Amyloidogenicity and lipid interaction of antimicrobial anuran peptides

Submitted in partial fulfilment of the requirements

of the degree of

Doctor of Philosophy

of the
Indian Institute of Technology, Bombay, India
and
Monash University, Australia

by

Sourav Ray

Supervisors:

Prof. Ajay S. Panwar (IIT Bombay)

Prof. Lisandra L. Martin (Monash University)
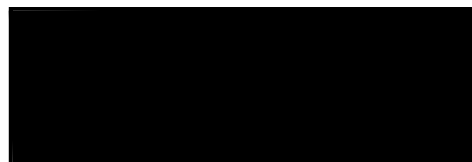
(Year 2019)

Thesis approval

This thesis entitled **"Amyloidogenicity and lipid interaction of antimicrobial anuran peptides"** by **Sourav Ray** is approved for the degree of **Doctor of Philosophy**.

(Prof. Sanjib Senapati)
External Examiner

(Prof. Ranjith Padinhateeri)
Internal Examiner

IITB Supervisor

(Prof. Lisandra L. Martin)
Monash Supervisor

(Prof. Jhumpa Adhikari)
Chairperson

Date: 25-July-2019
Place: IIT Bombay

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Notice 1

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Notice 2

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Student Name: Sourav Ray

IITB ID: 144114004

Monash ID: 26414686

Date: 25-July-2019

Place: IIT Bombay

Abstract

Amyloidogenic and antimicrobial peptides demonstrate preferential toxicity towards eukaryotic and prokaryotic cells, respectively. Whereas amyloidogenic peptides have varying sizes and predominantly consist of hydrophobic residues, the antimicrobial peptides usually contain less than 50 amino acid residues and have a balanced distribution of cationic and hydrophobic residues. However, both classes of peptides are amphipathic and manifest their cytotoxicity using similar modes of membrane action. Thus, relatively smaller antimicrobial peptides, such as 17 amino acid residues long uperin 3.*x* peptides, can serve as model peptides to study the effect of electrolyte concentration and cationic to hydrophobic residue substitution on the propensity of antimicrobial peptides to form amyloid-like fibrils.

Both coarse-grained and fully-atomistic simulations were employed to determine that sodium chloride salt promotes uperin 3.5 aggregation by screening electrostatic repulsion and facilitating increased α-helical content in the early stages of peptide aggregation. All uperin 3.*x* peptides and corresponding variants interacted more favourably with bacterial membrane-mimic, anionic sodium dodecyl sulphate micelles; than with eukaryotic membrane-mimic, zwitterionic dodecyl phosphatidylcholine micelles. Moreover, the alanine variants demonstrated relatively weaker interaction with both types of micelles, compared to the wild-type peptides. Also, peptide segments predominantly consisting of hydrophobic residues were found to be crucial in the interaction of uperin 3.*x* peptides and their variants with these bilayer analogues. Furthermore, fully-atomistic simulations were performed to elucidate the effect of seventh-position substitution of arginine or lysine residue with alanine on these uperin 3.*x* peptides. It was observed that peptide aggregates solely composed of coil and turn secondary structures demonstrated a relatively rapid β-aggregation than those containing helical signatures. In general, alanine substitution at seventh-position increases the β-aggregation propensity of the uperin 3.*x* wild-type peptides. However, the initial structural composition of the peptide conformers and the three-dimensional inter-peptide orientation at the beginning would also play a crucial role.

Contents

Chapter 1

Chapter 2

Chapter 3

Mechanistic insight into the early stages of amyloid formation using an anuran peptide **61**

Chapter 4

Chapter 5

Propensity of Uperin 3.*x* **peptides towards beta-aggregation: Analysis of structurally-diverse conformers** **141**

List of Figures

List of Tables

Abbreviations

Aβ	Amyloid-beta
AI	Antimicrobial Index
AMP	Antimicrobial Peptide
atm	atmosphere
CF	Conformer Family
CG	Coarse-Grained
CHARMM	Chemistry at Harvard Macromolecular Mechanics
Colvars	Collective variables
DLVO	Derjaguin-Landau-Verwey-Overbeek
DPC	Dodecyl Phosphatidylcholine
DSSP	Define Secondary Structure of Proteins
EM	Energy Minimisation
FA	Fully-Atomistic
HB	Hydrophobic
HD	Hex Docking
hIAPP	human Islet Amyloid Polypeptide
HL	Hydrophilic
IC_{50}	Half-maximal Inhibitory Concentration
LAC	Loosely Associated Cluster
LJ	Lennard-Jones
MD	Molecular Dynamics

MP	Matrix Propagation
NaCl	Sodium chloride
NAMD	Nanoscale Molecular Dynamics
NMR	Nuclear Magnetic Resonance
OPLS	Optimized Potential for Liquid Simulations
PMF	Potential of Mean Force
RDF	Radial Distribution Function
R_g	Radius of gyration
$RMSD$	Root-Mean-Square Deviation
SASA	Solvent-Accessible Surface Area
SDS	Sodium Dodecyl Sulphate
STRIDE	Structural Identification
TFE	2,2,2-Trifluoroethanol
TIP3P	Transferable Intermolecular Potential with 3 Points
U	Uperin
vdW	van der Waals
VMD	Visual Molecular Dynamics
Wt	Wild-type

Chapter 1

Introduction

1.1. Background

Amyloidogenesis involves transition of certain peptides or proteins from soluble functional states into highly organized fibrillar aggregates, via protofibrillar intermediates[1]. The 'amyloid hypothesis' implicates peptide or protein aggregates for the cytotoxic 'degenerative' events associated with neurodegeneration and other misfolding diseases (Table 1.1)[2-3]. Although protofibrils or diffuse amyloids are soluble intermediates without fibrillar substructure, they serve as precursors to mature amyloid fibrils; and are linked to the cytotoxicity observed in amyloid diseases[2, 4-6]. Especially, in neurological disorders involving the central nervous system, these oligomeric forms cause the pathogenicity, primarily due to their small size allowing faster diffusion and greater exposure of hydrophobic amino acid sidechains on their surface[7].

An increasing number of antimicrobial peptides (AMPs) have been identified with an inherent ability to form amyloid structures[8]. For example, indolicidin, temporin L, magainin 2, and plantaricin A are all AMPs that can form amyloid-like fibrils, and consist of 13, 13, 23, and 48 amino acid residues, respectively[9]. Furthermore, the mechanism of cytotoxic action of AMPs and, amyloidogenic peptides and proteins, have considerable similarities[8-9]. Therefore, relatively small AMPs, such as 17 residues long uperin 3.*x* peptides (naturally obtained from *Uperoleia mjobergii*), can serve as model peptides for understanding the process of amyloid formation[10]. The uperin (U) 3.5 wild-type (wt) is unusual, as it does not aggregate in pure water, but self-aggregates in NaCl buffer at neutral pH to form amyloid-like fibrils[8]. Electrolytes can

play a crucial role in the stabilization of peptide secondary structure and fibril formation by screening the charges on the residues, and thus reducing the surface tension[11-12]. Moreover, monovalent cations (Na$^+$, K$^+$) facilitate intra-peptide hydrogen bonds and turn structures, and hence promote peptide self-assembly[12-13].

Table 1.1. Examples of peptides or proteins that form extracellular amyloid deposits or intracellular inclusions with amyloid-like features in human diseases[7].

Peptide or protein name	Number of residues[a]	Structure[b]	Associated disease
Amyloid-β peptide	40 or 42[c]	Intrinsically disordered	Alzheimer disease
α-Synuclein[d]	140	Intrinsically disordered	Parkinson disease
Prion protein	208	Intrinsically disordered (1–102) + All-α, prion-like (103–208)	Creutzfeldt-Jakob disease, Kuru
Huntingtin exon 1[d]	~103–187[c]	Intrinsically disordered	Huntington disease
β2-microglobulin	99	All-β, Ig-like	Dialysis-related amyloidosis
Islet amyloid polypeptide	37	Intrinsically disordered	Type II diabetes
Insulin	(30 + 21)[e]	All-α, insulin-like	Injection-localized amyloidosis

[a]Lengths of the processed forms depositing into the aggregates, not the precursor proteins.

[b]Structural class and fold of the native, processed protein or peptide prior to aggregation, according to the Structural Classification of Proteins database.

[c]Fragments of various lengths were reported for *ex vivo* fibrils.

[d]Intracellular proteins, unlike the others that are extracellular.

[e]Lengths of the A and B chains linked by a disulphide bridge.

Uperin 3.*x* peptides are known to have antimicrobial activity against a host of Gram-positive bacteria[10]. Compared to U3.6 wt, seventh-position alanine variant U3.6 K7A exhibits drastically reduced antibiotic activity, as observed against a host of Gram-positive and Gram-negative bacterial cultures[14]. The membrane action of U3.5 wt and U3.5 R7A has been investigated[15], and it was observed that compared to U3.5 wt, membrane action was greatly reduced or almost inhibited when U3.5 R7A peptide solution was introduced to a membrane bilayer. Furthermore, molecular dynamics (MD) simulations of melittin and MG-H2 AMPs

indicate that the presence of charged residues at the N-termini region plays a crucial role in their membrane action[16]. In general, AMPs exert their cytotoxic activity by acting on the bacterial cell membranes, with membrane disruption and pore formation being the most common modes of action[17].

A lot of AMPs demonstrate a fibril-forming potential[8-9]. However, the formation of such amyloid structures would have different kinetic pathways for the diverse amyloidogenic peptides[7]. For instance, the investigation of U3.5 wt and U3.5 R7A revealed that U3.5 wt has lesser aggregation and β-sheet formation propensity, compared to U3.5 R7A[15]. The seventh-position substitution of a positively-charged amino acid with a nonpolar alanine residue has been found to increase the fibril content for the U3.5 R7A peptide. Substitution of isoleucine with arginine (I254R) in a p53$_{252–258}$ fragment adversely affected the propensity to form β-sheet structures[18]. Similarly, the introduction of charge also reduced the fibril formation of α-synuclein[19]. Furthermore, substitution of a lysine residue with alanine (K28A) increased the intra-peptide hydrophobic interactions in a Aβ$_{21–30}$ segment[20]. Significantly, most pathogenic mutations, common in hereditary amyloid diseases, disrupt the native structure of the protein and make it prone to aggregation[1, 7].

Table 1.2. Amino acid sequences of uperin 3.*x* wild-type peptides[10] and corresponding seventh-position alanine variants.

	Peptide	1	2	3	4	5	6	**7**	8	9	10	11	12	13	14	15	16	17	C-Ter
1.	U3.4 wt	G	V	G	D	L	I	**R**	K	A	V	A	A	I	K	N	I	V	-NH$_2$
2.	U3.4 R7A	G	V	G	D	L	I	**A**	K	A	V	A	A	I	K	N	I	V	-NH$_2$
3.	U3.5 wt	G	V	G	D	L	I	**R**	K	A	V	S	V	I	K	N	I	V	-NH$_2$
4.	U3.5 R7A	G	V	G	D	L	I	**A**	K	A	V	S	V	I	K	N	I	V	-NH$_2$
5.	U3.6 wt	G	V	I	D	A	A	**K**	K	V	V	N	V	L	K	N	L	F	-NH$_2$
6.	U3.6 K7A	G	V	I	D	A	A	**A**	K	V	V	N	V	L	K	N	L	F	-NH$_2$

In this thesis, we study the peptides from the uperin 3.*x* family and their corresponding alanine variants. The primary objective is to understand the effects of electrolyte concentration and cationic to hydrophobic residue substitution on the initial stages of aggregation, such as, the evolution of secondary structure. Moreover, the interaction of uperin 3.*x* peptides and variants with bilayer analogues is also investigated. The peptide sequences of U3.4 wt, U3.5 wt, and U3.6 wt, and their respective seventh-position variants, U3.4 R7A, U3.5 R7A, and U3.6 K7A,

are listed in Table 1.2. Uperin wild-type peptides from the 3.*x* family have innate amidation at the C-terminus[10], as shown in Table 1.2.

Uperin 3.*x* peptides are relatively small amyloidogenic AMPs, and therefore ideal for understanding the effect of electrolyte concentration on the amyloid formation process, and the associated secondary structure changes. Furthermore, these uperin 3.*x* peptides can provide important insights into the effect of cationic to hydrophobic residue substitution on the process of amyloidogenesis, the evolution of secondary structure, and the interaction with lipids. Therefore, both fully-atomistic and coarse-grained MD simulations were used in this study to examine the aggregation and lipid interaction properties of the uperin 3.*x* peptides and variants listed in Table 1.2.

1.2. Thesis objectives and roadmap

The central objective of the thesis was to explore the 'amyloidogenicity and lipid interaction of antimicrobial anuran peptides', and this has been addressed in the following chapters, briefly described below:

Chapter 3 - Mechanistic insight into the early stages of amyloid formation using an anuran peptide: *the objective was to understand the effect of NaCl concentration on the aggregation behaviour and the associated secondary structure evolution of U3.5 wt peptide in an aqueous medium.*

Chapter 4 - Interaction of anuran peptides with DPC and SDS micelles: A Molecular Dynamics study: *the objective was to obtain mechanistic insight into the effect of seventh-position alanine substitution on the interaction of uperin 3.x peptides and corresponding alanine variants with eukaryotic membrane-mimic zwitterionic dodecyl phosphatidylcholine (DPC) and bacterial membrane-mimic anionic sodium dodecyl sulphate (SDS) micelles.*

Chapter 5 - Propensity of Uperin 3.x peptides towards beta-aggregation: Analysis of structurally-diverse conformers: *the objective was to elucidate the effect of seventh-position alanine substitution on the self-aggregation phenomena of uperin 3.x wild-type peptides and corresponding alanine variants.*

The thesis also includes a literature review (Chapter 2), and a chapter on concluding remarks (Chapter 6). Chapter 2 (Literature review) is further divided into Amyloidogenesis (Section 2.1), Antimicrobial peptides (Section 2.2), Role played by ions and mutations (Section

2.3), An introduction to molecular dynamics theory (Section 2.4.), and Computational models and methods (Section 2.5). Chapter 6 provides a summary, and a cross-link between the Chapters 3–5 (as mentioned above). Moreover, it also provides suggestions for future research.

1.3. References

1.   Chiti, F.; Dobson, C. M., Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333-66.

2.   Lansbury, P. T.; Lashuel, H. A., A Century-Old Debate on Protein Aggregation and Neurodegeneration Enters the Clinic. *Nature* **2006**, *443*, 774-9.

3.   Hardy, J.; Selkoe, D. J., The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002**, *297*, 353-6.

4.   Nagy, Z.; Esiri, M. M.; Jobst, K. A.; Morris, J. H.; King, E. M.; McDonald, B.; Joachim, C.; Litchfield, S.; Barnetson, L.; Smith, A. D., The Effects of Additional Pathology on the Cognitive Deficit in Alzheimer Disease. *J. Neuropathol. Exp. Neurol.* **1997**, *56*, 165-70.

5.   Kirkitadze, M. D.; Bitan, G.; Teplow, D. B., Paradigm Shifts in Alzheimer's Disease and Other Neurodegenerative Disorders: The Emerging Role of Oligomeric Assemblies. *J. Neuro. Res.* **2002**, *69*, 567-577.

6.   Eisenberg, D. S.; Sawaya, M. R., Structural Studies of Amyloid Proteins at the Molecular Level. *Annu. Rev. Biochem.* **2017**, *86*, 69-95.

7.   Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27-68.

8.   Calabrese, A. N.; Liu, Y.; Wang, T.; Musgrave, I. F.; Pukala, T. L.; Tabor, R. F.; Martin, L. L.; Carver, J. A.; Bowie, J. H., The Amyloid Fibril-Forming Properties of the Amphibian Antimicrobial Peptide Uperin 3.5. *ChemBioChem* **2016**, *17*, 239-246.

9.   Zhang, M.; Zhao, J.; Zheng, J., Molecular Understanding of a Potential Functional Link between Antimicrobial and Amyloid Peptides. *Soft Matter* **2014**, *10*, 7425-7451.

10.   Bradford, A. M.; Bowie, J. H.; Tyler, M. J.; Wallace, J. C., New Antibiotic Uperin Peptides from the Dorsal Glands of the Australian Toadlet Uperoleia Mjobergii. *Aust. J. Chem.* **1996**, *49*, 1325-1331.

11.   Hong, Y.; Pritzker, M. D.; Legge, R. L.; Chen, P., Effect of Nacl and Peptide Concentration on the Self-Assembly of an Ionic-Complementary Peptide Eak16-Ii. *Colloids Surf. B* **2005**, *46*, 152-161.

12.   Aramvash, A.; Seyedkarimi, M. S., All-Atom Molecular Dynamics Study of Four Rada 16-I Peptides: The Effects of Salts on Cluster Formation. *J. Clust. Sci.* **2015**, *26*, 631-643.

13.   Smith, M. D.; Cruz, L., Effect of Ionic Aqueous Environments on the Structure and Dynamics of the Aβ21–30 Fragment: A Molecular-Dynamics Study. *J. Phys. Chem. B* **2013**, *117*, 6614-6624.

14.   Chia, B. C. S.; Bowie, J. H.; Carver, J. A.; Mulhern, T. D., The Solution Structure of Uperin 3.6, an Antibiotic Peptide from the Granular Dorsal Glands of the Australian Toadlet, Uperoleia Mjobergii. *The Journal of Peptide Research* **1999**, *54*, 137-145.

15.   Martin, L. L.; Kubeil, C.; Piantavigna, S.; Tikkoo, T.; Gray, N. P.; John, T.; Calabrese, A. N.; Liu, Y.; Hong, Y.; Hossain, M. A., Amyloid Aggregation and Membrane Activity of the Antimicrobial Peptide Uperin 3.5. *Pept. Sci.* **2018**, *e24052*.

16.   Mihajlovic, M.; Lazaridis, T., Charge Distribution and Imperfect Amphipathicity Affect Pore Formation by Antimicrobial Peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818*, 1274-1283.

17.   Shai, Y., Molecular Recognition between Membrane-Spanning Polypeptides. *Trends in Biochemical Sciences* **1995**, *20*, 460-464.

18.   Das, A.; Makarov, D. E., Effect of Mutation on an Aggregation-Prone Segment of P53: From Monomer to Dimer to Multimer. *The Journal of Physical Chemistry B* **2016**, *120*, 11665-11673.

19.   Xu, L.; Ma, B.; Nussinov, R.; Thompson, D., Familial Mutations May Switch Conformational Preferences in A-Synuclein Fibrils. *ACS Chemical Neuroscience* **2017**, *8*, 837-849.

20.   Tarus, B.; Straub, J. E.; Thirumalai, D., Structures and Free-Energy Landscapes of the Wild Type and Mutants of the Aβ21–30 Peptide Are Determined by an Interplay between Intrapeptide Electrostatic and Hydrophobic Interactions. *Journal of Molecular Biology* **2008**, *379*, 815-829.

Chapter 2

Literature Review

2.1. Amyloidogenesis

2.1.1. Introduction

A large number of human diseases occur due to the failure of a specific peptide or protein to adopt or remain in its native functional conformational state[1]. These pathologies are referred to as *protein misfolding or protein conformational diseases*. The largest group of misfolding diseases is associated with the conversion of specific peptides or proteins from their soluble functional states into highly organized fibrillar aggregates. These structures are generally described as amyloid fibrils or plaques when they accumulate extracellularly, whereas the term "intracellular inclusions bodies" is employed if such fibrils form inside the cell. The 'amyloid hypothesis' implicates protein aggregation as the trigger of a cascade of events that result in neurodegeneration and disease[2]. The diseases can be broadly grouped into neurodegenerative conditions, in which aggregation occurs in the brain; non-neuropathic localized amyloidosis, in which aggregation occurs in a single type of tissue other than the brain; and non-neuropathic systemic amyloidosis, in which aggregation occurs in multiple tissues[1]. However, most amyloid-forming proteins do not share any apparent sequence identity or structural homology. Moreover, considerably large variation is observed in secondary structure composition and the polypeptide chain length. Although, most of them have fewer than 100 amino acid residues[3].

Ribosome with
nascent chain

Functional
fibers

Functional oligomers

IDPs

Proteolyzed
fragment

Unfolded
state

Partially
folded state

Partially folded
(off-pathway)
state

Folded or
globular state

Ligand binding

Native-like state

Covalent
modification

Partially or fully folded

Disordered
aggregates

Disordered
aggregates

Native-like
aggregates

β-structured
aggregates

Initial aggregates

Amorphous deposits

Amyloid fibrils

Native-like deposits

Mature aggregates

Chiti F, Dobson CM. 2017.
Annu. Rev. Biochem. 86:27–68

Figure 2.1. A schematic representation of some of the many conformational states that can be adopted by polypeptide chains, and of the means by which they can be interconverted[3].

2.1.2. Fibril formation kinetics

During fibril formation, sigmoidal kinetics is observed with three distinct phases: a lag phase, an exponential phase (also termed as elongation, growth, polymerization, or fibrillation phase), and an equilibrium phase (also termed as plateau or saturation phase)[3].

A range of amyloid-prone proteins, such as α-synuclein (αs), tau, and the islet amyloid polypeptide (IAPP) are largely unstructured in solution, also described as natively unfolded or intrinsically disordered (Figure 2.1)[3]. However, they might fold into more well-defined structures after interaction with specific binding partners. Intrinsically disordered systems can also be generated following proteolysis from larger proteins that are otherwise folded, such as the amyloid-β peptide (Aβ) and the amyloidogenic fragment of gelsolin. In case of globular proteins like β2-microglobulin and transthyretin, the usually compact and highly cooperative structures bury the most aggregation-prone regions of the molecules within the core of the protein. Thus, fully folded proteins must convert into a partially unstructured ensemble that is competent for fibrillar aggregation. The different conformational states adopted by proteins involve a highly complex series of equilibria, whose thermodynamics and kinetics in a normally functioning living system are determined by their intrinsic amino acid sequences as well as through interactions with molecular chaperones, degradation processes, and other sophisticated quality control mechanisms. Here, the competition between intramolecular and intermolecular interactions needs to be considered, resulting in a dramatic increase in landscape ruggedness[4]. Multiple pathways are accessible and selected depending on experimental conditions, protein sequence, and conformational state adopted by the amyloidogenic monomer[3]. Although the amino acid sequences of proteins and peptides and, the surrounding biological environments in which they function have coevolved to maintain them in their soluble states, in some circumstances they can convert into non-functional and potentially damaging protein aggregates (Figure 2.1).

During an aggregation reaction, the initial species are generally largely unstructured, with the β-sheet–containing species appearing only later[3]. The large oligomers with high β-sheet content represent off-pathway species that need to dissociate prior to forming amyloid fibrils. The average β-sheet content generally increases with molecular weight, suggesting that an increase in size stabilizes their β-sheet structure. Such structures usually possess both antiparallel and parallel as well as out-of-register strands. However, in the $A\beta_{40}/A\beta_{42}$ fibrils. the β-strands are arranged in a parallel and in-register manner.

In nucleated polymerization mechanism, monomers that are completely or partially disordered convert into nuclei through a thermodynamically unfavourable process that takes place early in the lag phase[3]. These contain a relatively small number of molecules, and generally retain a structural memory of the monomeric states that have generated them (Figure 2.1). Such early aggregates are usually unstable, as only relatively weak intermolecular interactions are involved, and they may dissociate to regenerate soluble species. However, when the aggregation proceeds, such aggregates can undergo internal reorganization to form more stable species having β-sheet structure, a process that is often accompanied by an increase in compactness and also size. Fibrils then grow from these nuclei through the addition of monomers, often with structural reorganizations, to form well-defined fibrils with cross-β structure and a high level of structural order. This model can describe the aggregation of mammalian cellular prion protein (PrP), $A\beta_{40}$ and $A\beta_{42}$, αs, polyGln sequences, insulin, and others.

In some cases, however, monomers convert rapidly into misfolded aggregates that lack the structural characteristics needed to grow into organized fibrils. But these initial aggregates can undergo structural reorganization to generate nuclei, and other disorganized oligomers can interact with the end region of this aggregate to acquire the amyloid conformation through a templating or induced-fit mechanism, eventually leading to the formation of fibrils[3]. This nucleated conformational conversion mechanism has been used to describe amyloid formation by $A\beta_{40}$, αs, PrP, huntingtin exon 1, IAPP, and others.

2.1.3. Fibril structure

The amyloid fibrils are highly organized, with persistence lengths on the order of microns, and with a tensile strength that approaches that of steel[3]. The mechanical strength and stability of the amyloid fibrils can be primarily attributed to the formation of the extended arrangement of hydrogen-bonded β-sheets, that are characteristic of the amyloid architecture. This feature is enhanced by the presence of multiple closely interacting sheets as well as further association of protofilaments to form higher order structures. The morphological and structural similarity of the fibrils formed by different polypeptide chains include the similarities in the lengths of the β-strands, the low degree of twist in the β-sheets, and the fact that short peptides can assemble into three-dimensional crystals, whereas the most regular structures attainable by longer sequences are one-dimensional filamentous assemblies. Even under physiological conditions. the amyloid form of a protein molecule can be more stable than the native state,

whether intrinsically disordered or globular. Such a situation is particularly true in the case of short sequences, as longer sequences not only incorporate a fraction of their polypeptide chains into the cross-β core, but also have an increasing probability of being unable to generate highly ordered structures from unfolded segments of their sequences. Therefore, the polypeptides associated with amyloid formation have relatively short chain lengths, whereas proteins forming amorphous or native-like deposits are on average much longer (Figure 2.1). The regions found to correlate with a high level of hydrophobicity, a high propensity to form β-sheet structure, a small number of charged residues, and the presence of alternating hydrophilic and hydrophobic residue patterns, are frequently observed in these fibrils. The inherent architecture of the amyloid structure is not only determined by the intrinsic properties of the main chain, but also the optimization of the interactions among the various sidechains.

Even under the same solution conditions, a multitude of fibril morphologies can be formed simultaneously, indicating the complexity and multiplicity of aggregation pathways[4]. Because of this polymorphism, the same protein sequence can give rise to fibrils that differ both in the molecular structures of their protofilaments, and the overall morphology of the mature fibrils[3]. The conformational polymorphism of amyloid fibrils can also affect their biological properties[4]. The cause of these subtle but highly significant conformational changes is situated in the assembly precursor states further upstream in the aggregation landscape (Figure 2.1)[3]. Polymorphism is an inherent consequence of the generic ability of polypeptide chains to form amyloid fibrils. Unlike native folds that have been selected through evolution and are encoded in their amino acid sequences, the amyloid architecture is simply a consequence of the physicochemical properties of a polypeptide chain, and a given sequence can be incorporated in such structures in multiple ways (Figure 2.2). In the majority of cases, in the absence of evolutionary selection and optimization, a multiplicity of structural arrangements of closely similar energy is likely. However, once a given type of assembly has been initiated, it will generally propagate because of the inherent stability of repetitive structures in highly organized systems, and the high kinetic barriers that exist for the interconversion between polymorphs.
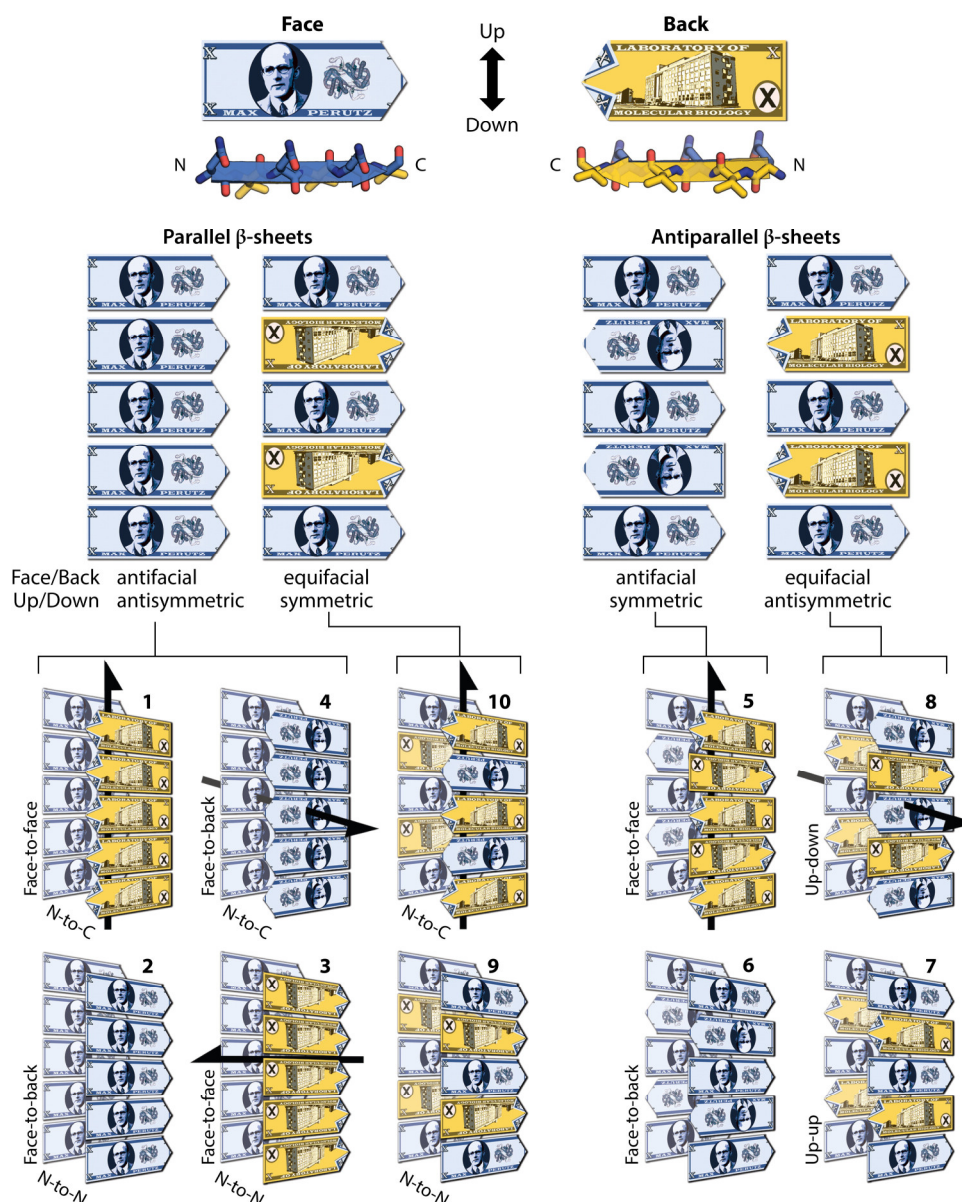
Figure 2.2. The ten amyloid symmetry classes of homo-steric zipper amyloid spines illustrated with the fictitious Max Perutz bank note. Single-bladed arrows signify 2_1 symmetry axes, meaning that the two sheets are related by a 180° rotation about the arrow and a translation along the arrow of one-half the distance between Max Perutz bank notes[5].

The strands are generally found to interact to form β-sheets with a parallel in-register arrangement in Aβ$_{40}$, Aβ$_{42}$, αs, tau, IAPP and β2-microglobulin fibrils[3]. The frequency of this parallel in-register arrangement relative to other alternatives can be attributed to both thermodynamic and kinetic factors, as it maximizes the number of hydrogen bonds and hydrophobic interactions along the fibril axis. The out-of-register arrangement necessarily generates alternative interactions of hydrophobic-hydrophilic residues in the fibril direction,

and so is likely to give rise to less stable structures. However, for small peptides, alternative organizations of the β-strands within fibrils, in which an antiparallel arrangement is sometimes preferred to a parallel one, are known to exist. The small number of residues involved means the energetic penalty to be paid in breaking the parallel in-register arrangement is relatively low and can be compensated by alternative interactions, for example, salt bridges between amino acid residues with opposite charges at the N-terminus and C-terminus, that are spatially close in the fibrils. Interestingly, in polyQ containing peptides, an antiparallel arrangement has been found for both long and short sequences. Antiparallel arrangements have also been found for long sequences without repetitive amino acid residues, such as in fibrils of calcitonin, some polymorphs of transthyretin, and the D23N mutant of $A\beta_{40}$. Except for peptides containing less than approximately 12 residues, only a fraction of the polypeptide chain is incorporated in the cross-β core of the fibrils.

2.1.4. Amyloid pathogenicity

The pathogenic species in non-neuropathic systemic or localized amyloidoses are both the extracellular amyloid deposits affecting the organ integrity, and the protein oligomers that are generated during the process of their formation or are released by mature deposits, causing direct cellular damage[3]. However, in neuropathic diseases involving the central nervous system, the pathogenicity arises from the oligomeric forms generated in the process of aggregation. Prefibrillar species or protofibrils lack fibrillar substructure and are a precursor to fibrillar amyloid[2]. These oligomers are thought to represent the most pathogenic species in the diseases associated with amyloid fibril formation[3]. $A\beta_{42}$ forms protofibrils more rapidly than $A\beta_{40}$, and is more strongly correlated with the pathology[2].

The fibrils can, however, deplete key components of the protein homeostasis network, serve as a reservoir of protein oligomers that can be released, and perhaps most importantly, act as potent catalysts for the generation of toxic oligomers through secondary nucleation[3]. Moreover, spreading of fibrils within organs such as the brain generates a number of fibrillar aggregates that amplify all such possibilities. In addition, at least some polymorphs of mature fibrils can have significant cytotoxicity.

The exposure of hydrophobic groups on the oligomer surface is a major determinant of oligomer-mediated toxicity[3]. Small-sized oligomers have a high diffusion coefficient, which allows them to diffuse more rapidly and have frequent aberrant interactions with the surrounding environment. Hence, the toxicity of the oligomers was found to correlate strongly

and inversely with their size, and directly with their surface hydrophobicity. Interestingly, the shape of the oligomers is not a crucial factor.

Different kinds of toxic aggregates are associated with these neuropathic diseases[3]. Further, the large network of aberrant interactions that such species can generate indicates that it is improbable to have a unique molecular interaction, a unique mechanism of action, or a unique cascade of cellular events, associated with a given disease. Instead, the toxicity of the protein aggregates, that results in diseases, is probably a result of their intrinsic misfolded nature and their structural heterogeneity. Such properties will cause them to engage in a multitude of aberrant interactions with a range of cellular components, including phospholipid bilayers, protein receptors, soluble proteins, RNAs, and small metabolites—any or all of which have the potential to cause cellular damage and ultimately cell death.

Fibrillar species with amyloid characteristics can serve a number of biological functions *in vivo*[1]. Such fibrillar structures have been observed in *Escherichia coli* and *Streptomyces coelicolor* bacteria, and also melanosomes in humans. However, these fibrils are formed under controlled conditions, and have extremely highly regulated assembly processes.

2.1.5. Causative agents
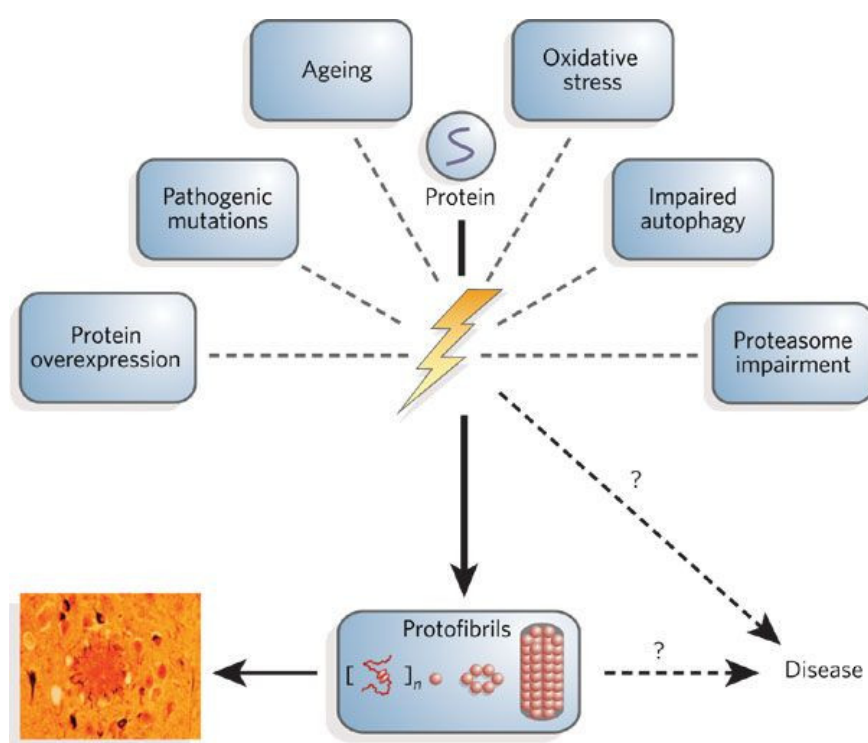
Figure 2.3. A number of factors have been proposed to trigger protein oligomerization (protofibril formation) and disease[2].

Half of the amyloid-related diseases, such as Alzheimer's and Parkinson's diseases, are predominantly sporadic, although hereditary forms are also well documented[3]. When sporadic, they have a relatively late age of onset, suggesting that the protein aggregation and the associated symptoms originate primarily from a progressive loss of regulatory control with aging. One-third of the diseases, such as the lysozyme and fibrinogen amyloidoses, arise from specific mutations and are hereditary. Therefore, they have an early age of onset. A small number of amyloid disorders occur as a result of medical treatment. Notable examples include dialysis-related amyloidosis, injection-localized amyloidosis, and PrP-associated Creutzfeldt-Jakob and Kuru diseases.

The genes linked to diseases are predominantly of two types: familial and susceptibility factors[2]. Several genes of the former type have been identified, and when mutated, these cause rare, early-onset forms of disease. The mutations are typically located in the polypeptide chain that undergoes aggregation, and are autosomal dominant; in other words, a mutation in just one copy of the gene is sufficient to cause disease[3]. Although susceptibility factors are observed in many patients, it is difficult to translate their subtle effects into a working hypothesis[2].

The pathogenic mutations often increase the propensity of the protein to aggregate, either directly or indirectly, as observed in transthyretin and lysozyme amyloidosis[3]. Formation of partially unfolded states facilitates aggregation, as hydrophobic groups and segments of the polypeptide backbone that are buried and engaged in structure formation in the native state become available for intermolecular interactions. Moreover, some mutations, such as those linked with familial amyloidosis, have been reported to increase the rate of aggregation of the fragments themselves. Most of these pathogenic mutations destabilize the native states, making them susceptible to aggregation either directly or following proteolytic cleavage. Mutations in tau protein cause alternative splicing of the pre-mRNA of tau, generating a highly amyloidogenic isoform in addition to the normal sequence. For Alzheimer's disease, mutations in the β-amyloid precursor protein (βAPP) increase the specificity of γ-secretase for cleavage sites so that the proportion of the $A\beta_{42}$ isoform is increased relative to the less amyloidogenic $A\beta_{40}$. Mutations can also become pathogenic through a change in the stop codon for translation, as observed in genes associated with hypotrichosis simplex of the scalp, familial British dementia, and others. In addition, there are several disorders associated with the elongation of the amyloidogenic CAG repeat expansion within a gene, as observed for the exon 1 of the huntingtin gene of the Huntington's disease. A

duplication or triplication of the gene, resulting in a higher concentration of the amyloidogenic protein, is observed in hereditary forms of Alzheimer's disease and Parkinson's disease.

Given the intrinsic propensity of proteins to form thermodynamically stable amyloid fibrils and the generally detrimental effects of misfolded protein oligomers, all living organisms have developed a dedicated network of cellular machineries to assist protein folding and counteract protein aggregation[3]. This network is generally referred to as the "proteostasis network" and is constituted by the translational machinery of the cell, molecular chaperones and cochaperones, and protein degradation mechanisms, such as the ubiquitin proteasome system (UPS) and autophagy. Molecular chaperones and cochaperones play a crucial role in protein folding, the assembly of complexes, inhibiting protein aggregation, and mediating protein degradation via the UPS or autophagy. Moreover, cells also use integrated processes in which sensors of the misfolded species activate the transcription of genes coding for proteostasis network components, such as the heat shock response in the cytosol and nucleus. Further, the aggregation of a polypeptide chain is also intrinsically inhibited by its amino acid sequence, that is carefully selected by evolution to promote folding into a compact stable structure, and generally possesses a very low propensity to self-assemble and aggregate when adopting a partially or fully unfolded state. Even transcriptional expression and half-lives of proteins are carefully regulated, depending on their intrinsic aggregation propensities.

A narrow boundary exists between health and diseases, and subtle changes caused, for example by genetic mutations, environmental stress, or increasing age, can result in the initiation of the pathogenic aggregation process, leading to the disruption of the proteostasis network (Figure 2.3)[3]. A progressive failure of the proteostasis network occurs with aging in humans: chaperone levels are reduced because they are sequestered by the increasing number of protein aggregates, and their expression is significantly downregulated with aging. An age-related decline of the UPS activity has also been observed, along with decreased efficiency of macroautophagy and chaperone-mediated autophagy. The age-associated increase of oxidative stress and decrease in mitochondrial function and ATP production further reduces the efficiency of the proteostasis network. Regulation of mRNA transcriptional levels determines the protein concentration, and thus the aggregation propensity as supersaturated environments are highly susceptible to aggregation.

Once the protein homeostasis network is overwhelmed, fibrils have an extraordinary ability to proliferate, not only as a result of their growth and the formation of new nuclei, but also

through secondary processes, such as fibril fragmentation and secondary nucleation, which increase the number of fibrils[3]. Moreover, fibrils or their precursors can spread from cell-to-cell within the same tissue and even across distinct regions of the same or different organs, contributing to the transmission of the histopathological traits of their associated disease to involve different locations.

2.1.6. Membrane action
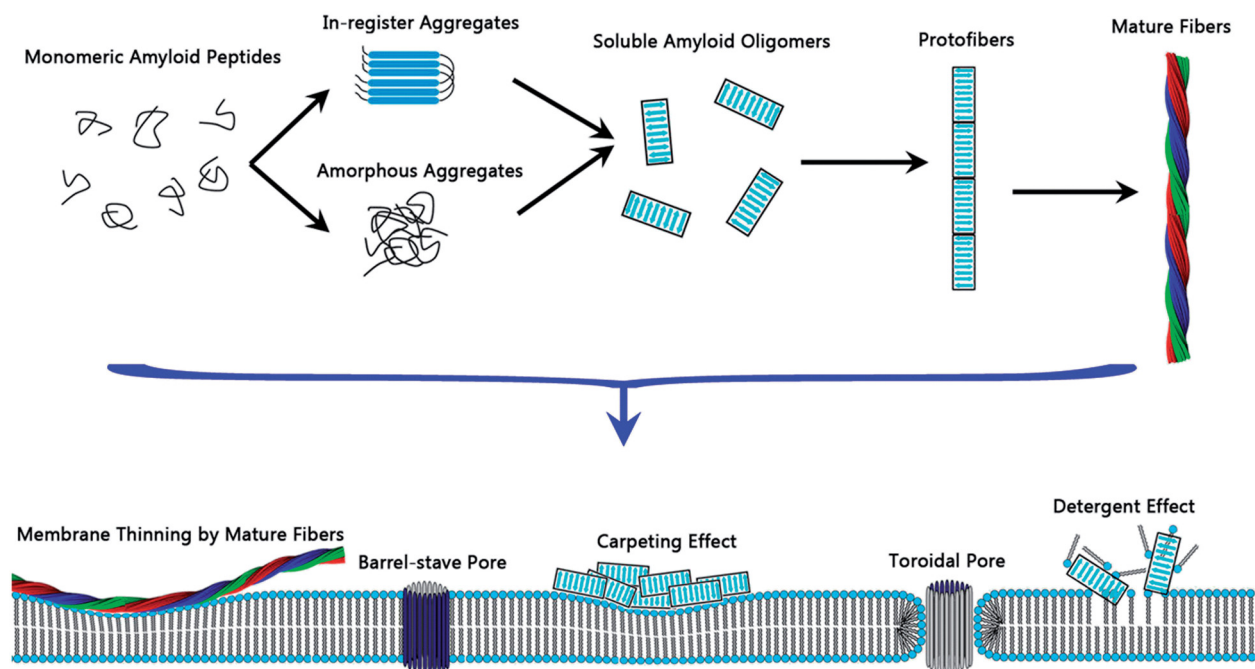
Figure 2.4. Schematic illustration of possible membrane-disruption mechanisms and aggregation pathways in solution by amyloid peptides[6].

Membrane disruption by amyloidogenic peptides usually occurs via two general mechanisms: the peptides can either penetrate into the membrane to form ion-channels/pores or associate with the membrane surface by adsorption/insertion to induce membrane thinning (Figure 2.4). Whereas both mechanisms are common for small monomers or oligomers, larger oligomers primarily act via membrane thinning[6]. Amyloid-forming peptides, such as the human IAPP, can also exert detergent-like effects on membrane permeation and disruption via amyloid oligomers and fibrils. Large membrane-bound amyloid aggregates and fibrils can severely distort the membrane curvature and cause membrane thinning, as observed for human IAPP fibrils. The carpet model can explain the exponential leakage kinetics and absence of a lag phase in both human IAPP- and rat IAPP-induced membrane leakage. Amyloidogenic peptides can induce both barrel-stave and toroidal pores, although computational studies of amyloid channels indicate that barrel-stave pores are more energetically favourable than toroidal pores.

Amyloid channels in cell membranes consist of several loosely contacting mobile oligomers. Further, these channels contain a U-shaped β-strand–turn–β-strand conformation, as observed in channels of Aβ, K3 and human IAPP oligomers. Thus, it may represent a general building block for both amyloid channels and amyloid fibrils.

2.1.7. Therapeutic avenues

The ability of antibodies to not only bind to unique sequences, but also to well-defined aggregation states, has led to considerable efforts to develop immunotherapies for amyloid-associated diseases[3]. In case of globular proteins, an evidently powerful strategy would be to stabilize the folded state to prevent both unfolding and the structural fluctuations that generate native-like states. An analogue of thyroxine, termed tafamidis, stabilizes the native tetramer of transthyretin and reduces its aggregation propensity. Therefore, it is widely utilized to treat familial amyloid polyneuropathy. Many rational design efforts have focused on inhibition of the process of aggregation itself. However, all early attempts to develop aggregation inhibitors have failed at one or other stage of clinical trials. Hence, efforts are instead being focused upon specific steps in the aggregation reaction by screening compounds *in vitro* and *in vivo* to identify potential inhibitors targeted towards neurodegenerative diseases. One stratagem is to develop inhibitors of the specific proteases that produce amyloidogenic fragments, such as the β- and γ-secretases that produce the N- and C-termini of the Aβ fragments, respectively. Upon inhibition of Aβ oligomerization by a PKC inhibitor, the capture and release of lipids by oligomeric Aβ are completely ceased, thus preventing Aβ oligomer-induced cell toxicity[6]. In non-neuropathic amyloidosis, strategies based on the removal or reduction of the source of amyloidogenic proteins are effective therapeutic avenues[3]. Moreover, βAPP and a large fraction of proteins coaggregating with amyloid plaques and intracellular neurofibrillary tangles, respectively, are transcriptionally downregulated in the central nervous system of Alzheimer's disease patients relative to healthy age-matched controls, presumably in an attempt to limit their further aggregation. This could be used as an early indicator of diseases and as a biomarker for their progression.

2.2. Antimicrobial peptides

2.2.1. Introduction

Antimicrobial peptides (AMPs) function as host-defence or host-offense components of many innate immune systems in plants, animals and humans[6]. Naturally occurring AMPs are

typically 20–40 residues in length, with a wide range of sequence diversity. Unlike native proteins, AMPs usually have a short and appropriate length, which helps them to fold into a relatively simple secondary structure and size approximating the membrane thickness. Most AMPs are overall cationic, with more abundant hydrophobic amino acids and less abundant polar residues. Due to natural selection, some AMPs preferentially kill particular pathogens, and therefore show certain amino acid compositional preferences. For instance, histatin and indolicidin are rich in histidine and tryptophan residues, respectively.

2.2.2. Membrane action

Bacterial membranes and human cell membranes have similar functions and structures, but different compositions[6]. Human cell membranes usually contain up to 45% cholesterol among the total lipids, while bacterial membranes do not contain cholesterol. The cholesterol in human cell membranes contributes to substantially lower adsorption of the peptides, resulting in lower membrane rupture and leakage. Most eukaryotic human cell membranes contain zwitterionic phospholipids, rendering the outer layer of the membrane uncharged. However, the outer membrane of prokaryotic bacterial cells is rich in anionic lipids. These compositional differences impart a structural basis for AMPs to preferentially bind to anionic lipids over zwitterionic phospholipids, which leads to cell selectivity. Positively-charged residues (e.g. lysine and arginine) of AMPs determine their cell selectivity for anionic bacterial membranes, and thus promote selective surface binding to target bacterial membranes. Upon adsorption, hydrophobic residues facilitate further penetration into the hydrophobic interior of the membranes, but they favour no tilt angle in particular. Therefore, polar and charged residues are crucial for determining the correct peptide orientation relative to the membrane surface. Antimicrobial activity is mainly governed by the appropriate balance of hydrophobic and electrostatic interactions between peptides, water, and lipids, as well as by the peptide sequence and structure.

In the carpet model, at a high peptide-to-lipid (P/L) ratio, AMPs intensively adsorb onto the membrane, with a preferentially parallel orientation, to reduce membrane thickness and increase membrane conductance (Figure 2.5c)[6]. Alternatively, AMPs at a high concentration act as "detergents" to extract lipids from the membrane to form peptide-lipid micelles, resulting in membrane disintegration (Figure 2.5d). In both models, the monomeric or oligomeric peptides accumulate at the membrane surface at high density to cause membrane damage. After

certain level of adsorption of peptides, secondary factors, such as membrane thinning, 2D phase separation, pore formation, and curvature strain, come into play.
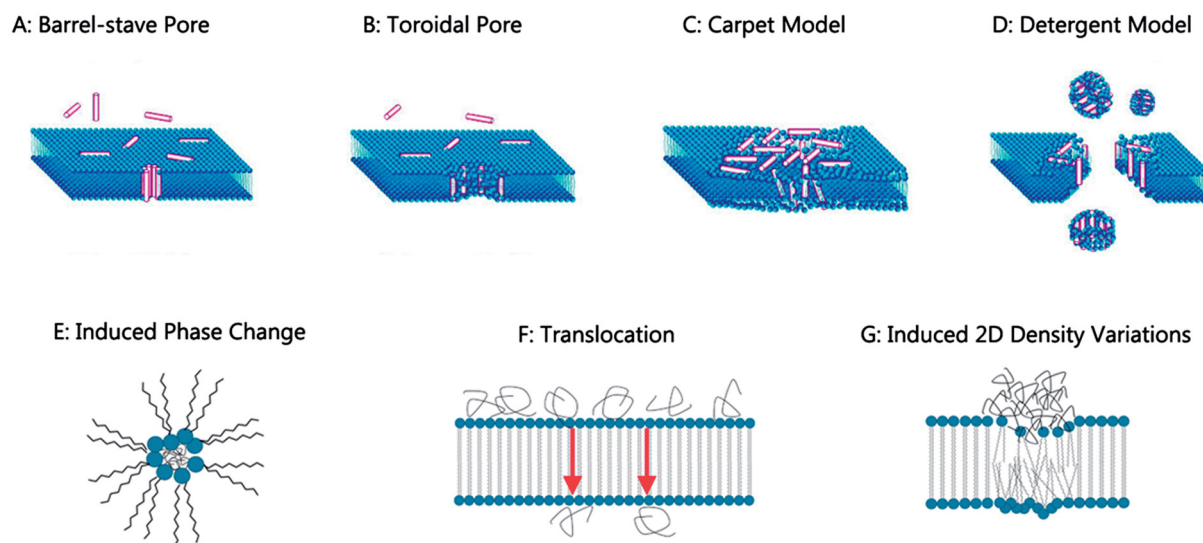
Figure 2.5. Different membrane-disruption models for antimicrobial peptides[6].

Peptides with a high threshold P/L^* value act via carpet and detergent models, while those with a low P/L^* appear to form leaking pores[6]. AMPs can insert into the cell membranes to form barrel-stave or toroidal transmembrane pores (Figures 2.5a and 2.5b, respectively). For example, distinctin and ceratotoxins can form barrel-stave pores, while viroporin and protegrin-1 (PG-1) create toroidal transmembrane pores. The barrel-stave pore-forming peptides stay at the membrane surfaces or inside the membrane, whereas the toroidal pore-forming peptides show stronger preference to stay at the water-lipid interface, while disfavouring but not fully excluding the centre of the membrane. Both pore models can cause similar membrane disruption, leakage of cytoplasmic contents, and concomitant bacterial cell death. Often both pore models are applicable to the same peptide under different conditions, as observed for bee-venom melittin and alamethicin.

2.2.3. Structural composition

The AMPs with known structures can be generally classified into four groups based on their secondary structure: α-helix, β-turn/sheet, mixed α/β structure and random coil[6]. Most α-helical and random coil AMPs lack disulphide bonds, while many AMPs with β-strand, β-hairpin or loop-strand conformation (e.g. PG-1 and defensins) often have disulphide bonds between cysteine residues to stabilize their peptide conformation.

AMPs can adopt completely different structures in solution and within the membrane[6]. Typically, AMPs undergo a complex structural transition from the bulk solution to the water-membrane interface to the membrane interior (if membrane insertion occurs), depending on the interplay among peptides, membranes, and environments. Many AMPs without disulphide bonds, such as melittin and cecropin A, undergo a random coil-to-helix transition upon interaction with the membranes. However, some unstructured AMPs lack the ability to convert to ordered secondary structure (e.g. α or β structures) even after adsorption onto a lipid-water interface.

2.2.4. Comparison with amyloidogenic peptides

Cytolytic peptides damage cell membranes and cause target cell death[6]. The two major categories of cytolytic peptides are amyloidogenic peptides (AMYs) and AMPs. Although the AMYs have a varied size, they typically contain less than 50 residues. Most AMYs have a high content of hydrophobic residues, whereas most AMPs are cationic, with abundant hydrophobic residues to counterbalance the positive charges. However, both AMYs and AMPs are largely amphipathic, with hydrophobic and charged hydrophilic regions. Amphipathicity is essential for the membrane disruption, as it facilitates membrane interaction. Both classes of peptides have a high content of hydrophobic residues, such as isoleucine and valine, because they are critical for membrane insertion, and stabilizing peptide insertion states with preferential hydrophobic match with hydrophobic lipids inside the membrane core. Positively-charged residues, such as arginine and lysine, play a crucial role in antimicrobial peptide-cell membrane interactions, but these residues rarely occur in amyloid sequences, as they prevent peptide aggregation. Interestingly, AMYs usually contain hydrophobic (P), hydrophilic (H) and charged (C) residues in a sequential manner, like CCCHHHPPPPPPPPPPPPPC, while in AMPs they are arranged in an alternate fashion, like PHCPHCPHCPHCPHCPHC.

While AMYs exhibit a characteristic cross-β-sheet structure, AMPs have diverse secondary structures[6]. For the β-rich structures of AMYs and AMPs, disulphide bonds usually occur less often in AMYs than in AMPs. Proline disfavours the β-structure in amyloid sequences but shows no obvious preference in AMP sequences. Many AMPs adopt a helical structure upon interaction with the membranes, and thus α-helical pores are observed more frequently than pores of the other conformations. The helical structure might be the active conformation of AMPs for membrane permeation. However, almost all computationally

modelled amyloid pores adopt β-sheet structures, probably because AMYs have an intrinsic ability to form β-sheet structures, both in solution and on the membrane.

2.2.5. Correlation between antimicrobial activity and amyloidogenicity

The serum amyloid A peptide can assemble into fibrils[6], and demonstrates antimicrobial and cytotoxic actions via formation of ion-permeable channels in cell membranes. Aβ can exert strong antimicrobial activity against some common and clinically relevant microbes, such as against the growth of *Candida albicans*. Amyloidogenic IAPP can cooperate with both antimicrobial magainin 2 and its D isomer to induce significantly greater membrane leakage and ultimately bacterial cell death, as compared to any of these individual peptides.

Some AMPs, such as temporin L, magainin 2, indolicidin, and plantaricin A, can form amyloid-like fibres[6]. In fact, PG-1 AMPs not only form fibrils with morphologies similar to the Aβ$_{42}$ fibrils in solution, but also PG-1 fibrils grow relatively faster than Aβ$_{42}$ fibrils. Human α-defensin 6 can self-assemble into functional amyloid-like fibrils, and then surround and entangle the target bacteria.

The cross-seeding with bacterial CsgA and CsgB reduced the lag-time of both IAPP and Aβ amyloid formation[6]. However, both CsgA and CsgB strongly inhibited IAPP elongation, exerted a concentration dependent effect on Aβ elongation. Therefore, AMYs might act synergistically with AMPs to exert different biological functions.

While some AMPs can self-assemble into amyloid-like fibrils with signature cross-β-sheet structures, certain AMYs can exert antimicrobial activity against several common microorganisms[6]. Such commonalities in their structural and functional properties indicate towards a potential functional link between amyloidogenic and antimicrobial peptides.

2.3. Role played by ions and mutations

2.3.1. Ions

2.3.1.1. Introduction

Most biomolecular MD simulations are performed in an aqueous medium with a certain ionic concentration to mimic the *in vivo* environment. Ions play a crucial role in not only modulating the intra- and inter-peptide interactions, but also the interaction of the peptides with the surrounding solvent molecules. In the absence of salt, uncharacteristically strong interactions between oppositely-charged sidechains could be observed because of negligible charge

screening[7]. However, placement of ions near the charged sidechain moieties at the outset of the simulation would significantly screen the electrostatic interactions amongst them, thus retarding translational diffusion and delaying equilibration. Since the Debye length for ions decreases with increasing ionic strength, a higher salt concentration would result in reduced Maxwell relaxation period. Thus, adding just enough ions to maintain charge-neutrality of the system might not be sufficient. However, high ionic strength would entail computation of electrostatic interactions without truncation. The salt ions also reduce the dielectric constants of the peptide and the surrounding water molecules in the solvation shell, as they hinder the movement of other atoms due to electrostatic interactions[8].

2.3.1.2. Effect on structure

Consisting of a group IIA cation, the $CaCl_2$ salt promotes and stabilizes random coil structures in $A\beta_{21-30}$, while reducing the intrapeptide hydrogen bonding[9]. This could be attributed to cation-peptide salt-bridges and accompanying interaction with the water molecules in the peptide hydration-shell. Similar but relatively diminished action is exhibited by the $MgCl_2$ salt also. Such random coil structures within full-length $A\beta$ might be prone to aggregation and subsequent fibril formation. Interestingly, the NaCl and KCl salts of group IA cations do not affect the secondary structure of the $A\beta$ fragment but accentuate intrapeptide hydrogen bonding and salt-bridge formation along with the stabilization of turn structures.

The N- and C-terminal capped peptide Ace-AEAAAKEAAAKA-Nme demonstrates diverse responses to very high salt concentrations (3–4 M) of different salts[10]. It loses its predominantly α-helical (71%) structure upon introduction of NaCl, although NaI proved to be a stronger denaturant. Interestingly, the potassium salts demonstrated negligible influence on the structural composition. The Na^+ cation possesses a far greater affinity to sidechain carboxylates and backbone carbonyls than K^+ cation, and thus it effectively attenuates the salt-bridges and the hydrogen bonds involved in secondary structure formation. Moreover, the comparatively bigger I^- anion has a strong affinity to small hydrophobic moieties such as nonpolar alanine, and therefore acts in sync with Na^+ cation in its destabilizing activity. In presence of the strongest denaturant NaI, the peptide is least hydrated, indicating towards a preferential solvation of the peptide backbone by the salt ions.

Transient salt-bridges observed in the Val-24–Lys-28 loop region of $A\beta_{21-30}$, involving Lys-28 and, Glu-22 or Asp-23, get stabilised in the presence of NaCl ions[11]. Further, at physiological

NaCl concentration (0.14 M), Aβ$_{42}$ demonstrates stable α-helical signatures along with transient 3$_{10}$-helix segments[12].

2.3.1.3. Effect on aggregation

RADA 16-I is a synthetic and amphiphilic peptide, consisting of four subsequent RADA repeats[13]. Subject to the surrounding solvent environment, this peptide can self-assemble in a controlled manner to generate fibrils and higher-order structures. Greatest aggregation or cluster formation was observed in 0.2 M NaCl solution, followed by in pure water. However, no clusters could be observed in solutions containing CaCl$_2$ salt. Increased aggregation in NaCl solution could be attributed to the transitional, expanded α-helical signatures observed. These structures might facilitate better inter-peptide interaction, and thus promote initial cluster formation and ultimately peptide self-aggregation. The α-helical structures might be an outcome of the screening of the charged amino acid residues by salt ions, with monovalent Na$^+$ cation playing a crucial role. Further, a stepwise assembly process was proposed, characterised by conformational changes from random coil or turn to α-helical intermediates, and culminating in β-scaffold structures. Lack of peptide clusters in CaCl$_2$ solution could be attributed to the divalent Ca^{2+} cation inducing unstructured peptide aggregates.

Larger Aβ$_{37-42}$ oligomers in increased proportion were observed in 2 M NaCl solution, as compared to pure water[14]. Further, NaCl not only amplifies the aggregation propensity of the polypeptide, but also alters the molecular mechanism of Aβ$_{37-42}$ self-association. Increased dimer formation and their subsequent aggregation is observed only in the presence of NaCl, but not in pure water systems. Moreover, even coarse-grained systems demonstrated faster assembly and increased stability of the diphenylalanine bilayer in 0.25 M NaCl solution, compared to pure water (Figure 2.6)[15]. Hence, NaCl salt could be used to modulate the overall fabrication process and achieve the precise design and desired properties in the targeted biomaterials.
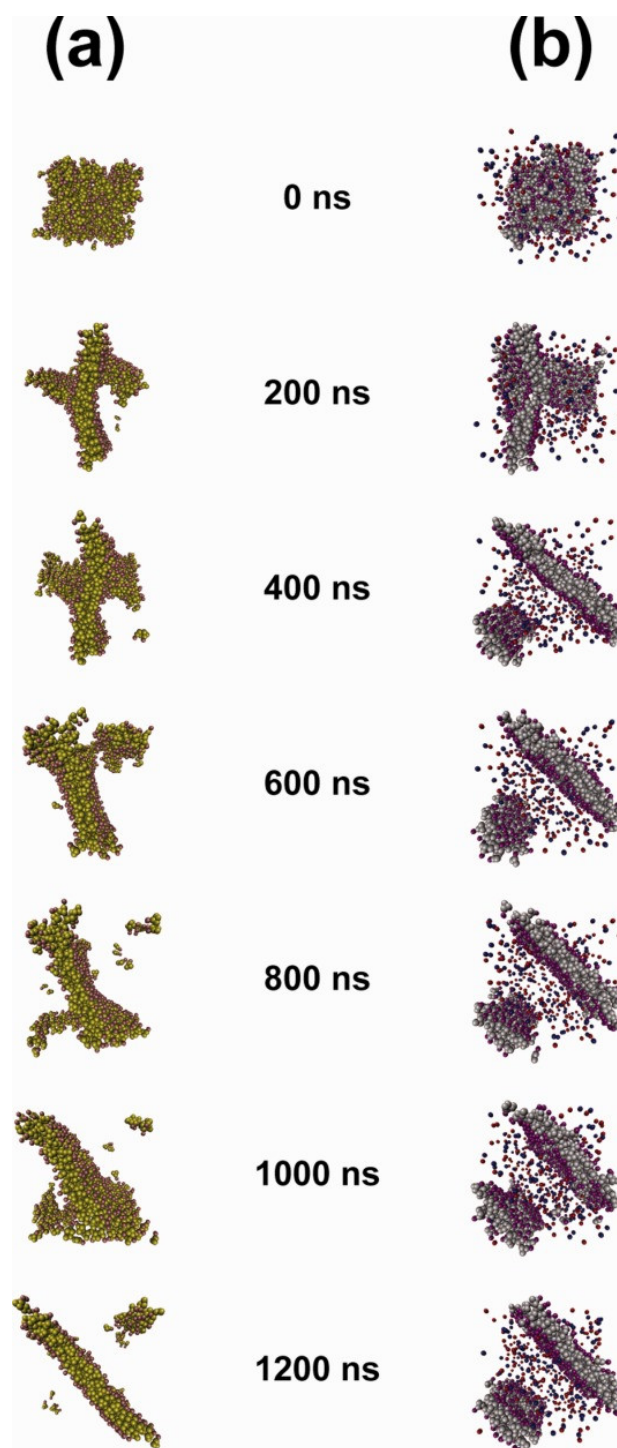
Figure 2.6. Simulation trajectories of two different coarse-grained molecular dynamic simulations. (a) Self-assembling simulation of diphenylalanine bilayer without sodium chloride (orange and yellow), (b) same simulation with sodium chloride ions (grey and magenta). Except for the existence of ions, all simulation conditions were identical[15].

2.3.1.4. Effect on lipid interaction

The binding of antimicrobial peptide magainin to POPC (palmitoyloleoyl phosphatidylcholine) lipid bilayers is stronger at lower NaCl concentrations[16]. This phenomenon could be attributed

to the sequential reduction in average area available per lipid head group with increasing NaCl concentration. The Na^+ cations strongly bind to the lipid ester oxygens, and thus cause denser packing of the lipids. Hence, the destabilizing effect of peptides on the lipids in immediate vicinity (within the bilayer) is diminished.

In experimental studies, presence of divalent Cu^{2+} or Zn^{2+} cations lead to the insertion of both $A\beta_{40}$ and $A\beta_{42}$ peptides into POPC and POPS (palmitoyloleoyl phosphatidylserine) bilayers within the pH range of 5.5–7.5[17]. However, membrane penetration could not be observed at higher pH because of changes in the Cu^{2+} coordination sphere. Significantly, in solvent environments devoid of $CuCl_2$ and $ZnCl_2$ salts, both peptides could enter the membranes only at pH < 5.5. Interestingly, increasing the cholesterol level to 0.2 mole fraction of the total lipid composition completely inhibited membrane insertion of both peptides under all conditions investigated. Further, peptides undergoing membrane insertion adopted α-helical structures, whereas those unable to penetrate the membrane formed β-sheet structures on the surface.

2.3.2. Mutations

2.3.2.1. Effect on structure

At physiological ionic strength (0.14 M), $A\beta_{42}$ wild-type (wt) adopts a partially folded but extended conformation[12]. Whereas the E22Q mutant attains a two-helix collapsed structure due to the clustering of hydrophobic residues. This might be attributed to the lack of a stable Val-24–Lys-28 loop in the E22Q mutant of $A\beta_{21-30}$ segment, as compared to the wild-type segment[11]. However, alternative turn structures stabilized by only a salt-bridge between Asp-23–Lys-28 were observed. Therefore, mutation-linked perturbations in Aβ folding pathway might lead to the formation of pathological structures.

Among the four mutations of $A\beta_{42}$ wt, viz. E22G, E22K, E22Q, and D23N, a reduction in α-helical propensity in $A\beta_{33-36}$ region was observed in all cases[18]. Further, E22K and E22Q mutations augment the helicity in $A\beta_{20-23}$ segment of the respective peptides, relative to the $A\beta_{42}$ wt peptide. Even though the α-helical propensity of peptides is inversely correlated with their tendency to aggregate, α-helical intermediates might be crucial in the process of amyloidogenesis. Further, structural modifications alter the aggregation process, and might lead to increased neurotoxicity of the aggregates. Whereas E22K and E22Q mutants aggregate more rapidly, E22G and D23N mutants demonstrate marginally slower aggregation than $A\beta_{42}$ wt. Moreover, for the $A\beta_{21-30}$ fragment, E22Q and K28A mutations increase the intra-peptide hydrophobic interactions, whereas the D23N mutation reduces them[19].

The A2V mutant demonstrates a fourfold increase in β-hairpin content along with a twofold reduction in intrinsic disorder, when compared with the Aβ$_{1-28}$ wt fragment[20]. This could also explain the higher aggregation propensity of A2V mutant relative to the Aβ$_{40}$ wt. However, due to appreciable differences in the conformational ensembles, an environment consisting of both Aβ$_{1-28}$ wt and Aβ$_{1-28}$ A2V peptides demonstrates lower aggregation and might protect against the deleterious effects of Alzheimer's disease.

2.3.2.2. Effect on aggregation

In hexamers of Aβ$_{16-35}$ wt, solvation of the buried Asp-23 and Lys-28 residues disrupted the amyloid-like oligomer structure, especially in the Glu22–Gly29 and Leu16–Ala21 regions[21]. However, in the case of E22Q/D23N and D23N/K28Q double mutants, and E22Q/D23N/K28Q triple mutant, the dehydrated neutral amide side-chains can form linear cross-strand hydrogen bond chains within mutant hexamers. These hydrogen bonds are also referred to as ''polar zippers'', and they can stabilize amyloid-like conformations. Oxidation of Met-35 decreases the β-strand content in Aβ$_{29-40}$ region, and additionally affects the secondary structure of the Aβ$_{33-35}$ region of Aβ$_{40}$ monomer[22]. Since such structural modifications could diminish Aβ fibril formation, oxidative stress might manifest its toxicity independent of peptide aggregation.

Amongst different mutations of α-synuclein (αs), ones incorporating an additional charge greatly influence the resultant amyloid polymorphs[23]. Further, mutations introducing more hydrophobic residues such as Q/T (A53T and H50Q) facilitate the early stages of αs aggregation. However, substitution with negatively-charged D/E residues (A53E and G51D) hinders the αs assembly due to electrostatic repulsion. Interestingly, replacing a negative charge with a positive one in E46K mutant augments the rate the aggregation. Since the proline residue preferentially adopts β-turn structure instead of β-sheet, the A30P mutation slows both the formation of oligomeric intermediates and subsequent fibrils. The A117V mutation in prion protein fragment PrP$_{113-120}$ facilitates aggregation by increasing the β-sheet and oligomeric contents[24]. The rise in intrinsic disorder due to the mutation is accompanied by an increase in inter-peptide backbone hydrogen bonding and sidechain hydrophobic interactions.

The aggregation of tumour suppressor p53 protein can cause cancer. For both p53$_{251-257}$[25] p53$_{252-258}$[26] fragments, I254R mutation acts as a suppressor of aggregation. Whereas the wild-type fragments demonstrate high aggregation propensity with characteristic β-sheet signatures, the I254R mutant displays a substantially reduced aggregation propensity, and preferentially adopts a disordered state. However, in a mixed system, the wild-type fragment can induce β-

sheet formation in the I254R mutant and can incorporate it into aggregates. Significantly, in a cross-interaction system, the wild-type p53[251–257] fragment has a greater propensity to interact with the I254R mutant than with itself[25]. Hence, while designing drugs to target protein aggregation, the drug molecule should not self-aggregate and be resistant to cross-interaction and subsequent aggregation with both target and other molecular species within the system.

The Transforming Growth Factor β-Induced Protein (TGFBIp) has been implicated in corneal dystrophies, where its aggregation and resulting fibril formation can adversely affect vision[27]. Most pathogenic mutations of TGFBIp are located in the fourth and ultimate fasciclin-1 (FAS1) domain. Although both A546T and R555W mutations affect the packing in the hydrophobic core of FAS1-4, only A546T mutant forms fibrils. Lack of a substantial population of fibrillar intermediates might explain the resistance of Arg-555 mutants to amyloid formation. In amyloid fibrils observed in the Alzheimer's disease, the diphenylalanine motif is a part of the central hydrophobic cluster in the fibrils[28]. Therefore, the diphenylalanine peptides were found to self-aggregate into twisted fibrils, with an inverse γ-turn structure. However, in similar experiments with tyrosine-substituted analogues, the peptides formed aggregates with parallel β-sheet structures, which later agglomerated as microspheres.

The human IAPP (hIAPP), implicated in Type II diabetes, forms fibrils *in vivo* and demonstrates cytotoxic effects towards the cultured pancreatic islet β-cells[29]. However, the rat IAPP (rIAPP), differing from hIAPP at only six amino acid residues in the central rIAPP[18–29] region, neither forms fibrils nor displays cytotoxicity. Single-residue substitutions in rIAPP sequence with amino acid residues from corresponding positions of the hIAPP sequence viz. R18H, L23F, or V26I, generated fibrils morphologically similar to the hIAPP fibrils, but with a relatively slower kinetics. Further, combination of two or three mutations increased fibril formation. The remaining three proline residues (Pro-25, Pro-28 and Pro-29) of rIAPP hinder the process of amyloidogenesis too and act in sync with the other three residues (Arg-18, Leu-23 and Val-26) in the rIAPP[18–29] region. Interestingly, the F15A and F23A mutations of hIAPP did not inhibit amyloid fibril formation, as the alanine substitution preserved the interactions in central and C-terminal regions of the peptide[30].

2.3.2.3. Effect on lipid interaction

The E22K and A21G mutants of native Aβ[11–28] fragments were studied in aqueous environments containing SDS micelle[31]. Whereas the E22K mutant attains a stable α-helix structure, glycine residue introduced by the A21G mutation causes the peptide to be more

flexible and thus adopt an unstructured conformation with a bend in the central region. Further, the E22K mutant demonstrates relatively stronger interaction with the anionic SDS micelle, as compared to the wild-type peptide and the A21G mutant. For hIAPP, the H18K mutation significantly alters the structure of the N-terminal region of the peptide, and hence diminishes its ability to disrupt the cell membrane of the pancreatic β-cells[30].

For the antimicrobial peptide (AMP) indolicidin, the P10A mutant demonstrates greater penetration into the DPC micelle than the wild-type peptide[32]. Further, even for the tryptophan residues in both peptides, deeper insertion into the hydrophobic core of the micelle was observed in case of the P10A mutant compared to the wild-type peptide. These observations can elucidate the relatively higher haemolytic activity of the P10A mutant than the native indolicidin, as its comparatively stronger interaction might disrupt the structural components of the cell membrane. The substitution of terminal VGR segment of protegrin-1 (PG-1) AMP with polar threonine (PC101), hydrophobic isoleucine (PC104), and negatively-charged glutamic acid (PC107) residues resulted in similar activities against Gram-negative bacteria, as compared with PG-1[33]. However, differences in cytotoxicity were observed, especially for PC107, which demonstrated significantly reduced toxicity towards the epithelial cells.

In case of ovispirin-1, the I10G (novispirin-G10) and I7T (novispirin-T7) mutants demonstrated reduced antimicrobial activity compared to the native peptide[34]. This phenomenon could be attributed to reduced hydrophobicity and altered structure of the mutants, relative to the wild-type peptide. The I7T mutation inhibits insertion of the peptide into the hydrophobic core of cell membrane. However, the I10G mutation introduces a bend in the central region of the peptide. Interestingly, this mutation still maintains the helical structure of the peptide, improves the binding of its positively-charged C-terminal motif to bacterial membranes, and yet diminishes its cytotoxicity towards host-cells. Piscidin 1 (Pis-1) is a potent AMP, but still demonstrates a low degree of haemolysis[35]. Both Pis-1 and its G8A/G13A double mutant (Pis-1AA) interact more strongly with the zwitterionic dioleoyl phosphatidylcholine (DOPC) membrane, as compared to the G8P mutant (Pis-1PG). This could be attributed to the introduction of a kink in the α-helical structure of Pis-1PG because of the proline residue, and the accompanying increase in flexibility of the peptide backbone (Figure 2.7)[35-36]. Significantly, in the absence of bilayer interaction, all the peptides in aqueous environment lose stability and unfold rapidly[35]. Further, Asp-4 to isoAsp-4 mutation of citropin 1.1 eliminates its pathogenicity towards Gram-positive *Staphylococcus aureus* and *Bacillus subtilis* bacteria[37].
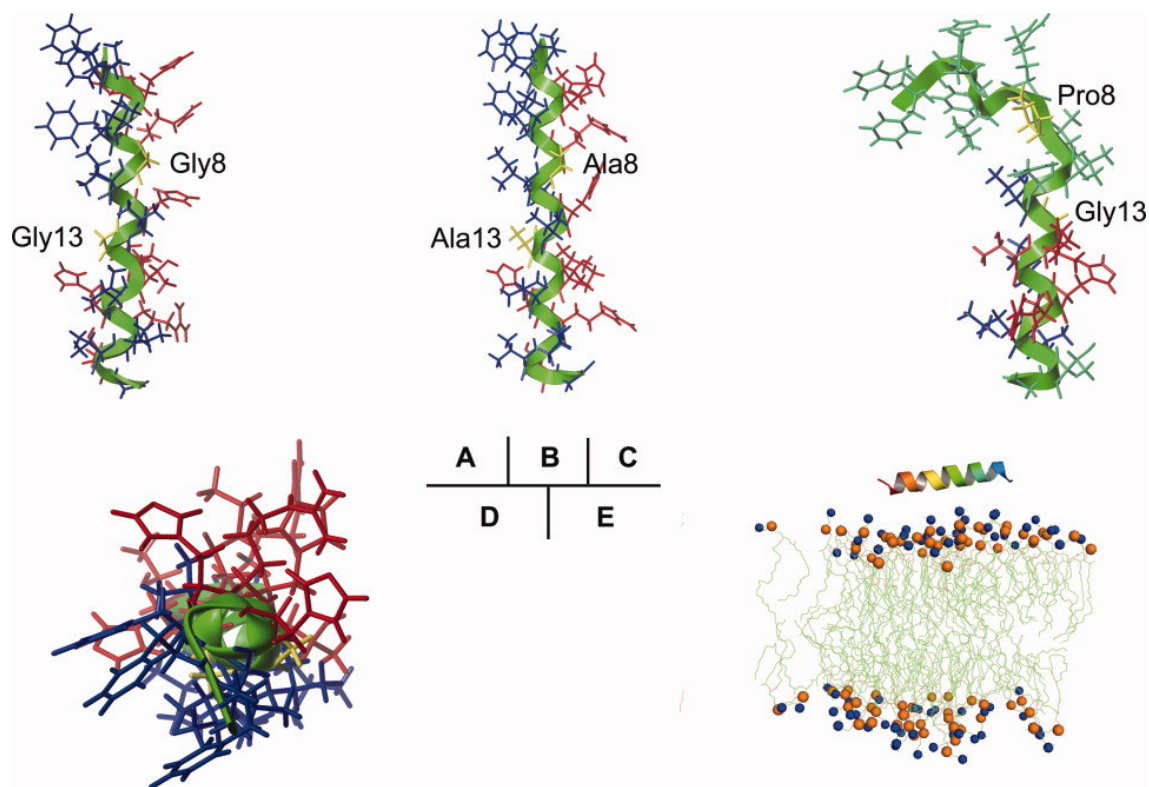
Figure 2.7. Structures of Pis-1 and its mutants. The backbones, hydrophobic, and hydrophilic sidechains are shown in green, blue, and red, respectively. Residues 8 and 13 are depicted with yellow. A: One of the NMR structures of Pis-1 in 300 mM SDS micelles. B: Mutated structure of Pis-1AA. C: Mutated structure of Pis-1PG. D: Head-on view of Pis-1. E: The diagram of piscidin-lipids model (Pis-1 paralleled with lipids)[35].

KR-12 is a small AMP consisting of residues 18–29 of the LL-37 peptide. It largely retains the antibiotic activity of LL-37 but demonstrates highly diminished cytotoxic and haemolytic activities[38]. R-KR12 and K-KR12 are two mutants of KR-12 with lysine residues mutated to arginine, and arginine residues mutated to lysine, respectively. None of these mutant peptides showed a stable interaction with the POPC (palmitoyloleoyl phosphatidylcholine) bilayer but demonstrated a strong interaction with the POPC/POPG (palmitoyloleoyl phosphatidylglycerol) bilayer. The latter bilayer additionally contains anionic lipids apart from the zwitterionic lipids, and this facilitates its initial interaction with the cationic peptides. Although these peptides remain unstructured in a purely aqueous environment, they adopt an amphipathic α-helical structure upon insertion into the headgroup region of bilayer. Due to the stronger cationic charge distribution of arginine compared to lysine, R-KR12 interacts more intensively with the POPC/POPG bilayer than K-KR12 and forms greater and stronger hydrogen bonds with it. This phenomenon corresponds to the observation that arginine-rich AMPs display higher toxicity towards both eukaryotic and prokaryotic cell lines than their

lysine-rich variants. Hence, R-KR12 demonstrates deeper insertion and subsequent disruption of the inner membrane of Gram-negative *Escherichia coli* bacterium than K-KR12.
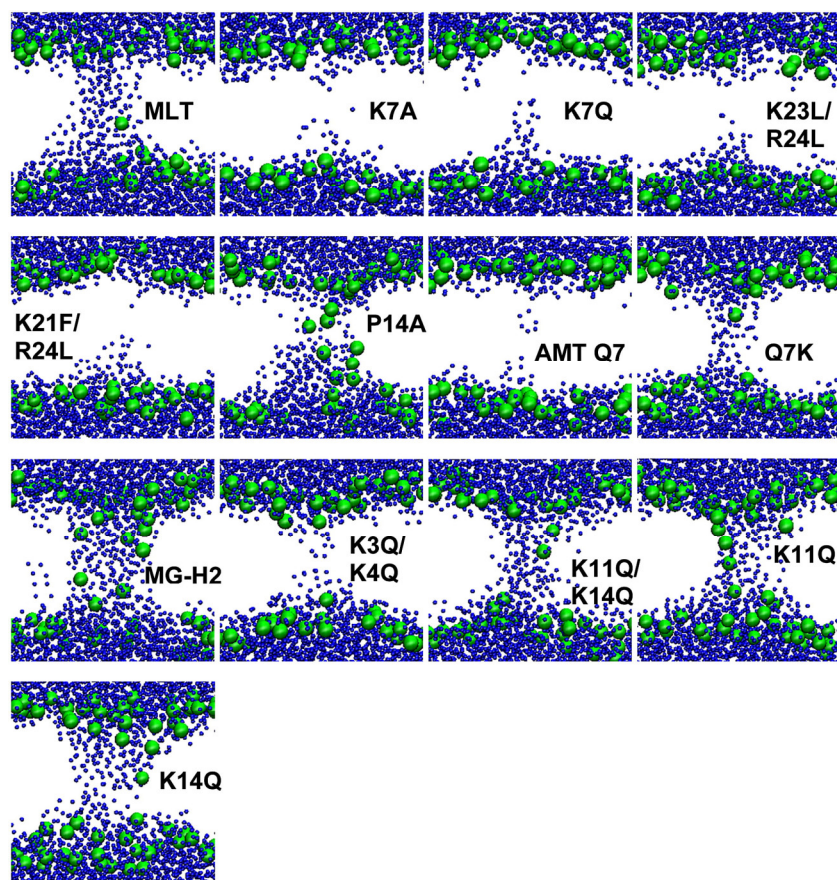
Figure 2.8. Pores at the end of MD simulations: melittin (MLT) at 140 ns, K7A at 216 ns, K7Q, K23L/R24L and K21F/R24L at 160 ns, P14A at 200 ns; alamethicin (AMT Q7) and Q7K at 160 ns; MG-H2 at 160 ns, K3Q/K4Q at 136 ns, and K11Q/K14Q, K11Q and K14Q at 160 ns. The peptides and ions are removed for clarity. The phosphocholines are shown as green balls and water as blue balls[36].

The net charge of a peptide, along with the distribution of charges, plays a crucial role in determining the pore shape[36]. MD simulations were performed with four monomers of different peptides and their mutants, placed in a pre-formed cylindrical pore within a DMPC bilayer, with the nonpolar side of the peptides towards the lipids. Upon substitution of Lys-7 in the N-terminal region of melittin with alanine (K7A) or glutamine (K7Q), the cylindrical pore remained stable and did not transform into a toroidal pore (Figure 2.8). To avoid electrostatically unfavourable interactions with the lipid head groups and the surrounding water molecules and ions, both K23L/R24L and K21F/R24L double mutations tilt the orientation of the respective peptides. This places the Lys-7 residue closer to the lower leaflet headgroups of the bilayer, and therefore inhibits the formation of a toroidal pore. Although the P14A mutant forms a toroidal pore, its structure differs from that of the wild-type melittin pore, and it is only

metastable. Thus, the positive-charge in the N-terminal region of the helical structure along with its imperfect amphipathicity are essential for melittin to form toroidal pores. Similarly, for MG-H2, whereas the wild-type peptide forms a toroidal pore, the K3Q/K4Q double mutant generates a semi-toroidal pore. On the other hand, although native alamethicin peptide prefers cylindrical pores, its Q7K mutant forms semi-toroidal pores. Therefore, cationic charges in the N-terminal region of the peptides might be critical in the formation of toroidal pores.

2.4. An introduction to molecular dynamics theory

2.4.1. Simulation algorithm

Molecular dynamics (MD) is a computational method that simulates the time evolution of a multi-particle system through the phase space, governed by the laws of Newtonian mechanics[39]. These simulations provide statistical predictions about the average behaviour of a macroscopic system. An MD simulation consists of different steps that are mentioned below:

1. Initialization: The parameters that specify the conditions of the simulation run, such as initial temperature, dimensions, density, boundary conditions, and time step, are defined at this stage.

2. Specification of initial atom parameters: The initial positions, velocities, and topologies of all the particles are defined.

3. Force-field selection: A force-field specifies the inter-atomic potentials, and along with the simulation parameters, defines the total potential energy of the system of interest. A detailed description of the force-fields can be found in section 2.4.2.

4. Force computation: Forces on all the particles in the system are computed. For a particle i at position r_i, the force F_i acting on it can be represented as $-\nabla_i U$, U being the total potential energy of the system.

5. Numerical integration: Forces calculated using the Newtonian equations of motion in the previous step are numerically integrated over the simulation time step with algorithms, such as, the velocity-Verlet algorithm.

6. Simulation loop: Steps 4 and 5 are repeated for the desired simulation time period until the required system characteristics have been achieved.

7. <u>Output</u>: Calculation of statistical values and rendition of trajectory snapshots using the simulation data, with intermediate post-processing steps as required.

2.4.2. Force-fields for molecular dynamics

In a typical MD simulation, the dynamics of a system of N particles evolves according to the Newtonian equations of motion[40]:

$$m_\alpha \ddot{\vec{r}}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{\text{total}}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \quad \alpha = 1, 2 \ldots N \tag{1}$$

Here, m_α is the mass of atom α, $\vec{r}_\alpha$ is its position, and U_{total} is the total potential energy, dependent on the positions of all the atoms and therefore, linked to their motion.

In a MD simulation setup, an individual molecule is generally defined as a series of charged points (atoms) linked by springs (bonds)[41-42]. Therefore, to describe the time evolution of bond lengths, bond angles and torsions, along with the non-bonding van der Waals and electrostatic interactions between the different atoms, a force-field is required. The force-fields consist of a collection of equations and associated constants modelled to reproduce molecular geometry and selected properties of pretested structures and systems. A force-field contains two principal components: (1) a functional form that defines the computation of the energies and forces on each particle of the system, and (2) a set of parameters that defines the relationship between the positions of the atoms, and their energy. Therefore, a force-field represents the total potential energy of the system.

The potential energy function, U_{total}, has the following contributions:

$$U_{\text{total}} = U_{\text{bonded}} + U_{\text{non-bonded}} \tag{2}$$

Here, U_{bonded} denotes the interactions between the chemically-bonded atoms. This energy component would be discussed in detail in section 2.5.1.1.

The interactions beyond chemically-bonded nearest neighbours are represented by the $U_{\text{non-bonded}}$ component, and it consists of two energy terms for van der Waals (vdW) and electrostatic (Coulombic) interactions, as shown below[40]:

$$U_{\text{vdW}} = \sum_i \sum_{j>i} 4\,\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] \tag{3}$$

$$U_{\text{Coulomb}} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{4}$$

Here ε_{ij}, σ_{ij}, and r_{ij} represent the well depth, the radius, and the distance between atoms i and j, respectively. The q_i and q_j are partial atomic charges on atoms i and j, respectively and, ε_0 denotes permittivity of free space. The vdW energy term is approximated by a Lennard–Jones (LJ) 6–12 potential, representing attractive and repulsive interactions, respectively.

2.4.3. Numerical integration

After the forces between the different particles have been calculated, numerical integration is performed on the Newtonian equations of motion over the specified time step using certain algorithms, such as Verlet algorithm.

For a particle within the system of interest, the Taylor expansion of its coordinate around time t can be represented as:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{r} + O(\Delta t^4) \tag{5}$$

Here, $f(t)$ and $v(t)$ denote the force acting on the particle of mass m, and its instantaneous velocity, respectively. The estimation of the particle's new position involves an error of the order Δt^4, where Δt denotes the time step of the MD simulation.

Similarly, the Taylor expansion of the particle's coordinate around time t can also be represented as:

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{r} + O(\Delta t^4) \tag{6}$$

The summation of Equations 5 and 6 results in:

$$r(t + \Delta t) + r(t - \Delta t) = 2r(t) + \frac{f(t)}{m}\Delta t^2 + O(\Delta t^4)$$

$$r(t + \Delta t) \approx 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 \tag{7}$$

It could be noted that the Verlet algorithm does not require the particle's velocity to compute the new position. Although, the velocity can be obtained from the particle's trajectory information:

36

$$r(t + \Delta t) - r(t - \Delta t) = 2v(t)\Delta t + O(\Delta t^3)$$

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + O(\Delta t^2) \tag{8}$$

The velocity expression involves an error of the order Δt^2.

After the calculation of new positions, the positions at the time $t - \Delta t$ can be discarded, and the numerical integration proceeds to the next time step.

The velocities generated can be used to calculate the kinetic energy of the particle, and therefore its instantaneous temperature. Further, the potential energy of the system is also calculated. For the Verlet-derived "velocity-Verlet" algorithm, the Taylor expansion of a particle's coordinate can be represented as:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 \tag{9}$$

However, compared to the Verlet algorithm, the velocities here are derived differently:

$$v(t + \Delta t) = v(t) + \frac{f(t + \Delta t) + f(t)}{2m}\Delta t \tag{10}$$

In the velocity-Verlet algorithm, the new velocities can be calculated only after the new positions have been obtained. Likewise, the new forces can be derived after the computation of the new velocities.

The Verlet and Verlet-derived methods are time reversible. Moreover, not only they conserve linear and angular momentum, but also just one force evaluation is required for each time step. Further, the total energy is constant throughout the simulation if the net linear momentum was zero at the onset of the simulation.

For large time steps, the accuracy of the calculations is very important because the longer the time step, the fewer evaluations of the forces are performed per unit of simulation time[39]. Thus, a sophisticated or higher-order algorithm is used that requires information about the higher-order derivatives of the particle coordinates. However, such algorithms require more storage due to the use of higher-order derivatives. Although higher-order algorithms tend to have very good energy conservation for short time periods, often the overall energy drifts over long time scales. Further, higher order Runge-Kutta type methods are not suitable for biomolecular simulations because they require several force evaluations for each time step, thereby increasing the duration of the corresponding MD simulations[40].

In contrast, Verlet-derived algorithms usually have only moderate short-term energy conservation but little long-term drift[39]. It should be noted that no algorithm has been formulated to date that can accurately predict the trajectory of all particles for both short and long time periods. The Verlet-derived algorithms are much faster than the sophisticated algorithms, even though forces on all the particles must be computed more frequently due to the smaller time steps required for these lower-order algorithms. Also, they need comparatively lesser memory storage than higher-order algorithms.

2.4.4. Simulation ensembles

In an MD simulation, a certain simulation environment or ensemble is usually maintained, and the various ensembles differ in the system variables that are regulated or conserved[43-44]. All these ensembles are characterised by a thermodynamic potential, that provides information about the most stable state of the system[45]. Some of the most commonly used ensembles are[43-44]:

1. Microcanonical or NVE: The simulations are characterised by a constant number of particles (N), constant volume (V), and constant energy (E). The sum of kinetic and potential energy is conserved and, the temperature and pressure are unregulated. The thermodynamic potential is the entropy (S):

$$Q_{NVE} = \frac{1}{N!\,h^{3N}} \int d\mathbf{r} d\mathbf{p}\ \delta\left(H(\mathbf{r},\mathbf{p}) - E\right)$$

$$S = k_{\mathrm{B}} ln Q_{NVE} \tag{11}$$

Here h and k_{B} denote the Planck's and Boltzmann's constants, respectively. Q_{NVE} is the partition function. $H(\mathbf{r},\mathbf{p})$ is the system Hamiltonian, with $\mathbf{r}$ and $\mathbf{p}$ representing the particle position and velocity, respectively. δ function selects the states where the total energy is E.

2. Canonical or NVT: The simulations are characterised by a constant number of particles (N), constant volume (V), and constant temperature (T). The temperature is regulated via a thermostat and the pressure is unregulated. The thermodynamic potential is the Helmholtz free energy (A):

$$Q_{NVT} = \frac{1}{N!\,h^{3N}} \int d\mathbf{r} d\mathbf{p}\ \exp\left(\frac{-H(\mathbf{r},\mathbf{p})}{k_{\mathrm{B}}T}\right)$$

$$A = -k_{\mathrm{B}} T ln Q_{NVT} \tag{12}$$

Here, Q_{NVT} is the partition function.

3. Isothermal-isobaric or NPT: The simulations are characterised by a constant number of particles (N), constant pressure (P), and constant temperature (T). The temperature is regulated via a thermostat and the pressure via a barostat. The thermodynamic potential is the Gibbs free energy (G):

$$Q_{NPT} = \frac{1}{N!\, h^{3N} V_0} \int d\mathbf{r} d\mathbf{p} \ \exp\left(\frac{-H(\mathbf{r}, \mathbf{p}) + PV}{k_\mathrm{B} T}\right)$$

$$A = -k_\mathrm{B} T ln Q_{NVT} \tag{13}$$

Here, Q_{NPT} is the partition function and V_0 is the initial volume.

4. Grand-canonical or µVT: The simulations are characterised by a constant chemical potential or molar Gibbs free energy of the selected components (µ)[45], constant volume (V), and constant temperature (T). The temperature is regulated via a thermostat and the pressure is unregulated. The thermodynamic potential is the grand potential (Ω):

$$Q_{\mu VT} = \sum_N \frac{1}{N!\, h^{3N}} \exp\left(\frac{-\mu N}{k_\mathrm{B} T}\right) \int d\mathbf{r} d\mathbf{p} \ \exp\left(\frac{-H(\mathbf{r}, \mathbf{p})}{k_\mathrm{B} T}\right)$$

$$\Omega = -k_\mathrm{B} T ln Q_{\mu VT} \tag{14}$$

Here, $Q_{\mu VT}$ is the partition function.

The NVE ensemble provides the real solution to the N-body problem, since it conserves the total energy of the simulation system. However, under experimental conditions, NVE ensemble is usually not applicable and therefore, NVT, NPT and µVT ensembles are used as necessary.

2.4.4.1. Thermostat

Constant temperature of the simulation system can be maintained using different schemes. For example, in the isokinetic MD scheme, the average kinetic energy per particle is kept constant[44]. Whereas in the velocity-scaling methods, the velocities of the particles are scaled according to the set temperature. The constant temperature can also be maintained by coupling the system to an external heat bath. The Andérsen method implements this coupling by applying stochastic impulsive forces on arbitrarily selected particles within the system. For the Nosé-Hoover thermostat, a fictious degree of freedom, consisting of both potential and kinetic energy components, is added to the system to control its temperature[46].

The constant temperature can also be maintained using Langevin dynamics[40]. This method uses a stochastic coupling, where two force terms are added to the Newton's second law in Equation 15 to account for the neglected degrees of freedom. Not only it is easier to implement, but the friction terms tend to enhance the dynamical stability of the system. Hence, Langevin dynamics can search conformations better than the Newtonian molecular dynamics. The typical Langevin equation is:

$$M\dot{v} \; = \; F(r) - \gamma v + \sqrt{\frac{2\gamma k_{\text{B}} T}{M}} \; R(t)$$

(15)

Here M is the mass, v is the velocity, F is the sum of all forces, r is the position, γ is the friction coefficient, k_{B} is the Boltzmann's constant, T is the temperature, and $R(t)$ is a univariate Gaussian random process. Coupling to the thermal reservoir is modelled by the fluctuating or random force (the last term), and the dissipative or frictional force ($-\gamma v$ term). Since friction opposes the motion, the first additional force is proportional to the particle's velocity but oppositely directed.

2.4.4.2. Barostat

For a multi-particle system, the constant pressure can be maintained using one of the various pressure-control methods or barostats. In the Andersén method, a fictious degree of freedom with a mass W (in units of mass × length^{-4}), and linked to the volume of the cubic simulation cell, acts as a piston[47]. This fictional piston equalizes the internal pressure of the system and the applied pressure by undergoing self-adjustment. The Parrinello-Rahman method also uses a fictious mass to control the rate of pressure adjustment[48]. However, unlike the Andersén method, this method permits variation in the shape of the simulation cell (allows non-cuboidal shapes).

A modified Nosé-Hoover method[49] can also be employed to maintain constant pressure during the simulation, in which Langevin dynamics is used to control the piston fluctuations in the barostat to allow partial damping[50]. In the Nosé-Hoover method, not only the isotropic volume fluctuations are maintained, but like the Parrinello-Rahman method, it is also ensured that the simulation cell is totally flexible[49]. The Langevin piston Nose-Hoover method is usually combined with a method of temperature control, such as Langevin dynamics, to simulate the NPT ensemble[40]. The equations of motion for the particles, along with the volume of the (cubic) simulation system, can be represented as[50]:

$$\dot{\boldsymbol{r}}_i = \frac{\boldsymbol{p}_i}{m_i} + \frac{1}{3}\frac{\dot{V}}{V}\boldsymbol{r}_i \qquad (16)$$

$$\dot{\boldsymbol{p}}_i = \mathbf{f}_i + \frac{1}{3}\frac{\dot{V}}{V}\boldsymbol{p}_i \qquad (17)$$

$$\ddot{V} = \frac{1}{W}[P(t) - P_{\text{ext}}] - \gamma\dot{V} + R(t) \qquad (18)$$

Here, V denotes the volume. $P(t)$ and P_{ext} represent the instantaneous and imposed pressures, respectively. Further, $\boldsymbol{r}$, $\boldsymbol{p}$, m, and $\mathbf{f}$ denote the position, momentum, mass, and force, respectively, for each particle in the system. The collision frequency and the random force are represented by γ and $R(t)$, respectively.

$$\langle R(0)R(t)\rangle = \frac{2\gamma k_{\text{B}}T\delta(t)}{W} \qquad (19)$$

$R(t)$ is derived from a Gaussian distribution with a zero mean and variance. Here, k_{B} and W denote the Boltzmann's constant and piston mass, respectively.

2.4.5. Periodic boundary conditions

In order to avoid surface effects at the boundary of a multi-particle system, and thus to mimic an infinite bulk environment, periodic boundary conditions are employed in the MD simulations[40, 51]. The defined volume containing all the particles is taken as a cell and replicated to an infinite lattice of identical cells by periodic translations. Any particle in the system that leaves the cell on one side is substituted by a copy entering the cell on the opposite side. Further, each particle is subject to interactions with all other particles in the system, including images in the surrounding cells. All the cells are identical to each other, and the image particles across different cells move together. It should be noted that although the origin of the periodic lattice can be chosen anywhere, the shape and orientation of the periodic cell is kept fixed. Periodic boundary conditions eliminate surface effects, but effects related to the finite-size of the periodic cell are still present.
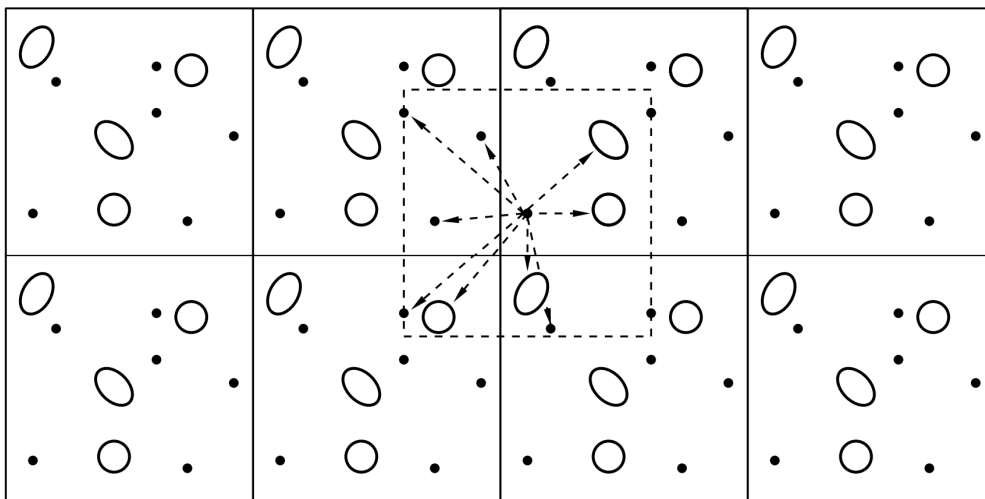
Figure 2.9. Schematic representation of periodic boundary conditions[51].

The van der Waals and electrostatic interactions exist between every nonbonded pair of atoms in the system (including those in neighbouring cells)[40, 51]. Hence, reliable computation of long-range interactions with periodic boundary conditions is not theoretically feasible because the calculation of the potential energy would involve an infinite sum of interactions to simulate the bulk behaviour. However, most interatomic interactions are short-ranged and can be truncated beyond a certain cut-off distance. Thus, the van der Waals interaction is spatially truncated, generally using a smoothing function, at a specified cut-off distance. The cut-off is set to be less than half the diameter of the periodic box to avoid double force calculations. Periodicity of the system plays an important role in the computation of nontruncated electrostatic interaction with minimal additional computational cost using the particle-mesh Ewald (PME) method.

2.4.6. Particle-mesh Ewald electrostatics

Ewald summation is used to calculate the long-range electrostatic interactions for the defined volume containing a multi-particle system, with periodic boundary conditions[40]. Compared to spatial truncation at a certain user-specified cut-off distance, the Ewald summation is considered more reliable. The Ewald sum consists of the following four terms:

$$U_{\text{Coulomb}} = U_{\text{Ewald}} = U_{\text{dir}} + U_{\text{rec}} + U_{\text{self}} + U_{\text{surface}} \qquad (20)$$

Here, U_{dir}, U_{rec}, U_{self}, and U_{surface} represent the direct sum, reciprocal sum, self-energy, and surface energy components, respectively[40]. The self-energy term is an inconsequential constant. Further, the surface energy term is generally not considered as ε_s is assumed to be infinity due to the high dielectric constant of water ($\varepsilon_s \approx 80 \gg 1$). Therefore, the N number of

point charges in the system, in Figure 2.10, can be represented as a set of screened charges whose interactions are calculated by direct summation, and the residual smoothly varying screening background is handled by the reciprocal sum component of the Ewald summation[52].

Direct sum (Screened point charges)

=

+

Ewald sum (Point charges)

Reciprocal sum (Gaussian screening background)

Charge distribution
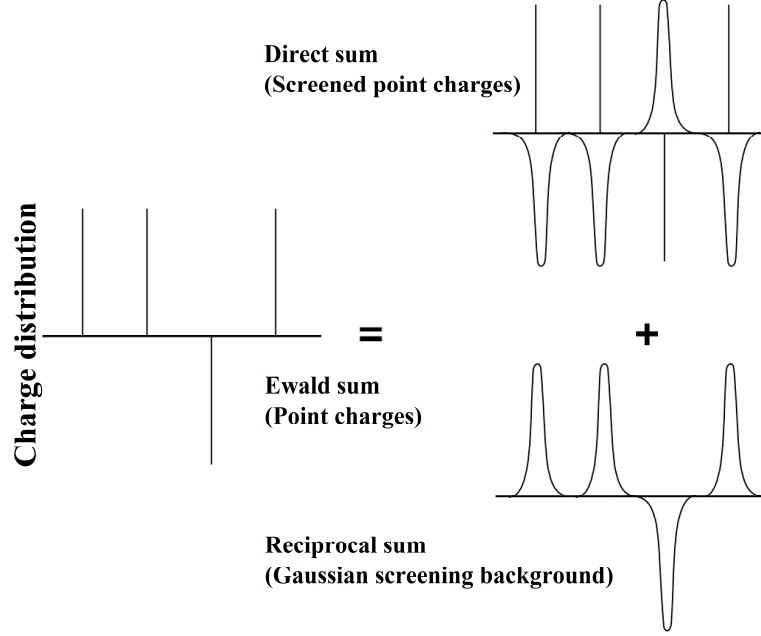
Figure 2.10. A set of point charges may be considered as a set of screened charges and a smoothly varying screening background[52].

After dropping the prefactor $1/4\pi\varepsilon_0$, the terms of Ewald summation can be represented as[40]:

$$U_{\mathrm{dir}} = \frac{1}{2} \sum_{i,j=1}^{N} q_i q_j \sum_{\vec{n}_r}{}' \frac{\mathrm{erfc}\big(\beta|\vec{r}_i - \vec{r}_j + \vec{n}_r|\big)}{|\vec{r}_i - \vec{r}_j + \vec{n}_r|} \qquad - \sum_{(i,j)\,\in\,Excluded} \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j + \vec{v}_{ij}|} \tag{21}$$

$$U_{\mathrm{rec}} = \frac{1}{2\pi V} \sum_{\vec{m}\neq 0} \frac{\exp(-\pi^2 |\vec{m}|^2/\beta^2)}{|\vec{m}|^2} \left| \sum_{i=1}^{N} q_i \exp(2\pi i \vec{m}\cdot\vec{r}_i) \right|^2 \tag{22}$$

$$U_{\mathrm{self}} = -\frac{\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \tag{23}$$

$$U_{\mathrm{surface}} = \frac{2\pi}{(2\varepsilon_s + 1)V} \left| \sum_{i=1}^{N} q_i \vec{r}_i \right|^2 \tag{24}$$

Here, q_i and $\vec{r}_i$ are the charge and position of atom i, respectively, and $\vec{n}_r$ is the lattice vector. For any simulation box defined by three independent base vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_3$, $\vec{n}_r = n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3$, where n_1, n_2, and n_3 are integers. Σ' denotes a summation over $\vec{n}_r$ that excludes the $\vec{n}_r$

= 0 term in case $i = j$; "excluded" denotes the set of atom pairs whose direct electrostatic interaction should be excluded. $\vec{v}_{ij}$ denotes the lattice vector for a (i, j) pair, that minimises $|\vec{r}_i - \vec{r}_j + \vec{v}_{ij}|$. β is a parameter adjusting the computational load between the direct sum and reciprocal sum terms. ε_s is the dielectric constant of the surroundings of the simulation box, which is water for most biomolecular systems. V is the volume of the simulation box. $\vec{m}$ is the reciprocal vector defined as $\vec{m} = m_1\vec{b}_1 + m_2\vec{b}_2 + m_3\vec{b}_3$, where m_1, m_2, m_3 are integers, and the reciprocal base vectors $\vec{b}_1$, $\vec{b}_2$, $\vec{b}_3$ are defined such that

$$\vec{a}_\alpha \cdot \vec{b}_\beta = \vec{\delta}_{\alpha\beta}, \qquad \alpha, \beta = 1, 2, 3 \tag{25}$$

The complementary error function $\text{erfc}(x)$ in Equation 21 is

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt \tag{26}$$

The infinite summation of charge-charge interactions for a charge-neutral system is dependent on the order of calculation[40]. For Ewald summation, this order can be specified as follows: first sum over each box, followed by sum over spheres of boxes of increasingly larger radii. However, the artificial periodicity involved in this summation can bias the free energy dynamics and for example, erroneously stabilise a protein structure that should have unfolded quickly. Moreover, this summation scheme is computationally expensive for large systems, because the computational cost scales as $N^{3/2}$, N being the number of charges in the system[52].

For intermediate-sized systems ($N \approx 10^3$–10^4), the relatively faster particle-mesh Ewald (PME) summation is an appropriate option[52]. However, to improve the accuracy of the particle-mesh method, the short-range and the long-range interactions are calculated separately[40]. The short-range contribution is directly derived from the particle-particle interactions, whereas the particle-mesh method is applied for the long-range component. In PME, the β parameter is calibrated such that the major computational load is directed towards the reciprocal sum, while the direct sum is calculated at a computational cost proportional to N. Using an interpolation scheme, PME distributes the charges present in the system to the nodes of a uniform grid to generate a discretised Poisson equation[40, 52]. Then, this equation is solved using the Fast Fourier Transform (FFT) to obtain the reciprocal sum. Since the computational load of PME scales to $N \log N$, it provides a major reduction in the simulation time, compared to the simple Ewald summation. For full electrostatic computations, the smooth PME (SPME) method is used where B-spline functions serve as the basis functions for charge interpolation[40]. The continuity

of B-spline functions and their derivatives reduces the computational time required to obtain an analytical expression of forces, and thus decreases the number of FFTs by approximately half, compared to the original PME method. Hence, SPME enables faster calculation of the reciprocal sum, and therefore a reduction in computational cost; resulting in a faster simulation.

2.4.7. NAMD

The Nanoscale Molecular Dynamics (NAMD) program is a popular MD simulation package suited for parallel computation and high-performance simulations of large biomolecular systems[40]. We have used NAMD for simulating the different systems in this study. Further, NAMD is compatible with the CHARMM force-fields employed in our simulations[41], discussed in detail in section 2.5.1.1. NAMD employs a multiple-time-stepping method to improve the integration efficiency[40]. It computes the slower-varying forces less frequently than the faster-varying ones in MD simulations. This stratagem is implemented in NAMD by three levels of integration loops. The inner loop uses only bonded forces to advance the system, the middle loop uses LJ and short-range electrostatic forces, and the outer loop uses long-range electrostatic forces.

2.5. Computational models and methods

2.5.1. Atomistic models

Atomistic models provide the greatest insight, but are accompanied by a massive computational cost[53]. Hence, they are limited to the study of monomeric and very small oligomeric complexes and analysing the stability of preformed fibril models and their interaction with dyes, small molecules, or peptide inhibitors.

The initial partially folded, aggregation-prone structures and prenucleus assemblies, present during the early stages of aggregation, are transient and unstable entities[53]. Hence, these are difficult to detect experimentally, and atomistic simulations play a crucial role in their study. Structural differences at monomeric level have been observed between the $A\beta_{40}$ and $A\beta_{42}$ alloforms. The information gleaned from the transient secondary structures could help in the identification of protein regions critical for initiating aggregation. Further, atomistic simulations have been used to study the growth phase of fibril formation and the role of point mutations in modulating aggregation. Also, the identification of aggregation-prone structures from a diverse family of existing structures has been achieved using these simulations.
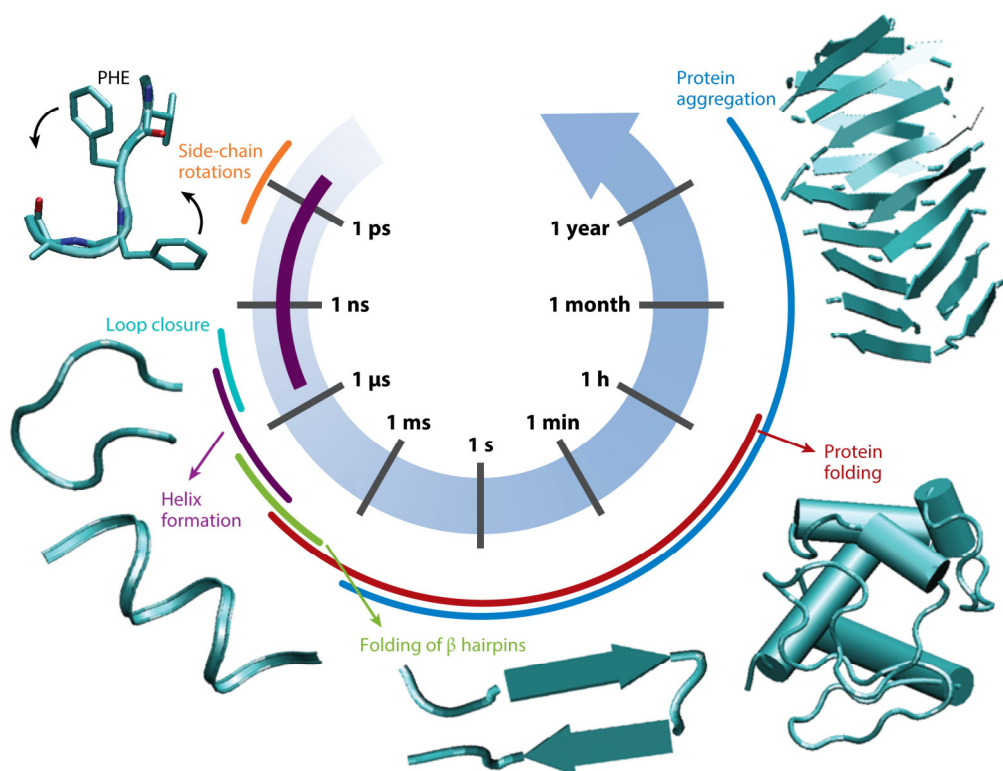
Figure 2.11. The contrasting breadth of timescales of protein rearrangement and assembly[53].

Simulations of IAPP peptide revealed significant differences in the monomeric structure, with the aggregating variants (e.g., human IAPP) generating both compact and extended conformations, and the nonaggregating variants (e.g., rat IAPP) showing only compact structures[53]. Atomistic simulations have been used to refine the fibril structures of Aβ and IAPP obtained from solid-state NMR. Moreover, it was observed that fibrils with a cross-β structure could act as template for the formation of additional β structure in Aβ monomers. Further, it was revealed that the inhibitory role of Congo Red (dye) in blocking fibril extension could be attributed to a new binding site in the molecule, which gets bound to the edge of the fibril, and this binding site is not observed for thioflavin T (dye) and its radioactive derivative Pittsburgh compound B (PiB).

Different atomistic force-fields can lead to relatively different secondary structure predictions, especially for the intrinsically disordered peptides such as Aβ and IAPP[53]. Since most force-fields have been parameterized on the basis of folded motifs, they may need reoptimization to better account for the unfolded/partially folded nature of the intrinsically disordered peptides.

2.5.1.1. CHARMM force-field

The CHARMM (or Chemistry at Harvard Macromolecular Mechanics) force-field is one of the most widely used force-field for biological systems. The potential function of a force-field consists of chemically-bonded and nonbonded energy terms, as discussed in section 2.4.2.

The interactions between chemically-bonded nearest neighbours can be represented by the bonded or intramolecular energy terms, as shown below[54]:

$$U_{\text{bonded}} = U_{\text{bonds}} + U_{\text{angles}} + U_{\text{dihedrals}} + U_{\text{improper}} + U_{\text{UB}} \qquad (27)$$
$$\phantom{U_{\text{bonded}} = U_{\text{bonds}} + U_{\text{angles}} + U_{\text{dihedrals}} + U_{\text{improper}}}{\text{dihedrals}}$$

$$U_{\text{bonds}} = \sum_{\text{bonds}} K_b (b - b_0)^2 \qquad (28)$$

$$U_{\text{angles}} = \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 \qquad (29)$$

$$U_{\text{dihedrals}} = \sum_{\text{dihedrals}} K_\phi \left(1 + \cos(n\phi - \delta)\right) \qquad (30)$$

$$U_{\substack{\text{improper} \\ \text{dihedrals}}} = \sum_{\substack{\text{improper} \\ \text{dihedrals}}} K_\varphi (\varphi - \varphi_0)^2 \qquad (31)$$

$$U_{\text{UB}} = \sum_{\text{UB}} K_{\text{UB}} \left(r_{1,3} - r_{1,3;0}\right)^2 \qquad (32)$$

Here, U_{bonds}, U_{angles}, $U_{\text{dihedrals}}$, $U_{\text{improper dihedrals}}$, and U_{UB} terms represent the covalent bond stretching, angle bending, dihedral, improper dihedral, and Urey-Bradley energy terms, respectively. Further, b_0, θ_0, φ_0, and $r_{1,3;0}$ are the bond, angle, improper and Urey-Bradley equilibrium values, respectively. The K's are the corresponding force constants, and n and δ are the dihedral multiplicity and phase, respectively. The term "dihedral" describes atom pairs separated by exactly three covalent bonds, with the central bond subject to the torsion angle. The "improper dihedral" term defines the geometry of four planar, covalently-bonded atoms[40]. The Urey-Bradley term is crucial for the in-plane deformations as well as for separating the symmetric and asymmetric bond stretching modes (for example, in aliphatic molecules)[41].

2.5.2. Coarse-grained models

Minimum timescales and length scales needed to study the aggregation process are typically higher than those required for protein folding studies (Figures 2.11 and 2.12)[53]. Hence, although atomistic simulations can capture the initial stages of aggregation, access to the total assembly process (from monomers to fibril) is not computationally feasible at present. The structural similarity between amyloid fibrils generated by diverse peptides supports the use of simplified peptide models that neglect certain molecular details but retain the crucial physical aspects of aggregation. Moreover, such models could also be employed to study the interaction of these peptides with membranes and other lipid structures and determine their influence on peptide aggregation.
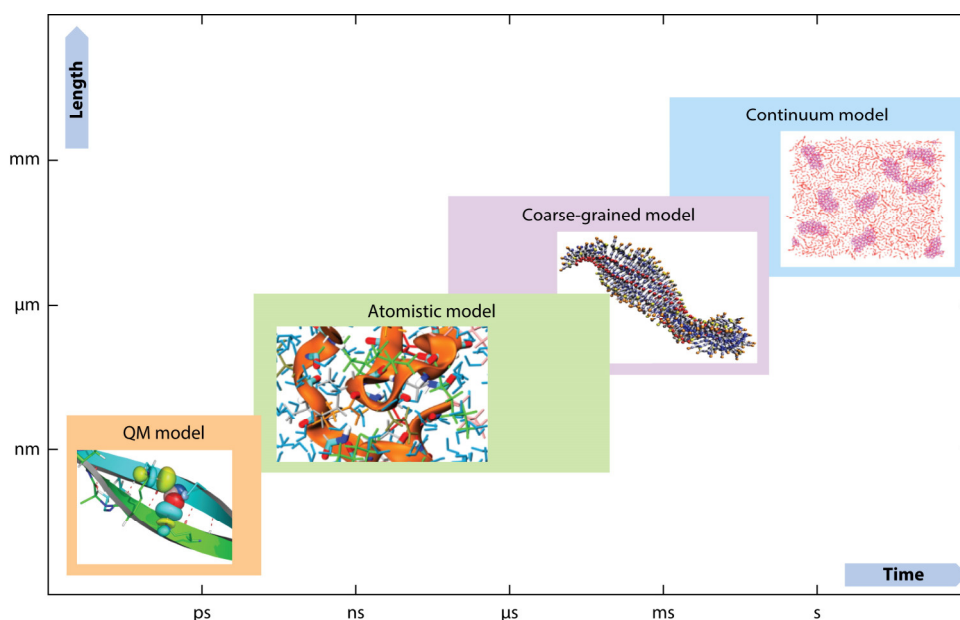
Morris-Andrews A, Shea J-E. 2015.
Annu. Rev. Phys. Chem. 66:643–66

Figure 2.12. The approximate timescales involved in different classes of molecular simulations: quantum mechanical (QM), atomistic, coarse-grained and continuum models[53].

Coarse graining entails trade-off between physical accuracy and computational efficiency[53]. Thus, the degree of coarse graining should depend on the smallest important length scale of the system of interest. Because of the extreme range of length scales and timescales involved, this is determined *a priori,* and an appropriate model is selected accordingly.

2.5.2.1. Lower-resolution models

Highly coarse-grained models forego sequence-level resolution[53]. However, tetrahedral, tube and cuboid models have been used to study the nucleation growth mechanism. By restricting

the moves to small increments, dynamic Monte Carlo method has been used with low-resolution models to achieve physically relevant kinetics. Coarse graining could also extend beyond the molecular length scale. For instance, the fibril itself could be modelled as a chain of coarse-grained beads. Such mesoscale model procures the elastic parameters from implicit water fully-atomistic simulations.
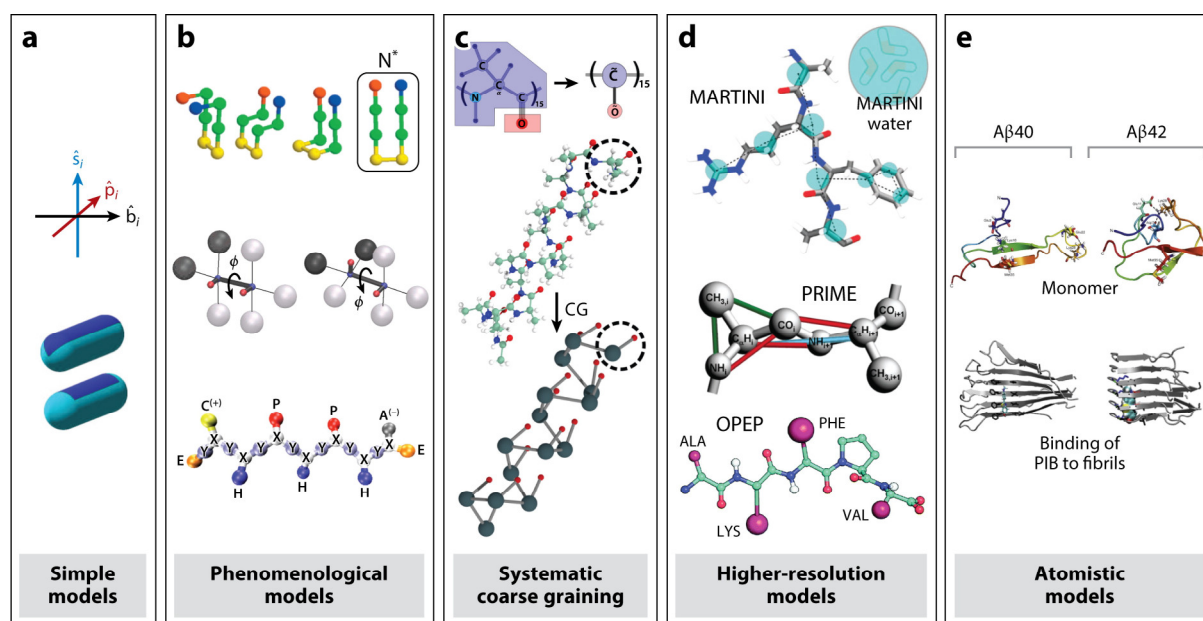
Morris-Andrews A, Shea J-E. 2015.
Annu. Rev. Phys. Chem. 66:643–66

Figure 2.13. Different resolution models for the study of protein aggregation, from coarse-grained to atomistic. (a) Simple models: the orientable stick model and the sphero-cylindrical model. (b) Phenomenological models: the lattice model, Caflisch model, and Shea model. (c) Systematic coarse graining: a coarse-grained polyalanine chain. (d) High-resolution models: the MARTINI model, PRIME model, and OPEP model. (e) Atomistic models of the Alzheimer amyloid-β peptide: monomers from replica exchange molecular dynamics simulations, and PIB bound to fibrils[53].

2.5.2.2. Phenomenological models

Coarse graining over atomic length scales, with one or more beads representing an amino acid, have been able to elucidate the physical properties of a protein that are crucial for aggregation[53]. However, these models typically use only generic amino acid types, such as charged, polar, hydrophobic, or neutral. The low-resolution lattice model uses a single bead to describe an amino acid and restricts the allowed coordinates to a cubic lattice (Figure 2.13b). Multi-bead description of the amino acids can identify the aggregate-prone conformations. In conjunction with implicit solvent models, these phenomenological models have been able to demonstrate the correlation between the peptide β-sheet propensity and the kinetics of fibril

formation. Although peptides with high β-sheet propensity formed fibrils via an ordered β-sheet nucleus, peptides with lower β-sheet propensity first formed disordered oligomers, and then the β structure emerged from this.

The mid-resolution Caflisch model utilizes two beads per residue[53]. It has been used to elucidate the crucial role played by peptide concentration, external conditions, and membranes surfactants in the fibril formation process. With this model, Phe-Phe, GNNQQNY, transthyretin, and Aβ_{40} were identified as highly amyloidogenic proteins, whereas Aβ_{42}, Sup35, prion proteins, and myoglobin were weakly amyloidogenic. The Shea peptide model uses three beads per residue: two for the backbone and one for the sidechain. In addition to the nucleation step, this model can also be used to study fibril growth.

2.5.2.3. MARTINI force-field

The MARTINI force-field is the most popular coarse-grained atom model[53]. It generally combines four heavy atoms into a single bead (Figure 2.13d). Likewise, MARTINI water model incorporates four water molecules into one bead. Although the bonded interactions are matched to an all-atom reference, the nonbonded interactions are based on the chemistry of the respective united atoms (i.e., their charge, polarity, and hydrogen bonding propensities). Since the nonbonded potentials are modelled on distinguishable chemical properties, instead of being system-specific, the MARTINI model can be optimized for all-atom simulations of water, lipids, proteins, and nucleic acids. However, this model neglects details of the peptide backbone, which are crucial to the study of protein aggregation. Nevertheless, the MARTINI model has been able to capture the amyloid assembly process, rapidly screen aggregation-prone sequences, and elucidate the effect of membranes and fluid interfaces on protein aggregation.

2.5.2.4. PRIME model and discontinuous molecular dynamics

PRIME model typically employs discontinuous molecular dynamics (DMD) simulations, which permit discontinuous breaks in the energy functional by computing the reflection or transmission of particles across the discontinuity[53]. This model reduces the number of particles to four per amino acid (Figure 2.13d) and can capture the hydrogen bonding energetics. Although the PRIME model reduces the number of interaction parameters, but each parameter is still physically relevant. Moreover, it also retains structural discrimination.

Using this model, it was observed that sequences containing long stretches of hydrophobic residues prefer a disordered, collapsed state, instead of a fibrillar structure[53]. For the $A\beta_{16-22}$ fibrils, nucleation growth mechanism was observed at lower temperatures, but templated assembly at higher temperatures. Apart from PRIME model, the DMD simulations have been used to study peptide aggregation with the tetrahedral protein model. In this model, the proteins were represented as a chain of hard spheres centred on the C_α atoms, with sequence-dependent hydrogen bonding.

2.5.2.5. OPEP model

OPEP is a high-resolution coarse-grained model with a high degree of chemical specificity[53]. Although the sidechain is represented as a single coarse-grained bead, the backbone heavy atoms are modelled with full atomic resolution (Figure 2.13d). Therefore, no artificial constraints are imposed on the protein secondary structure. This model has been used to elucidate the dimerization of various alloforms of Aβ. Further, it was determined that different drugs not only exhibit differences in their binding modes, but also may have varying effectiveness at different stages of Aβ oligomerization. The OPEP model was also used to study the early stages of oligomerization of yeast prion Sup35 fragments.

2.5.3. Computational methods

2.5.3.1. Systematic coarse graining

Systematic coarse graining provides a bottom-up methodology to obtain the optimum coarse-grained potentials corresponding to the behaviour of fully-atomistic simulations or experimental data (Figure 2.13c)[53]. It is extensively used to study protein aggregation and could be achieved by different computational methods. The relative entropy method uses information theory to find the potentials with minimal information loss in the configurational ensembles. This method was applied to study the self-assembly of polyalanine, where it was observed that the fibrillar structure could be achieved following internal reorganization within the disordered aggregate. Multiscale coarse graining is a variational technique that best matches all-atom momenta in the coarse-grained sites. For the polyglutamine peptides, increase in the aggregation propensity was observed with both concentration and chain length. The iterative Boltzmann inversion method can reproduce all-atom Boltzmann statistics. When applied to the aggregation process of oligoalanine peptides, it was found that certain details lost during the coarse graining process could be recovered by back-mapping to the atomistic coordinates.

Systematic coarse graining methods however require reliable all-atom statistics to parameterize the coarse-grained model[53]. Further, these potentials are highly system-specific, and would theoretically entail reparameterization of all the interactions; if the system changed in any manner. Therefore, these methods are ideal only for studying the self-assembly of numerous identical molecules.

2.5.3.2. Thermodynamic methods

To compute the free energy profiles more efficiently, increased sampling of Boltzmann-disfavoured morphologies is required[53]. These enhanced sampling methods can be used with both fully-atomistic and coarse-grained simulations. The replica exchange method enhances sampling by launching parallel simulations, such that each simulation explores a particular point in the parameter space. If this parameter is temperature, then the method is also referred to as parallel tempering. At regular intervals, each trajectory is given an opportunity to exchange trapped systems with trajectories at neighbouring temperatures to overcome free energy barriers (Figure 2.14b). Substantial energy overlap with neighbouring replicas is required to maintain a high temperature swap rate, and to ensure correct thermodynamics. However, because of this swapping procedure, each parameter has discontinuous trajectories and the correct kinetics are lost. This method can be used with both Monte Carlo and molecular dynamics simulations. The replica exchange statistical temperature molecular dynamics algorithm minimizes the number of replicas required for good sampling and provides a more even energy sampling. This method has been applied to study protein folding.

In umbrella sampling method, a predefined collective variable is kept at a particular target value[53]. Multiple trajectories are launched at varying target values, such that the statistics of neighbouring umbrellas, or Gaussian hills overlap. Umbrella sampling forces the system into regions of state space that might have poor sampling. This method has been applied to deduce the energy of separation of monomers, as the separation can be conveniently defined as a collective variable. Umbrella sampling was also used to study the binding of Aβ to a lipid bilayer, and it was observed that such binding facilitated the conversion of Aβ to aggregate-prone conformations.

a Metadynamics

b Replica exchange

T_1

T_2

T_N
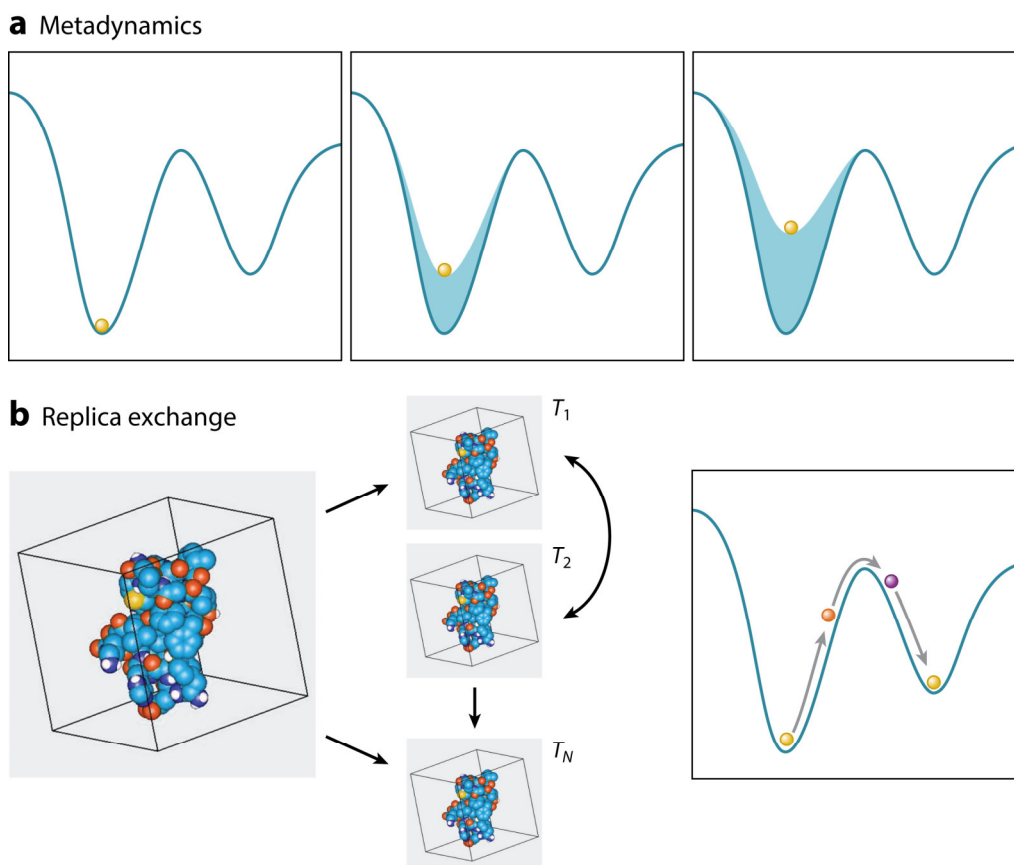
Figure 2.14. A schematic comparison of (a) metadynamics, and (b) replica exchange[53].

Metadynamics also involves enhanced sampling over collective variables, using a biased potential to force the system to sample low-probability states[53]. However, unlike umbrella sampling, metadynamics is an adaptive method. It automatically biases away configurations from the most visited regions of the state space by adding Gaussian hills (Figure 2.14a). Not only does this make the sampling more efficient, but also this provides access to rarer configurations. Like umbrella sampling, the bias is factored into while deducing the correct unbiased statistics of the system. It has been employed to study the diverse configurations of amyloidogenic proteins.

2.5.3.3. Kinetic methods

The previously described methods lose kinetic information by either biasing the sampling of collective variables (umbrella sampling and metadynamics), or giving discontinuous trajectories at constant temperature (replica exchange)[53]. However, certain sampling methods can preserve kinetics by launching multiple short, parallel trajectories. Markov state

model (MSM) increases sampling by deploying parallel trajectories. MSM bins sets of configurations in state space and models the system as a set of Markovian (history-independent) transitions between these configurations. Hence, it generates a kinetic map of transition probabilities between different states. This method adaptively selects starting configurations that require additional sampling, and therefore effectively samples the associated kinetic landscape. However, to ensure self-consistency, it needs to be verified that the transitions are indeed Markovian. It is also critical to bin the collective variables properly. This model has been applied to multiple biomolecular simulations.

Apart from MSM, other kinetic methods have also been used to study the biomolecular systems[53]. The free energy guided sampling method, applied to study protein folding, uses an approximate free energy surface instead of collective variables to bias the starting configurations. WExplore method biases the launching of trajectories towards poorly sampled regions of configuration space. However, unlike the MSM method, it accomplishes state transitions by using a weight for each trajectory with which it contributes to statistical averages. The WExplore method has been used to study the RNA conformational dynamics. Secondary nucleation data analysis methods represent the fibril formation kinetics as a function of an order parameter defining the degree of fibrillization. The data from experiments or simulations are fitted into a master equation that breaks down fibril growth into elongation, fragmentation, nucleation, and end-to-end association. The normal mode analysis method is applied to an existing (generally fully-atomistic) trajectory. It models the protein into coarse-grained sites, and then deduces collective motions of the biomolecule from the vibrational network of pairs of these sites. It has been employed to analyse the elastic properties of human IAPP fibrils and Aβ fibrils.

2.6. References

1.   Chiti, F.; Dobson, C. M., Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333-66.

2.   Lansbury, P. T.; Lashuel, H. A., A Century-Old Debate on Protein Aggregation and Neurodegeneration Enters the Clinic. *Nature* **2006**, *443*, 774-9.

3.   Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27-68.

4.   Jahn, T. R.; Radford, S. E., Folding Versus Aggregation: Polypeptide Conformations on Competing Pathways. *Arch Biochem Biophys* **2008**, *469*, 100-117.

5.   Eisenberg, D. S.; Sawaya, M. R., Structural Studies of Amyloid Proteins at the Molecular Level. *Annu. Rev. Biochem.* **2017**, *86*, 69-95.

6.   Zhang, M.; Zhao, J.; Zheng, J., Molecular Understanding of a Potential Functional Link between Antimicrobial and Amyloid Peptides. *Soft Matter* **2014**, *10*, 7425-7451.

7.   Pfeiffer, S.; Fushman, D.; Cowburn, D., Impact of Cl− and Na+ Ions on Simulated Structure and Dynamics of Bark1 Ph Domain. *Proteins: Structure, Function, and Bioinformatics* **1999**, *35*, 206-217.

8.   Yang, L.; Valdeavella, C. V.; Blatt, H. D.; Pettitt, B. M., Salt Effects on Peptide Conformers: A Dielectric Study of Tuftsin. *Biophysical Journal* **1996**, *71*, 3022-3029.

9.   Smith, M. D.; Cruz, L., Effect of Ionic Aqueous Environments on the Structure and Dynamics of the Aβ21–30 Fragment: A Molecular-Dynamics Study. *J. Phys. Chem. B* **2013**, *117*, 6614-6624.

10.   Dzubiella, J., Salt-Specific Stability and Denaturation of a Short Salt-Bridge-Forming A-Helix. *Journal of the American Chemical Society* **2008**, *130*, 14000-14007.

11.   Cruz, L.; Urbanc, B.; Borreguero, J. M.; Lazo, N. D.; Teplow, D. B.; Stanley, H. E., Solvent and Mutation Effects on the Nucleation of Amyloid Beta-Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18258-63.

12.   Bossis, F.; Palese, L. L., Amyloid Beta(1–42) in Aqueous Environments: Effects of Ionic Strength and E22q (Dutch) Mutation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, *1834*, 2486-2493.

13.   Aramvash, A.; Seyedkarimi, M. S., All-Atom Molecular Dynamics Study of Four Rada 16-I Peptides: The Effects of Salts on Cluster Formation. *J. Clust. Sci.* **2015**, *26*, 631-643.

14.   Yang, Y. I.; Gao, Y. Q., Computer Simulation Studies of Aβ37–42 Aggregation Thermodynamics and Kinetics in Water and Salt Solution. *J. Phys. Chem. B* **2014**, *119*, 662-670.

15. Kwon, J.; Lee, M.; Na, S., Sodium Chloride's Effect on Self-Assembly of Diphenylalanine Bilayer. *Journal of computational chemistry* **2016**, *37*, 1839-1846.

16. Kandasamy, S. K.; Larson, R. G., Effect of Salt on the Interactions of Antimicrobial Peptides with Zwitterionic Lipid Bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2006**, *1758*, 1274-1284.

17. Curtain, C. C.; Ali, F. E.; Smith, D. G.; Bush, A. I.; Masters, C. L.; Barnham, K. J., Metal Ions, Ph, and Cholesterol Regulate the Interactions of Alzheimer's Disease Amyloid-B Peptide with Membrane Lipid. *Journal of Biological Chemistry* **2003**, *278*, 2977-2982.

18. Lin, Y.-S.; Pande, V. S., Effects of Familial Mutations on the Monomer Structure of Aβ42. *Biophysical Journal* **2012**, *103*, L47-L49.

19. Tarus, B.; Straub, J. E.; Thirumalai, D., Structures and Free-Energy Landscapes of the Wild Type and Mutants of the Aβ21–30 Peptide Are Determined by an Interplay between Intrapeptide Electrostatic and Hydrophobic Interactions. *Journal of Molecular Biology* **2008**, *379*, 815-829.

20. Nguyen, P. H.; Tarus, B.; Derreumaux, P., Familial Alzheimer A2 V Mutation Reduces the Intrinsic Disorder and Completely Changes the Free Energy Landscape of the Aβ1–28 Monomer. *The Journal of Physical Chemistry B* **2014**, *118*, 501-510.

21. Han, W.; Wu, Y.-D., Molecular Dynamics Studies of Hexamers of Amyloid-B Peptide (16–35) and Its Mutants: Influence of Charge States on Amyloid Formation. *Proteins: Structure, Function, and Bioinformatics* **2007**, *66*, 575-587.

22. Brown, A. M.; Lemkul, J. A.; Schaum, N.; Bevan, D. R., Simulations of Monomeric Amyloid B-Peptide (1–40) with Varying Solution Conditions and Oxidation State of Met35: Implications for Aggregation. *Arch. Biochem. Biophys.* **2014**, *545*, 44-52.

23. Xu, L.; Ma, B.; Nussinov, R.; Thompson, D., Familial Mutations May Switch Conformational Preferences in A-Synuclein Fibrils. *ACS Chemical Neuroscience* **2017**, *8*, 837-849.

24.   Ning, L.; Wang, Q.; Zheng, Y.; Liu, H.; Yao, X., Effects of the A117v Mutation on the Folding and Aggregation of Palindromic Sequences (Prp113–120) in Prion: Insights from Replica Exchange Molecular Dynamics Simulations. *Molecular BioSystems* **2015**, *11*, 647-655.

25.   Lei, J.; Qi, R.; Wei, G.; Nussinov, R.; Ma, B., Self-Aggregation and Coaggregation of the P53 Core Fragment with Its Aggregation Gatekeeper Variant. *Physical Chemistry Chemical Physics* **2016**, *18*, 8098-8107.

26.   Das, A.; Makarov, D. E., Effect of Mutation on an Aggregation-Prone Segment of P53: From Monomer to Dimer to Multimer. *The Journal of Physical Chemistry B* **2016**, *120*, 11665-11673.

27.   Koldsø, H.; Andersen, O. J.; Nikolajsen, C. L.; Scavenius, C.; Sørensen, C. S.; Underhaug, J.; Runager, K.; Nielsen, N. C.; Enghild, J. J.; Schiøtt, B., Early Events in the Amyloid Formation of the A546t Mutant of Transforming Growth Factor B-Induced Protein in Corneal Dystrophies Compared to the Nonfibrillating R555w and R555q Mutants. *Biochemistry* **2015**, *54*, 5546-5556.

28.   Bera, S.; Jana, P.; Maity, S. K.; Haldar, D., Inhibition of Fibril Formation by Tyrosine Modification of Diphenylalanine: Crystallographic Insights. *Crystal Growth & Design* **2014**, *14*, 1032-1038.

29.   Green, J.; Goldsbury, C.; Mini, T.; Sunderji, S.; Frey, P.; Kistler, J.; Cooper, G.; Aebi, U., Full-Length Rat Amylin Forms Fibrils Following Substitution of Single Residues from Human Amylin. *Journal of Molecular Biology* **2003**, *326*, 1147-1156.

30.   Bouzakraoui, S.; Mousseau, N., Structural and Thermodynamical Properties of Early Human Amylin Oligomers Using Replica Exchange Molecular Dynamics: Mutation Effect of Three Key Residues F15, H18 and F23. *Phys. Chem. Chem. Phys.* **2017**, *19*, 31290-31299.

31.   Rodziewicz-Motowidło, S.; Juszczyk, P.; Kołodziejczyk, A. S.; Sikorska, E.; Skwierawska, A.; Oleszczuk, M.; Grzonka, Z., Conformational Solution Studies of the Sds Micelle-Bound 11-28 Fragment of Two Alzheimer's B-Amyloid Variants (E22k and A21g) Using Cd, Nmr, and Md Techniques. *Biopolymers* **2007**, *87*, 23-39.

32.   Fuzo, C.; Castro, J.; Degreve, L., Study of the Antimicrobial Peptide Indolicidin and a Mutant in Micelle Medium by Molecular Dynamics Simulation. *Genetics and Molecular Research* **2008**, *7*, 986-999.

33.   Langham, A. A.; Kaznessis, Y. N., Effects of Mutations on the C-Terminus of Protegrin-1: A Molecular Dynamics Simulation Study. *Molecular Simulation* **2006**, *32*, 193-201.

34.   Khandelia, H.; Kaznessis, Y. N., Molecular Dynamics Simulations of the Helical Antimicrobial Peptide Ovispirin-1 in a Zwitterionic Dodecylphosphocholine Micelle: Insights into Host-Cell Toxicity. *The Journal of Physical Chemistry B* **2005**, *109*, 12990-12996.

35.   Yuan, T.; Zhang, X.; Hu, Z.; Wang, F.; Lei, M., Molecular Dynamics Studies of the Antimicrobial Peptides Piscidin 1 and Its Mutants with a Dopc Lipid Bilayer. *Biopolymers* **2012**, *97*, 998-1009.

36.   Mihajlovic, M.; Lazaridis, T., Charge Distribution and Imperfect Amphipathicity Affect Pore Formation by Antimicrobial Peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818*, 1274-1283.

37.   Calabrese, A. N.; Markulic, K.; Musgrave, I. F.; Guo, H.; Zhang, L.; Bowie, J. H., Structural and Activity Changes in Three Bioactive Anuran Peptides When Asp Is Replaced by Isoasp. *Peptides* **2012**, *38*, 427-436.

38.   Rice, A.; Wereszczynski, J., Probing the Disparate Effects of Arginine and Lysine Residues on Antimicrobial Peptide/Bilayer Association. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2017**, *1859*, 1941-1950.

39.   Frenkel, D.; Smit, B., Chapter 4 - Molecular Dynamics Simulations. In *Understanding Molecular Simulation (Second Edition)*, Frenkel, D.; Smit, B., Eds. Academic Press: San Diego, 2002; pp 63-107.

40.   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005**, *26*, 1781-802.

41.   MacKerell, A. D., et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586-616.

42.   Becker, O. E., MacKerell Jr., A. (Ed.), Roux, B. (Ed.), Watanabe, M. (Ed.), *Computational Biochemistry and Biophysics*; CRC Press: Boca Raton, 2001, p 512.

43.   Frenkel, D.; Smit, B., Chapter 5 - Monte Carlo Simulations in Various Ensembles. In *Understanding Molecular Simulation (Second Edition)*, Frenkel, D.; Smit, B., Eds. Academic Press: San Diego, 2002; pp 111-137.

44.   Frenkel, D.; Smit, B., Chapter 6 - Molecular Dynamics in Various Ensembles. In *Understanding Molecular Simulation (Second Edition)*, Frenkel, D.; Smit, B., Eds. Academic Press: San Diego, 2002; pp 139-163.

45.   Alberty Robert, A., Use of Legendre Transforms in Chemical Thermodynamics (Iupac Technical Report). In *Pure and Applied Chemistry*, 2001; Vol. 73, p 1349. 10.1351/pac200173081349

46.   Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Physical Review A* **1985**, *31*, 1695-1697.

47.   Andersen, H. C., Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics* **1980**, *72*, 2384-2393.

48.   Parrinello, M.; Rahman, A., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52*, 7182-7190.

49.   Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

50.   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

51.   Frenkel, D.; Smit, B., Chapter 3 - Monte Carlo Simulations. In *Understanding Molecular Simulation (Second Edition)*, Frenkel, D.; Smit, B., Eds. Academic Press: San Diego, 2002; pp 23-61.

52. Frenkel, D.; Smit, B., Chapter 12 - Long-Range Interactions. In *Understanding Molecular Simulation (Second Edition)*, Frenkel, D.; Smit, B., Eds. Academic Press: San Diego, 2002; pp 291-320.

53. Morriss-Andrews, A.; Shea, J.-E., Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643-666.

54. Vanommeslaeghe, K.; MacKerell, A. D., Automation of the Charmm General Force Field (Cgenff) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144-3154.

Chapter 3

Mechanistic insight into the early stages of amyloid formation using an anuran peptide[†]

3.1. Introduction

Amyloidogenesis involves the conversion of specific peptides or proteins from soluble functional states into highly organized fibrillar aggregates with protofibrillar intermediates[1]. The 'amyloid hypothesis' implicates peptide or protein aggregation as the trigger for a cascade of events that result in neurodegeneration and other misfolding diseases[2-3]. Multiple pathways consisting of nucleation and growth phases with sigmoidal kinetics are accessible during the process of amyloidogenesis; depending on the physiological conditions, protein sequence and the surrounding molecular environment[4]. Protofibrils or diffuse amyloids are amyloid deposits that lack fibrillar substructure, but are precursors to mature amyloid fibrils and might be linked to cytotoxicity observed in amyloid diseases[2, 5-7]. Especially in neurological disorders involving the central nervous system, oligomeric forms cause the pathogenicity, primarily due to their small size allowing faster diffusion and greater exposure of the hydrophobic groups on their surface[4].

An increasing number of anti-microbial peptides (AMPs) have been identified with an inherent ability to form amyloid structures[8]. Uperin 3.5 (GVGDL^{5}IRKAV^{10}SVIKN^{15}IV-NH$_2$) is one such broad-spectrum AMP, obtained from the skin secretion of the toadlet *Uperoleia*

mjobergii[9]. Uperin 3.5 is unusual, as it does not aggregate in pure water, but self-aggregates in saline buffer at neutral pH to form amyloid fibrils[8]. Fibril formation occurs with the growth of oligomeric intermediates in a linear direction. These fibrils consist of a coiled-coil ultrastructure, rich in β-sheet content. Addition of co-solvent 2,2,2-trifluoroethanol (TFE) in small quantities is found to stabilize α-helical secondary structure intermediates and enhance fibril formation. However, fibril formation is not observed at higher concentrations of TFE due to an increased stability of the helical state.

The ability to form amyloid structures is an inherent property of certain polypeptide chains, and depends on overall charge, secondary-structure biases, and hydrophobicity of the peptide[4, 10]. However, there is no apparent sequence identity, structural homology or critical polypeptide chain length observed among these peptides[1, 11]. Still, most of these peptides are small in size (usually less than 100 residues). Most pathogenic mutations, common in hereditary amyloid diseases, disrupt the native structure of the protein and make it prone to aggregation. For example, the removal of a charged amino acid residue from a critical position in amyloid-beta (Aβ) protein (found in Alzheimer's disease) increases fibril formation in molecular dynamics (MD) simulations[12]. On the other hand, a synthetic mutant of amyloid-prone human islet amyloid polypeptide (hIAPP) showed significantly lesser membrane disruption activity compared to the wild-type polypeptide[13]. Peptide secondary structure is influenced by the surrounding environment. Simulations with Aβ$_{25-35}$, a portion of Aβ peptide, indicate that the solvent contributes to structural transitions that occur in amyloid structures. Notably, the random coil structure that dominates in aqueous environments is replaced by ordered secondary structures in membrane-mimicking environments[14]. Similar behaviour is observed for a host of amyloid-prone peptides *in vitro*, including uperin 3.5, which adopts an α-helical structure in membrane-like environments[8, 15]. Although, as observed for hIAPP, the surrounding lipid environment might also disrupt peptide aggregation[16]. Electrolytes can play a crucial role in the stabilization of peptide secondary structure and fibril formation by screening the charges on the residues and thus reducing the surface tension[17-18]. In addition, the valency of ions is also a critical factor in fibril formation. For instance, divalent cations (Ca^{2+}, Mg^{2+}) destabilize hydrogen bonds and turn structures in Aβ peptide and fragments through peptide-ion interactions and alterations to hydration shell, leading to the formation of unstructured aggregates[18-19]. But, monovalent cations (Na^+, K^+) facilitate intra-peptide hydrogen bonds and turn structures and hence promote peptide self-assembly.

There is increasing evidence that the mechanism of action of AMPs and amyloidogenic peptides and proteins have considerable similarity[8]. Thus, relatively small AMPs that form amyloid structures, such as uperin 3.5, can serve as model peptides for understanding the amyloid formation process. Further investigation of uperin 3.5 aggregation with MD simulations can provide significant insight into the structure of the transient and unstable oligomers observed in early stages of aggregation, at a resolution not accessible by current experimental facilities[20]. Examples of such phenomena and effects would include, the evolution of various oligomeric states adopted by uperin 3.5 peptides during aggregation, the influence of electrolyte concentration and solvent composition on uperin 3.5 aggregation, and the role played by charged residues in the process of amyloidogenesis. Finally, MD simulations can aid in the identification of regions crucial for inter-peptide interaction, and possibly the aggregation-prone structures[20].

In the current study, we utilize both fully-atomistic (FA) and coarse-grained (CG) MD simulations to understand the effect of salt concentration on the unusual aggregation behaviour of uperin 3.5 peptides in an aqueous medium. We use CG and FA simulations in a complementary way to gain insight into the peptide aggregation process in terms of both physical interactions as well as conformational transitions that take place in the early stages of aggregation. CG simulations provide large scale clustering and aggregation information over significantly longer time scales at relatively less computational period. In the absence of secondary structure details, CG simulations also allow us to understand the peptide aggregation process in simpler physical terms, analogous to aggregation in colloidal systems. The FA simulations, which are computationally more expensive, are intended to provide details of inter-atomic interactions and conformational changes that take place in the peptides.

3.2. Methods

Two interacting uperin 3.5 peptides with different inter-peptide orientations and aggregated in a dimeric state are shown in Figure 3.1, corresponding to both FA (Figure 3.1A) and CG (Figure 3.1B) representations. Whereas the FA representation is explicit and shows all the atomistic details, the CG representation corresponds to the MARTINI description. We note that the CG representation in Figure 3.1B is not the coarse-grained version of the FA representation in Figure 3.1A. The snapshots are purely representational and sourced from different time frames of completely independent simulations. In the MARTINI coarse graining scheme, the FA representation of every amino acid residue is substituted by one bead for the backbone and

a set of zero to four beads for the sidechain[21]. For bonded interactions, the MARTINI description correlates to an all-atom reference, but non-bonded interactions are determined by the chemistry of the united atoms (i.e., their charge, polarity, and hydrogen bonding capabilities)[20].

$$\text{GVGDL}^5\text{IRKAV}^{10}\text{SVIKN}^{15}\text{IV-NH}_2$$
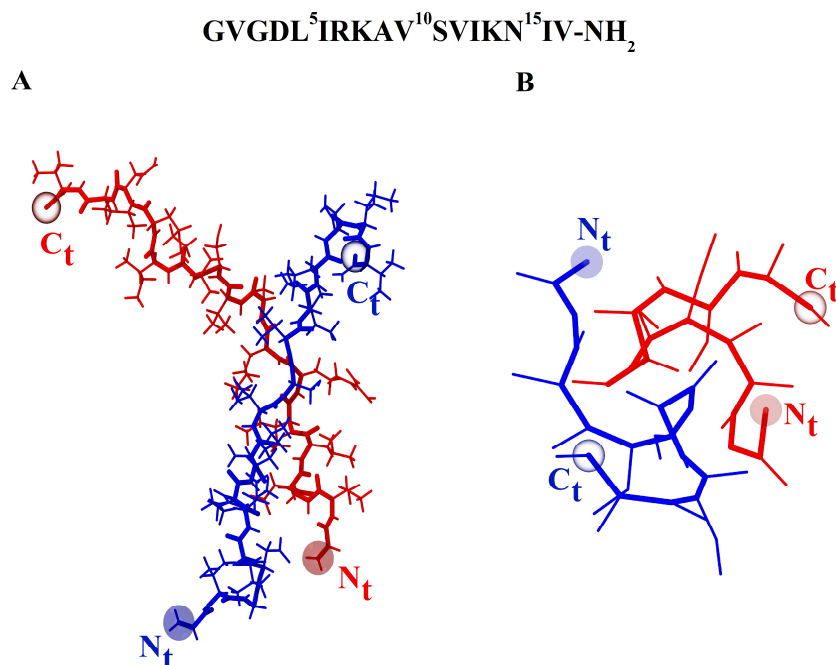
Figure 3.1. Representative snapshots of uperin 3.5 dimer in: (A) FA representation at 40 ns and (B) CG representation at 500 ns, obtained from corresponding uperin 3.5 dimer simulations in aqueous environment at 0.50 M salt concentration. Glycine (G) and valine (V) are the N-terminus (N_t) and C-terminus (C_t) amino acid residues respectively, in uperin 3.5 peptide.

In order to mimic the zwitterionic nature of peptides at physiological pH (7.4), a charge of +1 and -1 is usually assigned to the first and last backbone atoms of the peptide, i.e. the N- and C-terminus, respectively. However, in this case, the uperin 3.5 wild-type is amidated at the C-terminus[9], which means that no charge was assigned to the backbone atom at the C-terminus. Both FA and CG MD were used for simulating the following systems: (a) two uperin 3.5 peptides in a periodic box, and (b) twenty uperin 3.5 peptides in a periodic box, details of which are provided in Table 3.1. The peptide concentrations for two and twenty peptide systems (simulation numbers 1–9 in Table 3.1) are maintained at 3.32 mM and 20 mM, respectively.

Table 3.1. Setup details of CG and FA simulations performed.

Simulation number	Number of peptides	Simulation type	Box size ($Å^3$)	Number of water molecules	Salt concentration (NaCl, M)	Simulation time (ns)
1	2	CG	100 × 100 × 100	6518	0.00	1000
2	2	CG	100 × 100 × 100	6478	0.15	1000
3	2	CG	100 × 100 × 100	6382	0.50	1000
4	20	CG	118 × 118 × 118	10168	0.00	500
5	20	CG	118 × 118 × 118	10104	0.15	500
6	20	CG	118 × 118 × 118	9956	0.50	500
7	20	FA	118 × 118 × 118	50483	0.00	350
8	20	FA	118 × 118 × 118	50199	0.15	350
9	20	FA	118 × 118 × 118	49537	0.50	350

The CG MD simulations were performed using the MARTINI force-field for proteins, ions and water[21-24] in NAMD[25]. It is important to note that the MARTINI force-field description used here does not use polarizable models to describe either water or ions in the system. However, the MARTINI description appears to be adequate in correctly describing the qualitative changes in electrostatic effects at different salt concentrations. Periodic boundary conditions were applied along all three orthonormal directions. A switching function was implemented for the Lennard-Jones (LJ) interactions, with a smooth cut-off from 9 to 12 Å. Coulombic interactions were calculated directly with a shifting function throughout the interaction range, with a cut-off at 12 Å. Simulations were performed using a 10 fs timestep and a Langevin thermostat with a damping coefficient of 5 ps^{-1} [26-27]. A constant pressure of 1 atmosphere (1 atm = 1.01325 bar) was maintained using a Nosé-Hoover[28] Langevin[29] piston, with a period of 2 ps and a decay time of 1 ps. The systems were solvated with CG water molecules and ionized with hydrated sodium and chloride ions to get the desired salt concentrations[30]. All systems were equilibrated for 425 ps in a canonical (NVT) ensemble ($T = 310$ K), followed by 1 ns in an isothermal-isobaric (NPT) ensemble ($P = 1$ atm and $T = 310$ K). Then production runs were performed in NVT ensemble ($T = 310$ K).

The FA MD simulations were performed using the CHARMM36 force-field for proteins[31-33] and TIP3P water[34] in NAMD. Periodic boundary conditions were applied along all three orthonormal directions. A switching function was implemented for the LJ potential, with a smooth cut-off from 10 to 12 Å. Particle-mesh Ewald summation[35], with a grid spacing of 1 Å, was used for calculating the electrostatic interactions. Simulations were performed using a 2 fs timestep and a Langevin thermostat with a damping coefficient of 1 ps^{-1}. A constant pressure of 1 atm was maintained using a Nosé-Hoover[28] Langevin[29] piston, with a period of 100 fs and a decay time of 50 fs. The peptides were solvated with TIP3P water molecules. Then, Na$^+$ and Cl$^-$ ions were added as required. With only α-carbon of amino acid residues in the peptides restrained, equilibration was performed for 300 ps in an NVT ensemble ($T = 310$ K), followed by further equilibration for 500 ps in an NPT ensemble ($P = 1$ atm and $T = 310$ K). Finally, the systems were equilibrated for 1 ns in an NPT ensemble ($P = 1$ atm and $T = 310$ K) without any restraints. The resulting systems were then simulated in NVT ensemble ($T = 310$ K).

The effect of salt concentration on the free-energy of interaction between two CG uperin 3.5 peptides[36] is evaluated by calculating the potential of mean force (PMF) using the metadynamics method[37-41], as implemented by the Colvars module[42] in NAMD. The PMF is evaluated as a function of two variables, namely, the backbone root-mean-square deviation ($RMSD$) of the two peptides, with reference to an aggregated dimer state, and the radius of gyration (R_g) of the two peptides[30]. The PMF sampling was done in the range of 0–50 Å for both $RMSD$ and R_g reaction coordinates and the corresponding CG simulations were run for 5 μs.

Uperin 3.5 appears to have a random coil conformation in aqueous environment and a predominantly α-helical structure in membrane-mimicking environments[8]. However, no definitive structural information is available regarding the individual peptide conformations that progress to amyloid fibril formation. In a related study, theoretical models have shown that Aβ peptide conformations with predominantly random coil and turn secondary structures aggregate to form propagating oligomers that would lead to amyloid fibrils[43]. Hence, we start our simulations with random coil uperin 3.5 conformation[43]. The molecular editors PACKMOL[44] and VMD[45] were used to create the simulation systems. Tachyon ray-tracer[46], as incorporated in VMD, was used for visualization. Backbone RMSD clustering[47-48] was performed with the Stillinger algorithm[49]. Order parameters were derived using Wordom

program package[50-51]. Secondary structure in FA simulations was determined by STRIDE program[52] using DSSP definitions[53], as implemented by the Timeline plugin[54] in VMD.

3.3. Results

3.3.1. Aggregation of two peptides

3.3.1.1. Inter-peptide interaction

In order to examine the inter-peptide interaction in detail, we simulated the aggregation of two CG uperin 3.5 peptides under varying salt concentrations and analysed the peptide trajectories over the simulation period. We tested several different random coil conformations of uperin 3.5 monomers as different initial conditions and observed similar aggregation behaviours with respect to changing salt concentration. Thus, different starting configurations did not seem to have a statistically significant impact on the aggregation process. The centre of mass distance provides a measure of the average distance between the centre of mass of two peptides within the simulation system. The variation of centre of mass distance for two interacting uperin 3.5 peptides (denoted by P1 and P2) is plotted with respect to time for different salt concentrations in Figure 3.2, corresponding to a total simulation duration of 1000 ns. The N-termini (G1) and C-termini (V17) backbone atoms in uperin 3.5 have been represented as brown and violet coloured spheres, respectively in Figure 3.2 and elsewhere (as applicable). In the absence of salt (c_{NaCl} = 0.00 M), the peptide-peptide separation was found to be always more than 20 Å throughout the simulation duration, indicating negligible inter-peptide interaction (Figure 3.2A). With an increase in salt concentration, both the centre of mass distance values and fluctuations in the plots were found to decrease (Figures 3.2B and 3.2C); and were significantly lower at c_{NaCl} = 0.50 M for nearly 700 ns, clearly showing an increased attraction between the two peptides at higher salt concentrations. Increased hydrophobic interaction with an increase in salt concentration points to the screening of charged residues on the peptides by counter-ions from the salt, which results in the weakening of electrostatic repulsion between charged residues on the two peptides. Therefore, the plot for 0.50 M salt concentration has lesser fluctuations when compared to 0.15 M salt concentration. Hence, at c_{NaCl} = 0.15 M, weak attractions between the peptides result in a loosely associated dimer which is prone to disruption, as illustrated by the large fluctuations in the centre of mass distance plot (Figure 3.2B). At c_{NaCl} = 0.50 M, the dimeric state is significantly more stable for nearly 700 ns, and only after ~ 700 ns does the inter-peptide separation steadily increase. This suggests that

hydrophobic interactions between the two CG uperin 3.5 monomers at higher salt concentration remain susceptible to disruption.
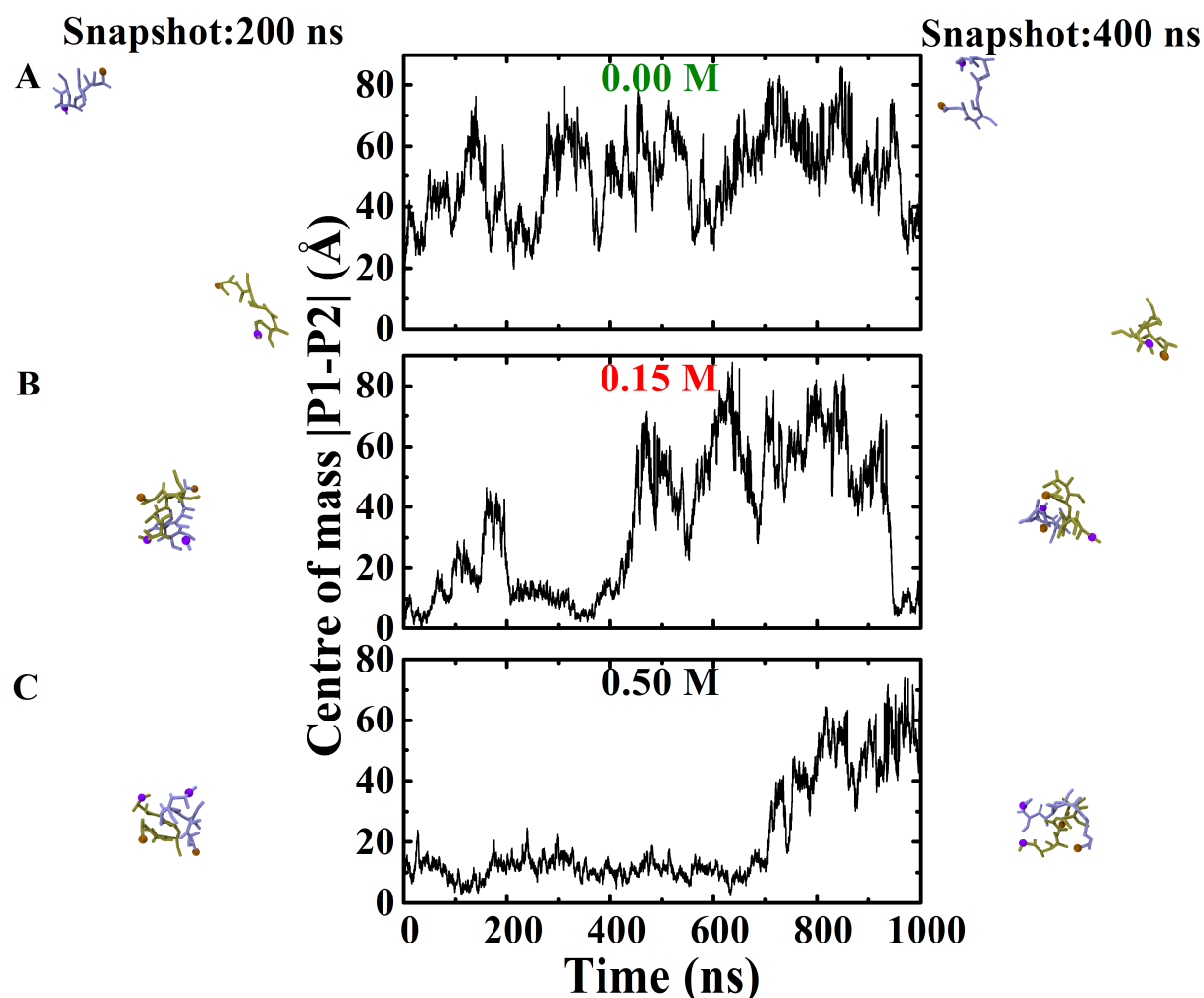
Figure 3.2. Centre of mass separation between the two peptides, P1 and P2, plotted as a function of time for different salt concentrations of (A) 0.00 M, (B) 0.15 M, and (C) 0.50 M NaCl. Peptide trajectory snapshots from the CG simulations, showing dimer interaction at different salt concentrations, and at two representative time instances of 200 ns (left-side panels) and 400 ns (right-side panels).

The CG uperin 3.5 simulations are characterized by large fluctuations of the peptide backbones, which could lead to an increased disruption of the aggregated dimers. These fluctuations could be inherent to the uperin 3.5 structure in aqueous media, arising from a lack of proline or aromatic residues in uperin 3.5[9]. Larger backbone fluctuations could also be an artefact of the reduced number of overall inter-atomic interactions in a CG model[21], which would otherwise be responsible for stabilizing conformations in a FA simulation. As a result, hydrophobic interactions between the two peptides might not be strong enough to keep the two peptides

associated for extended durations, even at high salt concentrations (Figure 3.2C). Trajectory snapshots of the two uperin 3.5 monomers (Figure 3.2) that were taken from the uperin 3.5 dimer simulations also show an increasing tendency of the monomers to aggregate in the defined salt concentrations. At 200 ns, the peptides are in close proximity in 0.15 M (Figure 3.2B) and 0.50 M (Figure 3.2C) systems, but further apart in 0.00 M system (Figure 3.2A). Similarly, at 400 ns, the uperin 3.5 monomers in 0.15 M (Figure 3.2B) and 0.50 M (Figure 3.2C) systems are closely associated, but well-separated from each other in 0.00 M system (Figure 3.2A).

At this stage, it is important to reiterate that several trajectories (corresponding to different initial peptide conformations and initial separations at a given NaCl concentration) were simulated, but the differences between the simulations were statistically insignificant. This means that the trajectories shown in Figure 3.2 can be considered as representative of dimer aggregation for a particular NaCl concentration. Whereas dimerization is never observed in the absence of NaCl, most trajectories lead to dimerization at 0.50 M NaCl. Large backbone fluctuations, inherent to uperin 3.5 CG simulations, can lead to disruption of the dimer (as seen in Figures 3.2B and 3.2C), and the lifetime of a dimer at a particular NaCl concentration depends on the initial peptide configuration and inter-peptide separation. The re-aggregation of the two peptides into a dimer after a disruption event is dependent on a collision probability, which is expected to increase with increasing peptide concentration. Since the peptide concentration is very small (3.32 mM) in the dimer simulations, a re-appearance of the dimeric state is a rare event and was not observed in the simulations. However, re-aggregation events following disruption of clusters are observed quite frequently for the CG:20 and FA:20 simulations, where the peptide concentration is much higher (20 mM) (see Figures 3.11 and 3.12).

For a more detailed analysis of the inter-residue contacts between peptide 1 (P1) and peptide 2 (P2) in the CG simulations of the uperin 3.5 dimers, contact maps[55-56] were generated as shown in Figure 3.3 for c_{NaCl} = 0.00 M, 0.15 M and 0.50 M. A contact map is generated by counting the total number of contacts between all possible pairs of residues along the two uperin 3.5 peptides[55]. A pair of residues, each located on a different peptide, is assumed to be in contact if the backbone residue atoms are separated by less than 12 Å. Using this distance criterion (based on the cut-off employed for non-bonded interactions in the CG simulations), the total

number of contacts are counted for all trajectory outputs (sampled every 100 ps) over the entire duration of the simulation.
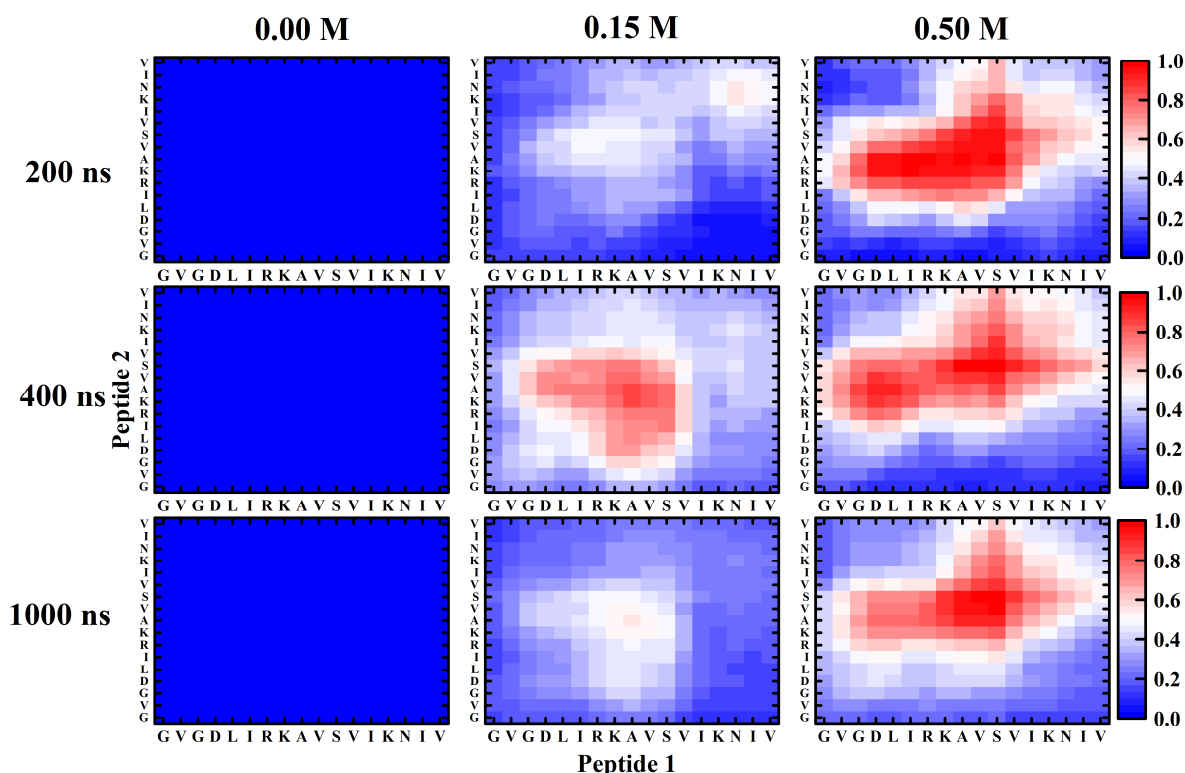
Figure 3.3. Contact maps showing inter-peptide interactions, calculated from CG simulations of two uperin 3.5 monomers as functions of NaCl concentration (increasing from left to right) and simulation time (increasing from top to bottom).

A visual inspection of the contact maps in Figure 3.3 clearly shows that most of the inter-peptide interactions occur along the stretches of predominantly hydrophobic residues 9–13 (AVSVI segment) on the two peptides. The highest number of interactions (or contacts) in this region were observed in 0.50 M system throughout the simulation period of 1000 ns. Based on this observation, we use the maximum value obtained for a contact pair in the AVSVI segment at 0.50 M NaCl as a reference value to normalize the number of contacts at all NaCl concentrations. Following this procedure, the number of contacts can be represented on a scale of 0 to 1 (as shown in Figure 3.3), where a value of 1 represents the highest value of contacts observed for a pair of residues in the AVSVI segment at 0.50 M NaCl. After 200 ns, the contact map shows negligible interaction (0.5%) at c_{NaCl} = 0.00 M, and interactions at only 47.2% of the maximum value were noted at c_{NaCl}= 0.15 M (as could be deduced by the increase in the white region of the contact map with respect to the colour scale, compared to the 0.00 M system). However, at c_{NaCl} = 0.50 M, significant interactions were observed along the AVSVI

stretch; as defined by the intense dark red regions. The contact maps in Figure 3.3 represent long time-scale peptide-peptide interactions by counting the cumulative number of residue-residue contacts up to a given time. The evolution of contact maps has been explained in detail in Figure 3.4, where contact maps for c_{NaCl}= 0.50 M are generated by counting residue-residue contacts over shorter 100 ns intervals at various stages of the CG uperin 3.5 dimer trajectory. Significantly, the shorter time-scale contact maps in Figure 3.4 clearly show that the peptide-peptide interactions are primarily confined to the AVSVI hydrophobic stretches of the two peptides.

At longer timescales (> 400 ns), the situation remains invariant at c_{NaCl} = 0.00 M, i.e., the interactions remain negligible (0.5%). In contrast, at c_{NaCl} = 0.15 M, there is an appreciable increase in the number of interactions (68.6% of the highest value) at 400 ns compared to 0.00 M NaCl. Once again, at the end of the simulations (1000 ns), fewer interactions (45.5 %) were observed at c_{NaCl} = 0.15 M in comparison to those observed at c_{NaCl} = 0.50 M. Fluctuations observed in contact maps at c_{NaCl} = 0.15 M and 0.50 M over the simulation period can be attributed to fluctuations in inter-peptide interaction over the duration of the simulation. These data are consistent with the role of salt as a facilitator of hydrophobic interactions between two uperin 3.5 peptides in the aqueous environment. Here, it is important to note that our simulation results are in agreement with the experimental observations of Calabrese et al.[8], where it was shown that introduction of salts facilitates aggregation of the uperin 3.5 peptides dissolved in water.

In order to validate the hypothesis that peptide-peptide interactions occur along the AVSVI hydrophobic stretches, we generated contact maps for the CG uperin 3.5 dimer trajectory at c_{NaCl} = 0.50 M, but with residue-residue contacts counted only over 100 ns intervals. Four such contact maps are shown in Figure 3.4 that are evaluated at different time intervals of the trajectory in Figure 3.2C corresponding to different stages of dimer aggregation, namely, tightly bound states (400–500 ns and 550–650 ns), during dimer dissociation (650–750 ns), and after dimer disruption (850–950 ns). It is very clear from the contact maps that the peptide-peptide interactions are primarily confined to their hydrophobic regions when the two peptides are in a tightly bound dimeric state. As expected, as the two peptides drift apart during dissociation, fewer residue-residue contacts are observed and the contact maps appear smeared out. No contacts are observed after the disruption event. Further, computational analysis with

the TANGO algorithm[57-59] indicates towards a crucial role played by the predominantly hydrophobic A^9VSVI^{13} segment in peptide aggregation, with increasing NaCl concentration.
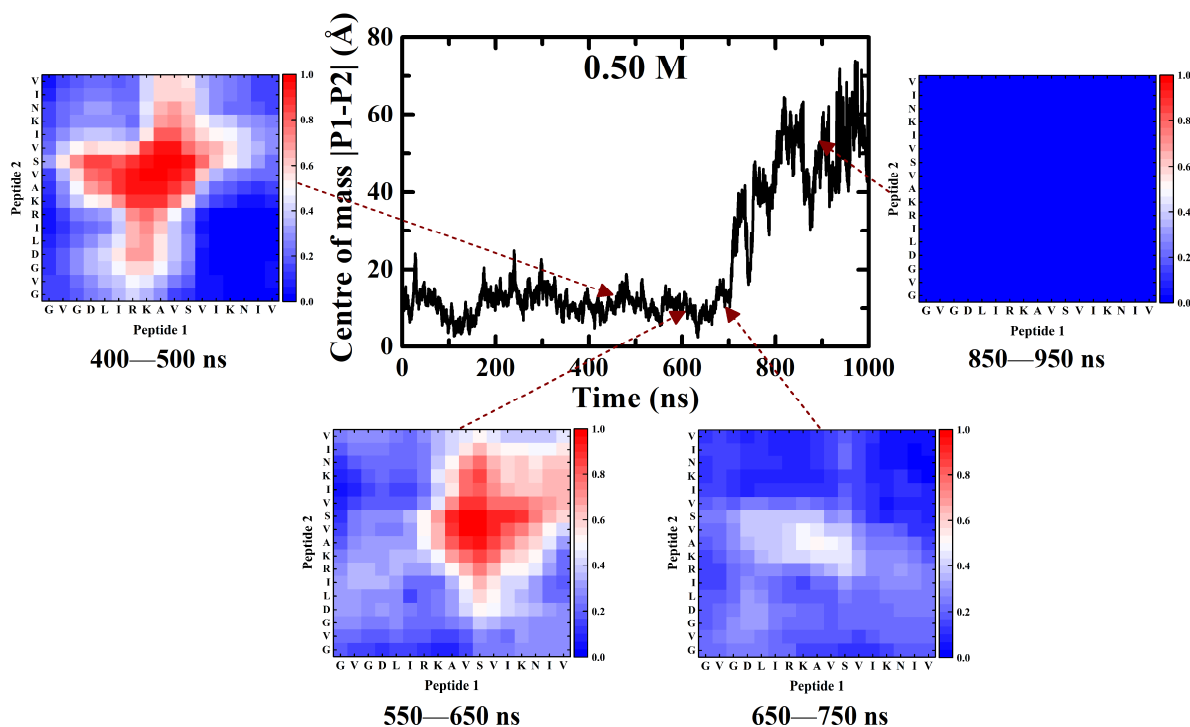
Figure 3.4. Centre of mass separation between the two uperin 3.5 peptides, P1 and P2, plotted as a function of time for salt concentration of 0.50 M NaCl, along with contact maps showing inter-peptide interactions calculated over different time intervals of the trajectory.

The TANGO algorithm[57-59] was employed in Figure 3.5 to analyse the effect of NaCl concentration on secondary structure and aggregation propensities of uperin 3.5. As with the multi-peptide simulations in this study, the temperature and the peptide concentration were kept at 310 K and 20 mM, respectively, for the analysis. It could be observed that the algorithm predicts similar β-sheet propensities of ≈ 11% in the $V^{10}SV^{12}$ region at all three salt concentrations in Figure 3.5A. However, in the A^9VSVI^{13} region of uperin 3.5 in Figure 3.5B, β-aggregation propensities of 0.00%, 0.67%, and 1.63% were predicted by the algorithm at c_{NaCl} = 0.00 M, 0.15 M, and 0.50 M, respectively. Therefore, this analysis indicates that the A^9VSVI^{13} region plays a crucial role in the self-aggregation phenomena of uperin 3.5, corroborated by the observations made in Figures 3.3 and 3.20 of the manuscript.
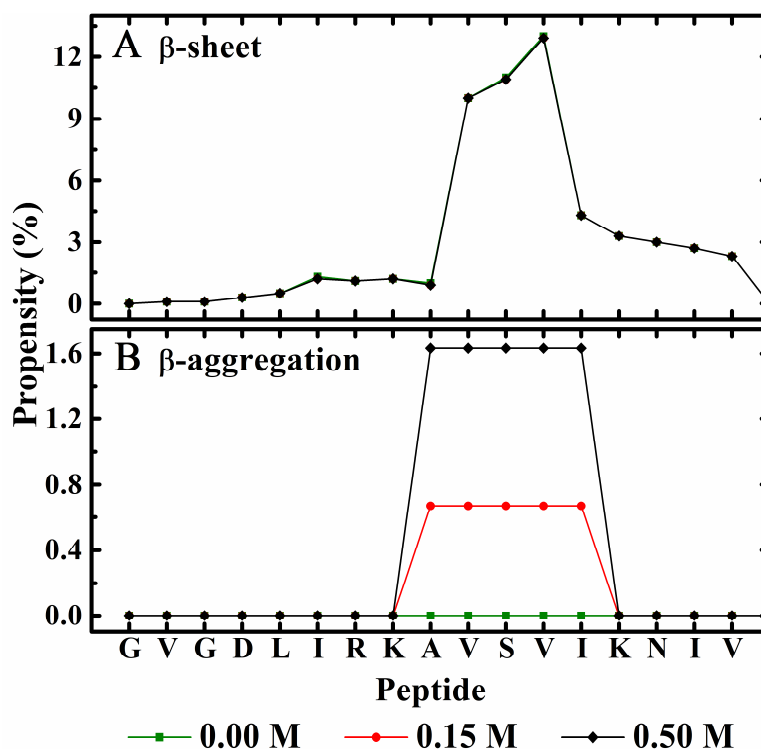
Figure 3.5. Plots showing the variation of (A) β-sheet secondary structure and (B) β-aggregation propensities along the peptide length at different salt concentrations, as predicted by the TANGO algorithm[57-59].

3.3.1.2. Energetics of interaction

Figure 3.6 shows the effect of increasing salt concentration on the interaction between two uperin 3.5 peptides, highlighting the positively-charged residues R7, K8 and K14 that are most influenced by charge screening. These charged residues are adjacent to the predominantly hydrophobic AVSVI segment (9–13). In addition, there is also a negatively-charged aspartic acid residue at position 4 (D4). The simulation snapshots (Figure 3.6A) clearly show an increase in the number of negatively-charged chloride counter-ions (Cl⁻) localized in the immediate vicinity of the positively-charged residues as the NaCl concentration increases. In contrast, almost no Na^+ counter-ions (in fact only one can be seen at both 0.15 M and 0.50 M NaCl) are observed next to the negatively-charged D4 residues.
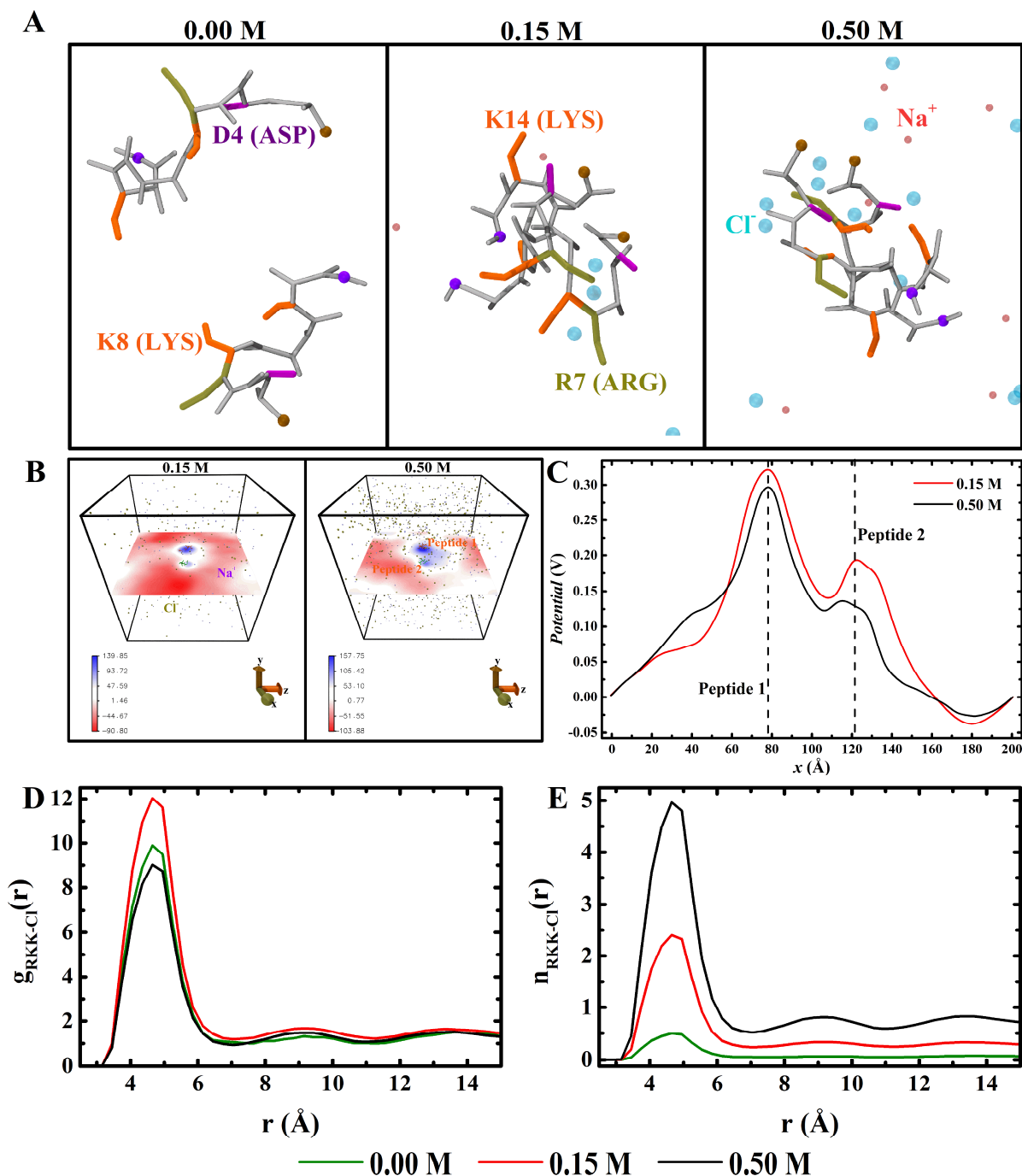
Figure 3.6. (A) Representative simulation snapshots taken at 400 ns from CG uperin 3.5 dimer simulations at different NaCl concentrations of 0.00 M, 0.15 M, and 0.50 M, with a focus on charged residues (D4, R7, K8 and K14) and the surrounding counter-ions. Negatively-charged chloride ions are represented as light blue spheres and positively-charged sodium ions are represented as light red spheres. (B) Electrostatic potential distribution in the *X-Z* plane for two uperin 3.5 peptides separated by 50 Å, and at 0.15 M and 0.50 M NaCl. (C) Variation of electrostatic potential along *X*-axis. (D) Radial distribution function and (E) number density of Cl⁻ counter-ions around positively-charged R7, K8 and K14 residues.

An increased electrostatic screening of the positively-charged residues with increasing salt concentration is also apparent from the variation of the electrostatic potential at $c_{NaCl} = 0.15$ M and 0.50 M in Figures 3.6B and 3.6C. The electrostatic potential variation was evaluated for two uperin 3.5 peptides held at a constant separation of 50 Å in a simulation box of volume $200 \times 200 \times 200$ Å^3, and at NaCl concentrations of 0.15 M and 0.50 M (Figure 3.6B). For these two simulations, the peptide centre-of-mass were constrained to remain at a constant separation of 50 Å over the 100 ns duration. The relevant planes and contour maps of the electrostatic potentials are presented in Figure 3.6B, clearly showing a positive electrostatic potential around the positively-charged residues which decays as one moves away from the peptides. Since the separation between the two peptides is along the X-axis, Figure 3.6C shows the variation of electrostatic potential along the X-axis, as averaged over the last 40 ns of the simulations. Higher electrostatic potentials are observed at the lower NaCl concentration of 0.15 M. For both salt concentrations, the electrostatic potentials decay as one moves away from the peptides. However, the electrostatic potential is lower at the higher salt concentration of 0.50 M NaCl, clearly pointing to the effect of greater electrostatic screening due to the presence of additional salt ions. Thus, a greater charge screening at higher salt concentration leads to reduced electrostatic repulsion between the two uperin 3.5 peptides.

Figures 3.6D and 3.6E discuss the counter-ion distribution (Cl$^-$ ion) in the vicinity of the positively-charged residues R7, K8, and K14, at different salt concentrations. The radial distribution function plot for the Cl$^-$ ion, $g_{RKK-Cl}(r)$ in Figure 3.6D, shows that the Cl$^-$ ion distribution peaks at 4.5 Å around the charged residues. The peak height increases from 0.00 M to 0.15 M NaCl, but decreases with further increase in the NaCl concentration to 0.50 M. However, if we calculate the number density of Cl$^-$ ions, $n_{RKK-Cl}(r)$ in Figure 3.6E, as a function of radial distance from the positively-charged residues), we find that the number density of Cl$^-$ ions in the immediate vicinity of the positively-charged residues increases with increasing NaCl concentration. This number density is significantly higher at the highest salt concentration of 0.50 M NaCl when compared to $c_{NaCl} = 0.00$ M (nearly 10 times) and $c_{NaCl} = 0.15$ M (almost double), and therefore conclusively shows that the electrostatic screening of positive charges on the peptides increases with increasing salt concentration. An increased electrostatic screening by Cl$^-$ counter-ions leads to a weakening of the electrostatic repulsion between the two peptides.
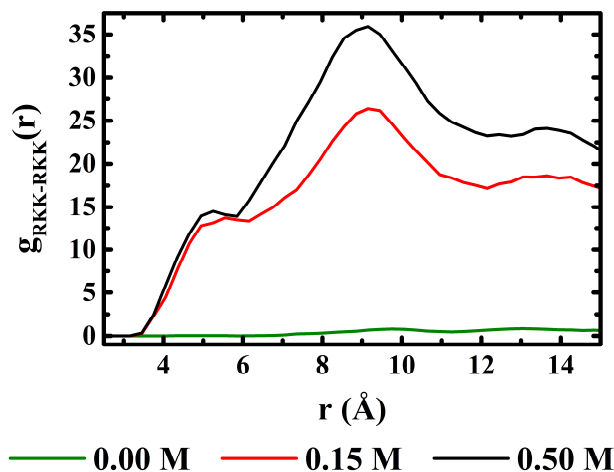
Figure 3.7. Radial distribution function between positively-charged R7, K8 and K14 residues on two different uperin 3.5 peptides.

Figure 3.7 discusses the radial distribution function $g_{RKK-RKK}(r)$, representing the inter-chain correlations between the positively-charged residues on the two uperin 3.5 peptides. The plot clearly shows that the inter-chain correlations between the positively-charged residues on the two peptides increase with increasing salt concentration, with the greatest correlation occurring at a separation of approximately 9 Å. This correlation peak appears at nearly double the distance of the correlation peak observed in $g_{RKK-Cl}(r)$ plot in Figure 3.6D, suggesting formation of ion bridges between positively-charged residues on either peptide chain, via the Cl⁻ counter-ions.
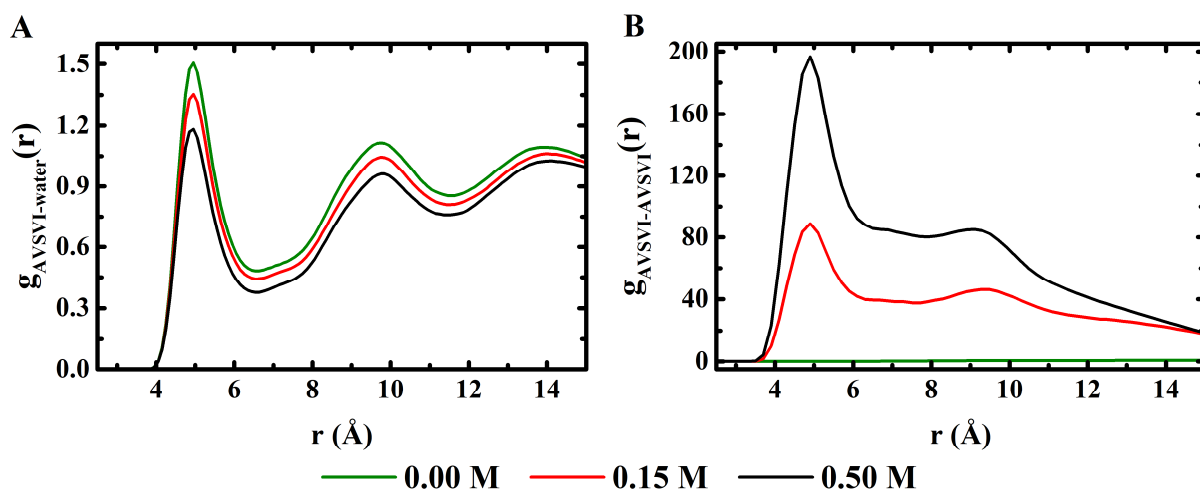
Figure 3.8. Radial distribution functions between (A) A^{9}VSVI^{13} segments and surrounding water molecules, and (B) A^{9}VSVI^{13} segments on two different uperin 3.5 peptides.

From the analysis of CG dimer simulations (specifically contact maps in Figure 3.3), we concluded that dimer formation occurs primarily along the A^{9}VSVI^{13} segments of the uperin

3.5 peptide. Hence, in Figures 3.8A and 3.8B, we calculated these radial distribution functions at different salt concentrations: $g_{AVSVI-water}(r)$ to quantify A^9VSVI^{13}-water correlations, and $g_{AVSVI-AVSVI}(r)$ to quantify inter-chain A^9VSVI^{13}-A^9VSVI^{13} correlations. A careful examination of the radial distribution function plots clearly shows that hydrophobic interactions between the two peptides increase with increasing salt concentration, which then leads to greater aggregation. Figure 3.8A shows that $g_{AVSVI-water}(r)$ values decrease for all distances with increasing salt concentration, demonstrating that the water is depleted from the vicinity of the A^9VSVI^{13} segments. In consonance with this, inter-chain A^9VSVI^{13}-A^9VSVI^{13} correlations increase significantly with increase in salt concentration, as evidenced from the $g_{AVSVI-AVSVI}(r)$ plots in Figure 3.8B. In fact, inter-chain A^9VSVI^{13}-A^9VSVI^{13} correlations are nearly non-existent at $c_{NaCl} = 0.00$ M, but show a significant increase at $c_{NaCl} = 0.15$ M, and more than double at $c_{NaCl} = 0.50$ M. The first correlation peak for the inter-chain A^9VSVI^{13}-A^9VSVI^{13} contacts occurs at nearly 5 Å separation for both 0.15 M and 0.50 M NaCl systems, suggesting dimerization of uperin 3.5 peptides by aggregation.

A build-up of counter-ions around each charged residue would screen that residue's charge, reducing the range of the repulsive electrostatic interactions between the charged residues on the two peptides. This would in turn reduce the electrostatic repulsion between the positively-charged residues and allow the neighbouring AVSVI stretches of predominantly hydrophobic residues on the peptides to interact with each other, thereby facilitating an attractive inter-peptide interaction. This picture of counter-ion condensation due to the addition of NaCl, leading to enhanced attraction between the two peptides, is also consistent with the observation made in Figure 2, where the centre of mass distance between the two peptides decreased with increasing salt concentration. With increasing NaCl concentration, radial distribution function plots show increased screening of charged residues (Figure 3.6E), increased inter-chain correlations between charged residues (Figure 3.7), and enhanced inter-chain hydrophobic interactions accompanied by depletion of water from the vicinity of hydrophobic A^9VSVI^{13} segments, leading to dimer formation (Figure 3.8).
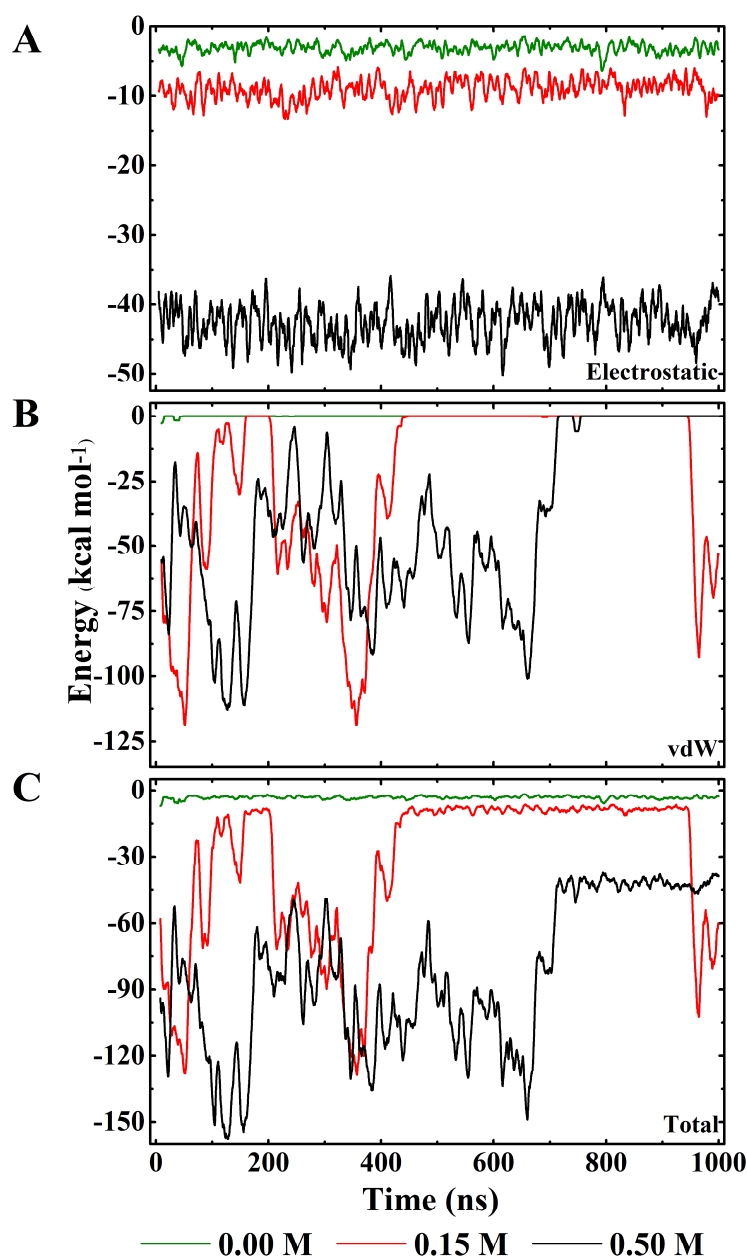
Figure 3.9. (A) Electrostatic, (B) van der Waals, and (C) total interaction energies between two CG uperin 3.5 peptides at 0.00 M, 0.15 M and 0.50 M NaCl.

Interaction energies between two CG uperin 3.5 peptides undergoing dimerization, at different salt concentrations, have been depicted in Figure 3.9. The electrostatic and van der Waals components of the interaction energy, along with the sum of these two contributions have been plotted in Figures 3.9A, 3.9B, and 3.9C, respectively, as functions of time. The interaction between the two uperin 3.5 monomers is modulated by the monovalent sodium and chloride ions present between these peptides and also the rest of the simulation system. As the salt concentration increases, the charges on the peptides get screened, allowing greater hydrophobic interaction and more stable association between the peptides. The electrostatic interaction

78

energy within the system increases with increasing NaCl concentration as observed in Figure 3.9A, resulting from reduced electrostatic repulsion between the peptides. This in turn allows closer association of the two peptides, resulting in more attractive van der Waals interaction between the two peptides, which can be observed from the highly negative van der Waals interaction energies in Figure 3.9B at $c_{NaCl} = 0.15$ M and 0.50 M. Upon taking both electrostatic and van der Waals effects into consideration, it can be clearly noted in Figure 3.9C that an attractive inter-peptide interaction is not favoured at no salt case. However, attraction between the two peptides is energetically favourable in 0.15 M and 0.50 M NaCl systems, with inter-peptide interaction at 0.50 M NaCl being comparatively more attractive than at 0.15 M NaCl.
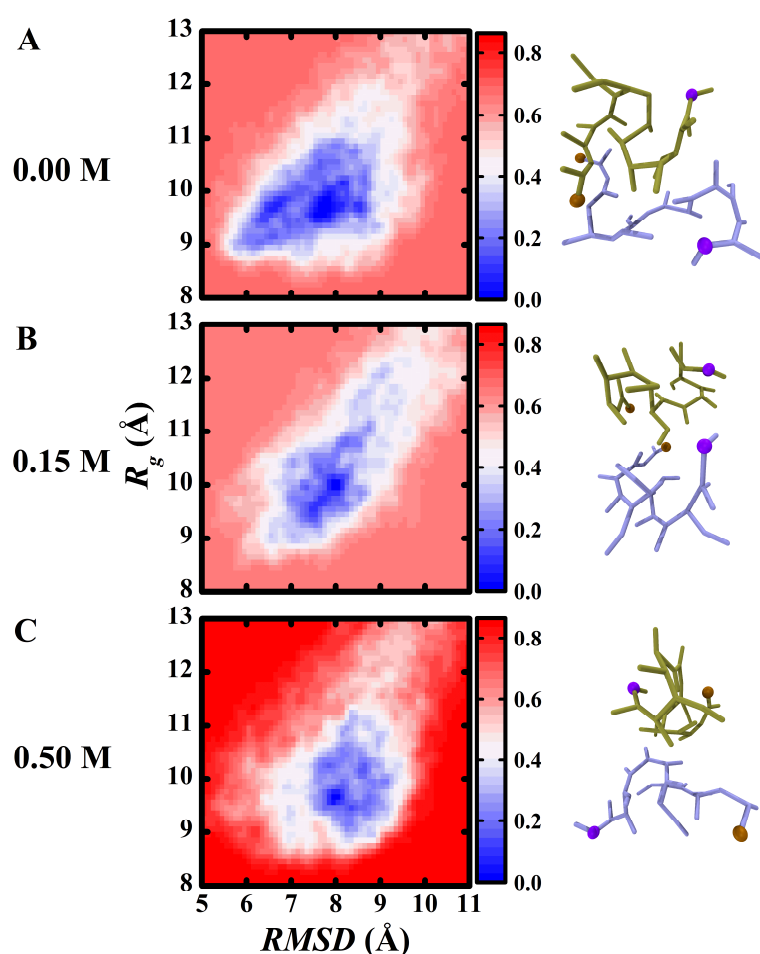
Figure 3.10. The PMF showing the interaction free energy between two uperin 3.5 peptides, plotted as a function of the backbone $RMSD$ and the R_g of the two peptides at NaCl concentrations of (A) 0.00 M, (B) 0.15 M, and (C) 0.50 M.

Figure 3.10 shows PMF plots for two CG uperin 3.5 peptides at different salt concentrations as a function of the backbone $RMSD$ of the two peptides with reference to an aggregated dimer, and the R_g of the two peptides. The PMF is an indicator of the interaction free energy between

the two peptides. Hence, PMF plots also indicate the propensity of two uperin 3.5 peptides to form a stable dimer at a particular salt concentration.

For the no salt case (0.00 M), the minimum PMF in Figure 3.10A was observed at *RMSD* and R_g values of 7.75 Å and 9.75 Å, respectively. At c_{NaCl} = 0.15 M in Figure 3.10B, the minimum PMF was at *RMSD* and R_g values of 7.95 Å and 9.95 Å, respectively. Similarly, at the highest salt concentration of 0.50 M in Figure 3.10C, the minimum PMF was noted at *RMSD* and R_g values of 7.95 Å and 9.65 Å, respectively. The well-depths observed at c_{NaCl} = 0.00 M, 0.15 M, and 0.50 M are 0.68, 0.66, and 0.86 kcal mol^{-1}, respectively. In fact, the PMF well depths at all three salt concentrations are very close to the thermal energy ($k_B T$) values at the simulation temperature of 310 K, where k_B is the Boltzmann's constant. Thus, the *RMSD* and R_g values at the minimum PMF, and the PMF well-depths are very similar at all three NaCl concentrations. This is most likely due to the fact that the peptides in these simulations are fluctuating random coils that sample a large number of conformations during the simulations. Using MD simulations, Tsigelny et al. have demonstrated that Aβ peptides have a diverse set of conformations, which result in dimers with large structural diversity and differing tendencies for amyloid formation[43]. However, the dimeric (or aggregated) state in our simulations most likely represents an average over all the possible uperin 3.5 conformations that may lead to aggregation. As a result, significant energy differences between dimerization processes at different salt concentrations are not visible. It is quite likely that specific uperin 3.5 conformations which form energetically stable dimers may well show significant differences in their dimerization PMFs at different salt concentrations. However, testing of such a hypothesis is beyond the scope of the present work. The overall PMF funnel is rugged and appears to shift towards higher values of *RMSD* as the salt concentration increases from 0.00 M to 0.50 M.

3.3.2. Aggregation of multiple peptides

3.3.2.1. Clustering analysis

Although uperin 3.5 dimer simulations provide significant insight into the nature of the interactions between two uperin 3.5 peptides, it remains difficult to predict how a fixed concentration (20 mM) of these peptides would associate during the initial stages of aggregation. In order to understand the initial aggregation process for a finite number of peptides, both CG (CG:20 for 1000 ns) and FA (FA:20 for 350 ns) simulations were carried

out at c_{NaCl} = 0.00 M, 0.15 M and 0.50 M, with 20 peptides in the simulation box corresponding to a peptide concentration of 20 mM (Table 3.1). In these simulations, the extent of peptide aggregation was quantified by introducing the concept of 'loosely associated clusters' (LAC), which defines a group of peptides that are in close proximity, and apparently in an aggregated state. For the purpose of a qualitative description of the extent of aggregation, simulation snapshots are shown in Figures 3.11A and 3.12A for CG:20 and FA:20, respectively, which convey an approximate picture of individual cluster sizes and the total number of clusters in the simulation box, based on visual inspection. Of course, a more rigorous quantitative description based on Stillinger's clustering algorithm is obtained and the results are shown in Figures 3.11B, 3.11C, 3.12B, and 3.12C.

A detailed *RMSD* analysis of the simulations was done for determining the number of peptides in a particular cluster. Backbone *RMSD* accounts for relative backbone orientation between two peptides and hence is a better criterion than the centre of mass separation between two peptides. More specifically, backbone *RMSD* for a given pair of peptides is obtained for every simulation frame by calculating the root mean-square distance between the backbone atoms of the two peptides. In this manner, *RMSD* values are calculated for all possible peptide pairs in a simulation frame, and *RMSD* is used as a clustering criterion in the Stillinger algorithm to identify all possible clusters in that frame. Therefore, using an *RMSD* cut-off value, each cluster is defined by the number of peptides that are associated with each other in a particular aggregate. The backbone *RMSD* cut-off used for these simulations was calculated based on uperin 3.5 dimer simulations for both CG and FA representations. RMSD cut-off values of 14.5 Å for CG:20 and 16.25 Å for FA:20, respectively, were used to define the cluster size. The clusters were categorized as monomers, dimers, trimers, tetramers, pentamers, hexamers and multimers; based upon the number of peptides in a specific cluster. For analytical clarity, only changes in the population of tetramers have been plotted here for all the multi-peptide simulations. This is because no clear patterns could be deduced from the monomer, dimer and trimer populations in these simulations. Moreover, a negligible number of pentamers and hexamers were observed, and no higher order clusters were found.

At 0.00 M NaCl, four LACs containing two to five peptides were observed in the CG:20 trajectory snapshot in Figure 3.11A, with the remaining peptides randomly scattered throughout the simulation volume. Moreover, at c_{NaCl} = 0.15 M, four LACs containing two to seven peptides were present. As anticipated, aggregation was greater at c_{NaCl} = 0.50 M, where

three LACs consisting of three to nine peptides were observed. Thus, by increasing the concentration of NaCl from 0.00 M to 0.50 M, there is a clear trend towards an increase in the size of the largest LAC observed. Figure 3.11B shows that tetramer occurrences are extremely rare at c_{NaCl} = 0.00 M NaCl but increase at higher NaCl concentrations. A plot of the cumulative tetramer population (collected over the duration of the simulation) as a function of salt concentration (Figure 3.11C) also shows that tetramer (aggregate) formation increases at higher NaCl concentrations.
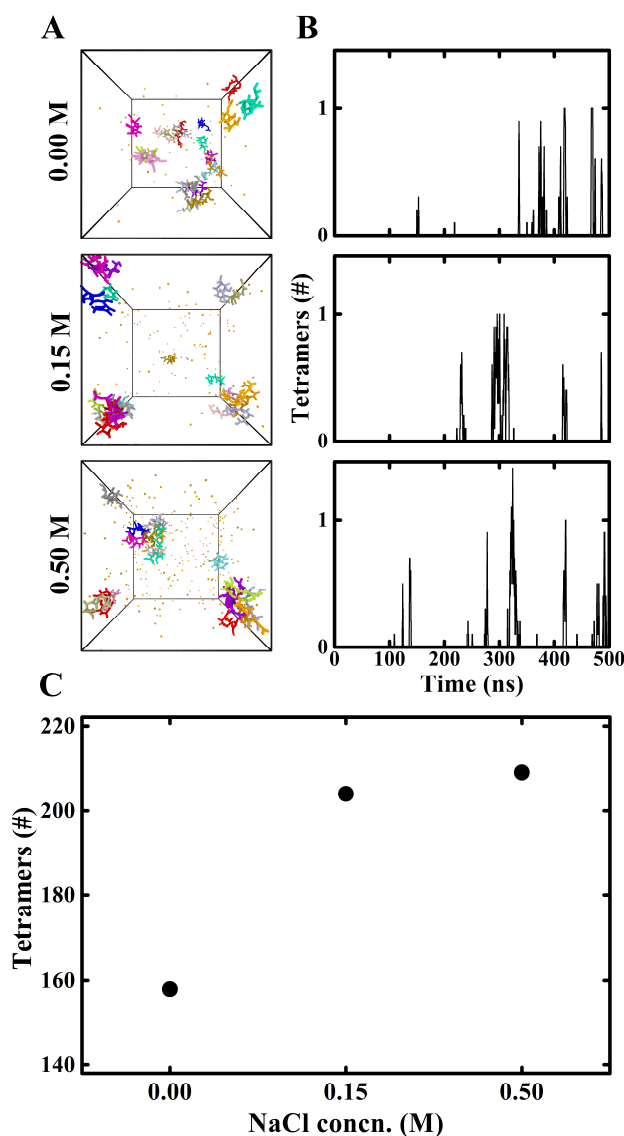
Figure 3.11. Aggregation of multiple peptides for CG:20. (A) Simulation snapshots showing peptide aggregation and cluster formation at different NaCl concentrations (increasing top to bottom). Sodium (Na$^+$) and chloride (Cl$^-$) ions have also been shown as brown and orange spheres, respectively. Water molecules have been omitted for visual clarity. (B) Evolution of tetramer populations over time for different NaCl concentrations (increasing top to bottom), and (C) cumulative tetramer population as a function of NaCl concentration.
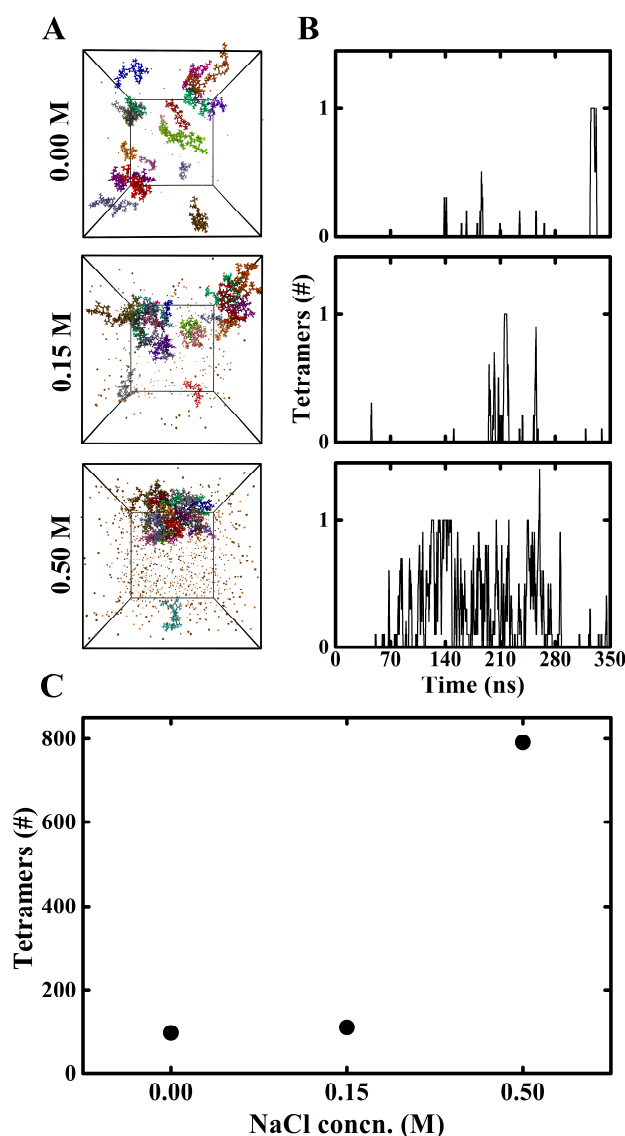
Figure 3.12. Aggregation of multiple peptides for FA:20. (A) Simulation snapshots showing peptide aggregation and cluster formation at different NaCl concentrations (increasing top to bottom). Sodium (Na^+) and chloride (Cl^-) ions have also been shown as brown and orange spheres, respectively. Water molecules have been omitted for visual clarity. (B) Evolution of tetramer populations over time for different NaCl concentrations (increasing top to bottom), and (C) cumulative tetramer population as a function of NaCl concentration.

The representative FA:20 trajectory snapshots in Figure 3.12A show approximately three LACs containing two to four peptides at c_{NaCl} = 0.00 M, with the remaining peptides distributed within the remaining volume of the simulation box. At 0.15 M NaCl, three LACs consisting of two to ten peptides were observed, with two remaining peptides not interacting. Finally, as anticipated, the maximum aggregation is observed with nineteen of the twenty peptides aggregating to form a single LAC at c_{NaCl} = 0.50 M, leaving only one peptide not associated. Thus, a progressive increase in the size of the largest LAC was observed as the salt

concentration increased from 0.00 M to 0.50 M NaCl. Figure 3.12B shows that tetramer occurrences are extremely rare at c_{NaCl} = 0.00 M, show a marginal increase at c_{NaCl} = 0.15 M, and a very large increase at c_{NaCl} = 0.50 M, clearly supporting the visual data of Figure 3.12A and showing that addition of salt leads to larger peptide aggregates in uperin 3.5. A plot of the cumulative tetramer population as a function of salt concentration (Figure 3.12C) also corroborates the same result that tetramer formation is favoured at higher salt concentrations.

3.3.2.2. Order parameter analysis

The CG simulations clearly showed that the AVSVI segment has a crucial role in inter-peptide interactions. The coordinates of α-carbon atoms for alanine (ninth position) and isoleucine (thirteenth position) residues were used to draw a unit vector to define the direction of the AVSVI segments. Vectors defined in this manner for individual peptides were used to calculate nematic and polar order parameters between multiple peptides.

The nematic order parameter, $\bar{P}_2$, in Equation 1 quantifies the relative orientation between different molecules or segments. A value of 1.0 denotes perfect alignment within a defined segment and 0.0 indicates no ordering[51, 60].

$$\bar{P}_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\left(\hat{z}_i.\hat{d}\right)^2 - \frac{1}{2} \quad (1)$$

Furthermore, the polar order parameter, $\bar{P}_1$, in Equation 2 defines the overall polarity of a system. Again, an absolute value of 1.0 would correspond to parallel orientation of vectors, whereas 0.0 would be the value for anti-parallel alignment within the system. Intermediate values would be obtained for a mixed strand orientation.

$$\bar{P}_1 = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{z}_i.\hat{d}\right) \quad (2)$$

Here, $\hat{d}$ is a unit vector that denotes the preferred direction of alignment, whereas $\hat{z}_i$ is a vector defined by a pair of suitably chosen three-dimensional coordinates obtained from the molecule or segment of interest, and N is the total number of molecules or segments being considered for the particular analysis.
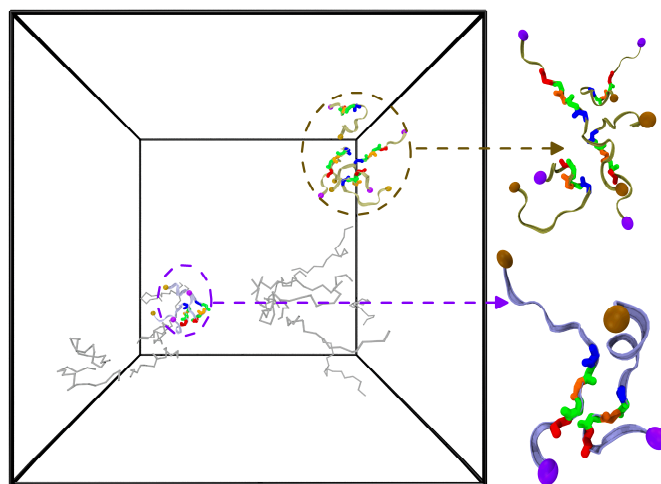
Figure 3.13. A representative trajectory snapshot of peptides showing several loosely associated clusters, taken at 100 ns from a multi-peptide FA:20 simulation with a salt concentration of 0.15 M. In particular, two stable LACs consisting of two and four peptides have been highlighted here.

In the multi-peptide FA simulations, LACs consisting of up to four peptides were observed; which remained stable over most of the simulation period. Hence, two stable LACs consisting of two and four interacting peptides were selected from each of the three salt concentrations of 0.00 M, 0.15 M, and 0.50 M NaCl to study the relative interaction between the constituent peptides with increasing salt concentration. Figure 3.13 shows two such LACs at c_{NaCl} = 0.15 M, a "dimer" (blue) and a "tetramer" (tan), which were selected and tracked over the course of the simulation. Then the relative ordering and interaction within these LACs was analysed over the simulation period.

At c_{NaCl} = 0.00 M, the peptides in the dimer have lower ordering (0.59) as per the nematic order parameter ($\bar{P}_2$), as shown in Figure 3.14A, compared with the dimers at 0.15 M (0.96) and 0.50 M (0.85) NaCl concentrations. The polar order parameter ($\bar{P}_1$) indicates that the peptides at 0.00 M NaCl have mixed strand orientations (0.85). However, the peptides at 0.15 M (0.99) and 0.50 M (0.00) NaCl concentrations are in a parallel and an anti-parallel arrangement, respectively. This indicates not only a higher level of ordering with increasing NaCl concentrations, but also a lack of preference for a parallel versus an anti-parallel alignment between the respective AVSVI segments of the two peptides. The time evolution of the two order parameters for the dimers also show similar trends (Figures 3.15A and 3.15B). The variation of the R_g values for the dimers at the three salt concentrations is shown in Figure 3.14B. The R_g plot in Figure 3.14B at c_{NaCl} = 0.00 M shows significant fluctuations, with an average R_g of ≈ 12.5 Å. At c_{NaCl} = 0.15 M, the average R_g is ≈ 11.5 Å, and at c_{NaCl} = 0.50 M,

the dimer has an average R_g of ≈ 10 Å. This clearly indicates the increase in hydrophobic interaction with increasing salt concentration due to the higher charge screening capacity.
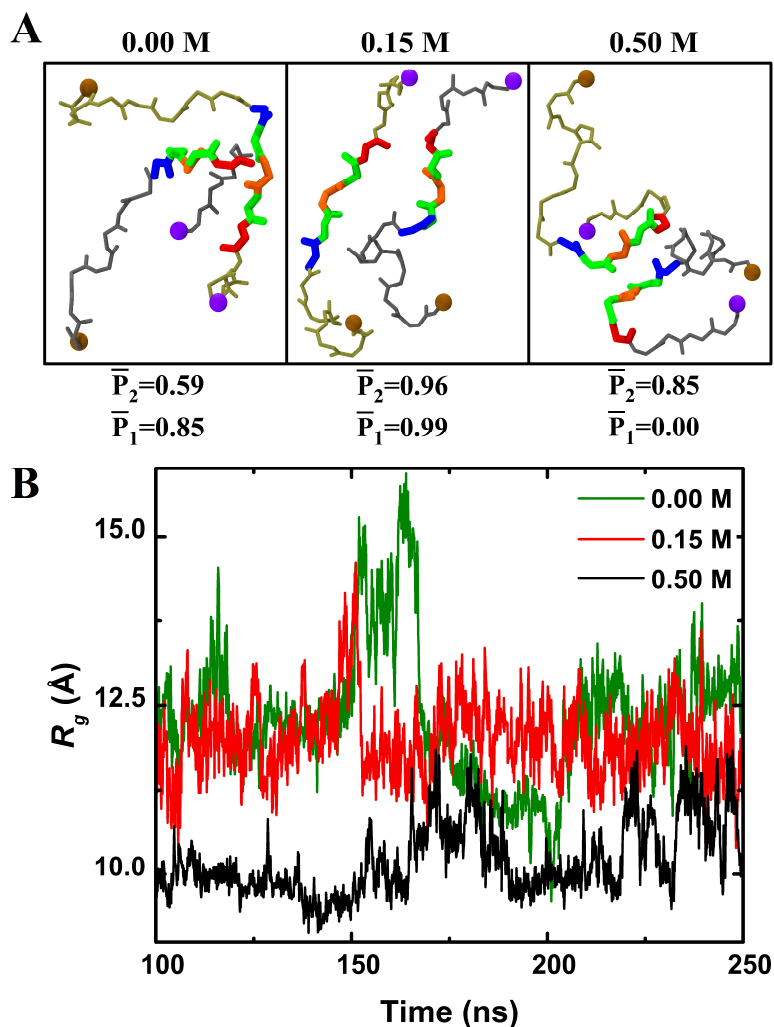
Figure 3.14. (A) Snapshots of dimer LACs at 200 ns, with nematic ($\bar{P}_2$) and polar ($\bar{P}_1$) order parameter values for all NaCl concentrations of 0.00 M, 0.15 M and 0.50 M. (B) R_g of the dimer LACs plotted as a function of time for all NaCl concentrations of 0.00 M, 0.15 M and 0.50 M.
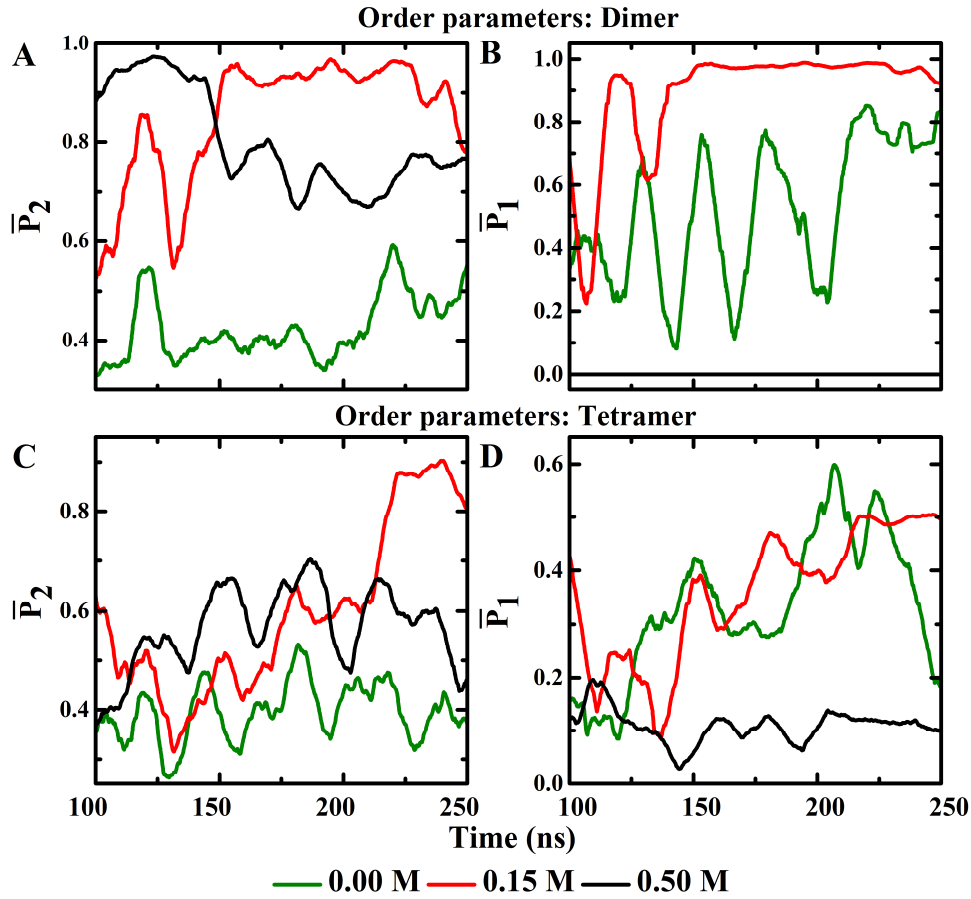
Figure 3.15. Nematic ($\bar{P}_2$) and polar ($\bar{P}_1$) order parameter plots for dimer (A, B) and tetramer (C, D) LACs; obtained from FA:20 simulations for NaCl concentrations of 0.00, 0.15 and 0.50 M. In the case of dimers, higher nematic and polar ordering is present in 0.15 and 0.50 M systems compared to no salt case. Similarly, for the tetramers, higher nematic ordering is observed in 0.15 M and 0.50 M cases compared to 0.00 M system. However, no conclusive trend could be noted in polar order parameter plots of the tetramers. Hence, as indicated by these plots, inter-peptide interaction increases upon introduction of NaCl and leads to higher ordering within the cluster.

If tetramers are considered (Figure 3.16), the constituent peptides have a relatively low nematic order parameter ($\bar{P}_2$) at 0.00 M NaCl (0.28) compared to those peptides at 0.15 M (0.55) and 0.50 M (0.54) NaCl, indicating less ordering of the peptides; as seen in Figure 3.16A. Furthermore, no discernible pattern could be derived from the polar order parameters ($\bar{P}_1$). Thus, no clear conclusion could be derived from the data in Figure 3.16A regarding the preferred alignment direction between AVSVI segments, i.e. whether parallel or anti-parallel. Similar inferences can be drawn from the time evolution plots of $\bar{P}_1$ and $\bar{P}_2$ for these tetramers (Figures 3.15C and 3.15D). The R_g average in the tetramer simulations is shown in Figure 3.16B, and is $\approx$ 22 Å for 0.00 M NaCl, $\approx$ 14 Å for 0.15 M NaCl, and $\approx$ 12 Å for 0.50 M NaCl.

Thus, the tetramer shows significantly lower values of R_g (and also smaller fluctuations in R_g) in presence of NaCl, clearly indicating that the presence of salt leads to enhanced attractive interactions between the peptides. Similar trends in R_g values were also observed in the case of the dimers in Figure 3.14B.
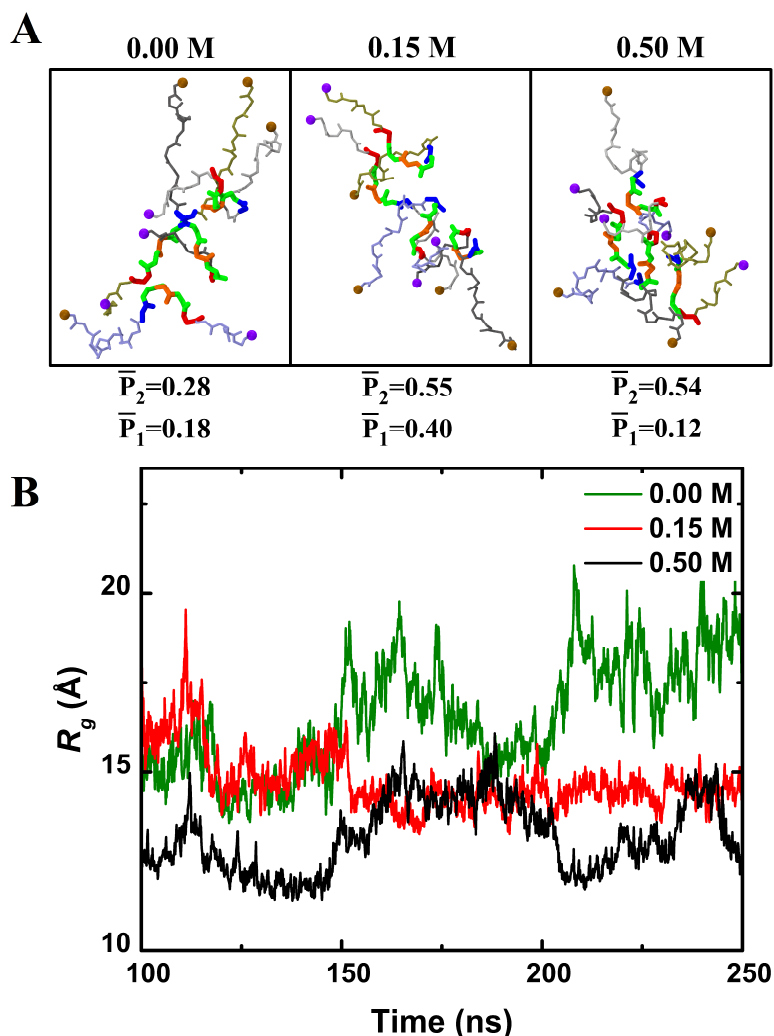
Figure 3.16. (A) Snapshots of tetramer LACs at 100 ns, with nematic ($\bar{P}_2$) and polar ($\bar{P}_1$) order parameter values for all NaCl concentrations of 0.00 M, 0.15 M and 0.50 M. (B) R_g of the tetramer LACs plotted as a function of time for all NaCl concentrations of 0.00 M, 0.15 M and 0.50 M.

3.3.2.3. Secondary structure evolution during aggregation

In the CG simulations based on the MARTINI model, the secondary structure is predefined and thus invariable[21]. Therefore, any conformational changes involving the evolution of secondary structure over the course of the simulation are not accessible using MARTINI

coarse-grained simulations. Moreover, the starting conformations in all our simulations are random coils. However, information can be obtained from the secondary structure analysis[61] of the peptide trajectories from twenty peptide FA simulations.
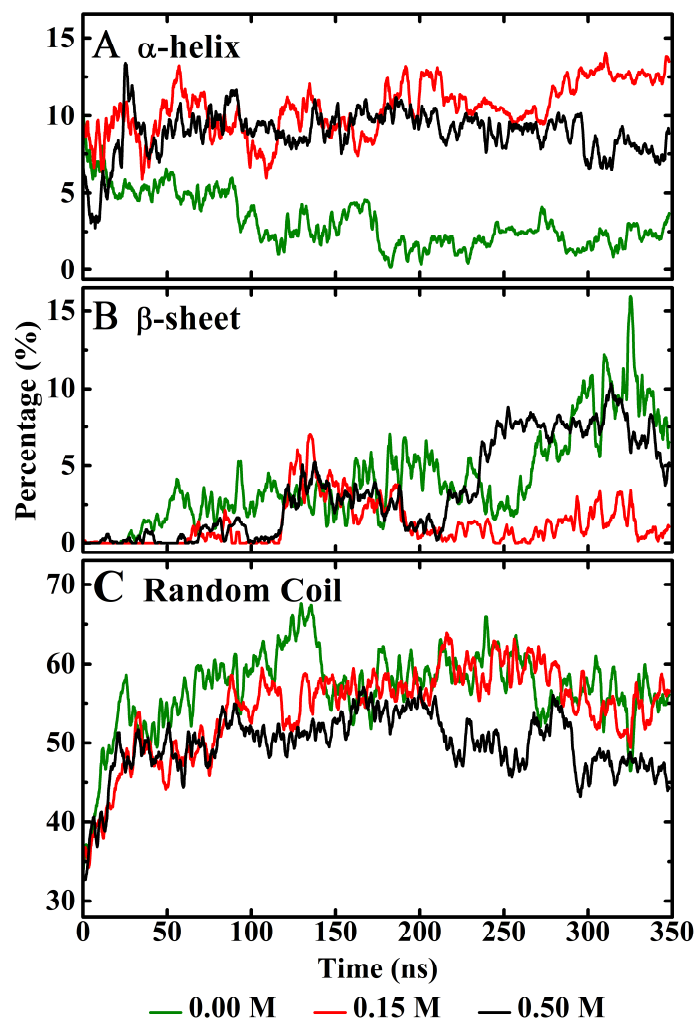
Figure 3.17. Plots showing time evolution of (A) α-helix, (B) β-sheet, and (C) random coil secondary structure components of the peptides, obtained from the FA:20 simulations at different NaCl concentrations.

Secondary structure analysis examines peptide conformational changes that take place over time in terms of the evolution of the various secondary structure elements, including α-helix, 3_{10}-helix, π-helix, β-sheet, isolated β-bridge, random coil and turn structures. The time evolution plots of α-helix and β-sheet elements over 350 ns long FA:20 simulations at the three different NaCl concentrations are shown in Figures 3.17A and 3.17B, respectively. Significant differences with respect to the salt concentration were observed only for the α-helix and the β-sheet elements (shown in Figure 3.17). In the absence of salt, the α-helical content is found to steadily decrease with time and settles to a very low value of nearly 2% after 180 ns. In contrast,

when salt is present, the α-helical content increases and reaches a steady-state value of close to 10% at an earlier time (≈ 50 ns). Although, post 200 ns, the α-helical content is somewhat higher at 0.15 M (≈ 12%) than at 0.50 M NaCl (≈ 9%). In comparison to the α-helical content, the evolution of the β-sheet element shows a reverse trend in Figure 3.17B. The β-sheet content remains negligible (≈ 1%) for the intermediate 0.15 M NaCl concentration. However, it is found to steadily increase for the 0.00 M case and attains an average value of approximately 10% towards the end of the simulation. These results seem to suggest that addition of salt to the solution promotes peptide conformations with higher α-helical content. In the absence of salt, peptide conformations with higher β-sheet content and negligible α-helical content are noted. Interestingly, at c_{NaCl} = 0.50 M, conformations with significant amounts of both α-helical and β-sheet content seem to be favoured.
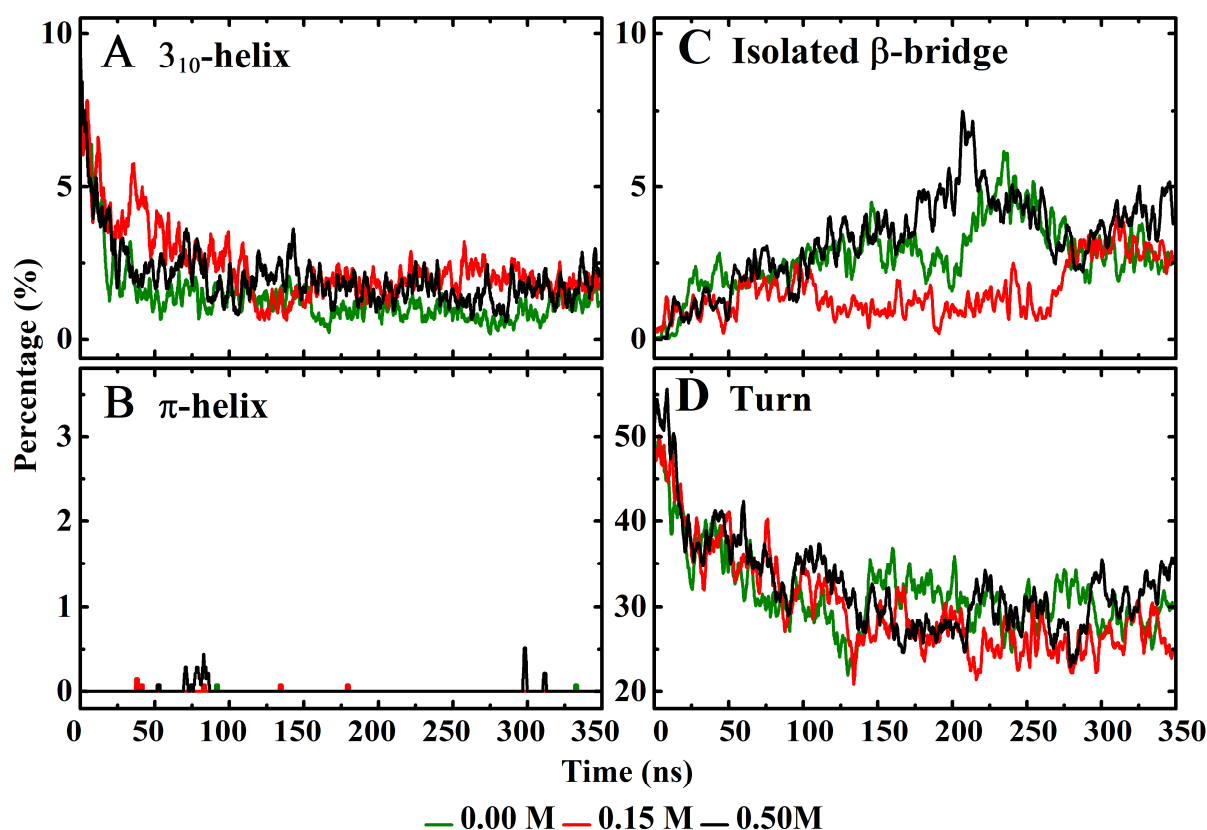
Figure 3.18. Plots showing time evolution of (A) 3_{10}-helix, (B) π-helix, (C) isolated β-bridge and (D) turn secondary structure components of the peptides, obtained from the FA:20 simulations at different NaCl concentrations. Negligible π-helical content could be noted at all three salt concentrations. Whereas after ≈ 100 ns, the 3_{10}-helix and turn structure fractions are at similar values for all three salt concentrations.

Isolated β-bridge plots for the three salt concentrations (Figure 3.18C) follow a trend similar to β-sheet plots in the 110–270 ns time range. A higher isolated β-bridge content (≈ 4%) is

observed for the salt-free (0.00 M) and high salt (0.50 M) cases compared to intermediate salt concentration of 0.15 M ($\approx$ 1%). Significantly after $\approx$ 210 ns, a lower random coil content ($\approx$ 49%) is noted at 0.50 M NaCl in Figure 3.17C compared to 0.00 M and 0.15 M NaCl systems ($\approx$ 58%). This could be due to the higher α-helical and β-sheet content at $c_{NaCl} = 0.50$ M relative to 0.00 M and 0.15 M NaCl, as noted earlier in Figures 3.17A and 3.17B, respectively.
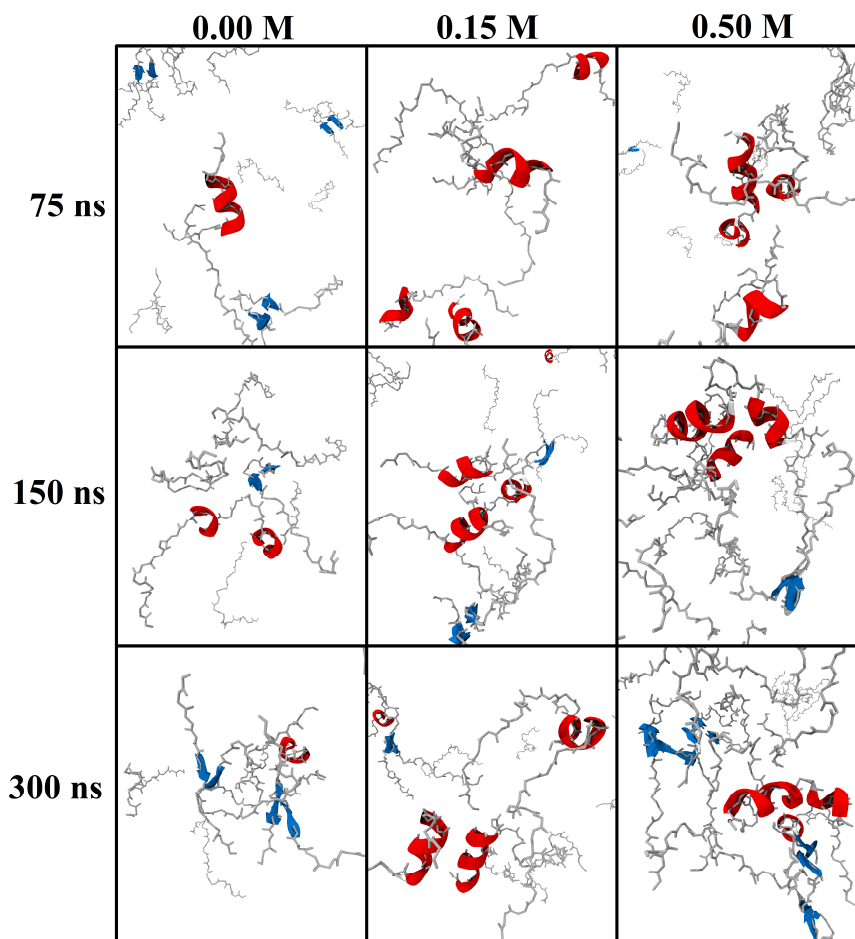
Figure 3.19. Simulation snapshots showing the time evolution (time increases from top to bottom) of α-helical and β-sheet components in the FA:20 simulations for different NaCl concentrations (salt concentration increases from left to right).

Figure 3.19 shows simulation snapshots emphasizing on the evolution of the α-helical and β-sheet regions at different time frames of the FA:20 simulations at the three salt concentrations. The results are consistent with the results of Figure 3.17 and clearly show that the α-helical content is low at 0.00 M NaCl. In contrast, the α-helical content is significantly higher at c_{NaCl} = 0.15 M and 0.50 M. However, the β-sheet content appears to be higher at both 0.00 M and 0.50 M NaCl, but nearly vanishes at 300 ns at the intermediate 0.15 M salt concentration. We note that the snapshots may not be at the same magnification and may not show all the twenty

peptides. However, the corresponding snapshots in Figure 3.19 and Figure 3.12A are from the same simulations, which showed an increase in aggregation between peptides and the aggregate size with increasing salt concentration. The results from Figures 3.12A, 3.17 and 3.19 suggest that a greater α-helical content with increasing salt concentration appears to promote greater aggregation between the peptides. Further, at much higher salt concentration (0.50 M NaCl), significant amounts of both α-helical and β-sheet content appear, which lead to still larger aggregate sizes.
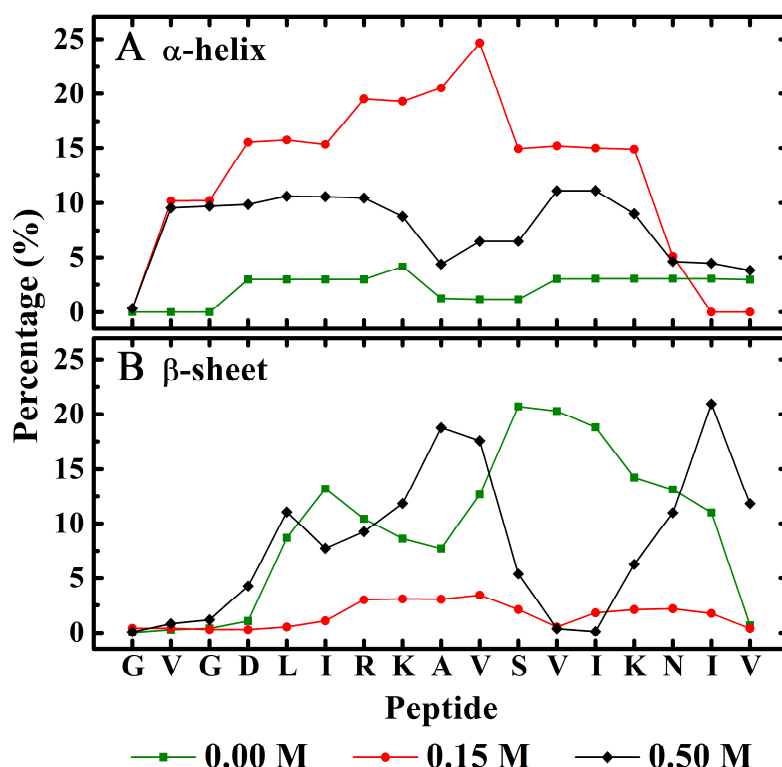
Figure 3.20. Plots showing the variation of (A) α-helix and (B) β-sheet secondary structure components along the peptide length at different salt concentrations.

The variation of α-helical and β-sheet content across the peptide sequence for different salt concentrations is shown in Figure 3.20. The percentage of these secondary structure elements at each peptide residue is calculated by taking an average over the 300–350 ns time period. Negligible α-helical content ($\approx 2\%$) was observed in the no salt case in Figure 3.20A. However, at the intermediate salt concentration of 0.15 M, a high α-helical content of 17% was observed in the D^4LIRKAVSVIK^{14} region. Interestingly, an overall decrease in the α-helical content is observed as the salt concentration is further increased to 0.50 M, although the α-helical content is still significant in the V^2GDLIRKAVSVIK^{14} region ($\approx 9\%$). The overall α-helical content at 0.50 M, though slightly reduced in comparison to the 0.15 M case, is still much higher when

92

compared to the no salt case. At c_{NaCl} = 0.00 M, a high β-sheet content of 13% is observed in the L[5]IRKAVSVIKNI[16] region in Figure 3.20B. In contrast to the increase in α-helical content observed with addition of salt, the β-sheet content reduces to less than 2% as the salt concentration increases to 0.15 M. However, a rise in β-sheet content is observed at c_{NaCl} = 0.50 M compared to the 0.15 M system, primarily localized at the D[4]LIRKAVS[11] (≈ 12%) and K[14]NIV[17] (≈ 12%) regions.

The residue-specific secondary structure analysis in Figure 3.20 shows that the addition of salt leads to screening of the charged residues (previously discussed for D4, R7, K8, and K14 residues in Figure 3.6) and is accompanied by a significant change to an α-helical structure for most of the peptide sequence. The increase in α-helical content, arising from addition of salt, is accompanied by a nearly equivalent loss in the β-sheet content for most of the peptide sequence when NaCl concentration increases to 0.15 M. However, at 0.50 M NaCl, both α-helical and β-sheet contents are significant. Since peptide aggregation is also seen to significantly increase with addition of salt, our results suggest a correlation between aggregation and conformational change to an α-helical structure. A conformational change to an α-helical structure may lead to a more amphipathic conformation for the peptide, which in turn would provide better shielding of the hydrophobic residues from water in an aggregated state. In contrast, a correlation between aggregation and evolution of β-sheet content does not emerge very clearly from the current simulation data. Longer simulations would provide more definitive insight into further evolution of the secondary structure during the aggregation process. However, the time scales are computationally prohibitive, since conformational transitions, such as helix formation in peptides, typically take place on the time scales of hundreds of nanoseconds to hundreds to microseconds[20].

Sidechains of hydrophobic amino acid residues (alanine, isoleucine, leucine and valine) of the peptide regions in α-helical conformation could be observed in the inter-helical region in Figure 3.21 at c_{NaCl} = 0.15 M and 0.50 M. In contrast, the charged acidic (aspartic acid) and basic (arginine and lysine) amino acid sidechains are excluded from the inter-helical core, as they preferentially interact with the surrounding water and ion molecules (not shown for the sake of visual clarity). These α-helical clusters could not be observed for the salt-free case (0.00 M NaCl) due to negligible charge-screening by counterions, as discussed earlier for Figure 3.6. Therefore, although the inter-peptide interaction is dominated by hydrophobic interactions

between the residue sidechains, effective charge-screening by counterions appears to facilitate the same.
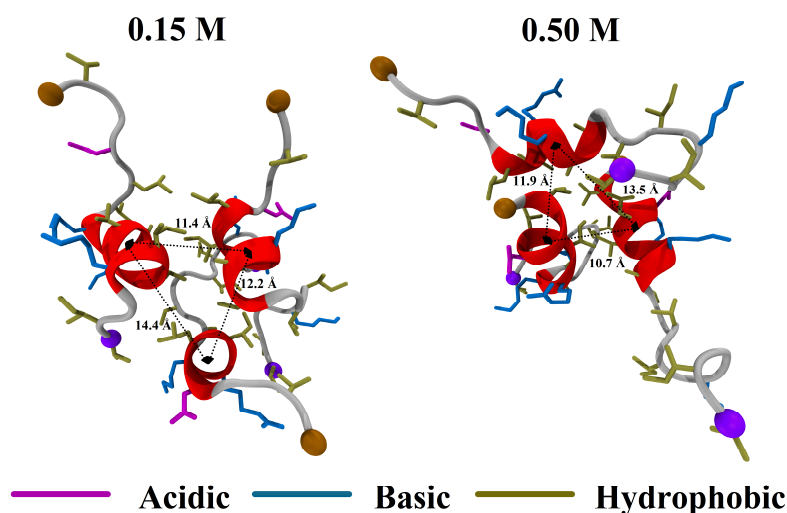
Figure 3.21. Snapshots at 150 ns from FA:20 simulations showing significant α-helical content in peptide aggregates comprising three uperin 3.5 monomers at 0.15 M and 0.50 M NaCl.

In the absence of salt, a large number of β-sheet structural elements could be observed. However, as electrostatic screening of charged residue sidechains is insignificant at 0.00 M NaCl, stable peptide aggregates were not observed. Introduction of salt screens the peptide charges, which reduces the range of electrostatic repulsion and leads to a greater contribution from hydrophobic interactions to the overall interaction between peptides. Addition of salt also results in peptide regions with greater α-helical content, which in an aggregated state would provide a water-excluded environment to the hydrophobic residues. Moreover, high salt concentration (0.50 M) also stabilizes the β-sheet structural elements on account of significantly reduced electrostatic repulsion between the charged peptide sidechains. Hence, there seem to be two aspects of increased aggregation with addition of salt in the case of uperin 3.5, namely, electrostatic screening of charged residues, and secondary structure changes.

3.4. Discussion

Addition of salt to a peptide solution can screen the positive and negative charges along the polypeptide, reduce electrostatic repulsion between charged residues, and lead to an effective increase in hydrophobic interactions between the peptides[18]. This should facilitate aggregation between the peptides, and the enhanced aggregation kinetics with increase in salt concentration can be understood in the context of the Derjaguin-Landau-Verwey-Overbeek (DLVO) theory

of colloidal stability[62]. For most colloidal systems, charges are more or less uniformly distributed on the surfaces, and colloidal particles rarely show conformational diversity in terms of secondary structures. However, charge distribution along a polypeptide sequence is usually non-uniform, and charged residues may appear in stretches, interspersed with stretches of hydrophobic residues. In case of uperin 3.5, there are only four charged residues non-uniformly distributed along the sequence; with stretches of hydrophobic residues in between. As salt concentration increases, peptides can undergo conformational change[30], which can also influence aggregation kinetics. Thus, the aggregation of peptides in the presence of salt would depend on a combination of factors, including electrostatic screening, modified peptide-peptide interactions due to conformational change[30], and enhanced aggregation arising from stretches of hydrophobic residues[63].

In the CG simulations of two uperin 3.5 peptides, addition of NaCl to the system lead to a steady decrease in the centre of mass distances between the two peptides (Figure 3.2), indicating a greater attraction between the peptides at higher salt concentration. A thorough analysis of residue-residue contacts between the two peptides (Figure 3.3) showed interactions between predominantly hydrophobic segments (AVSVI, residues 9–13) increase with increasing salt concentration, with the highest number of residue-residue interactions observed at the highest salt concentration of 0.50 M NaCl. As expected, electrostatic screening of the charges on the charged residues (positively-charged residues at R7, K8 and K14 and negatively-charged residues at D4) increased with increasing salt concentration (Figure 3.6). The positively-charged residues at R7, K8 and K14 border the hydrophobic AVSVI stretch along the peptide. Electrostatic screening of the charged residues should result in reduced repulsion between charged residues on the two interacting peptides, and therefore a corresponding increase in attraction between the hydrophobic AVSVI segments on the two peptides.

Qualitatively, the aggregation behaviour of the two uperin 3.5 peptides agrees well with a DLVO approach to colloidal aggregation. However, PMF plots for formation of dimer aggregates do not show significant quantitative differences at different salt concentrations (Figure 3.10). The well depths of the attractive minima in the free energy landscape do not change significantly with NaCl concentration. This is in contrast with observations of Figures 3.2 and 3.6, where there are clear indications of increasing attraction between uperin 3.5 peptides with increasing salt concentration. In addition, a comparison of CG and FA

simulations of systems of 20 uperin peptides (CG:20 vs. FA:20 in Figures 3.11 and 3.12, respectively) suggests that electrolyte effects are more pronounced in the case of FA simulation. Although, both CG and FA simulations show increased aggregation with increase in salt concentration, larger sized clusters are more frequently observed for the FA simulations in comparison to the CG simulations. The greater extent of aggregation observed in the case of FA:20 simulation could be due to the softer potentials employed in CG simulations[64]. Thus, the reduced molecular friction in CG systems might affect the peptide-peptide interactions, and result in greater backbone flexibility compared to the FA systems[64]. This is especially relevant for a relatively short peptide like uperin 3.5 that does not contain any proline or aromatic amino acid residues[9] that might stabilize backbone orientations. These two results, namely the PMF plots (Figure 3.10) and more pronounced aggregation for FA:20 simulations (Figure 3.12), suggest that conformational evolution of the peptide structure may have a significant influence on the aggregation kinetics of uperin 3.5.

A detailed analysis of the evolution of secondary structure components of peptides from the FA:20 simulations showed that the addition of salt lead to significant changes in the relative fractions of α-helical and β-sheet contents. Although approximately 85% of the peptide structure is comprised of random coils and turns, nearly 10–12% of the structure is distributed between α-helical and β-sheet components at all salt concentrations (Figures 3.17 and 3.18). Whereas in the absence of salt the β-sheet component is dominant with very little α-helical content, the opposite trend is observed at 0.15 M NaCl. With increase in salt concentration (0.15 M NaCl), the α-helical content increases with a corresponding decrease in the β-sheet component (Figure 3.17). At the highest salt concentration of 0.50 M NaCl, significant fractions of both α-helical and β-sheet components are observed, with a marginal decrease in the random coil component at 0.50 M NaCl. The increase in α-helical content with increase in salt concentration is a direct consequence of charge screening at the charged residues, and the largest increase in α-helical content is indeed observed in the D[4]LIRKAVSVIK[14] region (Figure 3.20). We find that a larger α-helical content at high salt concentrations would lead to more amphipathic peptide conformations, which would facilitate aggregation of peptides due to formation of a core region populated by sidechains of hydrophobic residues that is relatively well-shielded from the surrounding aqueous environment. This is clearly observed in simulation snapshots at both 0.15 M and 0.50 M NaCl in Figure 3.21. Thus, a correlation between an increase in α-helical content and enhanced aggregation for uperin 3.5 with addition of NaCl is observed. In a related study, Calabrese *et al.*[8] have also shown that addition of TFE

to a uperin 3.5 assay induces α-helical content by stabilizing hydrogen bonds within the peptide, and also enhances peptide aggregation at lower concentrations of TFE (< 20 %, v/v). In summary, addition of salt promotes transitions to α-helical conformations, which might facilitate further transformations to peptide structures rich in β-sheet conformations. However, amyloid-like fibrils composed of α-helices might be obtained instead, as noted for phenol-soluble modulin α3[65] and certain ionic-complementary peptides[66].

3.5. Conclusions

Both CG and FA MD simulations were used to study the effect of NaCl concentration on the aggregation process of an amyloidogenic peptide, uperin 3.5, in an aqueous environment. The simulations show that uperin 3.5 aggregates in the presence of salt, and its aggregation propensity increases with an increase in salt concentration, which is in agreement with previous experimental observations[8, 67]. A detailed analysis of residue-residue contacts shows that peptide aggregation is primarily driven by interactions along the hydrophobic AVSVI stretch (residues 9 – 13) of uperin 3.5. In the presence of NaCl media, chloride ions screen the positively-charged R7, K8 and K14 residues located at either side of the AVSVI region, and therefore reduce the electrostatic repulsion between two uperin 3.5 peptides. The extent of this screening effect was higher at higher NaCl concentrations, resulting in an increased hydrophobic interaction and enhanced aggregation between the peptides. Our results also show a strong correlation between increased α-helicity of the uperin 3.5 peptide and its aggregation in presence of NaCl. In contrast, in the absence of salt, uperin 3.5 conformations are dominated by random coil components, with very little α-helical content, and show little or no propensity to aggregate. This supports the hypothesis that an α-helical intermediate lies on the pathway towards amyloid formation, and facilitates the initial stages of peptide aggregation[8]. Thus, in an aqueous environment, the addition of salt enhances fibril formation by stabilizing the secondary structure of intermediates along the fibril pathway. Since uperin 3.5 wild-type is known to adopt a predominantly random coil conformation in water, the peptide lacks secondary structure components in its native state that may act as templates/precursors for further amyloid formation. Thus, in the context of relatively short, random coil peptides, this study highlights the crucial roles played by the peptide sequence and charged residues in the process of aggregation at molecular level as a function of salt concentration.

3.6. References

1.   Chiti, F.; Dobson, C. M., Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333-66.

2.   Lansbury, P. T.; Lashuel, H. A., A Century-Old Debate on Protein Aggregation and Neurodegeneration Enters the Clinic. *Nature* **2006**, *443*, 774-9.

3.   Hardy, J.; Selkoe, D. J., The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002**, *297*, 353-6.

4.   Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27-68.

5.   Nagy, Z.; Esiri, M. M.; Jobst, K. A.; Morris, J. H.; King, E. M.; McDonald, B.; Joachim, C.; Litchfield, S.; Barnetson, L.; Smith, A. D., The Effects of Additional Pathology on the Cognitive Deficit in Alzheimer Disease. *J. Neuropathol. Exp. Neurol.* **1997**, *56*, 165-70.

6.   Kirkitadze, M. D.; Bitan, G.; Teplow, D. B., Paradigm Shifts in Alzheimer's Disease and Other Neurodegenerative Disorders: The Emerging Role of Oligomeric Assemblies. *J. Neuro. Res.* **2002**, *69*, 567-577.

7.   Eisenberg, D. S.; Sawaya, M. R., Structural Studies of Amyloid Proteins at the Molecular Level. *Annu. Rev. Biochem.* **2017**, *86*, 69-95.

8.   Calabrese, A. N.; Liu, Y.; Wang, T.; Musgrave, I. F.; Pukala, T. L.; Tabor, R. F.; Martin, L. L.; Carver, J. A.; Bowie, J. H., The Amyloid Fibril-Forming Properties of the Amphibian Antimicrobial Peptide Uperin 3.5. *ChemBioChem* **2016**, *17*, 239-246.

9.   Bradford, A. M.; Bowie, J. H.; Tyler, M. J.; Wallace, J. C., New Antibiotic Uperin Peptides from the Dorsal Glands of the Australian Toadlet Uperoleia Mjobergii. *Aust. J. Chem.* **1996**, *49*, 1325-1331.

10.   Dobson, C. M., Protein Folding and Misfolding. *Nature* **2003**, *426*, 884-890.

11.   Pastor, M. T.; Kümmerer, N.; Schubert, V.; Esteras-Chopo, A.; Dotti, C. G.; de la Paz, M. L.; Serrano, L., Amyloid Toxicity Is Independent of Polypeptide Sequence, Length and Chirality. *J. Mol. Biol.* **2008**, *375*, 695-707.

12. Cruz, L.; Urbanc, B.; Borreguero, J. M.; Lazo, N. D.; Teplow, D. B.; Stanley, H. E., Solvent and Mutation Effects on the Nucleation of Amyloid Beta-Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18258-63.

13. Bouzakraoui, S.; Mousseau, N., Structural and Thermodynamical Properties of Early Human Amylin Oligomers Using Replica Exchange Molecular Dynamics: Mutation Effect of Three Key Residues F15, H18 and F23. *Phys. Chem. Chem. Phys.* **2017**, *19*, 31290-31299.

14. Wei, G.; Shea, J. E., Effects of Solvent on the Structure of the Alzheimer Amyloid-Beta(25-35) Peptide. *Biophys. J.* **2006**, *91*, 1638-47.

15. Zhang, M.; Zhao, J.; Zheng, J., Molecular Understanding of a Potential Functional Link between Antimicrobial and Amyloid Peptides. *Soft Matter* **2014**, *10*, 7425-7451.

16. Xing, Y.; Pilkington, E. H.; Wang, M.; Nowell, C.; Kakinen, A.; Sun, Y.; Wang, B.; Davis, T. P.; Ding, F.; Ke, P. C., Lysophosphatidylcholine Modulates the Aggregation of Human Islet Amyloid Polypeptide. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30627-30635.

17. Hong, Y.; Pritzker, M. D.; Legge, R. L.; Chen, P., Effect of Nacl and Peptide Concentration on the Self-Assembly of an Ionic-Complementary Peptide Eak16-Ii. *Colloids Surf. B* **2005**, *46*, 152-161.

18. Aramvash, A.; Seyedkarimi, M. S., All-Atom Molecular Dynamics Study of Four Rada 16-I Peptides: The Effects of Salts on Cluster Formation. *J. Clust. Sci.* **2015**, *26*, 631-643.

19. Smith, M. D.; Cruz, L., Effect of Ionic Aqueous Environments on the Structure and Dynamics of the Aβ21–30 Fragment: A Molecular-Dynamics Study. *J. Phys. Chem. B* **2013**, *117*, 6614-6624.

20. Morriss-Andrews, A.; Shea, J.-E., Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643-666.

21. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J., The Martini Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819-834.

22.   Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K., Coarse Grained Protein-Lipid Model with Application to Lipoprotein Particles. *J. Phys. Chem. B* **2006**, *110*, 3674-84.

23.   Shih, A. Y.; Freddolino, P. L.; Arkhipov, A.; Schulten, K., Assembly of Lipoprotein Particles Revealed by Coarse-Grained Molecular Dynamics Simulations. *J. Struct. Biol.* **2007**, *157*, 579-92.

24.   Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H., The Martini Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812-7824.

25.   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005**, *26*, 1781-802.

26.   Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Sligar, S. G.; Schulten, K., Assembly of Lipids and Proteins into Lipoprotein Particles. *J. Phys. Chem. B* **2007**, *111*, 11095-11104.

27.   Miao, L.; Schulten, K., Transport-Related Structures and Processes of the Nuclear Pore Complex Studied through Molecular Dynamics. *Structure* **2009**, *17*, 449-459.

28.   Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

29.   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

30.   Liao, C.; Esai Selvan, M.; Zhao, J.; Slimovitch, J. L.; Schneebeli, S. T.; Shelley, M.; Shelley, J. C.; Li, J., Melittin Aggregation in Aqueous Solutions: Insight from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119*, 10390-8.

31.   MacKerell, A. D., et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586-616.

32.   MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698-9.

33.   Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr., Optimization of the Additive Charmm All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Phi, Psi and Side-Chain Chi(1) and Chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257-3273.

34.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

35.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

36.   Miller, C. A.; Gellman, S. H.; Abbott, N. L.; de Pablo, J. J., Association of Helical B-Peptides and Their Aggregation Behavior from the Potential of Mean Force in Explicit Solvent. *Biophys. J.* **2009**, *96*, 4349-4362.

37.   Laio, A.; Parrinello, M., Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562-12566.

38.   Grubmüller, H., Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E* **1995**, *52*, 2893-2906.

39.   Huber, T.; Torda, A. E.; van Gunsteren, W. F., Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. *J. Comput. Aided Mol. Des.* **1994**, *8*, 695-708.

40.   Bussi, G.; Laio, A.; Parrinello, M., Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **2006**, *96*, 90601-90604.

41.   Iannuzzi, M.; Laio, A.; Parrinello, M., Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2003**, *90*, 238302.

42.   Fiorin, G.; Klein, M. L.; Hénin, J., Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111*, 3345-3362.

43.   Tsigelny, I. F.; Sharikov, Y.; Kouznetsova, V. L.; Greenberg, J. P.; Wrasidlo, W.; Gonzalez, T.; Desplats, P.; Michael, S. E.; Trejo-Morales, M.; Overk, C. R., Structural Diversity of Alzheimer's Disease Amyloid-B Dimers and Their Role in Oligomerization and Fibril Formation. *J. Alzheimer's Dis.* **2014**, *39*, 583-600.

44. Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M., Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157-64.

45. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.

46. Stone, J. E., An Efficient Library for Parallel Ray Tracing and Animation. *Masters Theses. http://scholarsmine.mst.edu/masters_theses/1747* **1998**, *1747*.

47. Brown, A. M.; Lemkul, J. A.; Schaum, N.; Bevan, D. R., Simulations of Monomeric Amyloid B-Peptide (1–40) with Varying Solution Conditions and Oxidation State of Met35: Implications for Aggregation. *Arch. Biochem. Biophys.* **2014**, *545*, 44-52.

48. Yang, Y. I.; Gao, Y. Q., Computer Simulation Studies of Aβ37–42 Aggregation Thermodynamics and Kinetics in Water and Salt Solution. *J. Phys. Chem. B* **2014**, *119*, 662-670.

49. Buff, F. P.; Stillinger Jr, F. H., Surface Tension of Ionic Solutions. *J. Chem. Phys.* **1956**, *25*, 312-318.

50. Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A., Wordom: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* **2007**, *23*, 2625-2627.

51. Katyal, N.; Deep, S., Inhibition of Gnnqqny Prion Peptide Aggregation by Trehalose: A Mechanistic View. *Phys. Chem. Chem. Phys.* **2017**, *19*, 19120-19138.

52. Frishman, D.; Argos, P., Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566-579.

53. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577-2637.

54. Isralewitz, B., Timeline: A Vmd Plugin for Trajectory Analysis. Tutorial. *University of Illinois at Urbana-Champaign c) http://www.ks.uiuc.edu/Research/vmd* **2012**.

55. Seo, M.; Rauscher, S.; Pomès, R. g.; Tieleman, D. P., Improving Internal Peptide Dynamics in the Coarse-Grained Martini Model: Toward Large-Scale Simulations of Amyloid-and Elastin-Like Peptides. *J. Chem. Theory Comput.* **2012**, *8*, 1774-1785.

56. Eargle, J.; Wright, D.; Luthey-Schulten, Z., Multiple Alignment of Protein Structures and Sequences for Vmd. *Bioinformatics* **2006**, *22*, 504-506.

57. Rousseau, F.; Schymkowitz, J.; Serrano, L., Protein Aggregation and Amyloidosis: Confusion of the Kinds? *Curr Opin Struct Biol* **2006**, *16*, 118-126.

58. Fernandez-Escamilla, A.-M.; Rousseau, F.; Schymkowitz, J.; Serrano, L., Prediction of Sequence-Dependent and Mutational Effects on the Aggregation of Peptides and Proteins. *Nature Biotechnology* **2004**, *22*, 1302.

59. Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L., A Comparative Study of the Relationship between Protein Structure and Beta-Aggregation in Globular and Intrinsically Disordered Proteins. *Journal of molecular biology* **2004**, *342*, 345-353.

60. Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A., Replica Exchange Molecular Dynamics Simulations of Amyloid Peptide Aggregation. *J. Chem. Phys.* **2004**, *121*, 10748-10756.

61. Tamamis, P.; Terzaki, K.; Kassinopoulos, M.; Mastrogiannis, L.; Mossou, E.; Forsyth, V. T.; Mitchell, E. P.; Mitraki, A.; Archontis, G., Self-Assembly of an Aspartate-Rich Sequence from the Adenovirus Fiber Shaft: Insights from Molecular Dynamics Simulations and Experiments. *J. Phys. Chem. B* **2014**, *118*, 1765-1774.

62. Butler, J., Theory of the Stability of Lyophobic Colloids. *Nature* **1948**, *162*, 315-316.

63. de la Paz, M. L.; de Mori, G. M.; Serrano, L.; Colombo, G., Sequence Dependence of Amyloid Fibril Formation: Insights from Molecular Dynamics Simulations. *J. Mol. Biol.* **2005**, *349*, 583-596.

64. Desikan, R.; Patra, S. M.; Sarthak, K.; Maiti, P. K.; Ayappa, K., Comparison of Coarse-Grained (Martini) and Atomistic Molecular Dynamics Simulations of\Alpha and\Beta Toxin Nanopores in Lipid Membranes. *J. Chem. Sci.* **2017**, *129*, 1017-1030.

65. Tayeb-Fligelman, E.; Tabachnikov, O.; Moshe, A.; Goldshmidt-Tran, O.; Sawaya, M. R.; Coquelle, N.; Colletier, J.-P.; Landau, M., The Cytotoxic Staphylococcus Aureus Psmα3 Reveals a Cross-A Amyloid-Like Fibril. *Science (N.Y.)* **2017**, *355*, 831-833.

66. Zou, D.; Cao, Y.; Qin, M.; Dai, W.; Wang, W., Formation of A-Helix-Based Twisted Ribbon-Like Fibrils from Ionic-Complementary Peptides. *ChemComm* **2011**, *47*, 7413-7415.

67. Martin, L. L.; Kubeil, C.; Piantavigna, S.; Tikkoo, T.; Gray, N. P.; John, T.; Calabrese, A. N.; Liu, Y.; Hong, Y.; Hossain, M. A., Amyloid Aggregation and Membrane Activity of the Antimicrobial Peptide Uperin 3.5. *Pept. Sci.* **2018**, *e24052*.

Chapter 4

Interaction of anuran peptides with DPC and SDS micelles: A Molecular Dynamics study

4.1. Introduction

Uperin 3.x peptides, naturally obtained from *Uperoleia mjobergii*, are known to have antimicrobial activity against a host of Gram-positive bacteria[1]. Hence, these peptides are also referred to as antimicrobial peptides (AMPs). In general, AMPs exert their cytotoxic activity by acting on the bacterial cell membranes, with membrane disruption and pore formation being the most common modes of action[2]. Compared to uperin (U) 3.6 wild-type (wt), seventh-position alanine variant U3.6 K7A exhibits drastically reduced antibiotic activity, as observed against a host of Gram-positive and Gram-negative bacterial cultures[3]. In a recent study, membrane action of U3.5 wt and U3.5 R7A has been investigated[4]. It was observed that compared to U3.5 wt, membrane action was greatly reduced or almost inhibited when U3.5 R7A peptide solution was introduced to a membrane layer. Molecular dynamics (MD) simulations of melittin and MG-H2 AMPs indicate that the presence of charged residues at the N-termini region plays a crucial role in their membrane action[5]. Thus, to obtain a greater mechanistic insight into the significance of the seventh-position residue of the uperin 3.4, 3.5 and 3.6 wild-type peptides, and the corresponding alanine variants using MD simulations, the eukaryotic membrane-mimic, zwitterionic dodecyl phosphatidylcholine (DPC) and the bacterial membrane-mimic, anionic sodium dodecyl sulphate (SDS) micelles were selected for these simulations.

The peptide sequences of U3.4 wt, U3.5 wt and U3.6 wt, and their respective seventh-position variants, U3.4 R7A, U3.5 R7A and U3.6 K7A, are listed in Table 4.1. Uperin wild-type peptides from the 3.*x* family have innate amidation at the C-terminus[1], as shown in Table 4.1. All the uperin 3.*x* wild-type peptides have a net positive charge of +3, thus as alanine is a neutral residue, the respective variants have a +2 charge. Pre-equilibrated DPC and SDS micelles consisting of 65 and 60 monomers, respectively, were obtained from previously performed simulations by Tielman et al.[6] and MacKerell et al.[7], respectively.

Table 4.1. Amino acid sequences of uperin 3.*x* wild-type peptides[1] and corresponding seventh-position alanine variants.

	Peptide	1	2	3	4	5	6	**7**	8	9	10	11	12	13	14	15	16	17	C-Ter
1.	U3.4 wt	G	V	G	D	L	I	**R**	K	A	V	A	A	I	K	N	I	V	-NH₂
2.	U3.4 R7A	G	V	G	D	L	I	**A**	K	A	V	A	A	I	K	N	I	V	-NH₂
3.	U3.5 wt	G	V	G	D	L	I	**R**	K	A	V	S	V	I	K	N	I	V	-NH₂
4.	U3.5 R7A	G	V	G	D	L	I	**A**	K	A	V	S	V	I	K	N	I	V	-NH₂
5.	U3.6 wt	G	V	I	D	A	A	**K**	K	V	V	N	V	L	K	N	L	F	-NH₂
6.	U3.6 K7A	G	V	I	D	A	A	**A**	K	V	V	N	V	L	K	N	L	F	-NH₂

The overall aim of this study is to find quantitative support for the differential action of uperin 3.*x* peptides and variants on membrane-mimic surfaces that model membranes. A rational explanation for the crucial role played by the seventh-position, positively-charged amino acid residues in the antibiotic activity of these uperin 3.*x* peptides is needed. Thus, simulations were performed to elucidate the disparity of these peptides interacting with DPC and SDS micelles.

4.2. Methods

The starting structure for U3.5 wt was based on the α-helical representation determined by NMR, with the peptide present in aqueous SDS detergent[8]. It should be noted that no definite peptide structure could be obtained in a solely aqueous environment[8]. However, in the presence of membrane-mimicking SDS molecules, U3.5 wt adopted a predominantly α-helical conformation like most AMPs, probably due to its amphipathic nature[4, 9]. Using SWISS-MODEL Workspace[10], the initial structure of U3.4 wt was generated using the U3.5 wt as a template. These two peptides differ by only two amino acids of their sequences at the eleventh and twelfth positions, as indicated in Table 4.1, so it is not surprising that U3.4 wt also has an

α-helical secondary structure in Figure 4.1a. The U3.6 wt peptide has been shown to adopt a predominantly α-helical amphipathic structure by NMR in the membrane-mimicking TFE-aqueous environment[3]. Hence, an α-helical structure was also imposed on the U3.6 wt peptide sequence, as shown in Figure 4.1c, using the *Protein Builder* module[11] available in VMD[12]. In order to generate an alanine variant of the three uperin 3.*x* wild-type peptides, the *Mutator* plugin of VMD was employed to execute a point substitution at the seventh-position of the wild-type peptides. Hence, like their wild-type templates, these three variants were also primarily α-helical at the onset of the respective simulations, as shown in Figures 4.1d–4.1f).
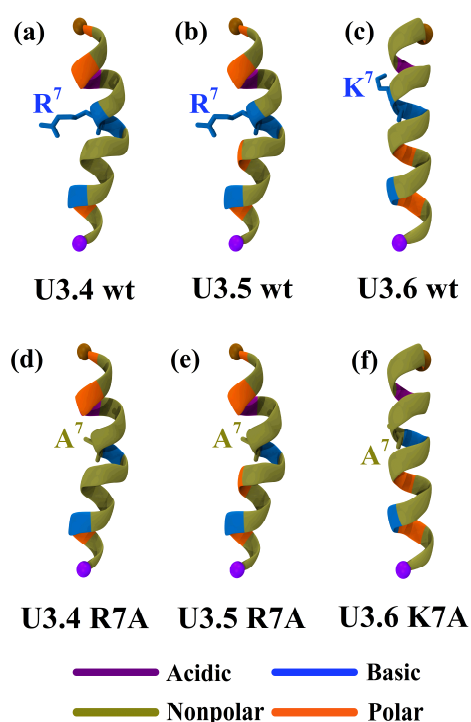
Figure 4.1. Molecular models of the uperin 3.*x* peptides in this study, represented as an α-helix. (a-c) shows the three uperin 3.*x* wild-type peptides and (d-f) their corresponding seventh-position variants. Heavy (non-hydrogen) atoms of the seventh-position residues have been shown with stick representation.

To diminish the computational time required for these peptides to stably bind to the micelles[13], the Autodock software[14] was deployed to obtain a peptide-micelle complex, and this complex is used as an initial structure for subsequent MD simulations later. The peptides were kept as ligands, and the micelles were receptor targets used for docking. The peptide backbone was kept rigid and the amino acid sidechains were allowed to be flexible. The docking calculations were based on Gasteiger-Marsili charges[15]. Autogrid was used for grid preparation, with a grid spacing of 0.375 Å. Ten peptide-micelle docking runs were performed using Lamarckian genetic algorithm, with the rates of gene mutation and crossover kept at 0.02 and 0.8,

respectively for the LUDI scoring function employed[16]. All other docking parameters were kept at their default values. Ten runs were undertaken and the peptide-micelle complex with most energetically favourable docking energy was selected for further simulations, as depicted for U3.5 wt peptide and, DPC micelle (Figure 4.2a) and SDS micelle (Figure 4.2b). The N-termini (G^1) and C-termini (V^{17}) backbone atoms in uperin 3.5 have been represented as brown and violet colored spheres, respectively in Figure 4.2 and elsewhere (as applicable).
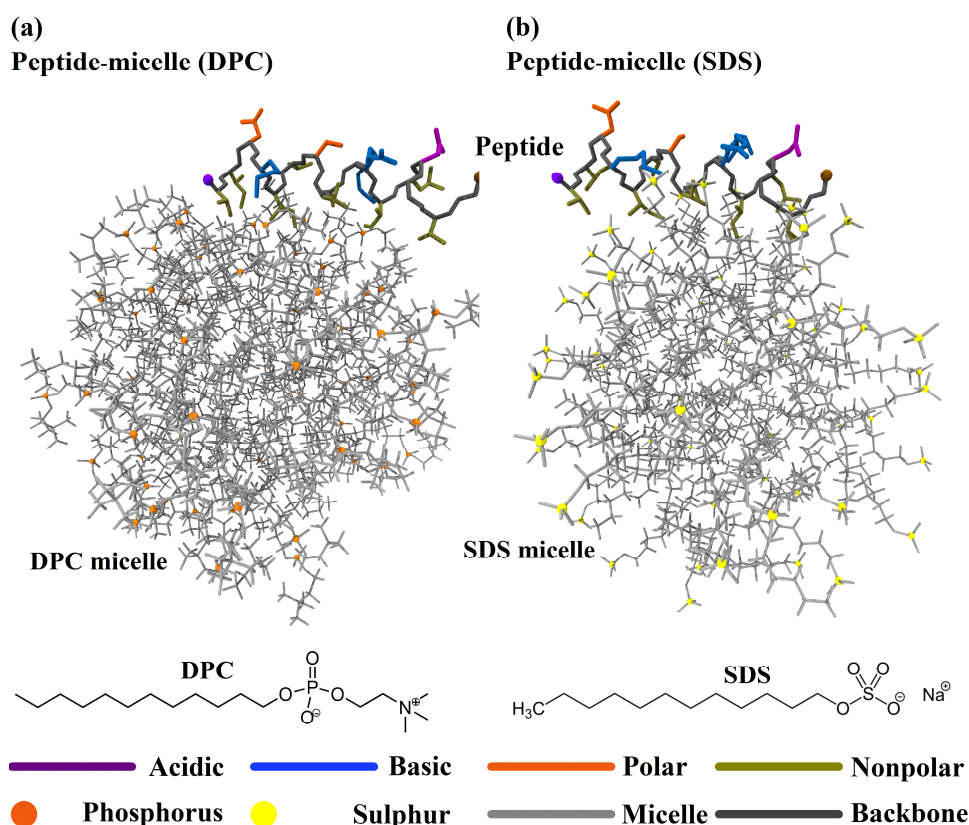
Figure 4.2. Docked complex of U3.5 wild-type peptide, and (a) DPC and (b) SDS micelles, along with the schematic of the respective lipid monomers.

The fully-atomistic (FA) MD simulations were performed using the CHARMM36m force-field for proteins[17] and lipids[18], and TIP3P water[19] in NAMD[20]. Periodic boundary conditions were applied along all three orthonormal directions. A switching function was implemented for the LJ potential, with a smooth cut-off from 10 to 12 Å. Particle-mesh Ewald summation[21] with a grid spacing of 1 Å was used for calculating the electrostatic interactions. Simulations were performed using a 2 fs timestep and a Langevin thermostat with a damping coefficient of 1 ps^{-1}. A constant pressure of 1 atmosphere (1 atm = 1.01325 bar) was maintained using a Nosé-Hoover[22] Langevin[23] piston, with a period of 100 fs and a decay time of 50 fs. The peptides were solvated with TIP3P water molecules. Then, Na$^+$ and Cl$^-$ ions were added to obtain an NaCl concentration of 0.15 M. The peptide and water molecules were initially kept under

harmonic constraints of 10 and 5 kcal mol^{-1} Å^{-2}, respectively, to restrict water diffusion into the micelle and allow the micellar lipid molecules to relax around the peptide[24]. The restraints were gradually removed with 400 ps minimization steps (steepest descent method) over a 2 ns period. Thereafter, the total simulation system was further minimized for 400 ps without any constraint. Further, equilibration was performed for 1 ns in an NVT ensemble (T =310 K), followed by final equilibration for 1 ns in an NPT ensemble (P = 1 atm and T = 310 K). The resultant DPC micelle-peptide and SDS micelle-peptide systems were then simulated in NPT ensembles (P = 1 atm and T = 310 K) for 50 ns and 40 ns, respectively.

The molecular editors PACKMOL[25] and VMD[12] were used to create the simulation systems. Tachyon ray-tracer[26], as incorporated in VMD, was employed for trajectory visualization. The antimicrobial propensity indices for the different peptides were calculated using the AMPA web server[27]. Helical wheel schematics for the different wild-type and variant peptides were constructed using the HELIQUEST server[28]. Secondary structure composition was determined by STRIDE program[29] using DSSP definitions[30], as implemented by the Timeline plugin[31] in VMD. The remaining analyses of data were performed with VMD and in-house scripts. The interaction energies, radial distribution function (RDF) plots and solvent-accessible surface area (SASA) values were derived from last 20 ns of the respective simulation trajectories. Trajectories were sampled every 10 ps.

4.3. Results and discussion

4.3.1. Peptide-micelle interaction

4.3.1.1. Peptide-micelle separation

As noted from the centre of mass distance plots of uperin peptides and the DPC (a–c) and SDS (d–f) micelles in Figure 4.3, a near steady-state is attained for most of the systems within the first 5 ns of the simulation. Interestingly, for U3.4 wt and U3.4 R7A (Figure 4.3a), the interaction with DPC micelle requires almost 18 ns to stabilize and the centre of mass distance reaches a value of ≈16 Å. For U3.5 wt and U3.5 R7A simulations (Figure 4.3b), a centre of mass separation of approximately 15 Å was observed. The U3.6 wt and U3.6 K7A peptides (Figure 4.3c) also reached a centre of mass distance of ≈15 Å, with respect to the DPC micelle. For the SDS micelle, the centre of mass distance from the U3.4 wt and U3.4 R7A peptides was at nearly 13 Å (Figure 4.3d), as was the case for U3.5 wt and U3.5 R7A (Figure 4.3e). The centre of mass distance for U3.6 wt and U3.6 K7A peptides (Figure 4.3f) was observed at ≈12

Å. Thus, to conclude from the centre of mass distance values for all peptides studied here, the peptides remain bound to the DPC and SDS micelles after the peptide-micelle interaction reaches a near steady-state. Moreover, only minor differences are found in the centre of mass distance plots for the uperin 3.x wild-type and variant peptides with the micelles. Interestingly, for the same set of peptides, there is a difference of 2–3 Å in the centre of mass separation values relative to DPC and SDS micelles. Proximity of the peptides to the centre of the SDS micelles, compared to that of the DPC micelle, could be attributed to the smaller number of lipids in the SDS micelle, along with the anionic nature of these lipids. Hence, further quantitative analyses would be required to glean further information regarding the differential interaction of these peptides with DPC and SDS micelles, if present.
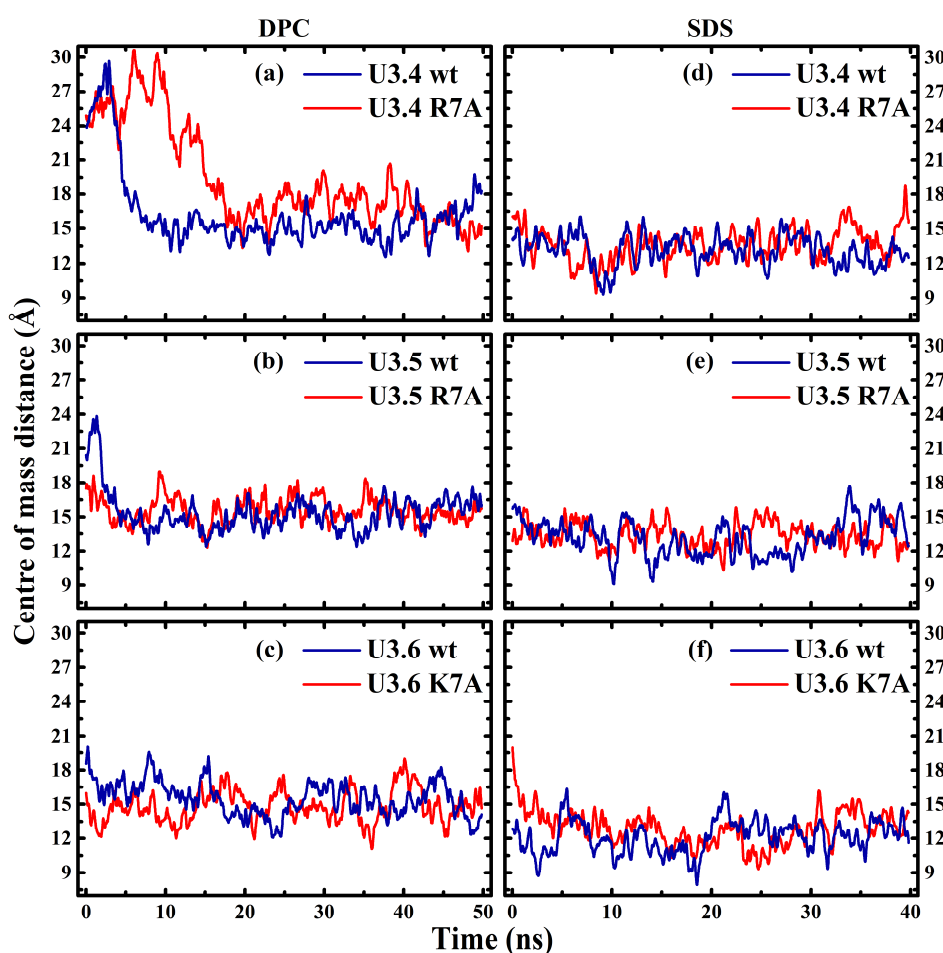
Figure 4.3. Centre of mass distance plots of wild-type and variant uperin 3.x peptides with respect to (a–c) DPC and (d–f) SDS micelles.

The interaction of U3.5 wt with both DPC and SDS micelles stabilized within the first 5 ns, as noted in Figures 4.3b and 4.3e, respectively. However, since both micelles have head groups of different atomic and electrostatic charge composition, analysis of the type of peptide

residues that preferentially interact with these micelles was undertaken. Figure 4.4 shows that the initial interaction has the nonpolar residues positioned towards the micelle, post docking with the Autodock program. Moreover, even at 25 and 50 ns, only the nonpolar residues preferentially interact with the DPC micelle, with the remaining acidic, basic and polar residues preferring greater contact with the surrounding water molecules and ions (omitted for clarity). Further, there is negligible difference between the peptide-micelle interactions at 25 and 50 ns trajectory snapshots, in agreement with the centre of mass distance plot in Figure 4.3b.
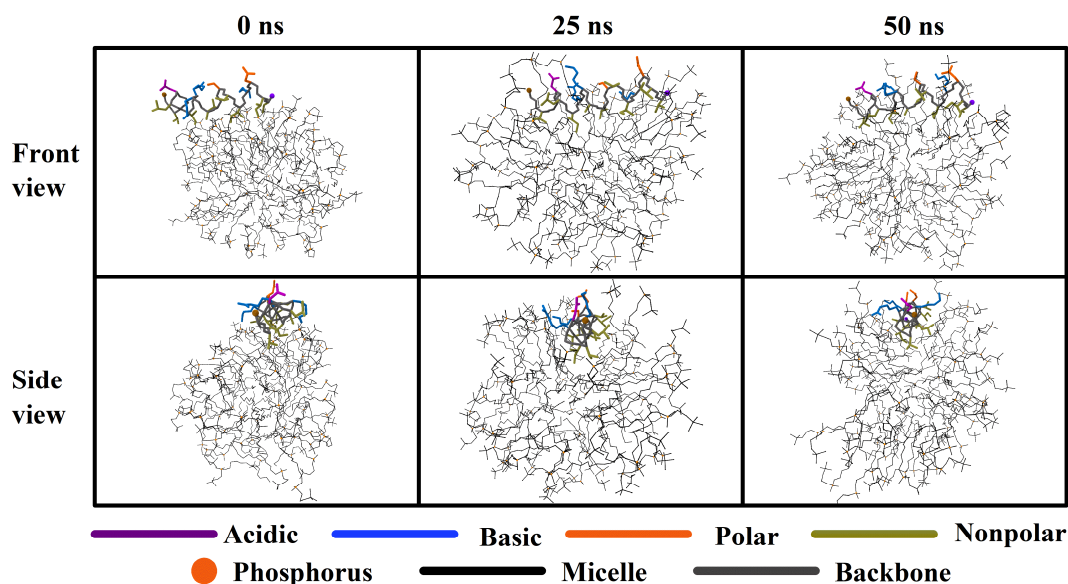
Figure 4.4. Trajectory snapshots of U3.5 wt and DPC micelle complex taken at 0, 25 and 50 ns instances of the simulation.

The interaction between U3.5 wt and SDS micelle in Figure 4.5 shows a similar pattern to that observed for U3.5 wt-DPC micelle interaction in Figure 4.4. The peptide-micelle interactions observed in 20 and 40 ns snapshots are indistinguishable, indicating that the equilibration of this interaction is achieved by 20 ns, as also noted in the centre of mass distance plot in Figure 4.3e. Remarkably, in the trajectory snapshots taken at the middle and end of both these simulations in Figures 4.4 and 4.5, formation of a groove in the micellar structure could be observed, especially in the front view images. This might be attributed to the penetration of the nonpolar amino acid sidechains towards the hydrophobic core of these micelles, away from the water molecules in vicinity. However, further quantitative analysis would be required to discern the disparity in the interaction between uperin 3.*x* peptides and variants on one hand, and DPC and SDS micelles on the other, as opined earlier during the discussion of Figure 4.3.

	0 ns	20 ns	40 ns

Front view

Side view

Acidic — Basic — Polar — Nonpolar
● Sulphur — Micelle — Backbone
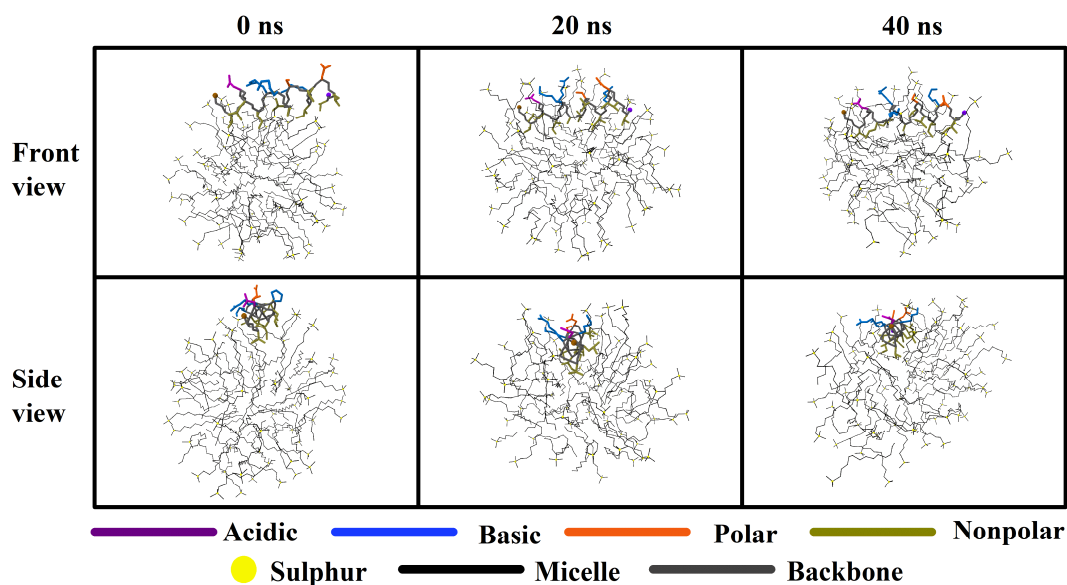
Figure 4.5. Trajectory snapshots of U3.5 wt and SDS micelle complex taken at 0, 25 and 50 ns instances of the simulation.

4.3.1.2. Energetics of interaction

Table 4.2. Autodock docking energies between uperin 3.*x* wild-type peptides and seventh-position variants, and DPC and SDS micelles.

Peptide		Docking energy (kcal mol^{-1})	
		DPC	SDS
1.	U3.4 wt	11.41	-2.51
2.	U3.4 R7A	10.26	-2.08
3.	U3.5 wt	8.83	0.35
4.	U3.5 R7A	9.20	-2.31
5.	U3.6 wt	8.10	-5.43
6.	U3.6 K7A	7.57	-3.14

The Autodock docking energies were obtained after docking the uperin 3.*x* wild-type and seventh-position variant peptides onto the DPC and SDS micelles using the Autodock software, as discussed in the Methods section. No definitive pattern could be obtained from the docking energies of uperin 3.*x* peptides and their corresponding seventh-position variant with either DPC or SDS micelle. However, all the peptides demonstrated a more favourable docking with the SDS micelle; compared to the DPC micelle.
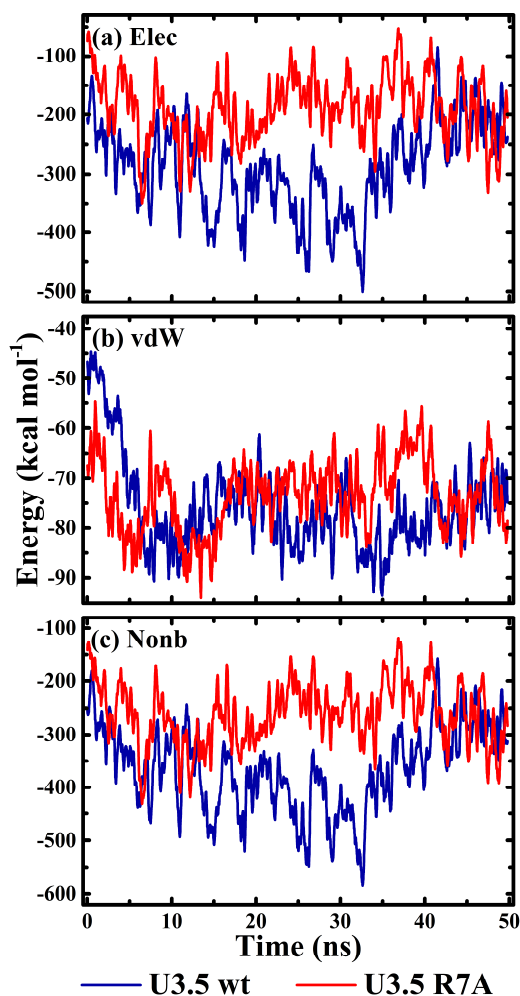
Figure 4.6. Interaction energies: (a) electrostatic, (b) vdW and (c) nonbonded, between U3.5 wt and U3.5 R7A, and DPC micelles.

To obtain a deeper insight into these peptide-micelle interactions, the interaction energies were calculated between uperin 3.x wild-type and seventh-position variant peptides, and DPC micelles. Over the 50 ns simulation period, it could be observed that in general, the U3.5 wt has a more favourable electrostatic interaction with DPC micelle, compared to U3.5 R7A (Figure 4.6a). Electrostatic energies of nearly -300 and -180 kcal mol^{-1} were observed for the U3.5 wt and the U3.5 R7A, respectively. No discernible difference could be noted in the van der Waals (vdW) interaction energies between these two peptides and the DPC micelle, with an approximate value of -75 kcal mol^{-1} observed (Figure 4.6b). The trend in nonbonded interaction energies is similar to the one noted for electrostatic energies (Figure 4.6c). The nonbonded energies can be defined as the sum of electrostatic and vdW interaction energies between two distinct entities. U3.5 wt has an appreciably higher total nonbonded interaction with the DPC micelle, relative to U3.5 R7A, with energy values around -370 and -250 kcal

mol^{-1}, respectively. It could be noted that due to the dynamic nature of these MD simulations, high fluctuation in energy values are observed throughout these simulations in Figure 4.6.

Figure 4.7. Interaction energies: (a) electrostatic, (b) vdW and (c) nonbonded of uperin 3.*x* wild-type peptides and seventh-position variants with DPC micelles.

A convenient comparison between the different interaction energy values for the uperin 3.*x* wild-type and variant peptides with DPC micelle was obtained with a statistical average of the energy values, calculated over the last 20 ns of the respective simulations in Figure 4.7. Then, the error bars were added using the standard deviation calculated over the 20 ns dataset. Electrostatic energies of these peptide-micelle interactions were -303, -265 and -253 kcal mol^{-1} for U3.4 wt, U3.5 wt and U3.6 wt simulations, respectively (Figure 4.7a). For U3.4 R7A, U3.5 R7A and U3.6 K7A, these electrostatic energies were -158, -180 and -227 kcal mol^{-1}, respectively. The vdW energies for these interactions were -75, -78 and -81 kcal mol^{-1} for U3.4

wt, U3.5 wt and U3.6 wt peptides, respectively (Figure 4.7b), and -66, -72 and -79 kcal mol^{-1} for U3.4 R7A, U3.5 R7A and U3.6 K7A, respectively. The nonbonded energies for peptide-micelle interactions were observed at -378, -343 and -334 kcal mol^{-1}, for U3.4 wt, U3.5 wt and U3.6 wt systems, respectively (Figure 4.7c). The U3.4 R7A, U3.5 R7A and U3.6 K7A had nonbonded energies of -224, -252 and -306 kcal mol^{-1}, respectively.

In summary, the interactions of wild-type peptides with DPC micelle were significantly stronger, compared to their respective seventh-position variants. The electrostatic interaction energy of U3.4 wt was higher by ≈44 kcal mol^{-1} than the other two wild-type peptides, with U3.6 wt showing the least electrostatic interaction. However, the trend is reversed for the uperin 3.*x* variants, with U3.6 K7A and U3.4 R7A demonstrating the highest and the least electrostatic interaction, respectively. These trends were also observed for vdW and nonbonded interactions of uperin 3.*x* wild-type peptides and their corresponding variants with DPC micelle. Among the uperin 3.*x* peptides, the U3.4 wt interacts the strongest with the DPC micelle and for the uperin 3.*x* variants, the U3.6 K7A had the strongest interaction. It should be noted that most dramatic effect of seventh-position substitution in electrostatic and nonbonded energies of interaction could be observed for U3.4 wt and U3.4 R7A, and the least for U3.6 wt and U3.6 K7A. This highlights the important role played by the peptide sequence in intermolecular interaction; even amongst peptides from the same family.

Similar to the data discussed in Figure 4.6 for simulations with DPC micelle, interaction energies were calculated between SDS micelle, and uperin 3.*x* wild-type peptides and variants. During the 40 ns simulation period, it was noted that the U3.5 wt has a significantly greater electrostatic interaction with the SDS micelle in Figure 4.8a, compared to U3.5 R7A. Electrostatic interaction energies of nearly -650 and -460 kcal mol^{-1} were observed for U3.5 wt and U3.5 R7A, respectively. However, the difference in the vdW interaction energies between U3.5 wt and U3.5 R7A, and SDS micelle in Figure 4.8b is smaller in magnitude, with approximate values of -87 and -66 kcal mol^{-1}, respectively. The trend in nonbonded interaction energies is similar to the one observed for electrostatic interaction energies, with U3.5 wt showing greater nonbonded interaction with SDS micelle relative to U3.5 R7A (Figure 4.8c), with energy values around -750 and -540 kcal mol^{-1}, respectively.
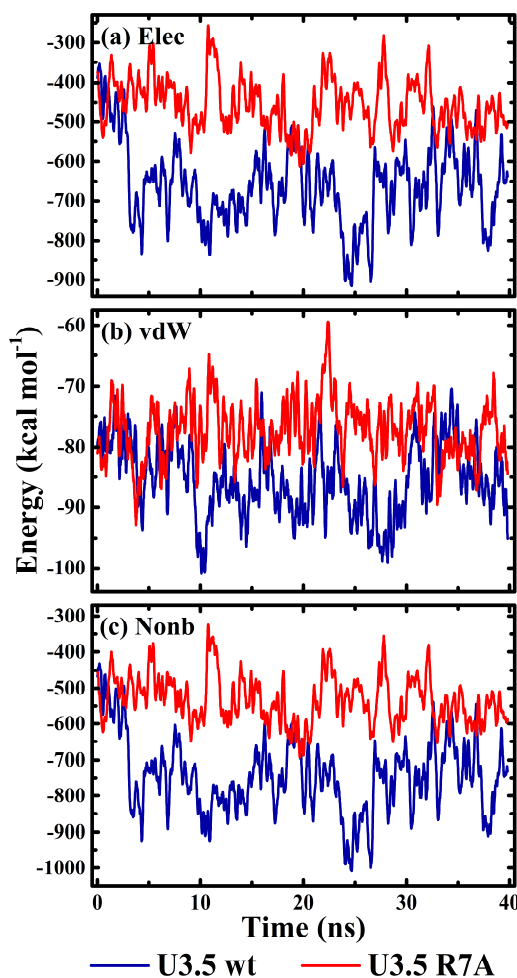
Figure 4.8. Interaction energies: (a) electrostatic, (b) vdW and (c) nonbonded, between U3.5 wt and U3.5 R7A, and SDS micelles.

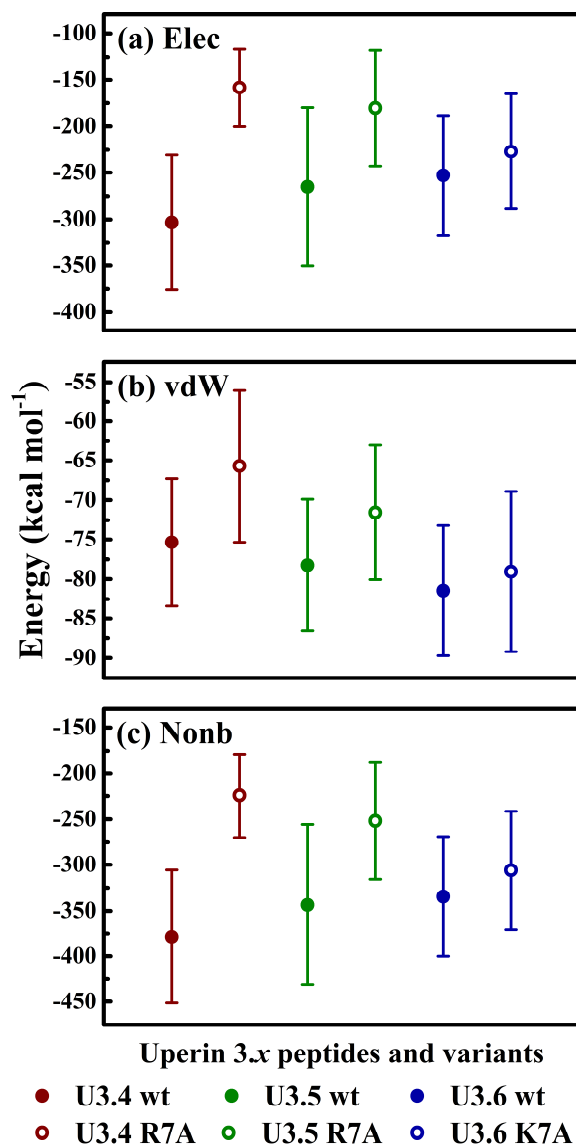Comparison between the interaction energies of uperin 3.*x* peptides and variants with SDS micelle was done in Figure 4.9, as described previously for the peptide-DPC micelle interactions. Electrostatic energies of these peptide-micelle interactions were noted at -588, -680 and -605 kcal mol^{-1} for U3.4 wt, U3.5 wt and U3.6 wt systems, respectively (Figure 4.9a). For U3.4 R7A, U3.5 R7A and U3.6 K7A, electrostatic interaction energies had a value of -458, -464 and -476 kcal mol^{-1}, respectively. vdW energies of interactions were observed at -86, -86 and -90 kcal mol^{-1} (Figure 4.9b) for U3.4 wt, U3.5 wt and U3.6 wt peptides, respectively. For U3.4 R7A, U3.5 R7A and U3.6 K7A, the vdW energies had a value of -80, -77 and -85 kcal mol^{-1}, respectively. Nonbonded energies of these peptide-micelle interactions were noted at -674, -766 and -695 kcal mol^{-1} for U3.4 wt, U3.5 wt and U3.6 wt systems, respectively (Figure 4.9c). For U3.4 R7A, U3.5 R7A and U3.6 K7A, the nonbonded energies had values of -538, -541 and -561 kcal mol^{-1}, respectively.

116
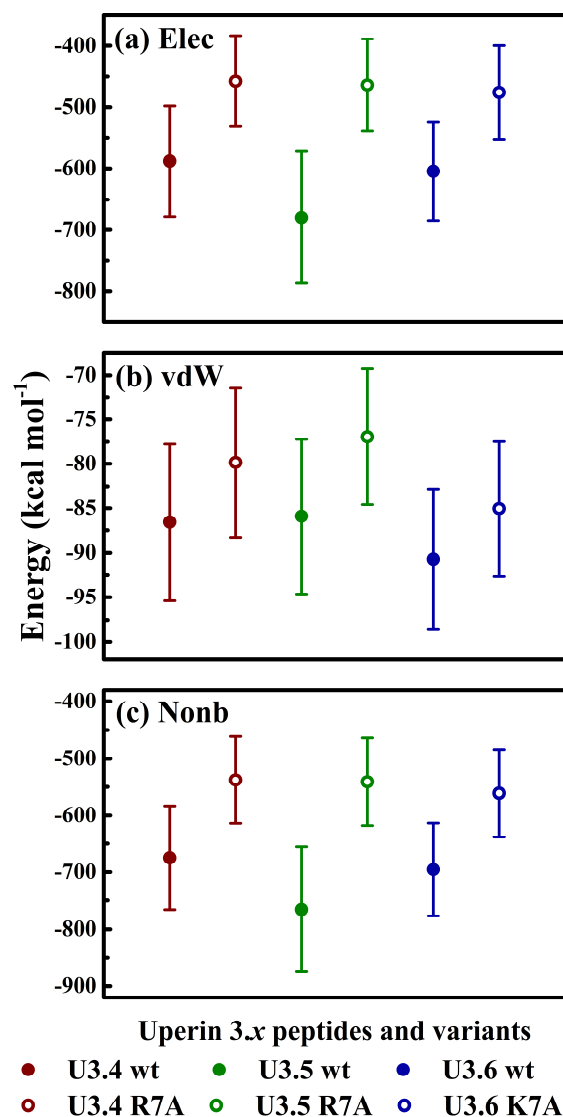
Figure 4.9. Interaction energies: (a) electrostatic, (b) vdW and (c) nonbonded of uperin 3.*x* wild-type peptides and seventh-position variants with SDS micelles.

In contrast to previously described interactions with DPC micelle, the electrostatic interaction energy is highest for U3.5 wt, distantly followed by U3.6 wt and U3.4 wt in that order i.e. U3.5 wt>U3.6 wt>U3.4 wt. Moreover, electrostatic interactions between uperin 3.*x* variants and SDS micelle were weaker, compared to the corresponding wild-type peptides. Among the variants, strongest electrostatic interactions were observed for U3.6 K7A>U3.5 R7A>U3.4 R7A. Similar vdW interaction energies were found for U3.4 wt and U3.5 wt simulations, with the U3.6 wt one being slightly higher than the other two. Among uperin 3.*x* variants, U3.6 K7A had the greatest vdW interaction with SDS micelle, and U3.5 R7A the least. For nonbonded interactions, similar trends to those noted for electrostatics energies were observed. Among the wild-type peptides and variants, strongest nonbonded interaction with SDS micelle was noted for U3.5 wt and U3.6 K7A, respectively. Therefore, all these six peptides had a significantly

stronger interaction with SDS micelle, when compared to DPC micelle. The trend for the interaction energies of wild-type peptides with SDS micelle could be contrasted with the one observed for DPC micelle, although the trend in different energy values for uperin 3.x variants are similar for both DPC and SDS micelles. The presence of two positively-charged amino acid residues next to each other (at seventh and eight positions) in uperin 3.x wild-type peptides might result in differential interaction with zwitterionic and anionic head groups of DPC and SDS micelles, respectively.

4.3.2. Computational analysis of peptide sequences

The crucial role of the seventh-position cationic residue towards the antimicrobial activity of the peptide was analysed using a predictive algorithm employed by the AMPA computational web server[27]. The AMPA server uses an algorithm to calculate an antimicrobial propensity scale for the given peptide sequence, based on the high-throughput screening results of the AMP bactenecin 2A. Half-maximal inhibitory concentration (IC$_{50}$) values indicate the concentration of a particular pharmacological agent required to inhibit a specific biological activity by half[32]. Therefore, it can be used to determine the cytotoxic activity of a pharmacological agent against bacterial cell cultures. The antimicrobial IC$_{50}$ values for all amino acid substitutions of bactenecin 2A have been calculated. Using these data, an antimicrobial index (AI) can be computed to provide an estimate about the antimicrobial propensity of a particular peptide sequence. Lower AI values, as shown in Figure 4.10, correspond to greater antimicrobial potency of that region on the peptide. As illustrated in Figures 4.10a–4.10c, the substitution of seventh-position cationic residue with alanine significantly reduces the antimicrobial activity of the corresponding peptides, i.e. a shift to higher AI values. Moreover, it is clear that one amino acid change at the seventh position can adversely affect the antimicrobial activity of the entire 4–10 amino acid stretch in the alanine variant. Although the AI values are predictive in nature, these plots might help explain the difference in the peptide-micelle interaction energies between wild-type and alanine variants of the respective uperin 3.x peptides.
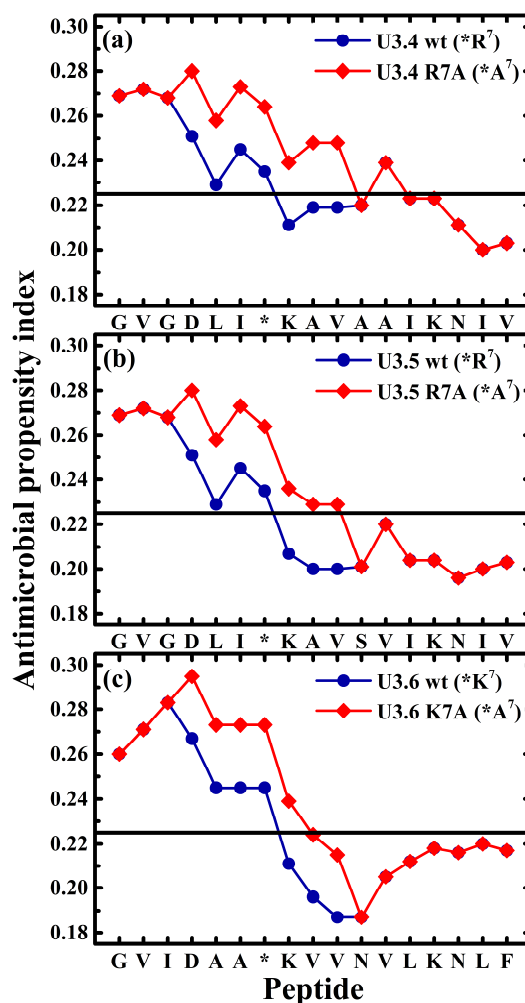
Figure 4.10. Antimicrobial propensity profiles of (a) U3.4 wt and U3.4 R7A, (b) U3.5 wt and U3.5 R7A, and (c) U3.6 wt and U3.6 K7A peptides created using AMPA[27].

The alanine substitution at seventh-position adversely impacts the amphipathic nature of the uperin 3.*x* AMPs, as could be observed in Figure 4.11. Helical AMPs can be roughly partitioned into two regions: one that preferentially interacts with the hydrophilic surfactant headgroups and the surrounding solvent environment, and the other region that favours interaction with the hydrophobic surfactant tails. Therefore, one region is predominantly hydrophilic, composed of acidic, basic or polar amino acid residues, and the other region contains hydrophobic or nonpolar amino acids. Hydrophobic moment provides a measure of the overall amphipathicity of these helical AMPs, and is defined as the mean vector sum of the hydrophobicities of the different amino acid sidechains on the helix[33]. Amongst all three pairs of wild-type peptides and their corresponding seventh-position alanine variants, uperin 3.5 pair has the greatest hydrophobic moment, followed by uperin 3.6, and finally uperin 3.4 pair. Furthermore, all wild-type peptides have a greater hydrophobic moment than their corresponding alanine variants. Since an alanine substitution introduces a nonpolar amino acid residue into the

119

hydrophilic region, as seen in Figures 4.11b, 4.11d and 4.11f, this is expected to reduce the extent of the hydrophilic region for these peptides.
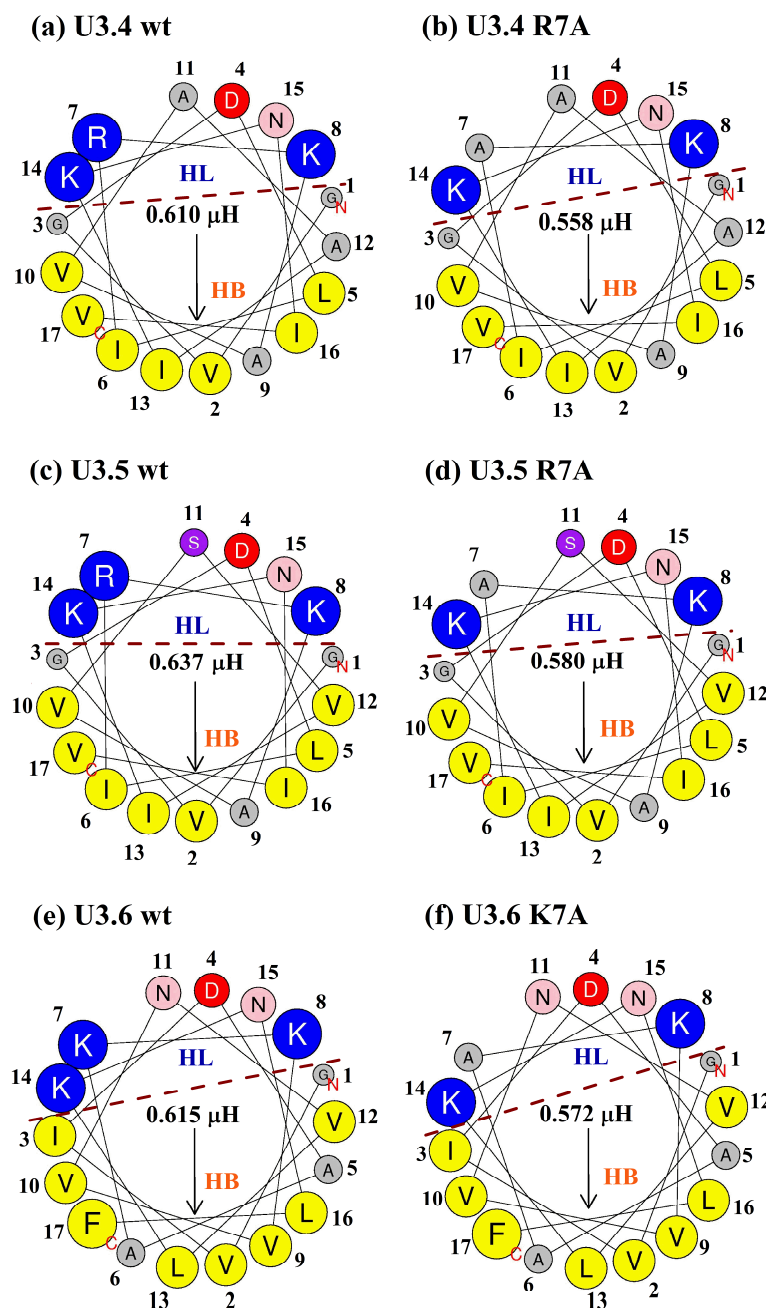
Figure 4.11. Helical wheel representation of wild-type and variant uperin 3.*x* peptides[28]. Individual amino acid residues have been numbered as per their location in the peptide sequence (from N- to C-terminus). The size and colour of these residues correspond to the relative size and, electrostatic and steric nature of their sidechains, respectively. The peptide residues have been grouped as: acidic (red), basic (blue), nonpolar (grey, yellow) and polar (light pink, purple). A dashed line roughly partitions the helical wheel of these peptides into predominantly hydrophilic (HL) and hydrophobic (HB) regions. The downward arrows and the numeric values indicate the direction and magnitude of the hydrophobic moment vector, respectively.

The α-helical secondary structure of U3.5 wt in Figure 4.11c is neatly partitioned into two distinct hydrophilic and hydrophobic regions. However, in comparison, the U3.4 wt has an alanine residue at the eleventh-position in Figure 4.11a, decreasing the polarity across the hydrophilic region and resulting in a lower hydrophobic moment, compared to U3.5 wt. The reduction in hydrophobic moment for the U3.6 wt, in Figure 4.11e, might be attributed to the relatively weaker atomic charge distribution of the lysine residue, compared to the arginine residue[34]. These strong partial charges could also explain the significant reduction in interaction energies of seventh-position alanine variants of U3.4 wt and U3.5 wt with the DPC and SDS micelles, as observed in Figures 4.6–4.9, compared to the corresponding wild-type peptides; where an arginine residue is present at the seventh-position. Whereas, the U3.6 wt peptide contains a lysine residue at the seventh-position, which is expected to cause a relatively smaller reduction in interaction energies, following substitution with alanine at the seventh-position of the U3.6 K7A variant.

4.3.3. Seventh residue-micelle interaction

4.3.3.1. Energetics of interaction

To elucidate the crucial role played by the seventh-position cationic residues in these peptide-micelle interactions, the interaction energies of these particular residues with both DPC and SDS micelles were calculated and analysed. Electrostatic energies of these seventh residue-DPC micelle interactions were noted at -106, -96 and -54 kcal mol^{-1} for arginine (R^7)-U3.4 wt, R^7-U3.5 wt and lysine (K^7) - U3.6 wt residues, respectively, in Figure 4.12a. For alanine (A^7)-U3.4 R7A, A^7-U3.5 R7A and A^7-U3.6 K7A residues, the electrostatic energies had a value of -0.5, 0.7 and 1.2 kcal mol^{-1}, respectively. The vdW energies for R^7-U3.4 wt, R^7-U3.5 wt and K^7-U3.6 wt residues were noted at -6, -5 and -4 kcal mol^{-1}, respectively, in Figure 4.12b. For A^7-U3.4 R7A, A^7-U3.5 R7A and A^7-U3.6 K7A, the vdW energies had a value of -1.6, -1.5 and -2.3 kcal mol^{-1}, respectively. Nonbonded energies of these seventh residue-micelle interactions were noted at -112, -101 and -58 kcal mol^{-1} for R^7-U3.4 wt, R^7-U3.5 wt and K^7-U3.6 wt, respectively, in Figure 4.12c; and for A^7-U3.4 R7A, A^7-U3.5 R7A and A^7-U3.6 K7A, the nonbonded interaction energies were -2.1, -0.8 and -1.1 kcal mol^{-1}, respectively. As observed earlier for the wild-type peptides in Figure 4.7, the electrostatic and nonbonded interaction energies are highest for R^7-U3.4 wt and least for K^7-U3.6 wt. The corresponding interaction energies between alanine residues at the seventh-position of variants, and DPC micelle, is comparatively negligible. It should be noted that the magnitude of error bars is very close to

the mean values of vdW energies of all six simulation systems in Figure 4.12, along with the electrostatic and nonbonded interaction energies of the alanine residues with DPC micelles. Hence, these values might lack statistical accuracy.
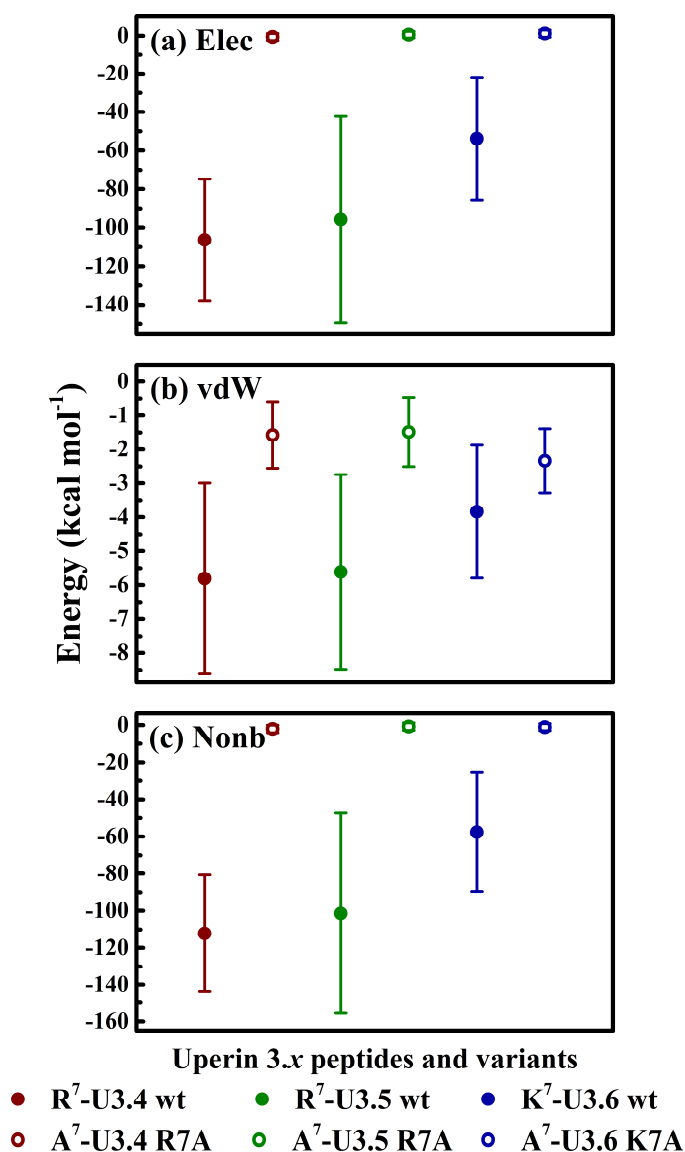
Figure 4.12. Interaction energies: (a) electrostatic, (b) vdW and (c) nonbonded of seventh-position arginine: R^7 or lysine: K^7 (wild-type peptides), and alanine: A^7 (variants) residues with DPC micelles.

Likewise, interaction energies between seventh-position residues and SDS micelles were also calculated. Electrostatic energies of these seventh residue-micelle interactions were noted at -161, -192 and -173 kcal mol^{-1} for R^7-U3.4 wt, R^7-U3.5 wt and K^7-U3.6 wt, respectively. in Figure 4.13a. For A^7-U3.4 R7A, A^7-U3.5 R7A and A^7-U3.6 K7A, the electrostatic energies had a value of 0, 0.9 and -1.2 kcal mol^{-1}, respectively. vdW energies of -6, -6 and -5 kcal mol^{-1} were observed in Figure 4.13b for R^7-U3.4 wt, R^7-U3.5 wt and K^7-U3.6 wt, respectively. For

A^7-U3.4 R7A, A^7-U3.5 R7A and A^7-U3.6 K7A, the vdW energies had a value of -2.2, -2.5 and -2.6 kcal mol^{-1}, respectively. Nonbonded energies of these seventh residue-micelle interactions were noted at -167, -198 and -178 kcal mol^{-1} for R^7-U3.4 wt, R^7-U3.5 wt and K^7-U3.6 wt respectively in Figure 4.13c. For A^7-U3.4 R7A, A^7-U3.5 R7A and A^7-U3.6 K7A, the nonbonded energies had a value of -2.2, -1.6 and -3.8 kcal mol^{-1}, respectively.
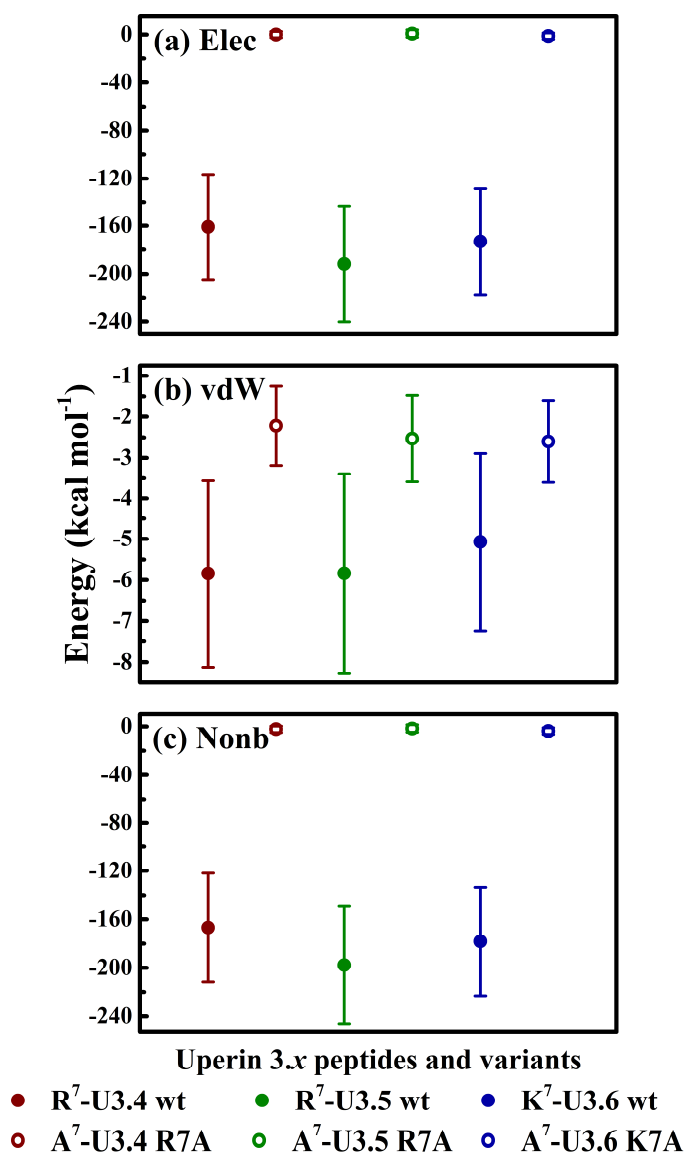
Figure 4.13. Interaction energies: (a) electrostatic, (b) vdW and (c) nonbonded of seventh-position arginine: R^7 or lysine: K^7 (wild-type peptides), and alanine: A^7 (variants) residues with SDS micelles.

As discussed earlier for the wild-type peptides in Figure 4.9, the electrostatic and nonbonded interaction energies are maximum for R^7-U3.5 wt and least for R^7-U3.4 wt. All vdW energies, and electrostatic and nonbonded interaction energies of the alanine residues with SDS micelles are comparatively negligible, also noted earlier in Figure 4.12. The electrostatic and nonbonded

interaction of positively-charged R^7 and K^7 residues of uperin 3.*x* wild-type peptides with anionic SDS micelle is more energetically favourable, than that observed with zwitterionic DPC micelle. This could explain the similar trend observed for the interaction of wild-type peptides with DPC and SDS micelles, as described earlier for Figures 4.7 and 4.9. Therefore, for both DPC and SDS micelles, the seventh-position cationic residue plays a significant role in their interaction with uperin 3.*x* peptides, which gets adversely affected by its substitution with a nonpolar alanine residue in the variants.

4.3.3.2 Radial distribution plots

Simulation trajectory snapshots shown in Figure 4.14 demonstrate the differential interaction of the positively-charged (arginine) and nonpolar (alanine) sidechain of the seventh-position amino acid residue of uperin 3.5 peptides, interacting with either DPC or SDS micelles. These snapshots were taken at 50 ns for the DPC systems and 40 ns for the SDS systems i.e. at the end of the respective simulations. Only the atoms of interest are represented as solid spheres in the figure. Whereas hollow spheres in Figures 4.14a and 4.14c depict the neighbouring aliphatic carbon atoms of micelle tails, those in Figures 4.14b and 4.14d represent the neighbouring phosphate and sulphate head group heavy (non-hydrogen) atoms. The surrounding water molecules and ions have been excluded for clarity. Interestingly, both peptides are in a predominantly α-helical conformation. Figure 4.14a shows the R^7-U3.5 wt preferentially interacting with the negatively-charged phosphate head group of DPC, whereas A^7-U3.5 R7A shows closest interaction with the aliphatic carbon atoms of the DPC tail (Figure 4.14b). Similarly, the negatively-charged sulphate head group of SDS aligns near to the R^7-U3.5 wt in Figure 4.14c, and the aliphatic carbon atoms of the SDS tail are closer to A^7-U3.5 R7A, as seen in Figure 4.14d.
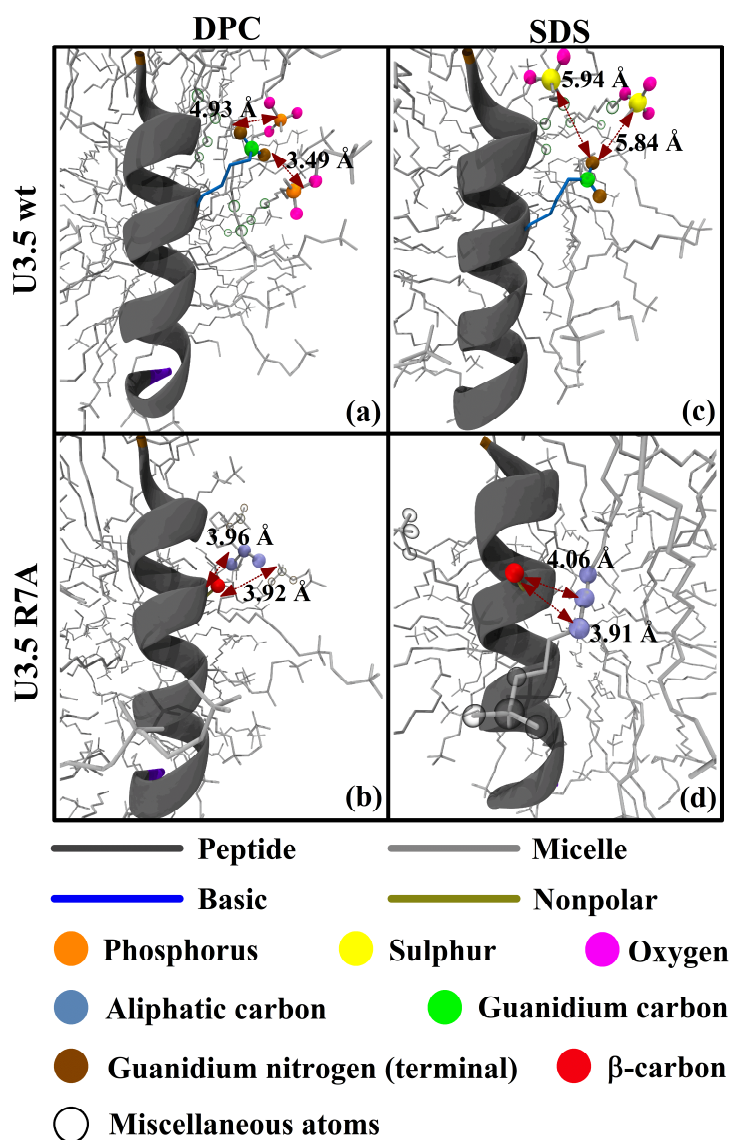
Figure 4.14. Trajectory snapshots of U3.5 wt and U3.5 R7A near (a–b) DPC and (c–d) SDS micelles. Images taken at the end of the respective simulations.

In order to analyse the seventh residue-micelle interactions, observed between the seventh-position residue sidechain and a specific chemical moiety from the micelle in Figure 4.14 in greater detail, radial distribution functions (RDFs) were calculated between these two groups of heavy atoms. Only the sidechain heavy atoms were incorporated in these analyses, along with the respective micellar heavy atom selections used as a reference. However, the heavy atom selection was restricted to positively-charged terminal region of guanidium ($-C-(NH_2)_2^+$), and ε-ammonium ($-NH_3^+$) region of arginine and lysine residues, respectively[35]. The specific RDF calculations were then normalized with respect to the number of atoms in the amino acid sidechain selection.
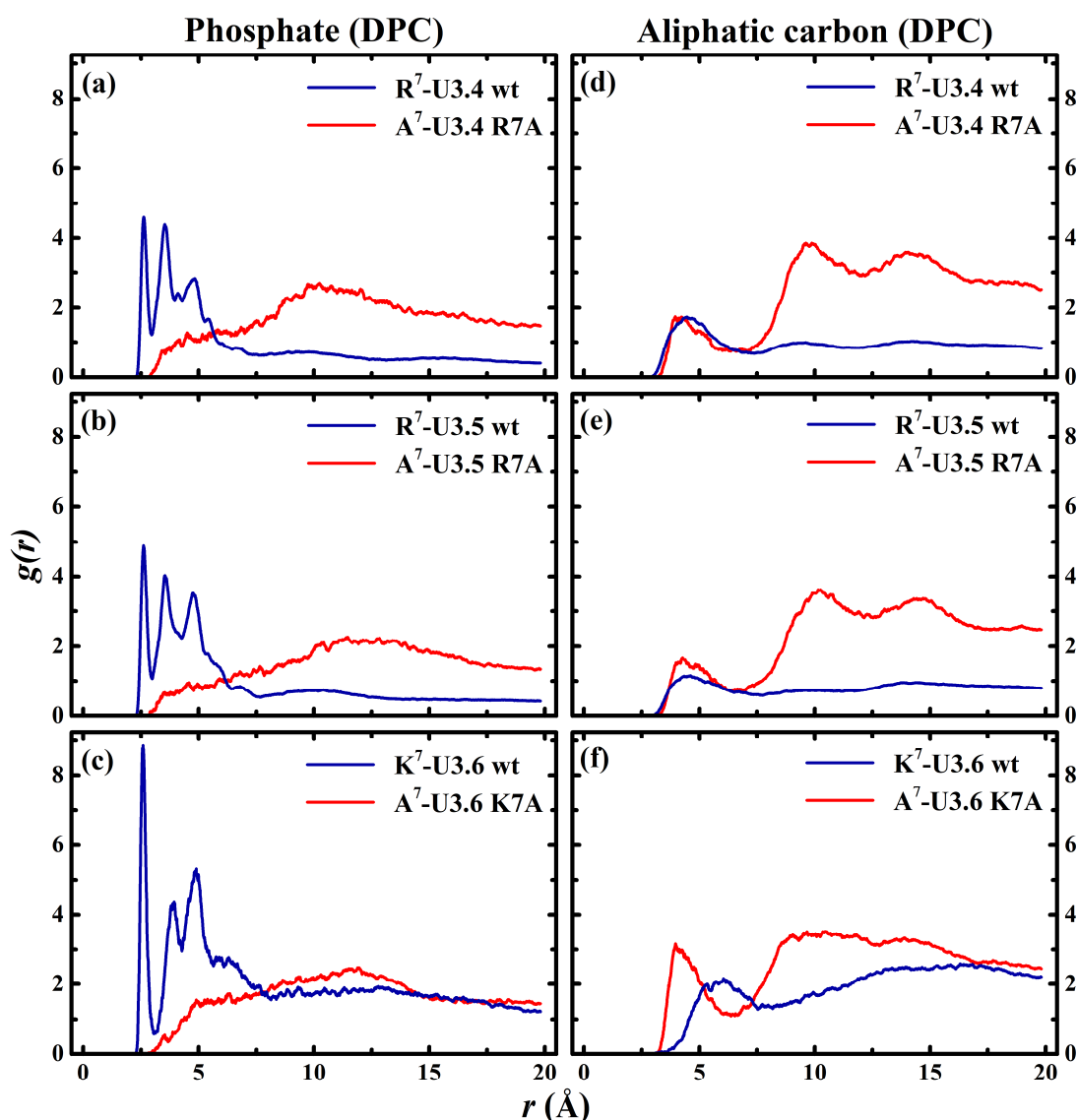
Figure 4.15. Radial distribution function plots for the sidechain heavy atoms of the seventh-position residue of the uperin 3.*x* peptides, with the phosphate head group (a–c) and the aliphatic carbon tail (d–f) heavy atoms of the DPC micelle.

For R[7]-U3.4 wt, with reference to the DPC phosphate head groups, two distinct peaks were observed at 2.6 Å, 3.5 Å and 4.8 Å in Figure 4.15a, with *g(r)* values of 4.6, 4.4 and 2.8, respectively. Whereas for R[7]-U3.5 wt, three peaks could be noted at 2.6 Å, 3.5 Å and 4.7 Å, in Figure 4.15b, with *g(r)* values of 4.9, 4.0 and 3.5, respectively. As observed in Table 4.1, U3.4 and U3.5 differ at only two residues of their respective wild-type and variant peptide sequences. For K[7]-U3.6 wt, a sharp peak was noted at 2.6 Å and *g(r)* of 8.8, in Figure 4.15c. Further, two peaks were observed at 3.9 Å and 4.9 Å, having *g(r)* values of 4.7 and 5.2, respectively. Interestingly, in Figures 4.15a–4.15c, for RDF plots of A[7]-U3.4 R7A, A[7]-U3.5 R7A and A[7]-U3.6 K7A, only one flat peak could be located at 10.7 Å, 13.4 Å and 11.9 Å, with *g(r)* values of 2.6, 2.2 and 2.5, respectively. Similarly, for RDF plots of R[7]-U3.4 wt, R[7]-U3.5 wt and K[7]-

126

U3.6 wt with reference to the DPC aliphatic carbon tails, a single flat peak could be noted at 4.4 Å, 4.6 Å and 6.0 Å and $g(r)$ values of 1.7, 1.2 and 2.2 respectively (Figures 4.15d–4.15d f). The data obtained for A^7-U3.4 R7A, in Figure 4.15d, show three flat peaks at 4.2 Å, 9.6 Å and 14.0 Å, having $g(r)$ values of 1.7, 3.9 and 3.6, respectively. Likewise, the RDF plot of A^7-U3.5 R7A is characterised by three flat peaks at 4.3 Å, 10.1 Å and 14.7 Å, with $g(r)$ values of 1.7, 3.6 and 3.4, respectively, in Figure 4.15e. However, for A^7-U3.6 K7A plot in Figure 4.15 f, only one distinct peak was found at 4.0 Å and a $g(r)$ of 3.1. Thus, as anticipated, arginine and lysine sidechains interacted most favourably with the phosphate head groups of DPC micelles. Similarly, for alanine sidechains, the favourable interaction is with the aliphatic carbon tails of the DPC micelles. For wild-type and R7A variants of both U3.4 and U3.5, the short-range interactions ($\approx$4.5 Å) are very similar, probably due to the zwitterionic DPC head group providing an overall neutral environment for the alanine sidechain. However, as the interaction range is increased ($\approx$10 Å), the alanine sidechain interacts more favourably with the aliphatic carbon tails that constitute the core of the DPC micelle, compared to the cationic sidechains. Although, for A^7-U3.6 K7A, even at short-range, a noticeable preference for interaction with the DPC micelle core is observed.

Similar RDF calculations were performed between atom selections of seventh-position residue sidechain and the SDS micelle, as described in the previous text. In case of R^7-U3.4 wt in Figure 4.16a, one distinct peak and two flat peaks were observed, with reference to the SDS sulphate head groups, at 2.7 Å, 3.6 Å and 4.8 Å, with $g(r)$ values of 2.4, 2.1 and 2.3, respectively. Whereas for R^7-U3.5 wt, three peaks at 2.7 Å, 3.6 Å and 4.8 Å could be noted in Figure 4.16b, having $g(r)$ values of 3.4, 3.0 and 3.2, respectively. For K^7-U3.6 wt in Figure 4.16c, a sharp peak was observed at 2.6 Å and $g(r)$ of 16.3. Apart from that, two additional peaks were noted at 4.0 Å and 4.9 Å, with a $g(r)$ value of 9.3. In Figures 4.16a–4.16b, just one flat peak could be located for the RDF plots of A^7-U3.4 R7A and A^7-U3.5 R7A, at 8.9 Å and 8.6 Å, with $g(r)$ values of 3.2 and 3.8, respectively. No recognizable peak could be observed for A^7-U3.6 K7A, in Figure 4.16c. Moreover, in Figures 4.16d–4.16f, only one flat peak at 4.7 Å, 4.7 Å and 5.9 Å could be noted in the RDF plots of R^7-U3.4 wt, R^7-U3.5 wt and K^7-U3.6 wt, with reference to the SDS aliphatic carbon tails, with $g(r)$ values of 1.5, 1.4 and 2.6, respectively. For A^7-U3.4 R7A in Figure 4.16d, one distinct peak and two flat peaks could be observed at 4.1 Å, 9.8 Å and 13.8 Å, having $g(r)$ values of 3.0, 4.0 and 3.6, respectively. Likewise, for RDF plot of A^7-U3.5 R7A too, one distinct peak and two flat peaks could be noted at 4.0 Å, 10.0 Å and 14.8 Å, in Figure 4.16e, with $g(r)$ values of 3.8, 4.0 and 3.7,

respectively. However, in case of A^7-U3.6 K7A in Figure 4.16f, one distinct peak is found at 4.2 Å and a *g(r)* value of 3.6, with two broad peaks also observed.
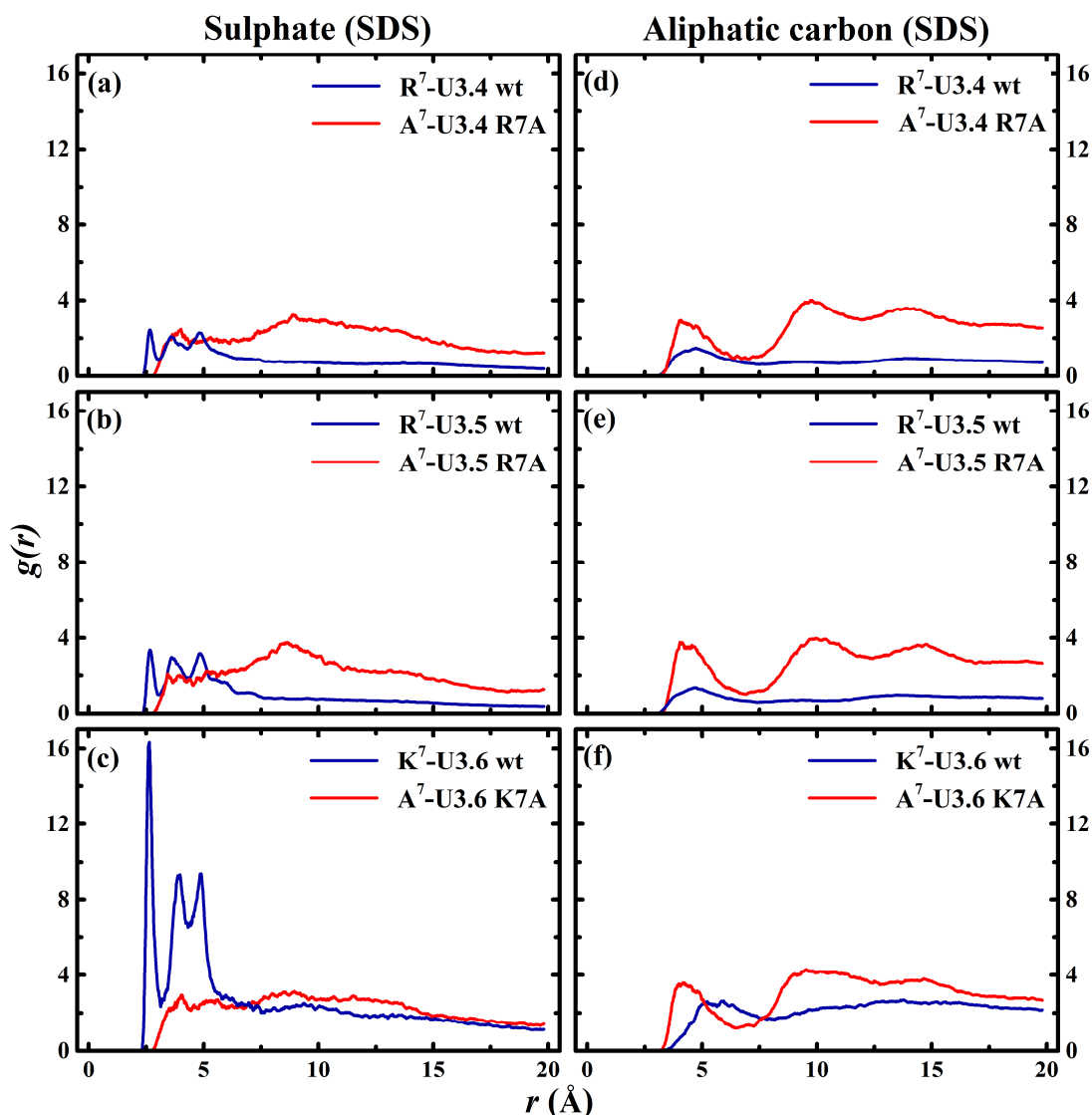
Figure 4.16. Radial distribution function plots for the sidechain heavy atoms of the seventh-position residue of the uperin 3.*x* peptides, with the sulphate head group (a–c) and the aliphatic carbon tail (d–f) heavy atoms of the SDS micelle.

For wild-type and R7A peptides of both U3.4 and U3.5, R^7 demonstrates marginally greater interaction with SDS sulphate head group, when compared to the A^7 residue. This could in part be explained by the presence of cationic residue at the neighbouring eight-position of both wild-type and variant peptides, which might slightly compensate for the loss of a neighbouring cationic residue at seventh-position of the variant. Moreover, the anionic SDS sulphate head group would interact more strongly with the positively-charged peptide residues, compared to the zwitterionic DPC head group. Also, peptide sequence would play a crucial role in peptide-

micelle interaction too. However, relative to A[7]-U3.6 K7A, K[7]-U3.6 wt has a significantly stronger interaction with the sulphate head groups. Compared to cationic sidechains, alanine sidechains interact more favourably with hydrophobic SDS micellar core.

4.3.3.3. Angular orientation analysis

The relative orientation of the seventh-position residue sidechain, with respect to the micelle centre of mass, was analysed. The angle between two vectors, one connecting the micelle centre of mass and α-carbon of the seventh-position residue, and the other between a specific side-chain atom and the α-carbon of seventh-position residue, was tracked over the course of the simulation, as depicted for the R[7]-U3.5 wt and SDS micelle in Figure 4.17. For the amino acid sidechain atom, the β-carbon atom of alanine, guanidium carbon atom of arginine, and ε-ammonium nitrogen atom of lysine were considered.
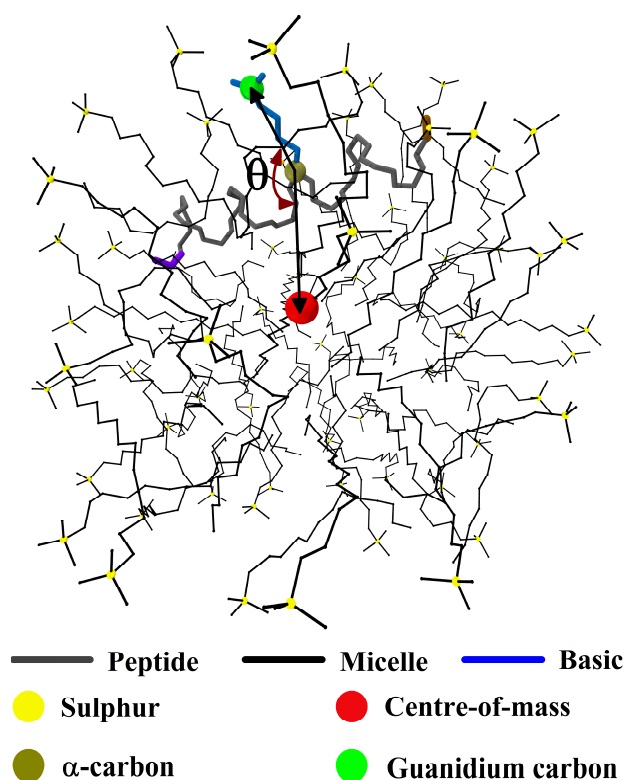
Figure 4.17. Angle between vectors connecting α-carbon of R[7]-U3.5 wt to its guanidium carbon, and to the SDS micelle centre of mass.

Angular orientation of R[7]-U3.4 wt relative to the centre of mass of the DPC micelle was observed at 100°, after ≈18 ns in Figure 4.18a, but at 140° for A[7]-U3.4 R7A. Likewise, for R[7]-U3.5 wt in Figure 4.18b, an angle of 115° was observed relative to the DPC micelle centre of mass. However, an angular orientation of 150° was found for A[7]-U3.5 R7A. Finally, the angular values of 135° and 150° were found for K[7]-U3.6 wt and A[7]-U3.6 K7A, respectively, with reference to the DPC micelle centre of mass in Figure 4.18c. Similar calculations were

performed for the simulations with the SDS micelles to determine the angle between the two vectors, as described earlier for the simulations with DPC micelles and shown in Figures 4.18 a–4.18c. The angular orientation of R^7-U3.4 wt relative to the centre of SDS micelle was observed at 120° in Figure 4.18d, but at 145° for A^7-U3.4 R7A. Similarly, for R^7-U3.5 wt, an angle of 135° was observed with respect to the DPC micelle centre of mass in Figure 4.18e. However, an angular orientation of 150° was found for A^7-U3.5 R7A. Further, in Figure 4.18f, angular values of 135° and 150° were noted for K^7-U3.6 wt and A^7-U3.6 K7A, respectively, relative to the DPC micelle centre of mass. Therefore, it could be noted that the seventh-position residues of wild-type peptides and corresponding alanine variants demonstrate different angular orientations relative to the respective micellar centre of mass. It should be noted that while alanine has a small sidechain, both lysine and arginine have relatively longer sidechains. Hence, these positively-charged sidechains show greater fluctuations in the angular orientation values.
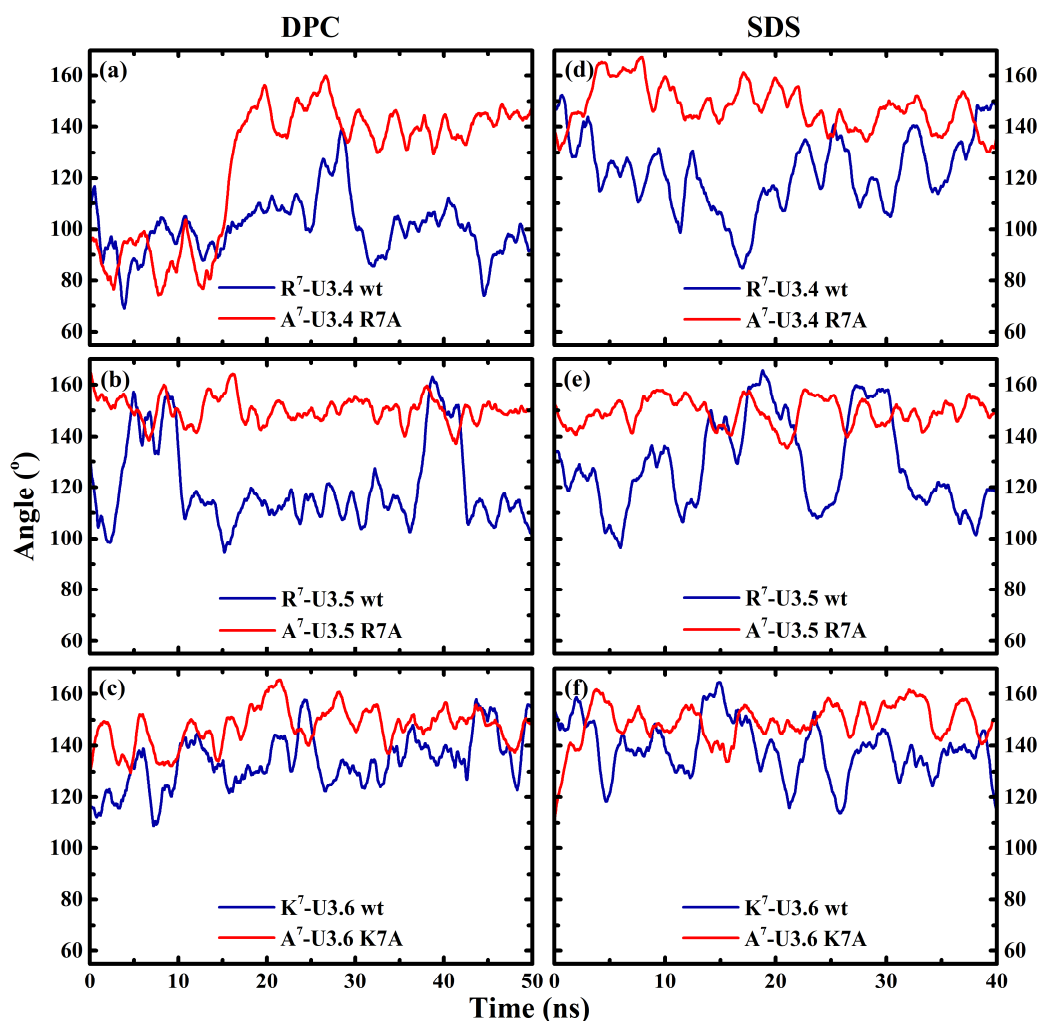
Figure 4.18. Angular orientation of selected sidechain heavy atoms of the seventh-position residue of the peptide, with respect to the centre of mass of (a–c) DPC and (d–f) SDS micelles.

4.3.4. Secondary structure evolution

Peptide secondary structure changes in these MD simulations, in the vicinity of micelles, are almost non-existent. In Figure 4.19, it is clear that the N- and C-termini amino acid residues are in a primarily coil secondary structure. Apart from these two terminal residues, the rest of the peptide is in a predominantly α-helical conformation throughout the simulation period. This trend was noted in simulations of both U3.5 wt and U3.5 R7A peptides near the DPC and SDS micellar environment. Quantification of the secondary structure content of these peptides, from the corresponding micelle simulations, demonstrated a similar pattern in Figure 4.20. The secondary structure composition is predominated by α-helical (88%) and coil content (12%), with occasional presence of 3_{10}-helix and turn structures being observed.
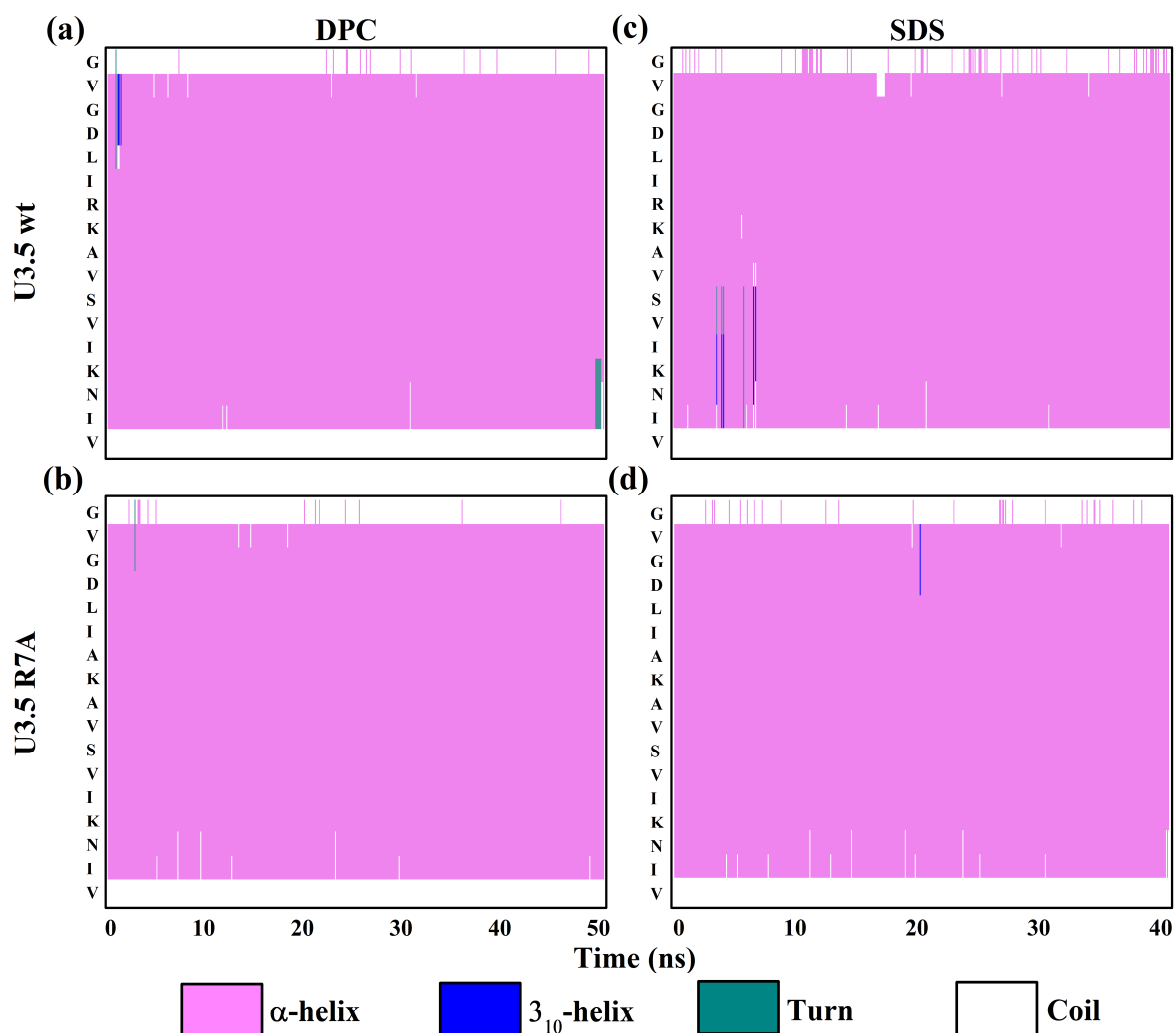
Figure 4.19. Secondary structure composition of U3.5 wt and U3.5 R7A peptides across the peptide length, obtained from the simulations with (a–b) DPC and (c–d) SDS micelles.

The lack of secondary structure changes upon replacement of a cationic residue with alanine at the seventh-position could be due to the high propensity of alanine residue to be in α-helical conformation in AMPs[36] despite the reduction of overall AMP amphipathicity, as discussed earlier for Figure 4.11. Moreover, the adjacent micellar molecules might restrict the degrees of freedom of these peptides, primarily due to the preferential interaction of the negatively-charged head groups and aliphatic tails of the micellar lipids with basic and nonpolar amino acid residues of these AMPs, respectively, as noted in Figures 4.14–4.16. Moreover, α-helical conformation of an amphipathic peptide might result in stronger binding and greater insertion into both DPC and SDS micelles, compared to the peptides in a different secondary structure conformation, such as a β-strand[37].
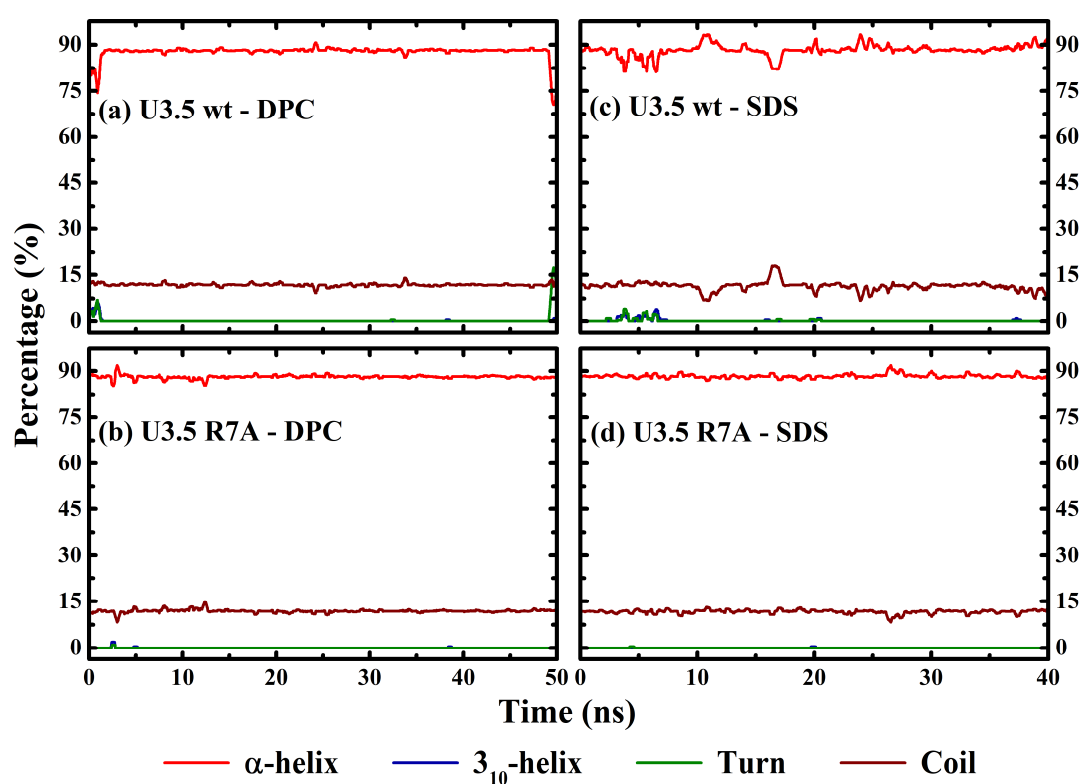
Figure 4.20. Average secondary structure composition of U3.5 wt and U3.5 R7A peptides, obtained from the simulations with (a–b) DPC and (c–d) SDS micelles.

4.3.5. Solvent-accessible surface area

Solvent-accessible surface area (SASA) could be defined as the peptide surface area accessed by a rolling sphere-probe of radius 1.4 Å, representing the aqueous solvent environment. The search parameters could be tuned to exclude solvent-interaction with the remaining peptide when calculating the SASA value for a particular peptide residue. Whereas relatively higher SASA values indicate greater solvent interaction, lower SASA values correspond to greater

micellar interactions. From the SASA plots of peptide-DPC micelle and peptide-SDS micelle simulation systems in Figures 4.21 and 4.22, respectively, it is noted that all plots follow a similar trend and have similar SASA values, irrespective of the nature of the micelle.
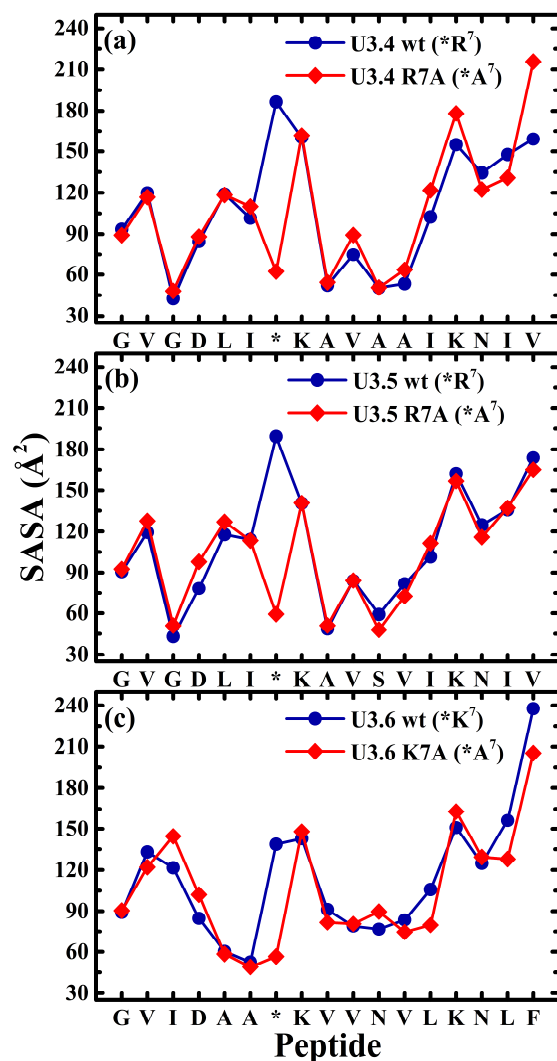
Figure 4.21. Solvent-accessible surface area (SASA) of individual residues of (a) U3.4 wt and U3.4 R7A, (b) U3.5 wt and U3.5 R7A, and (c) U3.6 wt and U3.6 K7A peptides, in complex with DPC micelle.

In all the plots, it was noted that there is a substantial difference between the SASA values of the seventh-position cationic and alanine residues of the wild-type peptides and their corresponding variants, respectively. The seventh-position cationic and alanine residues preferentially interact with the surrounding ionic aqueous environment and micellar head groups, and the micellar core, respectively, as discussed in Figures 4.14–4.16. For wild-type and R7A variants of U3.4 and U3.5, local maxima were noted at V^2 and K^{14} residues, and local minima were observed at G^3 and N^{15} residues. Furthermore, for U3.4 wt and U3.5 wt peptides, a local maximum was observed at R^7. Similarly, a local minimum was noted at A^7 for the

respective R7A variants. A low-solvation region was noted at A^9VAA^{12} segment of both U3.4 wt and U3.4 R7A, indicating towards the critical role played by this nonpolar segment in peptide-micelle interaction. Likewise, for U3.5 wt and U3.5 R7A, relatively lower SASA values were observed in the predominantly hydrophobic A^9VSVI^{13} region. Significantly, in a recent study, this region has been found to play a central role in the greater inter-peptide interaction of U3.5 wt, as a function of increasing NaCl concentration[38].
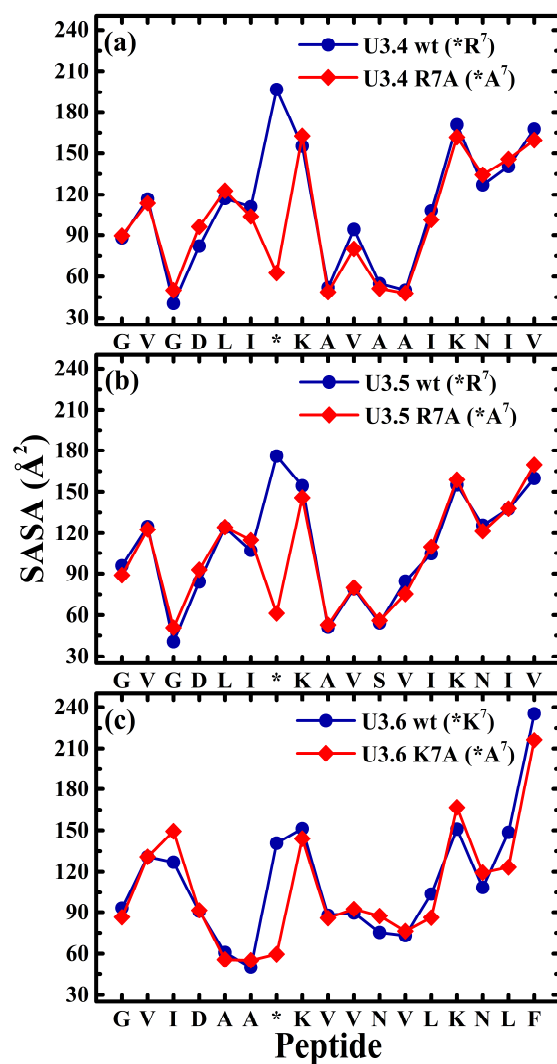
Figure 4.22. Solvent-accessible surface area (SASA) of individual residues of (a) U3.4 wt and U3.4 R7A, (b) U3.5 wt and U3.5 R7A, and (c) U3.6 wt and U3.6 K7A peptides, in complex with SDS micelle.

For U3.6 wt, local SASA maxima were noted at V^2, K^7K^8 and K^{14} segments, and local solvation minima were found at A^7A^8 and N^{15} regions. Similarly, in the case of U3.6 K7A, local SASA maxima were observed at I^3, K^8 and K^{14} positions, and local solvation minima was noted at $K^{14}N^{15}$ region. Both U3.6 wt and U3.6 K7A peptides demonstrate low SASA values in V^9VNVL^{13} region, primarily composed of nonpolar residues. Additionally, another low

solvation zone A^5AA^7, composed of three alanine residues, is observed in U3.6 K7A. This could explain the relatively stronger interaction of U3.6 K7A peptides with DPC and SDS micelles in Figures 4.7 and 4.9, respectively, relative to the other two alanine variants. Hence, when compared to U3.6 wt, the U3.6 K7A displayed slightly lower peptide-micelle interaction energies only. The other two wild-type - variant peptide pairs have appreciably greater difference in their interaction energies. Although as discussed earlier, the stronger charge-distribution of arginine relative to lysine also needs to be considered here[34]. Therefore, the hydrophobic effect drives these predominantly nonpolar segments to play an important role in the interaction of these peptides with both DPC and SDS micelles[39]. However, it should be noted that the solvation values of different residues are also dependent on the primarily helical secondary structure of the peptide, apart from the influence of the neighbouring residues and the micelle, in complex with that particular peptide.

4.4. Conclusions

Amongst all uperin 3.*x* peptides and seventh-position variants, compared to the DPC micelle, preferential interaction towards the SDS micelle was observed. Moreover, substitution of a seventh-position cationic residue with alanine significantly diminished the peptide interaction towards both DPC and SDS micelles. U3.4 wt and U3.6 wt displayed the strongest and weakest interactions, respectively, with DPC micelle. Whereas U3.5 wt interacted most favourably with the SDS micelle, U3.4 wt had the least favourable interaction. For alanine variants, U3.6 K7A demonstrated the greatest affinity towards both types of micelles, and U3.4 R7A the least. However, remarkable secondary structure changes could not be observed in both wild-type and variant peptides, and the α-helical starting structure of the peptide was mostly maintained. As expected, segments of predominantly nonpolar amino acid residues were found to preferentially interact with the hydrophobic interior of the micelles, thereby playing a critical role in the peptide-micelle interaction.

4.5. References

1.   Bradford, A. M.; Bowie, J. H.; Tyler, M. J.; Wallace, J. C., New Antibiotic Uperin Peptides from the Dorsal Glands of the Australian Toadlet Uperoleia Mjobergii. *Aust. J. Chem.* **1996**, *49*, 1325-1331.

2.   Shai, Y., Molecular Recognition between Membrane-Spanning Polypeptides. *Trends in Biochemical Sciences* **1995**, *20*, 460-464.

3.   Chia, B. C. S.; Bowie, J. H.; Carver, J. A.; Mulhern, T. D., The Solution Structure of Uperin 3.6, an Antibiotic Peptide from the Granular Dorsal Glands of the Australian Toadlet, Uperoleia Mjobergii. *The Journal of Peptide Research* **1999**, *54*, 137-145.

4.   Martin, L. L.; Kubeil, C.; Piantavigna, S.; Tikkoo, T.; Gray, N. P.; John, T.; Calabrese, A. N.; Liu, Y.; Hong, Y.; Hossain, M. A., Amyloid Aggregation and Membrane Activity of the Antimicrobial Peptide Uperin 3.5. *Pept. Sci.* **2018**, *e24052*.

5.   Mihajlovic, M.; Lazaridis, T., Charge Distribution and Imperfect Amphipathicity Affect Pore Formation by Antimicrobial Peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818*, 1274-1283.

6.   Tieleman, D. P.; van der Spoel, D.; Berendsen, H. J. C., Molecular Dynamics Simulations of Dodecylphosphocholine Micelles at Three Different Aggregate Sizes: Micellar Structure and Chain Relaxation. *The Journal of Physical Chemistry B* **2000**, *104*, 6380-6388.

7.   MacKerell, A. D., Molecular Dynamics Simulation Analysis of a Sodium Dodecyl Sulfate Micelle in Aqueous Solution: Decreased Fluidity of the Micelle Hydrocarbon Interior. *The Journal of Physical Chemistry* **1995**, *99*, 1846-1855.

8.   Rosengren, J., Martin, L. L., Ed. The University of Queensland, 2017.

9.   Calabrese, A. N.; Liu, Y.; Wang, T.; Musgrave, I. F.; Pukala, T. L.; Tabor, R. F.; Martin, L. L.; Carver, J. A.; Bowie, J. H., The Amyloid Fibril-Forming Properties of the Amphibian Antimicrobial Peptide Uperin 3.5. *ChemBioChem* **2016**, *17*, 239-246.

10.   Waterhouse, A., et al., Swiss-Model: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* **2018**, *46*, W296-W303.

11.   Johan Strumpfer, P. F., Christophe Chipot, Molefacture: A Tutorial to Build and Edit Molecules. *Centre National de la Recherche Scientifique, University of Illinois, Urbana–Champaign c)* *http://www.ks.uiuc.edu/Research/vmd* **2012**.

12.   Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.

13. Kar, R. K.; Mroue, K. H.; Kumar, D.; Tejo, B. A.; Bhunia, A., Structure and Dynamics of Antifreeze Protein–Model Membrane Interactions: A Combined Spectroscopic and Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2016**, *120*, 902-914.

14. Seeliger, D.; de Groot, B. L., Ligand Docking and Binding Site Analysis with Pymol and Autodock/Vina. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 417-422.

15. Hou, X.; Du, J.; Zhang, J.; Du, L.; Fang, H.; Li, M., How to Improve Docking Accuracy of Autodock4.2: A Case Study Using Different Electrostatic Potentials. *Journal of Chemical Information and Modeling* **2013**, *53*, 188-200.

16. Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S., Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein−Ligand Complexes. *Journal of Chemical Information and Modeling* **2006**, *46*, 380-391.

17. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **2016**, *14*, 71.

18. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W., Update of the Charmm All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **2010**, *114*, 7830-7843.

19. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

20. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005**, *26*, 1781-802.

21. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

22. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

23. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

24. Sayyed-Ahmad, A.; Khandelia, H.; Kaznessis, Y. N., Relative Free Energy of Binding between Antimicrobial Peptides and Sds or Dpc Micelles. *Molecular Simulation* **2009**, *35*, 986-997.

25. Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M., Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157-64.

26. Stone, J. E., An Efficient Library for Parallel Ray Tracing and Animation. *Masters Theses. http://scholarsmine.mst.edu/masters_theses/1747* **1998**, *1747*.

27. Torrent, M.; Di Tommaso, P.; Pulido, D.; Nogués, M. V.; Notredame, C.; Boix, E.; Andreu, D., Ampa: An Automated Web Server for Prediction of Protein Antimicrobial Regions. *Bioinformatics (Oxford, England)* **2012**, *28*, 130-131.

28. Gautier, R.; Douguet, D.; Antonny, B.; Drin, G., Heliquest: A Web Server to Screen Sequences with Specific A-Helical Properties. *Bioinformatics* **2008**, *24*, 2101-2102.

29. Frishman, D.; Argos, P., Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566-579.

30. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577-2637.

31. Isralewitz, B., Timeline: A Vmd Plugin for Trajectory Analysis. Tutorial. *University of Illinois at Urbana-Champaign c) http://www.ks.uiuc.edu/Research/vmd* **2012**.

32. Aykul, S.; Martinez-Hackert, E., Determination of Half-Maximal Inhibitory Concentration Using Biosensor-Based Protein Interaction Analysis. *Analytical Biochemistry* **2016**, *508*, 97-103.

33. Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C., The Helical Hydrophobic Moment: A Measure of the Amphiphilicity of a Helix. *Nature* **1982**, *299*, 371.

34. Rice, A.; Wereszczynski, J., Probing the Disparate Effects of Arginine and Lysine Residues on Antimicrobial Peptide/Bilayer Association. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2017**, *1859*, 1941-1950.

35. Khandelia, H.; Kaznessis, Y. N., Molecular Dynamics Simulations of the Helical Antimicrobial Peptide Ovispirin-1 in a Zwitterionic Dodecylphosphocholine Micelle: Insights into Host-Cell Toxicity. *The Journal of Physical Chemistry B* **2005**, *109*, 12990-12996.

36. Zhang, M.; Zhao, J.; Zheng, J., Molecular Understanding of a Potential Functional Link between Antimicrobial and Amyloid Peptides. *Soft Matter* **2014**, *10*, 7425-7451.

37. Wang, Q.; Hong, G.; Johnson, G. R.; Pachter, R.; Cheung, M. S., Biophysical Properties of Membrane-Active Peptides Based on Micelle Modeling: A Case Study of Cell-Penetrating and Antimicrobial Peptides. *The Journal of Physical Chemistry B* **2010**, *114*, 13726-13735.

38. Ray, S.; Holden, S.; Martin, L. L.; Panwar, A. S., Mechanistic Insight into the Early Stages of Amyloid Formation Using an Anuran Peptide. *Pept. Sci.* **2019**, *e24120*.

39. Ivanov, I.; Vemparala, S.; Pophristic, V.; Kuroda, K.; DeGrado, W. F.; McCammon, J. A.; Klein, M. L., Characterization of Nonbiological Antimicrobial Polymers in Aqueous Solution and at Water−Lipid Interfaces from All-Atom Molecular Dynamics. *Journal of the American Chemical Society* **2006**, *128*, 1778-1779.

Chapter 5

Propensity of Uperin 3.*x* peptides towards beta-aggregation: Analysis of structurally-diverse conformers

5.1. Introduction

Amyloidogenic diseases are characterized by the transformation of certain peptides or proteins from their native soluble states to highly organized fibrillar aggregates[1-2], typically marked by an extended arrangement of hydrogen-bonded β-sheets[3]. However, the intermediate protein oligomers or protofibrils, devoid of any fibrillar structure, are thought to be primarily responsible for the pathogenicity associated with these diseases[2-3]. Interestingly, many antimicrobial peptides (AMPs), such as indolicidin, magainin 2 and protegrin-1, also have the potential to generate amyloid-like fibrils[4]. The present study focuses on uperin 3.*x* peptides, known to have antimicrobial activity against Gram-positive bacteria[5] and mammalian cell lines[6].

The Uperin (U) 3.5 wild-type (wt) has been of interest, as it is stable in water; and only aggregates on addition of salts[6]. Also, it was observed that U3.5 wt aggregates less than U3.5 R7A variant[7]. Thus, the substitution of a positively-charged amino acid at the seventh-position with a nonpolar alanine residue has been found to increase fibril formation. Significantly, most pathogenic mutations, such as those observed in transthyretin and lysozyme amyloidosis, increase the aggregation-propensity of the target protein, either directly or indirectly[3]. Similarly, the K28A mutation of the Aβ$_{21-30}$ fragment increases its intra-peptide hydrophobic

interactions[8]. Furthermore, for both p53$_{251-257}$[9] and p53$_{252-258}$[10] fragments of the tumour suppressor p53 protein, the I254R mutation, which substitutes the nonpolar isoleucine residue with the positively-charged arginine residue, acts as an aggregation-suppressor. Although the wild-type fragments generate β-sheet structures, the I254R mutants demonstrate a significantly lower rate of aggregation and remain in an unstructured state[9-10]. Similarly, the introduction of negatively-charged residues in A53E and G51D variants hindered α-synuclein assembly due to electrostatic repulsion[11]. Hence, these 17 amino acid residues long uperin 3.x peptides[5] can serve as model peptides, and provide important insights into the effect of cationic to hydrophobic residue substitution on the manifestation of amyloidogenic diseases[1, 3].

Table 5.1. Amino acid sequences of uperin 3.x wild-type peptides[5] and corresponding seventh-position alanine variants.

	Peptide	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	C-Ter
1.	U3.4 wt	G	V	G	D	L	I	R	K	A	V	A	A	I	K	N	I	V	-NH$_2$
2.	U3.4 R7A	G	V	G	D	L	I	A	K	A	V	A	A	I	K	N	I	V	-NH$_2$
3.	U3.5 wt	G	V	G	D	L	I	R	K	A	V	S	V	I	K	N	I	V	-NH$_2$
4.	U3.5 R7A	G	V	G	D	L	I	A	K	A	V	S	V	I	K	N	I	V	-NH$_2$
5.	U3.6 wt	G	V	I	D	A	A	K	K	V	V	N	V	L	K	N	L	F	-NH$_2$
6.	U3.6 K7A	G	V	I	D	A	A	A	K	V	V	N	V	L	K	N	L	F	-NH$_2$

Uperin wild-type peptides from 3.x family, including U3.4 wt, U3.5 wt and U3.6 wt, were originally obtained from the skin secretions of *Uperoleia mjobergii*, and are found with C-terminus amidation[5]. The amino acid sequences of these three wild-type peptides; and their seventh-position alanine variants, U3.4 R7A, U3.5 R7A and U3.6 K7A, are shown in Table 5.1. Although, as mentioned earlier, experimental studies have demonstrated the greater aggregation of U3.5 R7A relative to U3.5 wt[7], but mechanistic details, especially of the initial stages of self-aggregation of these uperin 3.x wild-type peptides and their variants, are lacking. More specifically, substitution of the positively-charged residue at seventh position of these wild-type peptides with nonpolar alanine might alter the secondary structure evolution and the inter-peptide interactions of the respective variants. Tsigelny et al.[12] self-docked a diversity of amyloid-beta (Aβ)$_{42}$ peptide conformers to identify the ones capable of generating propagating structures. Using fully-atomistic (FA) molecular dynamics (MD) simulations, we employed a similar strategy to investigate the aggregation properties of the uperin 3.x (wt and R/K7A)

peptides, listed in Table 5.1, and to provide insight into the evolution of secondary structure of these peptides.

5.2. Methods

5.2.1. Simulation details

The fully-atomistic (FA) molecular dynamics (MD) simulations were performed using the CHARMM36m force-field for proteins[13] and TIP3P water[14] in NAMD[15]. Periodic boundary conditions were applied along all three orthonormal directions. A switching function was implemented for the Lennard-Jones potential, with a smooth cut-off from 12 to 14 Å. Particle-mesh Ewald summation[16], with a grid spacing of 1 Å, was used for calculating electrostatic interactions. Simulations were performed using a 2 fs timestep and a Langevin thermostat, with a damping coefficient of 1 ps^{-1}. A constant pressure of 1 atmosphere (1 atm = 1.01325 bar) was maintained using a Nosé-Hoover[17] Langevin[18] piston, with a period of 100 fs and a decay time of 50 fs. The peptides were solvated with TIP3P water molecules. The physiological environment was modelled by adding Na$^+$ and Cl$^-$ ions; to obtain an NaCl concentration of 0.15 M. All systems were equilibrated for 1 ns in an NVT ensemble ($T = 310$ K), followed by further equilibration for 1 ns in an NPT ensemble ($P = 1$ atm and $T = 310$ K), with the peptide backbone restrained by a harmonic constraint of 10 kcal mol^{-1}Å^{-2}. After gradual removal of the backbone constraints, the peptides were allowed to equilibrate for a period of 2 ns in an NPT ensemble ($P = 1$ atm and $T = 310$ K). The resulting systems were then simulated in NPT ensemble ($P = 1$ atm and $T = 310$ K).

The molecular editors PACKMOL[19] and VMD[20] were used to create the simulation systems. Tachyon ray-tracer[21], as incorporated in VMD, was used for visualization. Secondary structure was determined with the STRIDE program[22] using DSSP definitions[23], as implemented by the *Timeline* plugin[24] in VMD. Data analyses were performed using VMD and in-house scripts.

5.2.2. Simulating α-helix to coil transition

The starting structure for U3.5 wt was based on the α-helical representation shown in Figure 5.1b, determined by NMR, with the peptide present in aqueous sodium dodecyl sulphate (SDS) solution[25]. The peptide was unstructured in a pure aqueous environment, as measured with circular dichroism. Using SWISS-MODEL Workspace[26], the initial structure of U3.4 wt was generated using the U3.5 wt as a template. These two peptides differ at only two amino acids in the eleventh and twelfth positions of their sequences, as indicated in Table 5.1. Not

surprisingly, the U3.4 wt also consisted of a primarily α-helical secondary structure (Figure 5.1a). The U3.6 wt peptide has been shown to adopt a predominantly α-helical amphipathic structure by NMR in the membrane-mimicking TFE-aqueous environment[27]. Hence, an α-helical structure was imposed on the U3.6 wt peptide sequence, as shown in Figure 5.1c, using the *Protein Builder* module[28] in VMD[20]. In order to generate alanine variants for the three uperin 3.*x* wild-type peptides, the *Mutator* plugin of VMD was employed to execute a point substitution at the seventh-position of the wild-type peptides. Therefore, like the uperin 3.*x* wild-type peptides used as templates, the three alanine-containing uperin 3.*x* variants also consisted of α-helical secondary structures at the onset of the simulations, as shown in Figures 5.1d–5.1f. Each of these six α-helical uperin 3.*x* (wt and R/K7A) peptides, shown in Figure 5.1d–5.1f, were placed in an aqueous environment containing 0.15 M NaCl. All peptides showed a α-helix-to-coil transition, as shown for U3.5 wt in Figure 5.2.
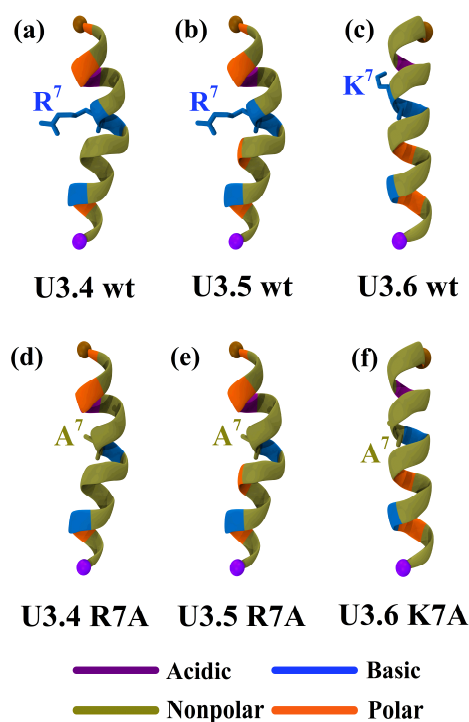
Figure 5.1. Molecular models of the uperin 3.*x* peptides in this study, represented as an α-helix. (a-c) shows the three uperin 3.*x* wild-type peptides and (d-f) their corresponding seventh-position variants. Heavy (non-hydrogen) atoms of the seventh-position residues have been shown with stick representation. A brown sphere represents the N-terminus and a purple sphere represents the C-terminus of the peptides.

5.2.3. Clustering of peptide conformers

A diversity of peptide conformers, with different backbone structures and secondary structure compositions, was observed during the simulations showing α-helix-to-coil transition for each of the uperin 3.*x* peptides and their corresponding alanine variants, as depicted for U3.5 wt in Figure 5.2. To achieve a better alignment, these peptide conformers were clustered into groups; based on their backbone structure, with *RMSD* (Root-Mean-Square-Deviation) relative to the backbone of the α-helical starting structure as reference. This *RMSD* fitting was performed with the *RMSD Trajectory Tool* of VMD.
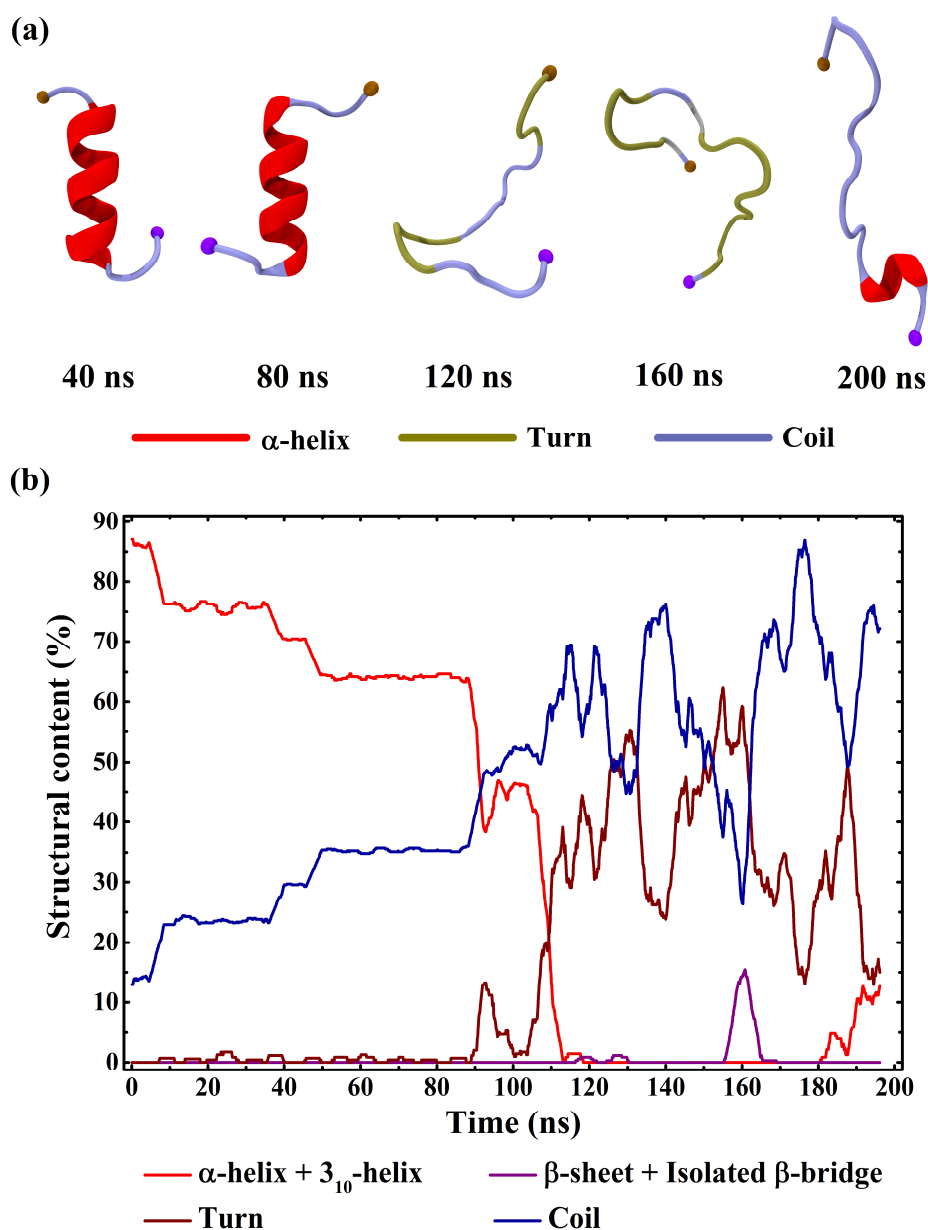
Figure 5.2. Secondary structure changes in U3.5 wt over the simulation period, represented by (a) peptide trajectory snapshots, and (b) a plot of variation of secondary structure elements with time.

145

After the *RMSD* fitting of the different peptide conformers, *RMSD* and Q_{res}[29] values were obtained for individual amino acid residues of each peptide conformer using the *Multiseq* plugin[29] in VMD. Q_{res} is a structural similarity parameter based on the Cα atom positions of the individual peptide residues. Using the *RMSD* and Q_{res} values discussed earlier, conformational analyses were performed to sort the peptide conformers from each of the six peptide trajectories into different structural families. First, hierarchical clustering was performed using the furthest neighbour method, as employed in Origin[30]. After this initial clustering, 19 conformer families were selected, as shown in Figure 5.3a for U3.5 wt. The individual cluster centres of each of these families was used as an input for subsequent k-means clustering, as shown in Figure 5.3b. After this operation, the corresponding centroid conformers from each of the conformer families were used for further analyses. Centroid conformer can be defined as the conformer, which is closest to the cluster centre after k-means clustering.
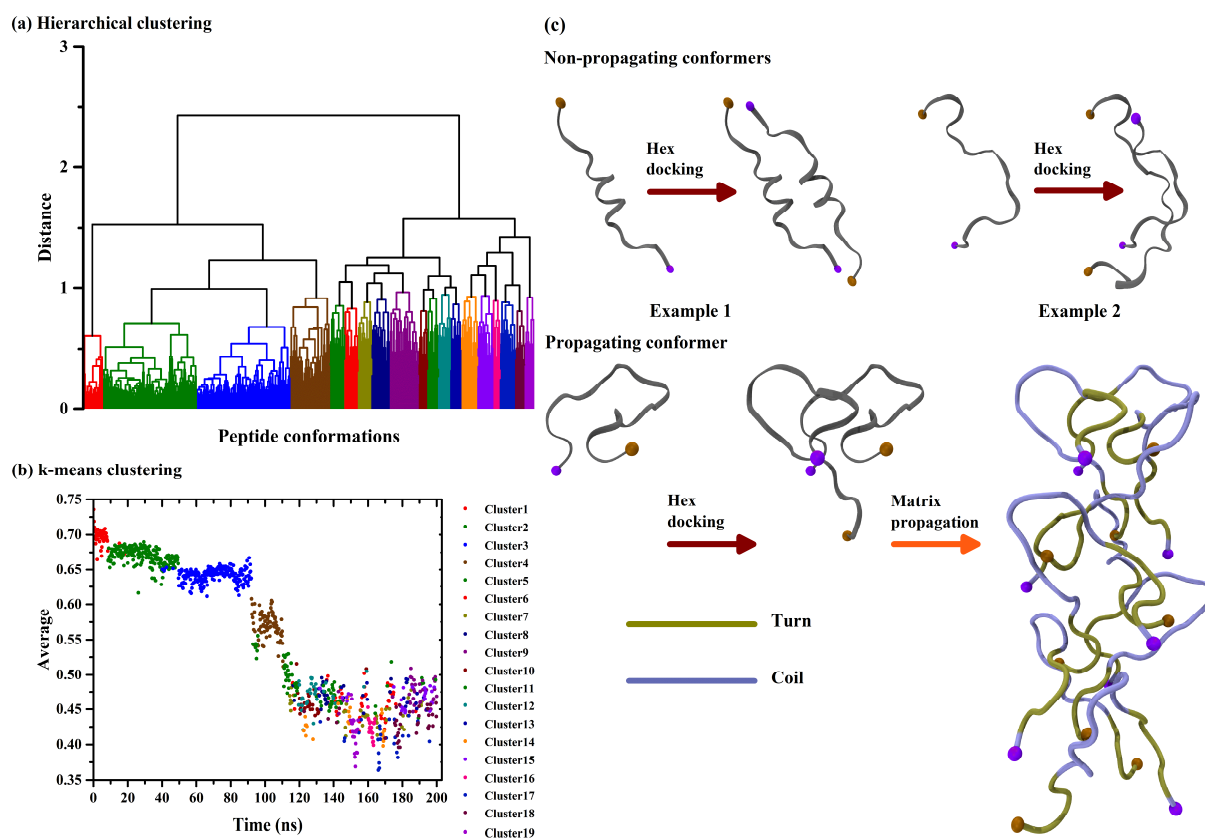
Figure 5.3. (a) Hierarchical clustering of U3.5 wt conformers, based on the *RMSD* and Q_{res} values of individual residues. (b) k-means clustering using the initial cluster centres from hierarchical clustering, and output plotted as the average of all *RMSD* and Q_{res} values for each conformer vs time. (c) An example of two of the non-propagating U3.5 wt conformers, and a propagating U3.5 wt conformer; with additional peptides docked to the dimer to generate an octamer aggregate structure.

146

5.2.4. Docking and propagation

The centroid conformers were docked with another copy of themselves (self-docking) to generate a two-peptide aggregate (dimer); using shape complementarity and electrostatics interactions as parameters, as employed by the Hex program[31]. Using the rotation and translation parameters i.e. the three-dimensional transformation matrix generated after the Hex dockings, more copies of that particular conformer were added and docked to the dimer, until the propagating conformers were revealed, as shown in Figure 5.3c. Unlike the propagating conformers, non-propagating conformers, such as those shown in Figure 5.3c, produced molecular structures with steric clashes upon docking additional copies to the previously docked dimer. For the propagating conformers, those with greater docking favourability and higher turn-coil content were selected for further docking and simulations (please refer to the section 5.6.1. of the Appendix for details). The aggregates consisting of eight docked peptides (octamers) were simulated for 100 ns using the parameters mentioned in the simulation details. The stability of these octamers was quantified using the radius of gyration (R_g), and the secondary structure evolution was analysed over the entire simulation trajectory.

5.3. Results

5.3.1. α-helix-to-coil transition

All the six peptides with α-helical starting structures were simulated till a complete loss of helical content was obtained. Here the sum total of α-helical and 3_{10}-helical content has been referred to as helical content. However, as observed in Figure 5.4, the rate of this α-helix-to-coil transition differed for each peptide. Moreover, all six simulations were allowed to proceed further for approximately 100 ns; even after the completion of α-helix-to-coil transition, to obtain a greater diversity of peptide conformers for further analyses. As a result, the simulation periods varied between 200–450 ns. In initial stages of all these six simulations, the increase in coil content is accompanied by a corresponding decrease in helical content. Whereas U3.4 wt and U3.5 wt demonstrated a rapid loss of helical content in Figure 5.4, the other four peptides displayed a more gradual α-helix-to-coil transition. Interestingly, transient re-emergence of some helical content could be observed at later stages of most of these simulations, except for those of U3.4 wt and U3.4 R7A peptides. Furthermore, the proportion of turn structure becomes prominent only in the later stages of these simulations. Negligible β-content could be observed over the entire simulation period. Post α-helix-to-coil transition, the peptide is predominantly

147

composed of coil and turn secondary structure elements. The U3.4 wt α-helical structure is significantly more stable than that of U3.4 R7A. However, the α-helical structure of U3.5 R7A was found to be appreciably more stable than that of U3.5 wt. Interestingly, U3.6 wt and U3.6 K7A α-helices demonstrate similar stability, with that of U3.6 wt being slightly higher. The following order of helical stability was noted amongst all the uperin 3.*x* peptides: U3.4 wt>U3.6 wt>U3.6 K7A>U3.5 R7A>U3.5 wt>U3.4 R7A.
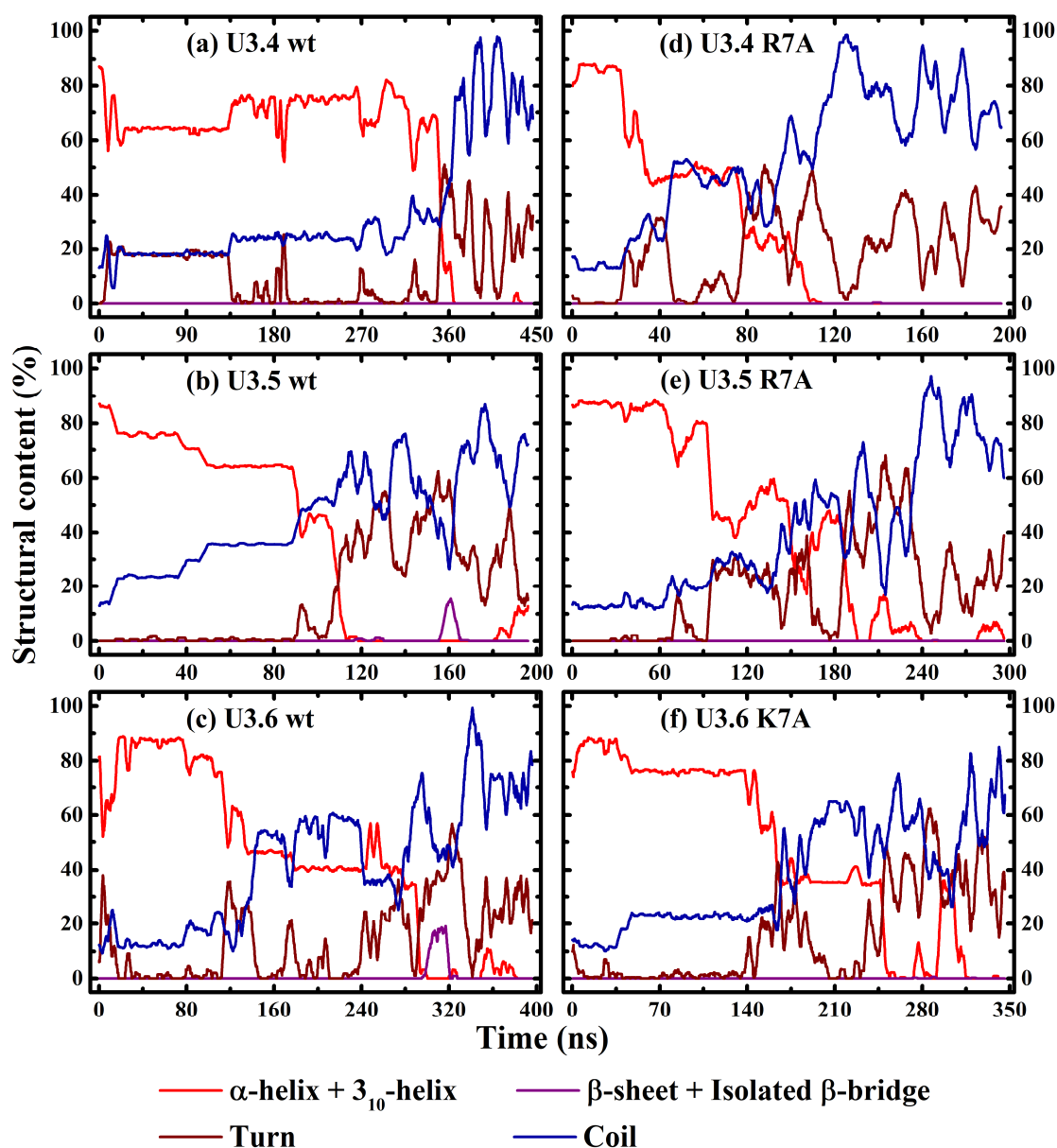
Figure 5.4. Secondary structure variation in (a) U3.4 wt, (b) U3.5 wt and (c) U3.6 wt peptides, and corresponding (d) U3.4 R7A, (e) U3.5 R7A and (f) U3.6 K7A variants, over the simulation time period.

5.3.2. Octamer analysis

As described earlier in section 5.2, the peptide conformers obtained during the α-helix-to-coil transition of the uperin 3.*x* peptides and variants were subjected to hierarchical; and then k-means clustering. For the simulated octamers, generated from different propagating conformers of each of the six peptides, only one conformer was selected for extensive analyses for each of these peptides (selection criteria explained in section 5.6.1. of the Appendix). Significantly, these octamers in Figure 5.5c were composed of only turn and coil secondary structures, except for those of U3.6 wt and U3.6 K7A, where some isolated β-bridges can also be observed. Such unstructured aggregates might act as incipient nuclei for the structured amyloid fibrils, generally observed down the amyloidogenesis pathway[1, 3].

(a) Monomer

(b) Dimer

Hex docking (HD)

(c) Octamer

Matrix propagation (MP)

| U3.4 wt | U3.5 wt | U3.6 wt | U3.4 R7A | U3.5 R7A | U3.6 K7A |

3¹⁰-helix — Isolated β-bridge — Turn — Coil
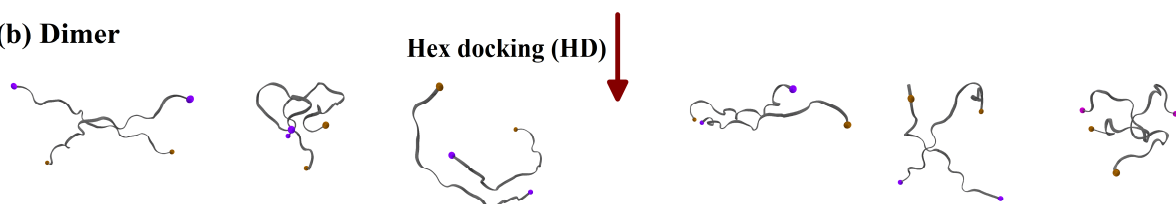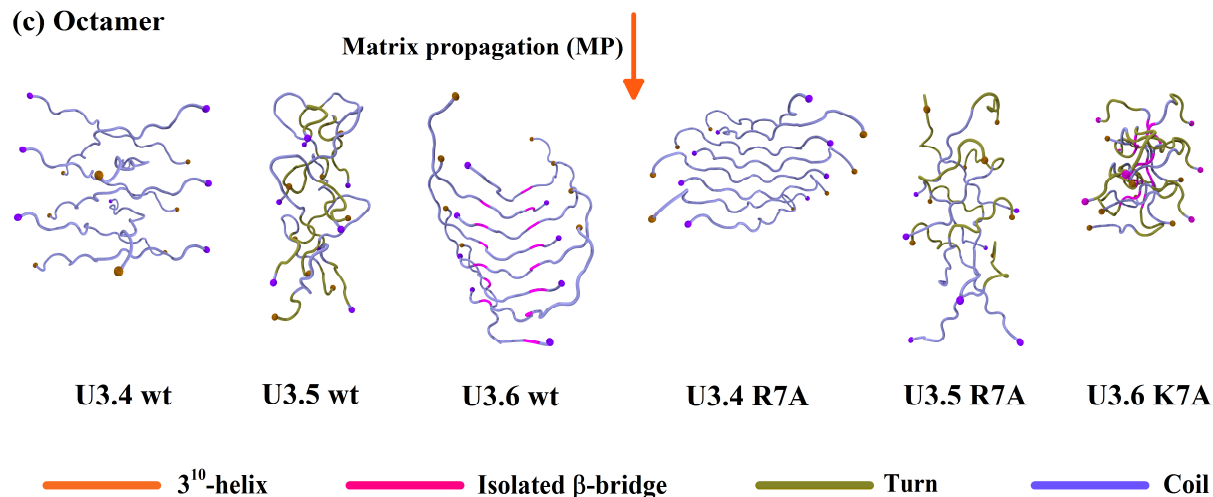
Figure 5.5. (a) Propagating conformers for all uperin 3.*x* peptides and variants, (b) the monomers are docked with another copy of themselves using Hex. (c) Transformation matrix, generated post Hex docking, used to dock further copies of the peptide conformer to the docked dimer to generate an octamer aggregate.

5.3.2.1. Octamer stability

Analysis of the octamer aggregates was undertaken after 100 ns simulation period, using the overall R_g as a parameter of interest. Smaller R_g values indicate that the octamer is more compact than those with larger R_g values. Greater stability of an octamer might be due to the stronger inter-peptide interactions within that particular octamer. The relatively minor fluctuations in R_g values are probably due to the positional rearrangements within that octamer. The U3.4 wt octamer, in Figure 5.6a, has a larger R_g value towards the end of the simulation; than that of the U3.4 R7A octamer. Although the U3.5 wt octamer disintegrates into smaller fragments over the course of the simulation, as observed from the high R_g values in Figure 5.6b, but the U3.5 R7A octamer remained stable. The U3.6 wt octamer, in Figure 5.6c, has a marginally higher R_g value at 100 ns, compared to the U3.6 K7A octamer. Therefore, it could be noted that the octamers of R/K7A uperin 3.x peptides remained more stable over the course of the simulation, relative to the respective uperin 3.x wild-type octamers.
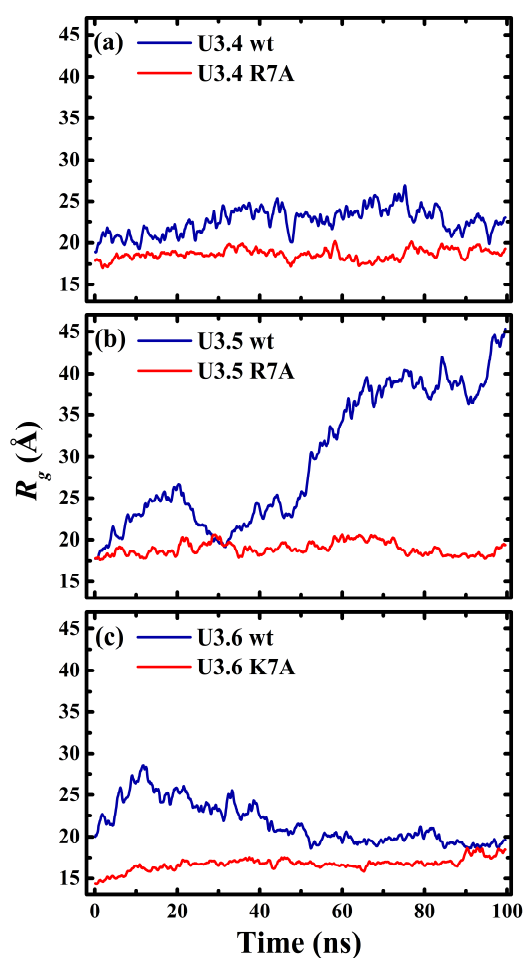
Figure 5.6. R_g plots of the octamers of (a) U3.4 wt and U3.4 R7A, (b) U3.5 wt and U3.5 R7A, and (c) U3.6 wt and U3.6 K7A, over the entire simulation period.

5.3.2.2. Beta-sheet content

The β-sheet secondary structure elements are composed of individual β-strand segments within certain peptides, as observed in Figure 5.7. For the sake of clarity and easier comparison between the different octamer trajectory snapshots, we discuss the individual β-strands within an octamer. Lesser number of β-strands could be observed, over the simulation period, within the U3.4 wt octamer in Figure 5.7, as compared to the U3.4 R7A octamer. Similarly, a higher β-strand content was observed for the U3.5 R7A octamer in Figure 5.7, compared to the U3.5 wt octamer.
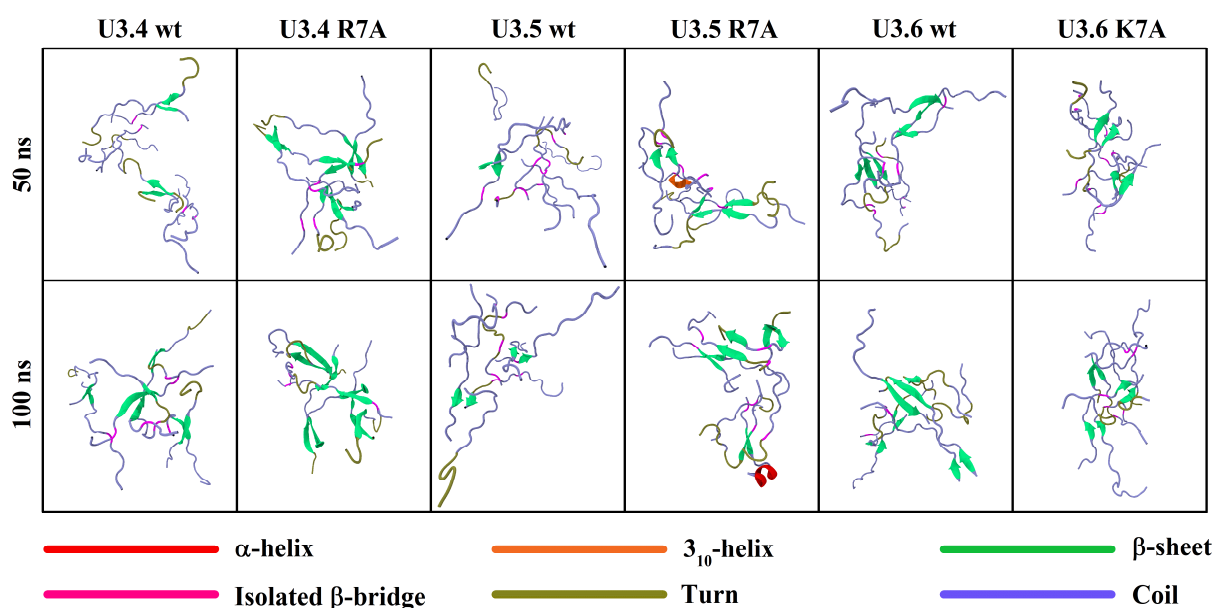
Figure 5.7. Trajectory snapshots of the uperin 3.*x* wild-type peptides and the corresponding seventh-position variants at the mid (50 ns) and end (100 ns) stages of the respective simulations.

In stark contrast to U3.4 and U3.5, the U3.6 wt octamer, in Figure 5.8, displayed higher β-strand content than the U3.6 K7A octamer. It should be noted that the U3.6 wt peptide has a vastly different amino acid sequence than those of the U3.4 wt and U3.5 wt (Table 5.1). Apart from the β-strands, appreciable number of isolated β-bridges, and significant turn and coil secondary structure content can be observed in Figure 5.8. Moreover, few 3_{10}-helix and α-helix segments can also be noted in some of the snapshots. Furthermore, greater number of β-strands were noted at 100 ns, compared to the 50 ns trajectory snapshots. This indicates towards certain level of β-sheet propensity among all these uperin 3.*x* (wt and R/K7A) peptides.

The evolution of β-sheet content for the selected uperin 3.*x* octamers, over 100 ns simulation period, is shown in Figure 5.8a. The U3.4 R7A octamer contains 16% higher β-sheet content

than the U3.4 wt octamer; towards the end of the simulation. Likewise, a 16% higher β-sheet content can be observed for the U3.5 R7A octamer, relative to the U3.5 wt octamer. A similar observation was made by Martin et al. in their experiments with U3.5 wt and U3.5 R7A[7]. However, as noted earlier in Figure 5.7, the U3.6 wt octamer has 11% higher β-sheet content than the U3.6 K7A octamer. The following order of β-sheet content was noted at the end of these 100 ns simulations amongst all the uperin 3.x peptides and variants: U3.4 R7A>U3.5 R7A>U3.6 wt>U3.4 wt>U3.5 wt>U3.6 K7A.
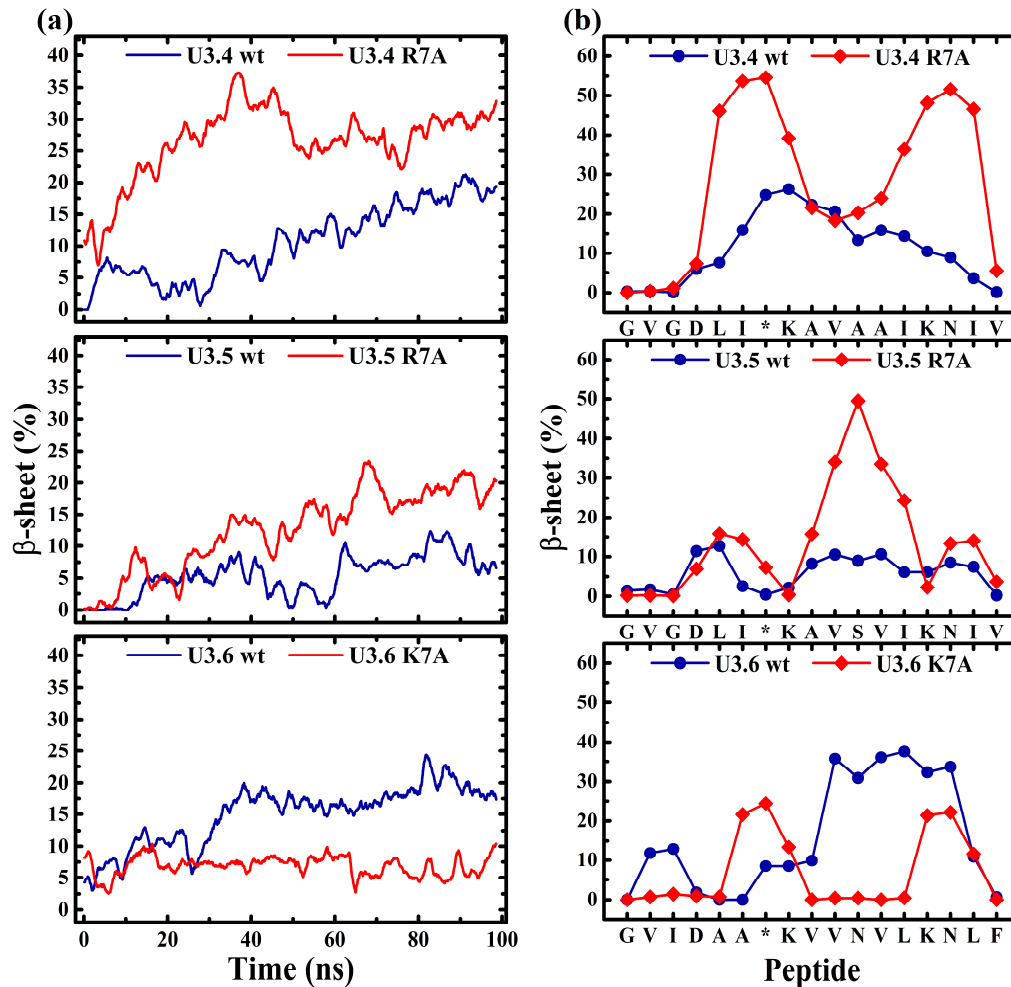
Figure 5.8. (a) The β-sheet content vs time dependence over the entire simulation period, and (b) the time-averaged β-sheet content of amino acid residues at different positions of the peptides within the octamers of uperin 3.x wild-type peptides and corresponding seventh-position variants.

To overcome any potential bias in the initial simulation setup, the β-sheet propensity of all uperin 3.x peptides and variants was calculated after discarding the initial 10 ns of the 100 ns simulation trajectories. A β-sheet content of 23% was observed in the R^7KAV^{10} region of U3.4 wt octamer in Figure 5.8b. Whereas, the L^5IAK^8 and $I^{13}KNI^{16}$ regions of U3.4 R7A octamer have a high β-sheet content of 47% and 48%, respectively. The β-sheet content of the A^9VSVI^{13}

152

region of U3.5 R7A octamer is 22% higher than that observed in the same region of the U3.5 wt octamer. For the U3.6 wt octamer, a high β-sheet content of 34% was observed across the V^{10}NVLKN^{15} region. However, for U3.6 K7A octamer, significant β-sheet content of 23% and 22% was observed in the A^{6}A^7 and K^{14}N^{15} regions, respectively. Once again, this is in contrast to the trends noted for the β-sheet content of the other two wild-type and variant octamer pairs, viz. U3.4 and U3.5, where the octamers of the seventh-position substituted peptides showed higher a β-sheet content in general across the entire peptide sequence. However, there is a good agreement with the observations made in Figures 5.7 and 5.8a.

5.4. Discussion

The different uperin 3.*x* wild-type peptides and corresponding seventh-position variants in Figure 5.1 uncoil at vastly different rates, as can be observed in Figure 5.4. The selected uperin 3.*x* wild-type octamers remained less compact during the simulation than the corresponding seventh-position variants, as can be observed in Figure 5.6. Further, although the U3.5 wt octamer dissociated over the simulation period, but the U3.5 R7A octamer remained stable. Significantly, the effect of the seventh-position alanine substitution is similar to the introduction of salt (NaCl), as greater salt concentration leads to the formation of more stable aggregates[32]. Moreover, the enhanced inter-peptide interaction, observed at the A^{9}VSVI^{13} region of the U3.5 wt peptides at 0.50 M NaCl, is similar to the phenomenon of high β-sheet content observed in that region of the U3.5 R7A octamer, but not the U3.5 wt octamer; in Figure 5.8b. Interestingly, the selected octamers demonstrate a structural transition from predominantly turn and coil to β-sheet secondary structure compositions in Figures 5.7–5.8. The U3.4 wt and U3.5 wt octamers have a lower β-sheet content than the corresponding U3.4 R7A and U3.5 R7A octamers, respectively. More favourable hydrophobic interactions and reduced electrostatic repulsions, due to the cationic to hydrophobic residue substitution at the seventh-position of the peptides in the R7A octamers, may contribute to this difference[7, 10]. The β-sheet content trends for U3.6 wt and U3.6 K7A octamers differ from those observed for the U3.4 and U3.5 wild-type and R7A peptide octamers, as shown in Figures 5.7–5.8. It should be noted that apart from the peptide sequence and structure of the peptide conformer, the rate of β-sheet formation would also depend on the initial Hex docking, i.e., the relative three-dimensional inter-peptide orientation within the dimer, which gets extended to the octamer assembly. This is likely to be critical to the stability of the octamer, as those trajectories with lesser inter-peptide interactions, and unfavourable steric and electrostatic interactions, might be unstable or have slower transitions into β-sheet rich structures. Further, as could be noted

for the selected U3.6 wt octamer in Figure 5.5, structural transformation might be restricted for octamers with compact initial structures. Another factor is that the seventh-position lysine in U3.6 wt has a relatively weaker partial charge distribution, compared to the arginine at the seventh-position of the U3.4 wt and U3.5 wt peptides[33]. This means that the electrostatic repulsion between the adjacent cationic (lysine-rich) regions might be lesser in the U3.6 wt octamer, relative to the U3.4 wt and U3.5 wt octamers. However, from section 5.6.2. of the Appendix, it could be observed that computational analysis indicates towards greater aggregation in the N-terminal region of all uperin 3.*x* R/K7A variants, relative to the corresponding uperin 3.*x* wild-type peptides. Importantly, we have not included the amyloid growth kinetics in our simulations, but instead focused on the structural transition phenomena within a peptide aggregate of a certain size, peptide backbone structure and inter-peptide orientation. Moreover, the secondary structure composition and the transition into β-sheet rich structures would be also influenced by the choice of solvent, ionic concentration, and temperature, in the experiments. Significantly, as noted in section 5.6.1. of the Appendix, some peptide conformers might self-associate into α-helical fibrils, rather than forming unstructured or β-sheet rich aggregates[34].

5.5. Conclusions

The propensity of uperin 3.*x* peptides and corresponding seventh-position variants towards β-aggregation was studied using simulations with preformed peptide aggregates. It was observed that the peptides with turn and coil starting structures can undergo structural rearrangements to generate β-sheet rich structures. However, helix-rich octamers demonstrated negligible to low β-sheet propensity over the 100 ns simulation period. The R7A substitution in U3.4 wt and U3.5 wt peptides resulted in a significant increase in the β-sheet propensity, and even structural stability. However, a similar trend was not observed for K7A substitution of U3.6 wt in the simulations. Therefore, the structural transformations are not only dependent on the peptide sequence, but also defined by the initial peptide structure and inter-peptide orientation. This study provides important insight into the structural diversity of the initial peptide aggregates, and the role of amino acid substitution on the β-aggregation of the antimicrobial peptides.

5.6. Appendix

5.6.1. Docking and propagation

Table 5.A1. Hex docking energies and secondary structure composition of propagating conformers of (a) U3.4 wt, (b) U3.5 wt, (c) U3.6 wt, (d) U3.4 R7A, (e) U3.5 R7A, and (f) U3.6 K7A peptides.

[#]Propagating conformer obtained without OPLS energy minimisation after Hex docking.
[*]Propagating conformer selected for further analyses in Figures 5.6–5.8.

Conformer Family (CF)	Docking energy	Secondary structure composition
(a) U3.4 wt		
10[#]	-290.17	Turn, Coil
10	-240.14	Turn, Coil
13	-258.43	Turn, Coil
17	-255.32	Turn, Coil
20[*]	-329.24	Coil
21	-303.18	Coil
(b) U3.5 wt		
19[#,*]	-307.96	Turn, Coil
11	-294.38	Turn, Coil
19	-229.01	Turn, Coil
(c) U3.6 wt		
08[#]	-359.17	3_{10}-helix, coil
06	-244.91	α-helix, turn, coil
07	-255.04	α-helix, coil
11[*]	-287.41	Isolated β-bridge, coil
17	-311.59	Turn, coil
19	-278.43	3_{10}-helix, turn, coil
(d) U3.4 R7A		
15[#]	-389.84	β-sheet, isolated β-bridge, turn, coil
04	-257.50	α-helix, turn, coil
08	-261.32	α-helix, turn, coil
14[*]	-260.88	Coil
(e) U3.5 R7A		
01[#]	-271.78	α-helix, coil
01	-225.75	α-helix, coil
03	-296.28	α-helix, turn, coil
04	-296.82	α-helix, turn, coil
09	-342.35	Turn, Coil
13	-276.75	α-helix, isolated β-bridge, turn, coil
15	-308.53	Turn, Coil
17	-323.68	α-helix, turn, coil
21[*]	-409.36	Turn, Coil
(f) U3.6 K7A		
22[#]	-467.28	Turn, Coil
21[*]	-257.29	Isolated β-bridge, turn, coil

After self-docking the centroid peptide conformers using the Hex program[31], docking energy values were obtained. It should be noted that these unitless energy values are relative to a reference energy point, where the peptides are at a supposedly infinite separation or zero interaction, and negative scores represent a favourable docking[35]. Although these values are irrelevant on an absolute scale, but they provide a relative measure of docking favourability. After Hex docking, search for propagating conformers was carried out both without and with an intermediate post-processing step of OPLS energy minimisation (EM)[36]. The propagating conformers for each of the six uperin 3.*x* (wt and R/K7A) peptides are listed in Table 5.A1, and it could be noted that post OPLS minimisation, the docked conformers had a greater propensity to extend to propagating conformers, than without OPLS minimisation.

(a) Monomer   **(b) Dimer**   **(c) Octamer**

CF: 10 | EM: No

CF: 10 | EM: Yes

CF: 13 | EM: Yes

CF: 17 | EM: Yes

CF: 21 | EM: Yes

	α-helix		3$_{10}$-helix		β-sheet
	Isolated β-bridge		Turn		Coil
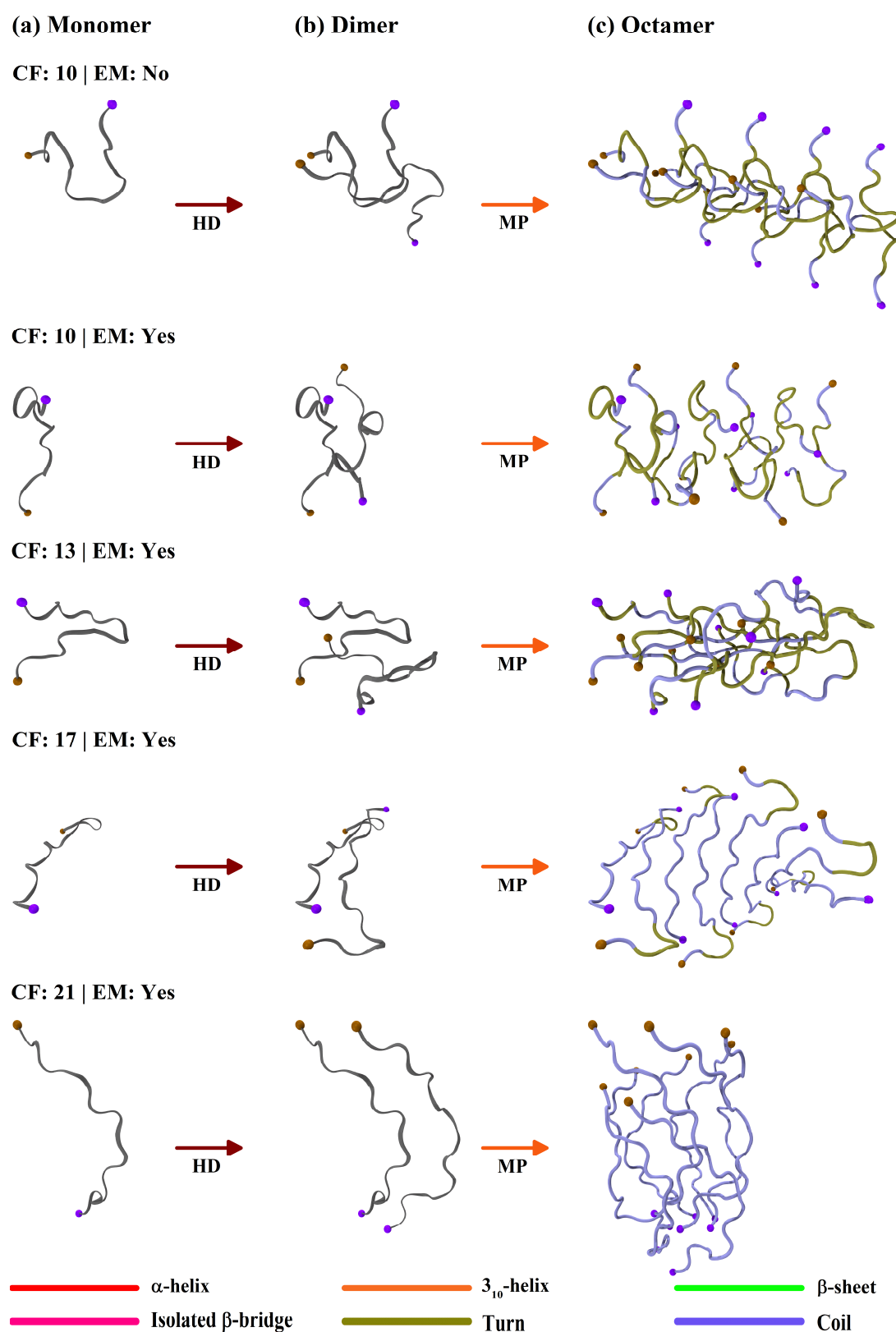
Figure 5.A1. (a) Propagating conformers of U3.4 wt, (b) self-docked using Hex. (c) Transformation matrix used to generate an octamer aggregate.

For the U3.4 wt propagating conformers in Figure 5.A1, the CF: 20 conformer has the lowest Hex docking energy, as noted in Table 5.A1a. Therefore, it was selected for further analysis. HD and MP denote Hex Docking and Matrix Propagation, respectively.

(a) Monomer **(b) Dimer** **(c) Octamer**

CF: 11 | EM: Yes

CF: 19 | EM: Yes

α-helix	3$_{10}$-helix	β-sheet
Isolated β-bridge	Turn	Coil
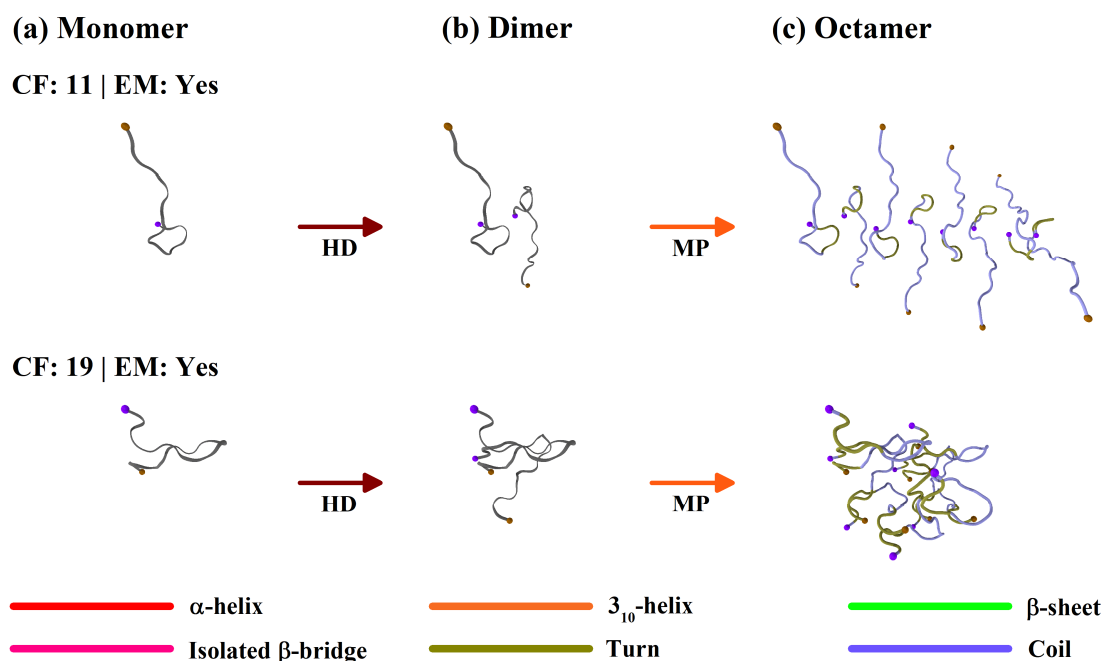
Figure 5.A2. (a) Propagating conformers of U3.5 wt, (b) self-docked using Hex. (c) Transformation matrix used to generate an octamer aggregate.

In case of the U3.5 wt octamers in Figure 5.A2, both CF: 19 octamers obtained without and with OPLS minimisation, were found to dissociate into smaller fragments. Therefore, the CF: 19 octamer obtained without OPLS minimisation was selected, since it has a more favourable docking energy, as observed in Table 5.A1b. CF: 11 octamer was not considered as such extended conformations were found to be unstable (please refer to the discussion for Figure 5.A7).

(a) Monomer　　**(b) Dimer**　　**(c) Octamer**

CF: 08 | EM: No

CF: 06 | EM: Yes

CF: 07 | EM: Yes

CF: 17 | EM: Yes

CF: 19 | EM: Yes

α-helix	3_{10}-helix
Isolated β-bridge	Turn
	β-sheet
	Coil

Figure 5.A3. (a) Propagating conformers of U3.6 wt, (b) self-docked using Hex. (c) Transformation matrix used to generate an octamer aggregate.

Both CF: 08 and CF: 17 octamers of U3.6 wt in Figure 5.A3 were found to disintegrate over the course of respective 100 ns simulations. Therefore, the CF: 11 octamer was selected for further analysis, as it had the next most favourable docking energy, as noted in Table 5.A1c.

(a) Monomer

(b) Dimer

(c) Octamer

CF: 15 | EM: No

CF: 04 | EM: Yes

CF: 08 | EM: Yes

HD

MP

—— α-helix   —— 3_{10}-helix   —— β-sheet

—— Isolated β-bridge   —— Turn   —— Coil

Figure 5.A4. (a) Propagating conformers of U3.4 R7A, (b) self-docked using Hex. (c) Transformation matrix used to generate an octamer aggregate.

For propagating conformers of U3.4 R7A in Figure 5.A4, the CF: 11 octamer has large segments with pre-existing β-sheet content, and hence it possesses an inherent bias to conformations with high β-sheet content. Further, the CF: 04 and CF: 08 octamers have appreciable α-helical content, and therefore these octamers were not considered either (explained in detail for Figure 5.A8). Therefore, CF: 14 octamer was chosen for further analysis.

(a) Monomer **(b) Dimer** **(c) Octamer**

CF: 01 | EM: No

CF: 01 | EM: Yes

CF: 03 | EM: Yes

CF: 04 | EM: Yes

	α-helix		3_{10}-helix		β-sheet
	Isolated β-bridge		Turn		Coil

(a) Monomer **(b) Dimer** **(c) Octamer**

CF: 09 | EM: Yes

CF: 13 | EM: Yes

CF: 15 | EM: Yes

CF: 17 | EM: Yes

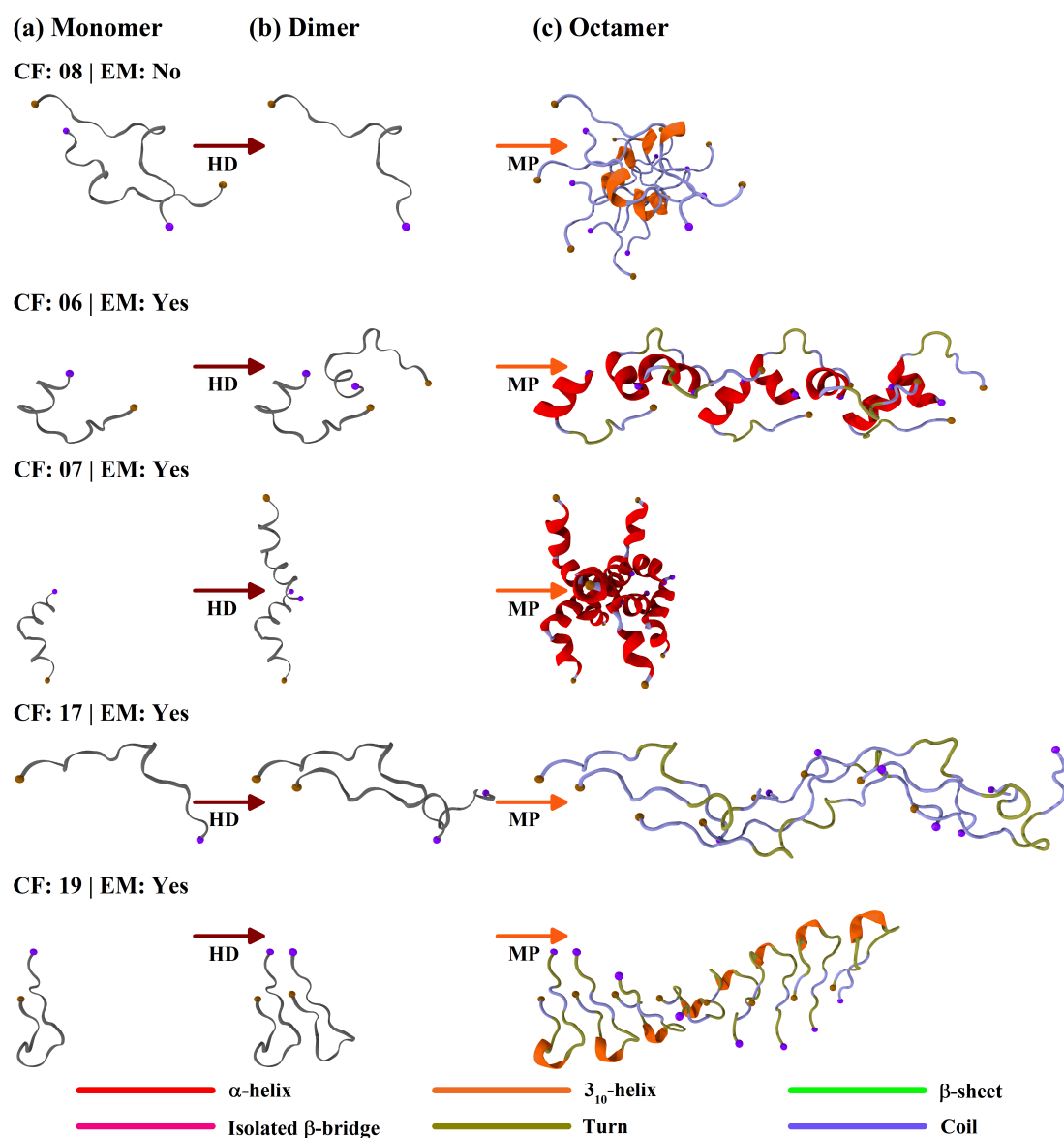|	α-helix		3_{10}-helix		β-sheet
	Isolated β-bridge		Turn		Coil

Figure 5.A5. (a) Propagating conformers of U3.5 R7A, (b) self-docked using Hex. (c) Transformation matrix used to generate an octamer aggregate.

In case of the U3.5 R7A conformers in Figure 5.A5, the CF: 21 one has the most favourable docking energy, as noted in Table 5.A1e. Hence, it was picked for further analysis.

(a) Monomer

CF: 22 | EM: No

(b) Dimer

(c) Octamer

HD   MP

— α-helix
— Isolated β-bridge
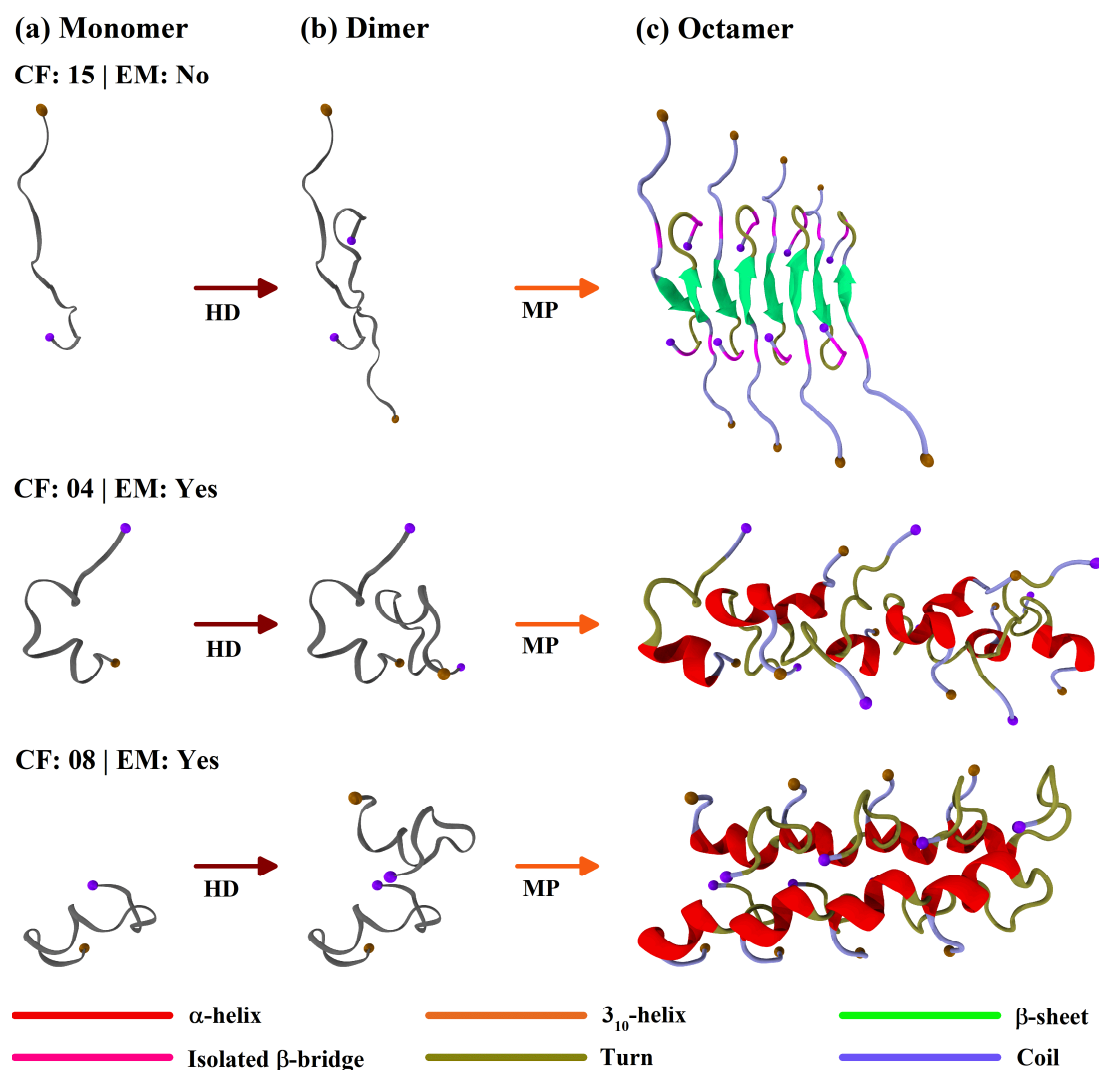— 3₁₀-helix
— Turn
— β-sheet
— Coil
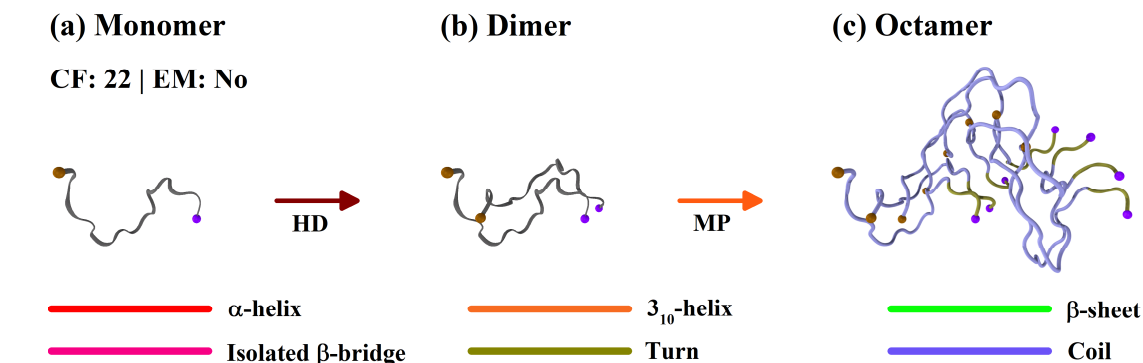
Figure 5.A6. (a) Propagating conformers of U3.6 K7A, (b) self-docked using Hex. (c) Transformation matrix used to generate an octamer aggregate.
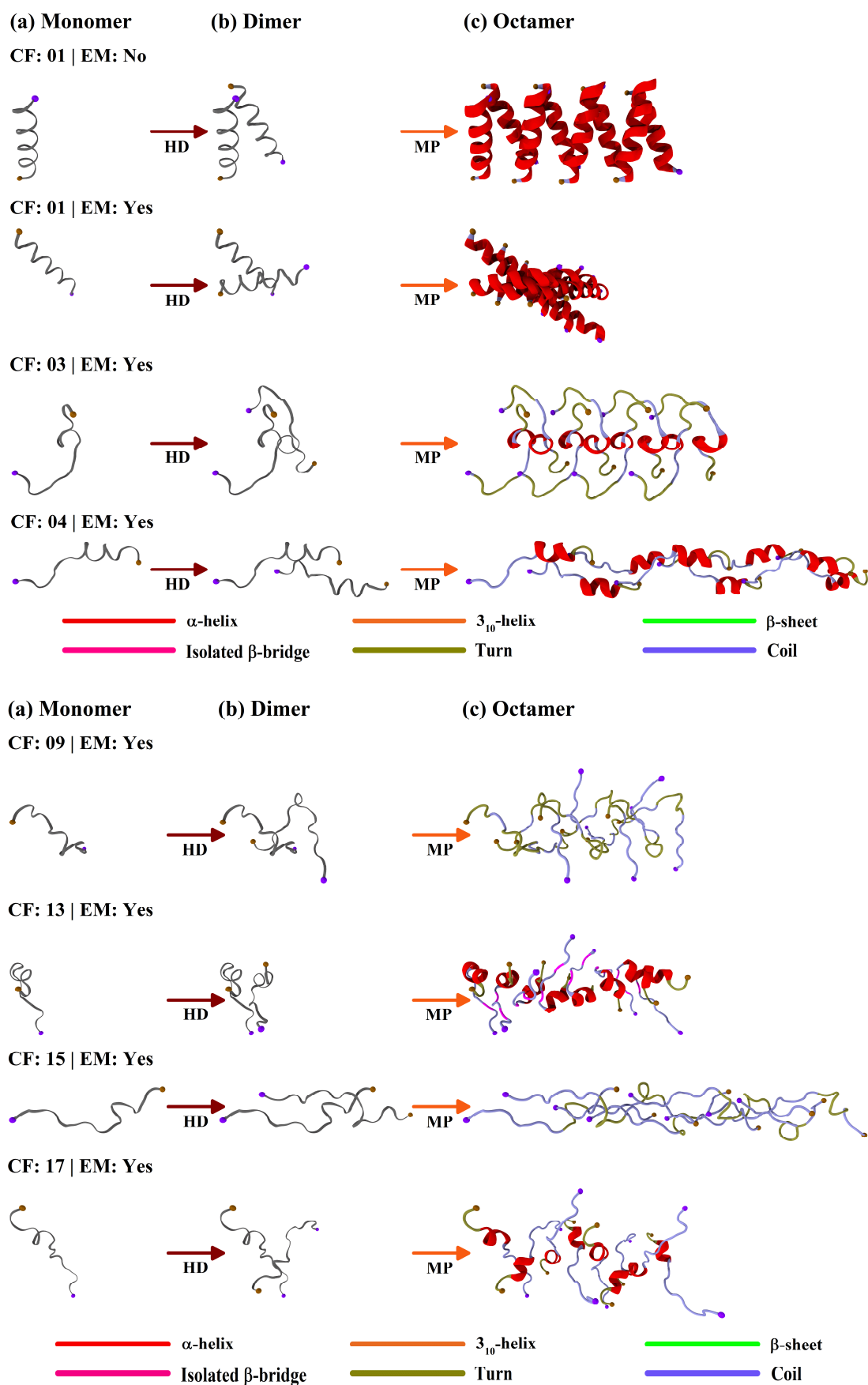
The CF: 22 octamer of U3.6 K7A, in Figure 5.A6, disintegrated over the 100 ns simulation period. Hence, the CF: 21 octamer, which remained stable, was selected.
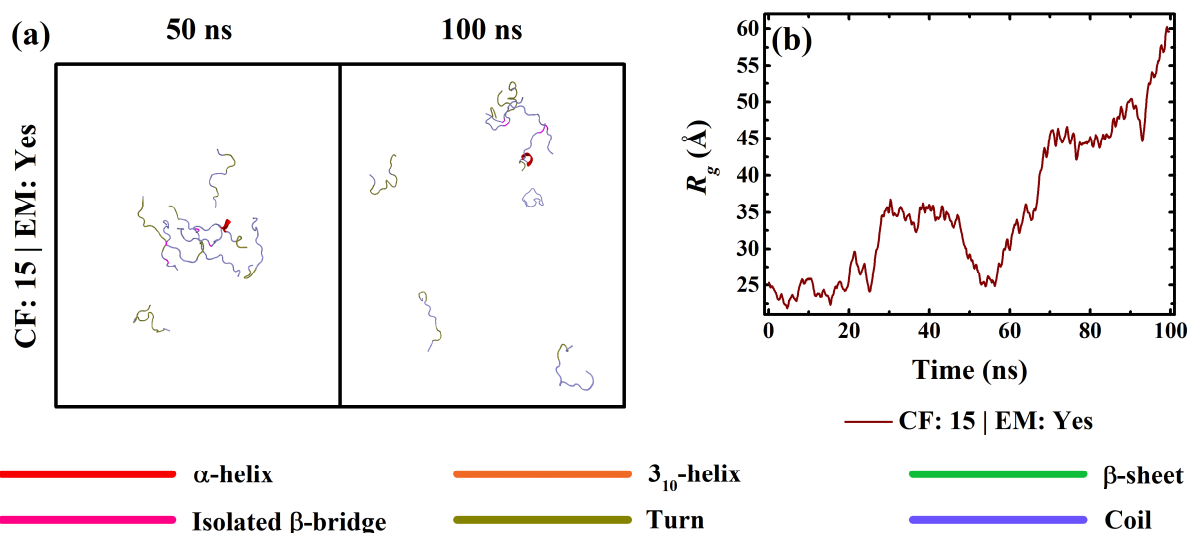
Figure 5.A7. (a) Trajectory snapshots of CF: 15 octamer of U3.5 R7A at the mid (50 ns) and end (100 ns) stages of the simulation, along with (b) R_g plot of that octamer over the entire simulation period.

Extended octamer structures were found to disintegrate over the course of the 100 ns long simulations. For CF: 15 octamer of U3.5 R7A, as could be noted in the 50 and 100 ns trajectory snapshots in Figure 5.A7a, the constituent peptides move further apart as the simulation progresses. At 50 ns, only two peptides are separated from the aggregate. But this increases to four at 100 ns, with the remaining four peptides part of the aggregate. Moreover, R_g of the octamer increases over the course of the simulation and reaches a value of 60 Å at 100 ns; in Figure 5.A7b.
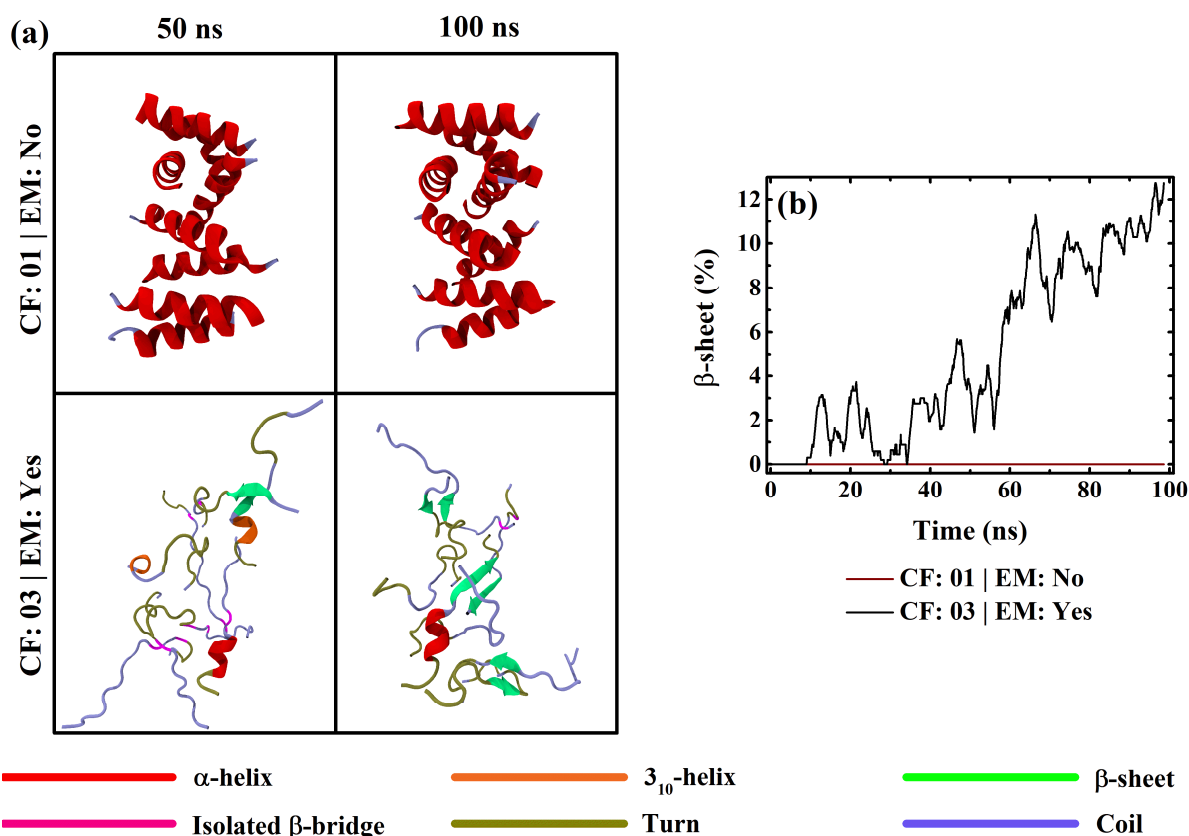
Figure 5.A8. (a) Trajectory snapshots of CF: 01 and CF: 03 octamers of U3.5 R7A (generated without and with OPLS minimisation, respectively) at the mid (50 ns) and end (100 ns) stages of corresponding simulations, along with (b) β-sheet content plot of these octamers over the entire simulation period.

Octamers with noticeable α-helical content were found to demonstrate negligible or slower transition to β-sheet rich secondary structure elements. Peptides in CF: 01 octamer of U3.5 R7A in Figure 5.A5 had a very high α-helical content at the onset. Hence, insignificant changes in the secondary structure composition could be observed in the trajectory snapshots in Figure 5.A8a, and the β-sheet content plot over the simulation period in Figure 5.A8b. Furthermore, even octamers with some α-helical content at the onset were found to display a slower transition to β-sheet structure, than the ones with exclusively turn and coil secondary structure content. This could be clearly noted for the U3.5 R7A octamer of CF: 03 conformers in Figures 5.A5 and 5.A8, and CF: 21 octamer in Figures 5.5cc, 5.7 and 5.8. For the CF: 03 octamer, only two β-strands were noted in Figure 5.A8a at 50 ns, along with two 3$_{10}$-helix segments and one α-helix segment. Also, isolated β-bridge, turn, and coil structural elements can be observed too. Further, six β-strands and one α-helix segment were present at 100 ns, and the overall β-sheet content in Figure 5.A8b reached 12% towards the end of the simulation. Hence, the transition to β-sheet content observed for this particular octamer is significantly slower than that observed

for the CF: 21 octamer of U3.5 R7A. Therefore, as noted previously in Figure 5.5c and Table 5.A1, the selected octamers of all uperin 3.*x* peptides predominantly consist of turn and coil secondary structure elements. However, intermediate conformers with α-helical signatures might play a crucial role in amyloid fibril formation[34].

5.6.2. Intrinsic residue solubility profiles

Greater aggregation in the N-terminal region of uperin 3.*x* R/K7A variants, relative to the uperin 3.*x* wild-type peptides, can be observed in the intrinsic residue solubility profiles in Figure 5.A9, calculated using the CamSol method[7, 37-38]. A score is assigned to each residue of the peptide sequence, with residues having a score greater than 1 being highly soluble, and those with a score lower than -1 being poorly soluble in aqueous environments. Higher solubility indicates lower aggregation propensity, and vice versa. Therefore, cationic to hydrophobic residue substitution results in a clear reduction of the solubility of the uperin 3.*x* R/K7A peptides, relative to the corresponding uperin 3.*x* wild-type peptides.
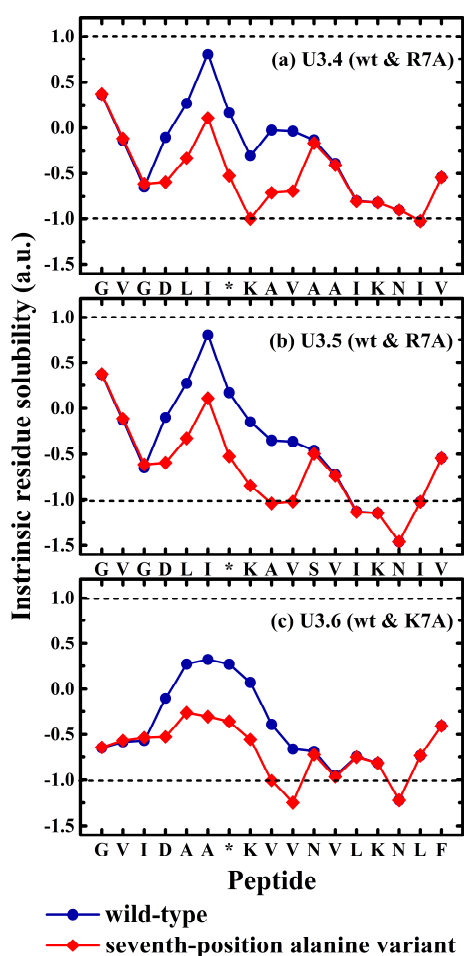
Figure 5.A9. Intrinsic residue solubility profiles of (a) U3.4 wt and U3.4 R7A, (b) U3.5 wt and U3.5 R7A, and (c) U3.6 wt and U3.6 K7A peptides, calculated using the Camsol method[37-38].

5.7. References

1. Chiti, F.; Dobson, C. M., Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333-66.

2. Lansbury, P. T.; Lashuel, H. A., A Century-Old Debate on Protein Aggregation and Neurodegeneration Enters the Clinic. *Nature* **2006**, *443*, 774-9.

3. Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27-68.

4. Zhang, M.; Zhao, J.; Zheng, J., Molecular Understanding of a Potential Functional Link between Antimicrobial and Amyloid Peptides. *Soft Matter* **2014**, *10*, 7425-7451.

5. Bradford, A. M.; Bowie, J. H.; Tyler, M. J.; Wallace, J. C., New Antibiotic Uperin Peptides from the Dorsal Glands of the Australian Toadlet Uperoleia Mjobergii. *Aust. J. Chem.* **1996**, *49*, 1325-1331.

6. Calabrese, A. N.; Liu, Y.; Wang, T.; Musgrave, I. F.; Pukala, T. L.; Tabor, R. F.; Martin, L. L.; Carver, J. A.; Bowie, J. H., The Amyloid Fibril-Forming Properties of the Amphibian Antimicrobial Peptide Uperin 3.5. *ChemBioChem* **2016**, *17*, 239-246.

7. Martin, L. L.; Kubeil, C.; Piantavigna, S.; Tikkoo, T.; Gray, N. P.; John, T.; Calabrese, A. N.; Liu, Y.; Hong, Y.; Hossain, M. A., Amyloid Aggregation and Membrane Activity of the Antimicrobial Peptide Uperin 3.5. *Pept. Sci.* **2018**, *e24052*.

8. Tarus, B.; Straub, J. E.; Thirumalai, D., Structures and Free-Energy Landscapes of the Wild Type and Mutants of the Aβ21–30 Peptide Are Determined by an Interplay between Intrapeptide Electrostatic and Hydrophobic Interactions. *Journal of Molecular Biology* **2008**, *379*, 815-829.

9. Lei, J.; Qi, R.; Wei, G.; Nussinov, R.; Ma, B., Self-Aggregation and Coaggregation of the P53 Core Fragment with Its Aggregation Gatekeeper Variant. *Physical Chemistry Chemical Physics* **2016**, *18*, 8098-8107.

10. Das, A.; Makarov, D. E., Effect of Mutation on an Aggregation-Prone Segment of P53: From Monomer to Dimer to Multimer. *The Journal of Physical Chemistry B* **2016**, *120*, 11665-11673.

11.	Xu, L.; Ma, B.; Nussinov, R.; Thompson, D., Familial Mutations May Switch Conformational Preferences in A-Synuclein Fibrils. *ACS Chemical Neuroscience* **2017**, *8*, 837-849.

12.	Tsigelny, I. F.; Sharikov, Y.; Kouznetsova, V. L.; Greenberg, J. P.; Wrasidlo, W.; Gonzalez, T.; Desplats, P.; Michael, S. E.; Trejo-Morales, M.; Overk, C. R., Structural Diversity of Alzheimer's Disease Amyloid-B Dimers and Their Role in Oligomerization and Fibril Formation. *J. Alzheimer's Dis.* **2014**, *39*, 583-600.

13.	Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **2016**, *14*, 71.

14.	Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

15.	Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005**, *26*, 1781-802.

16.	Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

17.	Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

18.	Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

19.	Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M., Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157-64.

20.	Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.

21. Stone, J. E., An Efficient Library for Parallel Ray Tracing and Animation. *Masters Theses.* http://scholarsmine.mst.edu/masters_theses/1747 **1998**, *1747*.

22. Frishman, D.; Argos, P., Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566-579.

23. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577-2637.

24. Isralewitz, B., Timeline: A Vmd Plugin for Trajectory Analysis. Tutorial. *University of Illinois at Urbana-Champaign c)* http://www.ks.uiuc.edu/Research/vmd **2012**.

25. Rosengren, J., Martin, L. L., Ed. The University of Queensland, 2017.

26. Waterhouse, A., et al., Swiss-Model: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* **2018**, *46*, W296-W303.

27. Chia, B. C. S.; Bowie, J. H.; Carver, J. A.; Mulhern, T. D., The Solution Structure of Uperin 3.6, an Antibiotic Peptide from the Granular Dorsal Glands of the Australian Toadlet, Uperoleia Mjobergii. *The Journal of Peptide Research* **1999**, *54*, 137-145.

28. Johan Strumpfer, P. F., Christophe Chipot, Molefacture: A Tutorial to Build and Edit Molecules. *Centre National de la Recherche Scientifique, University of Illinois, Urbana–Champaign c)* http://www.ks.uiuc.edu/Research/vmd **2012**.

29. Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z., Multiseq: Unifying Sequence and Structure Data for Evolutionary Analysis. *BMC Bioinformatics* **2006**, *7*, 382.

30. OriginLab: Northampton, MA.

31. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.-D.; Ritchie, D. W., Hexserver: An Fft-Based Protein Docking Server Powered by Graphics Processors. *Nucleic Acids Research* **2010**, *38*, W445-W449.

32. Ray, S.; Holden, S.; Martin, L. L.; Panwar, A. S., Mechanistic Insight into the Early Stages of Amyloid Formation Using an Anuran Peptide. *Pept. Sci.* **2019**, *e24120*.

33.   Rice, A.; Wereszczynski, J., Probing the Disparate Effects of Arginine and Lysine Residues on Antimicrobial Peptide/Bilayer Association. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2017**, *1859*, 1941-1950.

34.   Lin, Y.-S.; Pande, V. S., Effects of Familial Mutations on the Monomer Structure of Aβ42. *Biophysical Journal* **2012**, *103*, L47-L49.

35.   Ritchie, D. W., Hex 8.0.0 User Manual. *INRIA Nancy Grand Est, LORIA c)* *https://members.loria.fr/DRitchie/* **2013**.

36.   Ritchie, D. W., Evaluation of Protein Docking Predictions Using Hex 3.1 in Capri Rounds 1 and 2. *Proteins: Structure, Function, and Bioinformatics* **2003**, *52*, 98-106.

37.   Sormanni, P.; Amery, L.; Ekizoglou, S.; Vendruscolo, M.; Popovic, B., Rapid and Accurate in Silico Solubility Screening of a Monoclonal Antibody Library. *Scientific Reports* **2017**, *7*, 8200.

38.   Sormanni, P.; Aprile, F. A.; Vendruscolo, M., The Camsol Method of Rational Design of Protein Mutants with Enhanced Solubility. *Journal of Molecular Biology* **2015**, *427*, 478-490.

Chapter 6

Conclusions and Future Perspectives

6.1. Conclusions

In this thesis, we studied the uperin 3.4–3.6 peptides belonging to the uperin 3.*x* family, along with their corresponding seventh-position alanine variants. Uperin 3.*x* peptides are relatively small amyloidogenic antimicrobial peptides, and therefore ideal for understanding the effect of electrolyte concentration and cationic to hydrophobic residue substitution on the amyloid formation process and the associated secondary structure changes. Furthermore, the interaction of these uperin 3.*x* peptides and variants with bilayer analogues is also investigated. Hence, both fully-atomistic (FA) and coarse-grained (CG) molecular dynamics (MD) simulations were used in this study to examine the aggregation and lipid interaction properties of the uperin 3.*x* wild-type (wt) peptides and their alanine variants.

Both FA and CG simulations were employed to delve into the effect of NaCl salt on the initial stages of self-aggregation of uperin (U) 3.5 wt peptide; and the associated secondary structure transitions. The aggregation propensity of U3.5 wt increases with an increase in NaCl concentration and is primarily driven by interactions along the predominantly hydrophobic AVSVI stretch (residues 9–13) of the peptide. Chloride (Cl$^-$) ions of the NaCl salt screen the positively-charged R7, K8 and K14 residues, encompassing the A^{9}VSVI^{13} region, and therefore reduce the electrostatic repulsion between U3.5 wt peptides, leading to their aggregation. Moreover, peptide aggregation

in the presence of NaCl is marked by a conformational transition to an increased α-helical content.

To elucidate the effect of cationic to hydrophobic residue substitution on the micellar interaction of uperin 3.*x* peptides, FA simulations were performed on the respective peptide-micelle systems. All uperin 3.*x* peptides and seventh-position alanine variants demonstrate a preferential interaction towards the bacterial membrane-mimic, anionic sodium dodecyl sulphate (SDS) micelle, compared to the eukaryotic membrane-mimic, zwitterionic dodecyl phosphatidylcholine (DPC) micelle. Furthermore, alanine substitution at seventh-position significantly diminished the peptide interaction towards both DPC and SDS micelles. While the seventh-position cationic residues were more proximate to the phosphate and sulphate head groups of the DPC and SDS micelles, respectively, the alanine residues preferably interacted with the aliphatic carbon tails of both DPC and SDS micelles. Segments of primarily nonpolar amino acid residues, such as the A^9VSVI^{13} region of U3.5 wt and U3.5 R7A, preferentially interacted with the hydrophobic interior of the micelles, thereby playing a critical role in the peptide-micelle interactions.

Ultimately, we focused upon the conformational diversity, self-aggregation phenomena, and structural evolution of uperin 3.*x* wild-type peptides and variants. In the FA simulations performed, apart from U3.6 K7A, the U3.4 R7A and U3.5 R7A variants demonstrated a higher β-aggregation propensity, when compared to the corresponding wild-type peptide aggregates. Also, the β-sheet content of the A^9VSVI^{13} region of the U3.5 R7A aggregate was found to be substantially higher than that of the U3.5 wt aggregate. Furthermore, apart from coil and turn containing propagating conformers, which demonstrated high β-aggregation propensity, propagating conformers with very high α-helical content were also observed.

6.2. Significant contributions towards thesis objectives

- Aggregation of U3.5 wt is facilitated by NaCl salt due to the screening of charges and secondary structure transitions to α-helical intermediates.

- Uperin 3.*x* peptides and alanine variants demonstrate greater interaction with the anionic SDS micelle, than with the zwitterionic DPC micelle.

- Substitution of a cationic residue with nonpolar alanine diminishes the interaction of uperin 3.*x* wild-type peptides with both DPC and SDS micelles.

- The predominantly nonpolar segments within peptides stabilize the peptide-micelle interaction.

- Peptide conformers with only coil and turn structures have higher β-aggregation propensity than those with helical signatures.

- Alanine substitution at the seventh-position generally increases the β-aggregation propensity of the uperin 3.*x* wild-type peptides, although it would depend on the initial structural content of the peptide conformers and the three-dimensional inter-peptide orientation at the onset.

6.3. Suggestions for future research work

- In our study with U3.5 wt, we observed increased peptide aggregation with increasing NaCl concentration, marked by structural transitions to α-helical intermediates[1]. This study can be extended to analyse the effect of NaCl and other monovalent and divalent salts on aggregation of the other uperin 3.*x* wild-type and variant peptides, and similar amyloidogenic peptides. Furthermore, the interaction of the antimicrobial peptide magainin with palmitoyloleoyl phosphatidylcholine lipid bilayers was observed to be stronger at lower NaCl concentration, with reduced interaction noted at higher NaCl concentration due to the denser packing of the lipids[2]. Likewise, the effect of increased salt concentration on the interaction of the uperin 3.*x* peptides and alanine variants with different bilayers can also be studied using MD simulations.

- The U3.6 wt exhibits greater antimicrobial activity against Gram-positive and Gram-negative bacterial cultures, compared to the U3.6 K7A variant[3]. Therefore, further antimicrobial testing can be performed with the uperin 3.*x* wild-type and variant peptides, and the results could be compared with those obtained from the peptide-micelle simulations in this thesis; to corroborate the findings.

- After obtaining the initial multi-peptide aggregates post 100 ns or longer FA MD simulations[1], replica exchange MD simulations can be performed to obtain

well-equilibrated peptide aggregates[4] of the uperin 3.*x* wild-type peptides and corresponding alanine variants. Then, the interaction of these multi-peptide aggregates with different bilayer models and analogues can be studied with either a combination of, or standalone FA and CG MD simulations[5].

- The search for propagating conformers can be extended to other amyloidogenic peptides and variants, using a similar methodology as used in our studies, or the one used by Tsigelny et al. to find the propagating conformers of $A\beta_{42}$[6]. Furthermore, the stability and structural transitions of the multi-peptide aggregates thus generated can also be analysed using FA MD simulations.

6.4. References

1. Ray, S.; Holden, S.; Martin, L. L.; Panwar, A. S., Mechanistic Insight into the Early Stages of Amyloid Formation Using an Anuran Peptide. *Pept. Sci.* **2019**, *e24120*.

2. Kandasamy, S. K.; Larson, R. G., Effect of Salt on the Interactions of Antimicrobial Peptides with Zwitterionic Lipid Bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2006**, *1758*, 1274-1284.

3. Chia, B. C. S.; Bowie, J. H.; Carver, J. A.; Mulhern, T. D., The Solution Structure of Uperin 3.6, an Antibiotic Peptide from the Granular Dorsal Glands of the Australian Toadlet, Uperoleia Mjobergii. *The Journal of Peptide Research* **1999**, *54*, 137-145.

4. Nymeyer, H.; García, A. E., Simulation of the Folding Equilibrium of A-Helical Peptides: A Comparison of the Generalized Born Approximation with Explicit Solvent. *Proceedings of the National Academy of Sciences* **2003**, *100*, 13934-13939.

5. Gu, R.-X.; Ingólfsson, H. I.; de Vries, A. H.; Marrink, S. J.; Tieleman, D. P., Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2017**, *121*, 3262-3275.

6.   Tsigelny, I. F.; Sharikov, Y.; Kouznetsova, V. L.; Greenberg, J. P.; Wrasidlo, W.; Gonzalez, T.; Desplats, P.; Michael, S. E.; Trejo-Morales, M.; Overk, C. R., Structural Diversity of Alzheimer's Disease Amyloid-B Dimers and Their Role in Oligomerization and Fibril Formation. *J. Alzheimer's Dis.* **2014**, *39*, 583-600.

List of Publications

1. **Ray, S**.; Holden, S.; Martin, L. L.; Panwar, A. S., Mechanistic Insight into the Early Stages of Amyloid Formation Using an Anuran Peptide. *Pept. Sci.* **2019**, *e24120*.

2. **Ray, S**.; Martin, L. L.; Panwar, A. S. Interaction of anuran peptides with DPC and SDS micelles: A Molecular Dynamics study (To be Communicated).

3. **Ray, S**.; Holden, S.; Martin, L. L.; Panwar, A. S. Propensity of Uperin 3.*x* peptides towards beta-aggregation: Analysis of structurally-diverse conformers (To be Communicated).

Acknowledgements

First, I would like to express my acknowledgement and gratitude towards my IIT Bombay supervisor, Prof. Ajay S. Panwar. He introduced me to the field of molecular dynamics simulations and patiently guided me towards the completion of this doctoral study. Then, I would like to thank my Monash University supervisor, Prof. Lisandra L. Martin, for helping with the formulation of the thesis objectives and providing crucial feedback. My sincere gratitude to the research progress committee members: Prof. Arup R. Bhattacharyya, Prof. Ranjith Padinhateeri and Dr. David Turner, for providing their valuable feedback during the annual progress seminars and thoroughly reviewing my annual reports.

Utmost thanks to my colleagues at the Polymer Simulation and Characterization Laboratory, and the Martin Group: Rama Prasad, Raghuram Elancheran, Ajay K. Revelly, Torsten John and Clemens Kubeil, for their valuable insights and suggestions.

Gracious acknowledgements to the undergraduate students: Abhishek Mahato and Stephanie Holden, for generously assisting me.

I want to express my bountiful thanks to the system administrators (high-performance computing resources): Mr. Tarish Bairam (Dendrite, IIT Bombay), Mr. Tilak Raj (Spinode, IIT Bombay), NPSF support team (C-DAC, India), and MASSIVE support team (Monash University), for providing their support towards the successful completion of the simulations.

A lot of thanks to IITB-Monash Research Academy for accepting my doctoral candidature and providing financial support.

Last but not least, I express my heartfelt thanks towards my family and friends for their moral support during my journey of PhD.

Date: 25-July-2019

Place: IIT Bombay   (Sourav Ray)