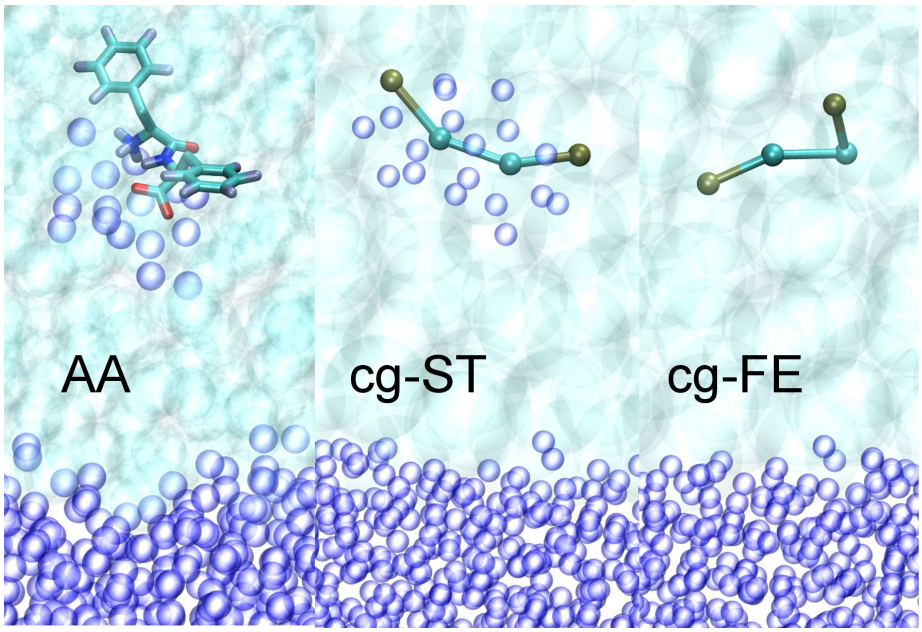


Figure 3.30: Snapshots showing the hydration shell when FF is pulled into the cyclohexane phase from the bulk water. The CG-ST solvent model (middle) where the surface tension matches that of the AA, correctly captures the hydration shell of the AA (left) whereas the cg-FE (right) which yields a higher surface tension value compared to AA, fails.

The results shown so far suggest that tuning the dihedral potential alone is not necessarily sufficient to capture the desired conformational behavior for the CG system in both bulk water and at the cyclohexane/water interface. Even though the affinity towards the interface (i.e. the PMF of the entire molecule) is reproduced fairly nicely with the fragment-based free-energy-tuned peptide-solvent interactions, the cis-trans conformational transition is not necessarily correctly represented by the CG model. Disconcertingly, in the present system the outcome appears to be depending on the solvent-solvent interactions, i.e. on parameters that do not involve the peptide at all and that may in some more complex scenarios of many-component systems not be free for parameter changes.

Therefore the options have been investigated (involving the peptide-solvent interactions) to reproduce the correct conformational behavior in *both* solvent scenarios. One possible solution involves increasing the strength of the interaction between the cyclohexane and phenylalanine side chain beads (see Fig. 3.15, dashed brown line, labeled mod-**P-C**), so that the segregation forces are further enhanced. Since the **P-C** interaction is used only in the interface simulation, the bulk water behavior of the molecule remains unaffected. For each of the two CG solvent models the depth of the modified **P-C** (listed in Table 3.4) is chosen such that at the interface the AA cis-dihedral distribution is more or less exactly reproduced – overruling the dihedral potential which favors a trans-like conformation. Interestingly, for both CG solvent models the interaction has to be substantially deepened to achieve the effect, even though in the CG-FE case the unmodified model had already shown a qualitatively quite acceptable distribution. As shown in Fig. 3.25 with dashed lines, after this modification the CG model correctly captures the interface dihedral distribution with both CG solvent models. The position with respect to the interface is also improved compared to the original CG model (Fig. 3.26 dashed brown line), surprisingly CG-ST model now outperforming the CG-FE model. The orientation of the molecule still yields a maximum at $\gamma = 90$ degrees (Fig. 3.27) and is much narrower compared to the AA result. However, this improvement takes place at the cost of the adsorption

free energy: the modified CG model (by construction – due to the stronger **P-C** affinity) exhibits a much larger driving force towards the interface (Fig. 3.28).

3.5 Concluding Remarks

Results presented here show that with respect to sampling of conformations, transferability of CG models cannot be taken for granted. The model system, diphenylalanine, displays distinctly different conformations in bulk water compared to the cyclohexane/water interface. The pseudo dihedral angle labeled as **PBBP** favors a trans-like conformation in bulk water, whereas it switches to cis-like conformation at the cyclohexane/water interface. The same conformational transition is also seen upon aggregation of FF molecules into nano-tubes. Hence, if one desires to capture the surface adsorption or aggregation of such a molecule via a CG model, the model should display an identical conformational transition.

Such a conformational cis/trans transition is not automatically captured by the CG model, even if the parametrization of the nonbonded interactions is done based on the partitioning forces that are the origin of this conformational transition. Furthermore, the ability to reproduce the conformational transition may depend on parameters of the solvent models that do not involve the peptide at all and that merely affect the behavior of the interface. Between the two different solvent models tested, the CG-ST solvent model, where the mixed solvent-solvent interactions had been parametrized according to surface tension, did not induce the conformational transition at the interface at all, even though the model showed quite good results for the surface affinity. The CG-FE model on the other hand, where the water-cyclohexane interaction is parameterized to mimic the solvation free energy of water in cyclohexane, yields a fairly good result in capturing the dihedral distribution at the interface and representing the adsorption free energy from water to the cyclohexane/water interface.

Finally, the study showed that the interface dihedral distribution for both of these CG solvent models can be corrected by reparametrizing the **P-C** interaction. This

modification not only affects the conformational behavior, it also significantly improves the position of the FF molecule at the interface. However, one sacrifices the correct adsorption strength of the FF to the interface, as seen in the PMF curves in Fig. 3.28. The crucial role that the correct representation of the water interface has on the conformational transitions of the peptide suggests that, improved representation of structural and thermodynamic properties of water could potentially eliminate the need for reparametrization of peptide-solvent interactions.

The entire study shows that one has to reconsider the separation of bonded and nonbonded interactions for the parametrization of CG models. The subtle interplay between the conformational preference of the molecule, the solvation properties of its fragments, and even interactions in the environment that do not directly affect the solute molecule at all have to be taken into account. This study showed that one way to correctly capture the conformational preference is to adjust the nonbonded interactions, rather than playing with the dihedral potential. This modification corrects the conformational preference of the molecule at the interface and reproduces the conformational transition, but one sacrifices the exact reproduction of the adsorption free energy to the interface.

The parametrization presented here was intended mainly to understand the interplay of the relevant driving forces for such an environmentally-driven conformational transition and how they can be correctly accounted for in the CG model. The four-bead CG model used may be problematic for reproducing the full aggregation behavior including the formation of microscopically accurate structures, since the backbone is likely to be too coarsely represented and sidechains might fail to capture the aromatic nature of the phenylalanine sidechains. Representation of sidechain conformations could potentially be improved by using an angle dependent potential [141]. Alternatively, a similar parametrization can be introduced for the MARTINI representation of FF molecule, which has been recently shown to exhibit interesting aggregation behavior for FF [142]. In the long run it would also be interesting to use a solvent free model to study large scale aggregation (without an interface). In that case the

crucial interactions that will have to be balanced are the hydrophilic and hydrophobic interactions between the various peptide beads.

Chapter 4

ENVIRONMENT INDUCED CONFORMATIONAL TRANSITION AND AGGREGATION OF LKα14 PEPTIDES

4.1 Introduction

The previous section showed the conformational transition of FF from bulk water to the hydrophilic/hydrophobic interface. One of the main reasons for studying FF was that it is a small molecule that is easy to simulate at atomistic scale. Furthermore, FF allowed a coarse-grain model with only four beads with a total of two different types which reduced the total number of interactions. This chapter expands the previous chapter by investigating the conformational transition between bulk water and the hydrophilic/hydrophobic interface using a more complex molecule, i.e. an LK peptide. LK peptides are composed of hydrophobic leucine (L) and hydrophilic lysine (K) residues and they display amphiphilic properties at hydrophilic/hydrophobic interfaces and are widely studied as a result.

The adsorption and the conformational preference of different LK peptides were examined at hydrophilic/hydrophobic interfaces by many studies. $(KL)_mK$ series of peptides have been shown to form anti-parallel β-sheets when bound to lipids. [143] In another study featuring several LK peptides that adopted α-helical conformations the hydrophilic/hydrophobic interface was found to be the main driving force that induces conformational transition. [144] Kang et. al. reported the antimicrobial effects of twelve helical LK peptides where a tryptophan was used as a residue that separates the hydrophilic and hydrophobic side in the helical wheel projection. [12] The experiments of Beven et. al. with LK peptides of various sizes argued the *carpet-model*

for the disruption of the bacterial cell where initially the peptide partitions at the membrane interface, then changes its conformation, and disrupts the cell. [145] Another study found that the antimicrobial activity helical LK peptides was augmented with increasing charge interactions whereas increasing the hydrophobic interactions resulted in enhanced hemolytic activities. [146] Experiments with LK_2 variants revealed a threshold length of 12 residues over which aggregation was detected. [147]

The LKα14 peptide used in this study is a short peptide of 14 amino acid residues with a hydrophobic periodicity of 3.5 to allow an α-helical structure. Degrado and Lear synthesized this peptide and analyzed its concentration dependent secondary structure and aggregation along with two more LK peptides, namely LKβ14 and LKα7. [148] In their work Degrado and Lear determined that one can engineer such peptides to obtain the desired secondary structure making use of the hydrophobic periodicity where a periodicity of 3.5 yielded α-helix and a periodicity of 2 produced beta-sheet. It has been shown that periodicity is more important than the propensities of amino acid residues for the secondary structure. [149] Degrado and Lear found that the peptide should be long enough to assume the α-helix where the 14 residue long LKα14 is α-helix but the 7 residue long is not. With another set of LK peptides twelve residue was found to be the threshold for the α-helix. [150] The requirement for β-sheet was found to be lower where a 7 residue long LKβ7 peptide yielded a β-sheet structure. In the same study they also found that the LKα14 aggregates to form a *four helix bundle* at high peptide and chloride concentrations. The α-helix allows the peptide to partition its hydrophobic and hydrophilic residues as seen in Fig. 4.1 where the green residues are hydrophobic leucine and the blue ones are charged lysine residues. This conformation favors a minimized exposure of the hydrophobic residues with the polar water molecules. Moreover, various spectroscopic techniques yielded an α-helical LKα14 on the polystyrene and fluorocarbon surfaces where the leucine sidechains adsorbed onto and the lysine sidechains positioned opposite to the surface. [151–157]

Few computational studies featured LKα14. Deighan and Pfaendtner investigated the dynamics of adsorption of LKα14 and LKβ15 on the self-assembled monolayer

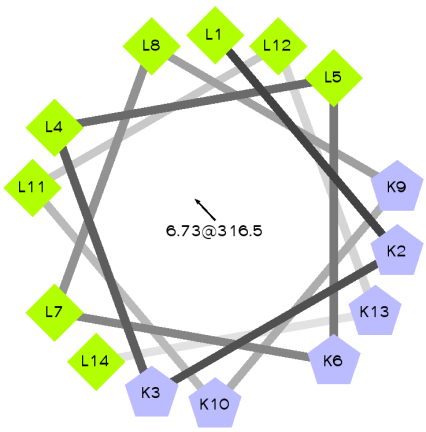


Figure 4.1: Helical wheel representation of LKα14 showing the separation of hydrophobic leucine residues (green) and hydrophilic lysine residues (blue).

surfaces using metadynamics and reported the adsorption free energy values for both peptides [158]. Another study compared the behavior of LK peptides in solution and at the functionalized surfaces for force-fields of CHARMM22, AMBER94 and OPLS-AA using replica exchange MD method where CHARMM22 yielded conformations that agree with the experimental results the most but the short simulation times of 12 ns in the study introduces the question of adequate sampling. [159] None of these studies addressed the question of the dynamics of folding and aggregation of LKα14.

This study explores the environment induced conformational transition of LKα14 as well as its aggregation properties both in bulk water and at the hydrophilic/hydrophobic interface, i.e. vacuum/water which is the simplest of such interfaces. The next section will present the methods and the necessary parameters used in this study. Then the results of the all-atomistic model of LKα14 will be discussed, the secondary structure will be compared in bulk water and at the vacuum/water interface. Then the LKα14 aggregates in bulk water will be investigated where different aggregate sizes will be compared in terms of stability. The aggregates will also be tested at the interface with the dimer having the focus.

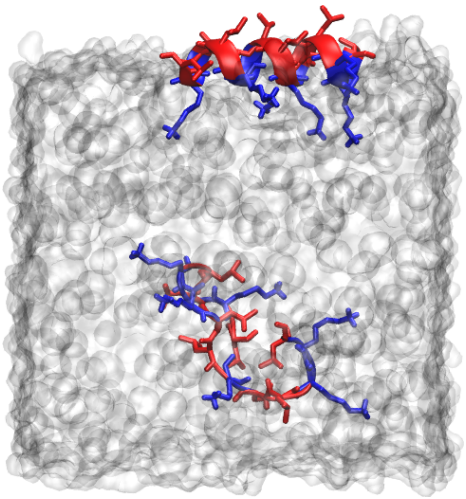


Figure 4.2: Atomistic view of KL peptide in bulk water (bottom) and at the vacuum/water interface (top). The translucent grey area represents the water slab. At the vacuum/water interface the lysine (blue) residues reside inside the water whereas the leucine (red) sidechains extend into the vacuum. The sidechains are shown with licorice and the backbones are shown with new cartoon representations.

Before the concluding remarks, preliminary results for coarse-graining the LKα14 will be provided.

4.2 Methods

The sequence of the peptide studied is Ace-$(LKKLLKL)_2$-Nme. To mimic the neutral pH environment, lysine sidechains are protonated and the system is neutralized using chloride ions. The amount of chloride ions inserted to neutralize the system increases the chloride concentration higher than the lowest experimental threshold value of concentration required to form helical structure [148]. Degrado et. al. reports that both the aggregate formation and the alpha helical secondary structure are unstable at low peptide and low chloride concentrations. Therefore if one wants to simulate the same level of low chloride concentration as in the study by Degrado et. al., one needs a much larger box size.

Simulations were conducted using the 4.5.6 version of GROMACS software package [160] with the leap-frog integrator employing the GROMOS 54a7 force-field [161]

and explicit SPC/E water [125]. All bonds were constrained with LINCS algorithm [56] to enable a time-step of 2 fs. Velocity-rescaling algorithm [59] was used to fix temperature at 298 K with a coupling time of 1 ps. For bulk water simulations isobaric-isothermal ensemble was utilized fixing the pressure at 1 atm with isotropic Berendsen pressure coupling [60] with a pressure relaxation time of 1 ps for both the solvent and the peptide. The isothermal compressibility of 4.5×10^{-5} bar^{-1} which corresponds to that of pure water was used for the system. Long range electrostatic interactions were calculated by particle mesh Ewald (PME) [127] with a Coulomb cutoff of 1 nm. Van der Waals interactions were truncated at 1.4 nm. Twin range cut-off method was applied with a neighbor-searching distance of 1.0 nm to 1.4 nm where neighbor list was updated every 10 steps.

Vacuum/water interface simulations were done in canonical ensemble. The single LKα14 simulations were run for 1 μs. The first 100 ps of the simulations were treated as equilibrium. The system was energy minimized with steepest-descent algorithm for 2000 steps, then three equilibration simulations were performed for 100 ps where the heavy atoms, the backbone atoms and the C_α atoms were constrained respectively.

16 replicas were set up for replica exchange simulations [80], with temperatures varying between 298 K and 375 K for the bulk water case, and 298 K and 366 K for the vacuum/water interface. The temperatures for the replicas were determined by the temperature predictor for parallel tempering simulations [162]. For the bulk water case, each replica was run for 1 μs with NPT ensemble where the reference pressure is set at 1 atm and for the vacuum/water interface each replica was run for 300 ns with NVT ensemble. An exchange between neighboring replicas was attempted every 1000 steps. The average exchange probability was determined to be 0.19 for the bulk water environment.

To enhance conformational sampling for both bulk water and vacuum/water interface environments, half of the peptides in replicas at low temperatures were initiated from alpha-helix, and the remaining replicas were initiated from random coil. The secondary structure of the peptide was determined from DSSP algorithm [163].

Potential of mean force (PMF) [164] calculations of umbrella sampling simulations [165] were done with weighted histogram analysis method [74] (WHAM) using g_wham [75] toolkit of GROMACS package. One of the LK peptides is pulled away from the remaining group with a pull rate of 0.01 nm/ps using a harmonic constant of 1000 kJ/mol^{-1} nm^{-2}. The snapshots taken at regular intervals are energy minimized and then equilibrated under NVT ensemble for 50000 steps. The resultant structures are then used as initial configurations of umbrella sampling windows. The distance between the centers of mass of the backbone atoms of the peptides is fixed in three dimensions for the bulk water simulations and in two dimensions for the vacuum/water interface simulations. For the cases of trimer, tetramer and pentamer, one of the peptides is pulled away relative to one of the remaining closest peptides.

All visualizations were produced by VMD [166] and secondary structures were detected by STRIDE algorithm [167].

4.3 Environment Driven Conformational Transition of LKα14 Peptide

In this section, firstly the effects of bulk water on the secondary structure of a single LKα14 is analyzed where the free MD simulations showed insufficient sampling for the molecular orientations. To scan the conformational space more effectively, advanced sampling technique of replica exchange is applied and the range of structures obtained from these simulations revealed an intrinsically disordered system. The next section discusses the induced stability by the vacuum/water interface which is the simplest of hydrophobic/hydrophilic interfaces. Similar to the bulk water system, replica exchange method is applied to confirm the equilibrium structure of α-helix.

4.3.1 Bulk Water Simulations

The full α-helix structure of LKα14 in bulk water exposes the hydrophobic leucine sidechains to polar water molecules which is expected to decrease the stability of α-helix. To observe the effects of water on the helical conformation, a single LKα14 peptide was simulated in bulk water with an initially α-helical secondary structure for 1 μs. The full-helical structure was not conserved and throughout the simulation the peptide adopted various conformations as shown in Fig. 4.3. The solvent accessible surface area (SASA) of the leucine sidechains (middle plot in Fig. 4.3) quantifies the premise that the full-helical conformation exposes hydrophobic regions to water molecules and therefore the LKα14 peptide seeks a compact conformation such as the snapshots in 575 ns and 811 ns which yield lower SASA values for leucine sidechains.

MD simulations containing structural reorientations such as folding and unfolding are strictly dependent on initial velocities as well as initial conformations. One simulation corresponds to a single point that may not be representative of the equilibrium structure of the system. To this end, we restarted the previous simulation with different initial velocities. Figure 4.4 shows the time evolution of the secondary structure of a single peptide in bulk water for four simulations with different initial velocities each started from full-α-helix structure. All of the simulations exhibit unfolding of α-helix to a degree, i.e. partial (top) or full (DSSP plots 2-4 in Fig. 4.4. The second

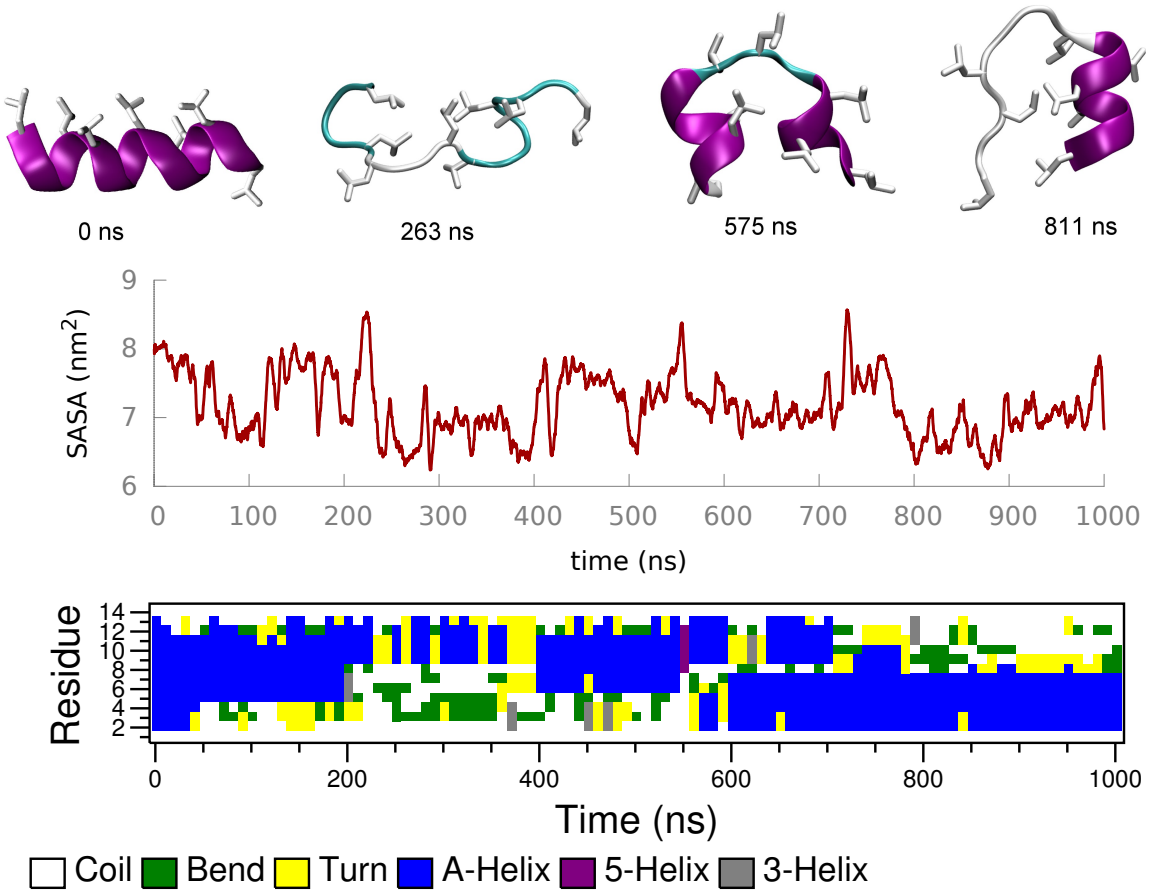


Figure 4.3: A single LKα14 in bulk water: The snapshots are hand-picked to represent different conformations observed throughout the simulation. The solvent accessible surface area of the hydrophobic leucine sidechains are plotted vs. time (middle). Time evolution of the secondary structure of a single LKα14 in bulk water color coded for each residue (bottom) along with the corresponding snapshots show that full-helical conformation was not stable and the peptide adopted various conformations including random-coil, and half-helix-half-random-coil.

simulation (second from top in Fig. 4.4) displayed a β-sheet formation unlike the other simulations. The third simulation showed full unfolding to a random coil secondary structure with no refolding to helix whereas in the third simulation refolding to first half-helix-half-random-coil and then to α-helix was observed.

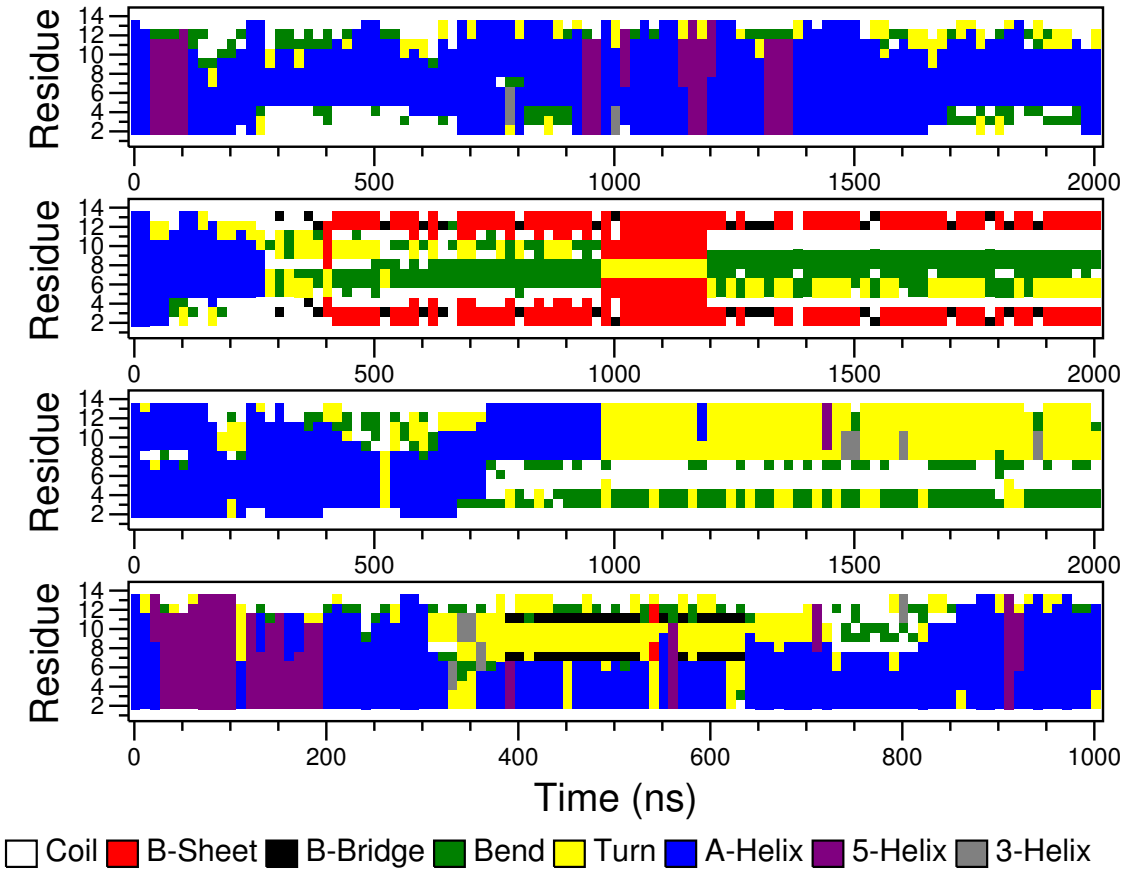


Figure 4.4: Time evolutions of the secondary structure of the LKα14 peptide in bulk water for the remaining 2 simulations with different initial seeds. For each simulation the initial secondary structure of α-helix was not stable, however for the figure in the middle a refolding into α-helix was observed. In general, the peptide assumed a variety of secondary structures including half-α-helix half-random coil, beta sheet and random coil.

Taking into account all of the five single LKα14 in bulk water simulations, no unique stable secondary structure was observed, i.e. all conformations displayed tran-

sitions between each other. Even the β-sheet observed for over 1 μs (second plot in Fig. 4.4) unfolded (1300-1500 ns). Although the dominant secondary structure was observed to be α-helix among all simulations, this is the result of the initial condition bias. Because when a single random coil LKα14 was simulated inside bulk water, no switch to a helix was observed (Figure 4.5). These results suggest insufficient sampling due to the ruggedness of the energy landscape of a single LKα14 in bulk water.

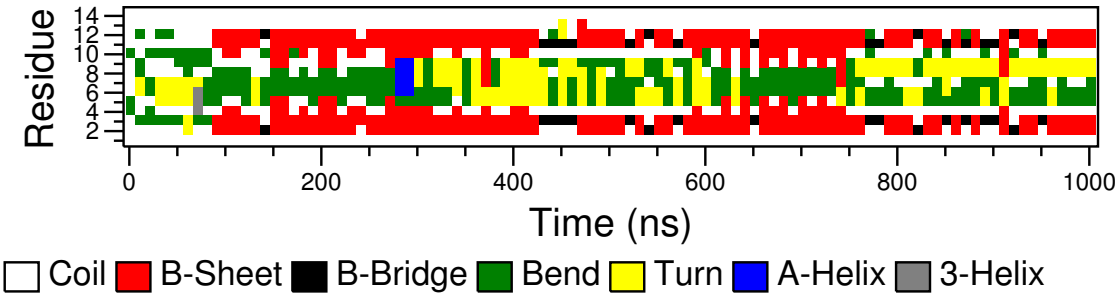


Figure 4.5: A single LKα14 peptide in bulk water initiated from a random structure. The peptide did not fold into α-helix, but instead formed an unstable beta-sheet.

To determine the range of conformations and the conformational preference for a single LKα14 peptide in bulk water the conformational space has to be properly sampled. However as shown the atomistic MD simulations fail to effectively sample the whole conformational space, at least in a reasonable amount of time. To overcome this sampling issue, one needs to use phase space enhancing methods.

Replica exchange molecular dynamics (REMD) method was applied to overcome any possible local minima in the potential energy surface and enhance conformational sampling. Several copies of the system with different temperatures were simulated in parallel for a single LKα14 peptide in bulk water. The temperature for the replicas were chosen to be between 298-520 K [162] for with 40 replicas with an average exchange probability of 0.22. The first eight replicas were initiated from an α-helical peptide whereas the remaining replicas started with a random coil peptide.

The time evolution of the secondary structure of the 40 replica conformations

which travel through different temperatures can be seen in Figure 4.6. The peptides adopt a variety of secondary structures throughout 1μs and do not show an unique stable equilibrium conformation. The peptides alternate between alpha-helix, random coil, beta-sheet and half-helix-half-coil. Lack of an unique stable secondary structure indicates a low transition energy between these conformations.

For each replica the mean number of residues with α-helix, beta-sheet, and random-coil secondary structures versus temperature was calculated (Figure 4.9). With increasing temperature the number of residues adopting random-coil increases that of helix and beta-sheet decreases as expected which is expected and one of the reasons why such high temperatures were chosen. At room temperature the helix turns out to be more likely than beta-sheet with a small difference where the probability of observing α-helix peaks at 335 K.

The REMD results for a single LKα14 in bulk water suggest that LKα14 lacks a unique conformation in bulk water, hence acts as an intrinsically disordered peptide. Note that conformations of the peptide are rather different from a random coil. In bulk water the conformational space of the peptide exhibits several distinct structures, none of them being fully stable and therefore.

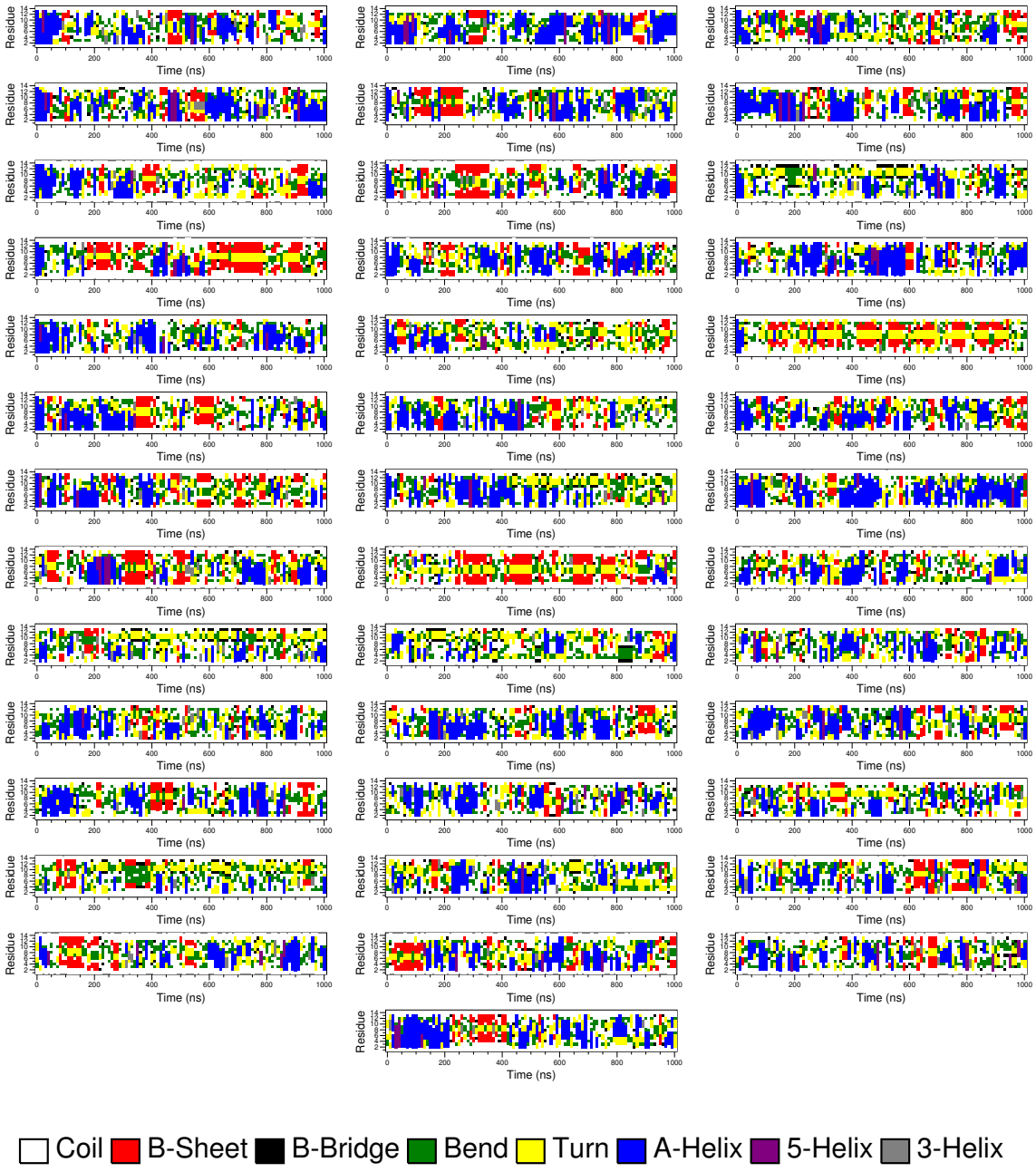


Figure 4.6: Replica exchange MD in bulk water: Time evolution of the secondary structure of each replica conformation for a single LKα14 peptide in bulk water. None of the simulations converge to a single secondary structure which indicates that for a single LKα14 in bulk water there is no unique equilibrium secondary structure. The plots are ordered in increasing temperature from left to right and top to bottom.

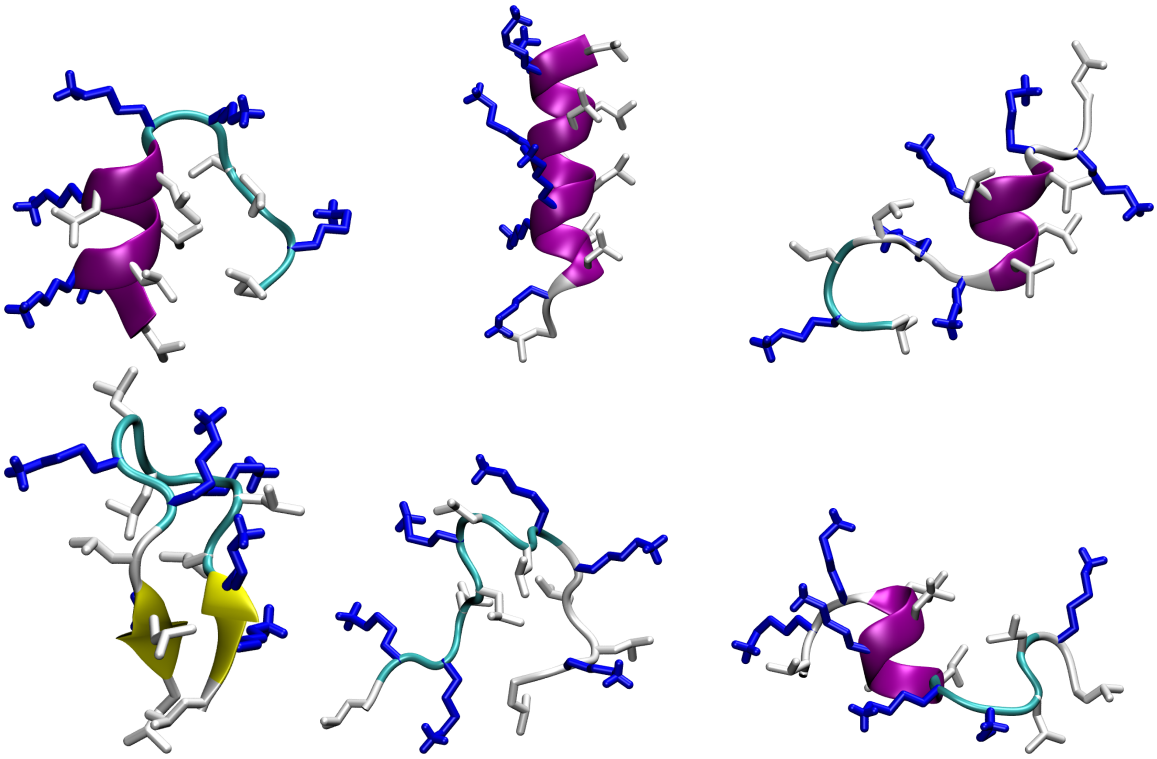


Figure 4.7: Most populated 6 conformations (from left to right and top to bottom) for the REMD simulation at 298 K for a single LKα14 in bulk water. The conformations are calculated using Jarvis-Patrick clustering algorithm [168] with a number of nearest neighbor cutoff of 60. The backbone is drawn with NewCartoon representation and colored according to secondary structure. The lysine (blue) and leucine (white) residues are shown as Licorice.

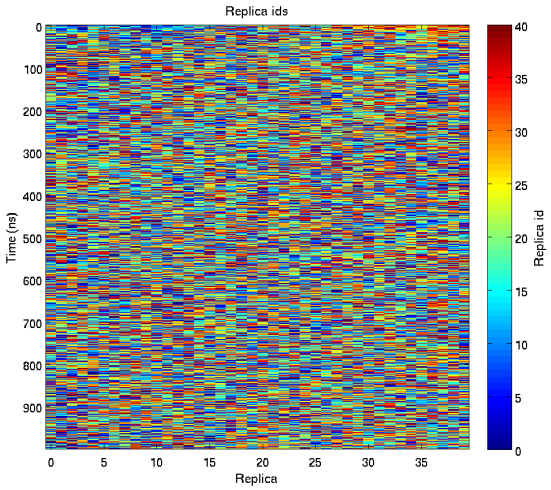


Figure 4.8: Matrix showing the exchanges between replicas for the REMD simulation in bulk water. The x scale represents the replica ids and the color bar represents the initial replica ids. The amount of mixing of the REMD simulations is correlated by the amount of mixing in this figure.

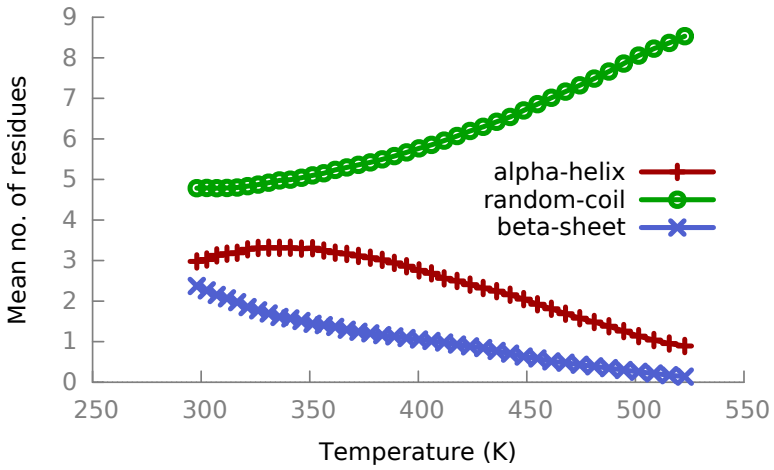


Figure 4.9: Mean number of residues with the shown secondary structure vs temperture.

4.3.2 Vacuum/Water Interface Simulations

Due to its amphiphilic property LKα14 acts as a surfactant changing the behavior of the interface. However, the interaction is mutual and the interface induces structural changes to the surfactants. In order to understand the effects of a hydrophobic/hydrophilic interface on the amphiphilic LKα14, a single LKα14 peptide was simulated at the vacuum/water interface. When a full α-helical LKα14 peptide was initially submerged in water, the peptide moved towards the interface in 20 ns and stayed there for the duration of the simulation with a stable alpha-helical secondary structure for 1 μs except for small unfolding of terminal groups.(Fig. 4.10)

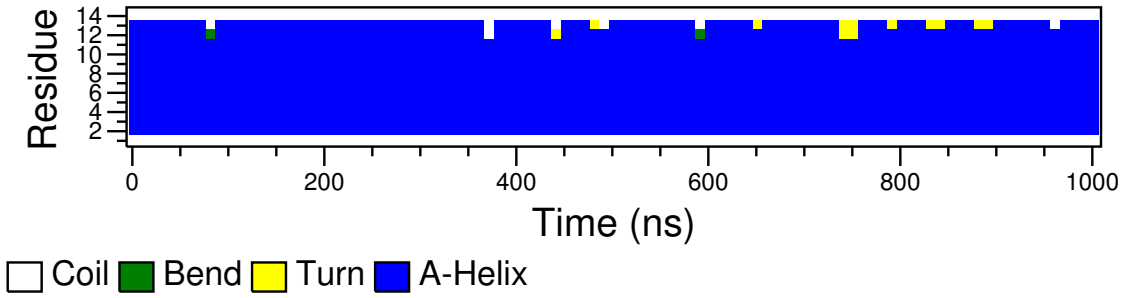


Figure 4.10: Time evolution of the secondary structure of the LKα14 peptide at the vacuum/water interface color coded per residue. The alpha helical secondary structure of the peptide is not disrupted for the duration of the simulation.

To alleviate the possibility for the stable LKα14 to be a local minimum kinetically trapped in the energy surface of the system, we restarted the vacuum/water interface simulations with a random coil peptide initially submerged in bulk water. Fig. 4.11 shows the snapshots and the time evolution of the secondary structure of the simulation. After several attempts, the peptide moved to the interface in 40 ns and stayed there for the whole simulation time. Around 450 ns partial folding was observed and in 800 ns the peptide folded into a full α-helix conformation.

The induced stability of the α-helical conformation is due to the amphiphilic nature of the LKα14 peptide. The hydrophobic leucine sidechains face the vacuum side and the polar lysine sidechains extend into water. This partitioning can be seen

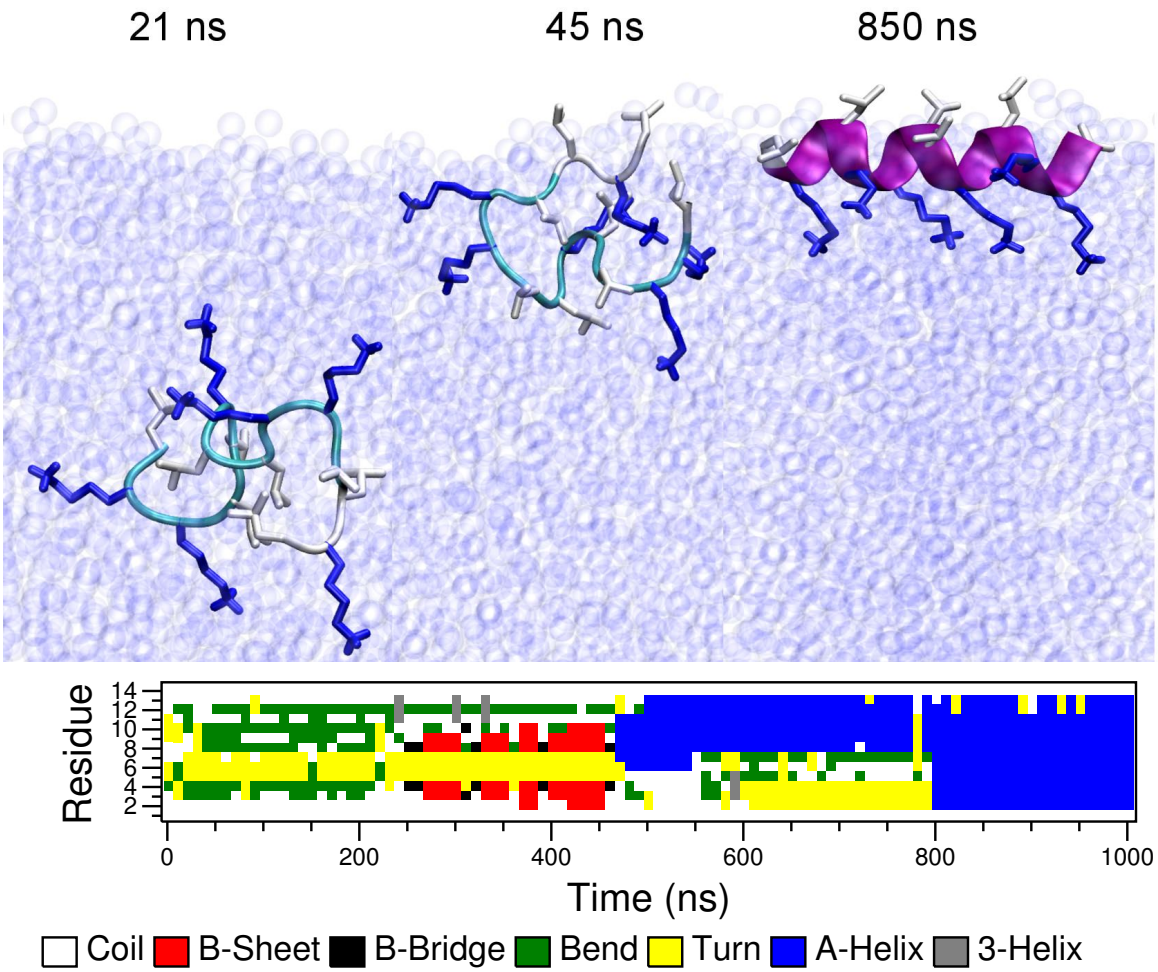


Figure 4.11: Time evolution of the secondary structure of single LKα14 simulation at the vacuum/water interface which is initially random and folds into α-helix.

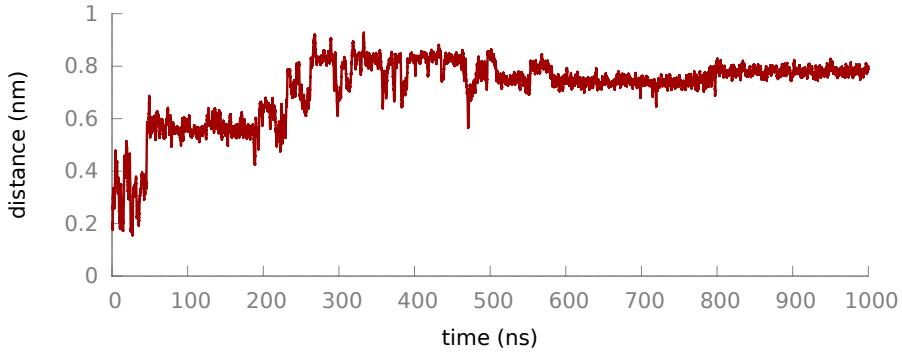


Figure 4.12: Single LKα14 at the vacuum/water interface: mean of the distance between the centers of mass of hydrophobic leucine sidechains and the charged NH_3^+ groups for a single peptide at the vacuum/water interface. The plot shows the amount of partitioning: as the distance between these groups increases partitioning increases. Initially the peptide is submerged inside the water slab, and moves to the interface in 40 ns (see Fig. 4.11). The α-helix conformation at the interface maximizes partitioning by preventing the hydration of leucine sidechains.

in the snaphost at the right in 4.11 where the polar lysine sidechains (blue licorice) extend into water and hydrophobic leucine sidechains (white licorice) face the vacuum side. The partitioning is quantified in Figure 4.12 by plotting the distance between the center of mass hydrophobic leucine sidechain and the positively charged NH_3 group. This distance increased as the peptide moves to the interface and is maximized when LKα14 adopts a full-α-helix secondary structure.

Fig. 4.13 shows the solvent accessible surface area for hydrophobic leucine sidechains for the simulation shown in Fig. 4.11. The adsorption of peptide to the interface shields the leucine sidechains (initial drop in Fig. 4.13) and the full α-helix minimizes the water-exposed area of the hydrophobic groups.

The LKα14 peptide displayed a strong adsorption force to the interface and the interface triggered a folding into full α-helix unlike the bulk water environment where the full α-helix peptide unfolded. To determine the range of conformations for the LKα14 in bulk water and to be sure if the folding observed in vacuum/water environment was not a rare event we performed replica exchange MD simulations for a single peptide in both environments.

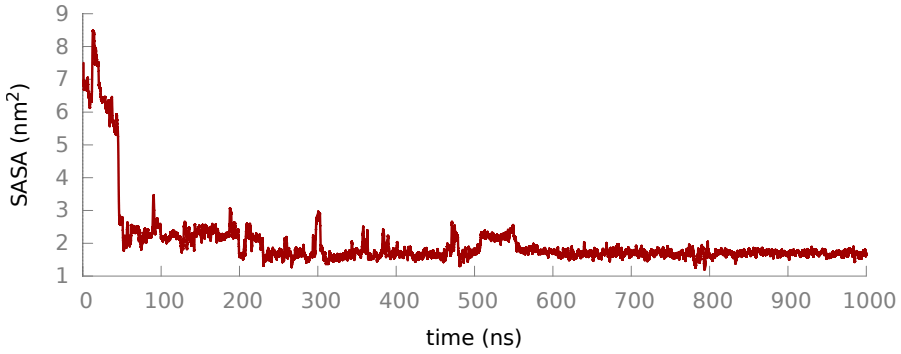


Figure 4.13: Mean of solvent accessible surface area for leucine sidechains for a single LKα14 at the vacuum/water interface. The peptide is initially random coil and submerged in bulk water. The initial drop corresponds to the peptide moving to the interface. The area of hydrophobic leucine sidechains exposed to water is minimised when the peptide adopts full α-helix (see Fig. 4.11). The area of the leucine sidechains facing vacuum is subtracted from the total area of the leucine sidechains accessible from scanning the protein surface to achieve the SASA.

To determine the equilibrium structure for a single LKα14 at the vacuum/water interface and remove the possibility of the full α-helical conformation to be a local minimum, we performed REMD simulations at the vacuum/water interface. The REMD simulations for single LKα14 at the vacuum/water interface contained 16 copies where the temperatures ranged between 298-366 K with an average exchange probability of 0.15. Similar to the replicas in bulk water REMD simulations, here the first 8 replicas were initiated from an α-helix LKα14 whereas in the remaining replicas the peptides were random coil.

Figure 4.14 shows the DSSP timeline data for each replica. The peptides were initially positioned to be at the vacuum/water interface and in none of the replicas drifting of the peptide away from the interface observed. The LKα14 peptides that started from a helical conformation did not change their secondary structures whereas the remaining replicas which were started with a random-coil folded to alpha-helix, where the last replica to fold did so in 300 ns.

In all the replicas, the peptides that adopted a full α-helix secondary structure displayed a partitioning of hydrophobic leucine and hydrophilic lysine residues. The

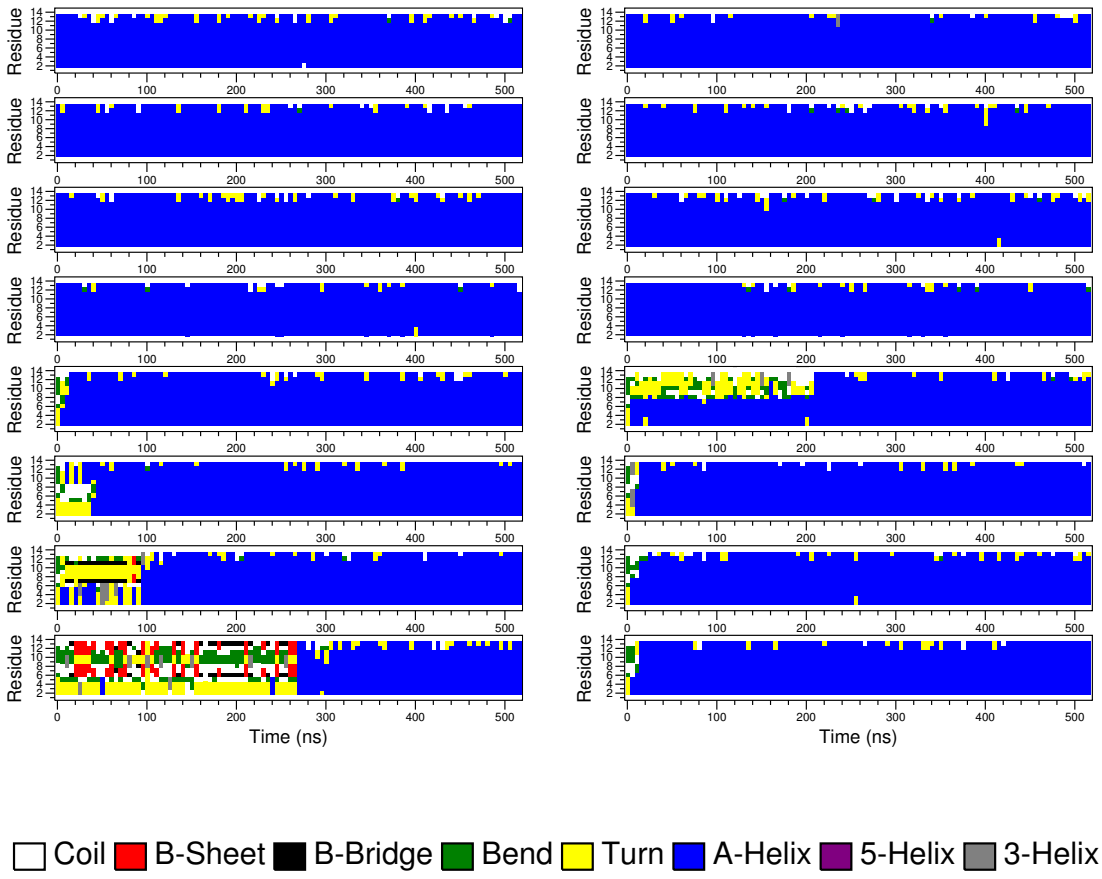


Figure 4.14: Replica exchange MD at the vacuum/water interface: Time evolution of the secondary structure of a single LKα14 peptide between temperatures 298-366 K. Secondary structure content was evaluated with the DSSP algorithm. The plots are ordered in increasing temperature from left to right and top to bottom.

amphiphilic properties of the LKα14 peptide favor the alpha-helix conformation at the vacuum/water interface whereas in bulk water the lack of strong driving forces decreases the probability of helical conformation. Therefore the α-helix is the equilibrium structure for a single LKα14 peptide at the vacuum/water interface compared to the bulk water case where no equilibrium secondary structure was observed.

The computational results lack direct comparison to experimental findings due to the absence of an experimental method that allows monitoring the conformation of a single peptide. However one can estimate the structural preference from the dilute peptide solution. Degrado and Lear suggested that in dilute solution LKα14 peptides converge to random coil. [148] The REMD simulations yielded unstable α-helices and have shown that a single LKα14 is intrinsically disordered without an unique stable conformation.

4.4 Aggregation of LKα14 Peptides

Degrado and Lear observed that LKα14 peptides form helical aggregates in bulk water at high chloride and peptide concentrations. [148] To determine the stability of aggregates for the atomistic model, we examined aggregates of LKα14 of different sizes. We started by analyzing the aggregation of two LKα14 peptides into a dimer both in bulk water and at the vacuum/water interface, and compared their secondary structures. Then using the PMF curves obtained through umbrella sampling simulations we investigated the stability of the aggregate.

To further examine the aggregation properties of LKα14 peptides, the larger size aggregates of trimer, tetramer, pentamer and octamer were simulated in bulk water where none of the aggregates dissociated. With the help of umbrella sampling simulations we compared the PMF curves for extracting one peptide from the aggregate. Among those aggregates, the tetramer was found to be the most stable aggregate size.

4.4.1 Bulk Water Simulations

Dimer Two LKα14 peptides can merge in two different fashions, namely the parallel or antiparallel orientation. The literature falls short on any suggestions regarding which orientation is the preferred one. Therefore we simulated the dimer of α-helix for each orientation for 1 μs. Fig. 4.15 provides the time evolution of these simulations as well as the snapshots from each simulation. A single full α-helical LKα14 peptide was unstable in bulk water whereas the dimer no unfolding was observed excluding the intermittent unfolding of the end residues. For 1 μs α-helix was stable for both the parallel and antiparallel orientations. Such stability of the α-helix was observed when a single peptide was placed at the vacuum/water interface where the dehydration of the hydrophobic leucine sidechains led to partitioning of the hydrophobic and the hydrophilic residues which stabilised the α-helix. Similarly, here, the added α-helix LKα14 peptide with its leucine sidechains all facing one side creates a micro-interface for the other peptide that delivers a hydrophobic surface which contributes to the helical stability. On this micro-interface a hydrophobic core forms that shields

hydrophobic leucine sidechains from water molecules pushing the lysine sidechains to gather on one side and thus to be fully hydrated in water (Fig. 4.15 snapshots). This partitioning of the hydrophobic and hydrophilic residues induces the stability of the α-helical secondary structure.

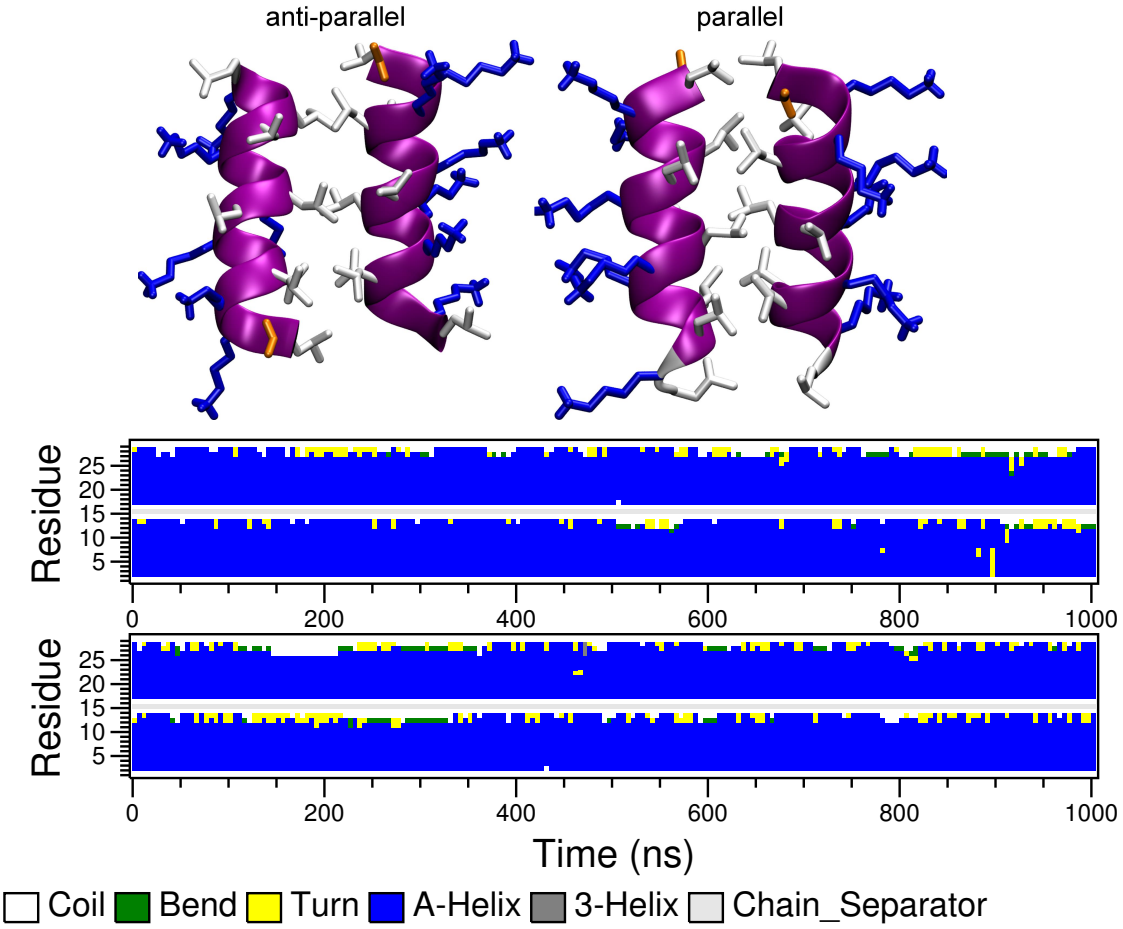


Figure 4.15: Time evolution of the secondary structure for a dimer of LKα14 in bulk water for antiparallel (top) and parallel (bottom) orientations. The α-helix secondary structure is stable for both orientations for the duration of the simulation. In the snapshots, the lysine sidechains (blue licorice) extend into bulk water whereas the leucine sidechains (white licorice) form a hydrophobic core. The peptide backbone is colored according to secondary structure with NewCartoon representation of VMD. The ACE cap residues are drawn with orange licorice.

Both of the simulations shown in Fig. 4.15 are started from an anti-parallel oriented dimer whereas one of those switched to parallel formation in the beginning of the

simulation as seen in the angle timeline plot in Fig. 4.16. The same plot reveals that the parallel one displayed an 80° of inter-dimer angle for about 100 ns. For more than 850 ns no major change was observed for the anti-parallel oriented dimer where after 850 ns the inter-dimer angle shifts to 30° converting the anti-parallel into a parallel. Therefore with these two simulations we observed that these two orientations can convert into each other on microsecond scales. Moreover, the peptides in the LKα14 dimer in bulk water are not perfectly aligned and are positioned with a non-zero angle with respect to each other. The histograms of the inter-dimer angle between the peptides are plotted in Fig. 4.16 where the parallel orientation peaks at 30° and the anti-parallel one at 150°.

When the dimers are oriented in parallel with respect to each other the distance between the center of mass of backbone atoms is lower than the one in anti-parallel yielding 0.97 nm for the former and 1.06 nm for the latter (Fig. 4.16), i.e. the dimers are positioned closer to each other when they are aligned in parallel. The distance of anti-parallel oriented dimer fluctuates between 0.8 and 2.2 nm, displayed as minor peaks in the distance histogram in bottom left of Fig. 4.16 where such peaks are non-existent for the parallel one. The correlation between the inter-dimer distance and the inter-dimer angle is provided in Fig. 4.17.

The peptides simulated in Fig. 4.15 were initially aggregated and folded. To alleviate the possibility for these aggregated forms with stable helical conformations to be a local minimum structure, we started another simulation with two separated (with 2.45 nm between centers of mass of backbone atoms) LKα14 peptides in bulk water each with a random conformation and simulated for 1 μs. Fig. 4.18 provides the snapshots, timeline evolution of the secondary structure, water exposed area of the leucine sidechains and the minimum distance between the peptides. As can be seen from the secondary structure plots (second plot from above) one of the peptides starts to fold into an α-helix conformation around 50 ns and finishes the folding process before the two peptides merge at 180 ns (the merging can be detected from the drop minimum distance plot at the bottom of Fig. 4.18). The peptides make contact from

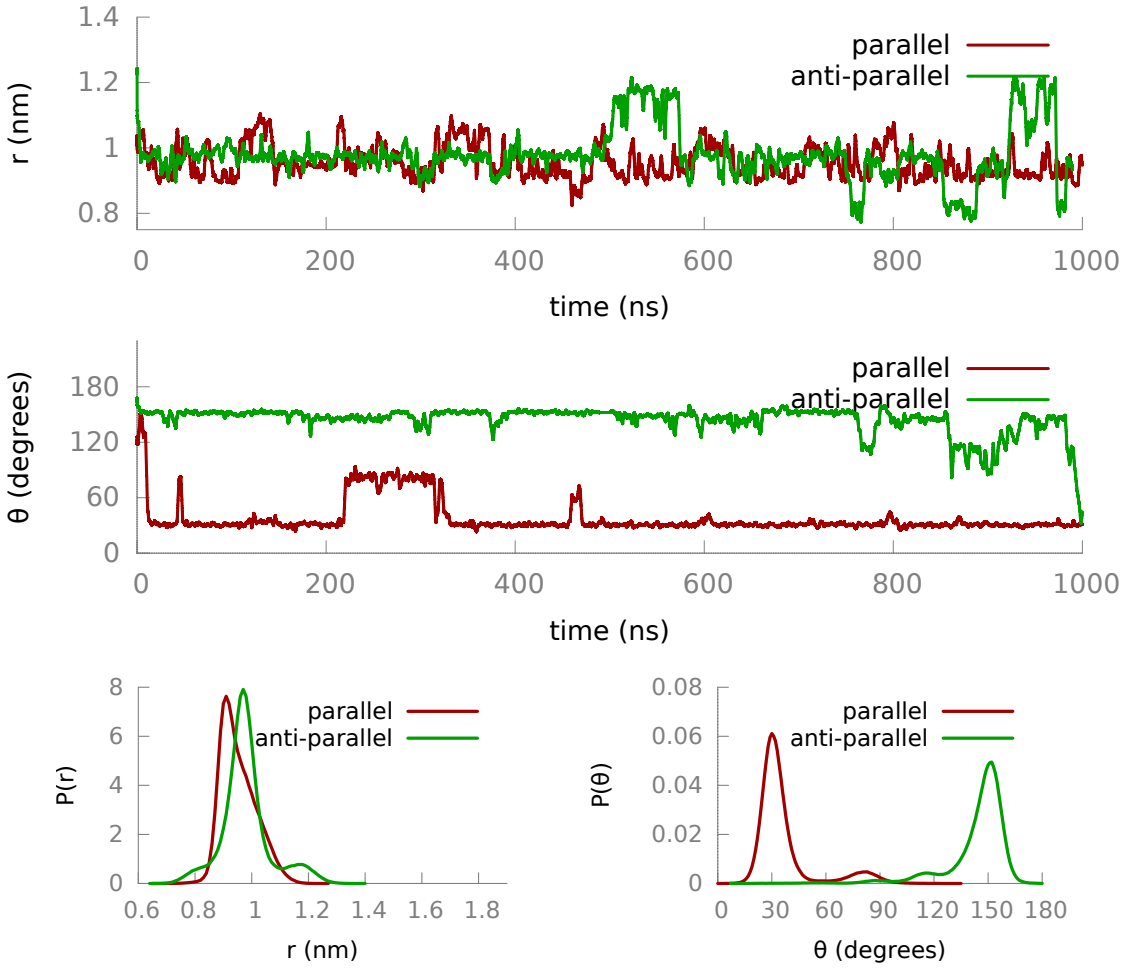


Figure 4.16: Comparison of parallel and anti-parallel orientation of LKα14 dimer in bulk water: Timeline of inter-dimer distance and inter-dimer angle (top) and the corresponding histograms (bottom). The distance is the distance between the centers of mass distance of the backbone atoms. The angle is the angle between the vectors passing through the backbone of the peptides, where the first and last 3 residues are not taken into account due to occasional unfolding of the end residues. The parallel orientation adopts an average inter-dimer angle of 30° and an inter-dimer distance of 0.97 nm where the anti-parallel one yields 150° and 1.06 nm.

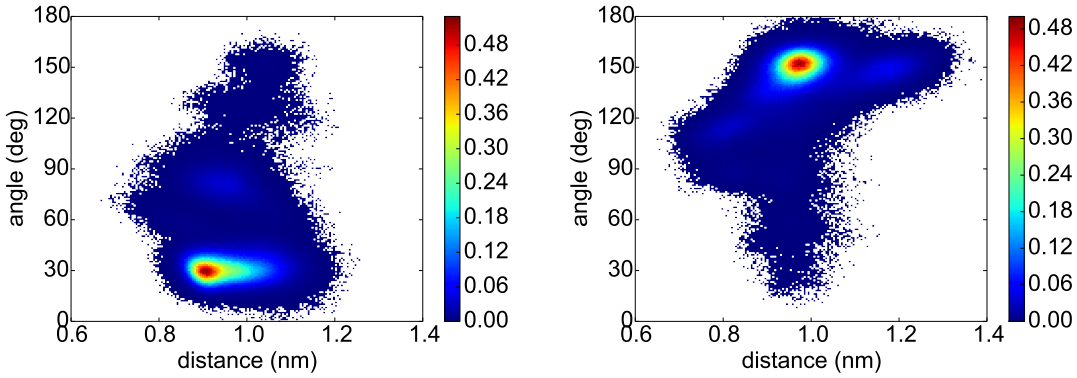


Figure 4.17: Distance angle correlation for a LKα14 dimer in bulk water for peptides aligned in parallel (left) and anti-parallel (right)

the end residues where the leucine sidechain of the one with the random conformation hooks onto the leucine sidechain of the folded peptide. After merging in parallel orientation the peptides do not dissociate and remain as an aggregated dimer for the rest of the simulation. The peptide that has the random coil structure firstly forms a half-helix at 580 ns, and at 750 ns folds into a full α-helix. The folding and aggregation observed in this simulation agrees with the results of the previous ones that the aggregate induces stability for the α-helix in bulk water. The solvent accessible surface area of the leucine sidechains (plot in the middle in Fig. 4.18) decreases once the peptides merge and form a hydrophobic core. This shielding of hydrophobic residues prevents the aggregate from dissociation throughout the simulation and the folding maximizes this shielding.

Both the anti-parallel and parallel orientations of the dimer displayed transitions between each other. To quantify the strength of these orientations in bulk water we performed umbrella simulations and pulled one peptide from the dimer relative to the other for each orientation and the corresponding potentials of mean force (PMF) can be seen in Fig. 4.19. Both of the conformations yield similar PMFs with a depth of 46 kJ/mol indicating that both are equally valid conformations for a dimer in bulk water. The PMF curves suggest that until the distance between the peptides is 2.5

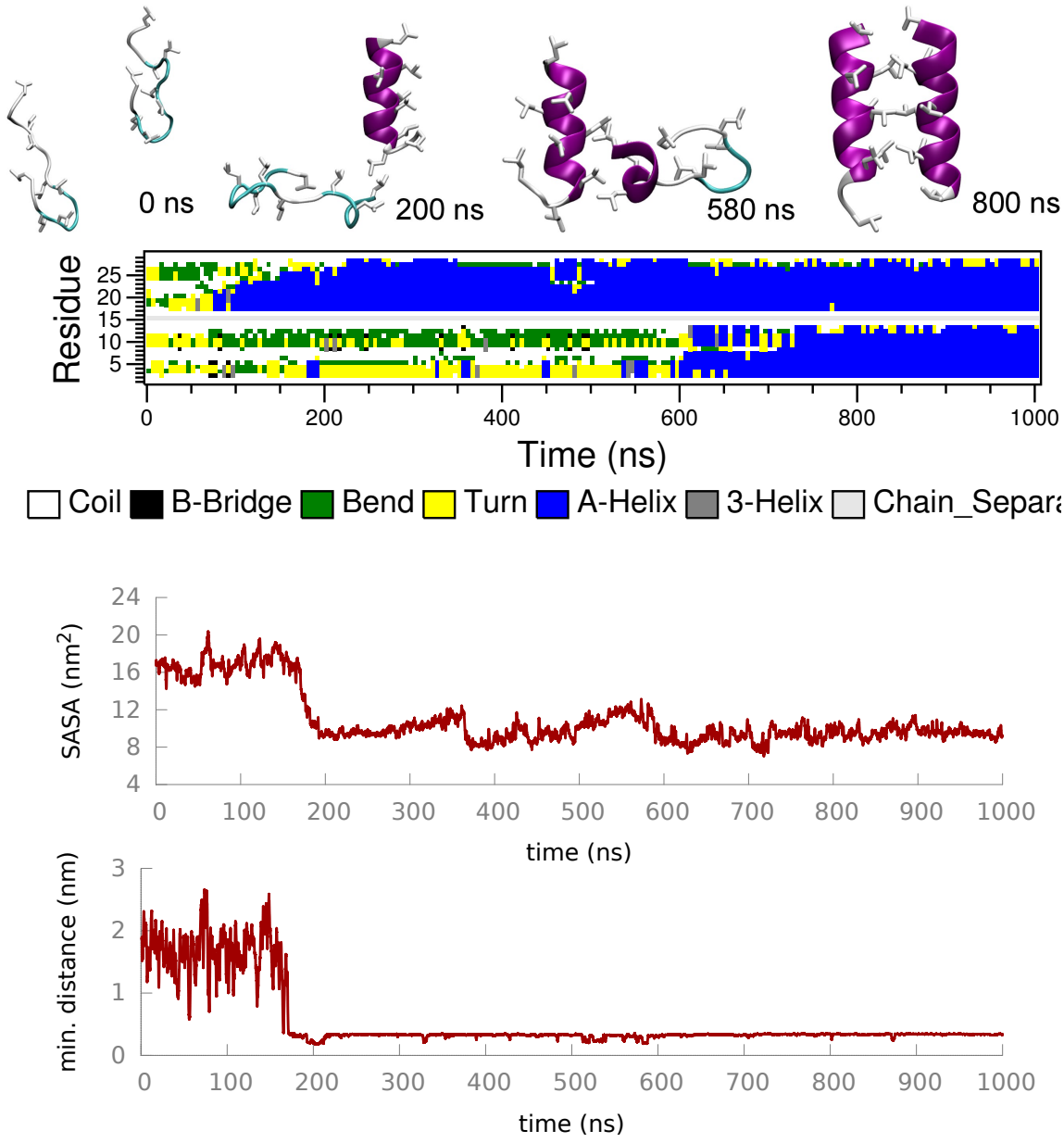


Figure 4.18: LKα14 dimer in bulk water: Two initially unfolded and separately positioned LKα14 peptides in bulk water first associate and then fold. Snapshots handpicked from the simulation are shown along with the time evolution of the secondary structure. Solvent accessible surface area for the leucine sidechains is plotted (third from above) and the minimum distance between the peptides is presented (bottom). The initial drop in both the minimum distance and SASA plots correspond to the association that takes around 180 ns.

nm the peptides attract each other, however if the inter-dimer distance is larger than 2.5 nm the peptides repel. Note that the initial distance was 2.45 nm in the previous simulation where the two initially separated random coil peptides aggregated and folded into α-helix. A different simulation that started with two full α-helix with a separation distance of 3.5 nm between the centers of mass of their backbone atoms, did not show any aggregation and the peptides for 250 ns and the full α-helical secondary structure was unstable.

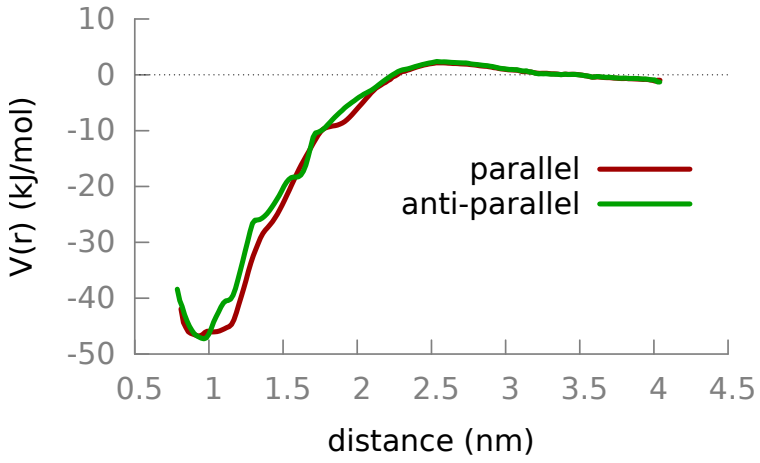


Figure 4.19: PMFs for separating one peptide from a dimer for parallel and anti-parallel orientations. The depth of the PMFs are the same for both orientations which suggests that they are equally preferable. The convergence for the PMF was checked by block analysis. Entropic correction was applied to the PMFs.

The source of the repulsion at large distances is the positively charged NH_3 groups at the end of the lysine sidechains creating an electrostatic repulsion at large distances which makes the aggregation of the molecules harder at low concentrations. This is tested by comparing the PMFs for pulling a peptide from an anti-parallel oriented dimer with the charged lysine sidechains and the uncharged ones. Removing the charges from the NH_3 group lends a flat PMF at large distances whereas the charged one yielded a negative slope (Fig. 4.20).

The repulsion that takes place at large distances highlights an important aggregation mechanism in bulk water. We found for two full α-helical LKα14 peptides that

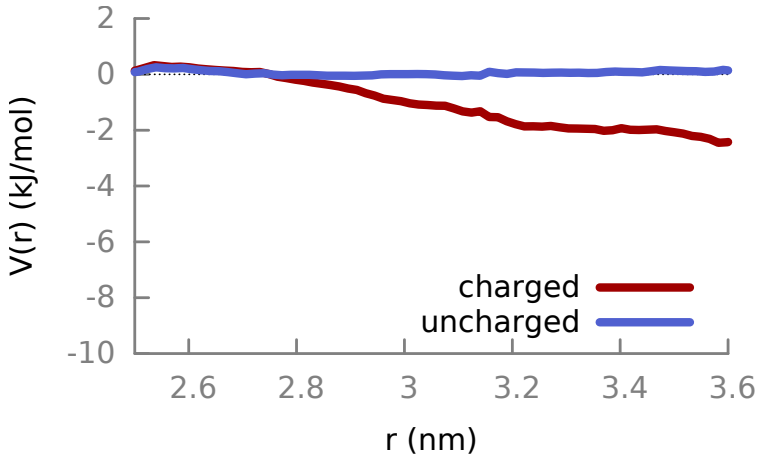


Figure 4.20: PMFs for separating one peptide from an anti-parallel oriented LKα14 dimer in bulk water obtained by umbrella simulations comparing the peptides where the lysine sidechains are charged and uncharged. The PMF of the charged peptides shows a decreasing slope after 2.5 nm which corresponds to a repulsion whereas the uncharged peptides display a flat PMF at large distances. Therefore the repulsion observed is of electrostatic origin. Entropic correction was applied to the PMFs.

are initially separated from each other, it took more than 300 ns to form a dimer. (Fig. 4.21) whereas the peptides that were initially random coil formed a dimer in 45 ns. The disordered structure of the LKα14 peptides covers a larger volume and thus increases the probability of making contact with the nearby peptides through the hydrophobic leucine sidechains. Whereas the full α-helical state is disadvantageous in that regard because all of the leucine sidechains are restricted to one side of the peptide and the repulsive electrostatic interaction hinders any association. This an example of the fly casting mechanism which proposes that the random structures have a higher probability to aggregate compared to ordered structures [169].

Frequency histograms were used to test the choice of windows in umbrella simulations (Fig. 4.22). Each window was chosen such that the distribution of the property of interest, i.e. distance between the center of mass of backbone atoms for the dimer, intersects with the histogram of the neighboring windows. The height of the histograms only show the length of the simulations and do not imply convergence which was determined by block analysis. The width of the histogram curves are affected by

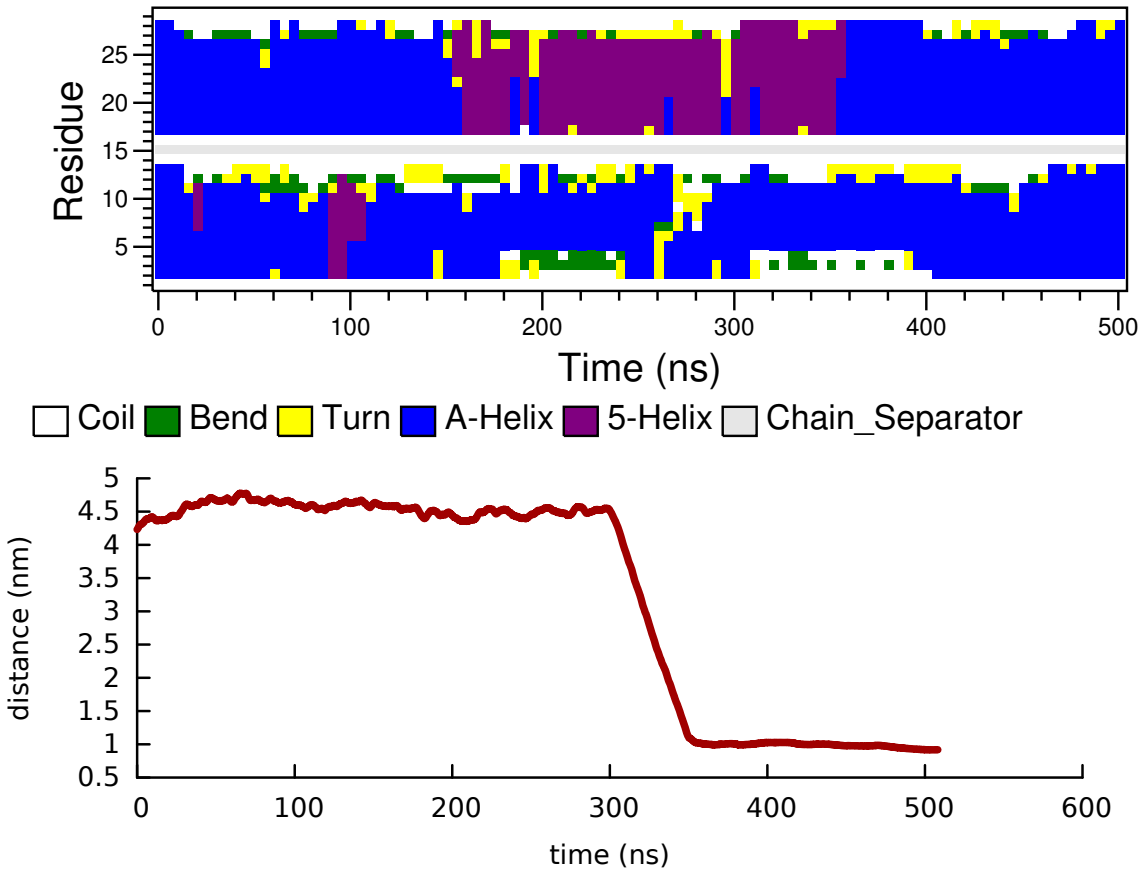


Figure 4.21: LKα14 dimer in bulk water: Initially, the peptides are full α-helix and separated. The peptides merged after 300 ns and α-helix for one of the peptides the full α-helix was not stable between 180-340 ns. Compared to the previous simulations where two randomly oriented peptides merged in 40 ns (Fig. 4.18) here the formation of the dimer takes more time. The reason behind this increased time for aggregation is the electrostatic repulsion observed at large distance as seen from the PMF curves in Figs. 4.19 and 4.20.

the choice of the bias factor k chosen as 1000 kJ/mol nm^2.

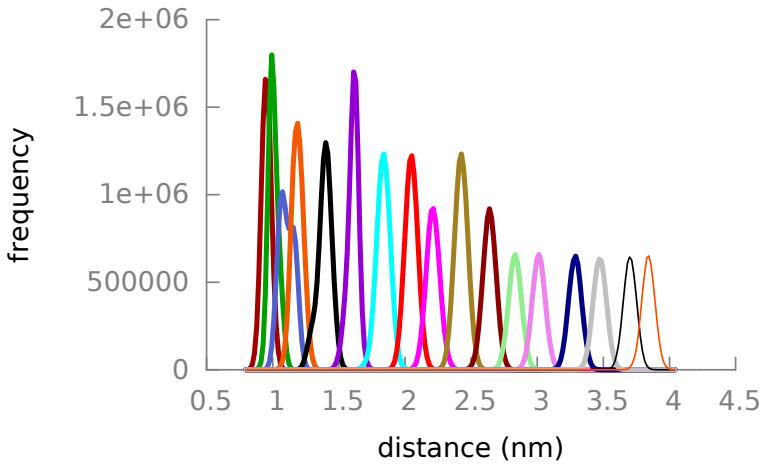


Figure 4.22: Frequency histograms for the umbrella simulations for pulling one peptide from the anti-parallel dimer. The histograms should intersect if one wants to achieve a reliable PMF curve.

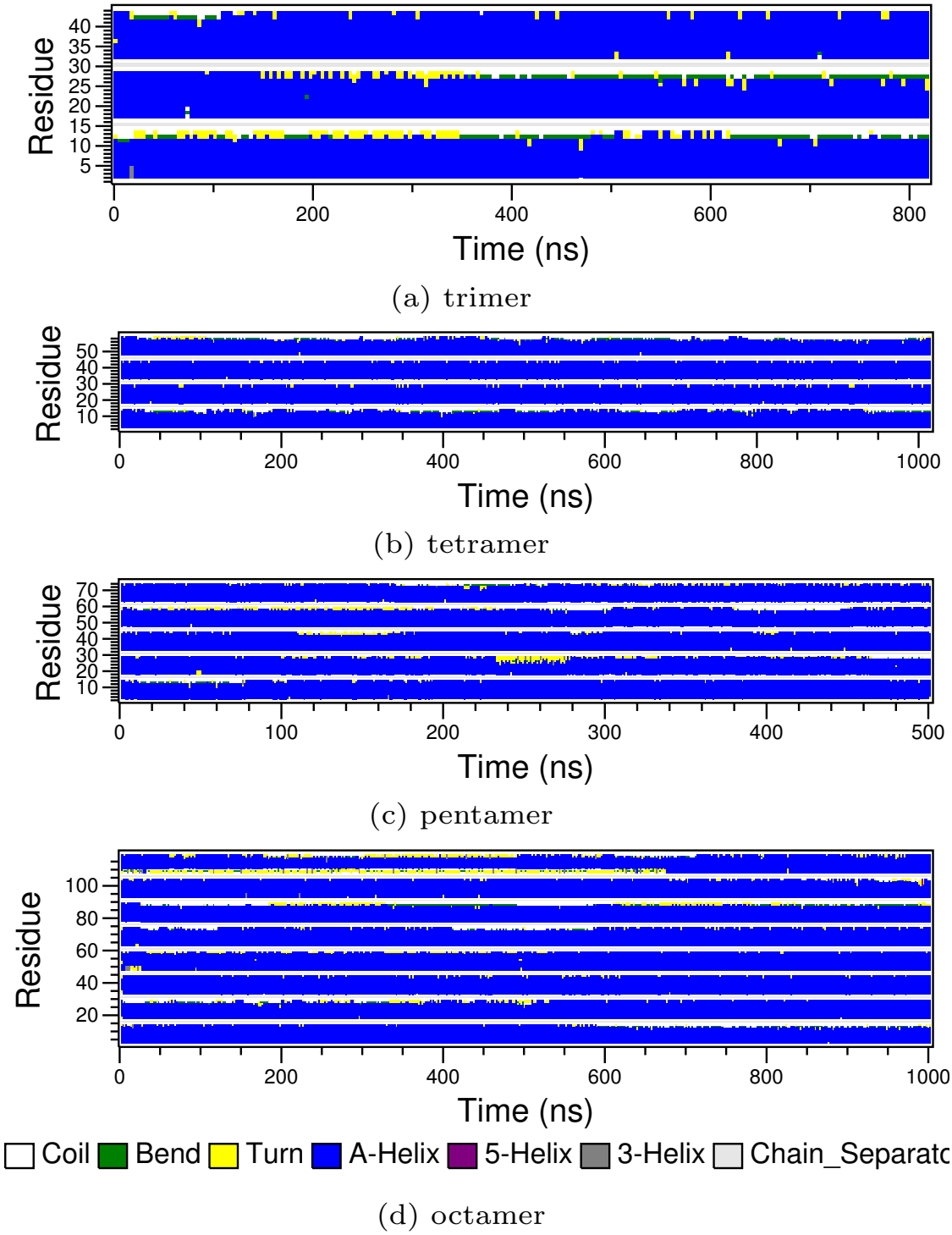
The LKα14 dimer in bulk water was stable and the dimer induced an ordered α-helical secondary structure unlike a single LKα14 in bulk water where the α-helix unfolded and there were no unique equilibrium structures. This stability was as a result of the leucine sidechains generating a hydrophobic core inside the dimer, shielding the hydrophobic residues from water molecules and the charged lysine sidechains extended into water. The PMFs results showed that the anti-parallel and parallel orientations were equally preferable. It is suggested that for two LKα14 peptides to aggregate random structures have a higher chance compared to the full α-helix conformations due to the increased probability of leucine sidechains making contact. The electrostatic repulsion at large distances because of the charged lysine sidechains hinder aggregation especially for full α-helix peptides where the short hydrophobic leucine sidechains are restricted to one side of the peptide. Next, the larger size aggregates in bulk water, namely the trimer, tetramer, pentamer and octamer will be discussed.

Larger Size Aggregates Experimental results suggest that LKα14 peptides fold into larger aggregate size than dimer. Degrado and Lear argued the most stable aggregate size being the four helix bundle in bulk water. [148] To analyze the aggregation properties of the larger aggregate size we simulated the full α-helical LKα14 aggregates of trimer, tetramer, pentamer and octamer in bulk water. The stability of the aggregates are analyzed and the secondary structures are monitored.

The secondary structure throughout the simulation trajectory for each simulation is portrayed in Fig. 4.23. The peptides are aligned in anti-parallel fashion where the neighboring peptides were aligned in head-to-tail. Throughout the simulations, none of the peptides in the aggregates displayed unfolding except for brief partial unfolding at the end residues which is expected due to hydration. The peptides maintained their aggregated form although the initial orientation of the peptides with respect to the neighboring peptides in the aggregate changed. The hydrophobic leucine sidechains formed a hydrophobic core and the lysine sidechains stretched into bulk water.

The snapshots of the aggregates can be seen in Figure 4.24. Each peptide in the trimer have a non-zero angle with its neighbor peptide. The third peptide added to the dimer shields leucine sidechains compared to a dimer. The neighboring peptides in the tetramer are positioned in anti-parallel fashion. During the simulation the two dimers in the tetramer sometimes display shifts in angle from a parallel orientation. There are cases where the dimers are oriented at a perpendicular angle. These shifts increase the distance between the charged NH_3^+ groups.

The added peptide to the pentamer forces the leucine sidechains away from the hydrophobic core the tetramer formed to the added peptide. However, this orientation brings the lysine sidechains closer introducing an instability for the tetramer and a continuous orientation and positioning during the simulation. Both the relative orientation and the positioning of peptides in octamer seems off where the leucine sidechains of the peptides at one end of the aggregate are oriented towards bulk water. Furthermore, the tetramers inside the octamer are not aligned and each one does not form a hydrophobic core because of the incorrect orientation of the leucine



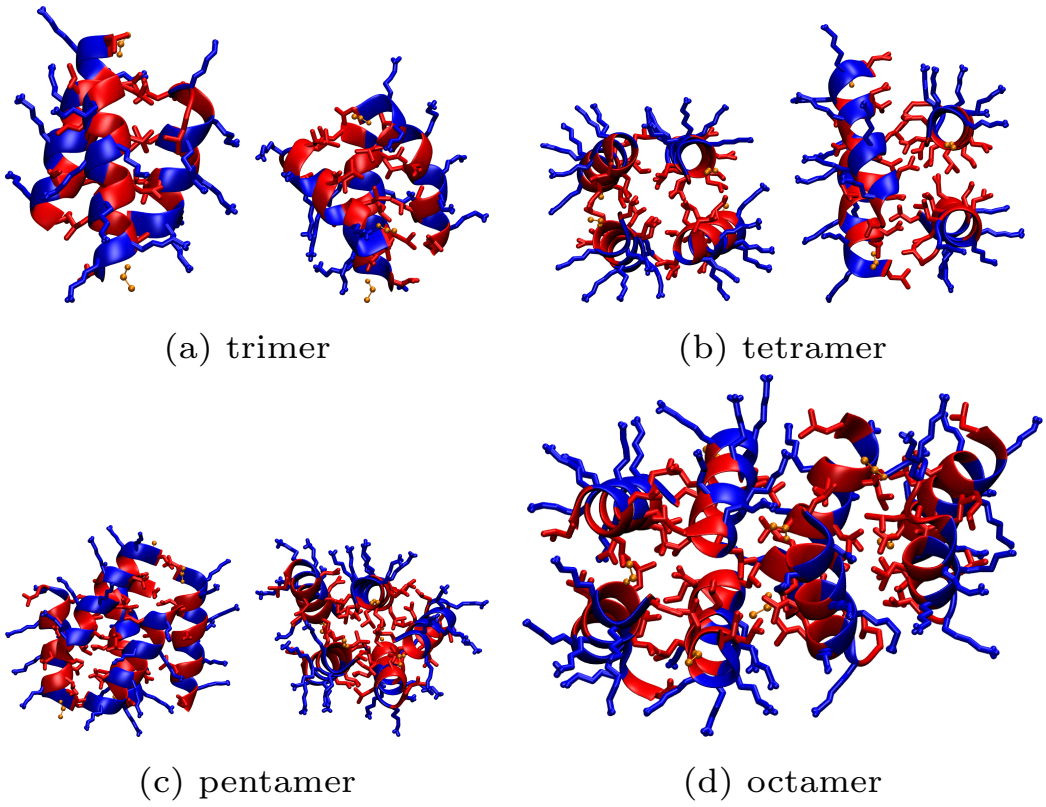
(a) trimer

(b) tetramer

(c) pentamer

(d) octamer

Figure 4.23: Time evolution of the secondary structure for each residue for trimer, tetramer, pentamer and octamer in bulk water color coded for each secondary structure.



(a) trimer (b) tetramer

(c) pentamer (d) octamer

Figure 4.24: Snapshots for the trimer, tetramer, pentamer and octamer simulations in bulk water.

sidechains. Considering that the charged lysine sidechains cover the outer region of the tetramer, additional tetramer near the one should be repelled.

As all the aggregates were stable during the simulations, to quantify the strength of each aggregate size we performed umbrella sampling simulations. The aggregation free energy of a single LKα14 peptide to the aggregates of trimer, tetramer and pentamer in bulk water was calculated. The octamer was omitted because the strength of the aggregate is expected to be lower than the ones simulated here because in the free simulation it adopted random orientations as discussed. Furthermore simulating the octamer has a high computational cost. In the pulling simulations, a single LKα14 peptide is pulled away relative to a peptide in the aggregate at a constant pull rate and at regular intervals snapshots of the system are taken. These snapshots formed the initial configuration of the umbrella sampling simulations where a harmonic spring was added between the center of mass of the backbone atoms of the pulled peptide and a reference peptide inside the remaining aggregate. Potential of mean force for aggregation was calculated using weighted histogram analysis.

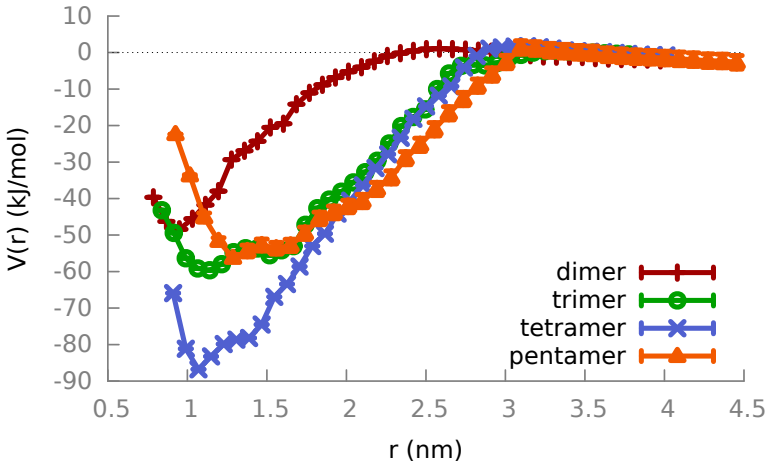


Figure 4.25: PMF curves obtained by umbrella sampling simulations for pulling away a single LKα14 peptide from the LKα14 aggregate. The tetramer has the highest depth among the four aggregates with the trimer and the pentamer having the next highest depth. Error bars are smaller than the point sizes, therefore may not be clearly visible.

Figure 4.25 shows the PMF curves obtained from umbrella sampling simulations where the pulling of one peptide from the tetramer has the highest depth of -88 kJ/mol making the tetramer as the most stable aggregate size. The contribution of a single peptide to the aggregate is similar for trimer with a depth of -62 kJ/mol and pentamer with a depth of -58 kJ/mol. The dimer had a depth of -50 kJ/mol and thus was found to be the least stable aggregate among the four aggregate sizes.

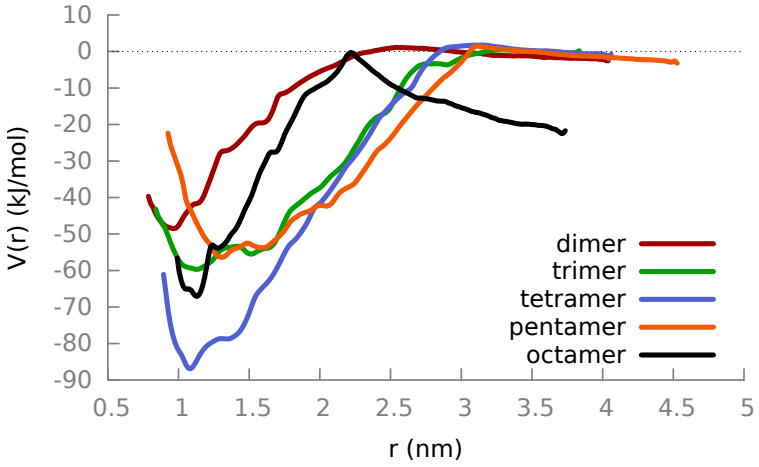
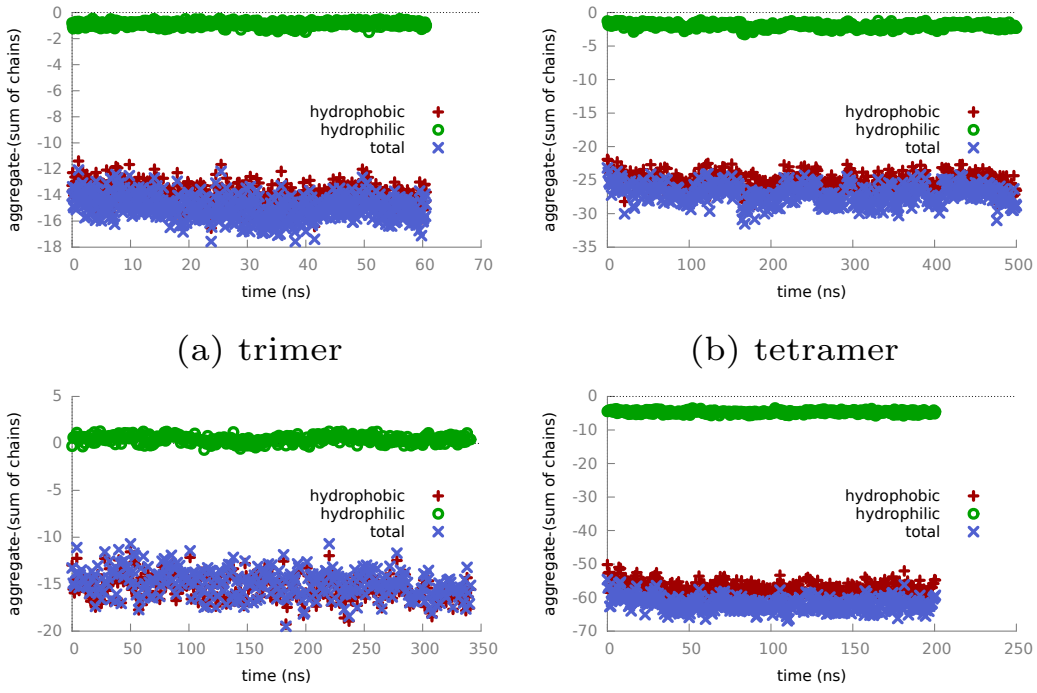


Figure 4.26: Umbrella PMF curves for pulling away a single LKα14 from the dimer, trimer, tetramer and pentamer compared to dividing the octamer into two tetramers.

If we pull the octamers and divide them into two tetramers, we obtain a PMF as seen in Figure 4.26 which has a depth around 65 kJ/mol but is still lower than pulling away one peptide from the tetramer.

Figure 4.27 shows the difference in SASA between the aggregate and the sum of individual peptides. Dividing the SASA difference with the number peptides in the aggregates octamer yields highest shielding of hydrophobic residues. However as discussed before, Figure 4.24 shows that the aggregate structure of octamer seems distorted and it is possible that the octamer will split into two tetramers. However further simulation is hampered by high computational demand. Ignoring the problematic octamer SASA results, tetramer displayed the second highest shielding with an average of 6.25 nm^2 per peptide whereas the trimer and pentamer had 5 nm^2 and



(a) trimer (b) tetramer

(c) pentamer (d) octamer

Figure 4.27: Difference in solvent accessible surface area between the aggregate and the sum of the individual chains, compared for different aggregate sizes of trimer, tetramer, pentamer and octamer. That is the sum of the SASA for individual chains is subtracted from the SASA of the whole aggregate and plotted according to time.

3 nm^2 respectively.

To sum up, among the aggregate sizes from two to five, the tetramer allows the highest shielding of the hydrophobic residues as well as the highest distance between the charged NH_3^+ groups. The peptides are oriented to form a hydrophobic core inside the aggregate and the charged lysine sidechains are extended outwards into water. The hydrophobic core of the dimer exposes the hydrophobic leucine sidechains to water molecules from the sides of the aggregate. Adding another peptide to the dimer reduces this exposure, the dehydration of hydrophobic leucine sidechains is maximized in the tetramer formation.

Larger size aggregates are shown to be stable in bulk water and similar to the dimer case, full α-helical secondary structure was conserved. In agreement with the experimental results, tetramer has been shown to be the most stable aggregate size displaying highest shielding for the hydrophobic leucine sidechains from polar water molecules. Next, the effect of vacuum/water interface on the LKα14 aggregates will be examined.

4.4.2 Vacuum/Water Interface Simulations

In the previous section we showed that the vacuum/water interface induced stability of α-helix for a single LKα14. And in bulk water the dimer was stable and aggregation induced a stable α-helix secondary structure. To test the stability of the aggregate at the vacuum/water interface we simulated the LKα14 dimer and observed that the aggregate was not as stable as it was in bulk water. The secondary structure for the LKα14 dimer is documented in Fig. 4.28. The α-helix is stable throughout the simulation time of 1 μs and the peptides exhibit the partitioning of the hydrophobic leucine and hydrophilic lysine residues (snapshots in Fig. 4.28), similar to the single LKα14 at the vacuum/water interface simulations.

The timeline for the inter-dimer distance and inter-dimer angle along with the histograms for each are compared with those of the bulk water simulations in Fig. 4.29. For the vacuuum/water environment, the distance between the centers of mass of the

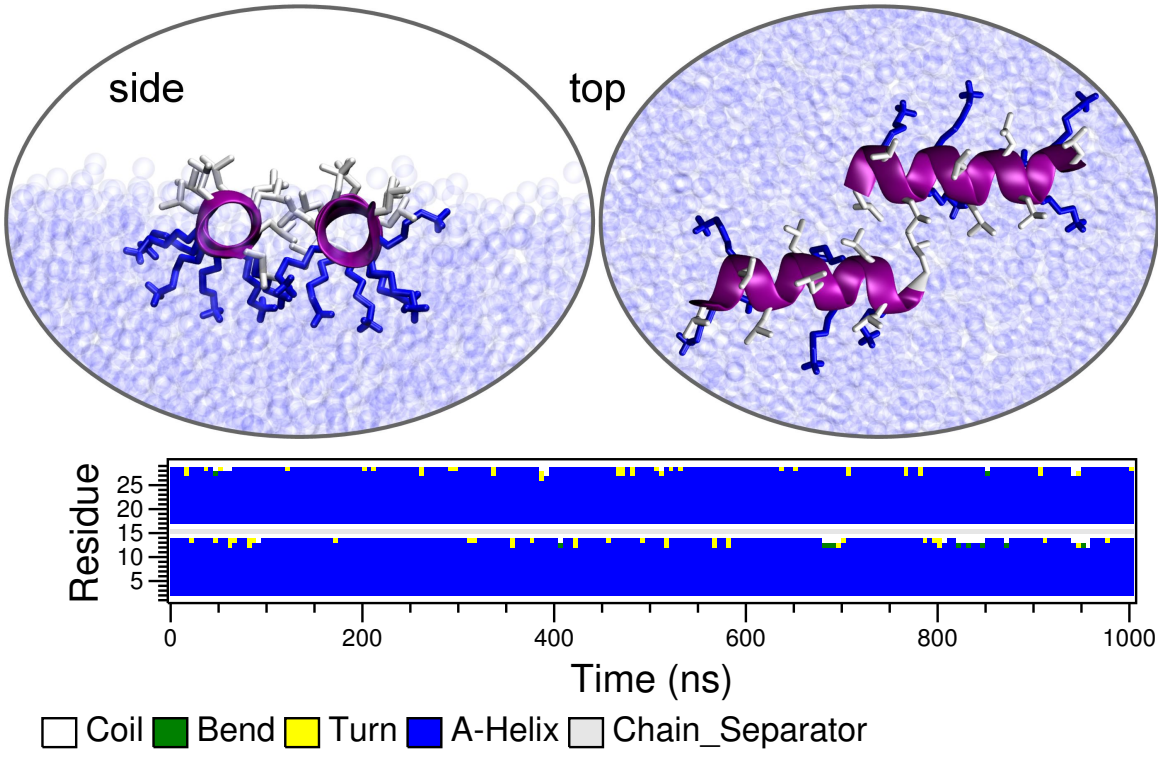


Figure 4.28: Time evolution of the secondary structure for a dimer at the vacuum/water interface.

backbone atoms ranged between 1 and 4.5 nm with three distinct peaks at 1.05, 1.39 and 1.66 nm in the distance histogram whereas in bulk water, the distance remains stiffly around 1 nm. Compared to the bulk water, here the peptides prefer to be positioned at a larger distance making the inter-peptide interaction weaker but not to the point that causes a stable dissociated structure. Although the dimer dissociates several times (inter-dimer distances larger than 3 nm), in each case they return to form an aggregate. The histogram of the inter-dimer angle has a broader peak compared to that of the bulk water simulation and there were random short-lived switches to parallel orientation.

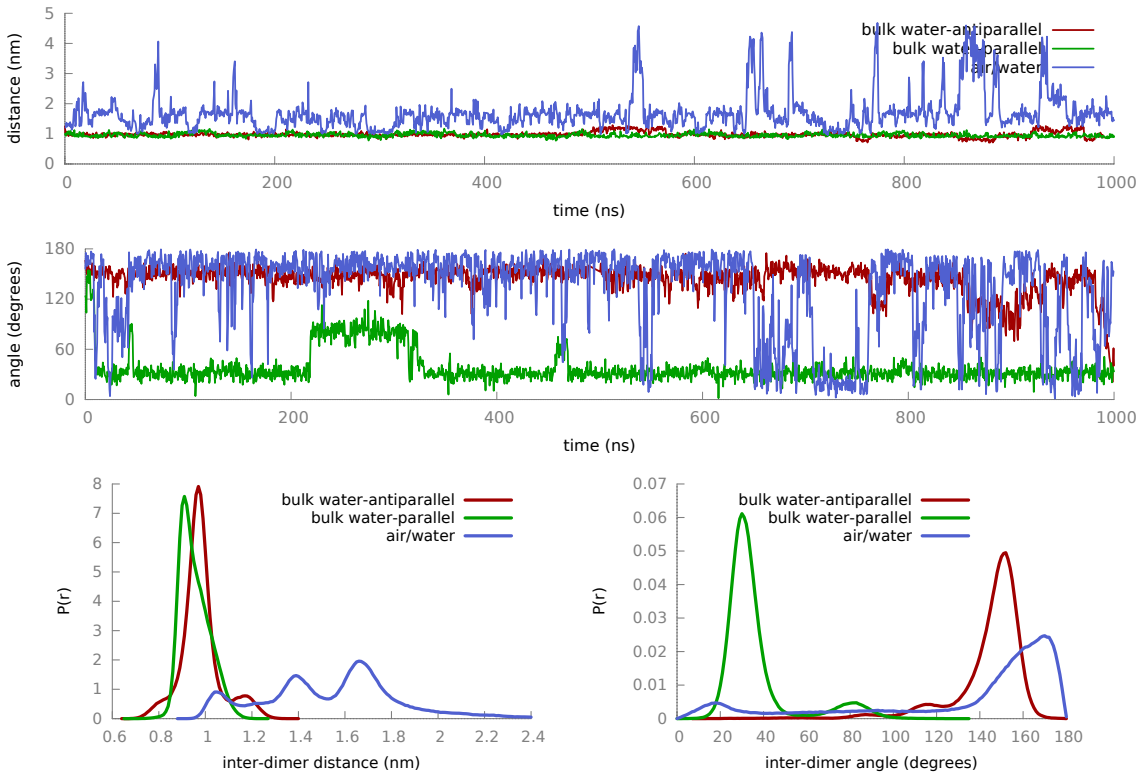


Figure 4.29: LKα14 Dimer in bulk water (red) and vacuum/water (green) interface. Timelines: Inter-dimer distance (top), inter-dimer angle (2nd from above). Histograms: inter-dimer distance (bottom, left), inter-dimer angle (bottom,right). Unlike bulk water the aggregate is loosely positioned with a weaker drive to form a stable aggregate.

The histogram for the inter-dimer distance showed three distinct peaks and the

correlation between the inter-dimer distance and inter-dimer angle in Figure 4.30 reveals 5 highly populated conformations defined in Table 4.1. The snapshots for these conformations are displayed in Fig. 4.31. The anti-parallel orientations are heavily populated compared to the parallel ones. This is not the result of an initial conformational bias because switches to parallel orientation are observed which switch back to anti-parallel. If the LKα14 dimer is started in a parallel fashion, an identical distance-angle correlation plot is obtained (data not shown). Among anti-parallel orientations, the conformations with larger distances between the center of mass of the peptide backbone are more probable compared to smaller distances which is thought to be the result of repulsive charge-charge interactions between the lysine sidechains.

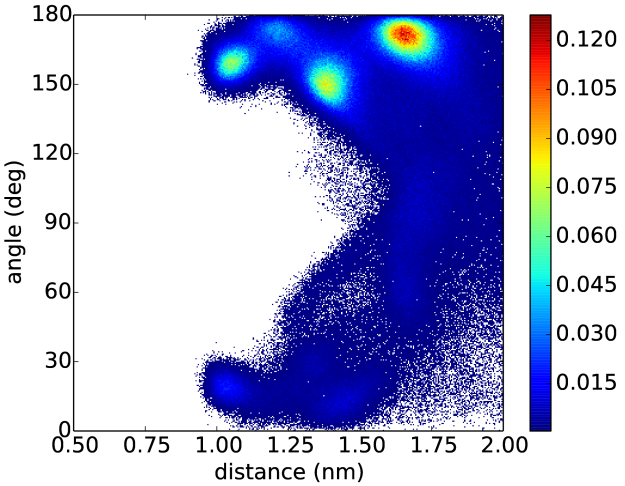


Figure 4.30: Inter-dimer distance vs inter-dimer angle correlation for a LKα14 dimer at the vacuum/water interface. The anti-parallel orientation is more populated compared to the parallel one and the dimer does not dissociate but instead prefers to position with larger distances. The three distinct peaks are can be seen in Fig. 4.31.

To quantify the dimer stability we generated PMF curves using umbrella sampling simulations and compared the strength of the dimer for the bulk water and vacuum/water interface. Figure 4.32 exhibits the PMF curves for dissociation of a dimer in bulk water and at the vacuum/water interface. The deeper PMF curve (48 kJ/mol) for bulk water agrees with the free MD simulations that the dimer is more

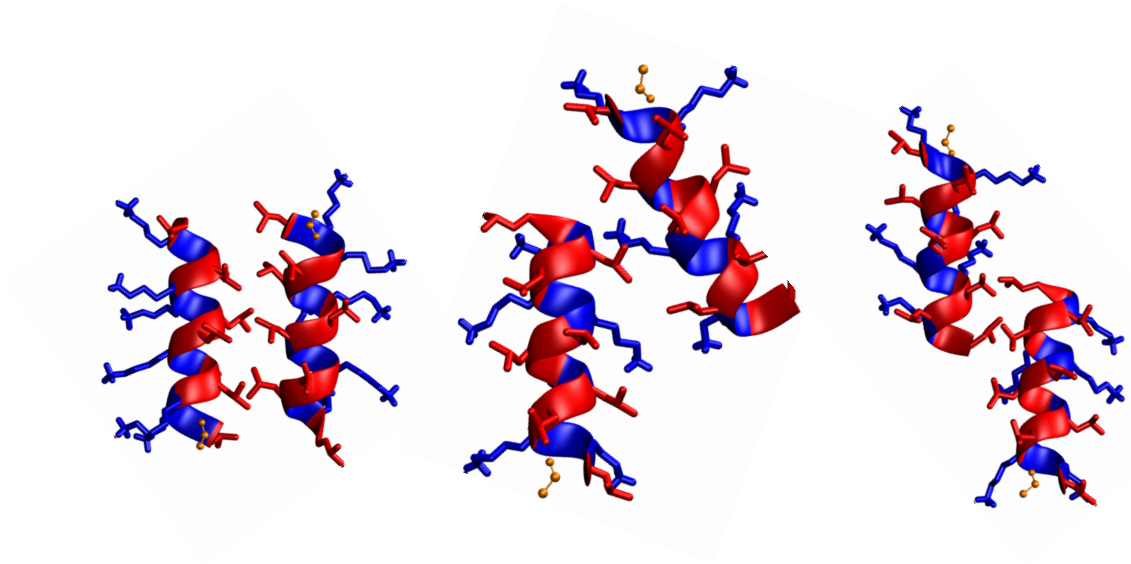
Table 4.1: Highly populated conformations for a LKα14 dimer at the vacuum/water interface defined in terms of inter-dimer distance and inter-dimer angle obtained from the correlation plot in Fig. 4.30.

	basin 1	basin 2	basin 3	basin 4	basin 5
distance	0.98-1.27	1.27-1.49	1.53-1.8	0.95-1.18	1.26-1.6
angle	159-175	138.5-166	153.5-179	4.80-24	2.5-33.5

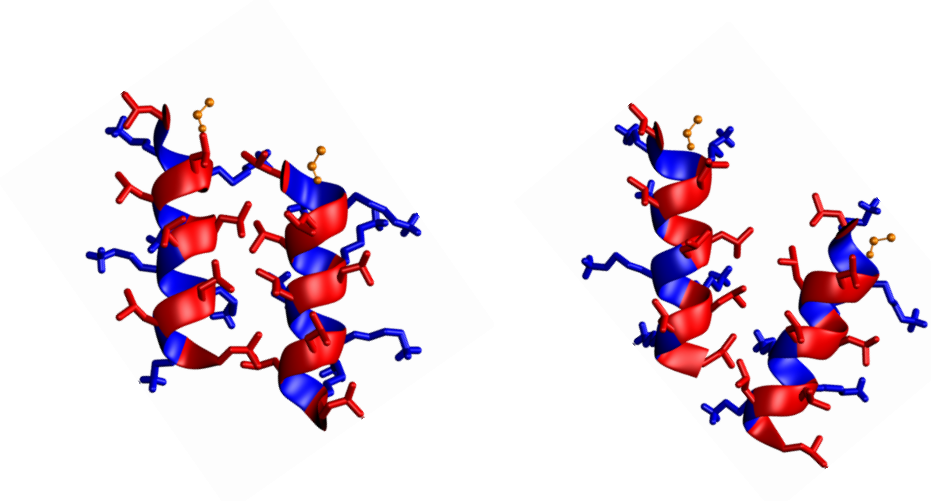
stable compared than the vacuum/water interface. The preference of anti-parallel orientations were documented for the free simulations with the distance-angle correlation plots in Fig.4.30 and the umbrella simulations yield a deeper curve for the anti-parallel. The peptides displayed electrostatic repulsion at inter-peptide distances larger than 2.5 nm whereas the PMF curves of the vacuum/water pulling simulations are flat at distances larger than 2.5 nm.

The decrease in dimer stability upon interacting with the vacuum/water interface is the result of the disruption of the hydrophobic core. In bulk water the hydrophobic leucine residues face each other and shield from the water molecules whereas at the vacuum/water interface the leucine sidechains extend into vacuum, thus making the formation of such a stable core unnecessary because the leucine sidechains are already dehydrated. This conformational change decreases the number of hydrophobic interactions and thus weakens the attraction between the peptides compared to the bulk water case.

The weaker attraction within the dimer at the vacuum/water interface is the result of the reduced distance between the charged lysine sidechains. Figure 4.33 indicates the histogram of the minimum distance between charged lysine sidechains. Note that the distance between the center of mass of backbone atoms between the anti-parallel orientation in bulk water and the basin 1 (inter-dimer distance: 1.05 nm, inter-dimer angle 160°) of the vacuum/water interface are similar, however the lysine sidechains of peptides are positioned closer to each other at the vacuum/water interface. Therefore electrostatic repulsion between the charged NH_3 groups drives the peptides away to



(a) 1.05 nm, 160° (b) 1.38 nm, 150° (c) 1.65 nm, 170°



(d) 1.04 nm, 19° (e) 1.45 nm, 15°

Figure 4.31: Representative snapshots of high probability conformations for a LKα14 dimer at the vacuum/water interface. The top three figures are the anti-parallel orientations and correspond to the peaks observed in distance angle correlation plot in Fig. 4.30. The ones at the bottom are parallel orientations. Lysine and leucine residues are colored blue and red, respectively, and the ACE cap is colored gold with CPK representation.

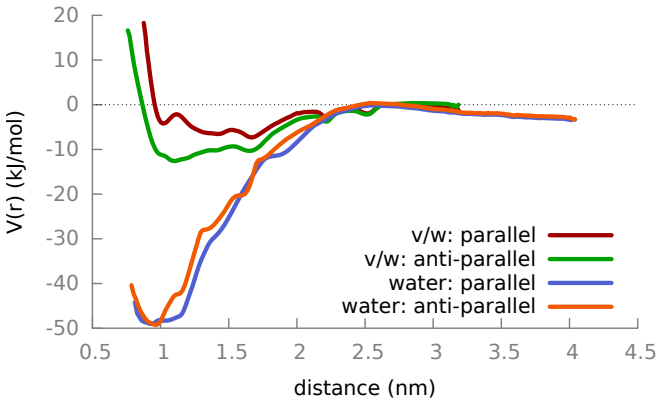


Figure 4.32: Comparison of PMF curves for separating one LKα14 peptide from a dimer in bulk water and vacuum/water interface obtained with umbrella sampling method. The bulk water PMF is deeper than those for the dimer at the vacuum/water interface. The anti-parallel orientation has a higher depth compared to that of the parallel in vacuum/water interface. Note that the at distances larger than 2.4 same simulations were used as windows for the vacuum/water simulations because at these distances the anti-parallel and parallel orientations are not meaningful and only the separation distance matters.

increase this distance, increasing the probability of the orientation in basin 3 (inter-dimer distance: 1.65 nm, inter-dimer angle 170°). In bulk water the lysine sidechains are positioned with a larger distance which contributes to the stability of the dimer.

The attraction between the hydrophobic leucine sidechains stabilised the aggregate in bulk water. Therefore as the number of contacts between these residues increases the conformation should be more stable. To see the effect of interactions of leucine sidechains between the different conformations observed in the correlation plot in Fig. 4.30 we calculated the number of contacts between the leucine sidechains for each conformation seen in Fig. 4.31. Using the RDF between the leucine sidechains for the tetramer, found as the most stable aggregate in bulk water, we chose a threshold distance of 0.67 nm which includes the first peak (Fig. 4.35). Fig. 4.34 shows the average number of inter-peptide contacts where with increasing inter-peptide distance the number of contacts decreases and the parallel conformations yield aligns leucine sidechains closer. However the most populated conformation of all, basin 3,

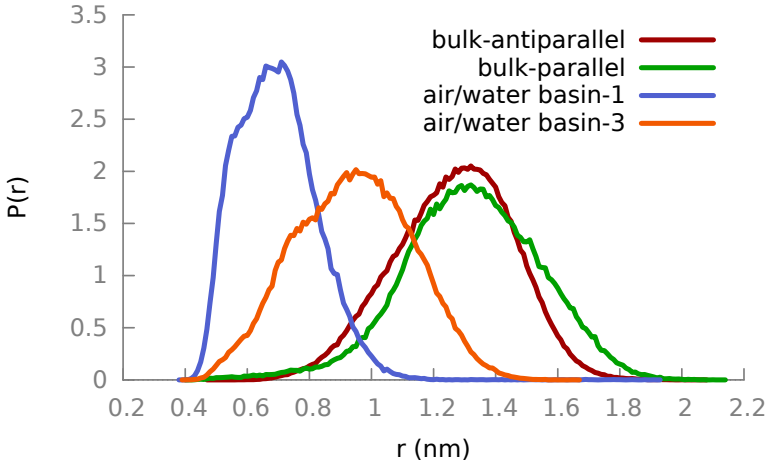


Figure 4.33: Histogram of the minimum inter-peptide distance between charged N atoms at the end of lysine sidechains for bulk water and basin 1 and 3 of vacuum/water interface.

has the second lowest number of contacts, suggesting the distance between the leucine sidechains have very little effect on the final conformation for the vacuum/water interface simulations. The leucine sidechains are already dehydrated by facing the vacuum side and to increase the number of interactions one has to position the peptides closer to each other which at the same time reduces the distance between the charged lysine sidechains increasing the electrostatic repulsion. Therefore the electrostatic repulsion and dehydration of the leucine sidechains are the dominant forces at play in vacuum/water interface simulations.

Figure 4.36 shows the SASA of the dimer minus the sum of the SASA for the individual peptides for an LKα14 dimer in bulk water and at the vacuum/water interface. Note that for the vacuum/water interface the figure illustrates the effect of dimer formation on hydrophobic leucine and hydrophilic lysine residues and does not consider the penetration of the water molecules to the residues extending into vacuum.

For all dimers there is no difference in the SASA of the hydrophilic lysine residues between the dimer and the individual peptides except for the last 50 ns of the parallel oriented dimer in bulk water where the difference is small compared to the

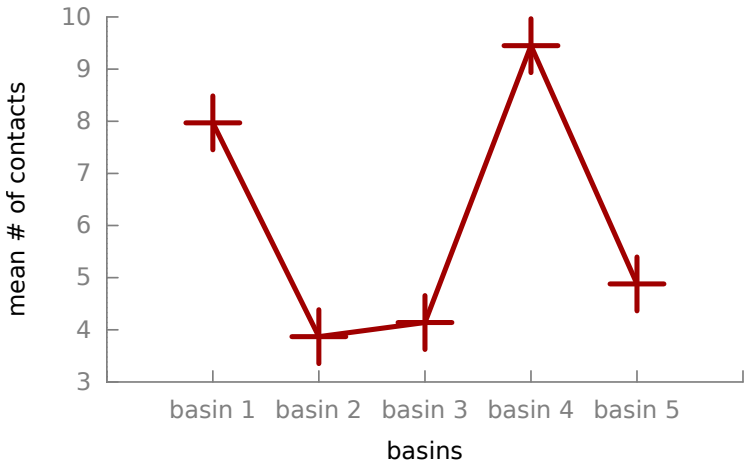


Figure 4.34: Mean number of inter-peptide leucine-leucine sidechain contacts for each highly populated conformation defined in Table 4.1. The snapshots for these conformations are shown in Fig. 4.31. The threshold distance was determined to include the first peak of the RDF (shown in Fig. 4.35) for the tetramer which was the most stable aggregate size in bulk water.

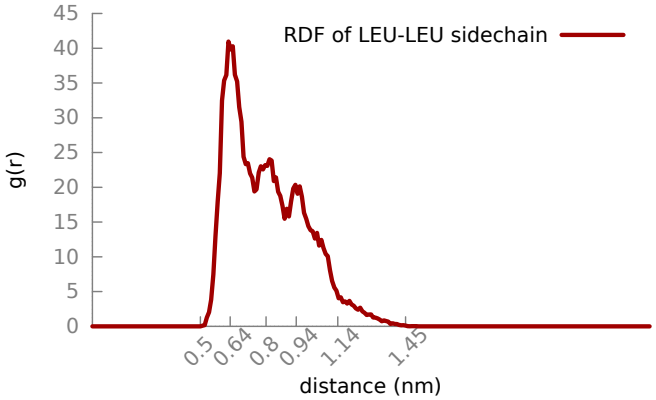


Figure 4.35: RDF of leucine-leucine sidechains

hydrophobic residue. The SASA plots suggest that the dimer formation is dominated by hydrophobic leucine residues. Table 4.2 compares the solvent accessible surface area for leucine sidechains for each simulation system, single and dimer in bulk water and at the vacuum/water interface. The decrease in SASA when a single LKα14 in bulk water moves to the vacuum/water interface is because of the partitioning of the hydrophobic and hydrophilic residues shielding the hydrophobic leucine sidechains as discussed before. Compared to the single in bulk water, the dimer reduces the exposure of leucine sidechains to the water molecules by the aid of the hydrophobic core. The difference of SASA between the single and dimer at the vacuum/water interface simulations is not significant and therefore the SASA values can be treated as equal.

Overall, the SASA plots are in accordance with the previous findings. The aggregated LKα14 peptides form a hydrophobic core in bulk water preventing water penetration and thus stabilizing the aggregate.

bulk water		vacuum/water	
single	dimer	single	dimer
7.11	4.83 (68%)	1.70 (24%)	2.26 (32%)

Table 4.2: Area of hydrophobic leucine sidechains that are exposed to water molecules for LKα14 in nm^2. For the cases involving a dimer, exposure of leucines per peptide is calculated. Comparison of the SASA values with the single peptide in bulk case is listed as percentages. The errors for each case are lower than 0.01.

When eight LKα14 peptides are placed at the vacuum/water interface and aligned in anti-parallel fashion on the surface in a row as shown in Fig. 4.37, and simulated for 100 ns, the aggregate immediately disintegrates and the peptides are positioned randomly. The final snapshots of the simulation are denoted in Fig. 4.38. There are two loosely interacting dimers in the final snapshot where one adopts the conformation shown in basin 3 and the other adopts basin 1. However these dimers are short-lived and the peptides disassociate to either form another dimer with or roam freely at the interface. The peptides remain at the interface for the duration of the simulation

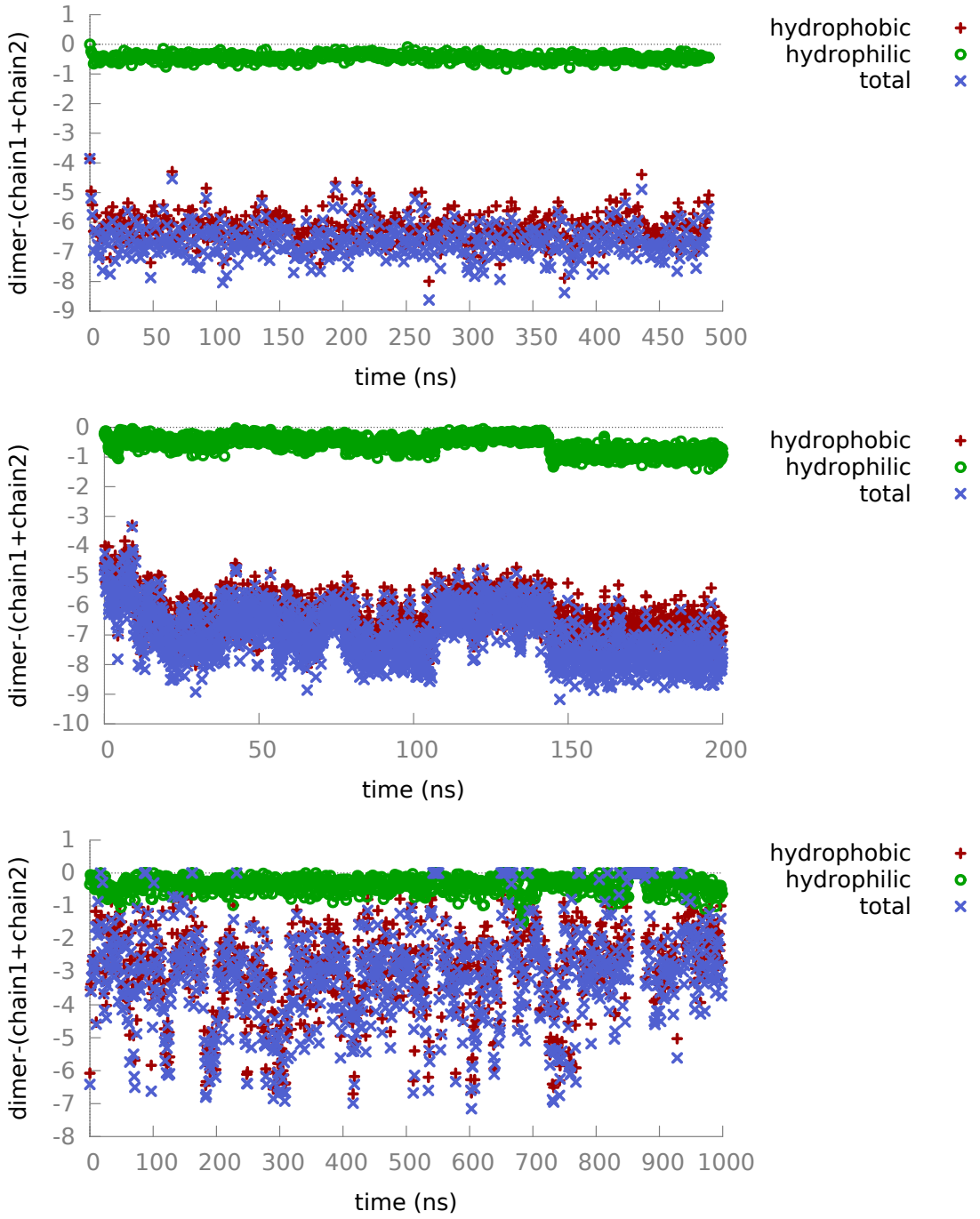


Figure 4.36: Difference between solvent accessible surface area of dimer and sum of individual peptides in bulk water for anti-parallel (top) and parallel (middle) orientations and for dimer at vacuum/water interface (bottom).

and the hydrophobic and the hydrophilic residues are partitioned with the α-helix secondary structure remaining stable (Fig. 4.39). This simulation shows that at the vacuum/water interface the larger size aggregates observed in bulk water are not stable.

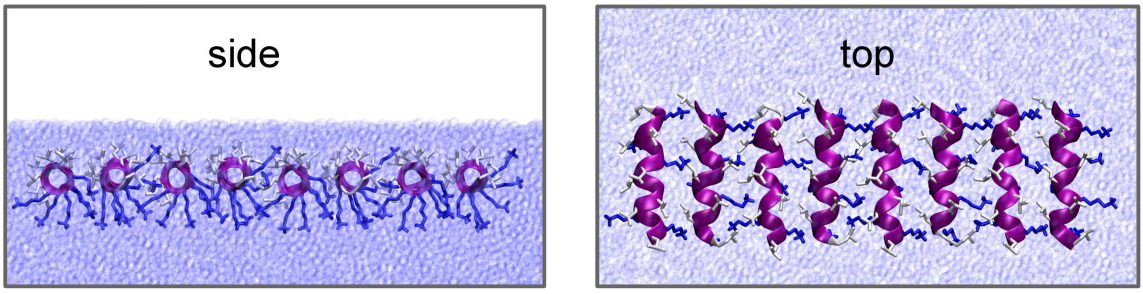


Figure 4.37: The initial configurations for eight LKα14 peptides at the vacuum/water interface

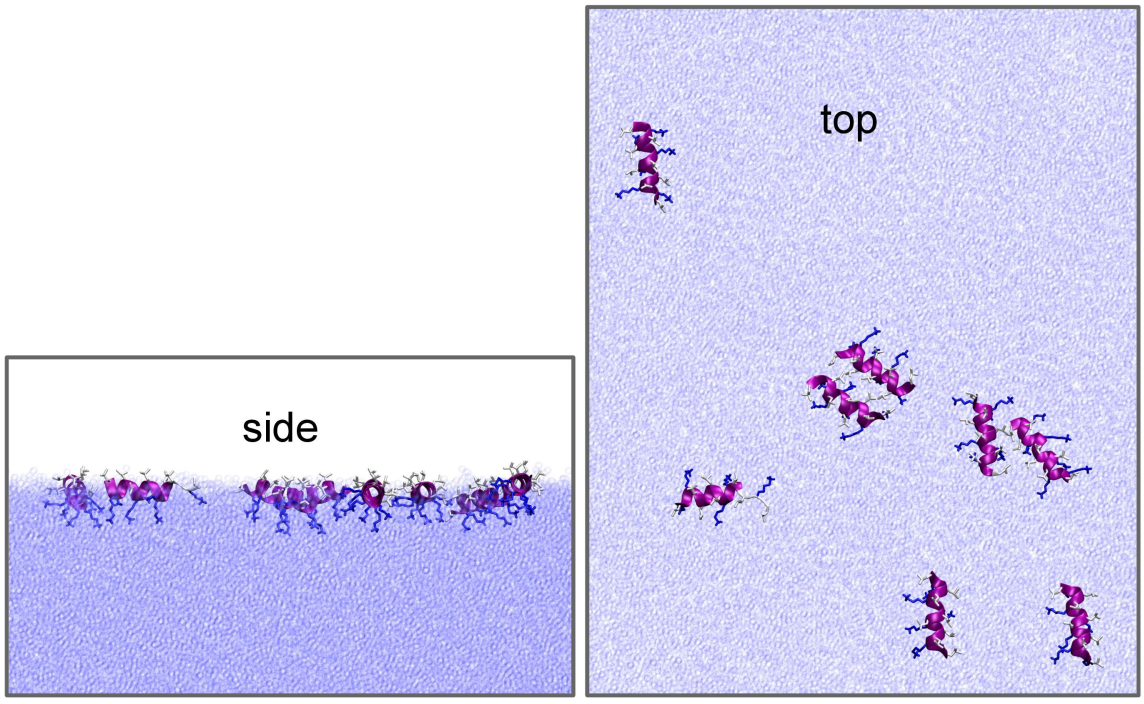


Figure 4.38: The final conformations for eight LKα14 peptides at the vacuum/water interface. All the peptides remain α-helix and are positioned at the vacuum/water interface with hydrophobic and hydrophilic residues partitioned as discussed before.

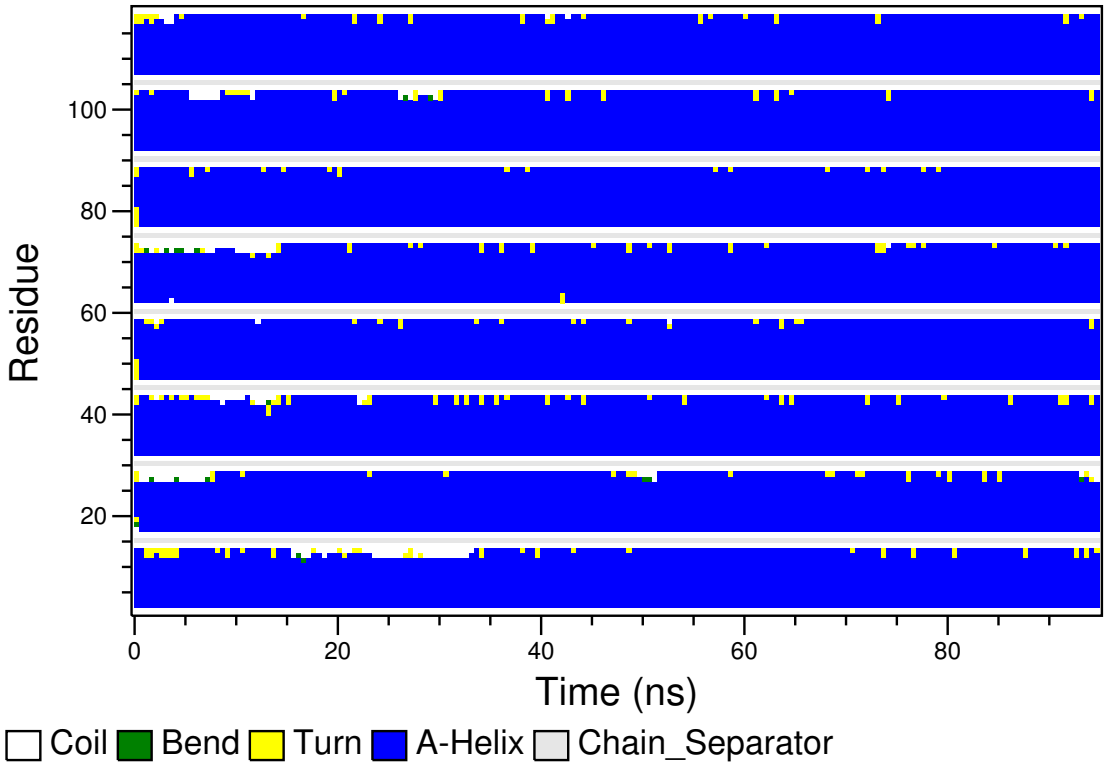


Figure 4.39: Time evolution of the secondary structure for eight LKα14 peptides simulated at the vacuum/water interface. Except intermittent unfolding of the end residues, α-helix structure is stable for all peptides.

This section analyzed the aggregation properties of LKα14 in bulk water and at the vacuum/water interface. Unlike the single LKα14 in bulk water, here, two LKα14 peptides adopted stable full α-helices. The driving force for the aggregation was the hydrophobic core formed by the leucine sidechains that shielded hydrophobic regions from the water molecules. Therefore, the peptides act as molecular interfaces that induce conformations similar to the hydrophilic/hydrophobic interfaces. Larger size aggregates of trimer, tetramer and pentamer were also stable with each LKα14 adopting a full α-helix. Among the aggregates of dimer, trimer, tetramer and pentamer the four helix bundle, i.e. the tetramer was found to be the most stable one via umbrella pulling simulations. However the vacuum/water interface decreased the stability of the aggregates replacing the hydrophobic core with the partitioning of the hydrophobic and hydrophilic residues. Two LKα14 peptides at the vacuum/water interface formed a weakly interacting aggregate that were positioned at such a distance that allows hydrophobic interaction while at the same time minimizes the electrostatic repulsion between the charged lysine sidechains.

Systems that exhibit conformational transition and aggregation phenomena require a large system and/or a long amount of time. For example, it took more than a month to simulate a single LKα14 in bulk water for 1 μs using 24 threads and the simulations of four helix bundle in bulk water lasted more than two months using 40 threads. The computational requirements for simulating larger size systems increases exponentially. Therefore, to study such large system, the LKα14 can be coarse-grained to reduce the number of degrees of freedom. To this end, next section discusses the preliminary results for coarse-graining the LKα14 peptide.

4.5 Coarse-grain Model of the LKα14 Peptide

In this section the preliminary results for the coarse-grain (CG) model of the LKα14 peptide will be introduced. The CG model is able to reproduce the helical nature observed in the atomistic tetramer simulations. The hydrophobic/hydrophilic interface was mimicked by a wall that favors the superatoms corresponding to the leucine sidechains and disfavors the remaining supertoms.

Mapping Scheme The atomistic LKα14 molecule was mapped to the CG model where the details are presented in Table 4.3. The leucine sidechains were represented with one superatom, **L**, the lysine sidechains were represented by two beads **KC** and **KN** where the former contains the $CH_2CH_2CH_2$ and the latter includes the CH_2-NH_3 atoms. **CA** superatom is chosen as the C_α atom and the remaining main chain *C-O-N-H* atoms are represented with **CN** superatom. All of the superatoms are placed at the center of mass of the atom groups they represent.

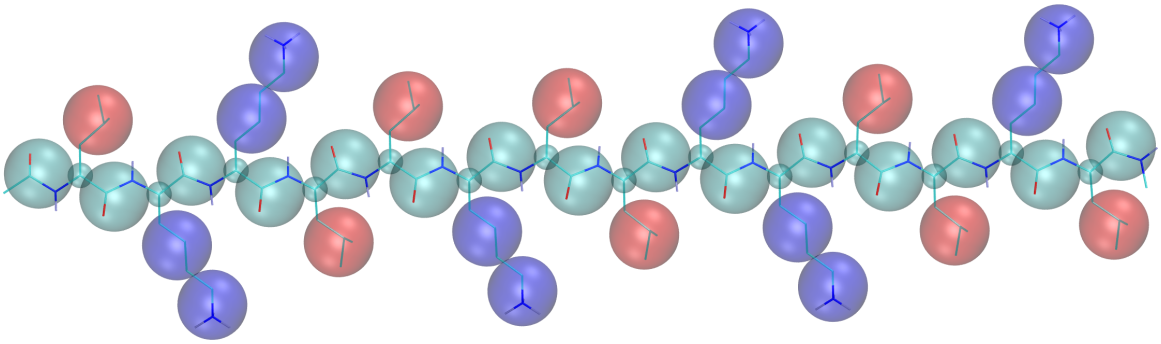


Figure 4.40: The CG mapping scheme for the atomistic LKα14 peptide. The coloring of superatoms are as follows: **L** are red, **KC** and **KC** are blue, **CA** and **CN** are cyan. The radius of the superatoms were tweaked to contain their atomistic counterparts and are arbitrary.

Bonded Interaction Potentials A single peptide was simulated in implicit water with all the non-bonded interactions removed to obtain the contribution of bonded distributions minus the contributions from the non-bonded interactions. The resulting

beads	mass	atoms
CN	43.025	C O N H
CA	13.019	C_α
L	57.116	C_β C_γ $C_{\delta 1}$ $C_{\delta 2}$
KC	42.081	C_β C_γ C_δ
KN	42.081	C_η NH_3

Table 4.3: Definition of the beads of the CG model of LKα14 and their masses

distributions will be compared with the distributions of the atomistic tetramer in bulk water. The bonded interaction distributions are listed in Figures 4.41-4.44. Based on these distributions the bonded interaction potentials were constructed.

Except the *CA-CN* and the *KC-KN* bonds (Fig. 4.41), the remaining bond-stretching distributions yield close curves. The difference in *KC-KN* is expected due to the relatively long sidechain and therefore this difference will not be addressed. However the more important distribution that hints the difference between the α-helix in the tetramer and the extended conformation obtained in implicit water is the *CA-CN* bond which forms the backbone of the peptide. One should keep in mind that additional parameters also influence the *CA-CN* as will be discussed.

Considering the backbone-backbone angles, there is a slight shift in *CA-CN-CA* angular distributions. The contrast between the helix and the extended is visible in the *CN-CA-CN* angles with its broader peak having a long tail. Both of the sidechain-backbone angles of *CN-CA-L* and *CN-CA-KC* display a broader peaks due to the extended behavior in implicit water as expected and the position of the major peaks are slightly shifted. However the reverse of these angles are almost identical for the two atomistic simulations. The difference in the *CA-KC-KN* angle distribution is small which concerns the lysine sidechain.

Fig. 4.43 compares the distributions for the proper dihedral angles of the tetramer (red curve) and the implicit water simulations (green curve). The backbone dihedral

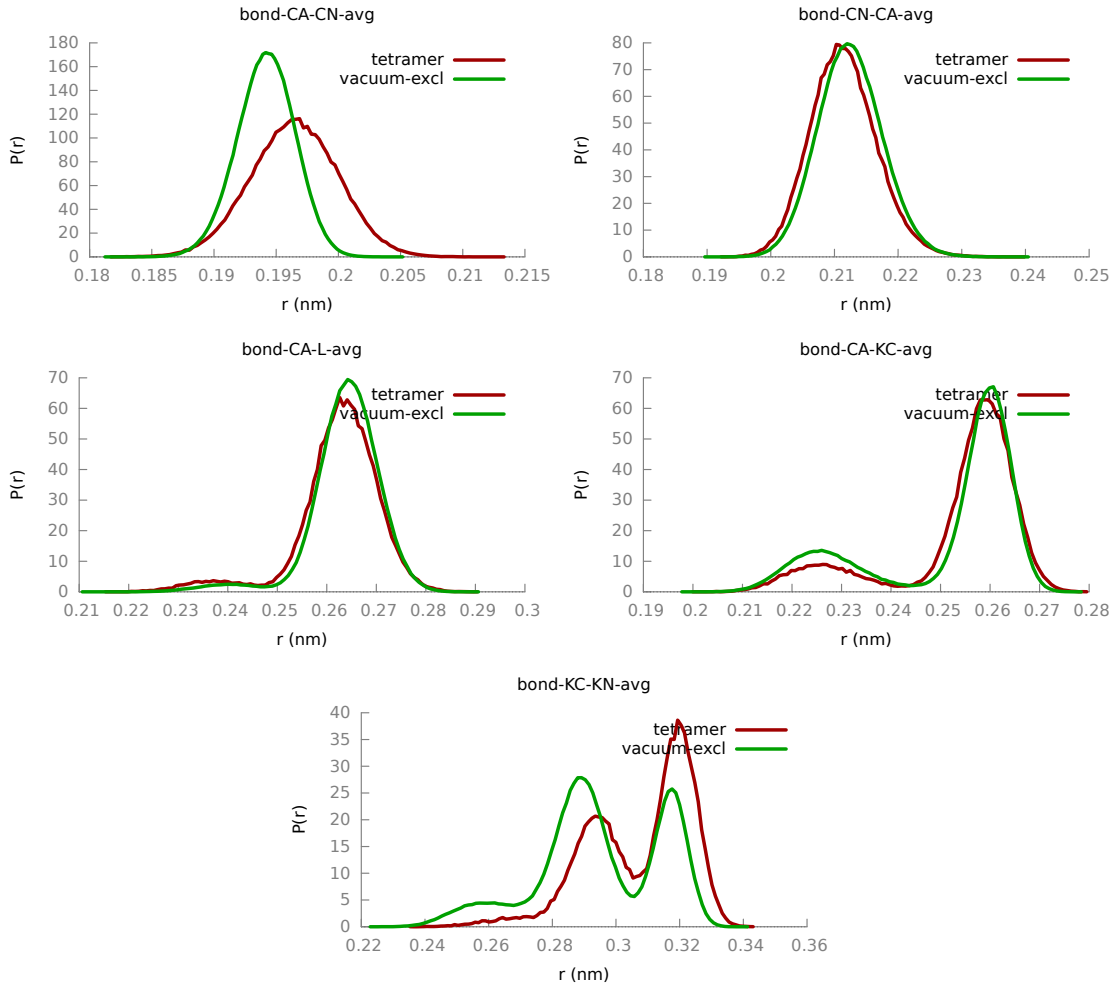


Figure 4.41: The coarse-grain mapped bond-stretching distributions for the tetramer (red) and the single peptide in implicit water with the non-bonded interactions excluded (labeled as *vacuum-excl* with green).

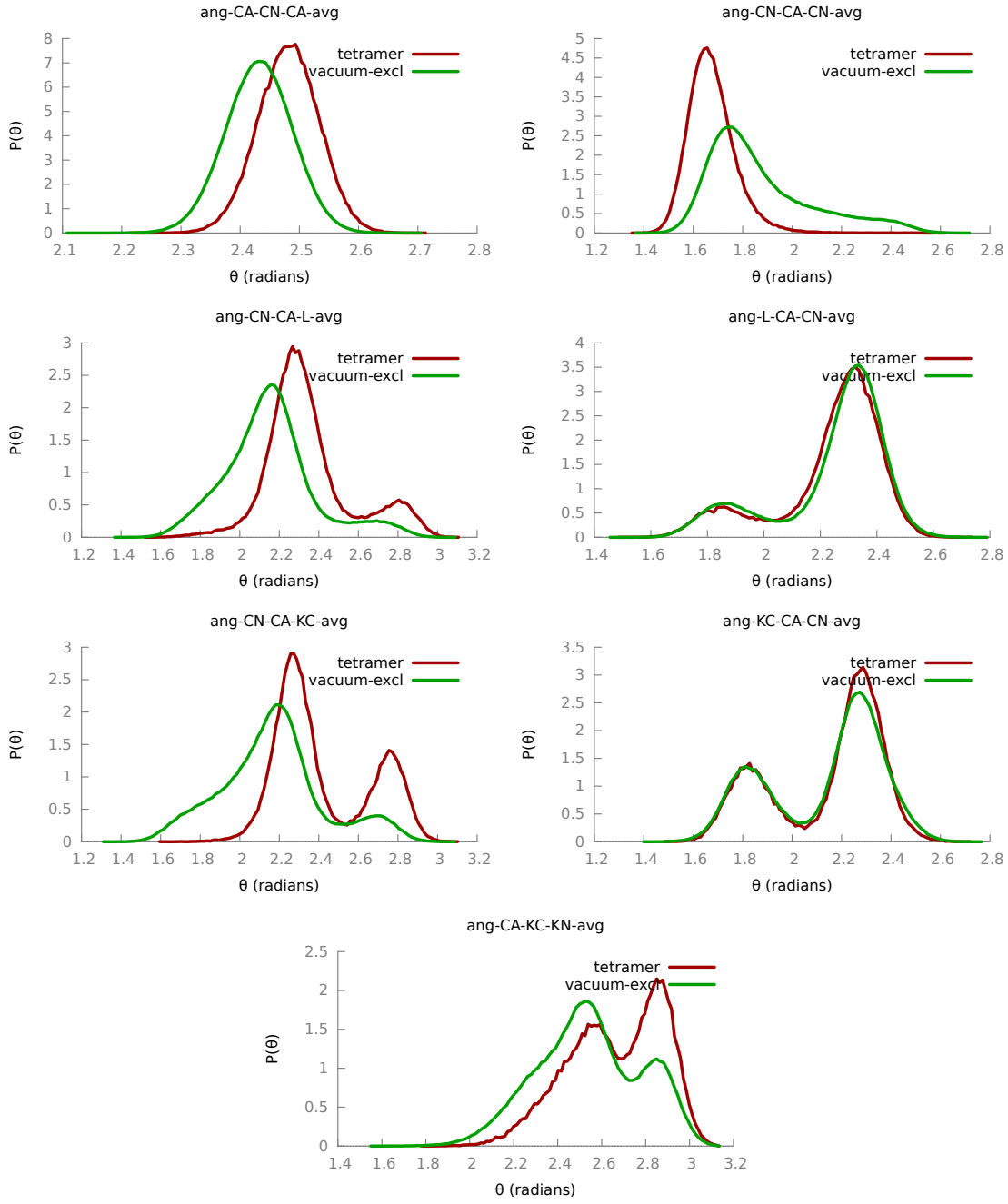


Figure 4.42: The CG mapped angle-bending distributions for the tetramer (red) and a single peptide in implicit water with non-bonded interactions excluded (green labeled as *vacuum-excl*).

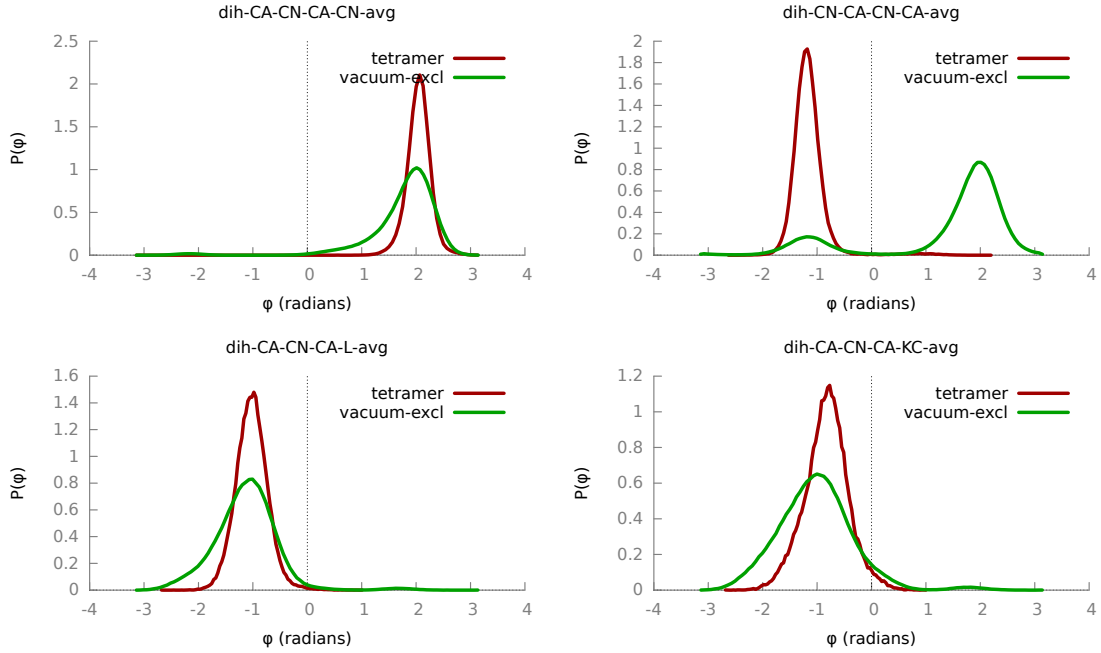


Figure 4.43: The coarse-grain mapped dihedral distributions for the tetramer and the single peptide in vacuum with non-bonded interactions excluded.

distributions for *CA-CN-CA-CN* and *CN-CA-CN-CA* reveal the difference between the helical conformation in the tetramer and the extended one in implicit water. To this end, our main aim will be to reproduce the transition in distribution between the dihedral angles and then, if necessary, tune the bond-stretching and angle-bending interactions. Similar curves obtained by both simulations for the improper dihedral angle distributions (Fig. 4.44) allow us to focus on matching the dihedral angles instead.

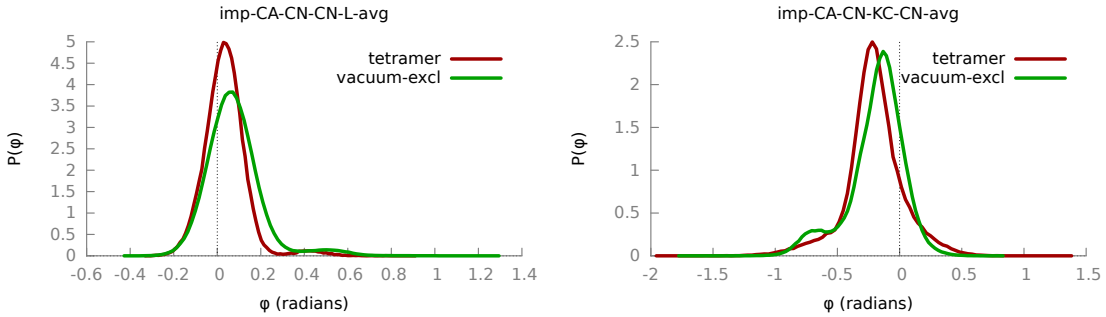


Figure 4.44: The coarse-grain mapped improper dihedral distributions for the tetramer (red) and a single peptide in implicit water with non-bonded interactions excluded (green labeled as *vacuum-excl*).

Non-bonded Interaction Potentials The backbone dihedral angle (*CA-CN-CA-CN* and *CN-CA-CN-CA*) and the distance between the CN_i-CN_{i+4} atoms contain the difference regarding the transition between the random structure adopted in bulk water and the α-helix assumed at the hydrophilic/hydrophobic interfaces. These distributions obtained from the atomistic tetramer and the implicit water simulations are shown in Figure 4.45. The tetramer adopts a *CN-CA-CN-CA* dihedral angle at -65° and a small peak at -155° is also visible whereas the distribution in implicit water yields a major peak at -155° and a minor one at -65°, i.e. the values pertaining to the peaks are identical but the height of the distributions are reversed. The distribution of the distance between CN_i-CN_{i+4} atoms in the tetramer has a major peak at 0.6 nm which is the characteristic i, i+4 hydrogen bonds between the backbone N-H and C-O groups in the α-helix. For the CG model to imitate the atomistic α-helix conformation then the CN_i-CN_{i+4} distances should be brought to 0.6 nm. To sum up, the switch between the random structure and α-helix is visible in the distance distributions of CN_i-CN_{i+4} and the *CN-CA-CN-CA* dihedral distributions and using these distributions as a guide we will aim to reproduce the atomistic tetramer distributions in our CG model.

The bonds in the CG model were defined as harmonic springs and their parameters can be found in Table 4.4. Fig. 4.46 plots the tabulated angle potentials that are determined by Boltzmann-inverting the angular distributions obtained from the atomistic LKα14 in implicit water without any defined nonbonded interactions. In terms of dihedral interactions, only the proper backbone dihedral angles are defined and represented by tabulated potentials as shown in Fig. 4.47. No potential interaction is defined for the proper dihedral angles corresponding to the sidechains, instead these are determined via implicit dihedral potentials which are plotted in Fig. 4.48.

When the CG model is simulated in implicit water, the distribution of the *CA-CN-CA-L* dihedral angle yielded a minor peak at 120° whereas such a dihedral angle was not observed in atomistic simulations. To prevent the CG model sampling this unwanted region we assigned a dummy improper dihedral potential to this angle as

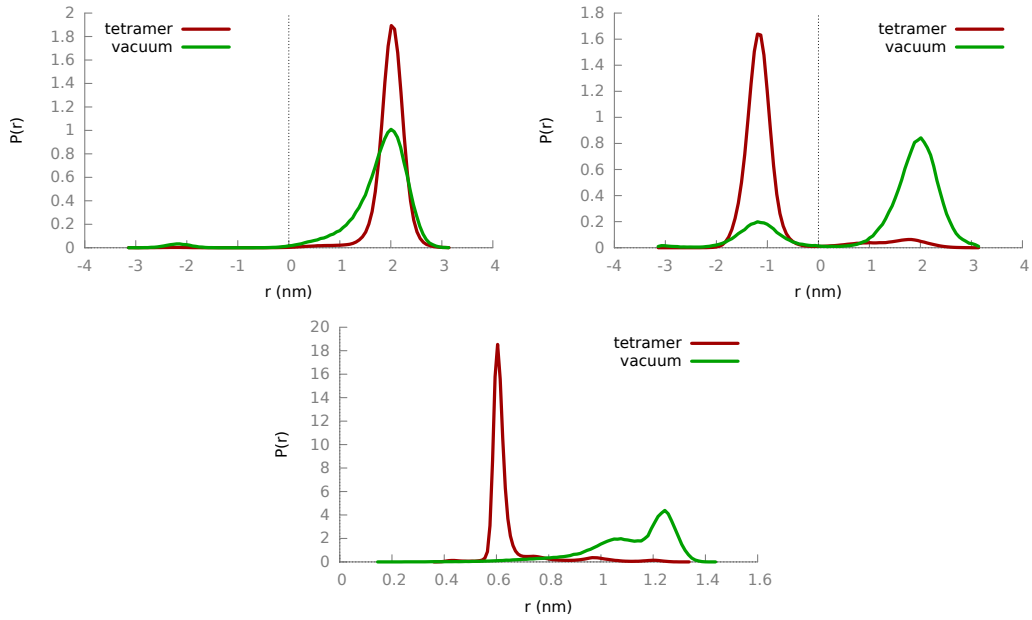


Figure 4.45: Distributions from tetramer and vacuum atomistic simulations for backbone dihedrals (top): CA-CN-CA-CN (left) and CN-CA-CN-CA (right) and CN_i-CN_{i+4} distance distribution (bottom). The difference between a random coil and α-helix is highlighted in CN-CA-CN-CA dihedral angle and the CN_i-CN_{i+4} distance distributions.

Bond	$_0$ (nm)	k_x (kJ/mol $\times nm^2$)
CN-CA	0.2112	95210
CA-CN	0.1952	457481
CA-L	0.263	62476.3
CA-KC	0.259	85700.8
KC-KN	0.3194	72640.4

Table 4.4: Equilibrium bond length and spring constant for the harmonic potentials connecting superatoms in the LKα14 CG model.

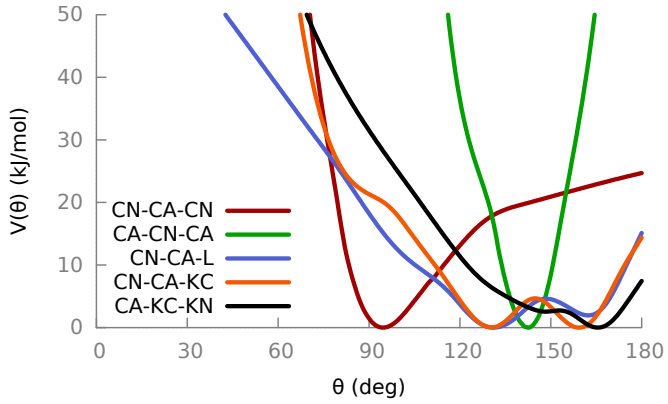


Figure 4.46: Tabulated potentials for angle interactions used in the LKα14 CG model obtained via Boltzmann-inverting the related angular distributions of the atomistic tetramer simulations in bulk water.

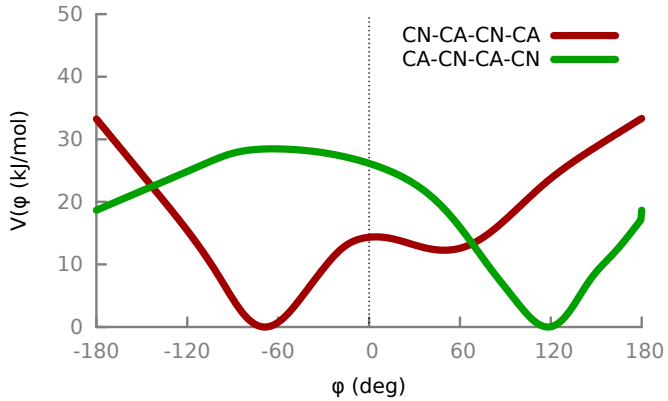


Figure 4.47: Tabulated potentials for the proper backbone dihedral interactions. These potentials are the Boltzmann-inverted forms of the corresponding distributions of atomistic tetramer simulations in bulk water.

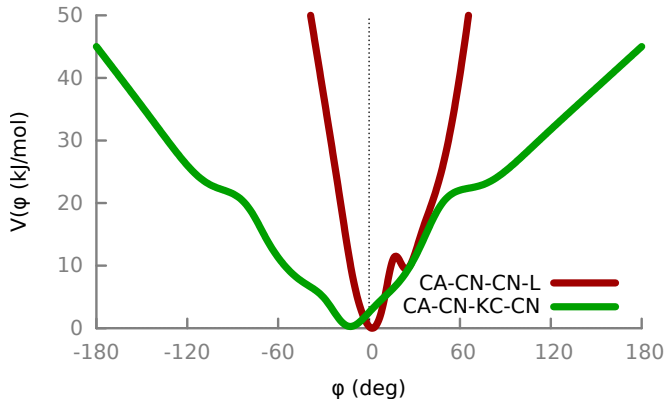


Figure 4.48: Tabulated potentials for implicit dihedral interactions.

displayed in Fig. 4.49 at the right. After applying this potential the modified CG model produces a correct distribution.

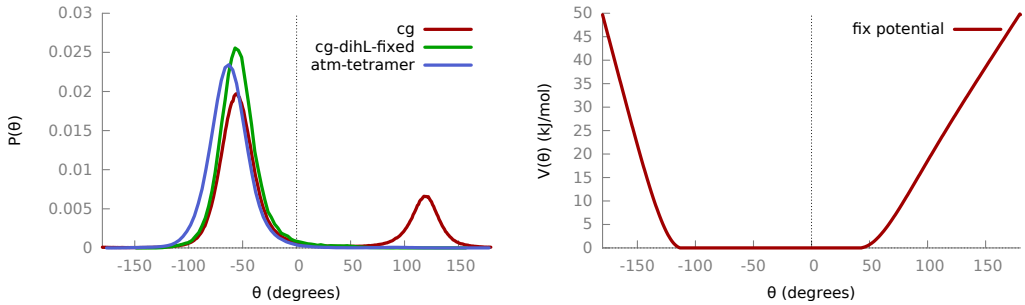


Figure 4.49: Dihedral distribution of CA-CN-CA-L for the CG without the dummy improper dihedral potential (red) and with it (green), compared to that of the atomistic tetramer. The improper fix potential is displayed at the right.

Using the interaction potentials shown above the CG model fails to reproduce the CN_i-CN_{i+4} distance distribution of the atomistic tetramer, i.e. the α-helix structure. Therefore one needs additional interactions such as pair potentials between the CN_i-CN_{i+4} beads. The pair potential is chosen to be the Boltzmann invert of the CN_i-CN_{i+4} distance distributions from the atomistic tetramer simulation where the LKα14 peptides adopted stable α-helix conformations. Fig. 4.50 presents the distributions obtained with different pair potentials. The pair interaction generates the correct

CN_i-CN_{i+4} distribution (green curve) compared to the CG model without any pair interactions (red curve). Reducing the depth of the pair potential had the same effect as smoothing the curve, i.e. they both increased the width of the distribution curve. Because the area of the potential well larger than 0.6 nm reduces probability of the peak value in CN_i-CN_{i+4} distribution by allowing higher distances favored by the CG model without the pair interactions. The CG model with the pair interaction potential (green curve in the left plot in Fig. 4.50) mimics the helical conformation.

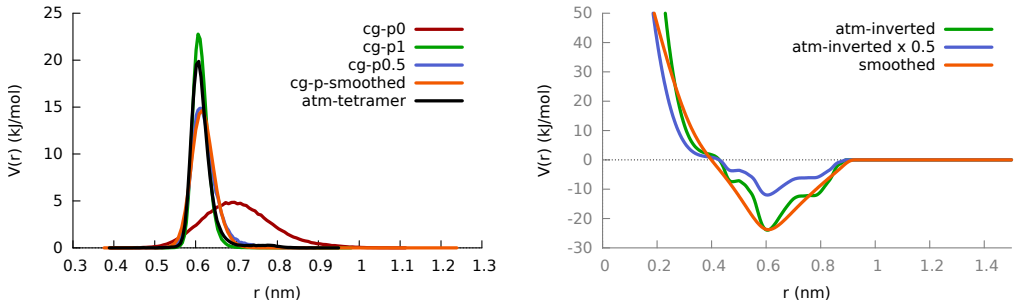


Figure 4.50: Different tabulated potentials used for the CN_i-CN_{i+4} interaction (right) and the corresponding CN_i-CN_{i+4} distance distributions (left). Without any defined pair interactions the CG model fails to reproduce the α-helix (*cg-p0*) and the Boltzmann inverted form of the CN_i-CN_{i+4} distance distribution for the atomistic tetramer reproduces the correct distribution of the atomistic tetramer in bulk water (*cg-p1*).

Up until now, we used the dihedral interaction potentials obtained from the atomistic tetramer distributions, although these potentials yield α-helix in bulk water, the distributions these potentials are derived from contain the contributions from other nonbonded interactions. Therefore when using these backbone dihedral potentials, we consider some of the nonbonded interactions twice: the first is with the dihedral potential, and the second with those interactions explicitly defined. On the contrary, deriving the dihedral potentials from the atomistic implicit water simulations with the excluded nonbonded interactions, yields potentials obtained only by the atoms considered. Furthermore, we also want our CG model to reproduce the random structure obtained in bulk water. To this end, we will examine the effects of using the backbone dihedral potentials obtained from the atomistic implicit water simulations.

The backbone dihedral distributions are compared for the CG models that employ dihedral potentials derived from the implicit water and the tetramer simulations (Figure 4.51). As expected, the former yields identical backbone distributions to the atomistic water simulations whereas the CN_i-CN_{i+4} distance is slightly shifted to the left.

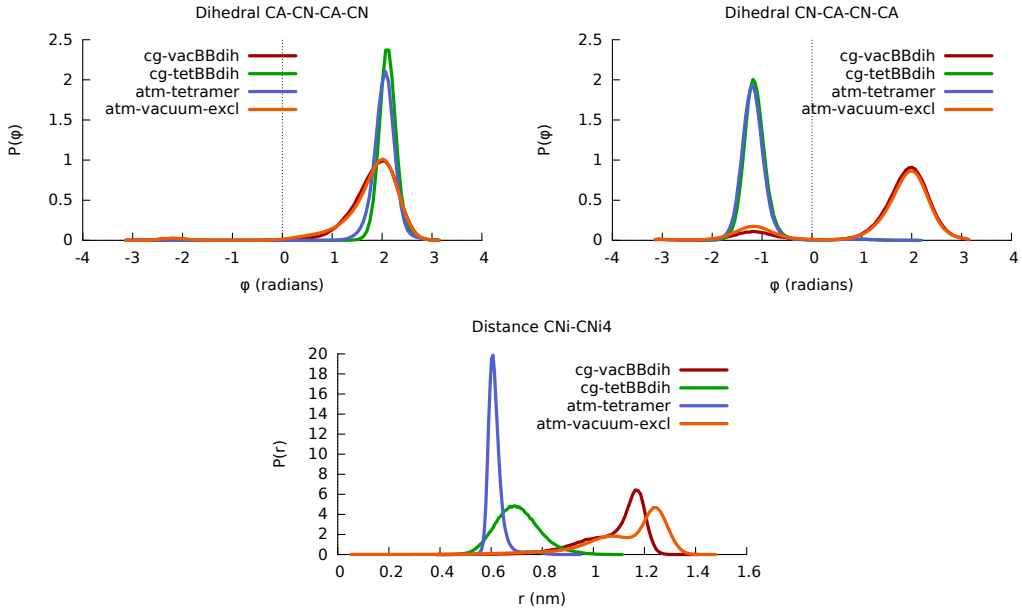


Figure 4.51: Distributions comparing the CG models with the tetramer derived backbone dihedrals and the implicit water derived ones. No pair potential between the CN_i-CN_{i+4} beads were used. The CG models display similar distributions to to those of their atomistic counterparts.

The CG models shown in Fig. 4.51 were defined without any CN_i-CN_{i+4} pair potentials. If one adds the CN_i-CN_{i+4} pair interaction potentials onto the CG model with the backbone dihedral potentials derived from the implicit water then one attains the exact CN_i-CN_{i+4} distance distributions as that of the atomistic tetramer but the backbone dihedral angles are not correctly reproduced and fall between the distributions of the two atomistic simulations (Fig. 4.52).

To mimic the hydrophilic/hydrophobic interface at the CG scale, one can use a wall that selectively interacts favorably with the L beads (imitating hydrophobic leucine

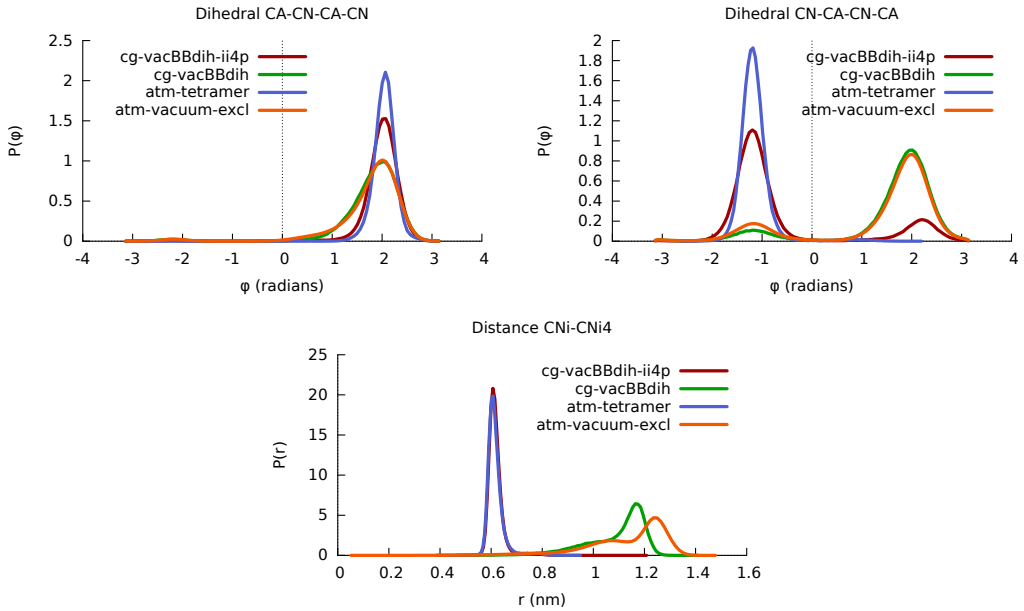


Figure 4.52: Backbone dihedral angle (top) and CN_i-CN_{i+4} distance distributions (bottom) comparing the CG models with and without the CN_i-CN_{i+4} pair potentials for the CG model with the backbone dihedral potentials derived from the atomistic implicit water simulations.

sidechains) and repels KC and KN beads (imitating hydrophilic lysine sidechains). The CG model should yield a helical structure that reproduces the atomistic tetramer CN_i-CN_{i+4} distance distribution. We already showed that the pair interaction potential in the CN_i-CN_{i+4} beads produces an α-helix. Therefore first we analyze the effect of the wall on the CG model without the pair potentials. Fig. 4.53 shows the distance distribution for the CG models with different wall-L potential interactions. The interaction between the L beads and the wall do not have any effect on the CN_i-CN_{i+4} interactions. Although the peptide adsorbs onto the wall, there is no change in secondary structure without the CN_i-CN_{i+4} interaction potentials. The alignment of the CG model at the wall is depicted in Fig. 4.54.

The inability of the wall to induce a change in secondary structure complicates the results. As can be seen in Fig. 4.55 the pair potential seems to be the only parameter that induces a stable helical conformation where the L-wall interaction potential be-

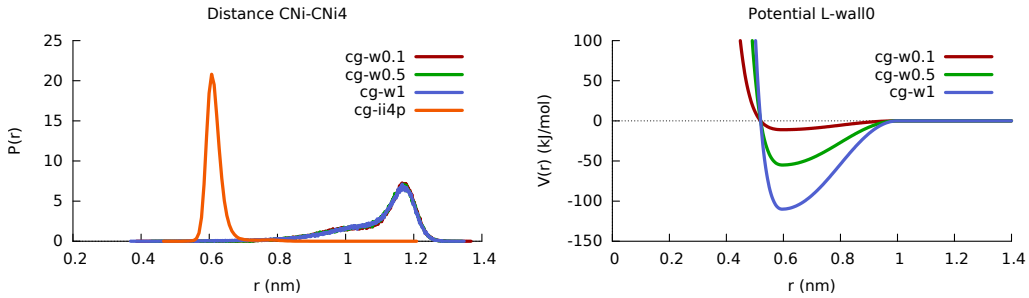


Figure 4.53: Distribution of the distance between CN_i-CN_{i+4} compared to different depths of wall potentials without the CN_i-CN_{i+4} defined (*cg-w0.1, cg-w.5, cg-21* where *cg-w0.1* means the wall potential obtained from that is a counterpart for the interaction for the atomistic *leucine-interface* is multiplied by 0.1). The CN_i-CN_{i+4} pair potential (*cg-ii4p*) is defined as the inverse of the atomistic tetramer distributions.

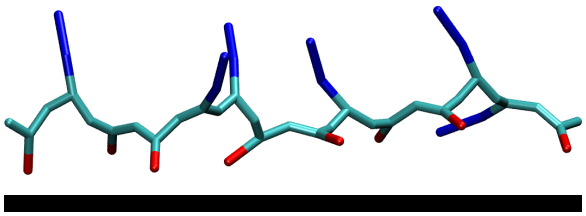


Figure 4.54: Snapshot of CG LKα14 on the wall where the wall is shown with a black line.

tween the was decreased to 1/10th of that of the previous simulation. The changes in L-wall potential do not effect the resulting CN_i-CN_{i+4} distance distributions. Furthermore, the radii of the interaction of the remaining beads with the wall was also found to play no part in triggering a change in CN_i-CN_{i+4} distance distributions (Fig. 4.56).

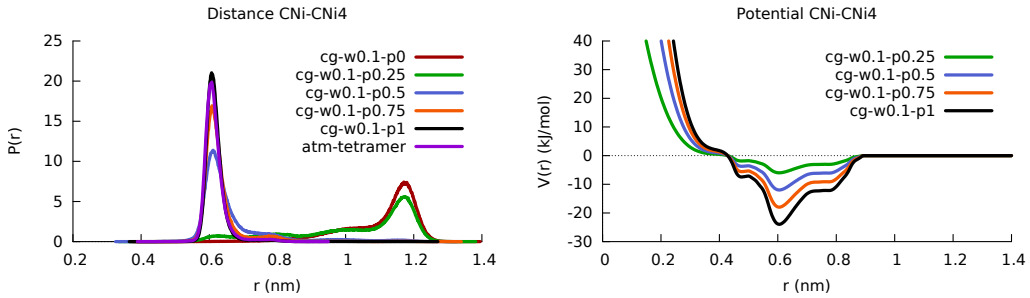


Figure 4.55: Histograms of the CN_i-CN_{i+4} distances (left) obtained by the corresponding tabulated potentials of used to represent the CN_i-CN_{i+4} pair interactions. Here the wall potential is multiplied by 0.1 and the CN_i-CN_{i+4} potentials are scaled.

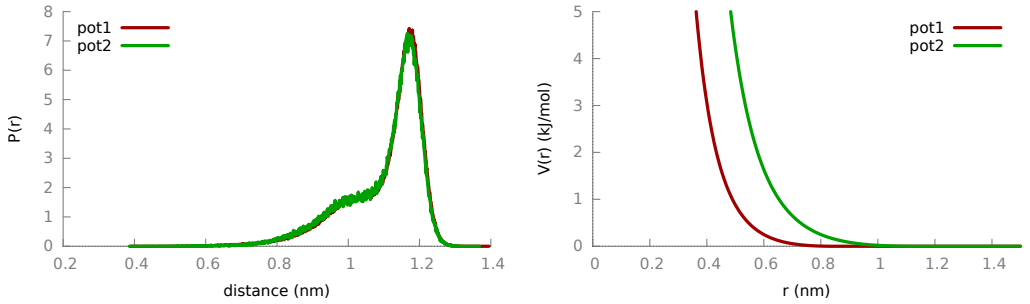


Figure 4.56: Tabulated interaction potentials of superatoms (right) with the wall except the L bead(i.e. CA-wall, CN-wall, KC-wall, KN-wall) and the distribution of their distances from the wall (left).

Preliminary results for a LKα14 CG model were presented where introducing the CN_i-CN_{i+4} pair potentials yielded a helical conformation. A wall was introduced which attracts L beads that are the CG counterparts of the atomistic hydrophobic leucine sidechains, that generates a surface mimicking a hydrophobic/hydrophilic interface. However, the interaction potential of the L beads was found to play no part in the CN_i-CN_{i+4} distance distributions, and therefore the helical structure.

4.6 Concluding Remarks

This study documents the different behaviors of LKα14 in solution and at hydrophilic/hydrophobic interfaces. Using 1 μs long MD simulations and replica-exchange MD simulations a single LKα14 peptide in solution was found to yield a variety of structures including α-helix, random-coil, β-sheet and in-between variants of these whereas at the vacuum/water interface the random structures of the peptide folded into a stable α-helix and the helices remained folded. The α-helix adopted at the vacuum/water interface was enabled by the partitioning of hydrophobic and hydrophilic residues, the former accumulating on the vacuum side and the lysine sidechains extended into water. The stability of this conformation was because of the dehydration of the hydrophobic leucine sidechains.

The aggregation of LKα14 in bulk water was investigated, where unlike the lack of an unique structure for a single peptide, two LKα14 peptides were shown to aggregate and form a stable α-helix. Therefore the added peptide behaved as a micro-interface that induced folding into α-helix with similar partitioning observed at the vacuum/water interface simulations, i.e. the hydrophobic leucine sidechains formed a hydrophobic core and the charged lysine sidechain faced the water molecules. The dimer was shown to have identical preferences for the anti-parallel and parallel orientations where MD simulations featuring the LKα14 dimer yielded transitions between these orientations.

Comparing the stability of the larger aggregate sizes of trimer, tetramer, pentamer and octamer, it was found that the α-helical conformation was stable with leucine residues concentrating in the inner region of the aggregate shielded from the water molecules. The free energy of removing a peptide from the aggregate was computed via umbrella sampling simulations for aggregates of dimer, trimer, tetramer and pentamer and the tetramer was found to have the deepest PMF curve suggesting the four helix bundle as the more stable aggregate among. The solvent accessible surface area of the hydrophobic leucine sidechains supported the findings.

The simulation of two LKα14 peptides yielded weakly interacting peptides with

larger inter-peptide distances, in contrast to the two peptides in bulk water. Furthermore, the umbrella simulations supported the free MD simulations with lower depth PMF curves for separating the peptides compared to the findings in bulk water environment. The disruption of the hydrophobic core was proposed as the culprit for this decreased inter-peptide interaction strength where at the interface the lysine sidechains were forced to be closer to each other if the LKα14 dimer adopted the bulk water conformation. Thus, the interplay between the charge repulsion of NH_3 groups and the attraction of the hydrophobic sidechains resulted in an unstable inter-peptide distance throughout the simulation.

Finally, the study showed a preliminary coarse-grain model of the LKα14 where the atomistic helical conformation was mimicked only by the pair potentials between the i,i+4 backbone atoms.

The entire study reports environment dependent folding of LKα14. The helical stability was allowed by preventing the exposure of the hydrophobic regions. Similarly, for the α-helix in the aggregates was made possible by the peptides acting as hydrophobic micro-surfaces. The main driving forces for conformations were found to be the electrostatic repulsion of lysine sidechains and the attraction between the leucine residues.

Chapter 5

CONCLUSION

Structure of a protein depends not only on the sequence but also on the environment. For example, hydrophilic/hydrophobic interfaces promote folding and aggregation exploiting the amphiphilic nature of proteins by partitioning the hydrophobic and hydrophilic residues. Studying these systems help us to understand the mechanisms and driving forces behind these phenomena by determining the effects of specific interactions that pariticipate in these processes. Here, molecular simulations are useful tools that enable monitoring the system at atomistic scale.

To this end we chose a short peptide that allows an easier modeling and simulation. We analyzed the environment-driven conformational transition for FF and showed that at the cyclohexane/water interface the molecule adopts a different conformation than the one in bulk water. Furthermore, we observed a similar conformational transition when FF formed aggregates where the neighboring peptides created a molecular interface mimicking the effects of a macromolecular one. However we also showed that even for such a simple peptide, observing these processes via atomistic MD simulations required a long amount of time making these studies harder for larger size systems.

Using the solvation free energy values and the radial distribution functions as a starting point we developed a two-state CG model that successfully mimics the bulk water and the interface behaviors. The success of the free energy approach indicated that the underlying forces between these transitions are thermodynamic. Furthermore, we showed that the solvent interactions between the solvent molecules that do not include the peptide interactions influence the final conformation of the peptide.

In the second study, a more complex amphiphilic LKα14 peptide which was designed to have a separated hydrophobic and hydrophilic residues when adopted a helical conformation was studied. Using microsecond long advanced sampling techniques we showed that in solution the peptide does not adopt a stable secondary structure. This intrinsically disordered behavior is alleviated when the interface restricts the conformations which allow hydrated hydrophobic residues and thus the peptide adopts a stable α-helix. This environment dependent behavior was also observed when the peptides self-assemble to form aggregates. The formation of the hydrophobic core between the peptides favored the accumulation of the peptides stabilising the helical secondary structure. Therefore it has been shown that the neighboring peptides act as molecular interfaces and thus similar to macromolecular interfaces, induce more ordered conformations.

Larger size aggregates of LKα14 were observed to be stable where each peptide adopted helical conformations. We measured the stability between the aggregates of different sizes and compared the PMFs for dissociating each bundle. Among the studied aggregates, the four helix bundle displayed the highest stability and which was confirmed by comparing the amount of dehydrated hydrophobic sidechain residues. Then we illustrated the importance of i,i+4 hydrogen bonds using the CG model of LKα14 where the α-helical conformation was made possible only via pair potential interactions between the relevant beads.

In this thesis we have analyzed the complex interactions behind the environment dependent behaviors of amphiphilic peptides and the difficulty of observing the folding and aggregation processes via atomistic resolution simulations. We achieved a free energy based two-state CG model that is transferable between two distinct environments and we highlighted the significance of thermodynamic forces behind the structural changes.

BIBLIOGRAPHY

[1] Alexei V Finkelstein and Oleg Ptitsyn. *Protein physics: a course of lectures*. Academic Press, 2002.

[2] David Whitford. *Proteins: structure and function*. John Wiley & Sons, 2005.

[3] Ken A Dill, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. The protein folding problem. *Annual review of biophysics*, 37:289, 2008.

[4] R Cerpa, F E Cohen, and I D Kuntz. Conformational switching in designed peptides: the helix/sheet transition. *Folding & design*, 1(2):91–101, January 1996.

[5] W. Kabsch and C. Sander. On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proceedings of the National Academy of Sciences*, 81(4):1075–1078, February 1984.

[6] D V Waterhous and W C Johnson. Importance of environment in determining secondary structure in proteins. *Biochemistry*, 33(8):2121–8, March 1994.

[7] L. Zhong and W C Johnson. Environment affects amino acid preference for secondary structure. *Proceedings of the National Academy of Sciences*, 89(10):4462–4465, May 1992.

[8] Manfred Mutter, R. Gassmann, U. Buttkus, and K. H Altmann. Switch Peptides: pHInduced αHelix to βSheet Transitions of Bisamphiphilic Oligopeptides. *Angewandte Chemie International Edition in English*, 30(11):1514–1516, November 1991.

[9] Manfred Mutter and René Hersperger. Peptides as Conformational Switch: MediumInduced Conformational Transitions of Designed Peptides. *Angewandte Chemie International Edition in English*, 29(2):185–187, February 1990.

[10] Chris Van Der Walle. *Peptide and protein delivery*. Academic Press, 2011.

[11] Jeffrey J Gray. The interaction of proteins with solid surfaces. *Curr. Opin. Struct. Biol.*, 14(1):110–5, February 2004.

[12] Su-Jin Kang, Hyung-Sik Won, Wahn-Soo Choi, and Bong-Jin Lee. De novo generation of antimicrobial lk peptides with a single tryptophan at the critical amphipathic interface. *Journal of Peptide Science*, 15(9):583–588, 2009.

[13] Manfred Mutter, Arunan Chandravarkar, Christine Boyat, John Lopez, Sonia Dos Santos, Bhubaneswar Mandal, Richard Mimna, Karine Murat, Luc Patiny, Lydiane Saucède, and Gabriele Tuchscherer. Switch Peptides In Statu Nascendi: Induction of Conformational Transitions Relevant to Degenerative Diseases. *Angewandte Chemie International Edition*, 43(32):4172–4178, August 2004.

[14] Sonia Dos Santos, Arunan Chandravarkar, Bhubaneswar Mandal, Richard Mimna, Karine Murat, Lydiane Saucède, Patricia Tella, Gabriele Tuchscherer, and Manfred Mutter. Switch-Peptides: Controlling Self-Assembly of Amyloid β-Derived Peptides in vitro by Consecutive Triggering of Acyl Migrations. *J. Am. Chem. Soc.*, 127(34):11888–11889, 2005.

[15] Per Hammarström, R. Luke Wiseman, Evan T Powers, and Jeffery W Kelly. Prevention of Transthyretin Amyloid Disease by Changing Protein Misfolding Energetics. *Science*, 299(5607):713–716, January 2003.

[16] Soto Claudio. Plaque busters: strategies to inhibit amyloid formation in Alzheimer's disease. *Molecular Medicine Today*, 5(8):343–350, August 1999.

[17] Biotek Website. Biotek Website Folding-Misfolding-Aggregation.

[18] Magnus Gustafsson, Guy Vandenbussche, Tore Curstedt, Jean-Marie Ruysschaert, and Jan Johansson. The 21-residue surfactant peptide (lysleu¡ sub¿ 4¡/sub¿)¡ sub¿ 4¡/sub¿ lys (kl¡ sub¿ 4¡/sub¿) is a transmembrane α-helix with a mixed nonpolar/polar surface. *FEBS letters*, 384(2):185–188, 1996.

[19] William F DeGrado, Holly Gratkowski, and James D Lear. How do helixhelix interactions help determine the folds of membrane proteins? Perspectives from the study of homooligomeric helical bundles. *Protein Science*, 12(4):647–665, April 2003.

[20] Martin B Ulmschneider, Jacques PF Doux, J Antoinette Killian, Jeremy C Smith, and Jakob P Ulmschneider. Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides. *Journal of the American Chemical Society*, 132(10):3452–3460, 2010.

[21] Aimee L Boyle and Derek N Woolfson. De novo designed peptides for biological applications. *Chemical Society Reviews*, 40(8):4295–4306, 2011.

[22] Fabrizio Chiti and Christopher M Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333–366, 2006.

[23] Christopher M Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, 2003.

[24] Hanna Rapaport, Haviv Grisaru, and Tova Silberstein. Hydrogel scaffolds of amphiphilic and acidic β-sheet peptides. *Advanced Functional Materials*, 18(19):2889–2896, 2008.

[25] Robert Fairman and Karin S Åkerfeldt. Peptides as novel smart materials. *Current opinion in structural biology*, 15(4):453–463, 2005.

[26] Karthikan Rajagopal and Joel P Schneider. Self-assembling peptides and proteins for nanotechnological applications. *Current opinion in structural biology*, 14(4):480–486, 2004.

[27] Hanna Rapaport. Ordered peptide assemblies at interfaces. *Supramolecular Chemistry*, 18(5):445–454, 2006.

[28] Teresa Head-Gordon and Scott Brown. Minimalist models for protein folding and design. *Current Opinion in Structural Biology*, 13(2):160–167, April 2003.

[29] Alessandra Villa, Christine Peter, and N.F.A. van Der Vegt. Self-assembling dipeptides : conformational sampling in solvent-free coarse-grained simulation. *Phys. Chem. Chem. Phys.*, 11:2077–2086, 2009.

[30] Peter L Freddolino, Christopher B Harrison, Yanxin Liu, and Klaus Schulten. Challenges in protein-folding simulations. *Nature physics*, 6(10):751–758, 2010.

[31] John C Shelley, Mee Y Shelley, Robert C Reeder, Sanjoy Bandyopadhyay, and Michael L Klein. A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B*, pages 4464–4470, 2001.

[32] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica*, 49(2-3):61–74, February 1998.

[33] Alessandra Villa, N.F.A. van Der Vegt, and Christine Peter. Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model. *Phys. Chem. Chem. Phys.*, 11(12):2068–2076, February 2009.

[34] Helmholtz Institute Ulm Website. Helmholtz Institute Ulm Multiscale Modelling.

[35] Florian Müller-Plathe Dirk Reith, Mathias Pütz. Deriving Effective Mesosclae Potentials from Atomistic Simulations. *J. Comput. Chem.*, 24(13):32, 2002.

[36] Teemu Murtola, Emma Falck, Michael Patra, Mikko Karttunen, and Ilpo Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer. *The Journal of chemical physics*, 121(18):9156–65, November 2004.

[37] Furio Ercolessi and James B Adams. Interatomic potentials from first-principles calculations: the force-matching method. *EPL (Europhysics Letters)*, 26(8):583, 1994.

[38] Joseph F. Rudzinski and W. G. Noid. Coarse-graining entropy, forces, and structures. *Journal of Chemical Physics*, 135(21):214101, December 2011.

[39] WG Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal of chemical physics*, 139(9):090901, 2013.

[40] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert J Marrink. The MARTINI Coarse-Grained Force Field : Extension to Proteins. *J. Chem. Theory Comput.*, 4:819–834, 2008.

[41] W. Shinoda, R. DeVane, and M. L. Klein. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, 33(1):27–36, January 2007.

[42] Clara D Christ, Alan E Mark, and W. F. van Gunsteren. Basic Ingredients of Free Energy Calculations : A Review. *Journal of Computational Chemistry*, 2009.

[43] Christophe Chipot and David A. Pearlman. Free Energy Calculations. The Long and Winding Gilded Road. *Molecular Simulation*, 28(1-2):1–12, 2002.

[44] a.K. Soper. Empirical potential Monte Carlo simulation of fluid structure. *Chemical Physics*, 202(2-3):295–306, January 1996.

[45] Victor Rühle and Christoph Junghans. Hybrid approaches to coarse-graining using the votca package: Liquid hexane. *Macromolecular Theory and Simulations*, 20(7):472–477, 2011.

[46] Ozge Engin, Alessandra Villa, Christine Peter, and Mehmet Sayar. A Challenge for Peptide Coarse Graining: Transferability of Fragment-Based Models. *Macromolecular Theory and Simulations*, 20(7):451–465, August 2011.

[47] Pritam Ganguly, Debashish Mukherji, Christoph Junghans, and Nico FA van der Vegt. Kirkwood–buff coarse-grained force fields for aqueous solutions. *Journal of Chemical Theory and Computation*, 8(5):1802–1807, 2012.

[48] Pritam Ganguly and Nico FA van der Vegt. Representability and transferability of kirkwood–buff iterative boltzmann inversion models for multicomponent aqueous systems. *Journal of Chemical Theory and Computation*, 9(12):5247–5256, 2013.

[49] Motoo Fukuda, Hedong Zhang, Takahiro Ishiguro, Kenji Fukuzawa, and Shintaro Itoh. Structure-based coarse-graining for inhomogeneous liquid polymer systems. *The Journal of chemical physics*, 139(5):054901, 2013.

[50] Cahit Dalgicdir, Ozge Sensoy, Christine Peter, and Mehmet Sayar. A transferable coarse-grained model for diphenylalanine: how to represent an environment driven conformational transition. *J. Chem. Phys.*, 139(23):234115, December 2013.

[51] Henry D Herce and Angel E Garcia. Molecular dynamics simulations suggest a mechanism for translocation of the hiv-1 tat peptide across lipid membranes. *Proceedings of the National Academy of Sciences*, 104(52):20805–20810, 2007.

[52] Leonor Saiz and Michael L Klein. Computer simulation studies of model biological membranes. *Accounts of chemical research*, 35(6):482–489, 2002.

[53] Leonor Saiz and Michael L Klein. Structural properties of a highly polyunsaturated lipid bilayer from molecular dynamics simulations. *Biophysical journal*, 81(1):204–216, 2001.

[54] KV Damodaran and Kenneth M Merz Jr. A comparison of dmpc-and dlpe-based lipid bilayers. *Biophysical journal*, 66(4):1076–1087, 1994.

[55] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, page btt055, 2013.

[56] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, September 1997.

[57] Jacob N Israelachvili. *Intermolecular and surface forces: revised third edition*. Academic press, 2011.

[58] Andrew R. Leach. *Molecular modelling: principles and applications*. Pearson Education, 2001.

[59] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, January 2007.

[60] H J C Berendsen, J. P. M. Postma, and W. F. van Gunsteren. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[61] Justin A. Lemkul and David R. Bevan. Assessing the Stability of Alzheimers Amyloid Protofibrils Using Molecular Dynamics. *J. Phys. Chem. B*, 114(4):1652–1660, 2010.

[62] Andrew S Paluch, David L Mobley, and Edward J Maginn. Small molecule solvation free energy: Enhanced conformational sampling using expanded ensemble molecular dynamics simulation. *Journal of Chemical Theory and Computation*, 7(9):2910–2918, 2011.

[63] Carl Caleman, Jochen S Hub, Paul J van Maaren, and David van der Spoel. Atomistic simulation of ion solvation in water explains surface preference of halides. *Proceedings of the National Academy of Sciences*, 108(17):6838–6842, 2011.

[64] Justin L MacCallum and D Peter Tieleman. Calculation of the water-cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field. *J. Comp. Chem.*, 24(15):1930–5, November 2003.

[65] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *The journal of physical chemistry. B*, 114(32):10235–53, August 2010.

[66] Charles H Bennett. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *Journal of Computational Physics*, (22):245–268, 1976.

[67] M Mezei. The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water. *J. Chem. Phys.*, 86:7084, 1987.

[68] Christophe Chipot and Andrew Pohorille. *Free energy calculations: theory and applications in chemistry and biology*. Springer, February 2007.

[69] D. Frenkel. Free-energy calculations. 1991.

[70] WK Den Otter. Thermodynamic integration of the free energy along a reaction coordinate in cartesian coordinates. *The Journal of Chemical Physics*, 112(17):7283–7292, 2000.

[71] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Academic press, 2001.

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

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

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

[75] Jochen S Hub, Bert L de Groot, and David van der Spoel. g_wham a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation*, 6(12):3713–3720, 2010.

[76] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1):275–282, 1995.

[77] M Chernick. Bootstrap methods: A guide for practitioners and researchers. *Wiley, New York*, 2007.

[78] Bradley Efron and Robert Tibshirani. *An Introduction to the Bootstrap*, volume 57. Springer Science, 1993.

[79] Simulating the dynamics of proteins to understand protein functions - Frontlines - RIKEN RESEARCH.

[80] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(12):141–151, November 1999.

[81] Daniel J Sindhikara, Daniel J Emerson, and Adrian E Roitberg. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.*, 6(9):2804–2808, September 2010.

[82] M Gustafsson, G Vandenbussche, T Curstedt, JM Ruysschaert, and J Johansson. The 21-residue surfactant peptide (LysLeu(4))(4)Lys(KL(4)) is a transmembrane alpha-helix with a mixed nonpolar/polar surface. *Febs Lett*, 384(2):185–188, 1996.

[83] WF DeGrado, H Gratkowski, and JD Lear. How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci.*, 12(4):647–665, 2003.

[84] Martin B. Ulmschneider, Jacques P F Doux, J Antoinette Killian, Jeremy C. Smith, and Jakob P. Ulmschneider. Mechanism and Kinetics of Peptide Partitioning into Membranes from All-Atom Simulations of Thermostable Peptides. *J. Am. Chem. Soc.*, 132(10):3452–3460, 2010.

[85] C. M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884 – 890, 2003.

[86] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333 – 366, 2006.

[87] C. Wu and J.-E. Shea. Coarse-grained models for protein aggregation. *Curr. Opin. Struct. Biol.*, 21:209–220, 2011.

[88] C Schladitz, E P Vieira, H Hermel, and H Möhwald. Amyloid-beta-sheet formation at the air-water interface. *Biophys. J.*, 77(6):3305–3310, 1999.

[89] H. Rapaport. Ordered peptide assemblies at interfaces. *Supramol. Chem.*, 18(5):445 – 454, 2006.

[90] R. Fairman and K. S. Akerfeldt. Peptides as novel smart materials. *Curr. Opin. Struct. Biol.*, 15(4):453–463, 2005.

[91] K. Rajagopal and J. P. Schneider. Self-assembling peptides and proteins for nanotechnological applications. *Curr. Opin. Struct. Biol.*, 14(4):480–486, 2004.

[92] Hanna Rapaport, Haviv Grisaru, and Tova Silberstein. Hydrogel Scaffolds of Amphiphilic and Acidic beta-Sheet Peptides. *Adv. Funct. Mater.*, 18(19):2889–2896, 2008.

[93] WJ Li, F Nicol, and FC Szoka. GALA: a designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery. *Adv. Drug Deliver. Rev.*, 56(7):967–985, 2004.

[94] Monica C Branco, Dina M Sigano, and Joel P Schneider. Materials from peptide assembly: towards the treatment of cancer and transmittable disease. *Curr. Opin. Chem. Biol.*, 15(3):427–434, 2011.

[95] Aimee L Boyle and Derek N Woolfson. De novo designed peptides for biological applications. *Chem. Soc. Rev.*, 40(8):4295, 2011.

[96] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein. A coarse grain model for n-alkanes parameterized from surface tension data. *J. Chem. Phys.*, 119:7043–7049, 2003.

[97] S. J. Marrink, A. H. deVries, and A. E. Mark. Coarse grained model for semi-quantitative lipid simulations. *J. Phys. Chem. B*, 108:750–760, 2004.

[98] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111:7812–7824, 2007.

[99] B. M. Mognetti, L. Yelash, P. Virnau, W. Paul, K. Binder, M. Mueller, and L. G. Macdowell. Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide. *J. Chem. Phys.*, 128:104501, 2008.

[100] R. DeVane, W. Shinoda, P. B. Moore, and M. L. Klein. Transferable Coarse Grain Nonbonded Interaction Model for Amino Acids. *J. Chem. Theory Comput.*, 5:2115–2124, 2009.

[101] W. Tschöp, K. Kremer, J. Batoulis, T. Burger, and O. Hahn. Simulation of polymer melts. i. coarse-graining procedure for polycarbonates. *Acta Polym.*, 49(2-3):61 – 74, 1998.

[102] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial-distribution functions - a reverse Monte-Carlo approach. *Phys. Rev. E*, 52:3730 – 3737, 1995.

[103] F. Müller-Plathe. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *Chem. Phys. Chem.*, 3:754 – 769, 2002.

[104] D. Reith, M. Putz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J. Comp. Chem.*, 24:1624 – 1636, 2003.

[105] C. Peter, L. Delle Site, and K. Kremer. Classical simulations from the atomistic to the mesoscale: coarse graining an azobenzene liquid crystal. *Soft Matter*, 4:859–869, 2008.

[106] T. Murtola, M. Karttunen, and I. Vattulainen. Systematic coarse graining from structure using internal states: Application to phospholipid/cholesterol bilayer. *J. Chem. Phys.*, 131:055101, 2009.

[107] A. Lyubartsev, A. Mirzoev, L. J. Chen, and A. Laaksonen. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss*, 144:43–56, 2010.

[108] A. Savelyev and G. A. Papoian. Molecular renormalization group coarse-graining of electrolyte solutions: application to aqueous NaCl and KCl. *J. Phys. Chem. B*, 113:7785–7793, 2009.

[109] G. Megariotis, An Vyrkou, A. Leygue, and D. N. Theodorou. Systematic Coarse Graining of 4-Cyano-4 '-pentylbiphenyl. *Ind. Eng. Chem. Res.*, 50:546–556, 2011.

[110] B. Mukherje, Delle Site L., Kremer K., and C. Peter. Derivation of a Coarse Grained model for Multiscale Simulation of Liquid Crystalline Phase Transitions. *J. Phys. Chem B*, 116(29):8474–8484, 2012.

[111] S. Izvekov and G. A. Voth. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, 109:2469–2473, 2005.

[112] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, 128:244114, 2008.

[113] M. S. Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.*, 129:144108, 2008.

[114] Margaret E Johnson, Teresa Head-Gordon, and Ard A Louis. Representability problems for coarse-grained water potentials. *J. Chem Phys.*, 126:144509, 2007.

[115] J. R. Silbermann, S. H. L. Klapp, M. Schoen, N. Chennamsetty, H. Bock, and K. E. Gubbins. Mesoscale modeling of complex binary fluid mixtures: Towards an atomistic foundation of effective potentials. *J. Chem. Phys.*, 124:074105, 2006.

[116] E. C. Allen and G. C. Rutledge. Evaluating the transferability of coarse-grained, density-dependent implicit solvent models to mixtures and chains. *J. Chem. Phys.*, 130:034904, 2009.

[117] J. W. Mullinax and W. G. Noid. Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.*, 131:104110, 2009.

[118] A. Villa, C. Peter, and N. F. A. van der Vegt. Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water. *J. Chem. Theory Comput.*, 6:2434–2444, 2010.

[119] S. Izvekov, P. W. Chung, and B. M. Rice. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.*, 133:064109, 2010.

[120] J.-W. Shen, C. Li, N. F. A. van der Vegt, and C. Peter. Transferability of Coarse Grained Potentials: Implicit Solvent Models for Hydrated Ions. *J. Chem. Theory Comput.*, 7:1916–1927, 2011.

[121] E. Brini, V. Marcon, and N. F. A. van der Vegt. Conditional reversible work method for molecular coarse graining applications. *Phys. Chem. Chem. Phys.*, 13:10468–10474, 2011.

[122] Carl Henrik Görbitz. Nanotube Formation by Hydrophobic Dipeptides. *Chemistry - A European Journal*, 7(23):5153–5159, December 2001.

[123] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der

Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[124] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and W. F. van Gunsteren. 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*, 25(13):1656–76, October 2004.

[125] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, November 1987.

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

[127] Ulrich Essmann, Lalith Perera, ML Berkowitz, and T. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):31–34, 1995.

[128] Victor Rühle and Christoph Junghans. Hybrid approaches to coarse-graining using the votca package: Liquid hexane. *Macromolecular Theory and Simulations*, 20(7):472–477, 2011.

[129] E Apol, R Apostolov, HJC Berendsen, A Van Buuren, P Bjelkmar, R Van Drunen, A Feenstra, G Groenhof, P Kasson, P Larsson, et al. Gromacs user manual version 4.5.4. *Royal Institute of Technology and Uppsala University: Stockholm and Uppsala, Sweden*, 2010.

[130] W Humphrey, A Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–38, 27–28, February 1996.

[131] Di Wu and David a Kofke. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *The Journal of chemical physics*, 123(8):084109, August 2005.

[132] W. F. van Gunsteren and H. J. C. Berendsen. A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation*, 1(3):173–185, 1988.

[133] Glenn M Torrie and John P Valleau. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.*, 28(4):578–581, 1974.

[134] Ozge Engin, Alessandra Villa, Mehmet Sayar, and Berk Hess. Driving forces for adsorption of amphiphilic peptides to the air-water interface. *The journal of physical chemistry. B*, 114(34):11093–101, September 2010.

[135] Joohyun Jeon, Carolyn E Mills, and M Scott Shell. Molecular insights into diphenylalanine nanotube assembly: all-atom simulations of oligomerization. *J. Phys. Chem. B*, 117(15):3935–43, April 2013.

[136] V. Tozzini. Minimalist models for proteins: a comparative analysis. *Q. Rev. Biophys.*, 43:333–371, 2010.

[137] O. Bezkorovaynaya, A. Lukyanov, K. Kremer, and C. Peter. Multiscale simulation of small peptides: Consistent conformational sampling in atomistic and coarse-grained models. *J. Comp. Chem*, 33:937, 2012.

[138] C. Peter and K. Kremer. Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back. *Soft Matter*, 5:4357–4366, 2009.

[139] O. Engin, A. Villa, C. Peter, and M. Sayar. A Challenge for Peptide Coarse Graining: Transferability of Fragment-Based Models. *Macromol. Theory Simul.*, 20:451–465, 2011.

[140] John G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.*, 3(5):300, 1935.

[141] N. V. Buchete, J. E. Straub, and D. Thirumalai. Development of novel statistical potentials for protein fold recognition. *Curr. Opin. Struct. Biol.*, 14(2):225 – 232, 2004.

[142] Cong Guo, Yin Luo, Ruhong Zhou, and Guanghong Wei. Probing the self-assembly mechanism of diphenylalanine-based peptide nanovesicles and nanotubes. *ACS Nano*, 6(5):3907–3918, 2012.

[143] Sabine Castano, Bernard Desbat, and Jean Dufourcq. Ideally amphipathic β-sheeted peptides at interfaces: structure, orientation, affinities for lipids and hemolytic activity of (kl)¡ sub¿ m¡/sub¿ k peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1463(1):65–80, 2000.

[144] Sylvie E Blondelle, John M Ostresh, Richard A Houghten, and Enrique Perez-Paya. Induced conformational states of amphipathic peptides in aqueous/lipid environments. *Biophysical journal*, 68(1):351–359, 1995.

[145] Laure Beven, Sabine Castano, Jean Dufourcq, Åke Wieslander, and Henri Wroblewski. The antibiotic activity of cationic linear amphipathic peptides: lessons from the action of leucine/lysine copolymers on bacteria of the class mollicutes. *European Journal of Biochemistry*, 270(10):2207–2217, 2003.

[146] Margitta Dathe, Michael Schümann, Torsten Wieprecht, Anett Winkler, Michael Beyermann, Eberhard Krause, Katsumi Matsuzaki, Osamu Murase, and Michael Bienert. Peptide helicity and membrane surface charge modulate the balance of electrostatic and hydrophobic interactions with lipid bilayers and biological membranes. *Biochemistry*, 35(38):12612–12622, 1996.

[147] Sabine Castano, Isabelle Cornut, Klaus Büttner, JL Dasseux, and Jean Dufourcq. The amphipathic helix concept: length effects on ideally amphipathic likj (i=2j) peptides to acquire optimal hemolytic activity. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1416(1):161–175, 1999.

[148] W. F. DeGrado and J. D. Lear. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.*, 107(25):7684–7689, 1985.

[149] Huayu Xiong, Brian L Buckwalter, Hong-Ming Shieh, and Michael H Hecht. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proceedings of the National Academy of Sciences*, 92(14):6349–6353, 1995.

[150] Sabine Castano, Bernard Desbat, Michel Laguerre, and Jean Dufourcq. Structure, orientation and affinity for interfaces and lipids of ideally amphipathic lytic l¡ sub¿ i¡/sub¿ k¡ sub¿ j¡/sub¿(¡ i¿ i¡/i¿= 2¡ i¿ j¡/i¿) peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1416(1):176–194, 1999.

[151] Tobias Weidner, Julia S Apte, Lara J Gamble, and David G Castner. Probing the orientation and conformation of α-helix and β-strand model peptides on self-assembled monolayers using sum frequency generation and nexafs spectroscopy. *Langmuir*, 26(5):3433–3440, 2009.

[152] Tobias Weidner, Nicholas F Breen, Kun Li, Gary P Drobny, and David G Castner. Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *Proc. Natl. Acad. Sci. U. S. A.*, 107(30):13288–93, July 2010.

[153] Diana C Phillips, Roger L York, Ozzy Mermut, Keith R McCrea, Robert S Ward, and Gabor A Somorjai. Side chain, chain length, and sequence effects on

amphiphilic peptide adsorption at hydrophobic and hydrophilic surfaces studied by sum-frequency generation vibrational spectroscopy and quartz crystal microbalance. *The Journal of Physical Chemistry C*, 111(1):255–261, 2007.

[154] Tobias Weidner, Nicholas F Breen, Kun Li, Gary P Drobny, and David G Castner. Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30):13288–93, July 2010.

[155] Julia S Apte, Lara J Gamble, David G Castner, and Charles T Campbell. Kinetics of leucine-lysine peptide adsorption and desorption at -CH3 and -COOH terminated alkylthiolate monolayers. *Biointerphases*, 5(4):97–104, December 2010.

[156] Julia S Apte, Galen Collier, Robert A Latour, Lara J Gamble, and David G Castner. XPS and ToF-SIMS investigation of alpha-helical and beta-strand peptide adsorption onto SAMs. *Langmuir*, 26(5):3423–32, March 2010.

[157] Nicholas F Breen, Tobias Weidner, Kun Li, David G Castner, and Gary P Drobny. A solid-state deuterium NMR and sum-frequency generation study of the side-chain dynamics of peptides adsorbed onto surfaces. *J. Am. Chem. Soc.*, 131(40):14148–9, October 2009.

[158] Michael Deighan and Jim Pfaendtner. Exhaustively sampling peptide adsorption with metadynamics. *Langmuir*, 29(25):7999–8009, June 2013.

[159] Galen Collier, Nadeem A Vellore, Jeremy A Yancey, Steven J Stuart, and Robert A Latour. Comparison between empirical protein force fields for the simulation of the adsorption behavior of structured LK peptides on functionalized surfaces. *Biointerphases*, 7(1-4):24, December 2012.

[160] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[161] Nathan Schmid, Andreas P Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E Mark, and Wilfred F van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal : EBJ*, 40(7):843–56, July 2011.

[162] Alexandra Patriksson and David van der Spoel. A temperature predictor for parallel tempering simulations. *Physical chemistry chemical physics : PCCP*, 10(15):2073–7, April 2008.

[163] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[164] John G Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3:300, 1935.

[165] Glenn M Torrie and John P Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters*, 28(4):578–581, 1974.

[166] W Humphrey, A Dalke, and K Schulten. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 27–28, February 1996.

[167] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4):566–579, 1995.

[168] Raymond A Jarvis and Edward A Patrick. Clustering using a similarity measure based on shared near neighbors. *Computers, IEEE Transactions on*, 100(11):1025–1034, 1973.

[169] Benjamin A Shoemaker, John J Portman, and Peter G Wolynes. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences*, 97(16):8868–8873, 2000.

VITA

Cahit Dalgıçdir was born on January, 11th 1984 in Istanbul. He received his undergraduate and masters degree from Sabanci University from Materials Science and Engineering program.