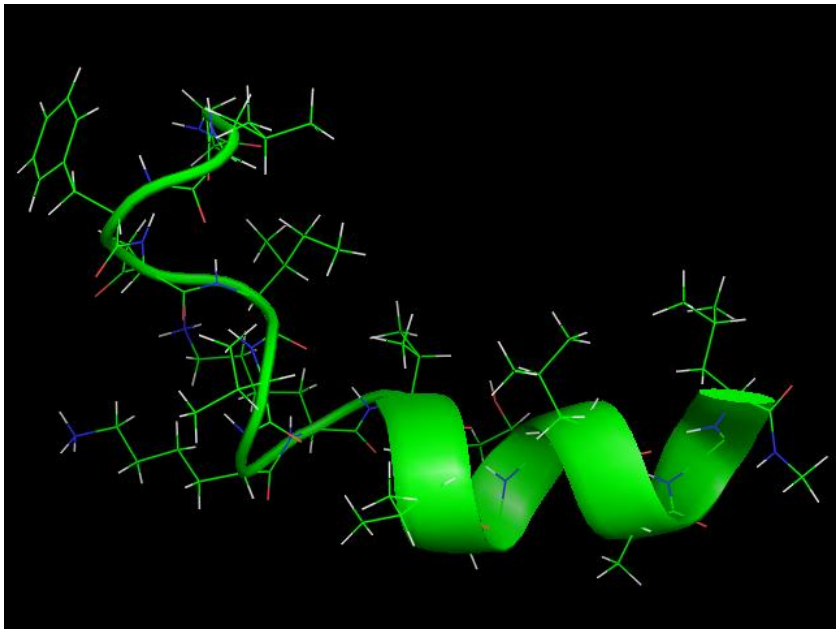
Computational studies of Anti-cancer Aurein Peptides

NEHA MANHAS



DURBAN UNIVERSITY OF TECHNOLOGY

2014

Computational studies of Anti-cancer Aurein Peptides

By

NEHA MANHAS

Submitted in fulfillment of the requirements of the degree of Master of Technology: Chemistry in the Faculty of Applied Sciences at the Durban University of Technology

Declaration

I Neha Manhas declare that the thesis submitted for the degree of Master of Technology (MTech): Chemistry at the Durban University of Technology has not been submitted to any other university and no portion of this or any other closely related work is under consideration for publication elsewhere in any medium. All the work was done by the Author.

_____ _____

Signature of student Date

_____ _____

Signature of promoter Date

_____ _____

Signature of HoD Date

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor K. Bisetty, for his guidance, understanding, patience and most importantly his friendly behavior during my master's project at Durban University of technology. His mentorship has been paramount in providing a well-rounded experience consistent with my long term career goals.

I also want to extend my sincere thanks to Prof G. Redhi (HOD, Chemistry Department) for all the departmental facilities in addition to his humble and co-operative nature.

I also would like to thank my colleagues, Dr Kanchi, Mr Myalo, Mr Khan, and Mr Ayyappa for their support and friendly environment.

I am also grateful to all the teaching and non-teaching staff members of Chemistry Department at Durban University of technology, for motivational chit-chat and great intimacy.

Finally and most importantly, I would like to express my gratitude to my husband Dr. Parvesh Singh for his never ending support, encouragement, patience and unwavering love. Literally, he encouraged me not only to grow as an experimentalist and a chemist but also as computationalist and an independent thinker.

I also want to thank my parents, brother (Honey) and my in-laws for their blessings, ample love and moral support. I would like to thank my source of inspiration my mom (Smt. Kiran manhas), it was under her watchful eye that I gained so much drive and an ability to tackle all challenges. She is the only person who always encouraged me and helped me to combat various hurdles in my life with courage. I feel so blessed to have her as my mother.

Abstract

Peptide folding is a very complicated and dynamic process taking place in all living systems. The understanding of a bioactive conformation of the peptides is very important to understand their biological functions and underlying mechanism of action. However, the high flexible nature of peptides makes this process difficult as they can adopt thousands of conformations within the fraction of a second. The usage of experimental techniques in the characterization process is also limited due to several associated complications including synthesis, isolation and crystallization of peptides. The present computational methodologies, on the other hand, are solid enough to provide detailed complementary information about the intrinsic conformational features of peptides by mimicking their physiological conditions. In the present work, molecular dynamics (MD) computational method was used to explore the configurational space of three Aurein peptides, namely Aurein 2.3, Aurein 2.4 and Aurein 2.5. These peptides are secreted by the amphibian skin when they are exposed to external stimuli. These peptides have been reported to possess anti-cancer and anti-bacterial activity with minimum resistance compared to the available drugs. However, despite their medicinal significance, the precise three dimensional structures of Aurein 2.4 and Aurein 2.5 are not as yet known. First, a validation study was performed on Aurein 2.3 to check the efficiency of the computational protocol. The results obtained revealed the presence of α-helicity in all residues of the Aurein 2.3, in accordance with its experimental structure. A similar protocol was further used to explore the conformational profiles of the remaining two peptides (Aurein 2.4 and Aurein 2.5) under implicit and explicit solvent conditions. The results obtained revealed that both these peptides exhibit α-helical character in all residues although in varying percentages. The α-helical region in the case of Aurein 2.4 was localized predominantly in the central residues extending towards its N-terminal residues, whereas it was flanked by N-terminal and the central residues in Aurein 2.5. However, α-helicity was completely absent in the explicit solvents, and the peptides preferred to stay either in β-turns or extended forms. Hence, the present work provides comprehensive information about the conformational preferences of Aurein peptides which could lead to a better understanding of their native conformations for future investigations and point the way towards developing their new agonists.

List of Contents

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MM:	Molecular mechanics
MD:	Molecular dynamics
3D:	Three-dimensional
NMR:	Nuclear magnetic resonance spectroscopy
SA:	Simulated annealing
PACAP:	Pituitary adenylate-activating polypeptide
VIP:	Vasoactive intestinal peptide
C-AMPs:	Cationic anti-microbial peptides
AA:	Amino acid
PBC:	Periodic boundary condition
PME:	Particle mesh ewald
GB/SA:	Generalized born/surface area
NOEs:	Nuclear overhauser effect intensities
CD:	Circular dichroism
RMSD:	Root mean square deviation
PDB:	Protein data bank
SANDER:	Simulated annealing with NMR-derived energy restraints
PE:	Poisson equation
FF:	Force field
ns:	Nanoseconds
ps:	Picoseconds
REMD:	Replica exchange molecular dynamics
OBC:	Onufriev, Bashford and Case
AMBER:	Assisted model building with energy refinement
HB:	Hydrogen bond
RS:	Representative structure

CHAPTER 1

INTRODUCTION

An overview of molecular modeling and computational chemistry is presented in this chapter. The peptide and protein folding problem is also briefly discussed with the basic notions needed to assess the possibility of solving the protein folding problem. The significance of the peptides selected for this work is described along with the aims and objectives.

1.1 Molecular modeling and computational chemistry

Molecular modeling **[1]** is a technique that encompasses both computational and theoretical techniques methods used to model and manipulate the performance of molecules. Several molecules ranging from small chemical systems to large biomolecules are being studied using molecular modeling techniques. Molecular modeling is also regarded as a field that employs a range of strategies to deduce structural features of a system at atomic level. This discipline includes several methodologies used in computational chemistry such as energy minimization, computation of energy of a molecular system, Monte Carlo (MC) methods and molecular dynamics (MD) **[2]**. Furthermore, the identification of structural functionalities involved in the interaction with receptors/proteins, using different computational methods, permits us to understand the molecular mechanism responsible for its specific biological activity. This information is used in the design of novel and more potent drugs. Due to the fact that simulation accuracy is limited to the precision of the constructed

models, computational simulations have to be compared with experimental results to confirm the accuracy and necessary modifications of the models.

1.2 Protein/Peptide folding

Generally, there is no fundamental difference between peptides and proteins. Both consist of a chain of amino acids linked by peptide bonds, with either a capped or charged end groups, and both may contain disulfide bridges. Nevertheless, the term “protein” is usually reserved for naturally occurring amino acid chains that have a well-defined folding pattern. A protein is referred to as a peptide when it is short, and when it has no well-defined structure or when it is an excised part of a protein, while peptides are regulators of the activity of other molecules (e.g. of proteins). This regulation is achieved by the interaction of the peptide with the target molecule. Some peptides are directly synthesized by the organism, while others are the products of the hydrolysis of proteins. The peptides originated by hydrolysis have a determined function, while others are just further degraded into smaller units. The structure of proteins and peptides in general can be classified into four basic levels (primary, secondary, tertiary and quaternary) depending upon their complexity.

Proteins perform several functions in the biological system depending on their structural features. Some common cellular functions catalyzed by proteins include immune response, catalysis of metabolic reactions and signal transduction under normal and disease conditions **[3-5]**. The proteins adopt one or more specific spatial conformations in order to perform their biological function. The structures of proteins are stabilized by a

number of non-covalent interactions such as hydrogen bonding **[6]**, ionic interactions **[7]**, van der Waals forces **[8]** and hydrophobic packing **[9]**.

A deeper understanding of structural features, chemical thermodynamics and mechanisms of peptide and protein folding, however, remains one of the current challenges in modern molecular biology **[3-4]**. In spite of several contributions from researchers, protein folding is still not adequately understood **[5]**. The emergence of post-genetic diseases due to misfolded proteins, as in the case of Alzheimer's, cystic fibrosis, Parkinson's and Mad cow disease, has generated a great deal of interest in the events happening at the proteomics level **[6-9]**.

During translation from m-RNA, each protein exists as unfolded or random coil conformation. Therefore, in order to perform a biological function, a polypeptide chain assumes its specific folded three-dimensional conformation although certain parts of it may remain unfolded. The intramolecular interactions between amino acid residues are responsible for this well-defined folded structure of a protein, also known as the native state. The protein folding process comprises of three aspects: thermodynamics, kinetics, and structure prediction. Owing to three possible combinations of the phi (ϕ) and psi (ψ) angles, a protein consisting of n amino acids can adopts 3^n conformations. According to the Levinthal's paradox **[10]**, it would take longer than the lifetime of the universe for a protein to randomly search its all available conformations. The Nobel laureate Christiane Anfinsen, reported the relationship between the amino acid sequence and the biologically active conformation **[11]**. Anfinsen demonstrated that the correct refolding of the unfolded Ribonuclease A into its native and enzymatically active

structure occur spontaneously in free solution [**11**]. All information determining the native structure is fully contained in the amino acid sequence of a protein [**12**].

Several experimental and theoretical studies have been performed to explain the folding and stability of proteins. Specifically the linear polypeptide chain is folded into a stable structure in a very short time. Thus, if a residue has five different conformations, then it requires 5^{100} conformational changes for a protein consisting of 100 residues to transform into its native form. As the conversion of an extended polypeptide into its folded structure require very short time, it is almost impossible for a protein to adopt many conformational changes within a short time. Based on this observation, Levinthal proposed the framework model theory stating that protein folding is a kinetically controlled process that proceeds through different stages instead of following all the conformational changes [**13, 14**]. Folding intermediates possess stabilized structural elements, mainly of secondary structural origin, in combination with the unstructured regions. In 1976, Karplus and Weaver proposed the diffusion collision model to explain the folding process of proteins. According to this model the initially formed stable secondary structural elements collapses into tertiary structures by diffusion and collision with other secondary structures [**15-16**]. The second referred to as the hydrophobic collapse model [**17-18**], is based on a rapid collapse of the hydrophobic polypeptide chain, upon which folding can proceed with significantly less possibilities for the formation of trapped intermediate folding states. Ptitsyn proposed a three step mechanism for the protein folding based on his experimental observations [**13**]. First step is very rapid (~0.01 seconds) and leads to the formation of molten globule in second step (~1 second). The last step of the protein folding, on the other hand, is very

slow (~10-40 minutes). The model proposed by Jackson and co-workers [**19**] suggests that protein folds without forming detectable intermediate structures or that they can form secondary and tertiary structures in parallel during the hydrophobic collapse [**20**]. These observations lead to the proposal of the nucleation condensation mechanism [**21**], which combine features from both the framework mechanism and the hydrophobic collapse model. The widely accepted theory of protein folding was proposed by Wolynes and co-workers in which they proposed the funnel shaped energy landscape as shown in Figure 1.1. The depth of the funnel represents enthalpy and the width represents the entropy. They suggested that the polypeptide navigates its different possible conformations through this complicated energy landscape or a folding funnel with a vast array of down-hill routes to the native state in a relatively rugged surface [**22-25**].

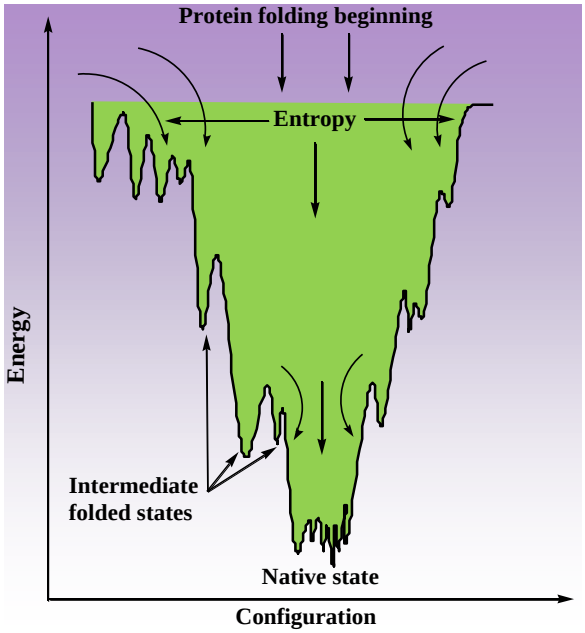


Figure 1.1 Funnel shaped energy landscape model for protein folding process.

The grooves in the funnel wall indicate the different local energy minima in which the proteins may get trapped during the folding process [**26**]. Typically the native state of a protein can be described thermodynamically as the free energy minimum of all possible structures populated according to the Boltzmann distribution. It is important to remember that the statistical weights of the different structures also involve entropic contributions.

Approximately, 52000 proteins with their amino acid sequences are currently available in the protein data base (www.rcsb.com). Nevertheless, the number of proteins with known three-dimensional protein structures is significantly low (~2000). The determination of protein structure using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy requires a considerable amount of work. Quite often, it is not possible to generate good protein crystals, because the process of protein crystallization is poorly understood and therefore, good conditions for protein crystallization cannot be predicted [**27**]. The presence of higher number of atomic nuclei and poor solubility of proteins in solvents create complications in their characterization process using NMR techniques. Protein threading is a method of protein modeling which is used to model those proteins which have the same fold as proteins of known structures, but do not have homologous proteins with known structures [**28-29**]. However, this method still cannot reliably predict the three-dimensional structure of a protein from an amino acid sequence.

Peptides offer a unique opportunity to bridge the gap between theoretical and experimental understanding of protein folding. Nowadays, the computational

methodologies used to investigate the propensities of a peptide to adopt different conformations are solid enough to provide detailed complementary information about their intrinsic conformational features.The utilization of different computational molecular dynamics (MD) simulations, for instance, has played a noteworthy role to describe the folding-unfolding pathways of proteins and peptides, and is widely explored in literature [**30**]. In these techniques, the conformational profile of a system is normally assessed either through a topographical exploration of the potential energy surface using methods like simulated annealing (SA) [**31**], or in the configurational space using Monte Carlo (MC) or MD simulation [**2**]. Particularly, MD simulations provide detailed information on conformational changes and fluctuations of proteins and nucleic acids, and have been extensively explored to investigate the structure, dynamics and thermodynamics of biomolecules [**32-35**].

Similarly, Generalized Born/surface area (GB/SA) implicit solvent models have recently been used to study protein–protein docking [**36**], the NMR supported structures [**37**] and the MD simulations [**38-39**] to sample the configuratonal space of proteins, peptides and nucleic acids. These models were found to be quite successful in the context of biological membranes described by multiple layers with different dielectric constants [**40-42**]. As the alcoholic solvents (TFE, Methanol and HFIP) are known to stabilize the helical contents in peptides and proteins [**43-44**], recent modifications to the standard GB implementations extend its applicability to a wider range of dielectric environments and thus play an imperative role to reproduce the environment induced by different explicit solvents [**45-46**].

Newly discovered peptides are becoming increasingly important, not only as molecular tools for the understanding several protein-protein interactions, but also as lead scaffolds for targeting different types of cancers. Over the past few years, cationic anti-microbial peptides (c-AMPs) have received considerable attention due to their potential activities against several microbes and cancerous cells. Amphibians are an important source of this category of peptides. Brevinins, aureins, magainins and maculatins are some common AMPs which are normally secreted by amphibian skin along with other defensive chemicals when they are exposed to an unfavorable environment [**47**]. The high efficiency and minimum resistance have made these peptides even more excellent targets for the development of novel anti-cancer and anti-microbial agents. However, in spite of having great medicinal benefits, the mechanism of action of these peptides is not adequately known. Over the past few years, a number of mechanisms showing the mode of action of AMPs were reported by researchers worldwide. However, the mechanism that is widely accepted for their biological action involves the increased plasma membrane permeability after their interactions, resulting in death of cells by leakage of the cytoplasmic components [**48-51**]. It is believed that electrostatic interactions between the positively charged peptide and the negatively charged membrane are the primary driving force for their binding propensities. Several studies suggesting an increase in anti-microbial action of these peptides with an increase in their positivity have been reported in the literature [**52**]. Apart from the positive charge, other factors of importance for anti-microbial activity of c-AMPs includes secondary structures in membrane, net charge, hydrophobicity, oligomerization ability and their amphipathic nature that segregate hydrophobic and basic residues [**53-57**].

Recently researchers have focused their attention on the cationic peptides which are potent anticancer agents in addition to their established roles as antimicrobial factors and modulators of innate immune systems [**58-60**]. Generally, these peptides have been found to exhibit anticancer activity at low micromolar levels with lesser haemolysis or toxicity to the mammalian cells, and thus possess greater potential for the design of novel and potent anti-cancer agents [**61**]. Aurein peptides belong to the category of cationic peptides that display a wide range of sequence diversity and activity with respect to the gram-positive bacteria and cancer. These peptides are predominantly isolated for the amphibian skin. A quick survey in the antimicrobial peptide database revealed that almost 400 peptides have been isolated from frogs [**62**]. Approximately 27 aurein peptides have been isolated only from the secretions of the granular dorsal glands of the Green and Golden Frog *Litoria aurea* and Southern Bell frog *L. raniformis* [**63**]. Of these peptides, 13 peptides have shown wide-spectrum antibiotic and anti-cancer activity.

Some aurein peptides such as Aurein 1.2, Aurein 2.2 and Aurein 2.3 are reported to be highly effective against bacterial infections and cancer [**64**]. The activities of these peptides are significantly affected by the nature of amino acid residues present in their structure. For instance, Aurein 2.3 gains anti-cancer activity when its isoleucine amino acid (AA) residue at position 13 is replaced with phenylalanine [**63**]. Aurein 2.2, on the other hand, loses its anti-bacterial activity [**63**] significantly when its leucine AA residue at position 13 is substituted with the isoleucine AA residue. Literature survey revealed [**47**,**65-67**] that several experimental studies have focused on the characterization of three dimensional (3D) structures of Aurein 1.2, Aurein 2.2 and Aurein 2.3 peptides [**65**].

However, despite being structural analogues of Aurein 2.3 and possessing similar biological activities, the precise characterization of 3D structures of Aurein 2.4 and Aurein 2.5 has not been investigated.

Although, several biophysical techniques are being used nowadays to investigate the interactions between different anti-microbial peptides and their biological receptors, the determination of the solution structure of the peptides remains one of the most significant pieces of information. The information collected from these studies is quite useful to identify key residues in the peptides responsible for their biological activity, and can be modified/mutated to alter their actions. Nowadays, circular dichroism (CD), nuclear magnetic resonance (NMR) and X-ray techniques are commonly used in the characterization of several proteins and peptides. However, tedious synthesis and crystallization protocols of proteins/peptides are few limitations of these techniques. Moreover, the greater flexibility of smaller peptides in solution phase also makes these spectroscopic techniques inefficient to provide all the information adequate for full characterization of their conformational preferences. The available advanced computational methods, on the other hand, have been quite successful to overcome these problems, and have been extensively used in parallel with experimental techniques to study the conformational preferences of a range of peptides and proteins.

Although obtaining the whole folding mechanism of proteins through MD has remained elusive until now [**68**], folding studies of peptides through MD are within the reach of computational power currently available. The reversible folding of peptides through MD has been described in past years [**69-70**]. Previous studies undertaken in our research

group [**71**], involved a series of MD simulations applied to small, medium-size polypeptides and neuropeptides to assess their conformational preferences. The validation of computational procedures was first performed on the 10-residue long chignolin-like synthetic peptide (CLN025) using the replica exchange molecular dynamics (REMD) sampling method [**72**]. Different case studies were subsequently performed on different peptides such as polypeptide 27 (PACAP27), the 28-residue vasoactive intestinal peptide (VIP) and a few neuropeptides (bombesin, neuromedin B and neuromedin C) to assess their folding characteristics [**73**]. One of the findings of these investigations was that implicit solvent models using force field ff94 does not reproduce the spectroscopically characterized structures of the peptides. The longer molecular dynamics (MD) simulations in combination with GB-OBC implicit solvent model and modified amber ff99 (ff99SB), on the other hand, were quite successful to reproduce the peptide secondary structures, in accordance with their experimentally characterized structures [**71-73**]. Accordingly in this study similar MD protocols were employed to fully explore the configurational space of Aurein peptides. Specifically, Aurein 2.3, Aurein 2.4 and Aurein 2.5 were investigated aimed at understanding their folding characteristics. Specifically, the molecular dynamics (MD) sampling method has been used to explore the conformational space of these peptides in conjunction with the Generalized Born/surface area (GB/SA) procedure [**74**] to treat the solvent implicitly. The primary reasons for using GB/SA solvent in the present study model includes reproducibility of the explicit solvent effects, faster tendency to sample the free energy of space of peptide and requirement of minimum computational power.

The present work has been organized in five chapters. The first two chapters comprise the introduction addressing the current knowledge on the general aspects of peptide/protein folding, brief summary of the different simulation techniques, followed by an overview of the experimental techniques currently used in the validation of the computational results. The main computational section of the thesis is organized in the next two chapters, and contains discussions on the computational results obtained for Aurein 2.3, Aurein 2.4 and Aurein 2.5. The overall conclusions and recommendations are concisely discussed in the last chapter.

1.3 Aims and Objectives

The aims of this study were to:

- use MD sampling technique to explore the conformational space of three Aurein peptides.
- predict the 3D structural features of Aurein 2.4 and Aurein 2.5 under implicit and explicit solvent conditions

The objectives of this study were to:

- validate the performance of the MD protocol to assess the features of Aurein 2.3 for which experimental structure is known.
- explore the conformational space of Aurein 2.4 and Aurein 2.5 using similar MD protocol aimed at understanding their preferred 3D-strucutres responsible for their biological actions both in implicit and explicit solvent models.

CHAPTER 2

COMPUTATIONAL TOOLS AND PROCEDURES

An overview of the molecular mechanics method used to predict the structures and properties of peptides and proteins is discussed in this chapter. A brief introduction to different computational procedures such as structure minimization, effect of charge and solvent, employment of periodic boundary conditions and the particle mesh ewald method normally used in the MD simulations are also presented. The experimental techniques including nuclear magnetic resonance (NMR), circular dichroism (CD) and X-ray crystallography are also presented. Finally, the computational chemistry software tools *viz.,* AMBER and CLASICO computer program used primarily in the present study are also presented.

2.1 Molecular Mechanics

Molecular mechanics (MM) exercise the laws of classical physics to predict detailed structure and physical properties of molecules. These properties include entropies, enthalpy of formation, strain energies, dipole moments and enthalpies. In the MM, a molecule is considered to be a collection of masses interacting with each other through harmonic forces. The atoms in molecules are, therefore, treated as balls of different sizes joined together by springs of different strength and equilibrium distances (bonds). Since, the electrons and protons of the atoms are not explicitly included in the MM calculations, this simplification allows for using MM as a fast computational model that

can be applied to molecules of any size. The total energy of a system is minimized with respect to the atomic coordinates, and represents a sum of different contributions that calculate the deviations from equilibrium values of bond lengths, angles and torsions plus non-bonded interactions [**75-77]**:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} \quad \text{..... (1)}$$

Where E_{tot} is the total energy of the molecule, E_{str} is the bond-stretching energy term, E_{bend} is the angle-bending energy term, E_{tors} is the torsional energy term, E_{vdw} is the van der Waals energy term, and E_{elec} is the electrostatic energy term.

The first term in equation (1) describes the energy change as a bond stretches and contracts from its ideal unstrained length. It is assumed that the interatomic forces are harmonic so the bond-stretching energy term can be described by the following the simple quadratic equation:

$$E_{str} = \frac{1}{2} k_b (b - b_0)^2 \quad \text{..... (2)}$$

Where k_b is the bond-stretching force constant, b_0 is the unstrained bond length and b is the actual bond length.

The van der Waals interactions between atoms that are not directly connected are usually represented by a Lennard-Jones potential:

$$E_{vdW} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \quad \text{.....(3)}$$

Where A*ij* is the repulsive term coefficient. The term B*ij* is the attractive term coefficient and r_{ij} is the distance between atoms *i* and *j*.

To describe the electrostatic forces an additional term with a Coulomb's interaction is used:

$$E_{elec} = \frac{1}{\varepsilon}\frac{Q_1 Q_2}{r_{ij}} \quad \text{..... (4)}$$

Where ε is the dielectric constant, and Q_1 and Q_2 are the atomic charges of interacting atoms and r_{ij} is the interatomic distance. Modest level quantum mechanical methods are adequate in accuracy and efficiency for describing intramolecular energy surfaces. Thus, the equilibrium values of bond lengths and bond angles are derived from the force constants used in the potential energy function defined in the force field, and define a set known as force field parameters **[75-76]**. Deviations from any of the equilibrium values inevitably result in increasing total energy of the molecule. As a result, the total energy of the system is taken as a measure of intramolecular strain relative to a hypothetical molecule with equilibrium values. On its own, the total energy has no strict physical meaning. However, differences in total energy between two different conformations of the same molecule can be compared **[75, 78-79]**.

2.2 Molecular Dynamics (MD)

Molecular dynamics (MD) is a computer simulation method in which the atoms and molecules interact periodically by approximations of known physical attributes, resulting in the simulation of the motion **[80]** for a system of particles **[81]**. Since, analytically it is

impossible to find the properties of such complex systems, MD simulation overcomes this problem by using numerical methods. It was originally conceived within theoretical physics in the late 1950s, but is applied today mostly in materials science and in the study of complex, dynamic processes that occur in biological systems, including protein folding, molecular recognition, etc. **[81-83]**. A graphical representation for the general MD procedure is depicted in Figure 2.1

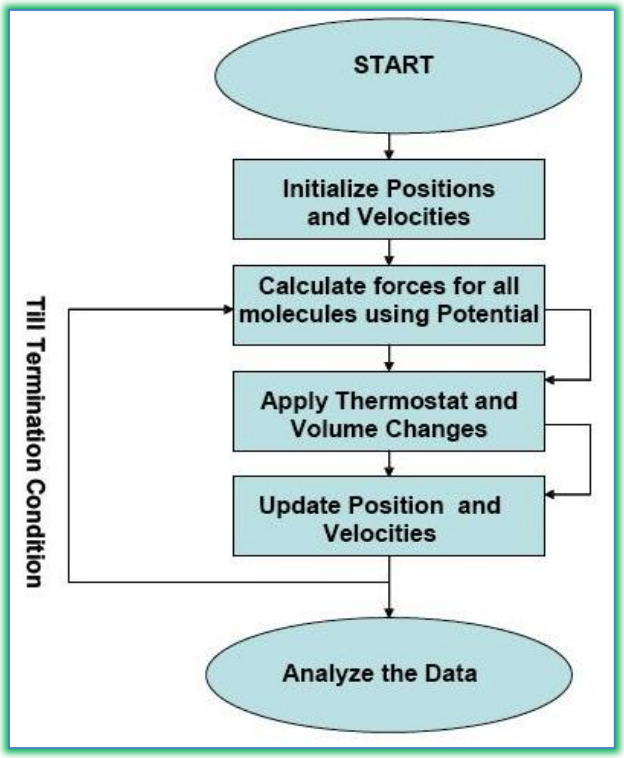


Figure 2.1 Flowchart showing the general molecular dynamics protocol.

It is assumed that the atoms in the molecule interact with each other according to the rules of the employed force field. Successive configurations of the system are generated by integrating Newton's Laws of motion. The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time.

This is done by first determining the force on each particle (F_i) as a function of time, which is equal to the negative gradient of the potential energy:

$$F_i = -\frac{\partial U}{\partial r_i}$$

where U is the potential energy function and r is the position of a particle. The acceleration, a_i of each particle can then be determined by dividing the force acting on it by the mass of the particle:

$$a_i = \frac{F_i}{m_i}$$

The change in velocities is equal to the integral of acceleration over time and the change in position is equal to the integral of velocity over time:

$$dv = \int a dt,$$
$$dr = \int v dt$$

Finally, the kinetic energy can be defined in terms of both the velocities and momentum of the particles:

$$K(v) = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2,$$

$$K(p) = \frac{1}{2}\sum_{i=1}^{N} \frac{p_i^2}{m_i}$$

The total energy of the system, called the Hamiltonian, is the sum of the kinetic and potential energies:

$$H(q,p) = K(p) + U(q)$$

where q is the set of cartesian coordinates, p is the momenta of the particles and U(q) represents the potential energy function. The velocities, $v_i(t)$ are the first derivative of the positions with respect to time:

$$v_i(t) = \frac{d}{dt} q_i(t)$$

where $q_i(t)$ refers to the atomic positions at a particular time, t. Based on the initial atom coordinates of the system, new positions and velocities of the atoms can be calculated at time t and the atoms will be moved to these new positions. As a result of this a new conformation is created. The cycle will then be repeated for a predefined number of steps. The collection of energetically accessible conformations produced by this procedure is called an ensemble.

As one of its many and varied applications, MD involves a study of the dynamics of large macromolecules, including biological systems such as proteins, nucleic acids (DNA, RNA), and biological membranes. Dynamical events play a key role in the controlling processes which affect the functional properties of biomolecules. Drug design is commonly used in the pharmaceutical industry to test the properties of molecules prior to the experimental work **[84]**.

2.3 Periodic Boundary Conditions (PBC)

A more realistic approach in simulations is to use the solvent explicitly, which is done by soaking the molecule in a box of solvent molecules, thus requiring additional

computational effort. Periodic Boundary Conditions (PBC) are normally employed to model the bulk solvent. In the case of an infinite PBC, the simulation box is infinitely replicated in all directions to form a lattice. In practice, most MD simulations evaluate potentials using some cutoff scheme for the computational efficiency. In these cutoff schemes, each particle interacts with the nearest images of the other n-1 particles (minimum-image convention). The use of cutoff methods, however, has been shown to introduce significant errors and artificial behaviour in simulations **[85]**.

2.3.1 Ewald Summation Techniques

In most MD simulations, the long-range Coulombic interactions are the most time consuming. Ewald summation was introduced in 1921 **[86]** as a technique to sum the long-range interactions between infinite particles and all their infinite periodic images efficiently. Long-range interactions are evaluated as sums that converge very slowly. The principle of obtaining the Ewald sum is by the conversion of the summation of the potential energy into two series as shown in Figure 2.2 **[87]**. A Gaussian charge distribution is commonly used. The sum over the point charges is converted to a sum of the interactions between the charges plus the neutralizing distributions according to the equation below:

$$U_{Ewald} = U^{r} + U^{m} + U^{0}$$

This part is the real space sum U^{r}. A second charge distribution is added to the system which exactly counteracts the first neutralizing distribution. This summation is performed in the reciprocal space and is termed U^{m}. The dipole term U^{0} includes the effect of the

total dipole moment of the unit cell, the shape of the macroscopic lattice, and the dielectric constant of the surrounding medium **[85]**.

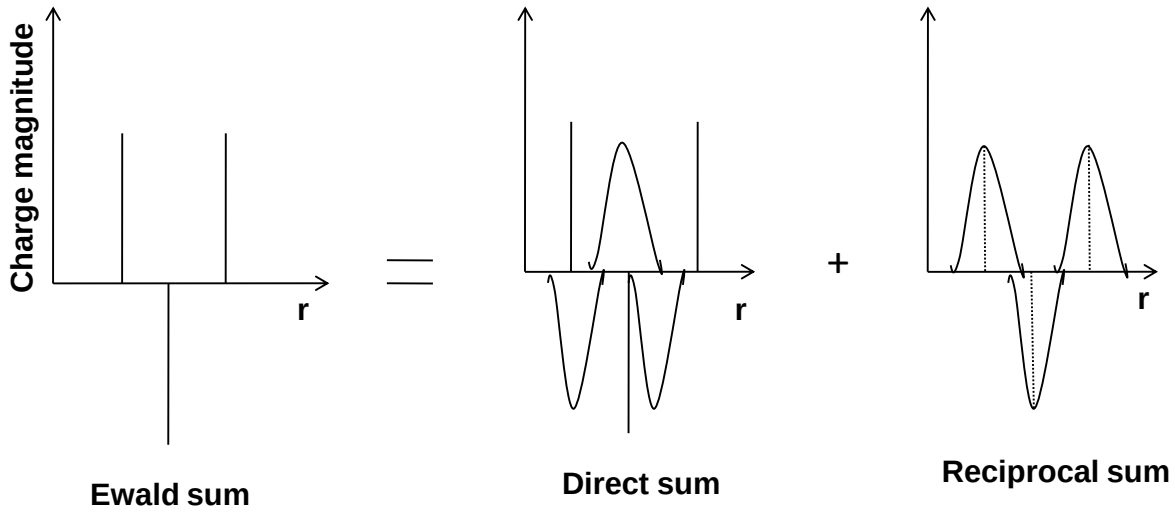


Figure 2.2 Splitting of charges into discrete and smeared distributions in the real and reciprocal space.

2.3.2 Particle-Mesh Ewald (PME)

The Particle-Mesh Ewald method (PME) divides the potential energy into Ewald's standard direct and reciprocal sums and uses the conventional Gaussian charge distributions **[88]**. The direct sum is evaluated explicitly using cutoffs while the reciprocal sum is approximated using fast Fourier Transform (FFT) with convolutions on a grid where charges are interpolated in the grid points (Figure 2.3). Furthermore, PME does not interpolate but rather evaluates the forces by analytically differentiating the energies, thus reducing memory requirements substantially **[85]**.

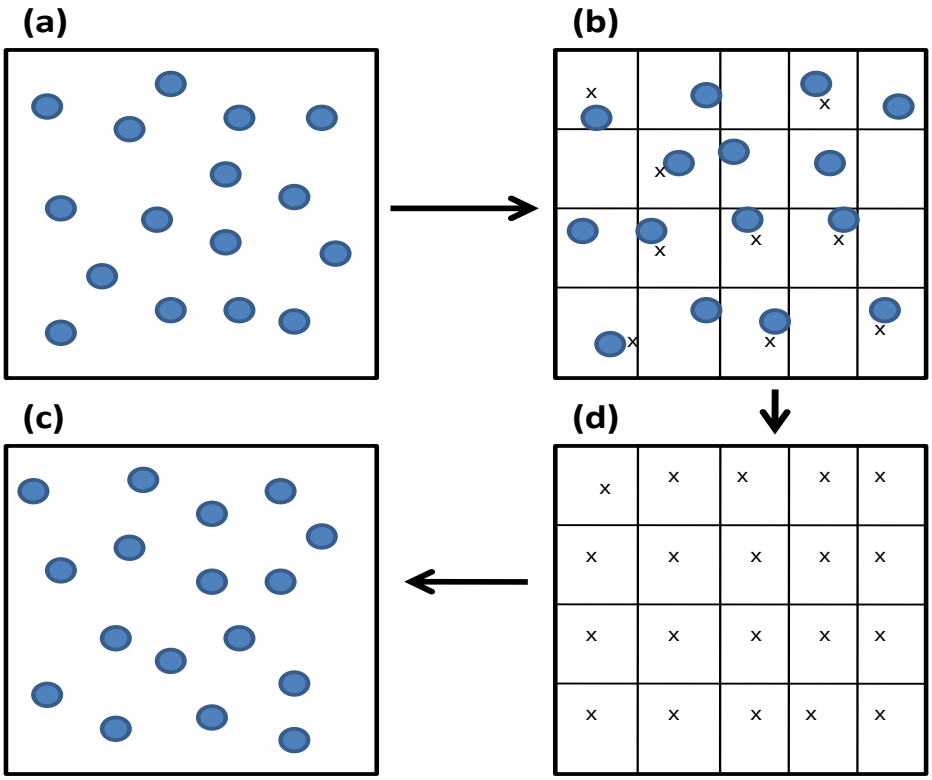


Figure 2.3 A 2D schematic of particle-mesh technique used in most Fourier-based methods **(a)** A system of charged particles. **(b)** The charges are interpolated on a 2D grid. **(c)** Using FFT, the potential and forces are calculated at grid points. **(d)** Interpolate forces back to particles and update coordinates.

2.4 Thermostats in MD

Numerous thermostat methods are available to add and remove energy from the boundaries of an MD system in a more or less realistic way, approximating the canonical ensemble. In the canonical ensemble, the number of particles (N), the volume (V) and the temperature (T) are conserved. In NVT, the energy of endothermic and exothermic processes is exchanged with a thermostat. However, since the temperature is defined by the ensemble average kinetic energies of all particles as Eq. (1), it is impossible to exactly fix *T* at a set point.

$$\left\langle \frac{1}{2} m v^2 \right\rangle = \frac{k_B T}{2} \text{ .....(1)}$$

Therefore, various types of thermostats, such as Berendsen, Langevin and Nosé-Hoover thermostats, have been proposed to control the particle motions. A Berendsen thermostat is a proportional type of thermostat, and corrects deviations of T from the set point T_0 by multiplying the velocities by a factor to control the value of T [**89**]. Nosé-Hoover thermostat is an integral type of thermostat, and introduces additional degrees of freedom (momentum) into the Hamiltonian of a system [**90-92**]. The Langevin thermostats follow the Langevin equation of motion instead of Newton's equation of motion [**93**]. In the Langevin equation, a frictional force added to the conservative force is proportional to the velocity, and it adjusts the kinetic energy of the particle so that the temperature matches the set temperature shown in Eq. (2).

$$ma = -\xi v + f(r) + f' \text{ .....(2)}$$

where m is mass of a particle, a is acceleration, $f(r)$ is conservative force acting on the particle, v is the velocity of the particle, ξ is a frictional constant, and f' is a random force. The frictional force -ξv decreases the temperature because ξ is a fixed positive value.

The force is randomly determined from a Gaussian distribution to add kinetic energy to the particle, and its variance is the function of set temperatures and time steps. The balance of random force with frictional force therefore maintains the system temperature at a set value.

2.5 Use of Charges and Solvents

Molecular mechanics calculations are normally carried out in vacuum conditions by setting the dielectric constant, ε =1. The investigation of molecules containing charges and dipoles however requires the consideration of solvent effects, otherwise the conformations are influenced by strong electrostatic interactions. Force fields try to maximize the attractive electrostatic interactions, resulting in energetically strongly preferred but unrealistic low-energy conformations of the molecule. This can be prevented by using the corresponding solvent dielectric constant. For example ε = 80 in the case of water.

The strength of the electrostatic interaction decreases slowly with r^{-1}. Therefore in some cases, the dielectric constant is chosen to be distance-dependent in order to decrease more rapidly, avoiding the need to consider atoms faraway from each other, simulating the effect of displacement of solvent molecules in the path of a ligand molecule approaching a macromolecular surface.

An accurate description of the solvent environment is essential for realistic biomolecular simulations, but may become very expensive computationally. The Generalized Born (GB) model is one of the implicit solvent models that treat the solvent as a dielectric continuum **[74]**. The electrostatic interactions in the implicit solvent model are rigorously described by the Poisson Equation (PE), and the electrostatic component of the solvation free-energy ΔG_{el}, can be obtained by solving PE numerically.

It efficiently describes the electrostatics of molecules in a water environment, by representing the solvent implicitly as a continuum with the dielectric properties of water, and includes the charge screening effects of a salt. There are several versions of the GB model **[74]** implemented in AMBER 9 **[94]**. In addition to the charges of the interacting particles, the algorithm also takes into account both the dielectric constant of the solvent (approximately 80 for water), and the smoothing function which depends on atomic radii and interatomic distances of the charged particles. Some of the useful features of GB models include:

- the computational cost associated with the use of these models in MD simulations which is much smaller than the cost of representing water explicitly.
- elimination of the need for lengthy explicit water simulations due to instantaneous solvent dielectric responses described by the model.
- faster exploration of conformational space of the molecules of interest due to the absence of viscosity associated with an explicit water environment.

2.6 Energy-Minimization Procedures in Simulations

Energy minimization methods can be divided into different classes depending on the order of derivative used for locating a minimum on the potential energy surface. Zero order methods are those that only use the energy function to identify regions of low energy through a grid search method. There are several first-derivative techniques including the steepest descent method or the conjugate gradient method, which makes use of the gradient of the energy function. Second-derivative methods, like the Newton-

Raphson algorithm make use of the Hessian function to locate the minima. In this study only the first-derivative methods have been used and will be described briefly **[75]**.

2.6.1 Steepest Descent Method

The steepest descent method is useful for quickly correcting bad starting geometries or for removing bad contacts. It moves directly down the steepest slope on the potential energy surface, with limited changes to the molecular structure, and is most effective when the molecular system is far from a minimum. Since the calculation does not readily converge and can oscillate, it is often recommended that a large step-size value be chosen. The Steepest Descent method computes the gradient at its current location, and then travels in the opposite direction of the gradient until it reaches a minimum in this direction. The energy is calculated for the initial geometry and after the movements of one of the atoms in a small increment in one of the directions of the coordinate system. This process is repeated for all the atoms which finally are moved to a new position downhill on the energy surface, due to the fact that every new step is at right angles to the one before it, making numerous smaller steps to proceed down along a narrow valley, and stops when a predetermined threshold condition is fulfilled. The optimization process is slow near the minimum, and consequently, the steepest descent method is often used for the structures far from the minimum as a first rough and introductory run, followed by a subsequent minimization employing a more advanced algorithm such as the conjugate gradient **[75, 95-97]**.

2.6.2 Conjugate Gradient Method

A conjugate gradient method is a first order minimizer and differs from the steepest descent technique by using both the current gradient and the previous search direction to drive the minimization. The advantage of the conjugate gradient minimizer is that it uses the minimization history to calculate the search direction, and converges faster than the steepest descent technique. The units of the gradient are kcal mol^{-1} $Å^{-1}$, due to the fact that it is the rate of change (first derivative) of the total energy with respect to atomic positions. The conjugate gradient method produces a set of directions that overcome the oscillatory behaviour of the steepest descents in narrow valleys. Successive directions are not at right angles to each other **[96],** and the conjugate gradient algorithm accumulates the information about the function from one iteration to the next. For each minimization step, the gradient is calculated and used as additional information for computing the new direction vector of the minimization procedure. Thus, each successive step refines the direction towards the minimum. The computational effort and storage requirements are greater than those for the steepest descent, but the conjugate gradient method for larger systems is the method of choice. The greater total computational expense and the longer time per iteration is more than compensated by the more efficient convergence to the minimum achieved in the case of conjugate gradients **[85, 96]**.

There are several ways in molecular minimization to define convergence criteria. In non-gradient minimizers, only the increments in the energy and the coordinates can be taken into account in order to judge the quality of the actual geometry of the molecular

system. However, in all gradient minimizers, the atomic gradients are used for this purpose. The best procedure in this respect is to calculate the root-mean-square gradients of the forces on each atom of a molecule **[75, 85]**. The value chosen as a maximum derivative will depend on the objective of the minimization. If a simple relaxation of a strained molecule is desired, a rough convergence criterion like a maximum derivative of 0.1 kcal $mol^{-1}Å^{-1}$ is sufficient, while for other cases 0.001 kcal $mol^{-1}Å^{-1}$ for a final minimum is sufficient **[85]**.

2.7 Experimental Techniques

2.7.1 X-ray Crystallography

X-ray crystallography is an experimental technique used for the study of the internal structure of crystalline materials, often known as X-ray diffraction. The technique is based on the interference pattern produced as X-rays pass through the three-dimensional, repeating pattern of atoms within a crystal lattice. A commonly used criterion for the validation of an MD simulation is the root-mean-square deviation (RMSD) obtained from the crystal structure of a protein; although proteins are usually simulated in water rather than a crystalline environment **[98]**.

2.7.2 Circular Dichroism (CD)

Circular dichroism is a spectroscopic method that is well established in the biochemical community. Theoretical aspects as well as applications to proteins, peptides, DNA and RNA have been reviewed thoroughly by Woody and coworkers **[99-100]**. The contribution of specific parts of a protein to the CD spectrum can be obtained using

molecular orbital calculations, as has been done for aromatic groups, peptide groups in poly-Gly helices and for methylated phenols complexed with β-Cyclodextrin **[101-102]**. A logical step in refining the methodology for calculating CD spectra seems to be the use of quantum chemistry at a high level of theory, rather than the customary molecular orbital calculations. It must be noted that the relatively new technique of vibrational circular dichroism (VCD) **[99]** has achieved quite some interest from theoretical chemists, e.g. **[103]**. However, the physical basis of VCD is very different from conventional CD, and therefore the theoretical frame work in the VCD field is not applicable to CD.

2.7.3 Nuclear magnetic resonance (NMR)

Nuclear Magnetic Resonance (NMR) is an experimental technique that exploits the magnetic properties of the atomic nuclei to predict the structure of molecules. It is also a powerful technique used to study the thermodynamics of biological macromolecules, because of a symbiotic relationship between MD simulations and NMR. The nuclear overhauser effects (NOEs) and transverse and longitudinal relaxation times are common observables used in the NMR experiments to study the peptide/protein folding. Relaxation effects can arise from dipolar interactions, quadrupolar interaction for nuclei with spin >1, and chemical shift anisotropy. The longitudinal and transverse relaxation rates T_1 and T_2 can be described in terms of the spectral density functions:

$$T_1^{-1} = A[J(\omega_i - \omega_j) + 3J(\omega_i) + 6J(\omega_i + \omega_j)] \qquad (2.1)$$

$$T_2^{-1} = \frac{1}{2} T_1^{-1} + \frac{A}{2}[4J(0) + 6J(\omega_j)] \qquad (2.2)$$

with,

$$A = \left(\frac{\hbar \gamma_i \gamma_j \mu_0}{8\pi} \right)^2 \qquad (2.3)$$

where $\gamma_x = g_x \mu_x / \hbar$ is the gyromagnetic ratio for nucleus x, g_x is the nuclear g-factor, μ_x is the nuclear magneton and μ_0 is the magnetic permeability of free space $\left(\mu_0 / (4\pi) = 10^{-7}\right)$. It should however be noted that deviations between computed longitudinal relaxation rates T_1 and the measured ones may occur due to spin diffusion **[97, 104]**. The cross-relaxation term σ_{ij}, i.e. the relaxation of nucleus i due to nucleus j, in terms of $j(\omega)$ is **[105]**:

$$\sigma_{ij} = (6J(\omega_i + \omega_j) - J(\omega_i - \omega_j)) \qquad (2.4)$$

Using the cross-relaxation and the longitudinal relaxation time T_1 the steady state NOE_{ij} between particles i and j can be written as:

$$NOE_{ij} = 1 + \frac{\gamma_i}{\gamma_j} A T_1 \sigma_{ij} \qquad (2.5)$$

Since T_1 occurs in the definition of the NOE, it is also influenced by the spin diffusion **[104]**. The steady state NOE is used primarily in heteronuclear NMR experiments. In homonuclear 1H-NMR experiments, $NOEs$ can be measured by multi-dimensional

experiments, and through the NOE intensity I_{ij}, the cross relaxation σ_{ij} can be determined directly **[106]**:

$$I_{ij} \propto \sigma_{ij} \tag{2.6}$$

A few limiting cases are of particular interest. The first, is that of a slowly tumbling macromolecule with fast internal motions. In this case the cross relaxation term is dominated by $J(0)$, and can be written using the model-free approach of Lipari and Szabo.

$$\sigma_{ij} = -\frac{2}{5}\left[\tau_c \left\langle r_{ij}^{-6} \right\rangle + (\tau_M - \tau_c)\frac{4\pi}{5}\sum_{m=-2}^{2}\left\langle \frac{Y_{2m}(\theta,\phi)}{r_{ij}^3} \right\rangle^2\right] \tag{2.7}$$

When $\tau_M >> \tau_c$ this reduces to:

$$\sigma_{ij} = -\frac{8\pi\tau_M}{25}\sum_{m=-2}^{2}\left\langle \frac{Y_{2m}(\theta,\phi)}{r_{ij}^3} \right\rangle^2 \tag{2.8}$$

As was originally shown by Tropp **[105]**, the distance dependence of σ_{ij} involves r^{-3} rather than r^{-6}, which means that particles at relatively long distances may contribute to cross relaxation. Hence, the NOE intensity depends on the time average of r^{-3}. In the case of the distance being a constant, the distance dependence can be removed from the averaging, and, using the addition theorem for spherical harmonics **[107]**, the cross relaxation term σ_{ij} can be written as:

$$\sigma_{ij} = -\frac{2}{5}\tau_M \left\langle r_{ij}^{-6} \right\rangle \left\langle P_2(\hat{r}_{ij}(0).\hat{r}_{ij}(t)) \right\rangle \qquad (2.9)$$

Another important effect is that of multiple protons (e.g. a methyl group) that contribute to cross relaxation at another proton. Some care is required, since in practice one usually works with effective distances r_{ij}^{noe} rather than σ_{ij} (here we omit the angular dependence for the sake of clarity):

$$r_{ij}^{noe} \propto \sigma_{ij}^{-1/6} \qquad (2.10)$$

The cross relaxation terms σ_{ij} can be added linearly, but in terms of distances this means that one first has to compute the $\left\langle r^{-3} \right\rangle$ average for each of the particles, and subsequently the sum of these squared (using eqn. 2.8, i.e. under the assumption that $\tau_M >> \tau_c$

$$r_{ij}^{noe} = \left[\sum_{j=1}^{3} \left\langle r_{ij}^{-3} \right\rangle^2 \right]^{-1/6} \qquad (2.11)$$

The distance information from NOEs can be used for refinement during MD simulations by introducing a penalty function for distance restraints such as:

$$V_{ij} = \begin{cases} 0 & r_{ij} \le r_{ij}^{noe} \\ \frac{1}{2} k_{dr} (r_{ij} - r_{ij}^{noe})^2 & r_{ij} > r_{ij}^{noe} \end{cases} \qquad (2.12)$$

where k_{dr} is a force constant, which is taken as the order of 1000 kJ mol^{-1} nm^{-2}. It may be clear from the preceding paragraph however, that internuclear distances derived from $NOEs$ (eqn. 2.10) have to be interpreted as effective average distances. Usually, structure refinement is done for proteins, so that we can use eqn. 2.8 to calculate an effective average distance; the explicit angular dependence has, to our knowledge, never been used. Rather, this term is therefore ignored and an effective distance $\bar{r}_{ij}$ is defined as:

$$\bar{r}_{ij} = \left\langle r_{ij}^{-3} \right\rangle^{-1/3} \qquad (2.13)$$

Torda et al. introduced an algorithm to take time-averaged distances into account in refinement **[108]**, and in a later paper applied it to the refinement of tendamistat, a protein which specifically inhibits mammalian alpha-amylases **[109]**. Since this original paper **[108]** the use of time-averaging has become standard practice in NMR refinement based on NOE restraints. A number of interesting studies of NMR relaxation from simulation data are present in scientific literature **[110-112]** as well as some older studies based on short simulations in vacuo **[113-114]**.

2.8 MD Analysis programs

2.8.1 The AMBER 9.0 computer program [94]

The term AMBER is a collective name for a suite of programs that allows users to carry out molecular mechanics calculations, particularly on biomolecules. It is also used to refer to the empirical force field that is implemented in the AMBER 9 computer program

[94]. Understanding where to begin in AMBER is primarily a problem of managing the flow of information in this program. Clearly, one needs to understand what information is needed when working with the simulation programs such as SANDER, PMEMD and NMODE within the AMBER program. The following information is needed by all the programs:

(i) Cartesian coordinates for each atom in the system.

This information usually comes from X-ray crystallography, NMR spectroscopy, or model-building programs. The program LEaP provides a platform for carrying out many of these modeling tasks, but other programs may be considered as well.

(ii) “Topology” Information regarding the connectivity, atom names, atom types, residue names, and charges. This “topology” information comes from the database. It contains information for the standard amino acids as well as N and C-terminal charged amino acids. The database contains default internal coordinates for these monomer units. However, coordinate information is usually obtained from the protein database (PDB) files.

(iii) The parameters for all of the bonds, angles, torsions and atom types in the system. These are the basic force field parameters, which are found in the database.

(iv) The procedural options and the desired parameters are provided to the programs in three separate files. The first contains the coordinates, the second contains the topology and parameters, and is called the “topology file” and the third contains the command or input file.

The primary aim of this study was to study the conformational profiles of Aurein peptides using force-field simulations within the framework of molecular mechanics. For this purpose, three additional components of the AMBER program extensively used in this study, are described below:

2.8.2 Preparatory programs in AMBER [94].

In this study, all the peptide sequences were built using the LEAP and ANTECHAMBER modules as part of the preparatory programs within the AMBER program.

LEaP is the primary program used to create a new system in Amber, or to modify old systems. It combines the functionality of prep, link, edit, and parm from earlier versions.

ANTECHAMBER is the main program from the Antechamber suite. For systems containing more than just standard nucleic acids or proteins, this may be helpful to prepare the input for LEaP.

2.8.3 Simulation programs in AMBER [94].

The AMBER program contains the energy programs SANDER, NMODE and PMEMD. For the purposes of this study, only the SANDER module was used, and thus a brief description is warranted.

SANDER

This part of the module carries out energy minimization, molecular dynamics, and NMR refinements. The acronym SANDER, refers to **S**imulated **A**nnealing with **N**MR-**D**erived

Energy **R**estraints. However, this module is used for a variety of simulations that have nothing to do with NMR refinements. This module provides standard protocols for energy minimization and molecular dynamics, and is used for everything except Gibbs free energy calculations. This program relaxes the structure by iteratively moving the atoms down the energy gradient until a sufficiently low average gradient is obtained. As a standard practice, structures are generally minimized before a molecular dynamics simulation. The molecular dynamics portion generates configurations of the system by integrating the Newtonian equations of motion. MD will sample more configurational space than minimization, and will allow the structure to cross over small potential energy barriers. Configurations may be saved at regular intervals during the simulation for later analysis, and basic free energy calculations using thermodynamic integration may be performed.

2.8.4 Analysis programs in AMBER [94].

Of the analysis modules incorporated in the AMBER program, PTRAJ module was relevant for the studies undertaken, and is discussed below.

PTRAJ: a general purpose utility module for analyzing and processing trajectory or coordinate files created from MD simulations (or from various other sources), carrying out superpositions, extractions of coordinates, calculation of bond/angle/dihedral values, atomic positional fluctuations, correlation functions, and analysis of hydrogen bonds.

2.8.5 Conformations Classification (CLASICO) [115]

CLASICO is a computer program designed specifically to group thousands of structures obtained from MD, REMD and SA trajectories into hundreds of patterns, easing the subsequent treatment of the information obtained. A secondary structure can then automatically be assigned to each of the structures left. For a given conformation the procedure consists of assigning a letter to each of the residues of the peptide according to the values of its backbone dihedral angles, following a partition of the space **[116]**. The procedure is based on splitting the Ramachandran plot into different regions as shown in Figure 2.4.

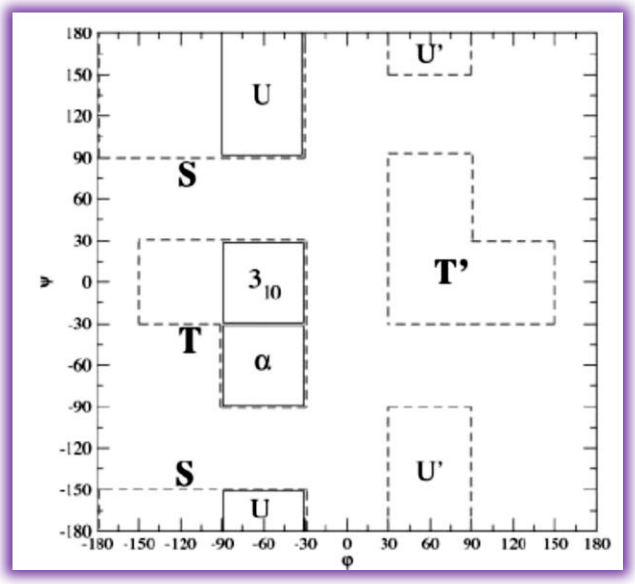


Figure 2.4 Conformational space partitioned into several regions **[115]**. The regions are named after the secondary structure motif that each encompasses, and labelled as: 3_{10}: 3_{10}-helix; α: α-helix; S: β-strand; T: type I β-turn, T': type I' β-turn; U: type II β-turn; U': type II' β-turn.

These regions are labelled S, T, T', U, U', 3_{10}, α and accordingly, from the dihedral angles of each residue, conformations can be described by a sequence of letters. In a second step, to each of the conformations already coded by a string of letters, successive two- or three-letters (Figure 2.4) are considered to assign conformational motifs, using a set of rules shown in Table 2.1.

Table 2.1 Conditions for secondary structure definition of three consecutive residues **[115]**, where 3_{10}: 3_{10}-helix; H: α-helix; S: β-strand; T: type I β-turn, T': type I' β-turn; U: type II β-turn; U': type II' β-turn and j: amino acid residue of peptide sequence.

Motif	Condition	Assignment	Code
3_{10}-helix	j and j+1 and j+2 x 3_{10}	j, j+1, j+2 = 3_{10}-helix	3_{10}
α–helix	j and j+1 and j+2 x H	j, j+1, j+2 = α-helix	α
β-strand	j and j+1 and j+2 x S	j, j+1, j+2 = β-strand	S
type I β-turn	j+1 x T and j+2 x T	j+1= type I β-turn (residue 1+1)	I1
	j x T and j+1 x T	j+1= type I β-turn (residue 1+2)	I2
type I' β-turn	j+1 x T' and j+2 x T'	j+1= type I' β-turn (residue 1+1)	i1
	j x T' and j+1 x T'	j+1= type I' β-turn (residue 1+2)	i2
type II β-turn	j+1 x U and j+2 x T'	j+1= type II β-turn (residue i+1)	II1
	j x U and j+1 x T'	j+1= type II β-turn (residue i+2)	II2
type II' β-turn	j+1 x U' and j+2 x T	j+1= type II' β-turn (residue i+1)	ii1
	j x U' and j+1 x T	j+1= type II' β-turn (residue i+2)	ii2
Coil	None of the above	j+1=coil	-

These new strings of conformational motifs are called patterns. This procedure enables the different conformational patterns attained by the peptide to be identified, as well as to compare differences of the conformational space sampled using different computational methods. Secondary structure patterns are then numbered and plotted along the trajectory, and frequency of the different patterns can be graphically assessed

by looking at the distribution of points. Following this procedure, the CLASICO algorithm **[115]**, can translate the thousands of snapshots stored during the MD, REMD or SA trajectories into 11 different possible motifs as shown in Table 3.1 (H, 3_{10}, S, I1, I2, i1, i2, II1, II2, ii1, and ii2), and the classification of β-turns on the basis of dihedral angles (ϕ and ψ) is depicted in Table 2.2.

Table 2.2 Definition ofβ-turns classified on the basis of dihedral angles **[115]**.

Type of β-turns	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	[-110, -10]	[-80, 20]	[-140, -40]	[-50, 50]
i	[10, 110]	[-20, 80]	[40, 140]	[-50, 50]
II	[-110, -10]	[70, 170]	[30, 130]	[-50, 50]
ii	[10, 110]	[-170, -70]	[-130, -30]	[-50, 50]
III	[-110, -10]	[-80, 20]	[-110, -10]	[-80, 20]
iii	[10, 110]	[-20, 80]	[10, 110]	[-20, 80]
VIa	[-110, -10]	[70, 170]	[-140, -40]	[-50, 50]
VIb	[-170, -70]	[70, 170]	[-110, -10]	[-50, 50]
VIb’	[-170, -70]	[70, 170]	[-110, -10]	[100, -160]

A flowchart showing the procedure and the commands used in CLASICO program **[115]** is depicted in Figure 2.5. The dihedral angles obtained from different trajectories are used as input in CLASICO to generate and classify the secondary structure motifs of the peptides.

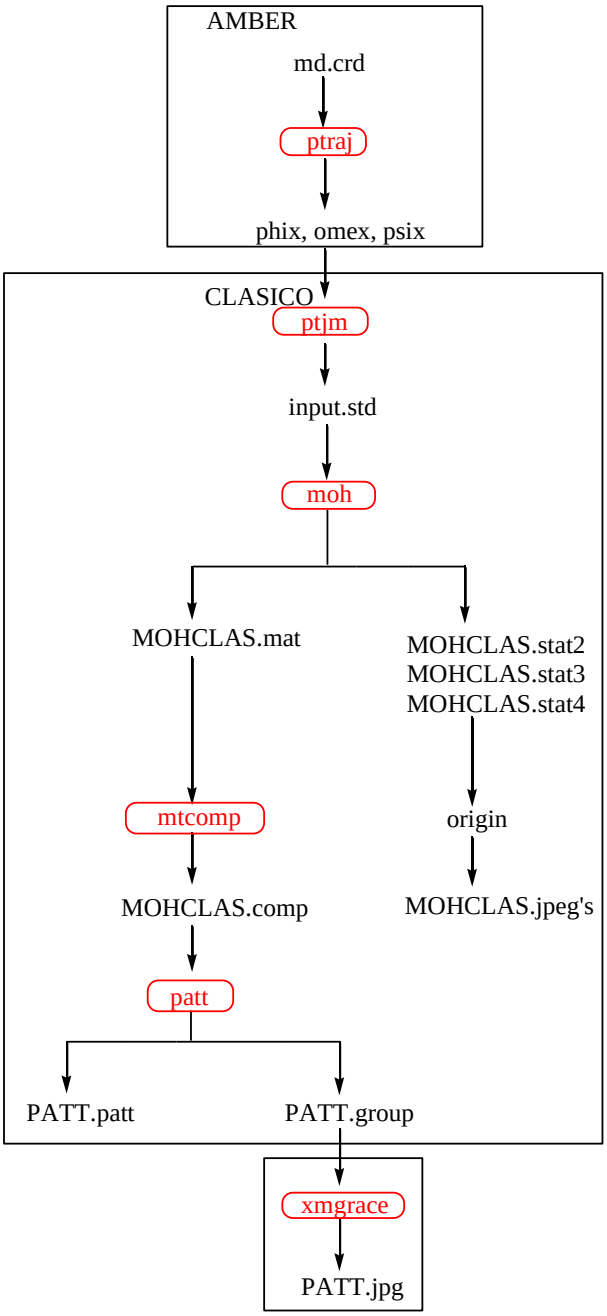


Figure 2.5 Flow chart for the protocol in the CLASICO program **[115]** (commands are shown in red).

2.8.6 Commands used in CLASICO program [115]

ptjm

In order to process the data the first step is to create the input for the CLASICO with ptjm. The program ptjm will convert the data from the files omex, phix, and psix obtained in ptraj to create an input.std file.

moh

This program classifies structures based on structural motifs. 3 consecutive residues which accomplish whatever helical condition (H, 3_{10} or PI) are considered an helix. The specific motif assignation is done in a local way depending on the condition accomplished. When different assignations can be done, the order of priority is H > 3_{10}> PI. The input.std file will be read by moh executable to create the MOHCLAS.mat.

mtcomp

The next step is to convert the MOHCLAS.mat into a compressed matrix with the mtcomp executable. The format of MOHCLAS.comp contains in the first column the number of the snapshot. The second column contains the number of following items in the same row.

patt

The program patt classifies the structures in patterns based on the conformational motifs present.

MOHCLAS.comp and PATT.sum are input files.

PATT.patt is the pattern file (identical to MOHCLAS.comp but only with non-repeated patterns)

PATT.group is the file for recording the appearance of new patterns- pattern numbers are assigned to the structure.

PATT.perclass records the number of times that a given pattern appears.

PATT.relev records the number of times that a given pattern appears if it appears more than the 0.1% of the structures.

CHAPTER 3

COMPUTATIONAL STUDY OF AUREIN 2.3 AND AUREIN 2.4

The work described in this chapter involves the study of the conformational profile of two Aurein peptides (Aurein 2.3 and Aurein 2.4) using the molecular dynamics computational method. First, a validation study was performed on Aurein 2.3 to check the efficiency of the computational procedure. The results showed the presence of an α-helical region distributed over all the amino acids residues of Aurein 2.3, in accordance with the experimental results. Subsequently, the conformational preferences of Aurein 2.4 were studied using the same computational procedure under the implicit solvent conditions. The results obtained revealed that the peptide adopts α-helicity predominantly in the central residues extending towards its C-terminus. Additionally, MD simulations performed in the polar solvents (water and methanol) indicated the absence of any characteristic structure of the Aurein 2.4.

3.1 Introduction

Cancer remains one of the several lethal diseases despite intensive research worldwide. The inadequate knowledge of biochemical mechanisms for carcinogenesis, non-selectivity of available drugs and multi-drug resistance (MDR) are few factors that are responsible for the exacerbation of the cancer therapy. The development of MDR protein kinase inhibitors and targeted anti-cancer drugs has made some progress to overcome these problems. However, there is still a need for new efficient drugs with novel mechanisms of action for cancer treatment, and one recent focus has been on anti-microbial peptides (AMPs). Generally, these peptides exhibit their biological activities at low micromolar levels with minimum toxicity to the normal cells. Some aurein AMPs have even shown anti-cancer activity against multiple types of human

cancers including renal, ovarian, colon, breast, lung and cancer with non-haemolytic action. These peptides are obtained from the amphibian skin and form an integral part of their defensive system.

Aurein peptides are very important class of cationic AMPs that have shown their potential against gram positive bacteria (*Leuconostoc lactis, Bacillus cereus, Listeria innocua, Staphylococcus aureus, Micrococcus luteus* and *Staphylococcus epidermidis*) and cancers [**63**]. Thus far, the most studied member of this class is Aurein 1.2, a 13 amino acid residue peptide, with 1+ charge [**67, 117-120**]. The anti-bacterial activity of this peptide has been reported to increase with increase in its positive charge in the presence of amide functionality at its C-terminus [**121-122**]. Aurein 2.2 (net charge +2) and Aurein 2.3 (net charge +2) are other peptides of this class with almost similar biological activity profiles. Several experimental studies focusing on the structural characterization of Aurein 1.2 [**63, 67**], Aurein 2.2 and Aurein 2.3 [**65**] have been reported in the literature. For instance, the circular dichroism (CD) and NMR studies performed on Aurein 2.3 in 25% TFE-water mixture revealed the presence of α-helicity in the peptide [**65**]. The peptide, however, remains unstructured in the presence of water solvent, and attains α-helicity only in the presence of membrane environment. Similarly, Aurein 1.2 also adopts α-helical conformation in membrane-mimic micelles **[63,67**]. However, despite having comparable medicinal and biological potential, the detailed three-dimensional structures of other peptides of Aurein family (Aurein 2.4 and Aurein 2.5) have not been characterized thus far.

Aurein 2.4 is a structural analogue of Aurein 2.3 sharing similar first eleven and last amino acid residues (Figure 3.1) in their structures. It differs from Aurein 2.5 due to the presence of four C-terminal residues; threonine (T), leucine (L), alanine (A) and glycine (G) are present at positions 12, 13, 14 and 15, respectively. There is a difference of only one amino acid at position 13 in the structures of Aurein 2.3 and Aurein 2.5. The biological activities of these peptides are significantly affected by the nature of amino acid residues in their structures. Because of the biological relevance of these aurein peptides, there is a considerable interest in the clinical potential of Aurein agonists, particularly in the fight against the cancer. Also due to their tendency to act as anti-microbial agents, they are interesting targets for drug discovery.

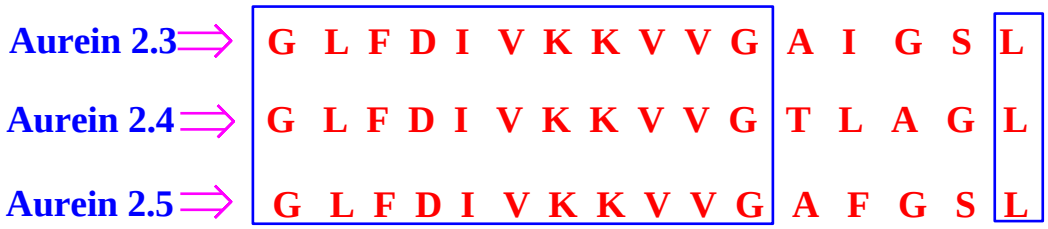


Figure 3.1 Single letter code for the amino acid sequences of peptides Aurein 2.3, Aurein 2.4 and Aurein 2.5. Conserved residues in all peptides are shown in the blue block.

In this chapter we report the results of the exploration of the conformational space of two medium-sized Aurein peptides (Aurein 2.3 and Aurein 2.4) using standard molecular dynamics (MD) calculations at 300 K and using the Langevin's thermostat [**93**, **123**]. The present work is intended to get some insight into the three dimensional (3D) structural features of these peptides by exploring their configurational space based on molecular dynamics. As longer MD simulations in combination with GB-OBC implicit solvent model [**74**] and modified amber FF99 (FF99SB) are quite successful to

reproduce the peptide secondary structures, this computational procedure was used in our all MD simulations. First, a validation study was performed on Aurein 2.3 whose experimental structure has already been known, followed by the application of same protocol to predict the 3D structures of Aurein 2.4 and Aurein 2.5 for which experimental data is not available.

3.2 Molecular dynamics simulation of Aurein 2.3 peptide: A validation study

3.3 MD methodology

The extended conformation of Aurein 2.3 peptide was generated using the LEaP module of AMBER 9 [**94**], and used as starting structures for the MD simulation. First, the energy minimization of peptide was performed using 500 steps of steepest gradient, followed by 500 steps of the conjugate gradient algorithm until a convergence of the gradient norm was lower than 0.001 kcal mol^{-1} $Å^{-1}$. Following this the structure was equilibrated for 20 ps by gradually increasing the temperature from 0 to 300 K. MD trajectory was conducted using the generalized Born (GB) approximation at constant temperature (300 K) regulated by the Langevin algorithm as a thermostat [**93, 123**]. The list of non-bonded interactions was updated every 10 steps. Internal dielectric constant for the peptide was set to 1, while an external dielectric constant of 80 corresponding to water was employed. The SHAKE algorithm was used for bonds involving hydrogen atoms and an integration time-step of 2 fs was employed. The analysis of MD trajectory reported in the present work was carried out with the AMBER 9.0 scripts [**94**]. Secondary structure analysis was performed using the CLASICO program [**115**].

For explicit solvent calculations, the extended conformation of Aurein 2.3 peptide was soaked in a box water and methanol containing 5171 and 2241 solvent molecules

respectively. The charge on peptide was neutralized by adding two chloride ions. The minimization of these new systems was performed until a convergence criteria (0.001 kcal mol^{-1} Å) was achieved. Thereafter, periodic boundary conditions were introduced and the structure was allowed to equilibrate for 500 ps at a temperature of 300 K [**93**], with the pressure set to 1 bar. The SHAKE algorithm was used to constrain all the bonds involving hydrogen atoms with a time step of 2 fs. After the first equilibration phase, the Particle-Mesh Ewald (PME) method was applied with a grid spacing of approximately 1Å. The production phase consisting of length 200 nanoseconds was then performed for both peptides at 300 K under similar conditions.

3.4 Results and discussion

The 300 nanoseconds (ns) MD trajectory was computed using the extended conformation of Aurein 2.3 peptide. A qualitative analysis was subsequently performed on the thermodynamic properties such as temperature (Figure 3.2a) and potential energy (Figure 3.2b) of the trajectory to assess the stability and accuracy of the MD simulation. Clearly, an average fluctuation of temperature around 300K indicated a stable nature of the MD trajectory over the entire simulation time (Figure 3.2a). Moreover, the initial drop in potential energy after few nanoseconds of the MD simulation (Figure 3.2b), and its constant average fluctuation around -180 kcal mol^{-1} suggested the considerable folding of the peptide during the progress of MD trajectory.

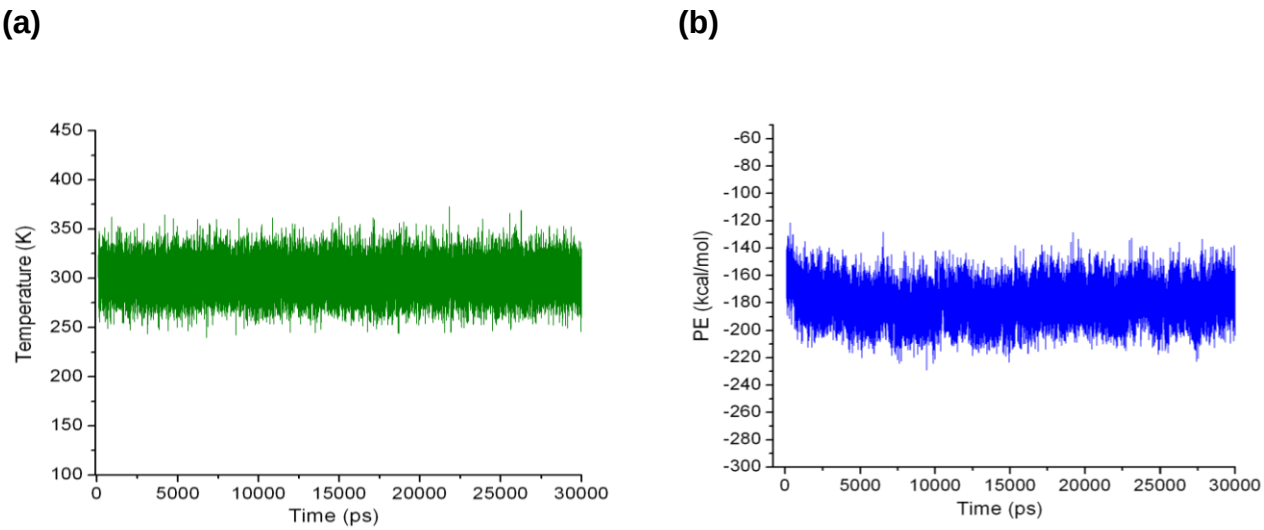


Figure 3.2 Temperature (a) and Potential energy (b) of the sampled conformations during the progress of MD simulation.

The atomic fluctuations in the peptide during the MD simulation were monitored by plotting the root mean square deviation (RMSD) of the backbone atoms of the conformations using their extended conformation as a reference, and are shown in Figure 3.3. The significant increase in RMSD (~ 12 Å) immediately after the start of MD simulation indicated the considerable deviation of the peptide structure probably due to the folding process. The constant average oscillations of the RMSD around 9.2 Å further suggested that the folding of the peptide to a particular structure. Moreover, the stable average fluctuation of the RMSD roughly indicated the adequate length of the MD simulation. Nevertheless, the peptide structure was not stable enough and was in a constant equilibrium with other structures during the progress of MD trajectory as evidenced by the RMSD oscillations between 6-12 Å, as depicted in Figure 3.3.

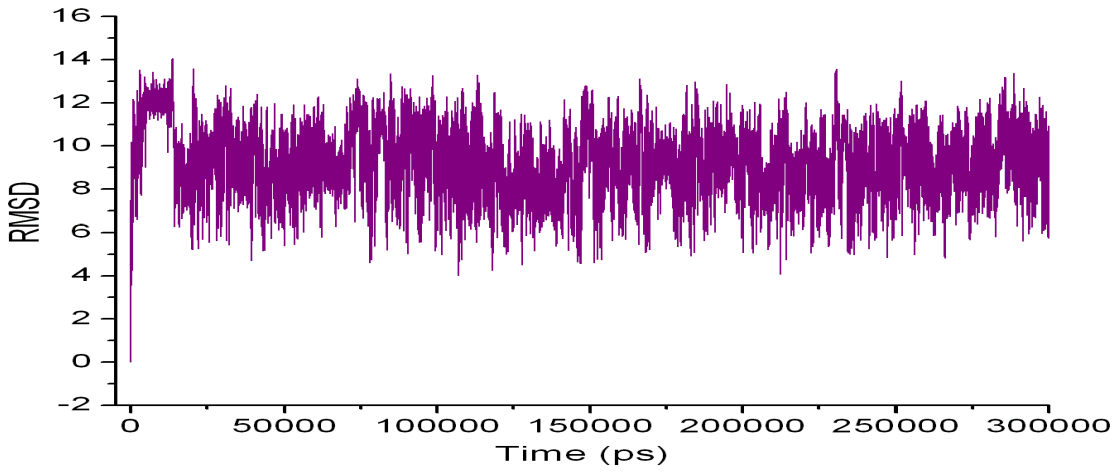


Figure 3.3 Root mean square deviations of sampled conformations relative to the starting structure considering their backbone atoms.

For every sampled conformation in the MD trajectory, conformational patterns were subsequently assigned following the procedure depicted in Chapter 2. Consequently, 64345 new patterns were characterized from the total of 300,000 sampled conformations, and their systematic development during the progress of MD trajectory is diagrammatically depicted in Figure 3.4. These plots provide a broad estimation of the performance of the MD protocols in sampling new patterns. As can be seen in Figure 3.4, the generation of patterns was quite low during the first 13 ns of the trajectory. However, a significant increase was observed after 14 ns, with new patterns sampled in a uniform fashion throughout the progress of the MD simulation. The peptide nevertheless experienced few trappings in the few regions of conformational space at certain intervals of the MD trajectory.

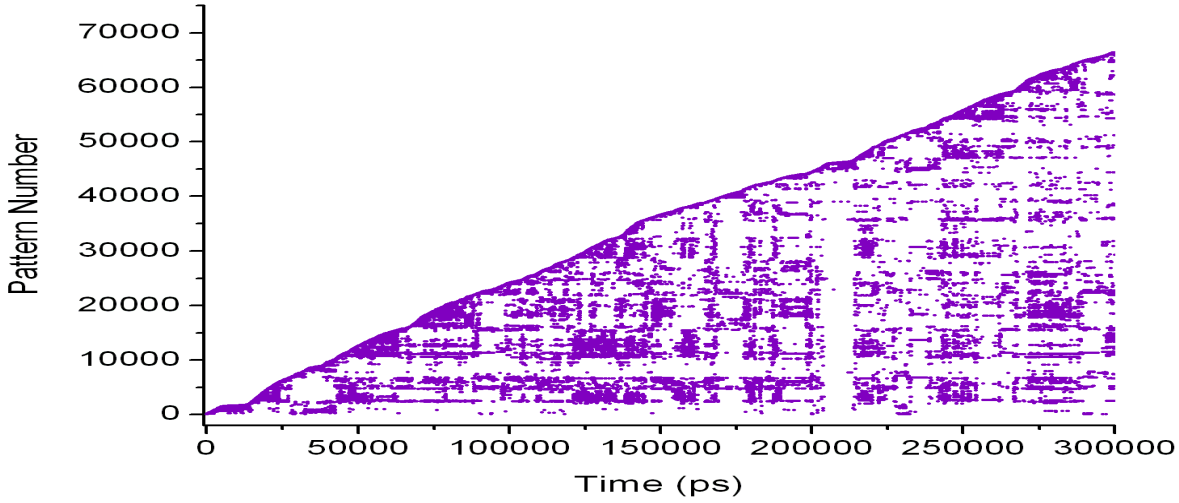


Figure 3.4 Evolution of patterns (new conformational motifs) during the progress of MD trajectory.

In order to get a deeper understanding of the secondary features of the peptide, a qualitative analysis of the conformational motifs was performed on the sampled structures of the MD trajectory using the CLASICO program [**115**]. In this computer program, each snapshot is translated into a string of letters followed by computing of backbone torsion angles for each residue, and a letter code is assigned to them following the Zimmerman partition [**116**] of the Ramachandran map. The secondary motifs are then assigned to each residue of the peptide by means of certain rules (details in Chapter 2). Figure 3.5 represents the different conformational motifs attained by the amino acid residues of the conformations along with their percentages. A closer inspection of Figure 3.5 revealed the presence of α-helical region supported by all residues (2-16) of the peptide in different percentages. Similarly, β-turns were the second most preferred conformational motifs adopted by the sampled conformations in the MD trajectory. Some conformations adopting 3_{10}-α-helicity between all residues (2-16) were also observed, but in lower percentages (2-3%). Since, the NMR studies

performed on Aurein 2.3 have also revealed the presence of a α-helical region distributed on all its amino residues [65], our computational results were in a good agreement with each other, and thus validated the computational procedure.

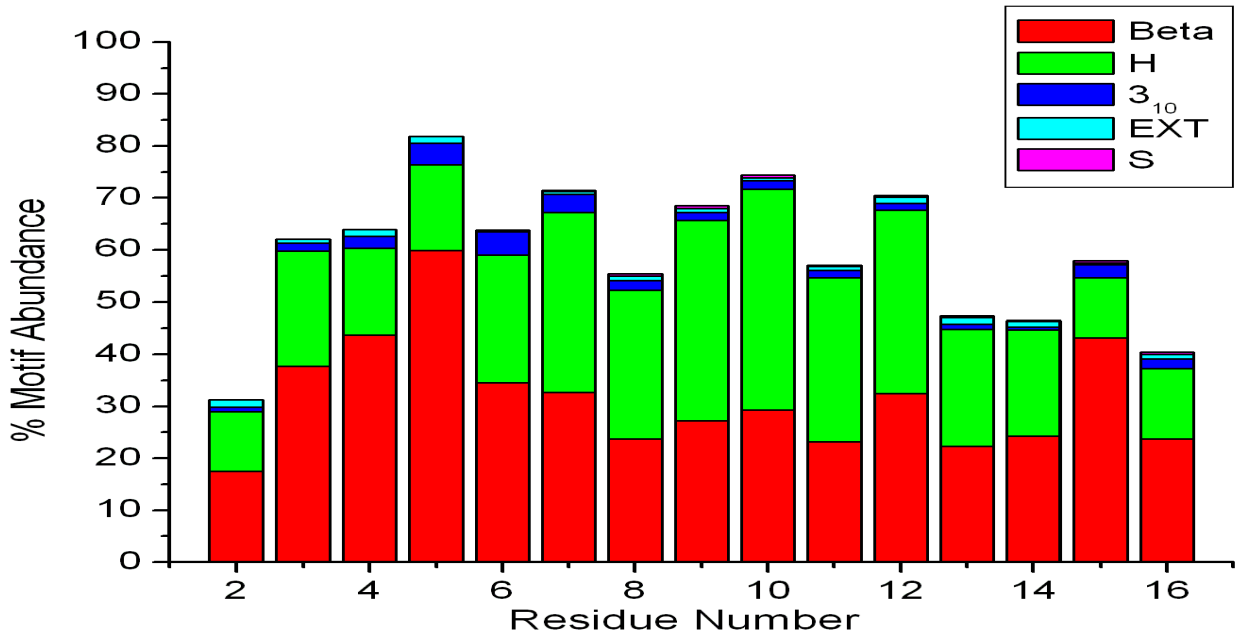


Figure 3.5 Different secondary motifs attained by amino acid residues of conformations in the MD trajectory for Aurein 2.3, obtained using the CLASICO program [**115**]. Conformational motifs are labelled: Beta (β -turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

3.5 Molecular Dynamics study of Aurein 2.4

Clearly, longer MD simulations (>300 ns) in combination with the GB_OBC solvent model and FF99SB force field were able to reproduce the NMR supported structure of Aurein 2.3 [65], where this peptide has been reported to exhibit α-helicity in its all amino acid residues. Consequently, this study was further extended to predict the three

dimensional structures of Aurein 2.4 (Figure 3.6) and Aurein 2.5. Accordingly, MD simulations of 300 ns length were performed on both these peptides using similar conditions as described in the method section for Aurein 2.3. The results of the MD simulation for Aurein 2.4 are discussed in this chapter whereas those corresponding to Aurein 2.5 are described in Chapter 4.

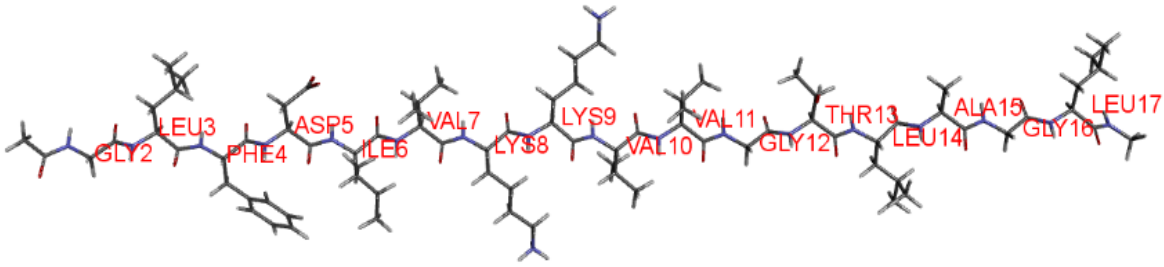


Figure 3.6 The extended conformation of Aurein 2.4 with its amino acid sequence.

First, the root mean square deviation (RMSD) of the sampled conformations in MD trajectory was calculated relative to the starting conformation of Aurein 2.4 using their backbone atoms and all atoms, and is pictorially depicted in Figure 3.7. Clearly, the atomic fluctuations in the case of all atoms (in red) were found to be higher as evidenced by the oscillations of the RMSD around 9 Å during the progress of the MD trajectory. The RMSD computed using the backbone atoms was comparatively lower, showing an average fluctuation around 5.3 Å. It is believed that the side-chains of the peptide relative to the backbone atoms were quite flexible during the progress of the MD simulations. Moreover, the greater computed RMSD values indicated that the sampled conformations were significantly deviated compared to their initial strucutres probably owing to their effective structural folding during simulation.

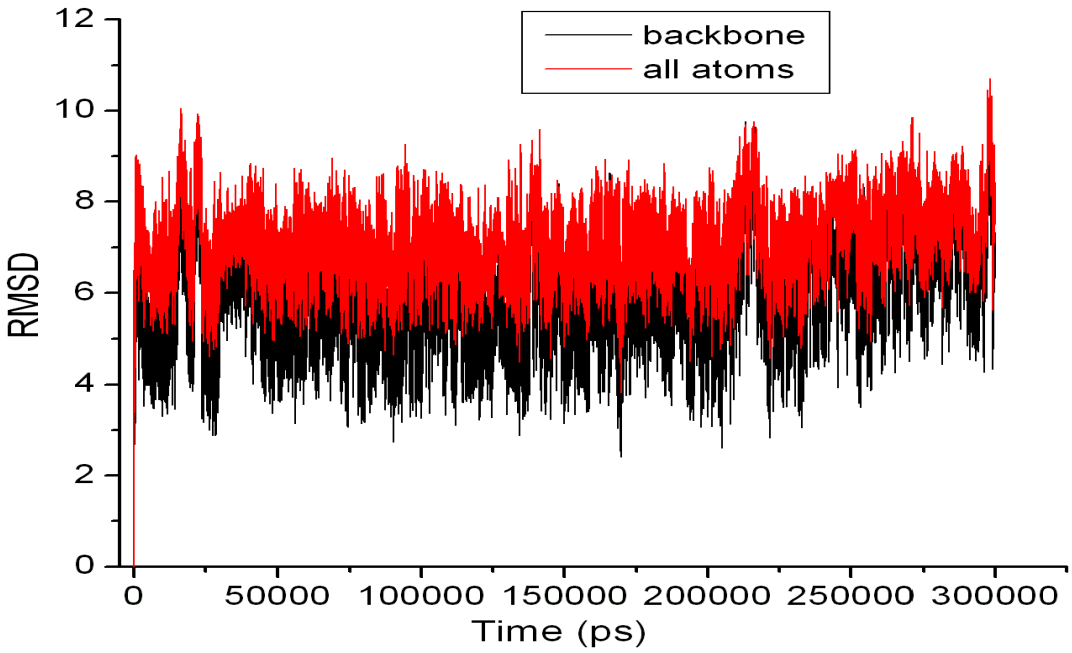


Figure 3.7 The root mean square deviation (in Å) of conformations relative to the starting structure calculated considering backbone atoms (in black) and all atoms (in red).

However, constant fluctuations of the RMSD (~6Å, Figure 3.7) indicated that the folded structures were not stable enough and were in a rapid equilibrium with their unfolded strucutral analogues during the sampling process throughout the progress of the MD trajectory. The constant average fluctuation of the RMSD also suggested the adequate length of the MD simulation chosen for this study.

In order to monitor the sampling efficiency of the MD simulation, the development of patterns during the progress of MD trajectory was monitored using the CLASICO program [**116**], and is pictorially shown in Figure 3.8. Of the set of 300000 sampled conformations, 69293 new patterns were obtained for the Aurein 2.4 peptide in the MD simulation performed under implcit solvent conditions. As can be seen in Figure 3.8 the sampling of new patterns started immediately after the start of the simulation and proceeded in a uniform fashion throughout the progress of the MD trajectory. A couple

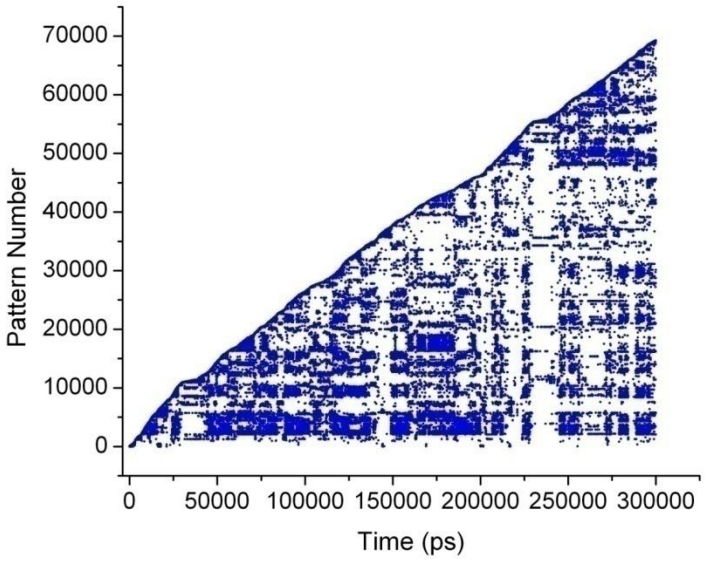


Figure 3.8 Evolution of the patterns during the MD simulation for peptide Aurein 2.4 obtained using the CLASICO program [**115]**.

of plateau formations suggested that the conformations were trapped in the segments between 34-36 ns and 228-230 ns of the MD simulation.

The existence of a large conformational variety in the MD trajectory prompted us to group the sampled conformations into different assembles by considering their structural resemblances. Accordingly, 5000 structures were selected as representative structures from 300 ns MD trajectory by choosing each conformation at an interval of 60 picoseconds (ps). These representative strucutres were subsequently subjected to cluster analysis using the Ward's hierarchical clustering algorithm [**124**]. The major clusters identified from the cluster analysis are shown as a dendrogram in Figure 3.9, while the identification of secondary features and other detailed results are depicted in Table 3.1. The absence of gaps at the bottom (corresponding to lower cutoff values) of

the dendrogram suggests a greater overlap of the structures in that region of the cluster, in contrast to the top end (corresponding to higher cutoff values). Similarly, a horizontal line across the dendrogram identifies the number and corresponding members of the cluster at any particular stage. Accordingly, four major clusters of sampled conformations (shown in different colours, Figure 3.9) were identified in the MD trajectory of Aurein 2.4 by selecting a cutoff value around 5. A closer inspection of Table 3.1 revealed that more than 60% of the strucutres of the total population were occupied by the major clusters C1 and C3. The major cluster C1 acquires 36% of the conformations whilst the second largest cluster C3 contributed 25% of the stuctures to the total population.The remaining two clusters, C2 and C4, acquired 22% and 17% strucutres of the total conformations of MD trajectory respectively.

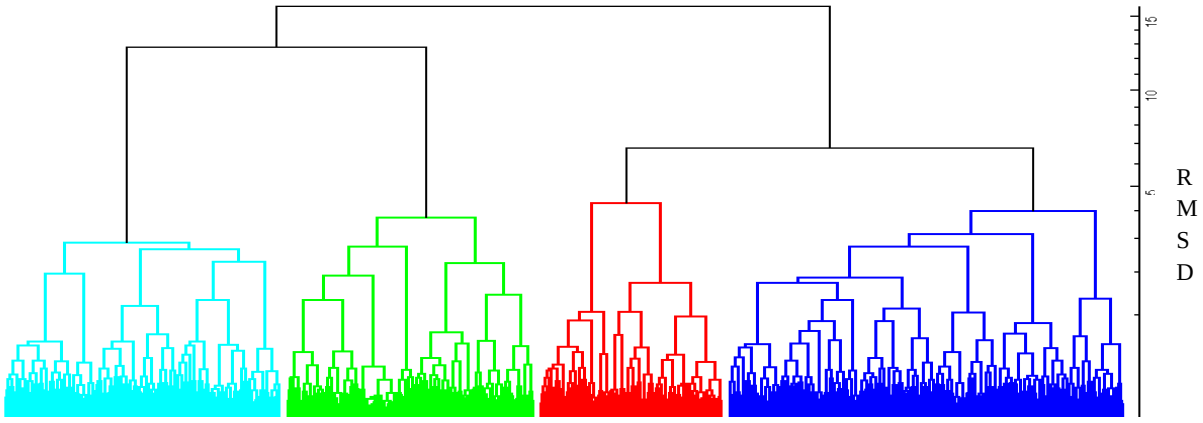


Figure 3.9 Dendrogram showing different clusters for the Aurein 2.4 in the MD trajectory classified using the Ward's clustering method [**124**]. Different number of clusters can be ascertained by considering different cutoff values of the RMSD across Y-axis.

Table 3.1: The results of cluster analysis for four major clusters identified in the MD simulation of Aurein 2.4.

Sr. No.	Cluster Number	Number of Structures	% of Structures	RS Number in cluster
1	C1	1799	36	1619
2	C2	1122	22	1903
3	C3	1253	25	1687
4	C4	826	17	1124

The distribution of the conformations in each cluster became more clear when the incidence of each cluster was plotted as a function of MD simulation time, and is depicted in Figure 3.10. As it can be seen, the first two major clusters, C1 and C3, start adding conformations immediately from the start of the MD simulation, and continue to enrich the new conformations during the sampling process throughout the progress of the MD trajectory. The remaining clusters C2 and C3, on the other hand, start adding conformations preferably after few nanoseconds of the MD trajectory and added new structures randomly as the MD trajectory proceeded.

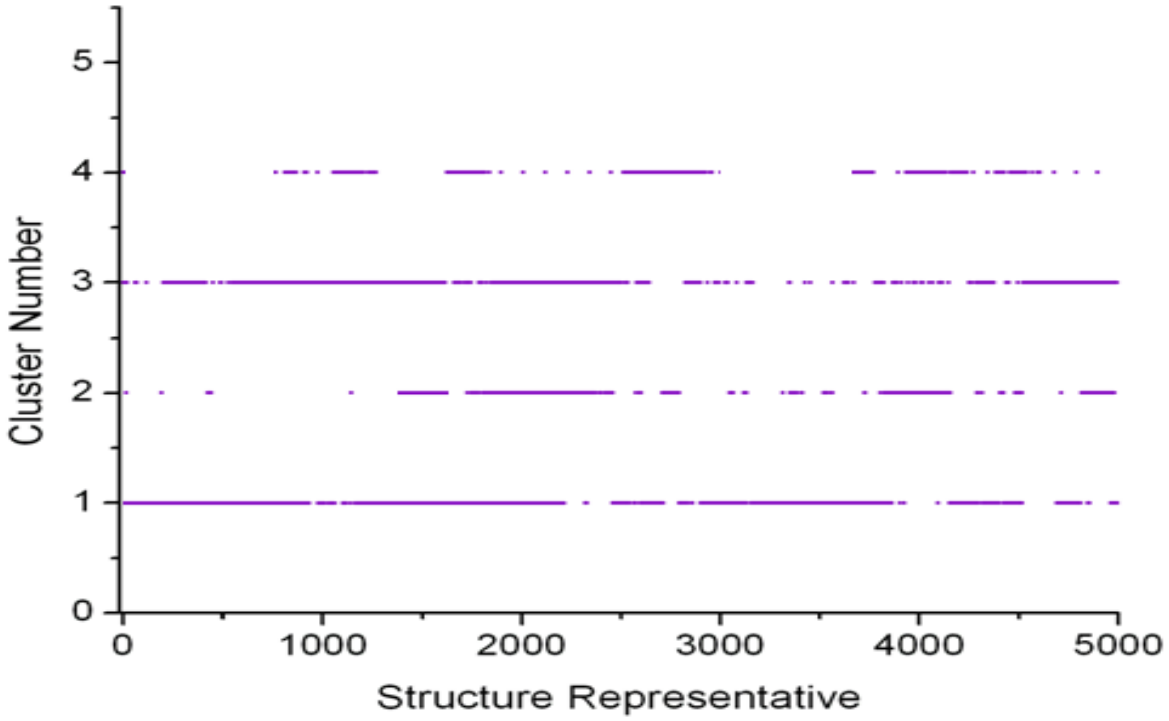
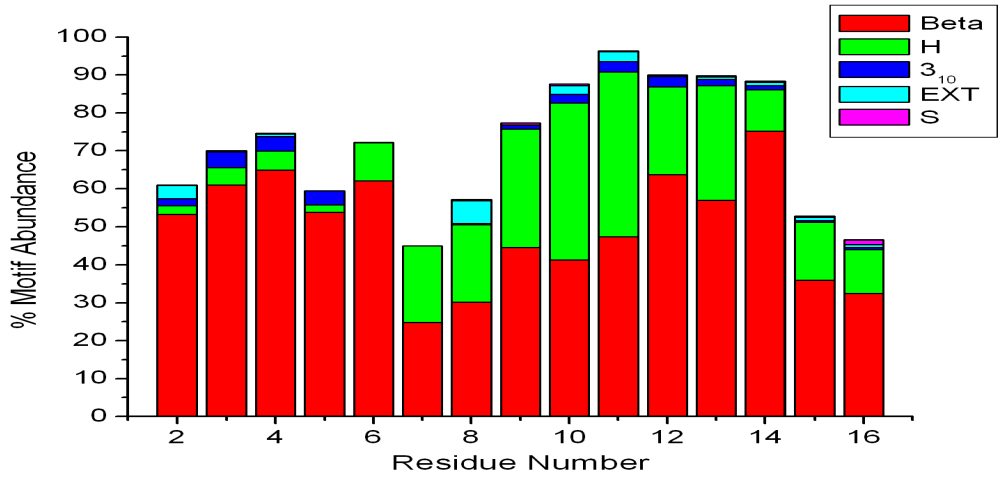


Figure 3.10 Evolution of four clusters during the progress of MD for representative structures (RS) obtained using Ward's clustering method [**124**]. The greater number of closely related structures in each cluster can be seen as solid lines.

A qualitative analysis of the secondary features was subsequently performed on the conformations of each cluster using the CLASICO program [**115**]. The histograms showing the statistics of the secondary motifs of each residue for Aurein 2.4 conformations in four clusters are diagrammatically represented in Figures 3.11a-b and 3.12a-b. A closer inspection of Figure 3.11a revealed that all residues of Aurein 2.4 conformations in cluster 1 (C1) adopted α-helicity to a certain extent preferably from the central region (Val^{7}-Gly^{16}) of the peptide extending towards the C-terminus. However, these folded conformations were not stable enough and were in a rapid equilibrium with the unfolded structures (β-turns). To some extent, 3_{10}-α-helicity between residues Gly^{2}-Asp^{5} and Lys^{9}-Ala^{15} was also observed. The characterization of secondary motifs was

(a)



(b)

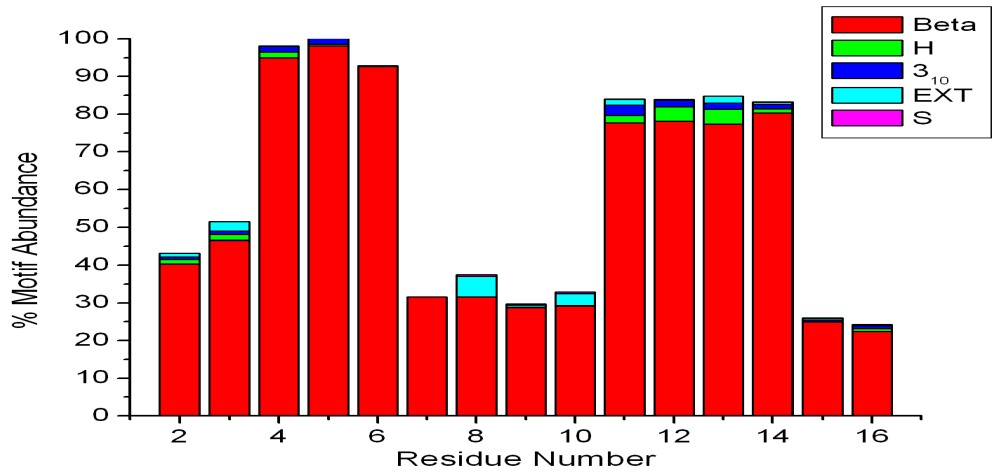
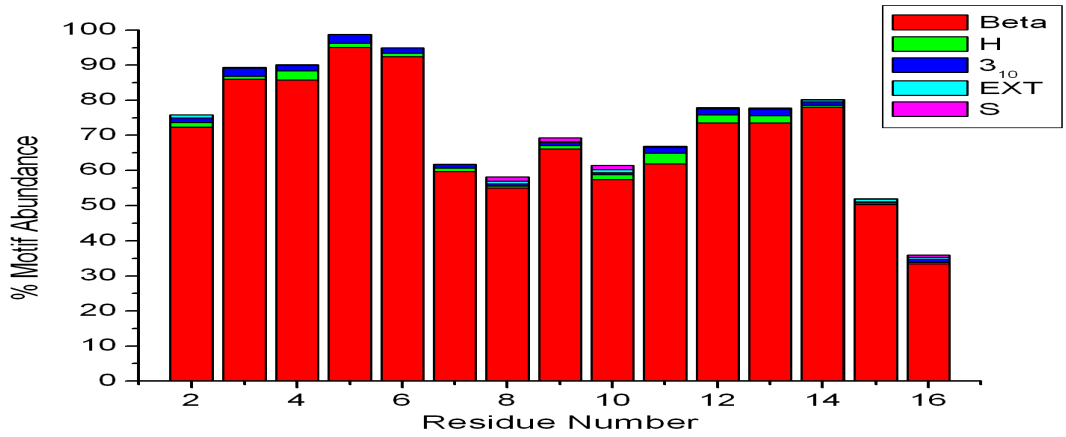


Figure 3.11 Secondary motifs attained by different residues of conformations in cluster C1 and C2. Conformational motifs are labelled: Beta (β-turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

based on the type of dihedral angles adopted by the peptide residues, as depicted in Table 2.1 in Chapter 2. The conformations in second most abundant cluster (C2) adopted the beta turns (Figure 3.11b) as their preferred secondary motif. To some

(a)



(b)

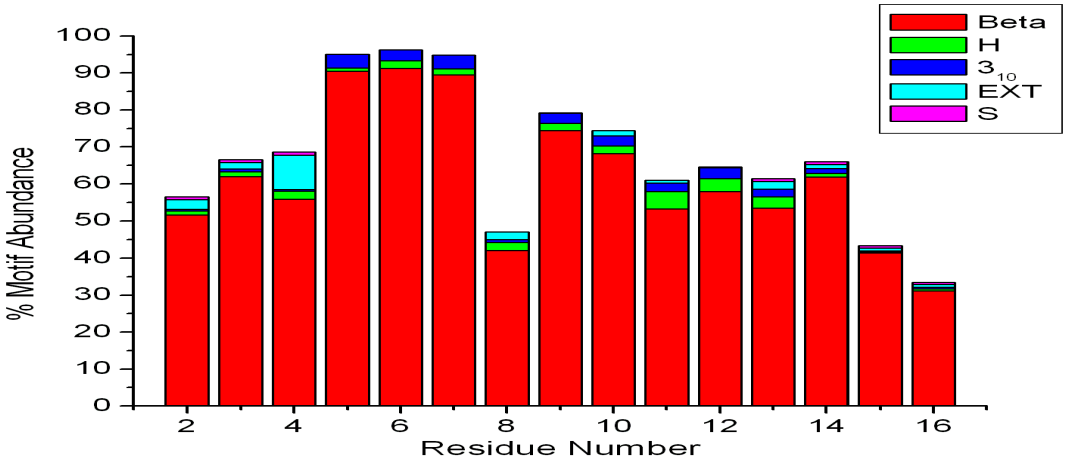


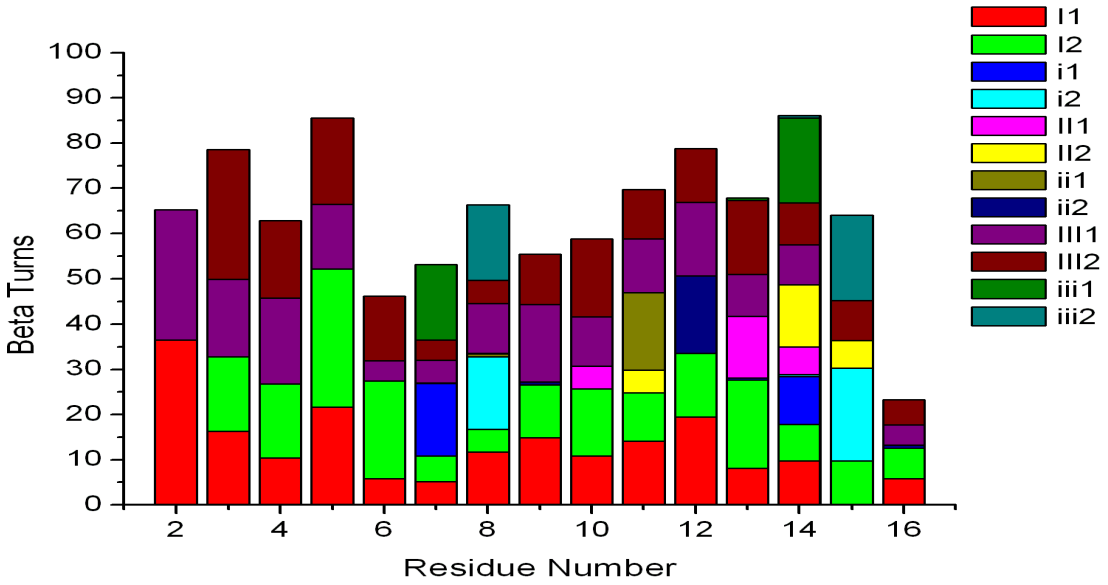
Figure 3.12 Secondary motifs attained by different residues of conformations in cluster C3 and C4. Conformational motifs are labelled: Beta (β-turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

extent, the α-helicity and 3_{10}-α-helical secondary motifs supported by residues Gly^{2}-Asp^{5} and Val^{11}-Leu^{14} were also observed in the same cluster. The remaining clusters, C3 (Figure 3.12a) and C4 (Figure 3.12b) were also occupied by the similar conformations showing the predominance of beta turns as their preferred secondary

features although the α-helical regions flanked by residues Leu^{3} and Thr^{13} were also observed in lower percentages.

Considering the abundance of β-turns in the sampled conformations, it was thought worthwhile to classify them according to their types using the CLASICO program [**115**], and are pictorially shown in the Figures 3.13 and 3.14. The assignment of different types of β-turns was based on the classification as described in Table 2.2 (see Chapter 2). A closer inspection of Figure 3.13a revealed that the conformations of the most abundant cluster (C1), adopted preferably β-turn type I in their residues although in different percentages. To some extent, β-turn type III between residues Gly^{2}-Leu^{14} was also observed. In case of cluster C2 (Figure 3.13b), the β-turn type I was present in the majority of the residues with higher propensity between residues Gly^{2}-Ile^{6} and Lys^{9}-Leu^{14} along with β-turn type i (mirror image of β-turn type I) between residues Val^{7}-Lys^{8} and Leu^{14}-Ala^{15} in lower percentages.

(a)



(b)

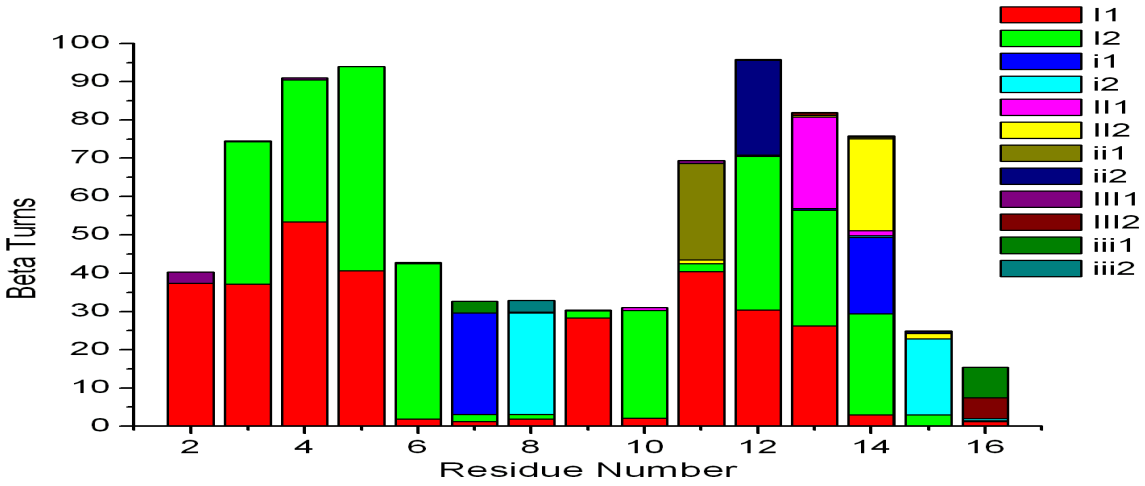
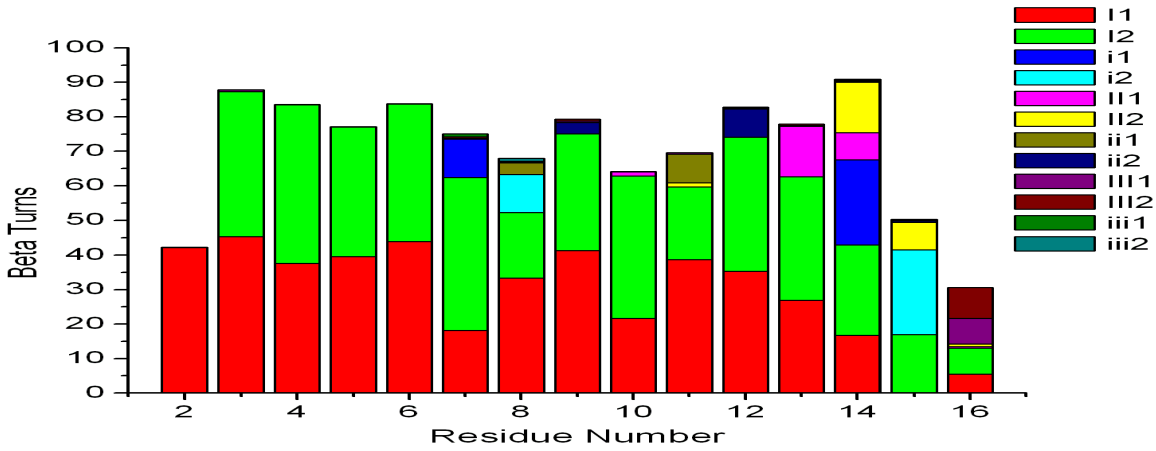


Figure 3.13 Different types of β-turns attained by the residues of peptide Aurein 2.4 in clusters (a) C1 (b) C2, obtained using the CLASICO program [**115**].

(a)



(b)

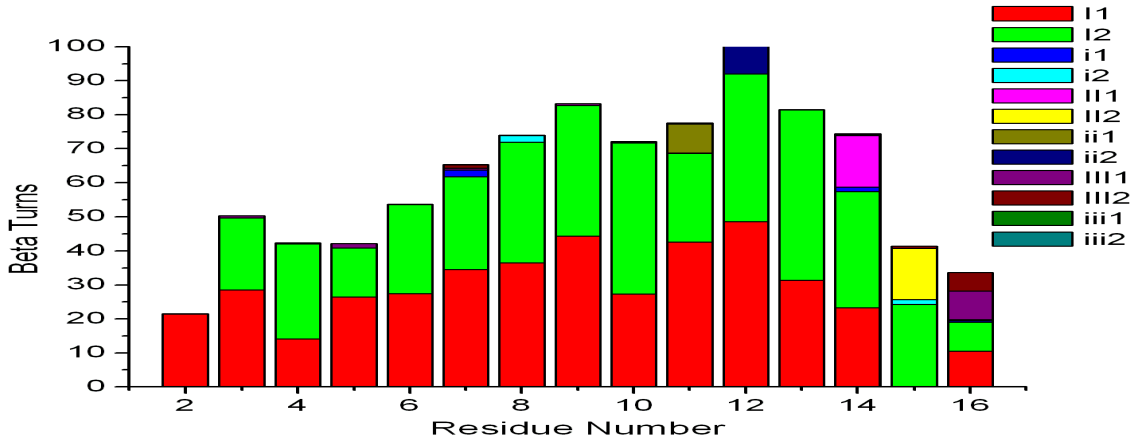
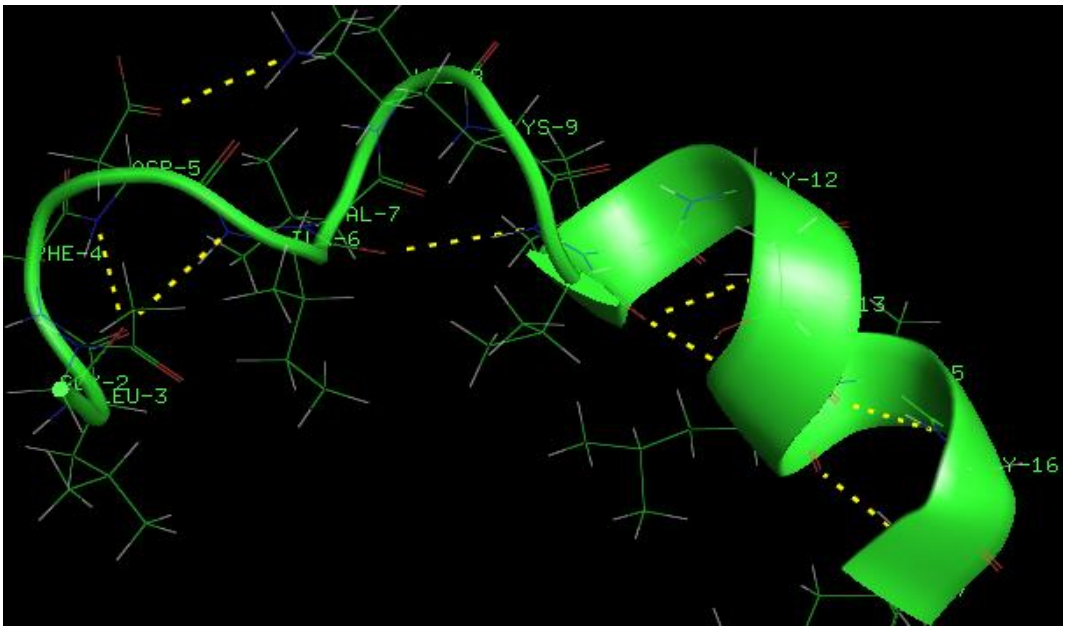


Figure 3.14 Different types of β-turns attained by the residues of peptide Aurein 2.4 in clusters (a) C3 (b) C4, obtained using the CLASICO program [**115**].

The conformations in Cluster 3 (Figure 3.14a) and Cluster 4 (Figure 3.14b) exhibited predominantly β-turn type I in their all residues. To a certain extent, β-turn type i between residues Val^{7}-Lys^{8} and β-turn type II between residues Thr^{13}-Ala^{15} in cluster 3, and β-turn type II between residues Leu^{14}-Ala^{15} in cluster 4 was also observed.

Figures 3.15a-b and Figures 3.16a-b depict the representative structures (RS) of each cluster showing important interactions between the amino acid residues. Clearly the RS of the largest cluster, C1 (Figure 3.15a) adopts a well-defined α-helicity flanked by the residues Val^{10}-Gly^{16}. This α-helical region was supported by two concurrent hydrogen bonds between the backbone oxygen atom (C=O) of the Val^{10} and the backbone nitrogen atom of Leu^{14}, one hydrogen bond between the oxygen atom of Thr^{13} and the nitrogen atom of Gly^{16}, and one hydrogen bond between Leu^{14} and amine nitrogen. A distorted α-helical region supported by the intramolecular hydrogen bonds between the residues Gly^{2}-Ile^{6}, Asp^{5}-Lys^{8} and Ile^{6}-Val^{10}, and a β-turn between the residues Gly^{2}-Asp^{5} was also observed. The RS for C2 (Figure 3.15b) showed the presence of a hydrogen bond between the carbonyl oxygen of Val^{10} and Leu^{14} accounting for an α-helical region between these residues. A reverse turn as evidenced by a single intramolecular hydrogen bond between the side chains of Lys^{9} and Thr^{13} was also observed. However, this structure remained in the unfolded form. For cluster C3 (Figures 3.16a), the RS exhibited a distorted α-helical region between the residues Phe^{4}-Lys^{8} supported by two intramolecular hydrogen bonds. Additionally, four β-turns between the residues Gly^{2}-Asp^{5}, Phe^{4}-Val^{7}, Asp^{5}-Lys^{8} and Thr^{13}-Gly^{16} were also present in the structure. The RS in C4 adopted two well-defined α-helical regions between residues Ile^{6}-Lys^{9} and Gly^{12}-Ala^{15} intervened by unstructured valine residues (Val^{10}-Val^{11}), as depicted in Figures 3.16b. The N-terminal residues (Gly^{2}-Asp^{5}) in RS, on the other hand, did not adopt any secondary feature and preferred to stay either in the extended or unfolded conformation.

(a)



(b)

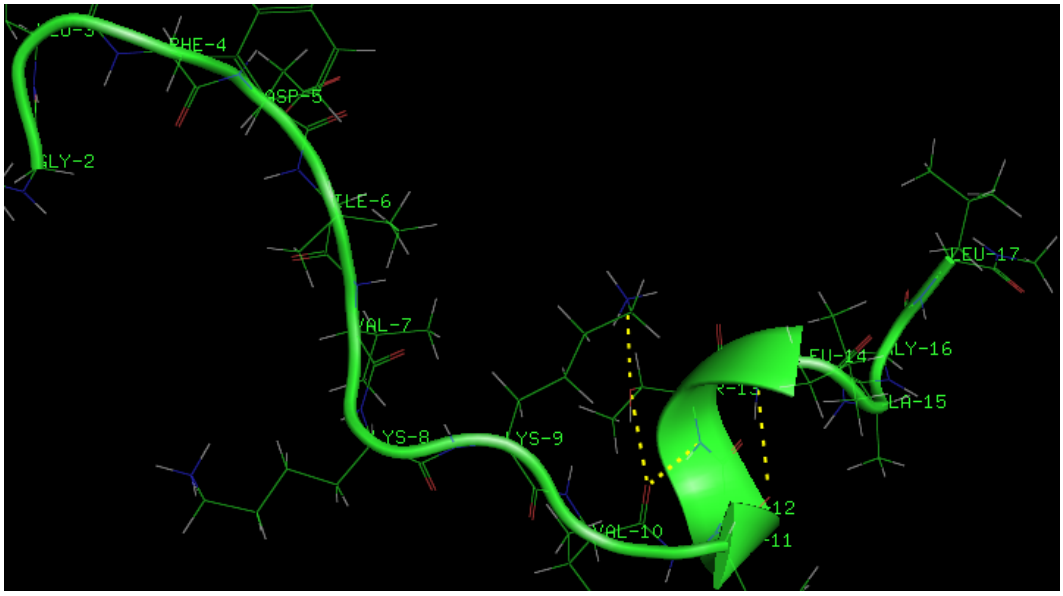
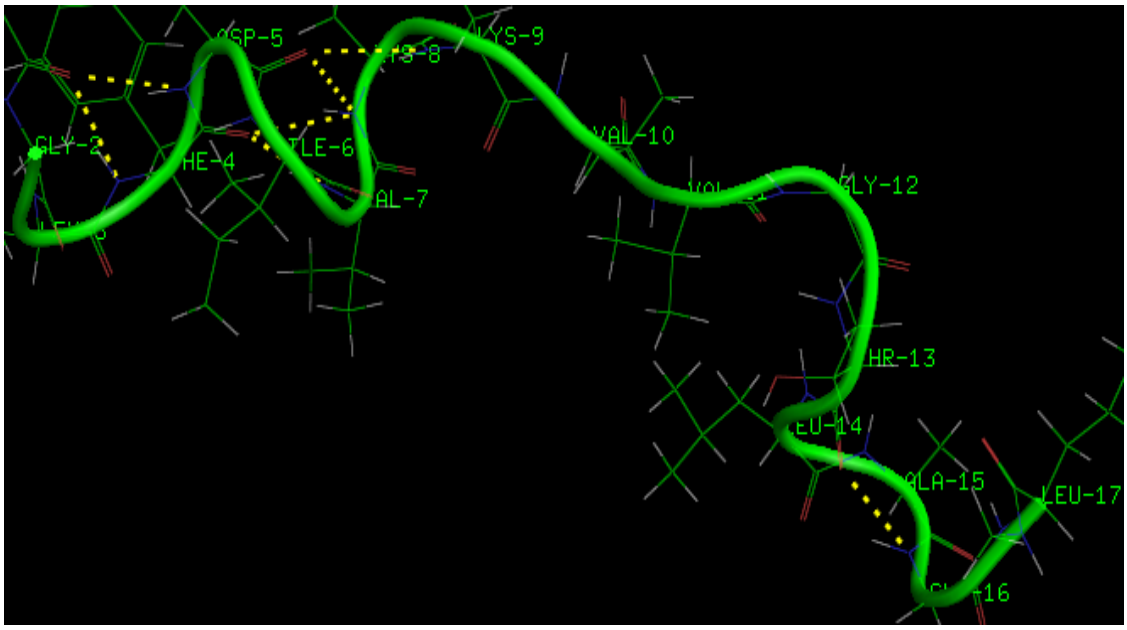


Figure 3.15 Representative structures of cluster 1 (a) and cluster 2 (b) showing important interactions between the residues, for Aurein 2.4. The side-chain atoms are presented in line format, while the back bone atoms are depicted in ribbon form. Hydrogen bonds are shown as yellow dotted lines.

(a)



(b)

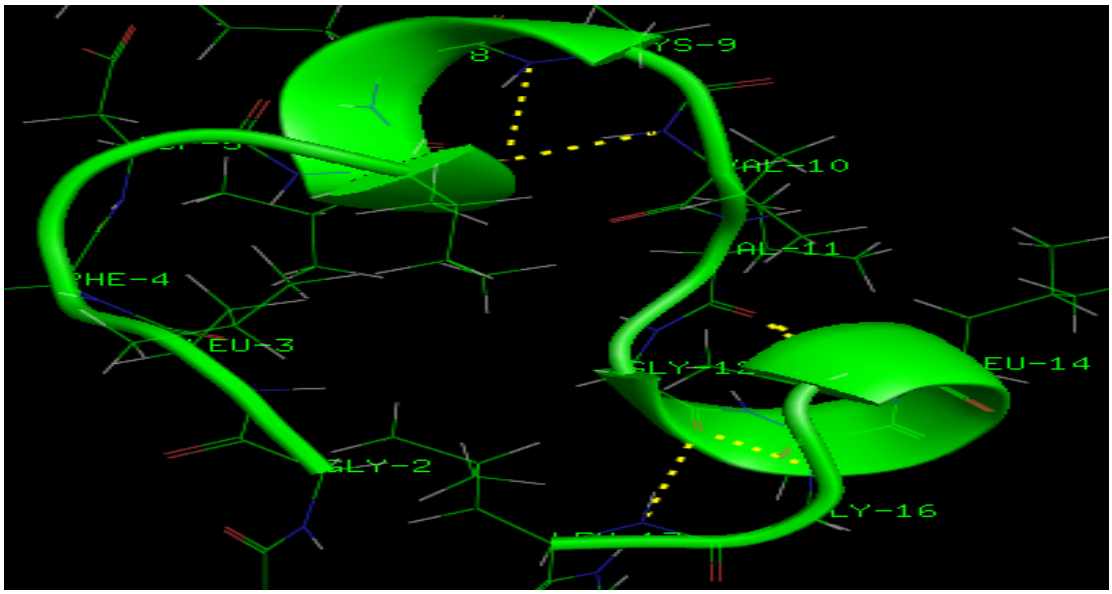


Figure 3.16 Representative structures of cluster 3 (a) and cluster 4 (b) showing important interactions between the residues, for Aurein 2.4. The side-chain atoms are presented in line format, while the back bone atoms are depicted in ribbon form. Hydrogen bonds are shown as yellow dotted lines.

Hydrogen bonding has been known to play a significant role in the stability and biological functions of macromolecules. In view of its importance in the structure of peptides and proteins, the hydrogen bond analysis was subsequently performed on the largest cluster, C1 as the conformations of this cluster exhibited the α-helical characteristics. Figure 3.17 represents the appearance/disappearance of hydrogen bonds in the sampled conformations of C1 during the progress of the MD trajectory, whereas the detailed results including percentage of hydrogen bonds are summarized in Table 3.2. It should be noted that the first percentage value specified in Table 3.2 corresponds to the first hydrogen bond distance depicted in Figure 3.17.

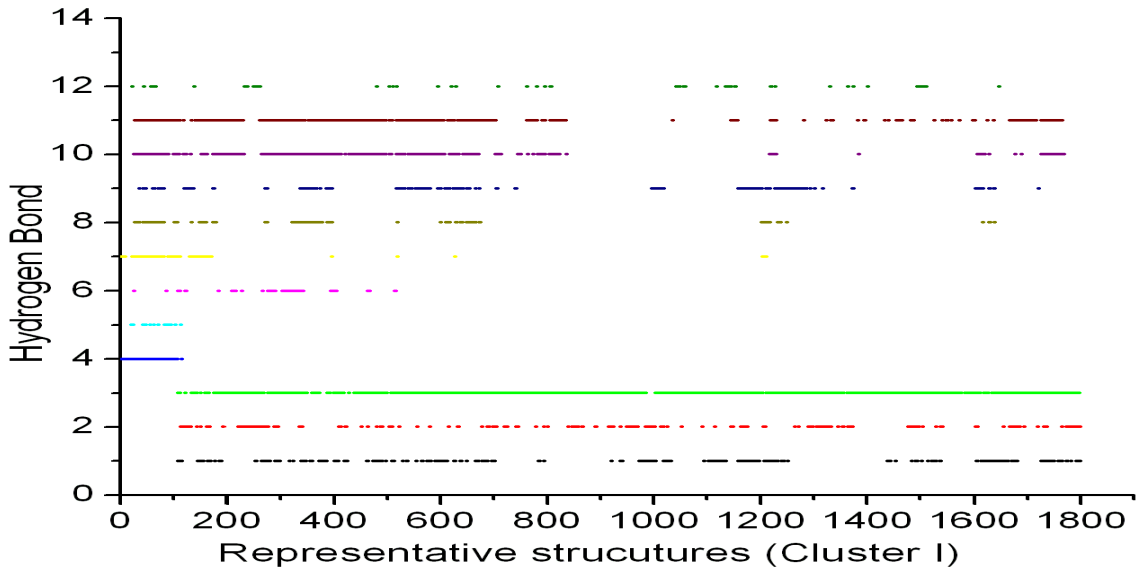


Figure 3.17 Progress of intramolecular hydrogen bonds observed between residues in the sampled conformations of cluster 1 (C1).

The hydrogen bonds present in more than 1% were only considered in the analysis. The geometrical criterion used in hydrogen bond analysis involved the donor (A)-acceptor (B) distance <3.0 Å (sum of van der Waals radii) and the angle AHB ≥ 120°, where A

was considered as a donor (i.e. N-H) and B as an acceptor (i.e. O=C). A closer inspection of Table 3.2 revealed that an α-helical region between residues Gly^{2}-Lys^{9} was present in majority of the conformations of cluster 1 with a high propensity between residues Phe^{4}-Lys^{8}. The percentage of hydrogen bonds in the central residues (Ile^{6}-Val^{11}) was comparatively less and was present only in the initial few conformations (entries 4, 5, Figure 3.17) of the cluster. Although, the number of hydrogen bonds responsible for α-helicity between the residues Lys^{8}-Leu^{14} (entries 7-9, Figure 3.17) was slightly increased, their progress was not uniform and was interrupted at several places during the sampling process. The α-helical region flanked by residues Val^{11}-Gly^{16} was comparatively stable and was adopted by most of the conformations in cluster as evidenced by higher percentages of hydrogen bonds (entries 10-11, Figure 3.17) in this region. Only a few conformations exhibiting the hydrogen bond interactions between the residues Thr^{13}-Leu^{17} were observed in the cluster. Overall, the present results indicated the distribution of α-helical character, although in different percentages, on all residues of the peptide. However, the folded conformations were not considerably stable and were in a rapid equilibrium with the unfolded conformations throughout the sampling process.

Table 3.2: The α-helical features observed due to backbone-backbone hydrogen bond interactions and their percentages in the largest cluster of MD for Aurein 2.4.

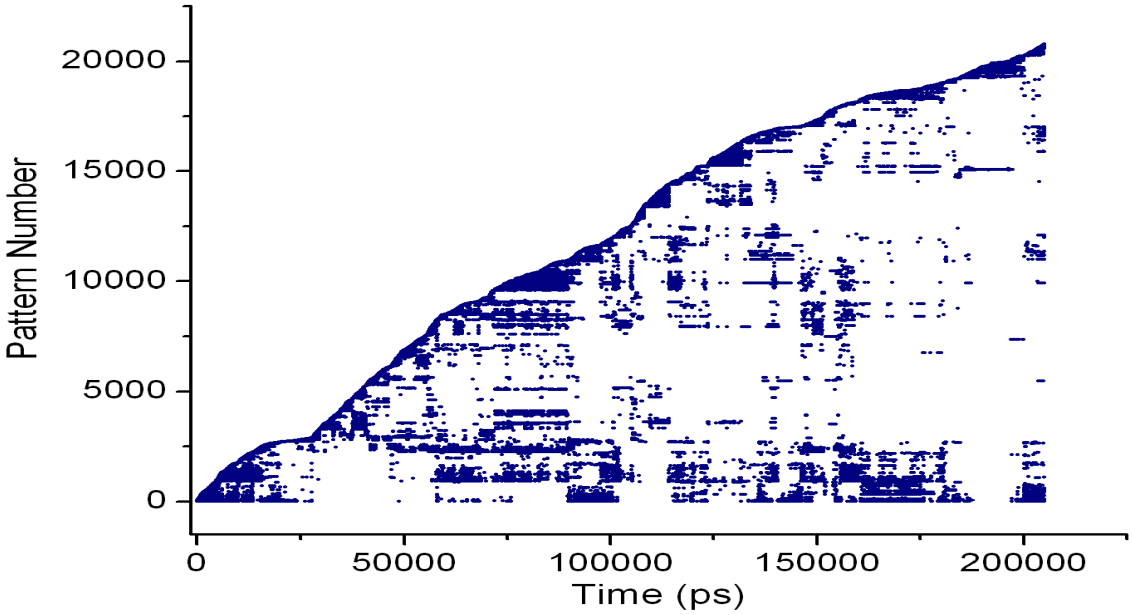
Entry	Donor-acceptor	Cluster I
1	$(Gly^{2})O\ldots N(Ile^{6})$	13.2 %
2	$(Leu^{3})O\ldots N(Val^{7})$	13.3 %
3	$(Phe^{4})O\ldots N(Lys^{8})$	67.0 %
4	$(Asp^{5})O\ldots N(Lys^{9})$	5.0 %
5	$(Ile^{6})O\ldots N(Val^{10})$	1.5 %
6	$(Val^{7})O\ldots N(Val^{11})$	3.2 %
7	$(Lys^{8})O\ldots N(Gly^{12})$	5.6 %
8	$(Lys^{9})O\ldots N(Thr^{13})$	6.6 %
9	$(Val^{10})O\ldots N(Leu^{14})$	12.3 %
10	$(Val^{11})O\ldots N(Ala^{15})$	25.3 %
11	$(Gly^{12})O\ldots N(Gly^{16})$	28.5 %
12	$(Thr^{13})O\ldots N(Leu^{17})$	3.5 %

3.6 Molecular dynamics study of Aurein 2.4 in explicit solvents

Literature studies have revealed that solvents play a pivotal role in decreasing the free energy difference between the conformations that differ by large energies on the vacuum surface, and thus lower the energy barrier separating these conformations. Consequently, the structure, stability and biological functions of the proteins and peptides are severely affected in the presence of different solvent environments. For

instance, a remarkable change in the conformational stability of the alanine dipeptide was observed when MD simulations performed in the gas phase were compared with those obtained in the solvent phase [**125**]. Considering the signficance of solvent molecules in the structure and function of peptides, it was thought worthwhile to investigate the conformational profile of Aurein 2.4 in the presence of explcit solvent molecules. Water and methanol solvents were selected in the present study as they are well known to mimic the dynamic behaviour of peptides under physiological conditions. Accordingly, two MD simulations of length 200 ns each were performed on Aurein 2.4 in water (MD^{wat}) and methanol (MD^{meth}), and 200,000 snapshots were stored from each trajectory. First, the efficiency of both MD trajectories were monitored by plotting the evolution of pattern number as a function of simulation time, and is shown diagrammatically in Figure 3.18a-b. Total number of patterns identified in case of trajecotries MD^{wat} and MD^{meth} were 20781 and 19551, respectively, clearly suggesting the lesser restrictive nature of the former in searching new conformational motifs. As it can be seen in Figure 3.18a that the appearance of new patterns in MD^{wat} started immediately after the start of the simulation and proceeded in a regular fashion untill 16 ns. Although, a small hinderance (plateau formation) in search of new patterns was observed between 16-28 ns segment of the trajectory, new patterns were explored in an uniform manner throughout the remaining section of the MD simulation. In case of methanol (Figure 3.18b), just like water, new patterns were sampled from the start of the trajectory (MD^{meth}), however, plateau formation was observed in the segments 8-15 ns, 19-24 ns, 47-49 ns, 97-99 ns and 145-147 ns of the simulation, again suggesting the restrictive nature of methanol compared to the water solvent in sampling new motifs.

(a)



(b)

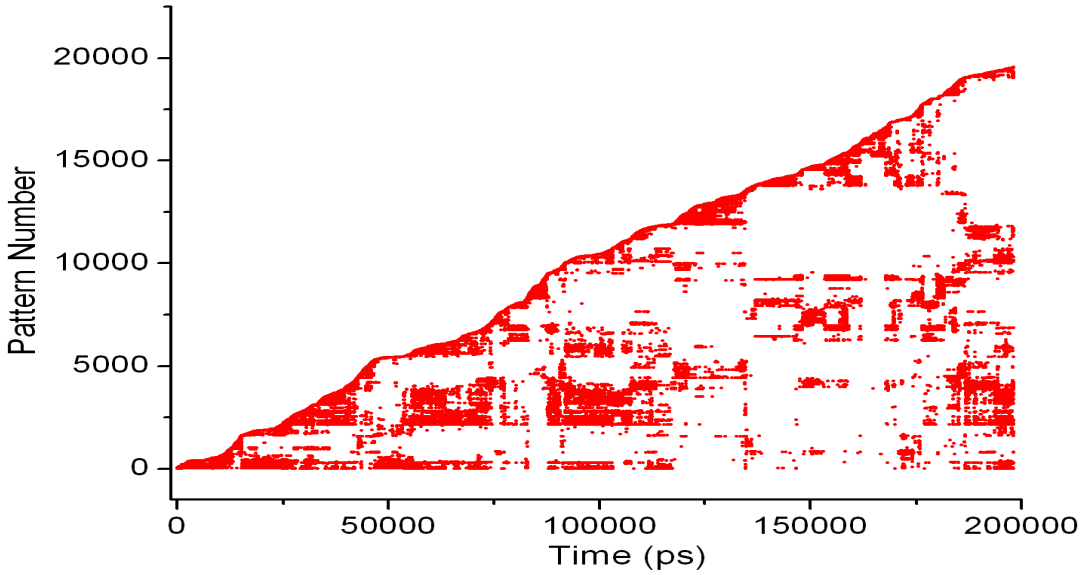
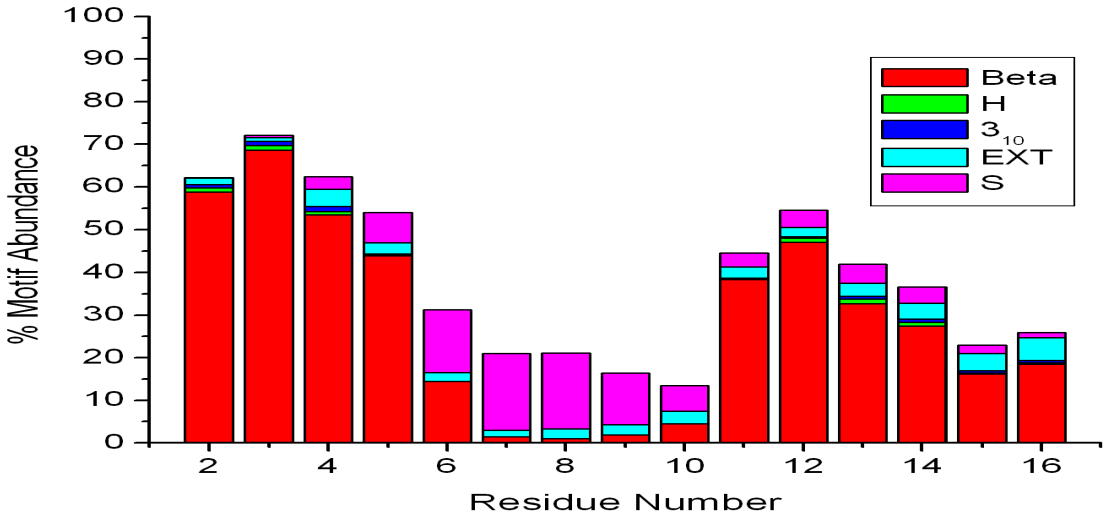


Figure 3.18 Evolution of new patterns during the progress of MD trajectory in (a) water (b) methanol, obtained using the CLASICO program [**115**]. Each pattern represents the combination of conformational motifs.

In order to get a deeper understanding of the secondary structure features of the conformations, both MD trajectories (MD^{wat} and MD^{meth}) were analyzed using the three residue window of CLASICO program [**115**]. Histograms showing different types of conformational motifs and their percentages obtained in water and methanol solvents are pictorially depicted in Figure 3.19a and Figure 3.19b respectively.

A closer inspection of Figure 3.19a revealed that majority of the conformations of Aurein 2.4 in water adopted β-turns with high propensity between residues Gly^2-Ile^6 and Val^{11}-Gly^{16}, whereas low propensity in the central residues (Val^7-Val^{10}). To some extent, β-strands between residues Phe^4-Gly^{16} with strong propensity between residues Ile^6-Lys^9 were also observed in the sampled conformations. Both β-turns and β-strands were preferred secondary motifs of the conformations in methanol solvent as depicted in Figure 3.19b. β-turns were predominantly adopted by N-terminal (Gly^2-Asp^5) and C-terminal residues (Val^{11}-Gly^{16}) in methanol similar to those observed in water. The β-strands, on the other hand were localized in the central residues (Ile^6-Val^{10}) in higher percentages. No α-helical character was observed in the residues of Aurein 2.4 peptide in these solvents.

(a)



(b)

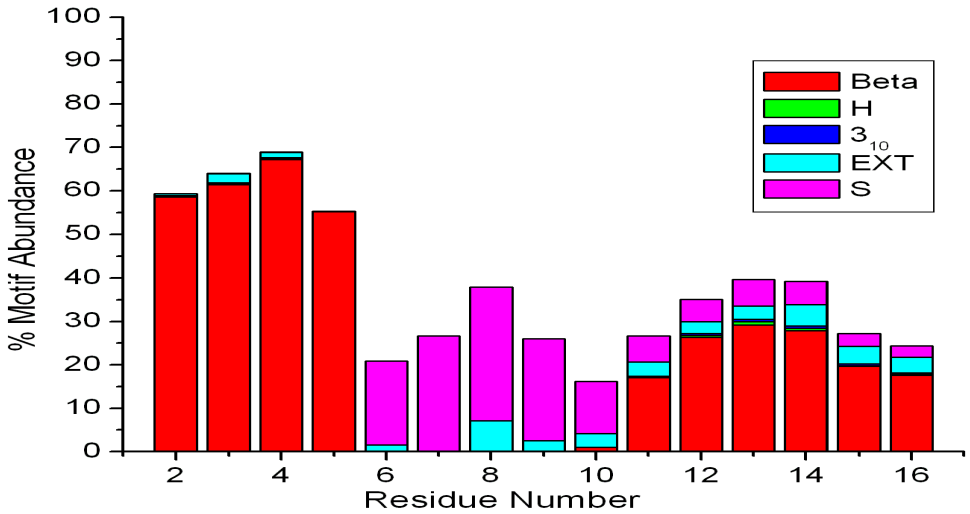
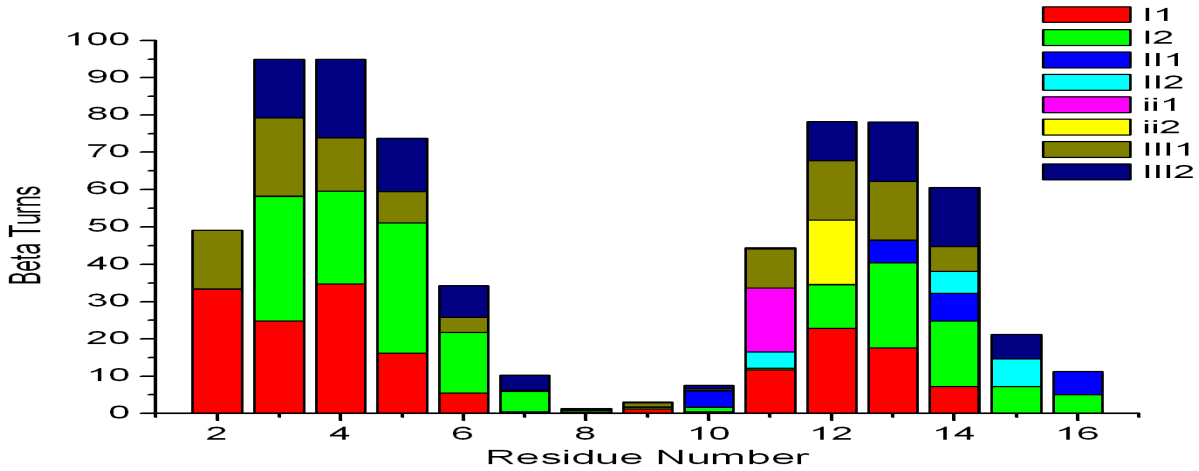


Figure 3.19 Different secondary motifs attained by different residues of conformations in trajectory (a) MD^{wat} (b) MD^{meth}. Conformational motifs are labelled: Beta (β-turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

Different types of β-turns were further identified for the conformations sampled in trajectories MD^{wat} and MD^{meth} using the CLASICO program [**115**], and are pictorially depicted in Figure 3.20a and Figure 3.20b, respectively. Clearly, β-turn type I and β-turn type III were the preferred secondary motifs adopted by the amino acid residues of the conformations in water, as shown in Figure 3.20a. Specifically, β-turn type I was adopted by N-terminal residues (Gly^2-Asp^5) in higher percentages, although, residues Ile^6-Val^7 and Val^{11}-Gly^{16} also exhibited similar type of β-turn in lower percentages. Similarly, β-turn type III was also flanked by N-terminal residues (Gly^2-Ile^6) and C-terminal residues (Val^{11}-Leu^{14}) in varying percentages. To some extent, β-turn type II between residues Val^{10}-Val^{11} and Thr^{13}-Gly^{16}, and its mirror image (β-turn type ii) between residues Val^{11}-Gly^{12} was also observed. The central lysine residues (Lys^8-Lys^9), on other hand, did not show any characteristic secondary motif and preferred to stay either in extended or some other unfolded form (Figure 3.20a). Almost similar β-turn profile showing the predominance of β-turn type I and β-turn type III was observed in the methanol solvent (Figure 3.20b). For instance, both types of β-turns (type I and III) were located in the N-terminal (Gly^2-Asp^5) and C-terminal (Val^{10}-Gly^{16}) residues of the conformations. The β-turn type II between residues Val^{10}-Val^{11} and Leu^{14}-Ala^{15} and its mirror image (type ii) between residues Val^{11}-Gly^{12}, however, was observed in lower percentages in methanol than in water (Figure 3.20b).

(a)



(b)

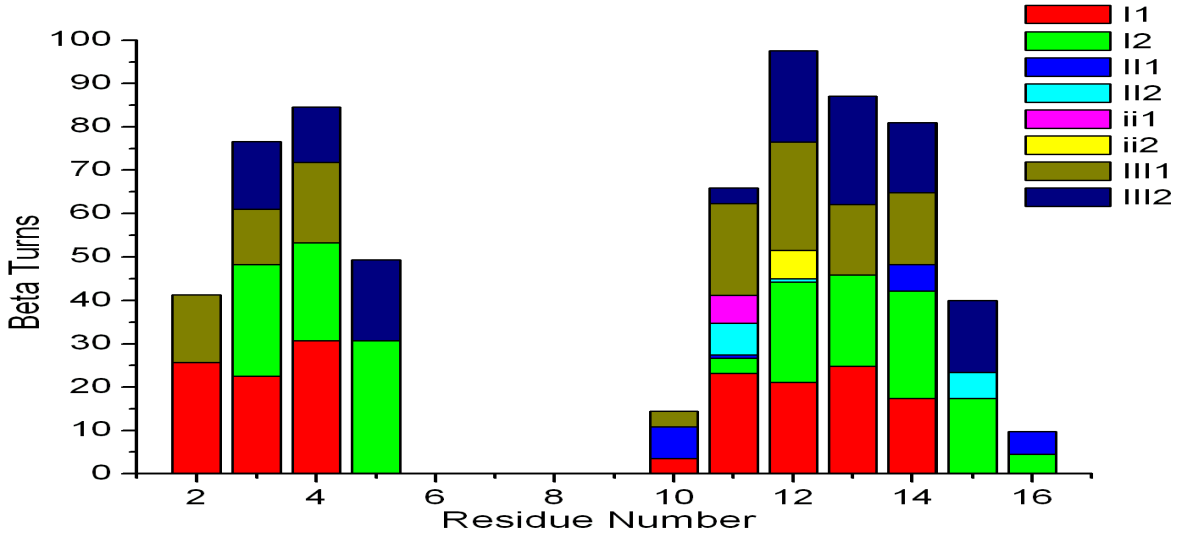


Figure 3.20 Different types of β-turns attained by the residues of peptide Aurein 2.4 in trajectories (a) MD^{wat} (b) MD^{meth}, obtained using the CLASICO program [**115**].

3.7 Conclusions

First, a validation study was performed by exploring the conformational profile of Aurein 2.3 using the MD method and the OBC implementation of the GB method to describe the solvent. The MD results obtained suggested that all amino acid residues of Aurein 2.3 adopt α-helicity, in agreement with the experimental results reported in literature [65]. Considering the similar biological relevance and structural resemblance, the same computational procedure was further employed to explore the conformational profile of Aurein 2.4, a structural analogue of Aurein 2.3. The results obtained revealed that the peptide adopts α-helical character in all its residues although the extent of helicity was more in the central residues extending towards its C-terminus. Moreover, the peptide has a tendency to attain folded, unfolded and helical conformations due to the low energy barrier between them, and therefore accounts for the higher flexibility of Aurein 2.3. The MD results performed under explicit solvent conditions indicated that the peptide has no specific folded conformation in the polar solvents, and prefer to stay either in the form of β-turn or β-strand.

CHAPTER 4

CONFORMATIONAL PREFERENCES OF AUREIN 2.5

In this chapter, the folding characteristics of Aurein 2.5 obtained using the exploration of the conformational preferences in implicit and explicit solvent models (water and methanol) using MD method, has been presented. The MD results suggest the presence of α-helical region distributed on its majority of the residues in equilibrium with β-turns due to low conformational energy barrier between them. The peptide remains un-structured in the polar solvents (water and methanol).

4.1 Introduction

Aurein 2.5 (Figure 4.1) is a 16 amino acid residue peptide that belongs to the Aurein family of anti-microbial peptides. The only difference between Aurein 2.5 and Aurein 2.3 is a single conservative mutation of a isoleucine to a phenylalanine at position 13. However, the anti-cancer activity of Aurein 2.3 is significantly altered by this single mutation [**63**]. In order to get a deeper understanding of the 3D structural features of Aurein 2.5, a 300 ns MD simulation was performed under implicit solvent conditions using the procedure described in the methods section in Chapter 3. For comparison purpose, the conformational profile of Aurein 2.5 was also explored in polar solvents (water and methanol) using MD method.

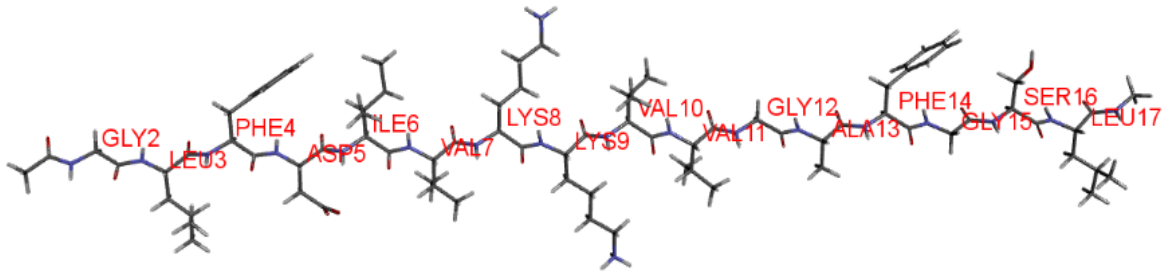


Figure 4.1 The extended conformation of Aurein 2.5 with three letter code for its amino acid sequence.

4.2 Results and Discussion

Initially, the atomic fluctuations in the conformations of the MD trajectory were roughly monitored by plotting its RMSD relative to the initial structure. Figure 4.2 represents the RMSD of the sampled conformations computed by considering their backbone atoms as well as all atoms, respectively. As expected, the side chains of the peptide experienced larger oscillations (in red, Figure 4.2) than its backbone atoms (in black, Figure 4.2) during the simulation process. The average fluctuations around 6.5 Å (for backbone atoms)

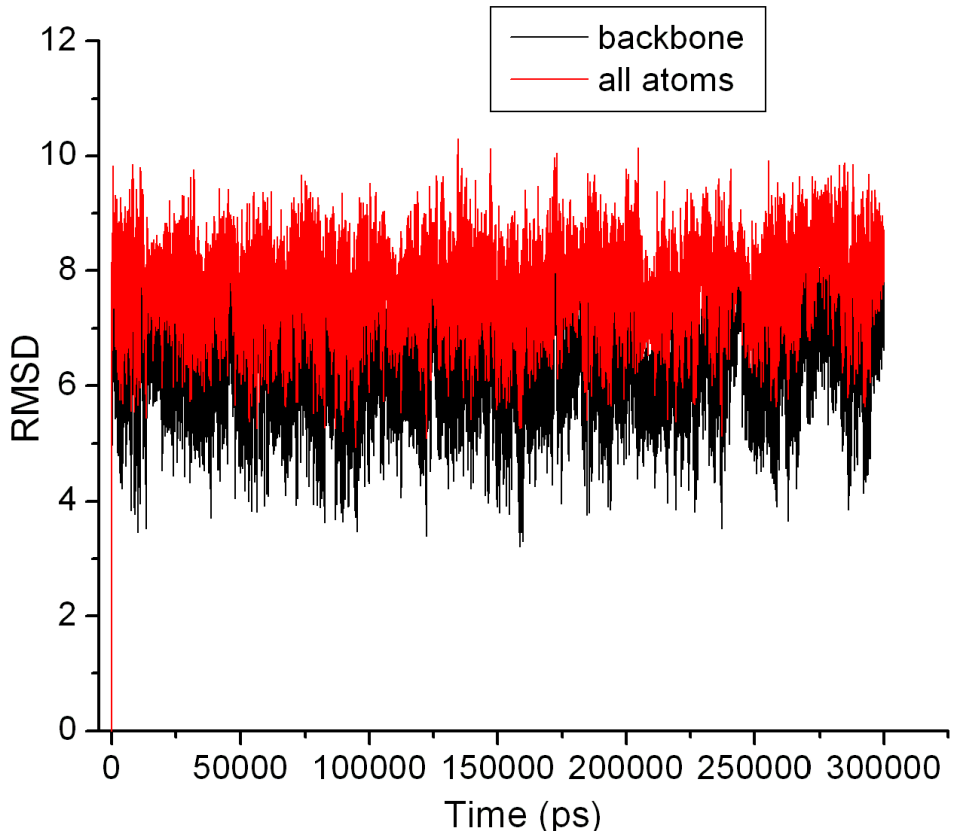


Figure 4.2 Root mean square deviation (in Å) of the sampled conformations relative to their backbone atoms (in black) and all atoms (in red).

and approximately 7.9 Å (for all atoms) further suggested that the significant deviations of the sampled conformations compared to their extended starting structure were probably be due to the sufficient folding of the peptide after few nanoseconds (ns) of the trajectory. However, the continuous RMSD fluctuations observed (Figure 4.2) indicated the unstable nature of the folded structures probably due to their rapid swapping with the unfolded structures. For every snapshot of the MD trajectory, the conformational patterns were identified using the CLASICO program [**115**] described in Chapter 2. The development of new patterns during the progress of the MD trajectory is pictorially depicted in Figure 4.3. After 300,000 snapshots, a total of 65873 new patterns were

identified for Aurein 2.5 peptide. A closer inspection of Figure 4.3 revealed that the appearance of patterns initiated at the beginning of the trajectory and proceeded in a uniform fashion until 74 ns. A slight reduction in the sampling of new patterns was observed between the 74-123 ns segment of the trajectory, followed by a regular increase until the end of the simulation process. Random trapping of the conformations was also observed throughout the progress of the MD trajectory as evidenced by the dark regions in the Figure 4.3.

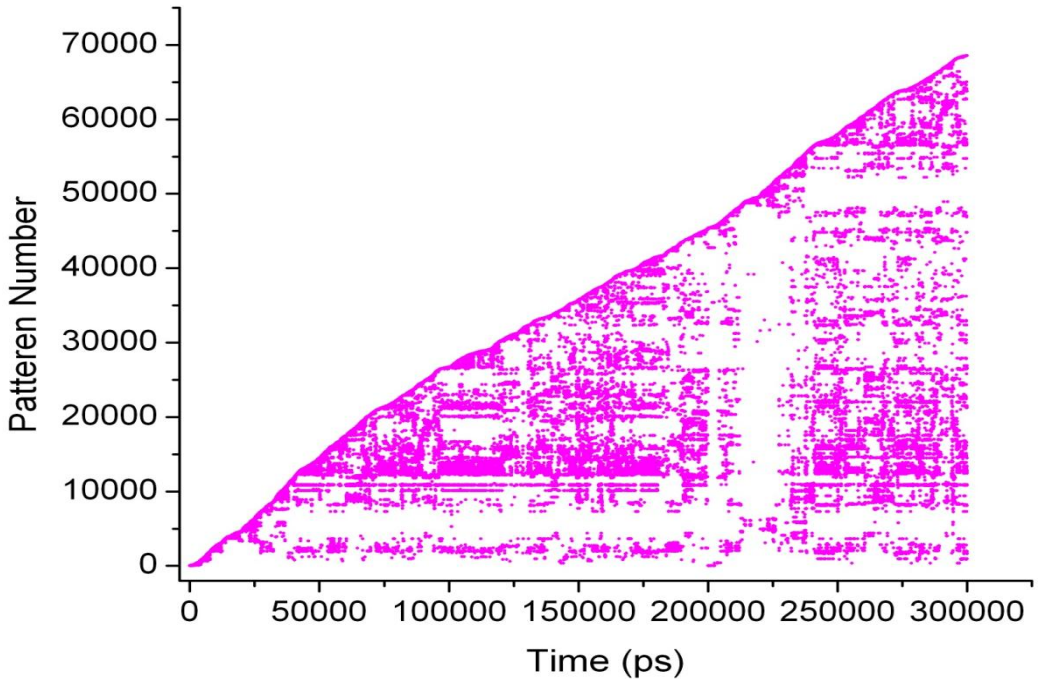


Figure 4.3 Evolution of new patterns during the MD simulation process for peptide Aurein 2.5, obtained using the CLASICO program [**115**].

The sampled conformations of the MD trajectory were further grouped into clusters taking their structural similarity into account, using the Ward's clustering method [**124**]. For this purpose, 5000 representative strucutres were selected from a total of 300,000 strucutres by sampling each conformation at the intervals of 60 ps. Four major clusters

were identified by selecting a RMSD cutoff around 7, as shown on the dendrogram in Figure 4.4 along with detailed results in Table 4.1. A closer inspection of Table 4.1 revealed that C3 was the most abundant cluster containing 49% of the total sampled conformations. The second most populated clusters, C1 and C2 acquired 20% and 19% conformations respectively, while the least abundant cluster, C4 contributed only 12% of the structures, as depicted in Table 4.1.

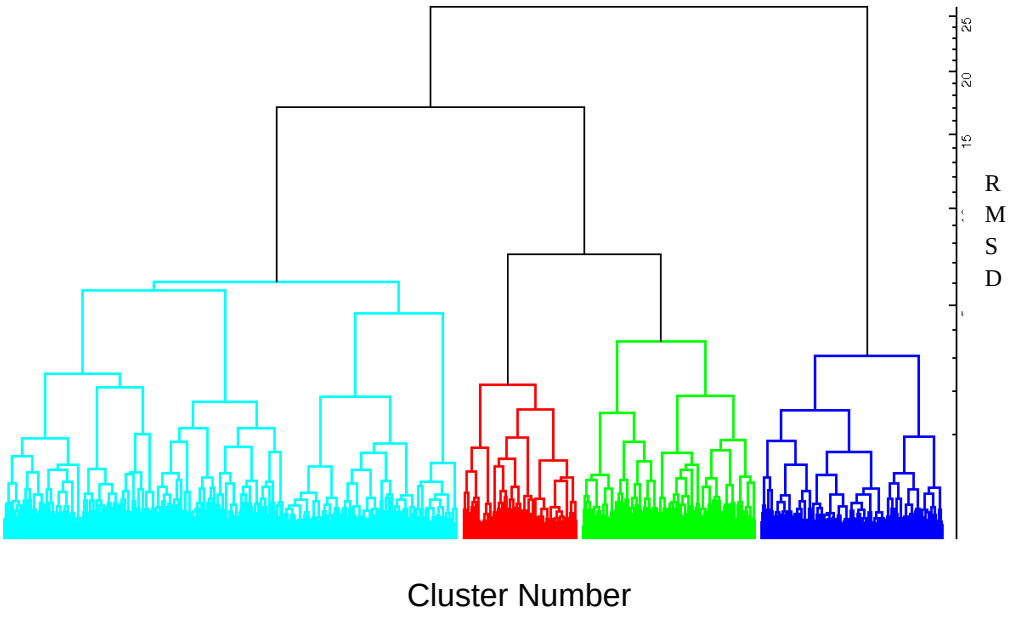


Figure 4.4 Dendrogram showing different clusters for the Aurein 2.5 in the MD trajectory classified using the Ward's clustering method [**124**]. Different number of clusters can be ascertained by considering different cutoff values of the RMSD across Y-axis.

Table 4.1: The results of cluster analysis for four major clusters identified in the MD simulation.

Sr. No.	Cluster Number	Number of Structures	% of Structures	RS Number in cluster
1	C1	985	20	3339
2	C2	935	19	4404
3	C3	2466	49	3658
4	C4	614	12	1731

To obtain a deeper understanding of the distribution of the sampled conformations of each cluster, the development of each cluster was monitored during the progress of the MD trajectory, and is depicted in Figure 4.5. Clearly, the development of the clusters C1 and C2 appeared after the middle of the MD simulation, and acquired new conformations randomly as the simulation proceeds thereafter (Figure 4.5). The most abundant cluster, C3, on other hand, started enriching with new conformations from the start of the simulation and continued to add to them in a consistent manner during the progress of the trajectory. The least occupied cluster C4, as opposed to C1 and C2, acquired all its structures before the middle of the simulation.

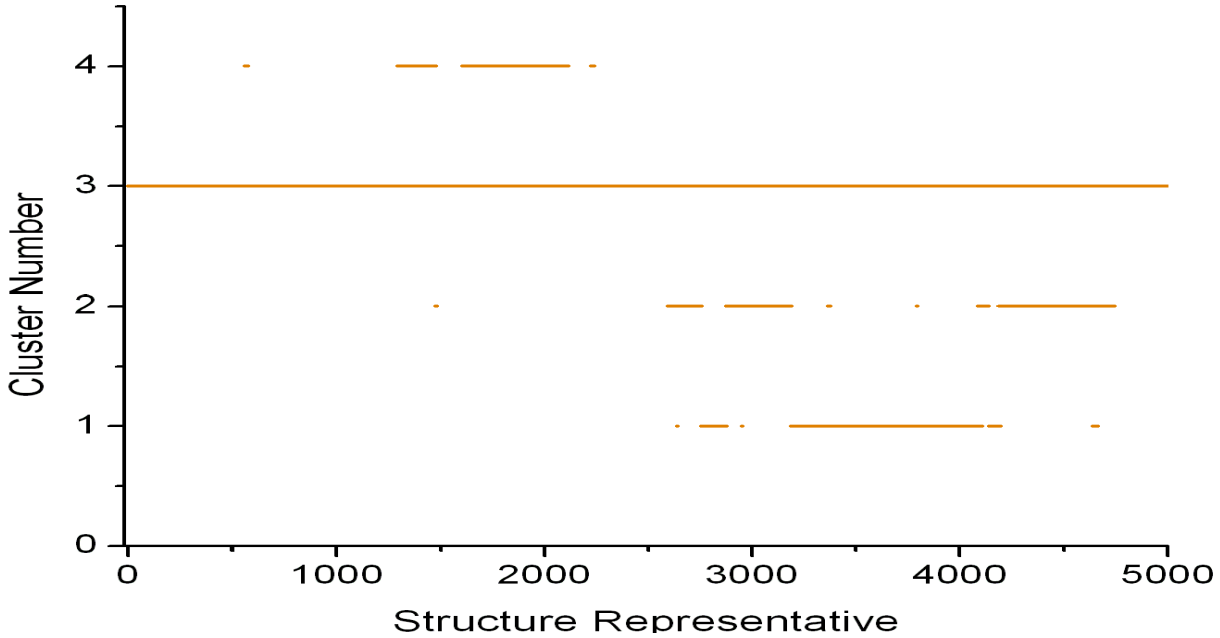
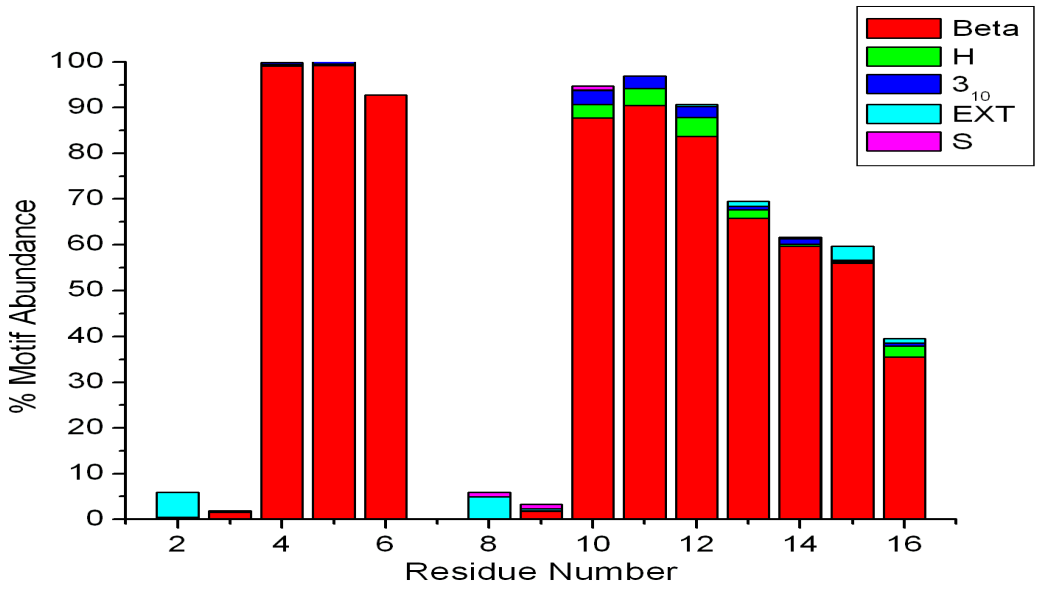


Figure 4.5 Evolution of four clusters during the progress of MD for representative structures obtained using Ward's clustering method [**124**]. The greater number of closely related structures in each cluster can be seen as solid lines.

Figure 4.6a-b and Figure 4.7a-b represent the conformational motifs of the conformations in major clusters identified using the CLASICO program [**115**] (details in Chapter 2). A closer examination of Figure 4.6a revealed that conformations in cluster C1 exhibited predominantly β-turns with a stronger propensity between the residues Phe^{4}-Ile^{6} and Val^{10}-Ser^{16}. To very small extent, α-helical character between residues Val^{10}-Ala^{13} was observed in the peptide. Residues Gly^{2}-Leu^{3} and Val^{7}-Lys^{9}, on the other hand, exhibited either extended or unspecific conformations. Although the α-helicity and 3_{10}-α-helices, in the case of cluster C2 (Figure 4.6b), was observed in lower percentages (<15%) between the residues Gly^{2}-Gly^{15}, β-turn was again the most preferred conformational motif in the conformations localized predominantly between the residues Gly^{2}-Ile^{6} and Val^{10}-Ser^{16}.

(a)



(b)

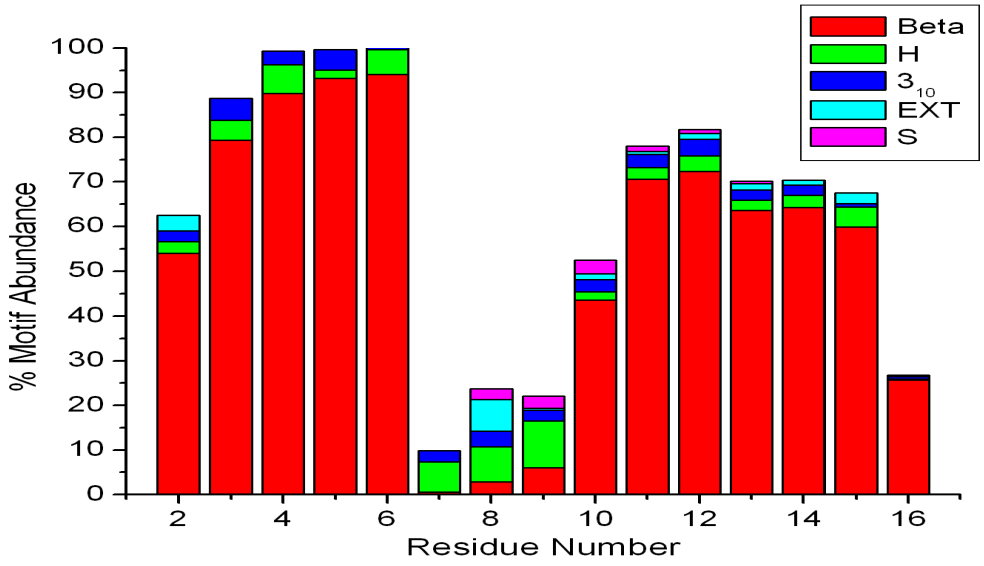
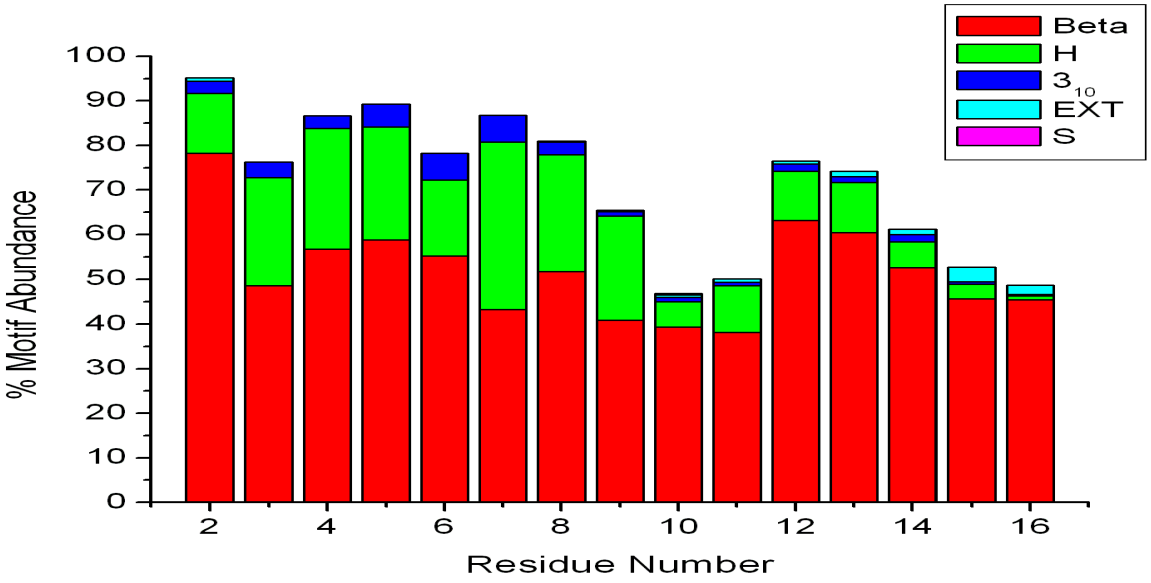


Figure 4.6 Secondary motifs attained by different residues of conformations of Aurein 2.5 in cluster C1 (a) and C2 (b). Conformational motifs are labelled: Beta (β -turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

The conformations in the highest occupied cluster, C3 adopted α-helicity flanked by residues Gly^{2}-Gly^{15} with a higher propensity between the residues Gly^{2}-Lys^{9} and with a weaker propensity between the residues Val^{10}-$Gly^{15,}$ as depicted in Figure 4.7a. The predominant conformational motif, β-turns, adopted by each residue of the peptide in varying percentages. The conformations in Cluster C4 (Figure 4.7b) also adopted β-turns in their residues with the residues Asp^{5}-Lys^{9} exhibiting α-helical character in very low percentages. The β-turns were predominantly occupied by the central residues (Asp^{5}-Val^{10}) and few C-terminal residues (Phe^{14}-Ser^{16}) with lower percentages between residues Gly^{2}-Phe^{4} and Val^{11}-Phe^{14}. Overall, these results suggest that the peptide was not stable to a particular conformation under the various solvation conditions, but was continuously interchanging between the folded and unfolded structures during the simulation process.

(a)



(b)

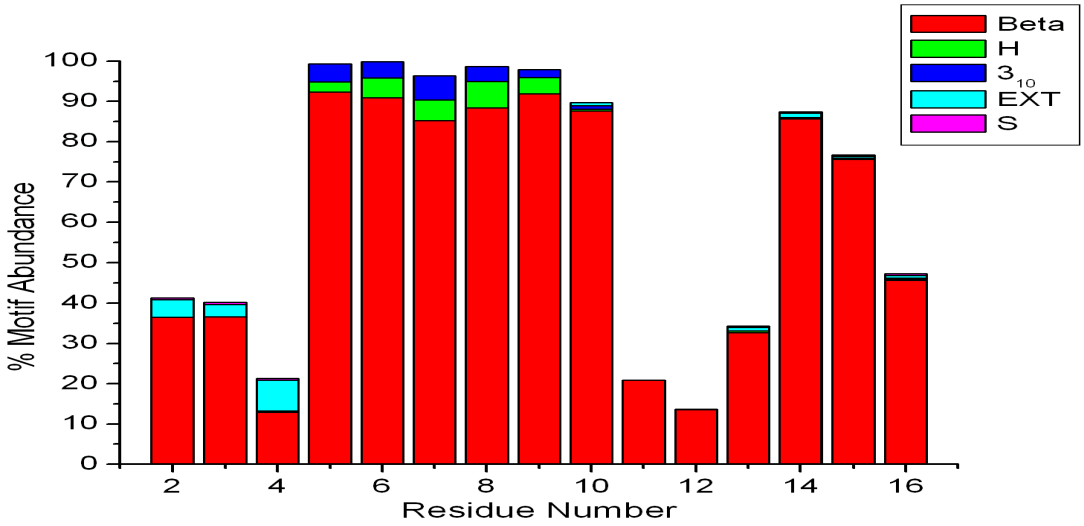
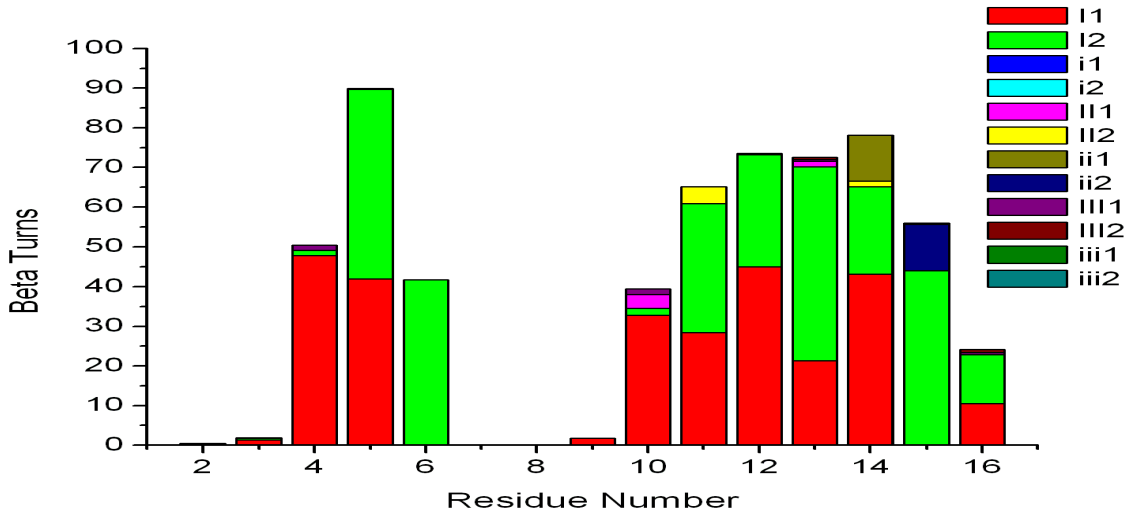


Figure 4.7 Secondary motifs attained by different residues of conformations of Aurein 2.5 in cluster C3 (a) and C4 (b). Conformational motifs are labelled: Beta (β -turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

The analysis was further extended by classifying the β-turns into different types using the CLASICO program [**115**]. The classification of β-turns was performed according to Table 2.2 as depicted in the Chapter 2.

Figure 4.8a-b and Figure 4.9a-b represent the statistics of different types of β-turns per residue for each major cluster. Generally, β-turn type I was the predominant structural motif, although in different proportions, adopted by the residues in each cluster. For instance, the residues Phe^{4}-Ile^{6} and Val^{10}-Ser^{16} in cluster C1 (Figure 4.8a) exhibited β-turn type I in higher percentages. To some extent, β-turn type II between the valine residues (Val^{10}-Val^{11}) was also observed. Residues Gly^{2}-Leu^{3} and Val^{7}-Lys^{9}, in contrast, did not exhibit any specific secondary motif. The conformations in cluster C2 adopted β-turn type I between the residues Gly^{2}-Ile^{6} and Val^{10}-Ser^{16} in higher percentages whereas between the residues Lys^{8}-Lys^{9} in lower percentages (Figure 4.8b). The β-turn type ii between the residues Val^{11}-Gly^{12} and Phe^{14}-Ser^{16} was also present in the structures of cluster C2 in lower percentages.

(a)



(b)

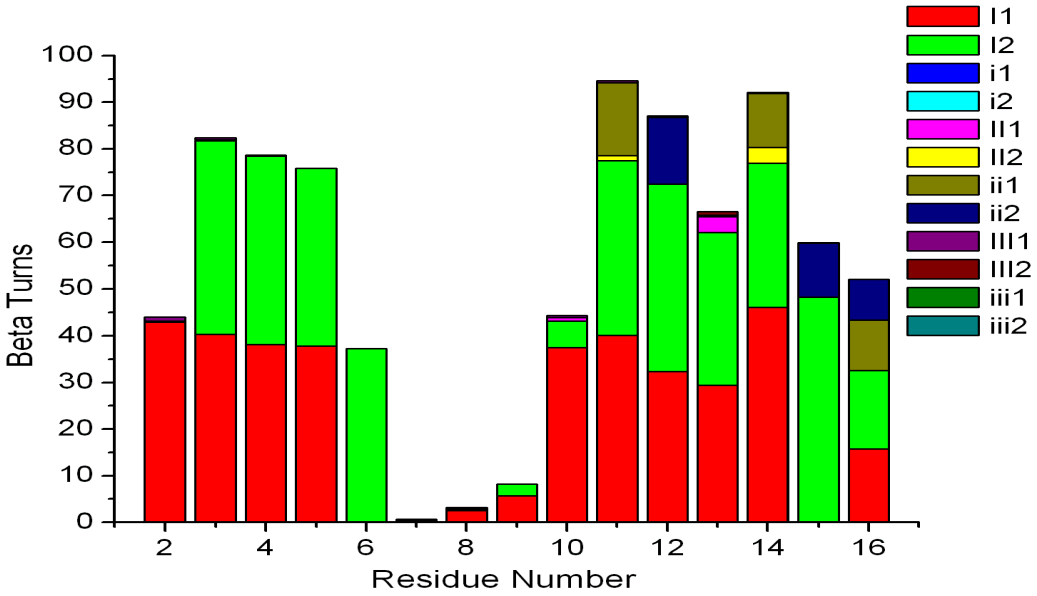
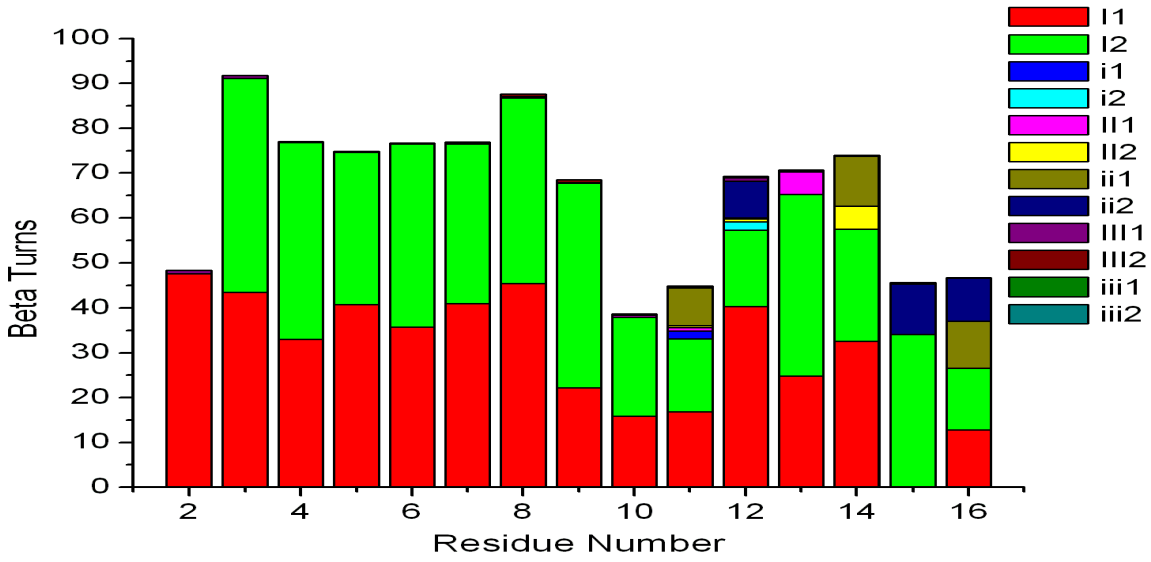


Figure 4.8 Different types of β-turns attained by the residues of peptide Aurein 2.5 in clusters (a) C1 and (b) C2, classified using the CLASICO program [**115**].

The largest cluster C3 (Figure 4.9a) exhibited β-turn type I in its amino acid residues of the peptide with a stronger propensity between the residues Gly^{2}-Lys^{9} and Gly^{12}-Gly^{15} whereas the lower propensity in the valine residues (Val^{10}-Val^{11}) and the serine residue at position 16. Residues Val^{11}-Gly^{12} and Phe^{14}-Ser^{16} exhibiting β-turn type ii (mirror image of β-turn type II) were also observed in C3 in lower percentages. Similarly, β-turn type II flanked by the residues Ala^{13}-Phe^{14} were also present in few conformations of the cluster C3. The conformations present in the least abundant cluster (C4) also exhibited β-turn type I supported by the residues Gly^{2}-Val^{11} and Ala^{13}-Leu^{16} with greater percentages between Gly^{2}-Leu^{3}, Val^{7}-Val^{10} and $Phe1^{4}$-Gly^{15} (Figure 4.9b). Residues exhibiting β-turn type II (Val^{10}-Val^{11}, Ala^{13}-Ser^{16}) and β-turn type i (Val^{11}-Gly^{12}) and β-turn type ii (Phe^{14}-Gly^{15}) were also observed in C4.

(a)



(b)

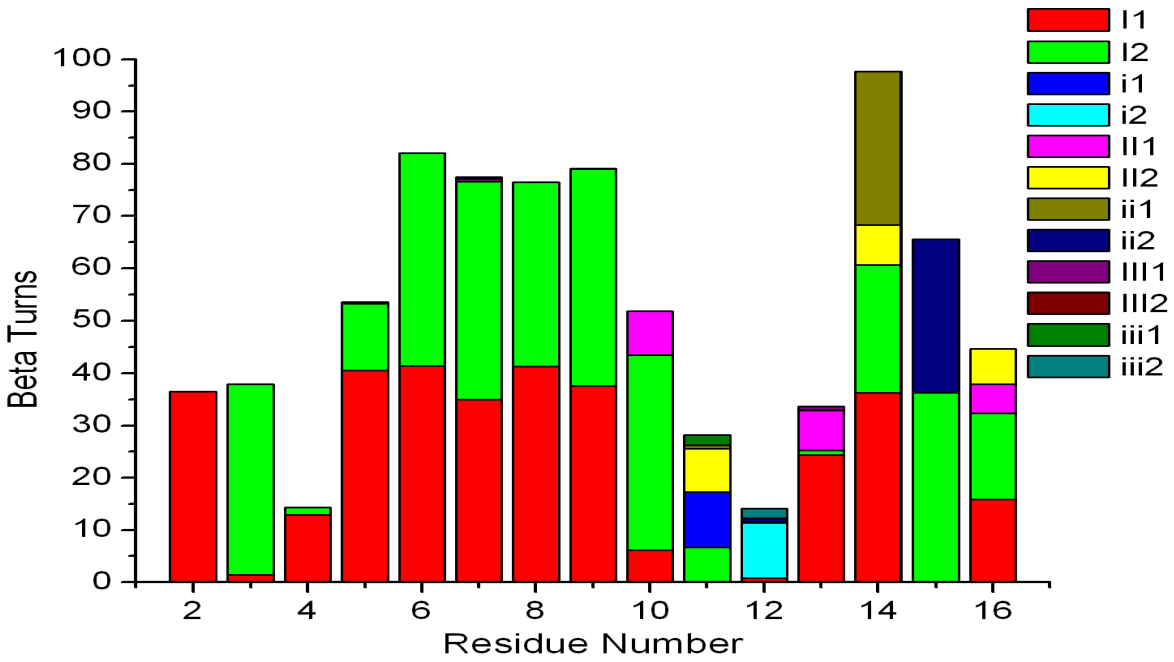
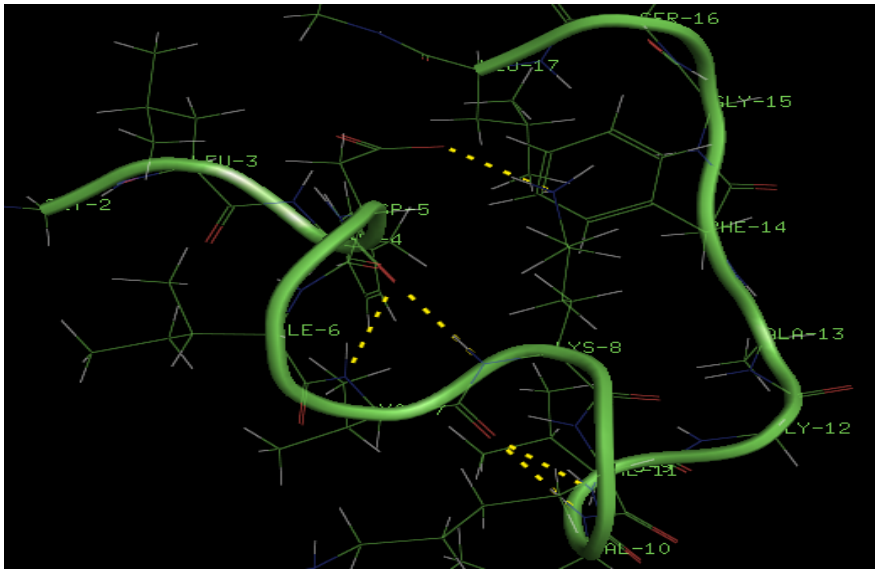


Figure 4.9 Different types of β-turns attained by the residues of peptide Aurein 2.5 in clusters (a) C3 and (b) C4, classified using the CLASICO program [**115**].

The representative structures (RS) of all clusters with their backbone atoms in ribbon form are diagrammatically depicted in Figures 4.10a-b and Figure 4.11a-b.

(a)



(b)

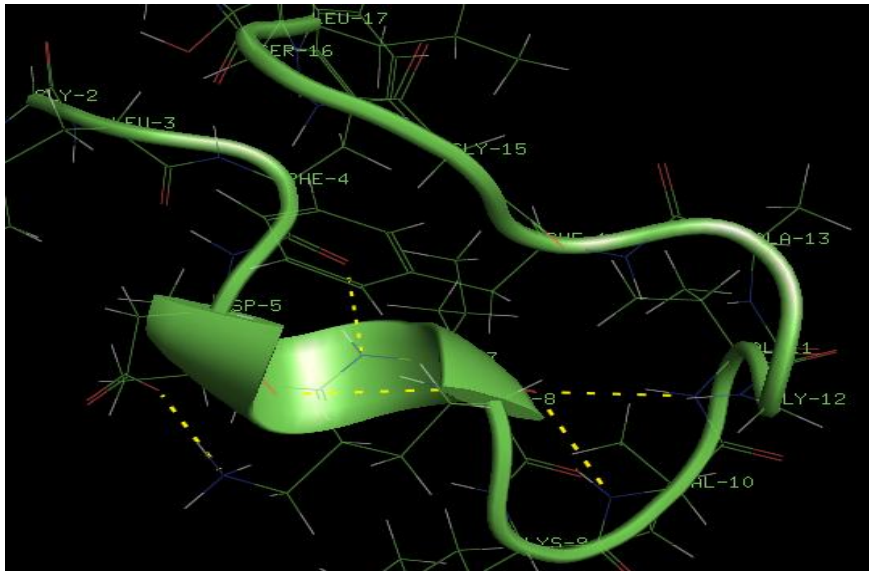
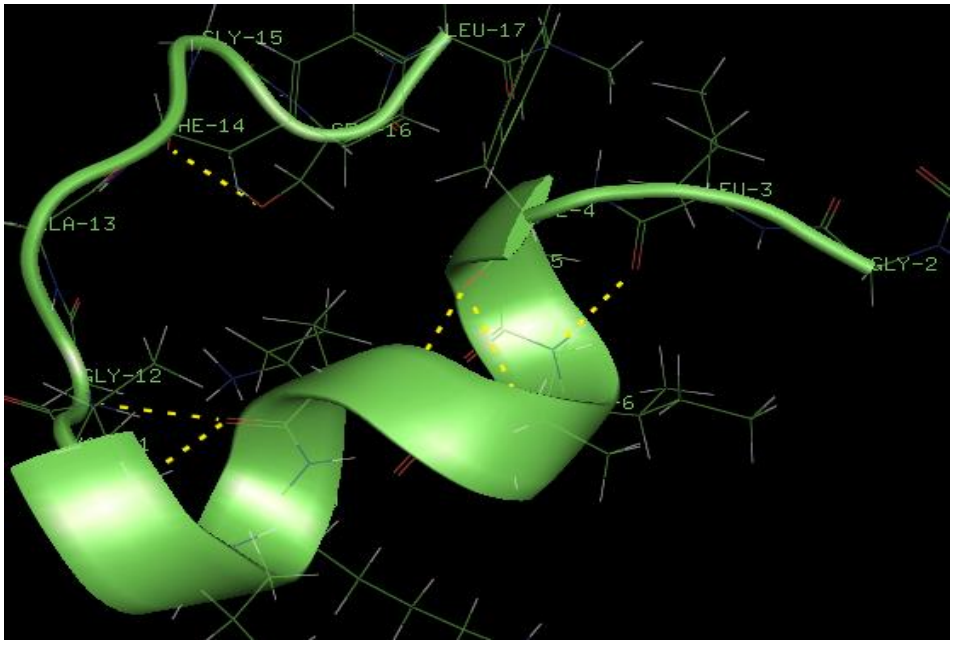


Figure 4.10 Representative structures of C1 (a) and C2 (b) showing important interactions between the residues, for Aurein 2.5. The side-chains are shown in lines format, whereas the backbone atoms are depicted in ribbon format. Hydrogen bonds are shown as yellow dotted lines.

A visual inspection of Figure 4.10a revealed the presence of an intramolecular hydrogen bond between the residues Phe^{4}-Lys^{8} accounting for a distorted α-helical region along with two β-turns between the residues Phe^{4}-Val^{7} and Val^{7}-Val^{10}. Additionally, a γ-turn flanked by the residues Val^{7}-Lys^{9} and side-chain interactions between the residues Asp^{5}-Lys^{8} were also present in the RS. In the case of cluster, C2 (Figure 4.10b), a distorted α-helical region between residues Asp^{5}-Val^{11}supported by three β-turns between residues Phe^{4}-Val^{7}, Asp^{5}-Lys^{8} and Lys^{8}-Val^{11} and a γ-turn between residues Lys^{8}-Val^{10}, were characterized on the basis of hydrogen bonding. Moreover, a side-chain interaction between residues Ile^{6}-Lys^{8} supported by a single hydrogen bond was also observed.

The RS of the largest cluster, C3, depicted in Figure 4.11a, exhibited a well-defined α-helical between the residues Phe^{4}-Val^{11}, on the basis of hydrogen bond interactions between its amino acid residues. Although, a single hydrogen bond interaction between the backbone oxygen atom of Phe^{14} and the side-chain oxygen atom of Ser^{16} was also observed in the RS, the peptide did not exhibit any characteristic secondary feature in the few N-terminal residues (Gly^{2}-Phe^{4}) and C-terminal residues Val^{11}-Leu^{17}. In case of the cluster C4 (Figure 4.11b), the RS adopted two β-turns between the residues Phe^{4}-Val^{7} and Asp^{5}-Lys^{8} along with distorted α-helical region supported by the hydrogen bonding between the residues Phe^{4}-Lys^{8} and Lys^{8}-Gly^{12}. A side-chain interaction between the oxygen atom of Asp^{5} and the aminic hydrogen atom of Lys^{8} was also observed in the RS.

(a)



(b)

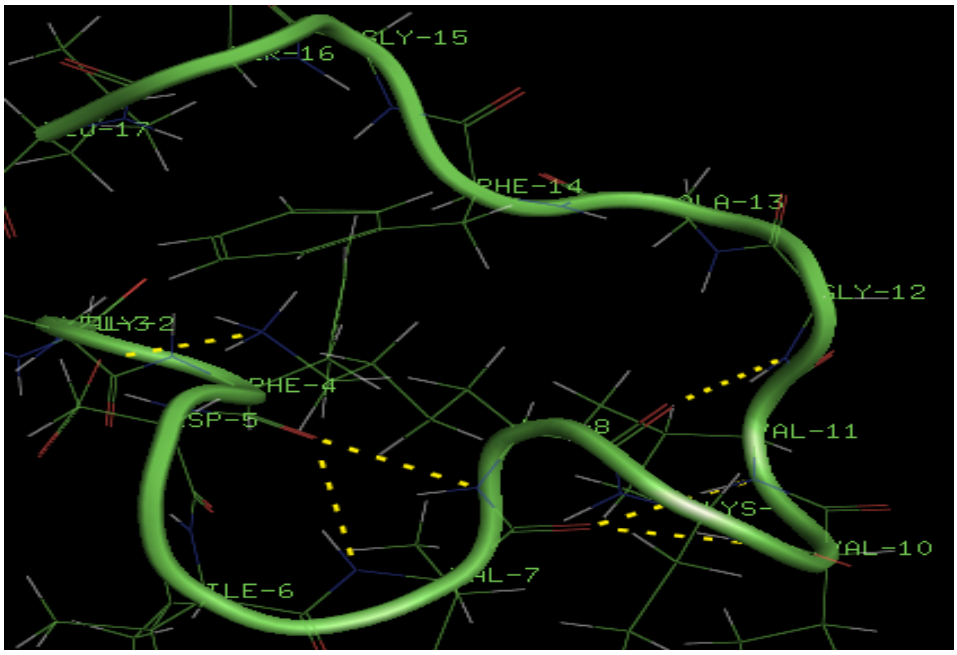


Figure 4.11 Representative structures of C3 (a) and C4 (b) showing important interactions between the residues, for Aurein 2.5. The side-chains are shown in lines format, whereas the backbone atoms are depicted in ribbon format. Hydrogen bonds are shown as yellow dotted lines.

The observed α-helical propensities in the largest cluster (C3) were further rationalized on the basis of hydrogen bond analysis. A description of the geometrical criterion used in hydrogen bond analysis has already been provided in Chapter 3. The systematic progress of the hydrogen bonds showing their formation/destruction during the progress of MD trajectory in the cluster C3 is pictorially depicted in Figure 4.12.

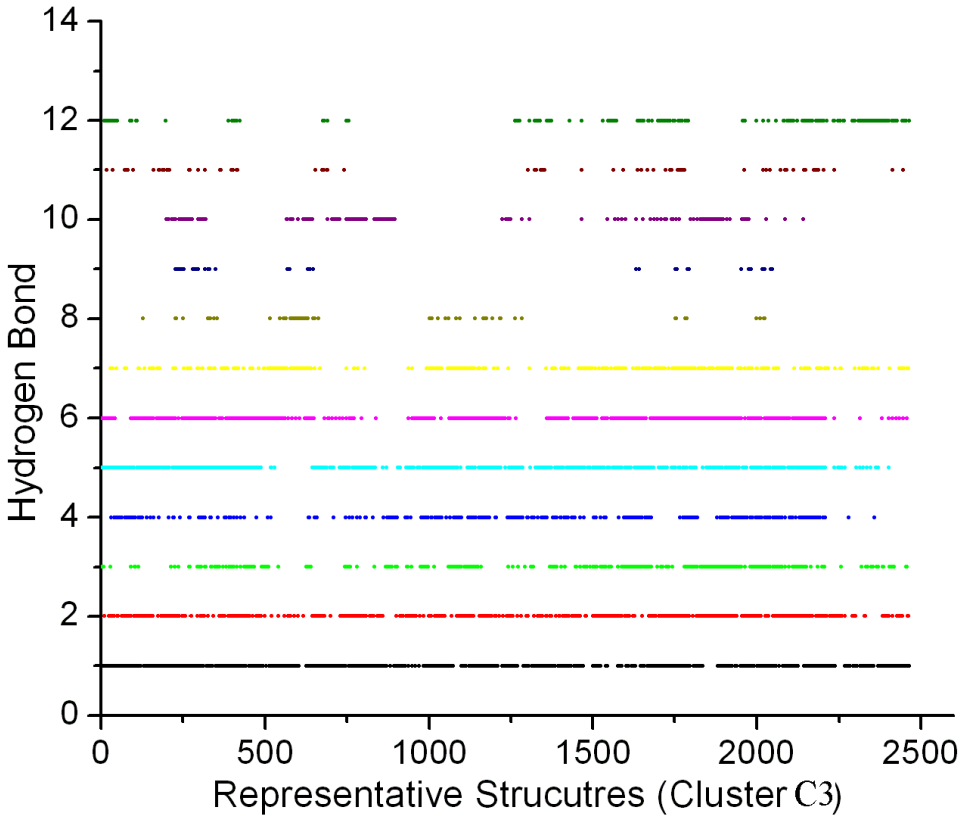


Figure 4.12 Progress of intramolecular hydrogen bonds observed between amino acid residues of the sampled conformations in the most abundant cluster (C3).

The results of the hydrogen bond analysis including the interacting residues and the percentage of hydrogen bonding are shown in Table 4.2. The first specified hydrogen bond distance depicted in Figure 4.12, corresponds to the first percentage value in Table 4.2. It is clear from Table 4.2, that the α-helical region was supported by almost all the residues (Gly^{2}-Leu^{17}) of the peptide although in different percentages. For instance, the hydrogen bonds (entries 1-3, Figure 4.12) responsible for the α-helical region flanked by the residues Gly^{2}-Lys^{8} were adopted from the start of the trajectory, and were randomly acquired by the conformations until the end of the simulation. The hydrogen bond accounting for the α-helical region between the residues Asp^{5}-Lys^{9}, however, was not observed in the last segment of the trajectory, as depicted in Figure 4.12 (entry 4). The α-helical character flanked by the residues Ile^{6}-Val^{11} was considerably increased, and was adopted by majority of the sampled conformations during the progress of the MD trajectory (entries 5-6, Figure 4.12). The percentage of hydrogen bonds accounting for the α-helicity between the residues Lys^{8}-Gly^{12} (entry 7, Figure 4.12) decreased due to its random adoption by the conformations. The progress of hydrogen bonds responsible for the α-helical region between the residues Lys^{9}-Leu^{17} (entries 8-12, Figure 4.12) was not uniform, and was regularly interrupted during the progress of the simulation resulting in lower percentages of hydrogen bonds (Table 4.2).

Table 4.2: The α-helical features observed due to backbone-backbone hydrogen bond interactions and their percentages in the most abundant cluster (C3) of MD for Aurein 2.5.

No	Donor-acceptor	Cluster 3 (C3)
1	$(Gly^{2})O\ldots N(Ile^{6})$	41.7 %
2	$(Leu^{3})O\ldots N(Val^{7})$	27.5 %
3	$(Phe^{4})O\ldots N(Lys^{8})$	19.1 %
4	$(Asp^{5})O\ldots N(Lys^{9})$	14.3 %
5	$(Ile^{6})O\ldots N(Val^{10})$	33.4 %
6	$(Val^{7})O\ldots N(Val^{11})$	36.2 %
7	$(Lys^{8})O\ldots N(Gly^{12})$	17.0 %
8	$(Lys^{9})O\ldots N(Ala^{13})$	3.0 %
9	$(Val^{10})O\ldots N(Phe^{14})$	3.1 %
10	$(Val^{11})O\ldots N(Gly^{15})$	6.5 %
11	$(Gly^{12})O\ldots N(Ser^{16})$	3.2 %
12	$(Ala^{13})O\ldots N(Leu^{17})$	6.0 %

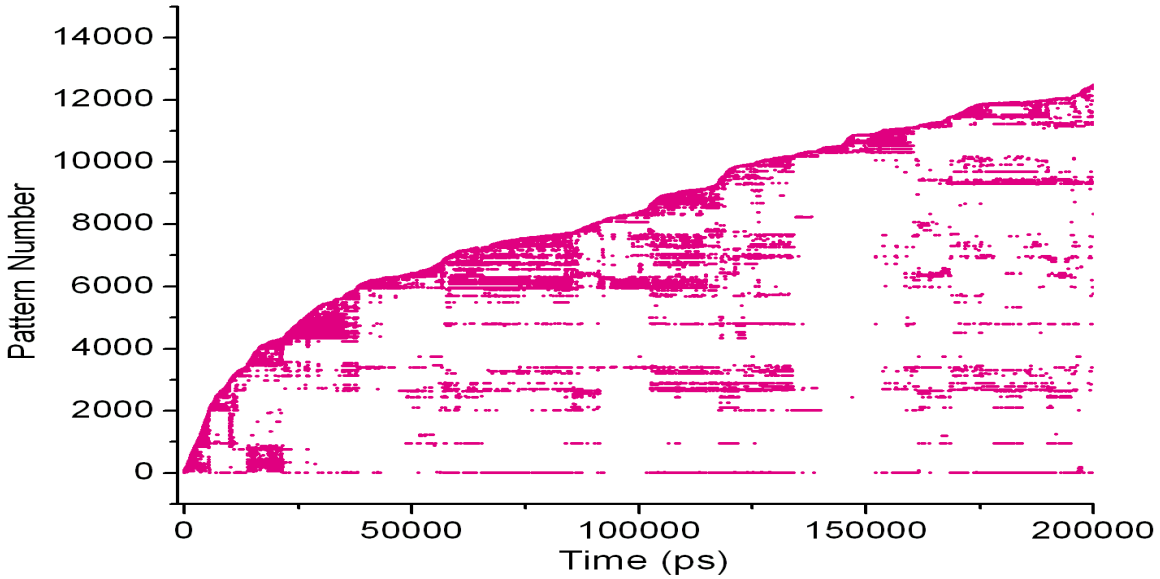
4.3 Molecular dynamics study of Aurein 2.5 in explicit solvents

In order to get a deeper understanding of the conformational preferences of Aurein 2.5 in different solvents, and to compare the results with those obtained using the implicit solvent model, we extended our MD study in the polar solvents *viz* water and methanol using their explicit solvent models in the AMBER [**94**]. The computational procedure used for the MD simulations was the same as employed for the Aurein 2.4, described in the methods section of Chapter 3. For convenience, the MD simulations performed for

water and methanol solvents are denoted as MD^{wat} and MD^{meth} respectively. The sampling efficiency of both MD simulations was first assessed by monitoring the evolution of new patterns as a function of simulation time using the CLASICO program [**115**], and is diagrammatically represented in Figure 4.13a-b. Of the 200,000 sampled conformations, a total of 121212 and 142312 patterns were identified for MD^{wat} and MD^{meth}, respectively. A closer inspection of Figure 4.13a revealed that the patterns were sampled very rapidly after the start of the simulation until 9 ns probably due to faster changes in the geometry of the peptide, and started to decrease thereafter as the system folds up due to the development of intramolecular hydrogen bonding between the amino acid residues of the peptide. A couple of random plateau formation particularly after 39 ns, 85 ns and 178 ns further indicated the entrapment of the conformations to the high energy states during the sampling process.

The conformations in the case of methanol solvent (MD^{meth}) adopted a greater number of patterns, and started acquiring them in a slow but steady pace from the beginning of the trajectory and proceeded in a similar manner throughout the progress of the simulation (Figure 4.13b). The absence of a plateau formation in the trajectory MD^{meth} suggests a less restrictive nature of the simulation in finding new patterns, and accounts for their greater number observed in the methanol solvent. In contrast to the water solvent, where the system was approaching a stable geometry after 200 ns, the peptide in methanol solvent was comparatively unstable and kept adding new conformations during the progress of the MD simulation.

(a)



(b)

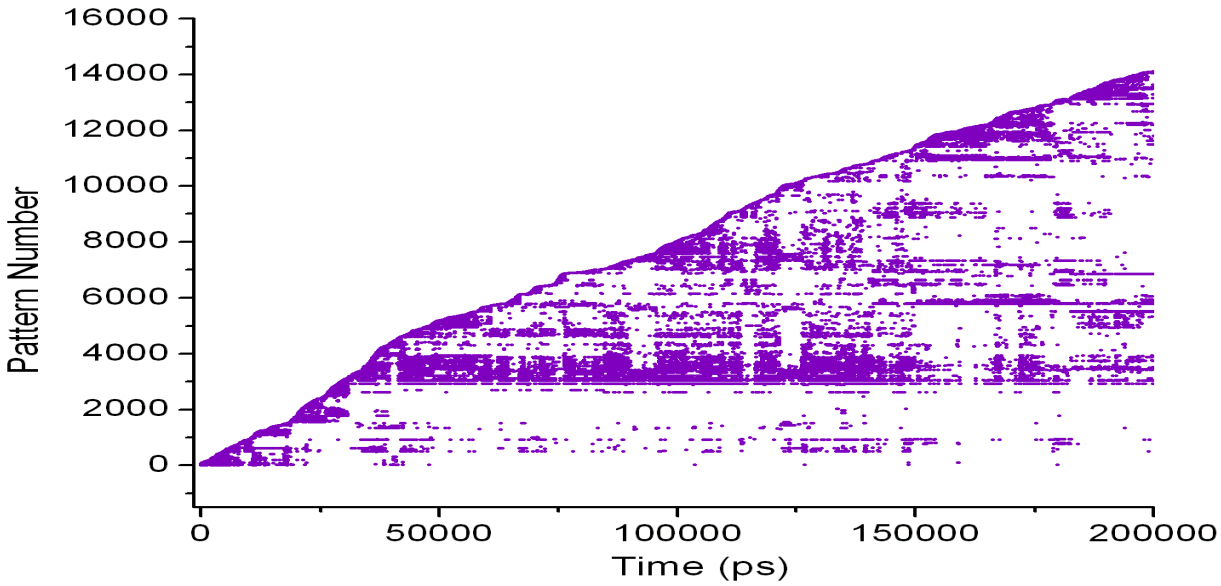
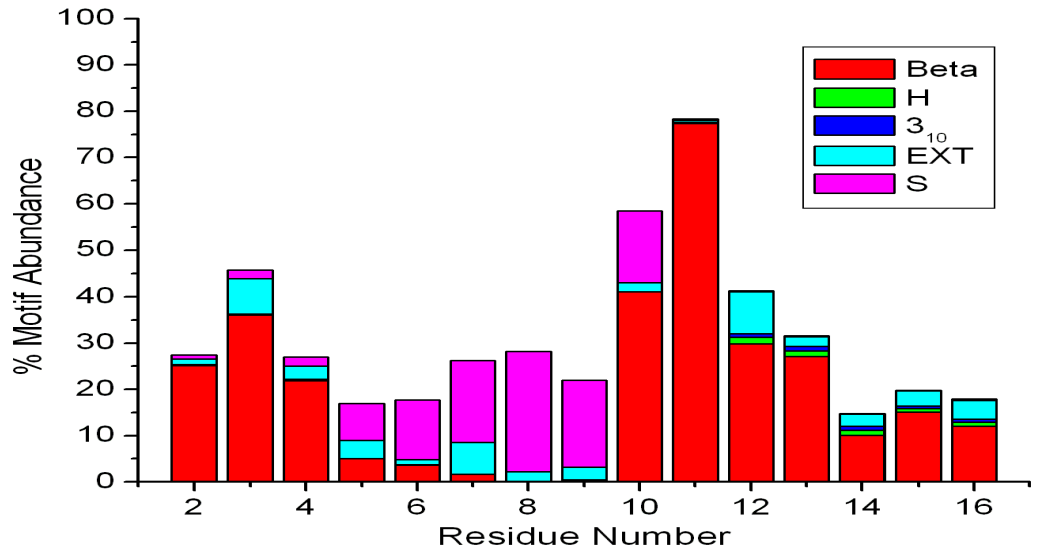


Figure 4.13 Evolution of new patterns during the MD simulation performed in water (a) and methanol (b) for peptide Aurein 2.5, obtained using the CLASICO program [**115**].

Different secondary features of the conformations sampled in the trajectories MD^{wat} and MD^{meth} were further identified using the CLASICO program [**115**], and are pictorially depicted in Figure 4.14a and Figures 4.14b respectively. A closer inspection of both figures revealed that the peptide adopted β-turns and β-strands as its preferred secondary motifs in both solvents. Specifically in water (Figures 4.14a), the peptide exhibited β-turns between residues Gly^2-Val^7 and Val^{10}-Ser^{16} with a higher propensity between the residues Val^{10}-Val^{11}. The residues Asp^5-Val^{10}, on other hand, adopted preferably β-strands with lower percentages. Nonetheless, the majority of the residues (>50%) of the peptide exhibited not any specific secondary feature in water. The extent of the secondary features of the peptide was comparatively higher in methanol solvent preferably in the N-terminal residues, as depicted in Figures 4.14b. More than 60% of the conformations adopted β-turns in their N-terminal residues (Gly^2-Ile^6), whereas C-terminal residues (Lys^9-Ser^{16}) exhibited β-turns in lower percentages. To some extent, β-strands flanked by the residues Ile^6-ser^{16} were also adopted by the conformations in the methanol solvent. Interestingly, no trace of α-helicity was observed in any residue of the peptide in the both solvents.

(a)



(b)

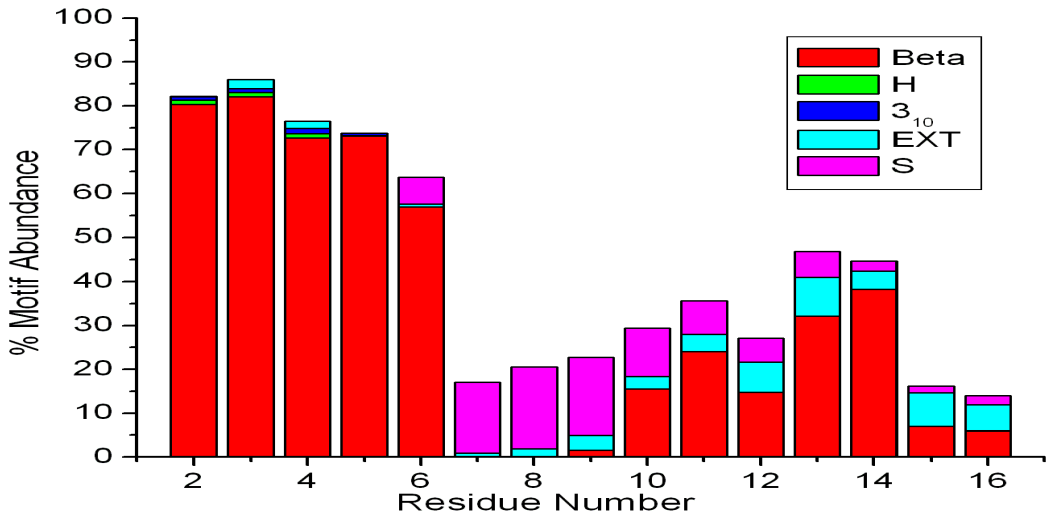
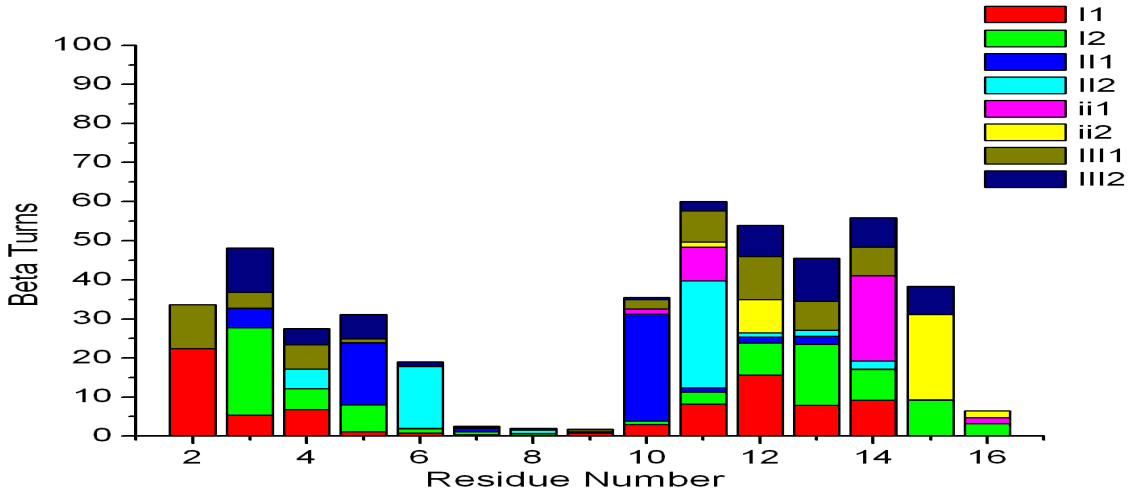


Figure 4.14 Different secondary motifs attained by different residues of conformations in trajectory (a) MD^{wat} and (b) MD^{meth}. Conformational motifs are labelled: Beta (β-turn), H (α-helical), 3_{10} (3_{10}-α-helix), EXT (extended) and S (β-strand).

Finally, the obtained β-turns were further classified into their turn-types using the CLASICO program [**115**] described in the methods section in Chapter 2. Histograms showing different types of β-turns adopted by each residue of the peptide along with their percentages are shown in Figure 4.15a-b. As can be seen in Figure 4.15a that only 60% of the sampled conformations exhibited different β-turns in water whereas the remaining did not shown any specific secondary feature. Specifically, β-turn type I was localized in the N-terminal (Gly^2-Asp^5) and C-terminal residues (Val^{10}-Gly^{15}) with a greater propensity between the residues Gly^2-Leu^3 and Gly^{12}-Phe^{14}. The central residues, on other hand, preferred to stay in an extended or unspecific conformation. To some extent the β-turn type III between the residues Gly^2-Asp^5 and Val^{10}-Gly^{15} were also observed in few conformations. Similarly, β-turn type II between residues Leu^3-Ile^6 and Val^{10}-Phe^{14} with stronger propensity between residues Val^{10}-Val^{11}, and its mirror image (β-turn type ii) between residues Val^{11}-Gly^{12} and Phe^{14}-Gly^{15} were also adopted by few conformations in the water solvent.

(a)



(b)

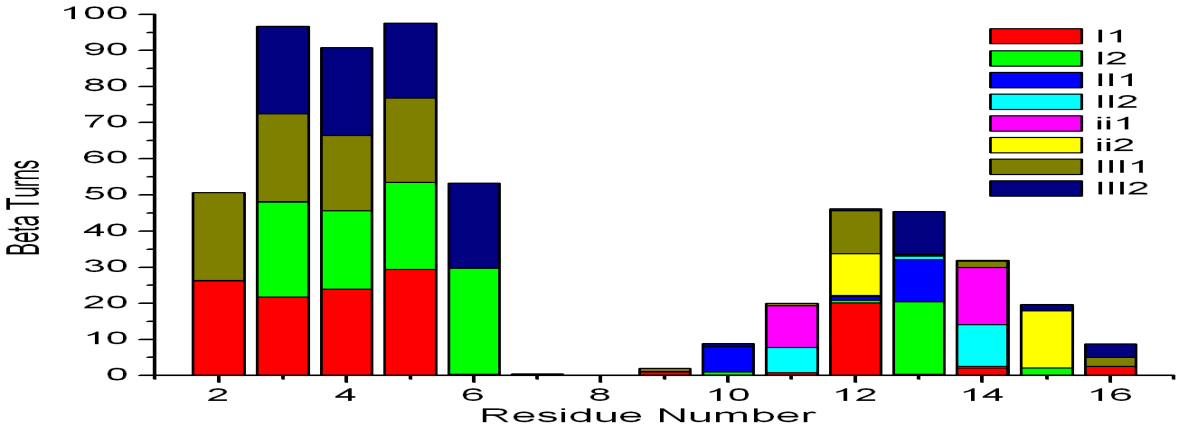


Figure 4.15 Different types of β-turns attained by the residues of peptide Aurein 2.5 in trajectories (a) MD^{wat} and (b) MD^{meth}, classified using the CLASICO program [**115**].

In the case of methanol (Figure 4.15b), more than 50% of conformations exhibited different types of β-turns in their N-terminal residues (Gly^2-Ile^6). The extent of β-turns was comparatively lower in the C-terminal residues with the residues Val^7-Lys^9 exhibiting unspecific secondary feature. Specifically, β-turn type I was adopted by the residues Gly^2-Ile^6 in higher percentages, whereas the residues Gly^{12}-Ala^{13} exhibited

similar type of β-turn in lower percentages. Similarly, β-turn type III was occupied by the residues Gly^{2}-Ile^{6} in higher percentages, whereas in lower percentages between the residues Gly^{12}-Ala^{13}. The conformations exhibiting β-turn type II between the residues Val^{10}-Val^{11} and Ala^{13}-Phe^{14}, β-turn ii between the residues Val^{11}-Gly^{12} and Phe^{14}-Gly^{15} and β-turn type III between the residues Gly^{12}-Ala^{13} were also observed in the methanol solvent. The observed higher percentages of secondary motifs in methanol suggested the greater stability of peptide in this solvent.

4.4 Conclusions

The MD simulations were performed on Aurein 2.5 to explore its conformational profile both under implicit (GB-OBC model) and explicit solvent conditions. The results obtained from implicit solvent MD simulation indicated that peptide attained α-helical conformation in all residues. However, the conformation of the peptide was not stable, and was in a rapid equilibrium with the un-folded conformations throughout the progress of MD simulation. The presence of the explicit polar solvent (water and methanol) molecules destabilized the α-helicity, and forced the peptide to stay either in the form of β-turns or β-strands. The comparison of secondary structure analysis for Aurein 2.4 and Aurein 2.5 revealed that the distribution of the α-helical content was different in both peptides. Unlike Aurein 2.4, where α-helical region was predominant in the central region extending towards C-terminus, the α-helicity in Aurein 2.5 was predominant in the N-terminal and central residues. Moreover, the extent of α-helicity in Aurein 2.3 was comparatively more than the other two peptides (Aurein 2.3).

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Aurein peptides, obtained primarily from frog skin **[63]**, have shown great potential against several microbial and cancer infections [**64**], and have promoted their studies worldwide for the development of new drugs and peptidomimetics. Although, some most active members (Aurein 1.2, Aurein 2.2 etc.) of this family have been extensively studied, some of their structural analogues have evaded the attention of researchers. The present work was aimed at getting a deeper understanding of the structure-activity relationships for three Aurein peptides: Aurein 2.3, and its two structural analogues Aurein 2.4 and Aurein 2.5. For this purpose the conformational profiles of these peptides were explored using longer molecular dynamics (MD) method. First, the performance of the MD protocols was assessed by performing simulations of Aurein 2.3 for which experimental structure is known [**65**]. Analysis of the results obtained for Aurein 2.3 revealed its tendency to exhibit a α-helix between residues 2-14, in accordance with the reported literature [**65**]. Similar MD protocol was further employed to explore the conformational profiles of Aurein 2.4 and Aurein 2.5. The results obtained reveal the propensity of both peptides to adopt helical conformations in equilibrium with their corresponding unfolded structures (β–turns). The comparison of MD results for Aurein 2.4 and Aurein 2.5 further revealed that the α-helical region in case of former was localized predominantly in the central residues extending towards its N-terminal residues, whereas it was flanked by N-terminal and central residues in case of later. Moreover, the α-helical content observed in case of both peptides (Aurein 2.4 and Aurein 2.5) was comparatively lower than those present for Aurein 2.3, and could be

related to their biological propensities. Additionally, the presence of explicit polar solvent molecules destabilized the α-helicity in both peptides (Aurein 2.4 and Aurein 2.5), and promoted their unfolded conformations (β–turns/strands and extended etc.).

Recommendations

The current simulation studies performed on the selected Aurein peptides (Aurein 2.3, Aurein 2.4 and Aurein 2.5) indicated that the AMBER ff99SB force field in combination with the Onufriev, Bashford and Case (OBC) implementation of the GB method is capable of reproducing the conformational profiles obtained with the explicit solvent simulations, and can be employed for future studies on the medium sized peptides. In order to get an efficient sampling for the peptides of this size, it is required to run MD trajectories at least 100 ns long when an implicit solvent model is used and in the case of explicit solvent greater than 200 ns. The application of replica exchange molecular dynamics method could also reduce the simulation time. In particular, the structures of peptides obtained from these simulations can be used to design agonists and antagonists of these peptides that can then be applied as potential lead compounds in the field of computer-aided drug design and peptidomimetics.

REFERENCES

1. Pensak, D. A., **1989**. Molecular modeling: scientific and technological boundaries, Pure & Applied Chemistry, (61): 601-603.

2. Scheraga, H. A., Khalili, M., and Liwo, A., **2007**. Protein-folding dynamics: Overview of molecular simulation techniques, Annual Review of Physical Chemistry, (58): 57-83.

3. Pineiro, A., Villa, A., Vagt, T, Koksch, B., and Mark, A.E., **2005**. A molecular dynamics study of the formation of, stability, and oligomerization state of two designed coiled coils: possibilities and limitations, Biophysical Journal, (89): 3701-3713.

4. Gnanakaran, S., Nymeyer, H., Portman, J., Sanbonmatsu, K.Y., and García, A.E., **2003**. Peptide folding simulations, Current Opinion in Structural Biology, (13): 168-174.

5. Daura, X., **2006**. Molecular dynamics simulation of peptide folding, Theoretical Chemistry Accounts, (116): 297-306.

6. Cecchini, M., Rao, F., Seeber, M., and Caflisch, A., **2004**. Replica-exchange molecular dynamics of amyloid peptide aggregation, The Journal of chemical physics, (21): 10748-10756.

7. Thomasson, W.A., **2008**. Unravelling the mystery of protein folding. http:/www.faseb.org/opa. Breakthroughs in science. Accessed on 25 May 2013.

8. Urbanc, B., Cruz, L., Ding, F., Sammond, D., Khare, S., Buldyrev, S.V., and Dokholyan, N.V., **2004**. Molecular dynamics simulation of amyloid β dimmer formation, Biophysical Journal, (87): 2310-2321.

9. Janson, J., Laedtke, T., Parisi, J.E., O'Brien P., Peterson, R.C., and Butler, P.C., **2004**. Increased risk of type 2 Diabetes in Alzheimer Disease, Diabetes, (53): 474-481.

10. Levinthal, C., **1968**. Are there pathways for protein folding? Journal of Chemical Physics, (65): 44-45.

11. Taniuchi, H. and Anfinsen, C.B., **1969**. An experimental approach to the study of the folding of staphylococcal nuclease, Journal of Biological Chemistry, (244): 3864-3875.

12. Anfinsen, C.B., **1973**. Principles that govern the folding of protein chains, Science, (18): 223-230.

13. Ptitsyn, O. B., **1991**. How does protein synthesis give rise to the 3D-structure? FEBS Letters, (285): 176-181.

14. Levinthal, C., **1968**. Are there pathways for protein folding? Journal of Chemical Physics, (65): 44-45.

15. Karplus, M. and Weaver, D. L., **1976**. Protein folding dynamics. Nature, (260): 404-406.

16. Karplus, M. and Weaver, D. L., **1994**. Protein folding dynamics: the diffusion collision model and experimental data. Protein Science, (3): 650-668.

17. Cuervo, A. M. and Dice, J. F., 2000. Age-related decline in chaperone mediated autophagy. Journal of Biological Chemistry, (275): 31505-31513.

18. Kurapati, R., Passananti, H. B.,Rose, M. R., and Tower, J.,**2000**. Increased hsp22 RNA levels in Drosophila lines genetically selected for increased longevity. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, (55): 552-559

19. Jackson, S.E., and Fersht, A.R., **1991**. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, Biochemistry, (30): 10428-10435.

20. Otzen, D.E., Itzhaki, L. S., Masry, N.F., Jackson, S.E. and Fersht, A.R., **1994**. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding, Proceedings of the National Academy of Sciences USA, (91): 10422-10425.

21. Fersht, A.R., **1997**. Nucleation mechanisms in protein folding, Current Opinion in Structural Biology, (7): 3-9.

22. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G.,**1995**. Funnels, pathways and the energy landscape of protein folding: A synthesis. Proteins Structure Function and Genetics, (21) 1619-1620.

23. Wolynes, P. G., Onuchic, J. N., and Thirumalai, D., **1995**. Navigating the folding routes, Science, (267): 1619-1620.

24. Radford, S.E., **2000**. Protein folding: progress made and promises ahead, Trends in Biochemical Sciences, (25): 611-618.

25. Schultz, C.P., **2000**. Illuminating folding intermediates, *Nature* Structural & Molecular Biology, (7): 7-10.

26. Onuchic, J.N., and Wolynes, P.G., **2004**. Theory of protein folding, Current Opinion in Structural Biology, (14): 70-75.

27. Drenth, J., **1994**. Principles of Protein X-ray Crystallography, New York: Springer.

28. Miller, R.T., Jones, D.T., and Thronton, J.M., **1996**. Protein fold recognition by sequence threading: Tools and assessment techniques, FASEB Journal, (10): 171-177.

29. Lathrop, R.H., and Smith, T.F., **1996**. Global optimum protein threading with gapped alignment and empirical pair score functions, Journal of Molecular Biology, (255): 641-666.

30. Liwo, A., Czaplewski, C., Oldziej, S., and Scheraga, H.A., **2008**. Computational techniques for efficient conformational sampling of proteins, Current Opinion in Structural Biology, (18): 134-139.

31. Corcho, F.J., Filizola, M., and Perez, J.J., **2000**. Evaluation of the iterative simulated annealing technique in conformational search of peptides, Chemical Physics Letters, (319): 65-70.

32. Arcangeli C., Bizzarri A. R., and Cannistraro S, **2001**. Molecular dynamics simulation and essential dynamics study of mutated plastocyanin: structural, dynamical and functional effects of a disulfide bridge insertion at the protein surface. Biophysical Chemistry, (92):183–199.

33. Legge, F. S., Budi A., Treutlein, H., and Yarovsky, I., **2006**. Protein flexibility: multiple molecular dynamics simulations of insulin chain B. Biophysical Chemistry, (119): 146–157.

34. Petraccone, L., Garbett, N. C., Chaires, J. B., and Trent, J. O., **2010**. An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. Biopolymers, (93): 533–548.

35. Stavrakoudis, A., Tsoulos, I. G., Shenkarev, Z. O., and Ovchinnikova, T. V., **2009**. Molecular dynamics simulation of antimicrobial peptide arenicin-2: *β*-hairpin stabilization by noncovalent interactions. Peptide Science, (92): 143–155.

36. Wang, T.,and Wade, R.C., **2003**. Implicit solvent models for flexible protein-protein docking by molecular dynamics simulation, Proteins Structure Function and Bioinformatics, (50):158–169.

37. Xia, B.,Tsui, B.,Case, D.,Dyson, H., and Wright, P., **2002**. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized born model, and with explicit water, Journal of Bimolecular NMR, (22): 317–331.

38. Dominy, B., and Brooks III, C.L.,**1999**. Development of a Generalized Born model parameterization for proteins and nucleic acids, Journal of Physical Chemistry, (103): 3765–3773.

39. Calimet, N., Schaefer, M., and Simonson, T.,**2001**. Protein molecular dynamics with the generalized Born/ACE solvent model, Journal of Molecular Graphics and Modeling, (19):136–145.

40. Feig, M., and Tanizaki, S. **2006**. Development of a heterogeneous dielectric generalized Born model for the implicit modeling of membrane environments, in: K. Naidoo (Ed.),

Modeling Molecular Structure and Reactivity in Biological Systems, Royal Society of Chemistry, Cambridge, UK, pp. 141–150.

41. Lin, J.-H., Baker, N.A., McCammon, J.A.,**2002**. Bridging implicit and explicit solvent approaches for membrane electrostatics, Biophysical Journal, (83): 1374–1379.

42. Roux, B., **1997**. Influence of the membrane potential on the free energy of an intrinsic protein, Biophysical Journal, (73): 2980–2989.

43. Hirota, N., Mizuno, K.,and Goto, Y.,**1998**. Group additive contributions to the alcoholinduced alpha-helix formation of melittin: implication for the mechanism of the alcohol effects on proteins, Journal of Molecular Biology, (275): 365–378.

44. Hong, D.,Hoshino, M., Kuboi, R.,and Goto, Y., **1999**. Clustering of fluorine-substituted alcohols as a factor responsible for their marked effects on proteins and peptides, Journal of American Chemical Society, (121): 8427–8433.

45. Feig, M., Im, W.,and Brooks 3rd, C.L.,**2004**. Implicit solvation based on generalized Born theory in different dielectric environments, Journal of Chemical Physics, (120): 903–911.

46. Sigalov, G., Scheffel, P., Onufriev, A., **2005**. Incorporating variable dielectric environments into the generalized born model, Journal of Chemical Physics, (122): 094511–094515.

47. Haney, E.F., Hunter, H.N., Matsuzaki, K., and Vogel, H.J., **2009**. Solution NMR studies of amphibian antimicrobial peptides: linking structure to function, Biochimica et Biophysica Acta, (1788): 1639–1655.

48. Chan, D.I., Prenner, E.J., and Vogel, H.J. **2006**. Tryptophan- and arginine-rich antimicrobial peptides: structure and mechanism of action, Biochimica et Biophysica Acta*,* (1758): 1184–1202.

49. Jenssen, H., Hamill, P., and Hancock, R.E., **2006**. Peptide antimicrobial agents, Clinical Microbiology Review*,* (19): 491–511.

50. Brogden, K.A., **2005**. Antimicrobial agents: pore formers or metabolic inhibitors in bacteria, Nature Reviews Microbiology, (3): 238–250.

51. Shai, Y., **1999**. Mechanism of the binding insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and non-selective membrane-lytic peptides, Biochimica et Biophysica Acta, (1462): 55–70.

52. Conlon, J.M., Abraham, N., Al-Ghaferi, B., and Leprince, J., **2007**. Strategies for transformation of naturally- occurring amphibian antimicrobial peptides into the therapeutically valuable anti-infective agents, Methods, (42): 349–357.

53. Hancock, R.E.W., **2005**. Mechanisms of action of newer antibiotics for gram-positive pathogens, Lancet Infectious Diseases, (5,): 209–218.

54. Chen, Y., Guarnieri, M.T., Vasil, A.I.,Vasil, M.L., Mant, C.T., and Hodges, R.S., **2007**. *Antimicrobial Agents and Chemotherapy*, (51):1398–1406.

55. Chen, Y., Mani, C.T.,Farmer, S.W.,Hancock, R.E.,Vasil, M.L., and Hodges, R.S., **2005**. Rational design of alpha helical antimicrobial peptides with enhanced activities and specificity/therapeutic index, Journal of Biological Chemistry, (280): 12316–12329.

56. Epand, R.M., and Vogel, H.J., **1999**. Diversity of antimicrobial peptides and their mechanisms of action, Biochimica et Biophysica Acta, (1462): 11–28.

57. Lee, D.L., Mant, C.T., and Hodges, R.S., **2003**. A Novel Method to Measure Self-association of Small Amphipathic Molecules: temperature profiling in reversed-phase chromatography, Journal of Biology Chemistry,(278): 22918–22927.

58. Courvalin, P., **2005**. Antimicrobial drug resistance: “prediction is very difficult, especially about the future”, Emerging Infectious Diseases, (11): 1503–1506.

59. Amaral, L., Engi, H., Viveiros, M., and Molnar, J., **2007**. Review. Comparison of multidrug resistant efflux pumps of cancer and bacterial cells with respect to the same inhibitory agents, In Vivo, (21): 237–244.

60. Tredan, O., Galmarini, C.M., Patel, K., and Tannock, I.F., **2007**. Drug resistance and the solid tumour microenvironment, Journal of the National Cancer Institute (99): 1441–1454.

61. David W. Hoskin, Ayyalusamy Ramamoorthy, Studies on anticancer activities of antimicrobial peptides, Biochimica et Biophysica Acta 1778 (2008) 357–375.

62. Wang, Z., Wang, G., **2004**. APD: the Antimicrobial peptide database, Nucleic Acids Research, (32): D590–D592.

63. Rozek, T., Wegener, K. L., Bowie, J. H., Olver, I. N., Carver, J. A., Wallace, J. C. Tyler, M. J., **2003**.The antibiotic and anticancer active aurein peptides from the Australian Bell Frogs *Litoria aurea* and *Litoria raniformis,* European Journal of Biochemistry, (267): 5330-5341.

64. Dennison, S. R. W., Harris, M., and David, F.P., **2006**. Anticancer alpha-helical peptides and structure-function relationship underpinning their interactions with tumour cell membranes, Current Protein and Peptide Science, (7): 487-499.

65. Pan, Y. L., Cheng, J. T.-J., Hale, J., Pan, J., Hancock, R. E. W., and Suzana,S.K.,**2007**. Characterization of the Structure and Membrane Interaction of the Antimicrobial Peptides Aurein 2.2 and 2.3 from Australian Southern Bell Frogs, Biophysical Journal (92): 2854–2864.

66. Mura, M., Dennison, S. R., Zvelindovsky, A. V., and Phoenix, D. A., **2013**. Aurein 2.3 functionality is supported by oblique orientated α-helical formation, Biochimica et Biophysica Acta, (1828): 586-594.

67. Li, X., Li, Y., Peterkofsky, A., Wang, G., **2006**. NMR studies of aurein 1.2 analogs, Biochimica et Biophysica Acta, (1758): 1203–1214.

68. (a) Duan, Y., and Kollman, P.A. **1998**. Pathways to a protein folding intermediate observed in 1μs simulation in aqueous solution, Science, (282): 740-744. (b) Duan, Y., Wang, L., and Kollman, P.A. **1998**. Proceedings of the National Academy of Sciences USA, (95): 9897-9902.

69. Daura, X., Jaun, B., Seebach, D., van Gunsteren, W.F., and Mark, A.E., **1998**. Reversible peptide folding in solution by molecular dynamics simulation, Journal of Molecular Biology, (280): 925-932.

70. Daura, X., van Gunsteren, W.F., and Mark, A. E. **1999**. Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations, Proteins (34): 269-280.

71. Mokoena, P. **2010**. Computational studies of the folding patterns of small and medium-size polypeptides, Thesis submitted at Durban University of Technology.

72. Rodriguez, A., Mokoena, P., Corcho, F., Bisetty, K., and Perez, J.J., **2011**. Computational study of the free energy landscape of the miniprotein CLN025 in explicit and implicit solvent, Journal of Physical Chemistry B, (115): 1440-1449.

73. Corcho, F.J., Mokoena, P., Bisetty, K., and Perez, J.J., **2009**. Molecular dynamics (MD) simulations of VIP and PACAP27, (91): 391-400.

74. Onufriev, A., Bashford, D., and Case, D.A., **2004**. Exploring protein native states and large-scale conformational changes with a modified generalized born model, Proteins: Structure, Function, and Bioinformatics, (55): 383-394.

75. Corcho, F. An introduction to molecular modeling and computer-aided drug design, http://www.tesisenxarxa.net/TESIS_UPC/AVAILABLE/TDX-0323104-11352//Chapter1.pdf.

76. Aleman, C., Karayiannis, N.C., Curco, D., Foteinopoulou, K., and Laso, M., **2009**. Computer simulations of amorphous polymers: from quantum mechanical calculations to mesoscopic models, Journal of Molecular Structure: Theochem, (898): 62-72.

77. Levitt, M., **1976**. A simplified representation of protein conformations for rapid simulation of protein folding, Journal of Molecular biology, (104): 59-107.

78. Shagidullin, R.R., Chernova, A.V., Katsyuba, S.A., Avvakumova, L.V., and Shagidullin Rif. R., **2004**. Energetics of intramolecular hydrogen bonds and conformations of ω-

diphenylphosphoryl- and ω-diphenylthiophosphoryl-substituted aliphatic alcohol molecules, Russian Chemical Bulletin, International Edition, (53): 55-59.

79. Wallin, S and Shakhnivich, E., **2008**. Understanding ensemble protein folding at atomic detail, Journal of Physics: Condensed Matter, (20): 283101-283112.

80. Post, C.B., and Dadarlat, V.M., **2006**. Molecular-dynamics simulations of biological macromolecules, International Tables for Crystallography, Vol F, Chapter 20.2: 489-495.

81. Marx, D. and Hutter, J., **2000**. Ab initio molecular dynamics: Theory and Implementation. Published in Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, Julich, NIC Series, Vol. 1, ISBN 3-00-005618-1, pp. 301-449.

82. http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html. **2009**. Theory of molecular dynamics simulations.

83. http://www-rohan.sdsu.edu/~spydell/md/md.html. **2009**. Molecular Dynamics and Visualization.

84. http://www.fisica.uniud.it/~ercolessi/md/md/node8.html. **2009**. Today's role of molecular dynamics.

85. Zhou, R., **2007**. Replica exchange molecular dynamics method for protein folding simulation, In Methods in molecular biology, Clifton, N.J. (350): 205-223.

86. Sugita Y., and Okamoto Y., **1999**. Replica-exchange molecular dynamics method for protein folding, Chemical Physics Letters, (314): 141-151.

87. Hukushima, K. and Nemoto, K., **1996**. Exchange Monte Carlo method and application to spin glass simulations, Journal of the Physical Society of Japan, (65): 1604-1608.

88. Okabe, T., Kawata, M., Okamoto, Y., and Mikami, M., **2001**. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble, Chemical Physics Letters, (335): 435-439.

89. Berendsen, H.J.C., **1984**. Molecular-Dynamics with Coupling to an External Bath, Journal of Chemical Physics, (81): 3684-3690.

90. Nosé, S., **1984**. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble, Molecular Physics, (52): 255-268.

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

92. Nosé, S., **1984**. A Unified Formulation of the Constant Temperature Molecular-Dynamics Method, Journal of Chemical Physics, (81): 511-519.

93. (a) Wu, X., and Brooks, B.R., **2003**. Self-guided Langevin dynamics simulation method, Chemical Physics Letters, (381): 512–518. (b) Adelman, S.A., and Doll, J.D., **1976**. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering Off Harmonic Solids, Journal of Chemical Physics, (64): 2375-2388.

94. Case, D. A., Darden, T. A. Cheatham, T .E. III., Simmerling, C .L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R .C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K .F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. **2006**. AMBER 9. University of California, San Francisco.

95. Toukmaji, A.Y., and Board. J.A., **1996**. Ewald summation techniques in perspective: A survey, Computer physics communications, (95): 73-92.

96. Darden, T., York, D., and Pedersen, L., **1993**. Particle Mesh Ewald: An N Log (N) method for Ewald sums in large systems, Chemical Physics, (98): 10089-10092.

97. Peng, J.W., and Wagner, G., **1992**. Mapping of spectral density functions using heteronuclear NMR relaxation measurements, Journal of Magnetic Resonance, (98): 308-332.

98. (a) van Gunsteren, W.F., and Karplus, M., **1981**. Effect of constraints, solvent and crystal environment on protein dynamics, Nature, (293): 677-678. (b) Heiner, A.P., Berendsen, H.J.C., and van Gunsteren, W.F., **1992**. MD simulation of subtilisin BPN′ in a crystal environment, PROTEINS: Structure Function and Genetics, (14): 451-464. (c) van Gunsteren, W.F., and Mark, A.E., **1992**. On the interpretation of biochemical data by molecular dynamics computer simulation, European Journal of Biochemistry, (204): 947-961 (d) Morikami, K., and Saito, M., **1994**. Molecular dynamics study on the stability of ions around human lysozyme in the crystal condition, Computer Physics Communications, (225): 196-201. (e) van Nuland, N.A. J., Wiersma, J.A., van der Spoel, D., de Groot, B.L., Scheek, R.M., and Robillard, G.T.,**1996**. Phosphorylation-induced torsion-angle strain in the active center of HPr, detected by NMR and restrained molecular dynamics refinement, Protein Science, (5): 442-446.

99. Woody, R.W., **1995**. Circular Dichroism, Methods in Enyzmology. (246): 34-70.

100. Bayley, P.M., Nielsen, E.B., and Schellman, J.A., **1969**. The rotary properties of molecules containing two peptide groups: Theory, Journal of Physical Chemistry, (73): 228-243.

101. (a) Manning, M.C., and Woody, R.W., **1989**. Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor, Biochemistry, (28): 8609-8613. (b) Manning, M.C., and Woody, R.W., **1991**. Theoretical CD studies of polypeptide helices: Examination of important electronic and geometric factors, Biopolymers, (31): 569-586. (c) Marconi, G., Monti, S., Mayer, B.,

and Khler, G.J., **1995**. Circular Dichroism of Methylated phenols Included in .beta.-Cyclodextrin. An Experimental and Theoretical study, Journal of Physical Chemistry, (99): 3943-3950.

102. Fleischhauer, J., Groetzinger, J., Kramer, B., Krueger, P., Wollmer, A., Woody, R.W., and Zobel, E., **1994**. Calculation of the circular dichroism spectrum of cyclo (L-tyr-L-tyr) based on a molecular dynamics simulation, Biophysical Chemistry, (49): 141-152.

103. Devlin, F.J., Finley, J.W., Stephens, P.J., and Frisch, M.J., **1995**. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields: a comparison of local, nonlocal, and hybrid density functionals, Journal of Physical Chemistry, (99): 16883-16902.

104. Kalk, A., and Berendsen, H.J.C., **1976**. Proton magnetic relaxation and spin diffusion in proteins, Journal of Magnetic Resonance, (24): 343-366.

105. Tropp, J., **1980**. Dipolar relaxation and nuclear overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances, Journal of chemical physics, (72): 6035-6043.

106. Ernst, R.R., Bodenhausen, G., and Wokaun, A., **1987**. Principles of nuclear magnetic resonance in one and two dimensions, Oxford: Clarendon Press.

107. Arfken, G., **1970**. Mathematical methods for physicists, New York: Academic Press.

108. Torda, A.E., Scheek, R.M., and van Gunsteren, W.F., **1989**. Time-dependent distance restraints in molecular dynamics simulations, Chemical Physics Letters, (157): 289-294.

109. Torda, A.E., Scheek, R.M., and van Gunsteren, W.F., **1990**. Time-averaged nuclear overhauser effect distance restraints applied to tendamistat, Journal of Molecular Biology, (214): 223-235.

110. Post, C.B., **1992**. Internal motional averaging and three-dimensional structure determination by nuclear magnetic resonance, Journal of Molecular Biology, (224): 1087-1101.

111. Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M., and Ernst, R.R., **1992**. Inuence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution, Journal of American Chemical Society, (114): 2289-2302.

112. Palmer III, A.G., and Case, D.A., **1992**. Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide, Journal of American Chemical Society, (114): 9059-9067.

113. Lipari, G., Szabo, A., and Levy, R.M., **1982**. Protein dynamics and NMR relaxation: comparison of simulations with experiment, Nature, (300): 197-198.

114. Olejniczak, E.T., Dobson, C.M., Karplus, M., and Levy, R.M., **1984**. Motional averaging of proton nuclear overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme, Journal of American Chemical Society, (106): 1923-1930.

115. (a) LaFargaCPL: CLASTERIT: Project Info. Available at: https://lafarga.cpl.upc.edu/projects/clusterit [Accessed June 12, 2008]. (b) Corcho, F., Canto, J., and Perez, J.J., **2004**. Comparative analysis of the conformational profile of substance P using simulated annealing and molecular dynamics, Journal of Computational Chemistry, (25): 1937-1952.

116. Srinivasan, R., and Rose, G.D., **1999**. A physical basis for protein secondary structure, Proceedings of the National Academy of Sciences USA, (96): 14258.

117. Ambroggio, E. E., F. Separovic, J. H. Bowie, G. D. Fidelio, and L. A. Bagatolli., **2005**. Direct visualization of membrane leakage induced by the antibiotic peptides: maculatin, citropin, and aurein. Biophysical Journal, (89): 1874–1881.

118. Balla, M. S., J. H. Bowie, and F. Separovic., **2004**. Solid-state NMR study of antimicrobial peptides from Australian frogs in phospholipid membranes. European Biophysical Journal, (33):109–116.

119. Marcotte, I., Wegener, K. L. Lam, Y. H., Chia, B. C., de Planque, M. R., Bowie, J. H., Auger, M., and Separovic, F., **2003**. Interaction of antimicrobial peptides from Australian amphibians with lipid membranes. Chemistry and Physics of Lipids, (122):107–120.

120. Rozek, T., Bowie, J. H., Wallace, J. C., and Tyler, M. J., **2000**. The antibiotic and anticancer active aurein peptides from the Australian bell frogs Litoria aurea and Litoria raniformis. Part 2. Sequence determination using electrospray mass spectrometry. Rapid Communications in Mass Spectrometry, (14): 2002–2011.

121. Boland, M. P., and Separovic, F., **2006**. Membrane interactions of antimicrobial peptides from Australian tree frogs. Biochimica et Biophysica Acta. (1758):1178–1183.

122. Pukala, T. L., Bowie, J. H., Maselli, V. M., Musgrave, I. F., and Tyler, M. J., **2006**. Host-defence peptides from the glandular secretions of amphibians: structure and activity. Natural Product Reports, (23):368–393.

123. Pastor, R.W., Brooks, B.R., and Szabo, A., **1988**. An analysis of the accuracy of langevin and molecular-dynamics algorithms, Molecular Physics, (65): 1409–1419.

124. Ward Jr, J.H., **1963**, Hierarchical grouping to optimized an objective function, Journal of American Statistical Association, (58): 236–244.

125. Tobias, D.J. and Brooks, C.L.I., **1992**. Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results, Journal of Physical Chemistry, (96): 3864–3870.