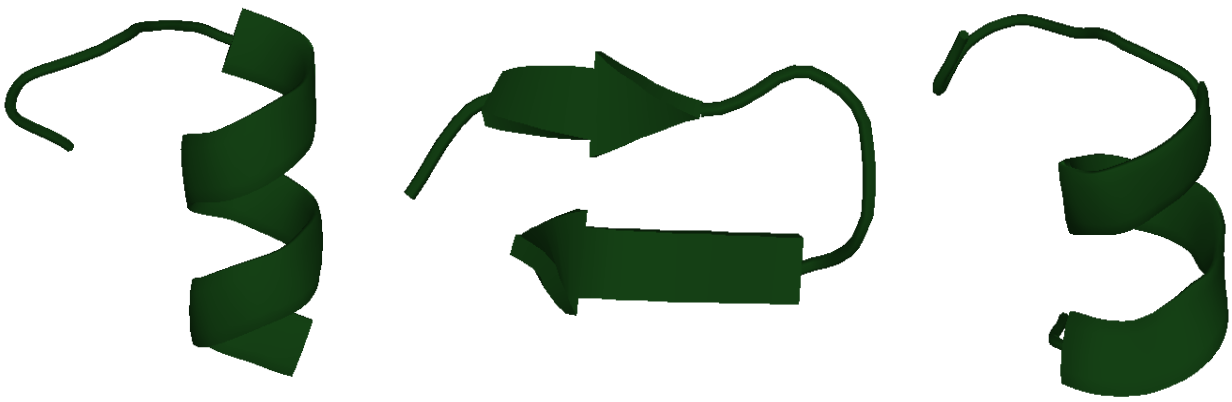
An Assessment of the Conformational Profile of Bombesin and Its Mammalian Analogues using Computational Chemistry Methods

Parul Sharma



DURBAN UNIVERSITY OF TECHNOLOGY

March 2011

An Assessment of the Conformational Profile of Bombesin and Its Mammalian Analogues using Computational Chemistry Methods

Submitted in fulfillment of the requirements of the Degree of Doctor of Technology: Chemistry, in the Faculty of Applied Sciences at the Durban University of Technology

Parul Sharma

Promoter: Prof Krishna Bisetty

PREFACE

The work described in this thesis was performed by the author under the supervision of Professor Krishna Bisetty at Durban University of Technology, Durban, South Africa, and co-supervision of Professor Juan J Perez at UPC, Barcelona, Spain, from 2008-2011. The study presents original work by the author and has not been submitted in any form to another tertiary institution or university. Where use is made of the work of others, it has been clearly stated in the text.

Signed: Date:

Parul Sharma

Signed: Date:

Prof. Krishna Bisetty (Promoter)

TABLE OF CONTENTS

DECLARATION

I declare that the thesis herewith submitted for the D Tech: Chemistry at the Durban University of Technology has not been previously submitted for a degree at any other University.

…….

ABSTRACT

Understanding the dynamics and mechanism of protein folding continues to be one of the central problems in molecular biology. Peptide folding experiments characterize the dynamics and molecular mechanisms of the early events of protein folding. However, generally the highly flexible nature of peptides makes their bioactive conformation assessment reasonably difficult as peptides fold at very fast rates experimentally, requiring probing on the nanosecond time resolution. On the other hand, determining the bioactive conformation of biological peptides is a requirement for the design of peptidomimetics in computer-aided drug design.

Peptides offer a unique opportunity to bridge the gap between theoretical and experimental understanding of protein folding. Therefore, the present work focuses on the exploration of the conformational space of biologically active neuropeptides with the aim of characterizing their conformational profile. Specifically, bombesin, neuromedin B (NMB) and neuromedin C (NMC), have been chosen for the current investigations. These peptides are widely distributed in the gastrointestinal tract, spinal cord and brain, and are known to elicit various physiological effects, including inhibition of feeding, smooth muscle contraction, exocrine and endocrine secretions, thermoregulation, blood pressure and sucrose regulations and cell growth. These peptides act as a growth factor in a wide range of tumours including carcinomas of the pancreas, stomach, breast, prostate, and colon. This work is intended to get some insight into the performance of different procedures used to explore the configurational space to provide an adequate atomic description of these systems.

Different methodological studies involving utilization of molecular dynamics (MD), multicanonical replica exchange molecular dynamics (REMD) and simulate annealing (SA) are undertaken to explore the folding characteristics and thermodynamics of these neuropeptides. MD and REMD calculations on bombesin peptide have revealed its dual conformational behaviour never discovered before and is described in chapter 3. These results explain the known structure-activity studies and open the door to the understanding of the affinity of this peptide to two different receptors: BB1 and BB2. In the case of NMC, REMD calculations are carried out in explicit and implicit solvents, using the Generalized Born (GB) surface area, and are then complemented with two additional MD simulations performed using Langevin and Berendsen thermostats. The results obtained clearly reveal that REMD, performed under explicit solvent conditions, is more efficient and samples preferentially folded conformations with a higher content of β and γ turns. Moreover, these results show good agreement with the experimental results supporting the role of two β-turns for its biological action, as reported in the literature. Finally, the results obtained from MD, REMD and SA calculations on NMB reveal that the peptide has a tendency to adopt both turns and helices suggesting its two different receptor recognizing and binding conformations during its biological action. Hence, the present work provides comprehensive information about the conformational preferences of neuropeptides which could lead to a better understanding of their native conformations for future investigations and point the way towards developing their new antagonists.

ACKNOWLEDGEMENTS

Pursuing a Ph.D. degree is both a painful and an enjoyable experience. It is just like climbing a high peak step by step, accompanied with bitterness, hardships and frustration on one hand and encouragement, trust and help of so many people on the other. When I found myself at the top enjoying the beautiful scenery, I realized that in fact, it was the teamwork that got me there. Though words cannot express my gratitude to all those people that helped me, I will still extend my heartiest appreciation to them.

It gives me immense pleasure to express my sincere gratitude and respect to my worthy guide, ***Prof K. Bisetty*** *(Department of Chemistry, DUT), for introducing me to the enthralling field of "Computational Chemistry". I proclaim my indebtedness to him for the dexterous guidance, untiring efforts, constructive criticism and providing continuous enthusiasm during the entire tenure of my research work. Being always empowered by his graceful and friendly behaviour, I enjoyed every moment of my research period. It was a delight to work under his kind supervision.*

I am indebted to ***Prof Juan J. Perez*** *(Molecular Engineering Laboratory, UPC, Barcelona) for his inspiring and knowledgeable guidance for my research work.*

I wish to extend my thanks to Prof G.G. Redhi (HOD) and all the faculty members, Department of Chemistry, DUT, for their moral support and encouragement.

I owe my thanks to my family for keeping their courage and allowing me to complete this research project overseas. Despite being far from me, their blessings and their trust in me motivated me to achieve my and my father's (used to call me Dr Parul even when I was just a postgraduate) dearest aspiration. Special thanks to my mother whose continuous suggestions, encouragement and moral support helped me to surpass all the hurdles of this journey. Also

there is an equal contribution of my little brother, Abhinav (Bhinni), for his regular sweet motivational chats which eased my stress at all times. Blessings of my grandparents (ammaji and babaji) are gratefully acknowledged. Also the support and best wishes of all my relatives and loved ones is highly appreciated.

It is high time to acknowledge to my fellow researchers, Paul and Myalo for sharing thoughts and special thanks to Dr Parvesh Singh for his invaluable suggestions and emotional support during the project. My sincere thanks to our Spanish researchers Dr F.J. Corcho, Alex Rodriguez and Patricia Gomez (Molecular Engineering Laboratory, UPC, Barcelona) for their assistance related to my research and fruitful stay in Spain.

Above all I express my heartfelt gratitude to "God" who has always been the impelling force behind man's efforts. Having unflinching faith, my sincere and earnest gratitude is extended to almighty, which helped me in the achievement of the given task.

(PARUL SHARMA)

LIST OF ABBREVIATIONS

MC : Monte Carlo

MD : Molecular Dynamics

3D : three-dimensional

NMR : Nuclear Magnetic Resonance spectroscopy

SA : Simulated Annealing

REMD : Replica exchange molecular dynamics

Cln : Chignolin

PACAP : pituitary adenylate-activating polypeptide

VIP : vasoactive intestinal peptide

NMB : Neuromedin B

NMC : Neuromedin C

GRP : Gastrin-releasing peptide

GPCR/7TM : G protein-coupled receptor/seven-transmembrane

MM : Molecular mechanics

E_{tot} : Total energy

E_{str} : Bond-stretching energy

E_{bend} : Angle bending energy

E_{tors} : Torsional angle energy

E_{vdw} : Van der Waals energy

E_{elec} : Electrostatic energy

IR : Infra-red

DNA : Deoxyribonucleic acid

RNA : Ribonucleic acid

PBC : Periodic boundary condition

PME : Particle Mesh Ewald

FFT : fast Fourier Transform

GB : Generalized Born

PE : Poisson Equation

NOEs : Nuclear Overhauser Effect Intensities

CD : Circular Dichroism

VCD : vibrational circular dichroism

RMSD : root mean square deviation

PDB : Protein Data Bank

BPTI : bovine pancreatic trypsin inhibitor

CHARMM : Chemistry at HARvard Molecular Mechanics

Sander : Simulated annealing with NMR-derived energy restraints

Glp : pyroglutamic acid

DMSO : dimethylsulfoxide

TFE : Trifluoroethanol

ff : force field

OBC : Onufriev, Bashford and Case

HB : hydrogen bond

Aib : α-aminoisobutyric acid

NMBR or BB1 : neuromedin B receptor receptor

GRPR or BB2 : gastrin releasing peptide receptor

2D : two dimensional

BRS3 or BB3 : bombesin receptor subtype 3

SAR : structure activity relationship

SDS : sodium dodecyl sulfate

LD : long distances

SD : short distances

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This chapter provides an overview of molecular modeling and computational chemistry. 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 which uses all the different strategies necessary to model and deduce information of a system at an atomic level, including simulation methodologies used in computational chemistry, such as the computation of the energy of a molecular system, Monte Carlo (MC) methods or molecular dynamics (MD) **[2]**. Moreover the identification of biomolecular moieties involved in the interaction with a specific receptor permits us to understand the molecular mechanism responsible for its specific biological activity using molecular modeling techniques. This knowledge is aimed specifically at designing new active molecules that can successfully be used as 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 accuracy of the model, and to modify them if necessary, in order to gain a better understanding of the system.

1.2 Proteins and Peptides

Proteins are biological macromolecules, shaped by millions of years of evolution into structures allowing them to perform specific functions. Thousands of different proteins exist in different organisms, and the functions they perform are diverse. They catalyze most cellular functions such as immune response, catalysis of metabolic reactions and signal transduction under normal and disease conditions **[3-5]**. To be able to perform their biological function, proteins fold into one or more specific spatial conformations, driven by a number of non-covalent interactions such as hydrogen bonding **[6]**, ionic interactions **[7]**, van der Waals forces **[8]** and hydrophobic packing **[9]**.

However, the understanding of the chemical thermodynamics and mechanisms of peptide and protein folding remains one of the current challenges in modern molecular biology **[3-4]**. Despite numerous contributions from different 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 **[10-13]**.

In principle 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.

The primary structure refers to the different amino acid sequence held together by covalent bonds generated during the course of protein biosynthesis or translation. Secondary structure corresponds to the highly regular local sub-structures, with the α-helix and the β-strand being the two main types of secondary structures. These secondary structures are defined by patterns of hydrogen bonds between the main-chain peptide groups **[14-16]**. An α-helix is formed when the spacing of the amino acid residues participating in a hydrogen bond occurs between the i and i + 4 positions. Smaller spacing between residues in the i and i + 3 positions result in a 3_{10}-helix. On the other hand a β-sheet is formed when two strands are bridged by hydrogen bonds involving alternating residues on each participating strand. A tertiary structure represents the three dimensional (3D) structure of a single protein molecule where the α-helices and the β-sheets are folded into a compact globule. Hydrogen bonds also play an important role in the formation of tertiary structures through the interaction of side chains. Quaternary structure is a larger assembly of several protein molecules or polypeptide chains stabilized by the non-covalent interactions and disulfide bonds similar to those present in tertiary structures.

1.3 Folding studies on proteins and peptides

Protein folding is the self-assembly process that turns a polypeptide chain of a foldable sequence into a stable three-dimensional conformation. The study of protein folding has three aspects: thermodynamics, kinetics, and structure prediction. A protein consisting of n amino acids can in principle have 3^n conformations, due to the three possible combinations of the phi (ϕ) and psi (ψ) angles. If the protein randomly searches all the available conformations, it would take longer than the lifetime of the universe, known as the Levinthal's paradox **[17]**. Christian Anfinsen **[18]**, was awarded the Nobel Prize in chemistry for his work on ribonuclease concerning the connection between the amino acid sequence and the biologically active conformation. Anfinsen demonstrated that the correct refolding of the unfolded Ribonuclease A into its native and enzymatically active structure occur spontaneously in free solution **[18]**. All information determining the native structure is fully contained in the amino acid sequence of a protein **[19]**.

Protein folding thus seems to occur along certain pathways, thereby simplifying the folding process by splitting it up into sequential steps. Stabilized folding intermediates were proposed, defining the individual steps of such a pathway **[20-22]**. Folding intermediates possesses stabilized structural elements, mainly of secondary structural origin, in combination with the unstructured regions. A mechanistic pathway of folding drastically reduces the amount of possible conformations during a folding process, thus allowing effective protein folding on biologically relevant timescales. Of the two main models used in the prediction of folding pathways **[23]**, one model predicts that the initially formed stable secondary structural elements collapse into

tertiary structures by diffusion and collision with other secondary structures, referred to as the framework model **[24]** or diffusion-collision model **[25]**. The second referred to as the hydrophobic collapse model **[26-27]**, 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. Jackson and co-workers **[28]** showed that proteins can actually fold without forming detectable intermediate structures or that they can form secondary and tertiary structures in parallel during the hydrophobic collapse **[29]**. These observations lead to the proposal of the nucleation condensation mechanism **[30]**, which combine features from both the framework mechanism and the hydrophobic collapse model. The existence of the multiple folding pathways for the different proteins led to the proposal of an energy landscape model for protein folding as depicted in Figure 1.

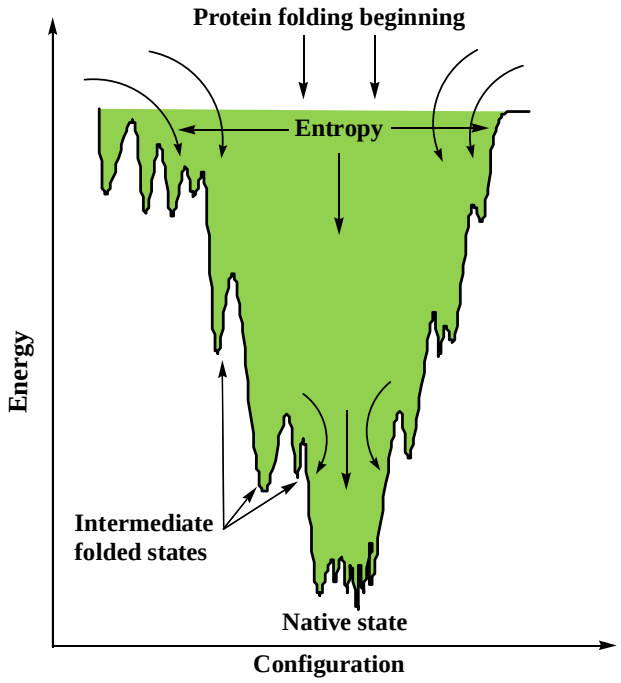


Figure 1.1 Folding-energy landscape for a protein molecule (Funnel type).

The folding process is described by an energy landscape or a folding funnel with a vast array of down-hill routes to the native state in a more or less rugged surface **[31-32]**. Dents in the funnel wall indicate the local energy minima in which the proteins may get trapped in unfavorable intermediate states during the folding process **[33]**. 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. Solvation effects may also be important. Various schemes are now available for calculating the solvation free energy of a conformation that may be added as an additional term to the intramolecular energy. Whether a denatured protein is prone to intramolecular aggregation or reaches the native state efficiently depends on the rate of the folding process, i.e., how fast a globular structure is reached in which the hydrophobic surfaces are minimally exposed. How a given amino acid sequence encodes a defined three-dimensional structure is however not fully understood.

The amino acid sequences of 50000 proteins are currently available, while the number of three-dimensional protein structures available in the protein databank (PDB) is about 2000. The amount of work required to determine a protein structure using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy is considerable. Moreover, it is not always 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 **[34]**. Also good quality NMR spectra of proteins are not easily available because of their size and limited

solubility. 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, and it works by using statistical knowledge of the relationship between the structures deposited in the PDB and the sequence of the protein to model **[35, 36]**. However, these methods cannot yet reliably predict the three-dimensional structure of a protein from an amino-acid sequence. On the other hand, the utilization of computational molecular dynamics simulations has played a significant role to describe the folding-unfolding pathways of proteins and peptides, and is widely discussed in literature **[37]**. The conformational profile of a system can be assessed either through a topographical exploration of the potential energy surface using methods like simulated annealing (SA) **[38]**, or in the configurational space using Monte Carlo or molecular dynamics (MD) simulation **[2].** However, the sampling efficiency of MD simulations is sometimes hampered by the very rugged free energy surface of proteins. One of the methods used to improve the sampling of the configurational space, is the Replica Exchange Molecular Dynamics method (REMD), where several loosely coupled copies or replicas of the peptide system are simulated at different temperatures simultaneously and the information between the different trajectories is exchanged according to an established swapping probability **[39]**. The process is fully parallelizable, taking advantage of present computer architectures available.

Although obtaining the whole folding mechanism of proteins through MD has remained elusive until now **[40]**, 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 **[41-42]**. Previous studies undertaken by our research group at Durban University of Technology (DUT) **[43]**, involved a series of MD simulations applied to small and medium-size polypeptides to assess the thermodynamics of their folding characteristics. The computational procedures were validated on the 10-residue long chignolin-like synthetic mini-protein (CLN025) using REMD calculations both in explicit and implicit solvents **[44]**. Subsequent case studies to assess the folding conformations of peptides of different lengths including, the 27-residue pituitary adenylate-activating polypeptide 27 (PACAP27) and the 28-residue vasoactive intestinal peptide (VIP) were undertaken **[45]**. Results from these case studies revealed that 200 ns MD trajectories and 100 ns REMD simulation for longer polypeptides were adequate in the exploration of the configurational space of medium-sized polypeptides **[43-45]**. Accordingly in this study similar MD and REMD protocols were employed to fully explore the configurational space of medium-sized neuropeptides. Specifically, the bombesin and its mammalian analogues neuromedin B (NMB) and neuromedin C (NMC) were chosen.

Bombesin and neuromedins (NMB and NMC) play a significant role in numerous physiological or pathological processes **[46]**. Generally these include diverse effects in the central nervous system **[47-48]**, potent developmental effects **[49-50]**, effects in the gastrointestinal tract **[51-52]** and effects on the immune system **[53-54]**. Binding studies, pharmacological studies and recent cloning studies have provided evidence that these peptides exert their biological effects by interacting with two different subtypes of bombesin receptors, a gastrin releasing peptide (GRP)-preferring bombesin receptor and a neuromedin B (NMB)-preferring bombesin

receptor **[55-57]**. Each bombesin receptor subtype is a member of the G protein-coupled seven-transmembrane (GPCR or 7TM) superfamily, activation of which in all cells examined, stimulates increases in phospholipase C, mobilization of cellular Ca^{2+} and generation of inositol phosphates **[58-59]**. However, the two receptor subtypes differ in the affinities for GRP and NMB as well as for different classes of bombesin receptor antagonists **[60-61]**.

Numerous structure function studies on bombesin-related peptides attempting to define the peptide binding domain and the requirements for receptor activation are reported in the literature **[62-64]**. These studies revealed that the carboxyl terminus is the biologically active portion of the molecule. However, a detailed structure-function information is lacking. Jenson and co-worker **[65]** demonstrated that some tissues used for structure-function studies such as Swiss 3T3 cells or pancreatic acinar cells possess only GRP-preferring bombesin receptors, however none of the interactions with these various tissues, or functions measured are now known to be due to occupation of NMB-preferring bombesin receptors. Therefore almost no structural information on this peptide is available for affinity or efficacy at the NMB-preferring bombesin receptor. Furthermore, there is almost no information on the important amino acids of the peptides for determining receptor specificity for any bombesin receptor subtype. Hence, the broader goals of this work was to get a deeper understanding of the structure-activity relationships of these neuropeptides and to understand the different states they exhibit by sampling their conformational profiles, which can eventually provide information in the design of novel peptidomimetics.

The present work is organized in six chapters. The first three chapters of the introduction addresses the current state of knowledge on the fundamental and general aspects of protein folding, an overview of the different simulation techniques used to predict the structures and properties of biomolecules, followed by a brief overview of the experimental techniques used in the validation of the computational results. The computational section of the thesis is organized in three other chapters, and reports on the main computational results obtained for bombesin and its mammalian analogues NMB and NMC respectively. The overall conclusions and recommendations are dealt with in the last chapter.

1.4 Aims and Objectives

The aims of this study were to:

- use different sampling techniques to explore the conformational space of neuropeptides.
- investigate the effect of different thermostat algorithms on the folding pattern of neuropeptides
- validate the performance of computational protocols with the available experimental results.

The objectives of this study were to:

- investigate the performance of the different thermostats and methodologies used in the implicit solvent model to assess the features of a fourteen residue bombesin peptide.

- study the conformational features of neuromedin C, a ten residue structural analogue of bombesin, also the performance of REMD and MD using implicit and explicit water models was investigated.
- study the folding characteristics and conformational profile of neuromedin B, a ten residue bombesin-like peptide, using MD, REMD and SA techniques.

CHAPTER 2

THEORETICAL AND EXPERIMENTAL TOOLS

This chapter provides an overview mainly of the molecular mechanics method used to predict the structures and properties of biomolecules. Different simulation techniques, such as molecular dynamics (MD), Replica exchange molecular dynamics (REMD) and simulated annealing (SA) are described. Energy minimization procedures, effect of charge and solvent, employment of periodic boundary conditions and the particle mesh ewald method normally used in the simulation process, are also explained. The experimental techniques including Nuclear Magnetic Resonance (NMR), circular Dichroism (CD) and X-ray crystallography are presented. A brief comparison of computational simulations with experimental methods is also presented.

2.1 Molecular Mechanics (MM)

Molecular mechanics (MM) use the laws of classical physics to predict the structure and properties of molecules. There are different MM formulations, each characterized by its own particular force field. The main components of a force field includes a set of equations defining how the potential energy of a molecule varies with the locations of its component atoms, and a series of atom types defined by the characteristics of an atom of the element within a specific chemical context **[66-67]**. The atom types, mentioned above, have different characteristics and exhibit different behaviour for atoms, depending upon their environment in the structure. For example, a carbon atom in a carbonyl group is treated differently from one bonded to three hydrogens. The atom types therefore depend on the hybridization, charge and the type of other atoms to which it is bonded. MM considers a molecule to be a collection of masses

interacting with each other through harmonic forces. Thus, the atoms in a molecule are treated as balls of different sizes joined together by springs of variable strength and equilibrium distances (bonds). This simplification allows for the use of molecular mechanics as a fast computational model that can be applied to molecules of any size. In the course of a calculation, the total energy is minimized with respect to the atomic coordinates, and is made up of a sum of different contributions that calculate the deviations from the equilibrium values of bond lengths, angles and torsions plus non-bonded interactions **[68-70]**:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} + \ldots.$$

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 the above equation 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:

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

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:

$$E_{vdw} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

Where A_{ij} and B_{ij} are the repulsive and attractive term coefficients respectively, and r_{ij} is the distance between atoms i and j.

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

$$E_{elec} = \frac{1}{\varepsilon}\frac{Q_1 Q_2}{r_{ij}}$$

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 the description of the accuracy and efficiency of intramolecular energy surfaces. Thus, the equilibrium parameters of bond lengths and bond angles are derived from the force constants used in the potential energy function defined in the force field **[68-69]**. Deviations from any of the equilibrium values inevitably result in the increase of the total energy of a molecule, resulting in the total energy of the system taken as a measure of intramolecular strain relative to a hypothetical molecule with equilibrium values **[68-69]**. 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 **[68, 71-72]**.

2.2 Molecular Dynamics (MD)

Molecular dynamics (MD) is described as a form of computer simulation in which the atoms and molecules are allowed to interact for a period of time by approximations of known physical attributes, resulting in the simulation of the motion **[73]** for a system of particles **[74]**. Since it is impossible to find the properties of such complex systems analytically, MD simulation circumvents this problem by using numerical methods. It

represents an interface between laboratory experiments and theory, and can be understood as a "virtual experiment". MD probes the relationship between molecular structure, movement and function. MD is a multidisciplinary method whose laws and theories stem from mathematics, physics, and chemistry, and it employs algorithms from computer science and information theory. 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. **[74-76]**. A graphical flowchart for the general molecular dynamics protocol is shown in Figure 2.1

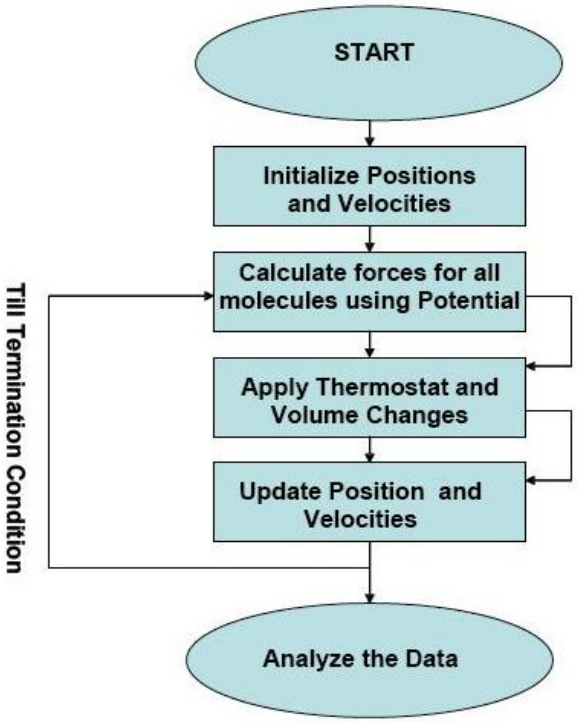


Figure 2.1 Flowchart showing the general molecular dynamics protocol.

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, 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 momenta 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 **[77]**.

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, have been shown to introduce significant errors and artificial behaviour in simulations **[68, 78]**.

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 **[79]** 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 **[80]**. 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 **[68, 78]**.

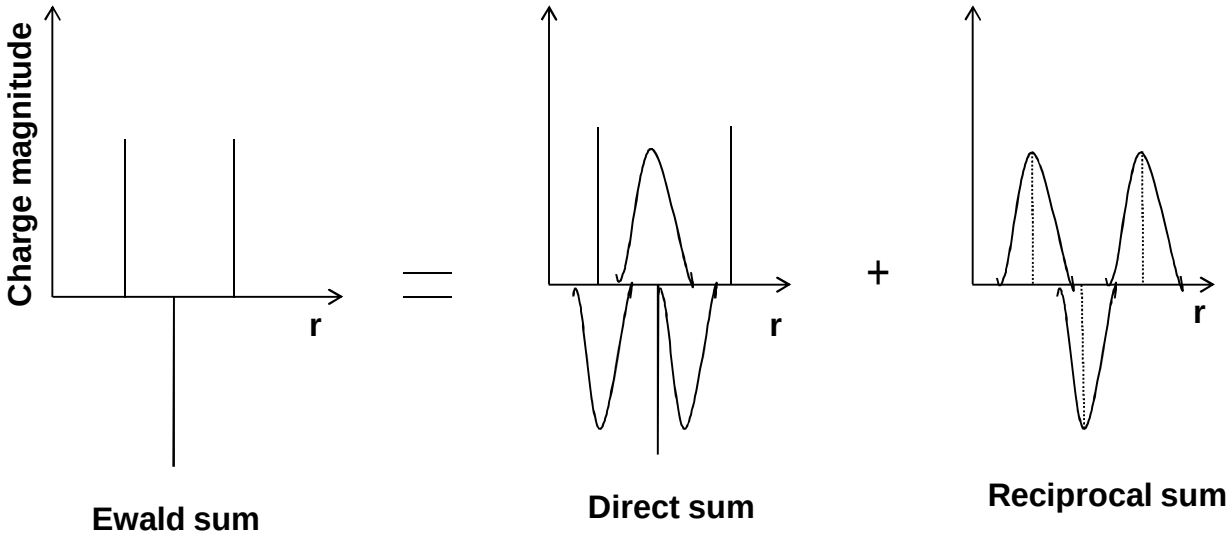


Figure 2.2 Splitting of charges into discrete and smeared distributions in the real and reciprocal space.

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 **[81]**. 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 **[68, 78]**.

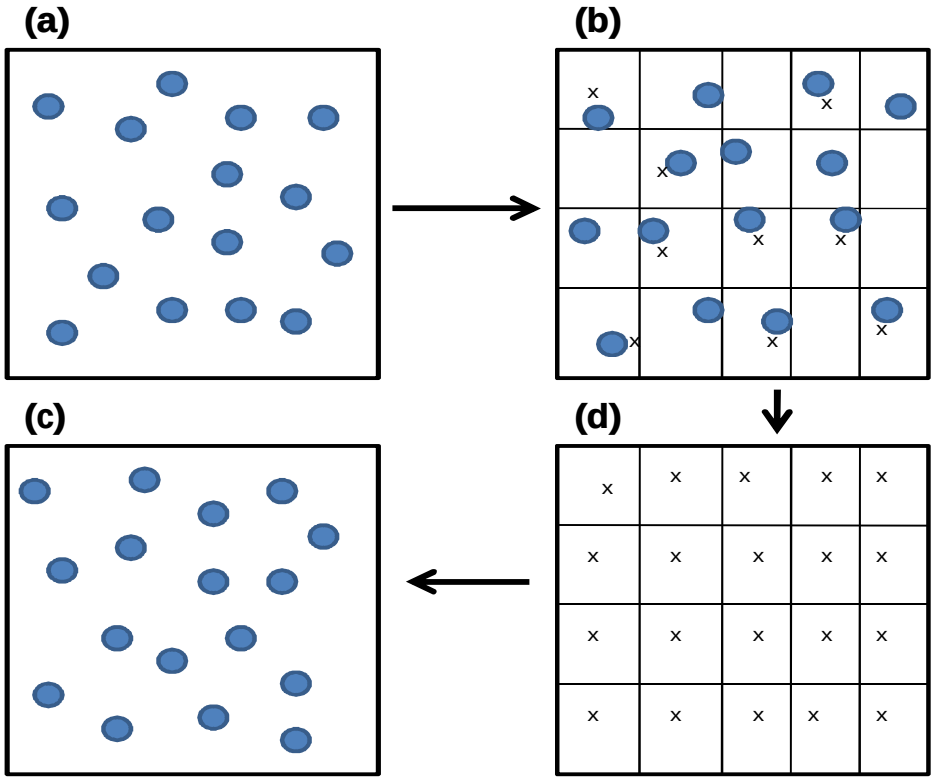


Figure 2.3 A 2D schematic of particle-mesh technique used in most Fourier-based methods **(a)** A system of charged particles. **(b)** The charges are interpolated on a 2D grid. **(c)** Using FFT, the potential and forces are calculated at grid points. **(d)** Interpolate forces back to particles and 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 **[82].** Nosé-Hoover thermostat is an integral type of thermostat, and introduces additional degrees of freedom (momentum) into the Hamiltonian of a system **[83-85]**. The Langevin thermostats follow the Langevin equation of motion instead of Newton's equation of motion **[86]**. 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 Replica Exchange Molecular Dynamics (REMD)

Replica-Exchange Molecular Dynamics (REMD) is a technique used to enhance sampling relative to a standard molecular dynamics simulation by allowing systems of similar potential energies to sample conformations at different temperatures, allowing for the exploration of new conformational space.

REMD is a computational method that couples MD trajectories at different temperatures with a random swapping probability process for efficient sampling of the conformational space **[87]**. The basic idea of REMD is to simulate different replicas of the system at the same time, but at different temperature values ranging from the desired temperature to a high temperature at which the replica can easily go over higher potential energy barriers **[87]**. Each replica evolves independently by MD, and only the neighbouring temperatures are swapped at fixed time intervals. This exchange is accepted or rejected based on a Metropolis acceptance criterion that guarantees the detailed balance **[10, 87]**. The direction of this exchange of neighbouring replicas is chosen at random **[87]**.

The biggest gap between MD models and biophysical reality exists in the accessible timescales and ensemble sizes. Biophysical experiments are increasingly performed on single molecules, but even with the availability of high end computational resources, the simulation of atomic-scale motion is far slower than movement in reality. Current MD simulations generate trajectories at a rate equivalent to a slow-motion movie playing with a speed approximately 10-15 orders of magnitude slower than reality. Even if this speed doubles every year for the next 30 years, simulations will not reach parity with real time. Therefore, no efforts are spared in making the best

possible use of the limited simulation time. In addition to speeding up the simulations, one may attempt to accelerate the process being simulated. REMD is a technique that exploits the physics of Brownian motion to achieve this. In any given MD simulation time, the warmer atoms thus explore a larger range of their available conformation space, which itself increases as the higher energy state become more accessible at higher temperatures. The downside to increased thermal motion is the reduced stability. Proteins unfold at elevated temperatures because their thermal energy exceeds the strengths of the interatomic interactions that constrain a structure. A warm polypeptide chain is thus more likely to encounter states similar to its cold folded conformation, but is likely to traverse this state quickly to continue exploring the vast conformational landscape accessible at higher temperatures. REMD **[87]** seeks to combine the benefits of high temperature rapid exploration with low-temperature stability. Multiple trajectories of identical systems (replicas) are simulated in parallel at different temperatures. At certain intervals, the potential energies of the replicas with neighbouring temperatures are compared. If the structure at the higher temperature has found a lower potential energy conformation, it is exchanged with the other structure and continues its trajectory at the lower temperature (Figure 2.4).

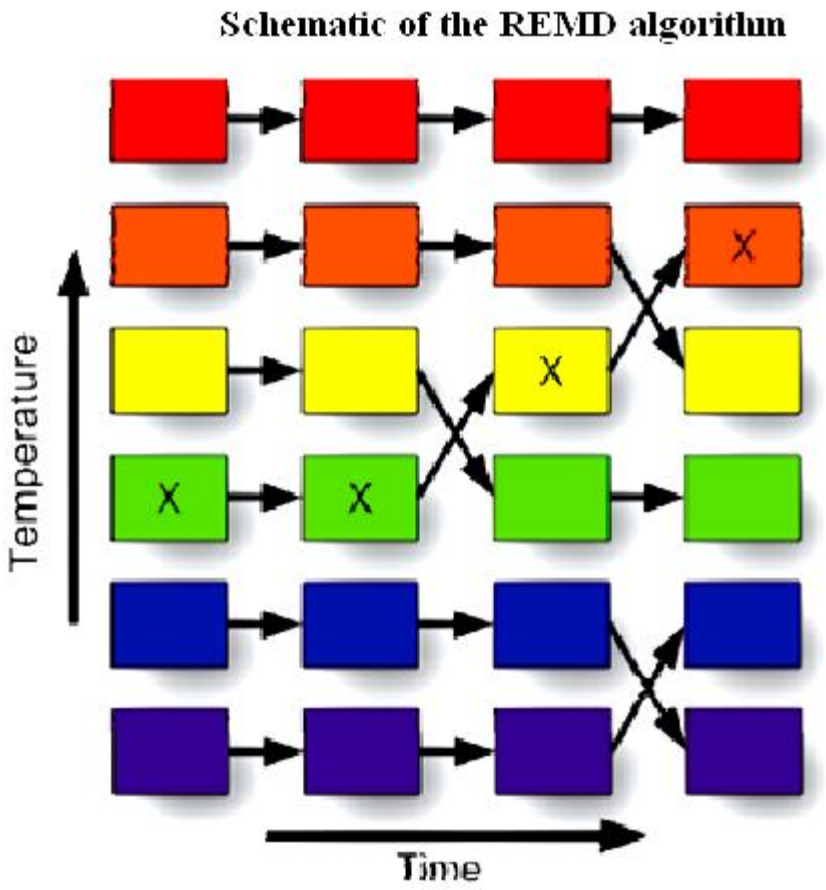


Figure 2.4 Replicas are simulated in parallel at different temperatures. At specified intervals, exchange probabilities are calculated based on a Metropolis acceptance criterion and neighbouring replicas are swapped accordingly.

The high temperature structures may still be exchanged, with a probability given by the potential energy difference between the two replicas, mathematically expressed by the Metropolis criterion **[87]**

$$P(\leftrightarrow) = \min(1, e^{-(\beta_2-\beta_1)(U1-U2)})$$

Where P is the probability of an exchange between two neighbouring replicas,

$$\beta_1 = 1/k_bT1 \text{ and } \beta_2 = 1/k_bT2$$

where k_b is Boltzmann's constant, $T1$ and $T2$ are the temperatures and $U1$ and $U2$ the potential energies of replicas 1 and 2, respectively. The REMD algorithm cycles the replicas through the entire range of temperatures used in the simulation set and effectively sorts low energy structures to low temperatures where they may be maintained and refined, and sends unfavorable, higher energy conformations to a warmer environment, where they may rapidly move through conformational space (Figure 2.4).

2.6 Simulated Annealing (SA)

Simulated Annealing (SA) is a conformational search method known to be particularly effective for the exploration of conformational space of peptides **[88]**. The SA method was first described in 1983 **[89]**, based on the similarity that exists between locating the global minimum of the potential energy function of a molecule, and the slow cooling required to obtain a perfect crystal. In fact, crystal growing will probably be perfect if the system is cooled very slowly by reaching the thermodynamic equilibrium when passing through restrained regions of the phase space. Application of this concept to the exploration of the conformational space can be translated in terms of starting the simulation at a sufficiently high temperature and subsequently decreasing it gradually until the system is frozen in the global minimum. All the studies carried out using the simulated annealing method have demonstrated that although the cooling scheme is not sufficiently slow to find global minimum, it is capable to find local minima of the regions explored. This means that SA combined with a searching strategy, which permits to cross different potential energy barriers and to reach the low energy regions, is a very efficient method to explore the conformational space. It is a heuristic global

optimization algorithm, which has successfully been applied to solve many difficult optimization problems **[90]**. In SA, a cost function takes the role of the free energy in physical annealing and a control parameter corresponds to the temperature. To use SA in conformational analysis the cost function would be the internal energy. At a given temperature the system is allowed to reach thermal equilibrium using MD or MC simulation. At higher temperatures, the system is able to occupy higher energy regions of the conformational space and to surmount higher energy barriers. As the temperature falls, the lower energy states become more attainable in accordance with the Boltzmann distribution. At absolute zero the system should occupy the lowest energy state (i.e. the global minimum energy conformation) **[90]**.

Figure 2.5 depicts the protocol generally used in the SA process where, the molecule is heated and then cooled very slowly so that conformational changes taking place will lead to a local minimum being located. This process is repeated many times until several very closely related, low energy conformations are obtained.

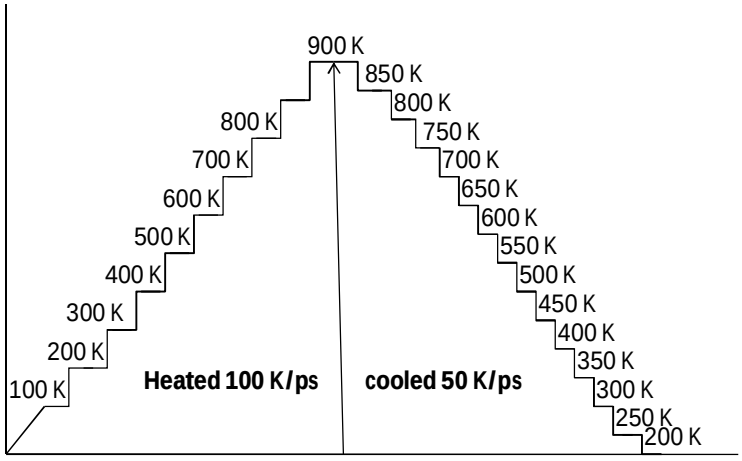


Figure 2.5 Schematic diagram of the iterative simulated annealing protocol.

2.7 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 **[91]**. 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 **[91]** implemented in AMBER 9 **[92]**. 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 **[91]** 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.

2.8 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 **[68]**.

2.8.1 Steepest Descent Method

The steepest descent moves directly down the steepest slope on the potential energy surface, with limited changes to the molecular structure. However, it is useful for quickly correcting bad starting geometries or for removing bad contacts, and it 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 **[68, 93-95]**.

2.8.2 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 **[94],** 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 **[94, 78]**.

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 **[68, 78]**. 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 **[78]**.

2.9 EXPERIMENTAL TECHNIQUES

2.9.1 Nuclear magnetic resonance (NMR)

Nuclear Magnetic Resonance (NMR) is an experimental technique that makes use of the magnetic properties of the nuclei of atoms to predict the structure of molecules. NMR is also a powerful technique used for the study of dynamics and thermodynamics of biological macromolecules, because of a symbiotic relationship between MD simulations and NMR. The use of nuclear magnetic relaxation experiments is of prime importance in the study of proteins and protein folding and observables from these experiments include the nuclear overhauser effects (NOEs) and transverse and longitudinal relaxation times. 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:

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

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

with,

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

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 **[95, 96]**. The cross-relaxation term σ_{ij}, i.e. the relaxation of nucleus i due to nucleus j, in terms of $j(\omega)$ is **[97]**:

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

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 **[96]**. 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 **[98]**:

$$I_{ij} \propto \sigma_{ij} \qquad (2.6)$$

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.

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

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

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

As was originally shown by Tropp **[97]**, 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 **[99]**, 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 $r\frac{noe}{ij}$ 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 **[100]**, and in a later paper applied it to the refinement of tendamistat, a protein which specifically inhibits mammalian alpha-amylases **[101]**. Since this original paper **[100]** 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 **[102-104]** as well as some older studies based on short simulations in vacuo **[105-106]**.

2.9.2 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 **[107, 108]**. 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 **[109, 110]**. 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) **[107]** has achieved quite some interest from theoretical chemists, e.g. **[111]**. 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.

2.9.3 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 **[112]**.

2.10 Comparison of the results of computer simulations with Experimental techniques

In the pioneering work, myoglobin was determined by X-ray crystallography **[113]** and since then many new structures with increasing complexity have been discovered. Nuclear Magnetic Resonance (NMR) has also contributed significantly to the understanding of protein structure in solution **[114-118]**. More recently, cryomicroscopy has been used to decipher the features of some protein complexes **[119]**. Proteomics, on the other hand has advanced beyond merely making catalogues of newly found proteins, and now focuses also on the comprehensive and detailed characterization of these proteins.

The contribution of these techniques is reflected by the increasing number of available structures in the Protein Data Bank (PDB). However, these experimental techniques have some limitations, including the difficulty to grow crystals, the size of the protein in the case of NMR or the resolution in the case of cryomicroscopy. As a result, little is known of the structure-function relationships of a significant fraction of proteins, including membrane proteins due to their inability to crystallize **[93, 120]**. Prediction of the 3D structure of a protein using first principles is still in its infancy. Understanding the factors regulating peptide folding is one of the key areas of research **[121]**. Although important advances have been achieved in the past regarding the process of protein folding, the available tools still require upgrading to produce enough accurate structures. Areas for improvement include the development

of a more robust force field, and the use of appropriate solvent models and solute charges. With the increase in computer power, different simulation methods such as molecular dynamics (MD), replica exchange molecular dynamics (REMD) and simulated annealing (SA) have contributed significantly in the determination of peptide structures and structures of small proteins. These techniques enable the study of the dynamic features of peptides, which can be used to supplement experimental NMR data **[5]**. Knowing the 3D-structure of a peptide is of great help to understand how it will interact with other molecules. This facilitates again the design of peptides with potential biotechnological or pharmaceutical use. MD simulations are always considered as an important tool for the understanding of the physical basis of the structure and function of biological macromolecules. The very first molecular dynamics study of the macromolecule of biological interest was done on bovine pancreatic trypsin inhibitor (BPTI) because of its small size, high stability and the relatively accurate X-ray structure available in 1975, but with unknown physiological function. The MD trajectory of 9.2 ps for this inhibitor changed the view of proteins as rigid structures **[122]**. During the following 10 years, a wide range of motional phenomena were investigated by MD simulations of proteins and nucleic acids, which were focused on internal motions and experimental interpretation *viz*., analysis of fluorescence depolarization of Trp **[123]**, the role of dynamics in measured NMR parameters **[124]**, inelastic neutron scattering **[125]**, the effect of solvent and temperature on protein structure and dynamics **[126]**, the widely used SA methods for X-ray structure refinement **[127]** and NMR structure determination **[128]**.

MD simulations can provide the ultimate detail concerning individual particle motions as a function of time. Thus, they can be used to address specific questions about the properties of a model system, often more easily than experiments on the actual systems. The applications of these simulations include the measurement of sampling configuration space which involves the utilization of MD often with SA protocols, the description of the system at equilibrium and values of thermodynamic properties and finally, the examination of the actual dynamics which requires the appropriate Boltzmann constant to correctly represent the development of the system over time. Currently with the increasing computer power it is possible to simulate larger systems such as explicit solvent or membrane environment for longer time lengths. The programs such as CHARMM **[129]**, AMBER **[130]** and GROMOS **[131]** are used for simulation studies and have a great range of capabilities, but to solve particular problems for a system, the development of new simulation methodologies are ongoing **[132]**. Experimentally it is not possible to determine whether or not solvent fluctuations have any effect or cause the internal motions of a protein **[133]** particularly at temperatures below the glass transition **[134]** when many proteins cease to be active, but it can be accomplished by simulating the one part of a system for example, the protein at one temperature and other part of system for example the surrounding solvent, at a different temperature. The amplitude of atomic fluctuations in carbonmonomyoglobin were calculated from simulations with temperatures of the protein and solvent at either 300 K or 180 K, i.e. above and below the protein glass transition approximately at 220 K. The results of the four possible combinations revealed that the magnitude of fluctuations in the protein are only weakly dependent

on the protein temperature, and the fluctuations were found to be large when the solvent is at 300 K, independent of protein temperatures. These results confirmed that the temperature of the solvent and thus its mobility is the dominant factor in determining the functionally important protein fluctuations in the temperature range of approximately 80 K for the solvent.

Although MD techniques perform very well in refinement procedures, the use of computational simulations is not always without problems. Therefore, experimental data is always essential to validate simulation methodology, to test the accuracy of the calculated results and to improve the methodology. In general, it is preferable to calculate experimental observables directly from the MD trajectory, and to do this, observables from different experimental techniques *viz*., NMR, CD, X-ray, fluorescence etc., are generated and compared with those obtained from the computational protocols.

CHAPTER 3

COMPUTATIONAL PROCEDURES

This chapter provides a brief discussion on the computational chemistry software tools *viz.,* AMBER and CLASICO computer program used primarily in the simulations and analysis of the trajectories studied.

3.1 The AMBER 9.0 computer program [92]

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 **[92]**. 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 medium sized neuropeptides 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:

3.1.1 Preparatory programs in AMBER [92].

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.

3.1.2 Simulation programs in AMBER [92].

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 **E**nergy **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.

3.1.3 Analysis programs in AMBER [92].

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.

3.2 Conformation Classification (CLASICO) [135]

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 **[136]**. The procedure is based on splitting the Ramachandran plot into different regions as shown in Figure 3.1.

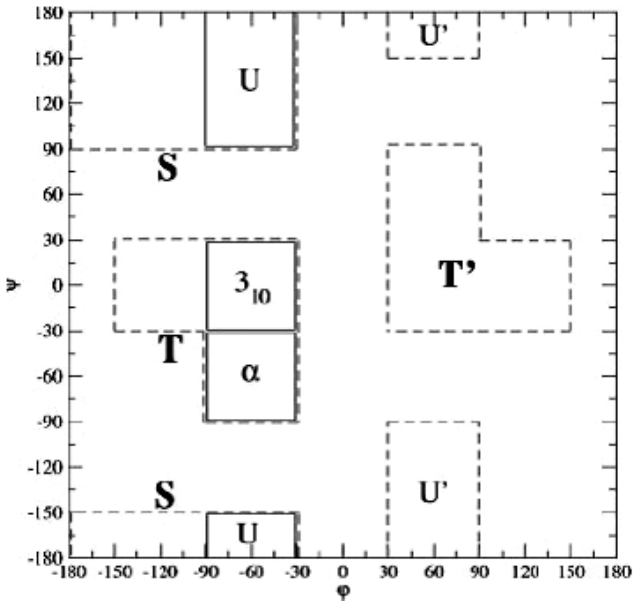


Figure 3.1 Conformational space partitioned into several regions **[136]**. The regions are named after the secondary structure motif that each encompasses (3_{10}: 3_{10}-helix; H: α-helix; S: β-strand; T: type I β-turn, T': type I' β-turn; U: type II β-turn; U': type II' β-turn).

These regions are labeled 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 3.1) are considered to assign conformational motifs, using a set of rules shown in Table 3.1. 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 **[135]**, 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 3.2.

Table 3.1 Conditions for secondary structure definition of three consecutive residues **[135]**, 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	H
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	-

Table 3.2 Definition of β-turns classified on the basis of dihedral angles **[135]**.

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 protocol and the commands used in CLASICO program **[135]** is depicted in Figure 3.2. The dihedral angles obtained from different trajectories are used as input in CLASICO to generate and classify the secondary structure motifs of the peptides.

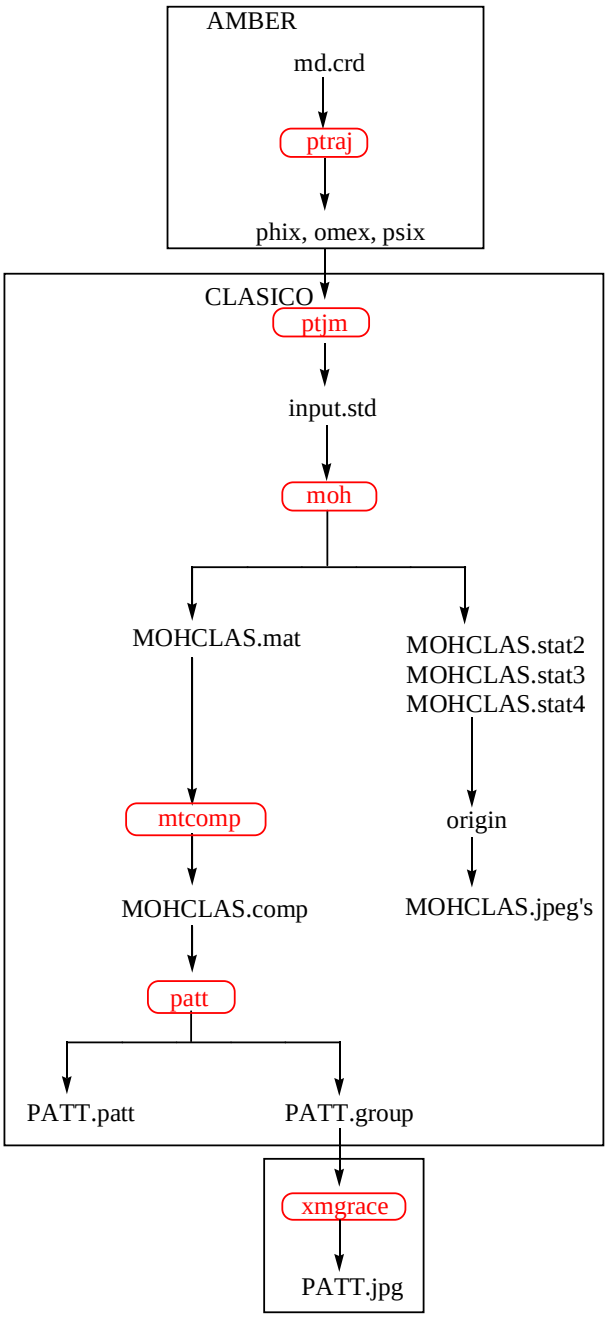


Figure 3.2 Flow chart for the protocol in the CLASICO program **[135]** (commands are shown in red).

3.2.1 Programs in CLASICO [135]

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 4

MOLECULAR DYNAMICS AND REPLICA EXCHANGE MOLECULAR DYNAMICS STUDIES OF BOMBESIN

The work described in this chapter involves the study of the conformational profile of bombesin using different computational procedures. Specifically, the present study describes on the one hand the effect of using the Berendsen's thermostat versus the Langevin's thermostat and on the other hand, the use of the multicanonical replica exchange molecular dynamics as compared with standard molecular dynamics. The results obtained from different computational protocols were finally compared with the previously reported NMR experiments.

4.1 Introduction

The majority of short polypeptide chains, due to their flexible nature, exhibit a complex conformational profile in solution as a result of a dynamical exchange between conformations at the microscopic level **[114]**. These peptides exhibit random coil structures in low viscosity solvents but attain a specific structural feature in the structuring solvents **[137]**. Accordingly, it is expected that experimental techniques reveal average features of the structure as a result of the superimposition of diverse coexisting structures of the ensemble. The conformational features of the polypeptide are dictated by a complex balance of molecular interactions determined by the amino acid sequence and modulated by the environment. In geometrical terms this leads to a rugged potential energy surface with multiple minima, whose characterization through atomistic simulations represents a complementary bottom-up approach to the understanding of the conformational features of a peptide in solution.

The reports indicating exploration of the conformational space of peptides are widely discussed in literature **[37]**. Sampling can be carried out either through a topographical exploration of the potential energy surface using methods like simulated annealing (SA) **[38]**, or in the configurational space using methods like Monte Carlo (MC) or Molecular dynamics (MD) **[2]**. However, due to the nature of the conformational energy surface, sampling engines can be trapped in a local minima. For methods that explore the potential energy surface, a qualitative knowledge of the extent of the space sampled can be assessed by inspection of the density of states characterized during the sampling process **[138]**. In contrast, when sampling the configurational space, convergence is achieved when a Maxwell-Boltzmann weighted ensemble is obtained **[41]**.

The present work is intended to get some insight into the performance of the different computational procedures used to explore the configurational space based on molecular dynamics simulations to provide an adequate atomic description of the system, compatible with the aggregated information provided by different experimental techniques. Specifically, in the present work we report the results of the exploration of the conformational space of a medium size peptide using standard molecular dynamics calculations at 300 K and using two different thermostats, the Berendsen's thermostat **[82]** and second, the Langevin's thermostat **[139]**. The former is widely used in biomolecular simulations because of its stability and efficiency, although in contrast to the latter, it does not produce canonical distributions. In addition, we want also to compare the results obtained from a molecular dynamics trajectory with those produced using the multicanonical replica

exchange molecular dynamics method for sampling purposes using the same thermostat **[79]**.

For the present study the tetradecapeptide of sequence: Glp-Gln-Arg-Leu-Gly-Asn-Gln-Trp-Ala-Val-Gly-His-Leu-Met-NH_2 (Glp= pyroglutamic acid) known as bombesin was considered, due to its pharmacological relevance. Originally isolated from the frog skin of an amphibian Bombina bombina **[140]**, belonging to a family of compounds that exhibit a variety of biological activities in numerous tissues and cell types **[141],** and is widely distributed in the different regions of the brain, lung and gastrointestinal tracts **[142-143]**. Bombesin acts as a neurotransmitter and a neuromodulator on the peripheral system by stimulating the muscles of the alimentary canal **[144-145]** and the secretion of pancreatic enzymes **[146]**. Thus, it triggers the release of some gastrointestinal hormones, enhances the proliferative activity of rat adrenocortical cells and acts as a potent mitogenic agent which displays a growth factor activity for human small-cell lung carcinomas **[147]**. Because of this wide spectrum of biological activities, there is a considerable interest in the clinical potential of both the agonist and antagonist molecules of bombesin, particularly in the fight against cancer **[148-149]**. Also, due to its role in the control of appetite, metabolism, and chronic itching they are interesting targets for drug discovery **[150-151]**. However, in order to develop new drugs a deeper understanding of its structure-activity relationships is necessary.

With regard to the known structure-activity relationships of bombesin, it was established early that the fragment 6-14 was the shortest sequence retaining the full agonist activity, identifying residues Gln^7, Trp^8 and His^{12} to be important for its

biological activity. Spectroscopic studies including NMR **[152-156]** Infrared (IR) **[157]**, Circular Dichroism (CD) and Fluorescence spectroscopy **[158]** and using different solvents like water, dimethylsulfoxide (DMSO) or trifluoroethanol-water mixtures have provided information of the conformational features of bombesin. Specifically, NMR reports in water and DMSO **[152-154]** describe the structure of bombesin as a random coil. In contrast, NMR experiments of bombesin carried out in a trifluoroethanol (TFE)/water mixture (30% v/v) **[155-156]** reports that the C-terminal segment of the peptide ranging from residue 6 to 14 displays a helical conformation, although less sharply structured with residues 11 to 14. Moreover, the first two N-terminal residues adopt an extended conformation, while the region between residues 3 and 5 exhibit a great deal of flexibility. The helical feature of the peptide has also been confirmed by IR **[157]**, CD and Fluorescence studies **[158]** when the peptide is incorporated into lipid environments. In view of these results it has been suggested that the C-terminal region of the molecule displays a helical structure into the hydrophobic lipid environment, whereas the rest of the molecule in the aqueous phase exists in a less ordered structure.

4.2 Computational Methods

In order to compare the outcome of the present simulations with the NMR results of a TFE/water mixture containing bombesin, whose secondary structure inducing capability is probably due to the formation of a solvation coating around the peptide **[156, 159],** implicit solvent models were considered as a good approximation as they allow the peptides to achieve equilibrium faster thus saving computer resources. Specifically, the Generalized Born surface area procedure **[91]** was selected to treat

the solvent in the present work since it offers higher computational efficiency and readily allows for the analytical evaluation of forces **[160-161]**. Consistent with this choice, the AMBER ff96 force field **[162]** was selected for the computation of the energy. This selection is based on the results reported by several authors suggesting that the ff96 set reproduces accurately the results of explicit solvent simulations when used with implicit solvent models **[44,163-164]**. Accordingly, ff66 parameters were developed for the pyroglutamic residue. Specifically, RESP charges were calculated by fitting the electrostatic potential computed at the Hartree-Fock level with a 6-31G(d) basis set following the Merz-Kollman procedure **[166]** and using two minimum energy conformations in the fitting procedure.

In the simulations, the C-terminal of the peptide was amidated while N-terminal was used without any protecting groups, with side chains of Glu and Arg residues considered charged. The salt concentration was adjusted to 0.2 M in order to mimic physiological conditions. An initial extended structure of the peptide was used and energetically minimized with a convergence criterion of 0.005 kcal $mol^{-1}Å^{-1}$. The SHAKE algorithm was used for bonds involving hydrogen atoms and an integration time-step of 2 fs was employed. All the calculations reported in the present work were carried out with the AMBER 9.0 suite of programs **[92]**.

4.2.1 MD simulations using the GB method and Langevin and Berendsen thermostat algorithms

Starting from the extended conformation, two sets of 200 ns MD simulations, using Langevin's and Berendsen's thermostats respectively, were performed at 300 K temperature using the Onufriev, Bashford and Case (OBC) implementation of the

Generalized Born approximation **[91]**. Internal dielectric constant in the peptide was set to 1, while an external dielectric constant of 80, corresponding to water was employed.

4.2.2 REMD Simulations

Starting from the extended conformation of bombesin, 100 ns REMD calculations were performed using the OBC implementations based on the Generalized Born approximation **[91]**. The dielectric constant around the peptide (internal dielectric constant) was set to 1 and the external dielectric constant was set to 80, corresponding to water. Prior to the REMD simulations, five standard MD simulations were run for a period of 5 ns at different temperatures ranging from 250-750 K to obtain the average value of the energy of the system. Time step was set to 0.2 fs, and the SHAKE method was used to constrain all hydrogen atoms. The results of these calculations were used to compute the temperature of the 14 replicas of the present calculation by setting swapping probability of 0.2 and using the condition to have one replica at 300 K. The replica temperatures used in the present calculation were: 280, 300, 321, 344, 369, 395, 424, 454, 487, 523, 561, 603, 648 and 697 K. Replicas were allowed to swap every 2 ps. The temperature during MD simulations was regulated by Langevin's thermostat **[139]**. Analysis of the REMD trajectories was similar to those applied to the MD simulations.

4.2.3 Classification of Structures (CLASICO) [135]

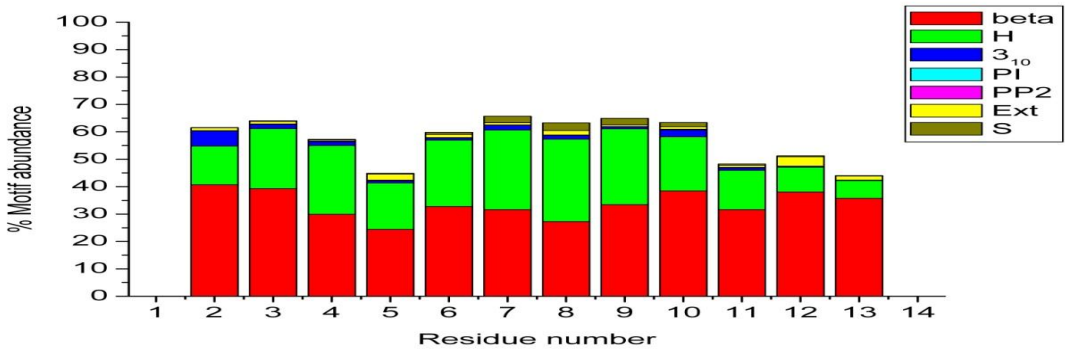
Analysis of all trajectories was carried out using the CLASICO software **[135]** which enables the formation/destruction of secondary structures to be monitored during the folding process, as well as in the characterization of the group of structures that

represents the folded molecule. The characterization of secondary structures for each of the trajectories was carried out using the procedure reported previously **[135]**. This procedure also permits the identification of different conformational patterns attained by the peptide, as well as to compare differences of the conformational space sampled using different computational methods.

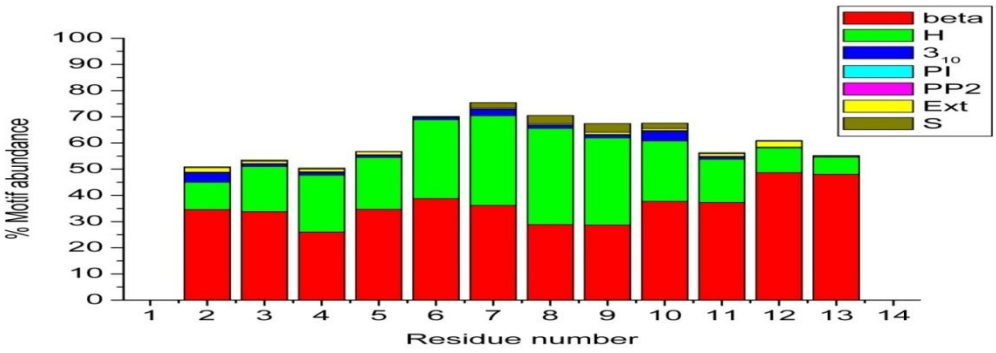
4.3 Results and Discussion

A qualitative analysis of the secondary motifs was performed in order to characterize the structural features of bombesin. For this purpose we used the in-house program CLASICO that permits the identification of secondary motifs from each snapshot, as described elsewhere **[135]**. The program translates each snapshot into a string of letters according to the following procedure: proceeding in a sequential manner, the program computes for each residue its backbone dihedral angles and assigns a letter to it following the Zimmerman partition of the Ramachandran map **[167]**. Following a set of rules, each string is analyzed using a three-letter string (Figure 3.1) to assign the corresponding secondary motif. Histograms of the secondary motifs per residue for each of the three procedures used in the present work are depicted in Figures 4.1a-c. It should be noted that the CLASICO program **[135]** does not include first and last residues of the peptide in secondary structure calculations which is why both the residues are not displaying any of the secondary structure features in Figures 4.1a-c. They show a very structured nature of the peptide: about 60-70% for the MD calculations and about 80% for the REMD calculation. As can be seen, all three calculations predict a high helical content for the whole sequence, with a higher percentage being obtained between residues 6 and 10.

(a)



(b)



(c)

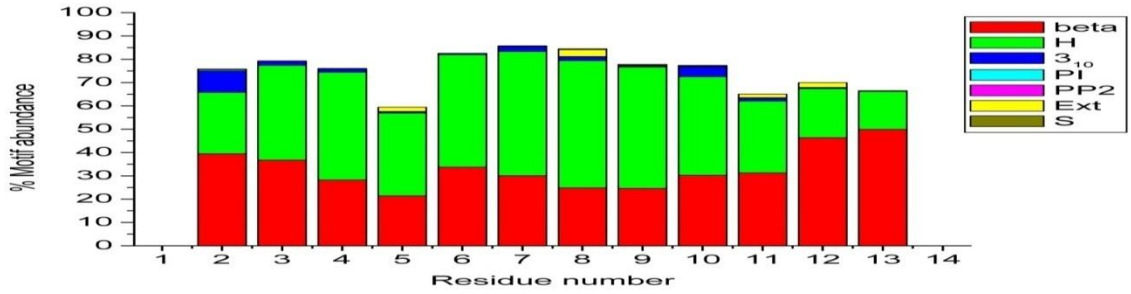


Figure 4.1 Motif abundance for the bombesin in (**a**) MD^{Lang} (**b**) MD^{Beren} and (**c**) REMD trajectories. Conformational motifs are labeled: H (α-helix), 3_{10} (3_{10}-helix), PI (π-helix), PP2 (polyproline II), Ext (extended), S (β-strand), beta (β-turn), as defined in Table 3.1 **[135]**. Only H, beta, 3_{10} and Ext are exhibited by the structures in the current MD studies.

In order to test the performance of the different procedures, the histograms needed to be compared in pairs. The difference in the helix content found between the REMD calculation (Figure 4.1c) and the MD trajectory performed using the Langevin thermostat (Figure 4.1a) must be due to a different sampling time and that the sampling is temperature biased. On the other hand, the MD performed using the Berendsen thermostat (Figure 4.1b) in comparison with the one performed using the Langevin thermostat (Figure 4.1a) shows a differential behaviour with regard to the N-terminus.

In order to gain a better insight into the structural features of the peptide we selected a helical structure from a set of configurations sampled and it was considered as a reference structure. The root-mean-square deviation (rmsd) of each of the snapshots was subsequently computed with regard to the reference structure and the data plotted along the trajectory, as depicted in Figure 4.2a-c. The two MD trajectories show a continuum, not observed in the REMD calculation, since the sampling produced in this method does not correspond to a temporal series. The snapshots were then classified into five sets according to the rmsd values of the reference structure: set 1 rmsd$\in$[0,1]; set 2 rmsd$\in$[1,2]; set 3 rmsd$\in$[2,3]; set 4 rmsd$\in$[3,4]; set 5 rmsd$\in$[4,7].

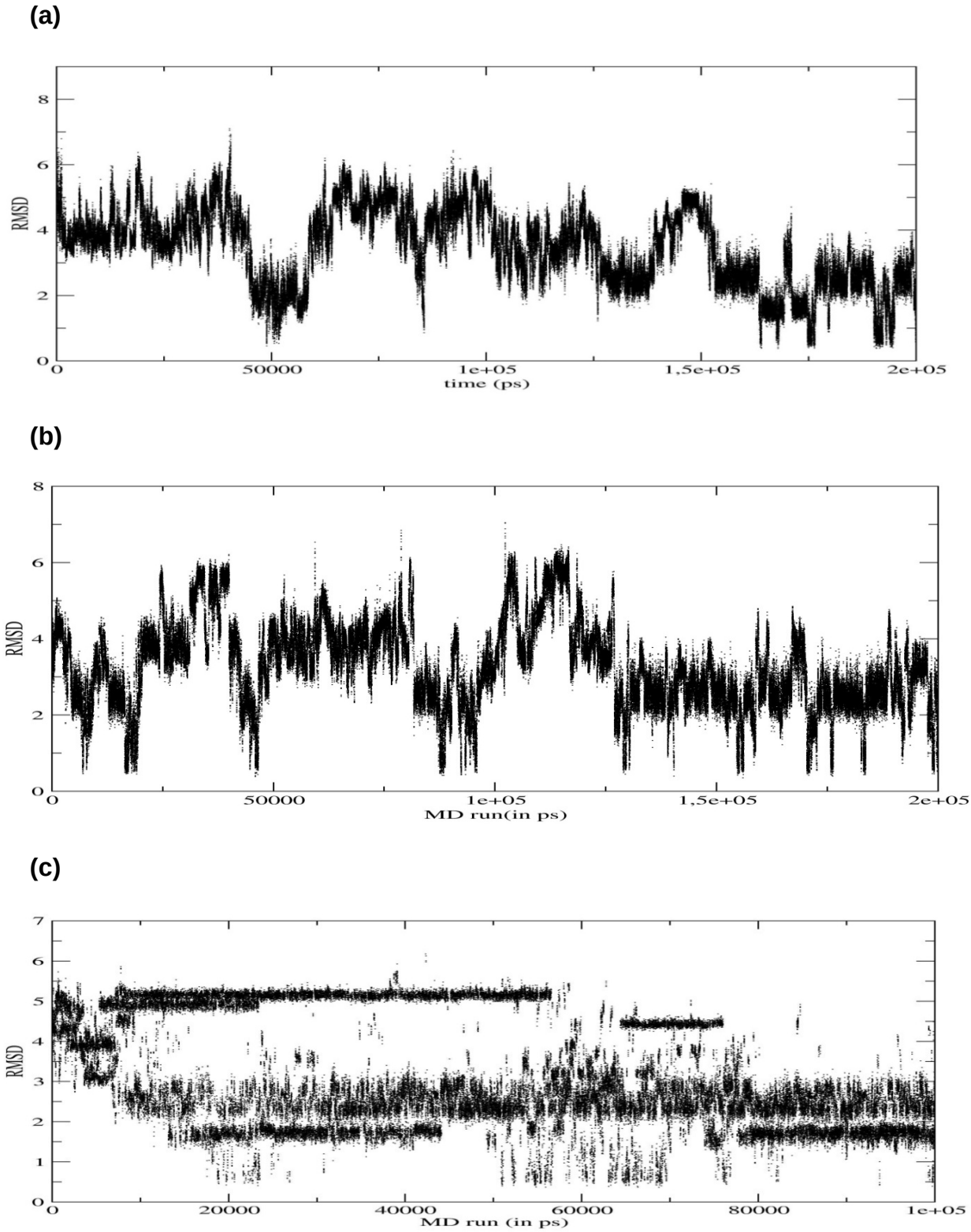
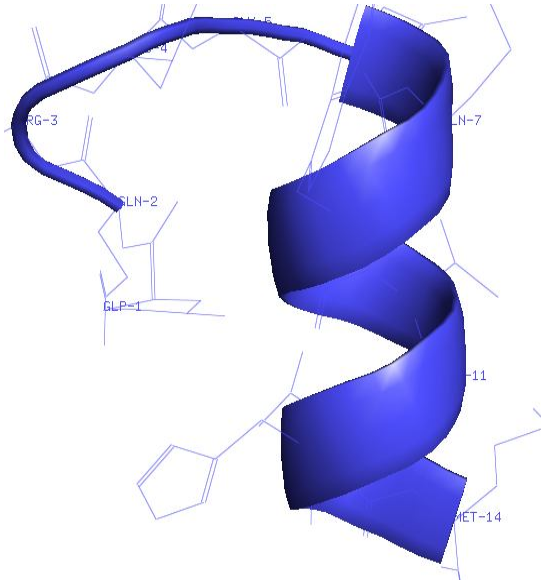


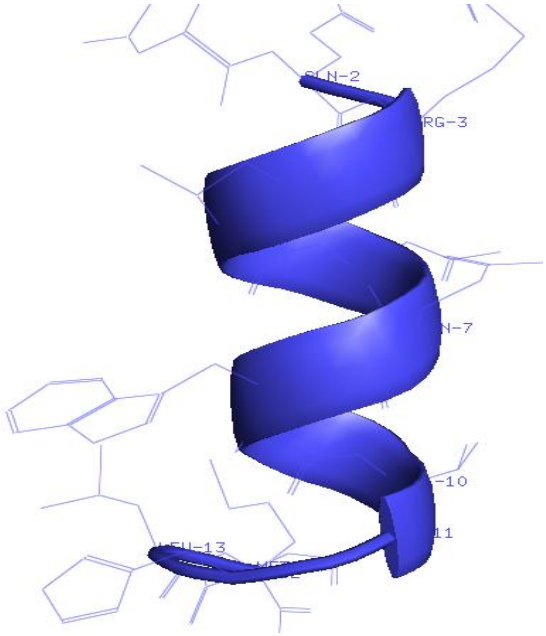
Figure 4.2 Main-chain RMSDs of the backbone atoms from the reference structure for (**a**) MD^{Lang}, (**b**) MD^{Beren} and (**c**) REMD' trajectories.

The percentage of structures in each of the sets are: for the MD (Langevin), 10%, 27%, 61%, 2%, 0%; MD (Berendsen), 8%, 36%, 45%, 2%, 9% and REMD calculation, 21%; 44%; 8%, 2%; 25%; (Table 4.1). The average structures of the different sets are shown pictorially in Figure 4.3. The average structure of set 1 exhibits a well defined α-helical structure on the segment ranging from residue 6 to 14. At the N-terminus on the other hand, the structure exhibits a type I β-turn between residues Glu^2 and Gly^5. Set 2 exhibits also a helical structure between residues 6 to 14, although modified into a π-helix in its C-terminus. This may cause a decrease in the stability that must be compensated by the favorable interaction of the aromatic side chains of Trp^8 and His^{12} that are aligned. Indeed this is an interesting structural feature, since these two residues are known to be important in the activity of the peptide. At the N-terminus, the structure also exhibits a type I β-turn between residues Glu^2 and Gly^5. With regard to set 3, the interaction between the side chains of Glu^2 and Asn^6 disrupts the helical structure at residues 6 and 7, although this is compensated with a type I β-turn between residues Glu^2 and Gly^5. In this structure the helix is distorted at the C-terminus. Set 4 exhibits a helical structure between residues 6 and 14 although the last two turns exhibiting a π-helix. This structure shows again the favorable interaction between residues Trp^8 and His^{12}. Moreover, in its structure, the N-terminus exhibits a bent conformation. Finally, set 5 clearly deviates from the helical structure, although with frequent turn conformations especially around the segment Gly^5- Ala^9 as will be discussed later.

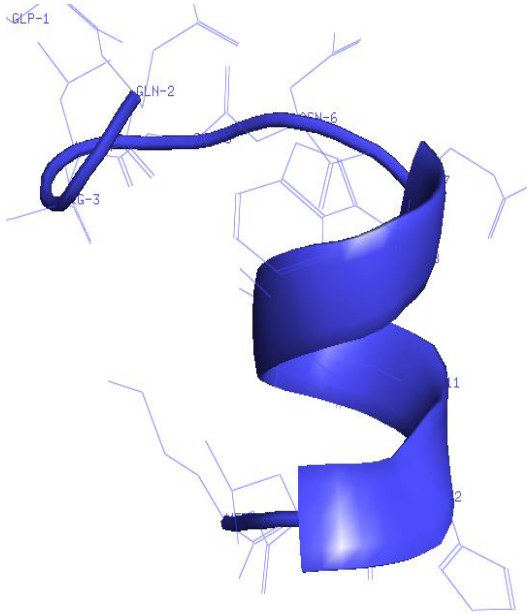
(a)



(b)



(c)



(d) **(e)**

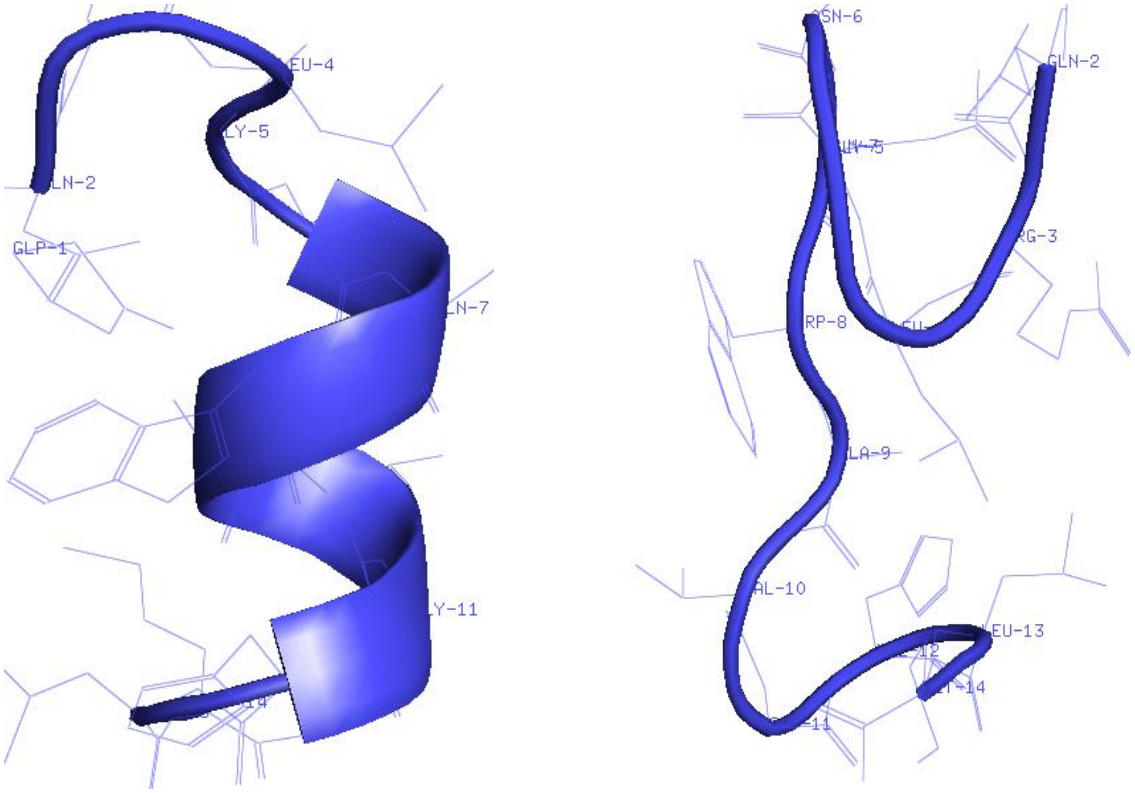


Figure 4.3 Average structures of the different intervals (Figures a-e) classified according to the rmsd w.r.t a helical structure between residues 6 and 10. Different intervals are classified according to the rmsd w.r.t the reference structure: interval 1 rmsd∈[0,1]; interval 2 rmsd∈[1,2]; interval 3 rmsd∈[2,3]; interval 4 rmsd∈[3,4]; interval 5 rmsd∈[4,7].

Table 4.1 Number of structures and their percentages observed in different intervals of the a) MD^{Lang} (b) MD^{Beren} and (c) REMD trajectories. Interval 1, Interval 2, Interval 3, Interval 4 and Interval 5 corresponds to 0-1, 1-2, 2-3, 3-4 and 4-7 regions, respectively of RMSD plots calculated from MD^{Lang}, MD^{Beren} and $REMD^{Lang}$ trajectories with respect to the reference structure.

		MD^{Lang}	**MD^{Beren}**	**$REMD^{Lang}$**
Set 1	Number of structures	20181	16086	20998
	%	10.09	8.04	21.00
Set 2	Number of structures	54023	72784	43837
	%	27.01	36.39	43.84
Set 3	Number of structures	122321	89666	7653
	%	61.16	44.83	7.65
Set 4	Number of structures	3145	3999	1721
	%	1.57	2.00	1.72
Set 5	Number of structures	330	17465	25791
	%	0.17	8.73	25.79

The observed tendencies in Table 4.1 suggest that the secondary structure can be further rationalized through the analysis of the backbone-backbone hydrogen bonding interaction. The statistical analysis of the hydrogen bonds (HBs) observed during the sampling process for the different calculations is shown in Table 4.2. The geometrical criterion used to consider two atoms [donor (A) and acceptor (B)] in the formation of the hydrogen bond (HB) were, to have a distance ≤ 3.0 Å and angle HAB ≤ 120^{o}. In order to obtain significant results, only the hydrogen bond with a percentage of existence ≥ 1.0 % was considered during the simulation time. Moreover, the hydrogen atoms of both terminal $–NH_2$ groups were considered indistinguishable.

Table 4.2 Secondary structures observed due to backbone-backbone hydrogen bond interactions and their percentages in different trajectories for bombesin. Secondary structures, α-helix, π-helix, γ-turn, reverse turn and loop, are assigned in terms of hydrogen bond interactions between i to i+4, i to i+5, i to i+2, i+n (n >2) to i and i to i+n (n >3) residues, respectively in the peptide **[14-16]**.

No.	Definition	2º Structure	MD^{Lang} (%)	MD^{Beren} (%)	$REMD^{Lang}$ (%)
1	(Gln^{2})CO…NH(Asn^{6})	α–helical	8.9	8.8	22.0
2	(Arg^{3})CO…NH(Gln^{7})	α–helical	7.3	12.4	21.7
3	(Leu^{4})CO…NH(Trp^{8})	α–helical	11.9	17.4	29.2
4	(Gly^{5})CO…NH(Ala^{9})	α–helical	16.4	16.7	39.3
5	(Asn^{6})CO…NH(Val^{10})	α–helical	10.5	11.0	24.4
6	(Gln^{7})CO…NH(Gly^{11})	α–helical	12.3	13.5	24.3
7	(Trp^{8})CO…NH(His^{12})	α–helical	15.0	19.6	32.0
8	(Ala^{9})CO…NH(Leu^{13})	α–helical	8.2	9.2	18.0
9	(Val^{10})CO…NH(Met^{14})	α–helical	6.0	6.1	16.0
10	(Gln^{7})CO…NH(Leu^{4})	Reverse turn	7.5	5.8	_
11	(Trp^{8})CO…NH(Leu^{13})	π–helical	5.4	7.1	9.1
12	(Ala^{9})CO…NH(Met^{14})	π-helical	9.2	12.7	16.3
13	(Gln^{2})CO…NH(Ala^{9})	Loop	5.6	_	_
14	(Gly^{11})CO…NH(Arg^{3})	Loop	_	_	7.7
15	(Ala^{9})CO…NH(Leu^{4})	Loop	_	_	5.4
16	(Asn^{6})CO…NH(Trp^{8})	γ-turn	_	_	6.1
17	(Val^{10})CO…NH(His^{12})	γ-turn	_	_	5.27

Figure 4.4a is a diagrammatic representation of the permanence of the difference in hydrogen bonding patterns along the dynamics for the trajectory performed using the Langevin thermostat. HBs responsible for the helicity are almost negligible up to 40 ns whereby the helical region between residues 5 to 10 can be observed during the next 20 ns. The region between 105-200 ns of trajectory, excluding the segment between 165-178 ns, shows conformations with helical structure between residues 6

to 14. A hydrogen bond formed between residues 2 to 9 corresponding to a loop conformation was also observed in some of the structures sampled during the initial 30 ns of the trajectory.

Table 4.2 lists the type of interactions along with the percentage of conformations in the currently investigated simulations, calculated in terms of backbone-backbone hydrogen bonding. Although most of the conformations sampled in each of the trajectories correspond to α-helical (residues 2 to 14) regions, REMD has proven to be better at sampling the helical conformations (Table 4.2). These results demonstrate that the formation of two β-turns between residues 6 to 8 and 10 to 12 along with two loops between residues 3 to 11 and 4 to 9 have been sampled in $REMD^{Lang}$, whereas they were absent in MD^{Lang} and MD^{Beren} trajectories. On the contrary, no reverse turns were observed in REMD whilst they were found in the MD^{Lang} and MD^{Beren} simulations between 4 to 7 residues. The corresponding HB diagram for the trajectory performed using the Berendsen thermostat is shown in Figure 4.4b. In this case, helical conformations are sampled from the beginning of the trajectory. The segments of trajectory showing regular distribution of hydrogen bonds and accounting for helical structures between residues 2 and 14 can be distinguished in the intervals 1000-5000, 11000-18000, 82000-84000 and 123000-200000 ps.

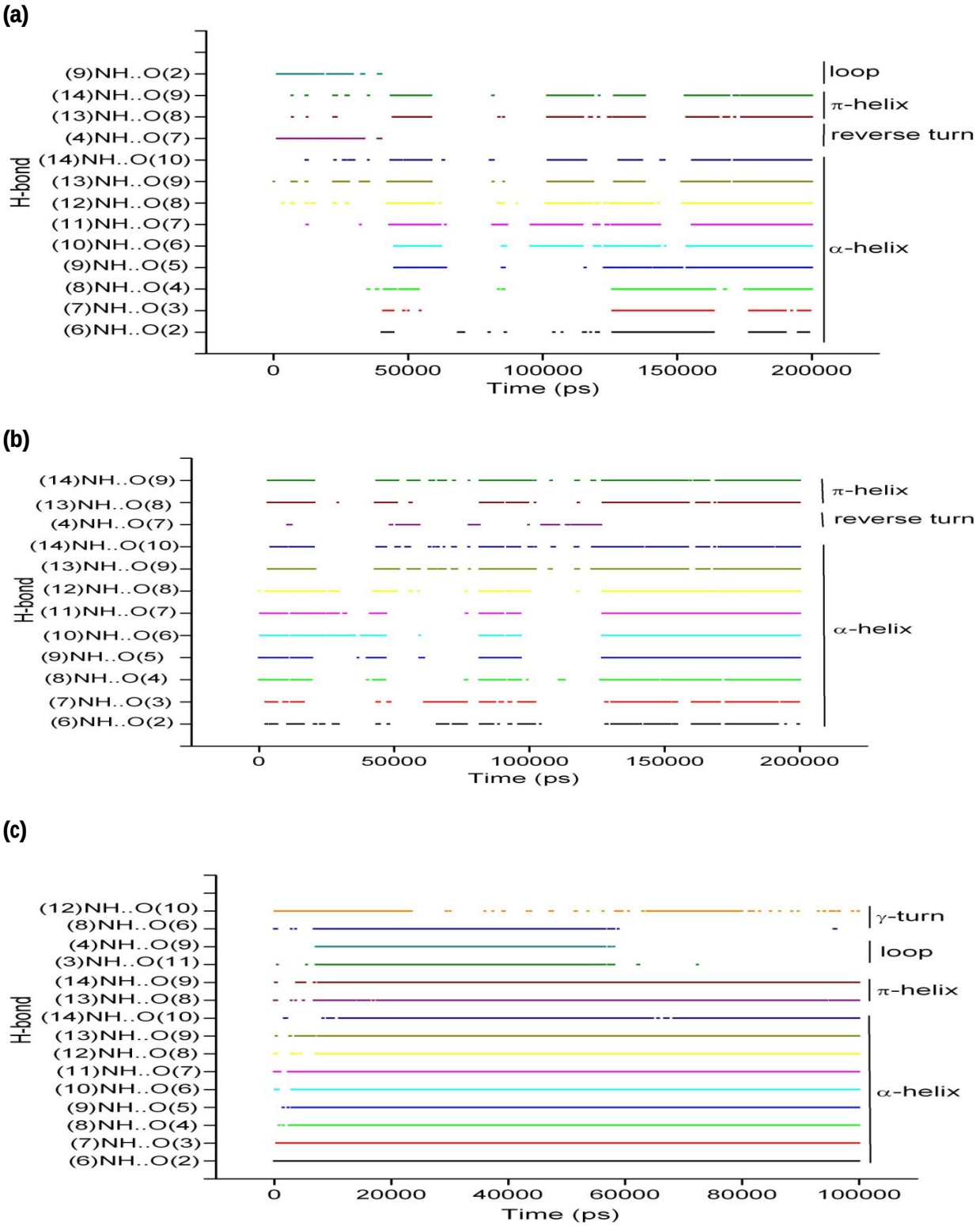


Figure 4.4 Progress of hydrogen bonds (Table 4.2) monitored between important residues for bombesin in (**a**) MD^{Lang} (**b**) MD^{Beren} and (**c**) REMD trajectories. Secondary structures, α-helix, π-helix, γ-turn, reverse turn and loop, are assigned in terms of hydrogen bond interactions between i to i+4, i to i+5, i to i+2, i to i+2, i+n (n >2) to i and i to i+n (n >3) residues, respectively in the peptide **[14-16]**.

To some extent, hydrogen bonding between residues 8 to 13 and 9 to 14 was responsible for the π-helicity which was also sampled mainly in the intervals 5000-15000, 130000-158000 and 170000-200000 ps.

Figure 4.4c shows the hydrogen bonding pattern in the sampling process using the REMD method. Inspection of Figure 4.4c reveals the extensive appearance of hydrogen bonds between residues 2 to 14. The existence of two π-helical turns between residues 8 to 13 and residues 9 to 14 can be observed in most of the conformations. Two consistent hydrogen bonds between residues 6 to 8 and residues 10 to 12 corresponding to a β-turn were also observed in most of the configurations sampled. Overall, these results suggest that the peptide has a high propensity to adopt helical conformations flanked by residues 2-14, under REMD conditions.

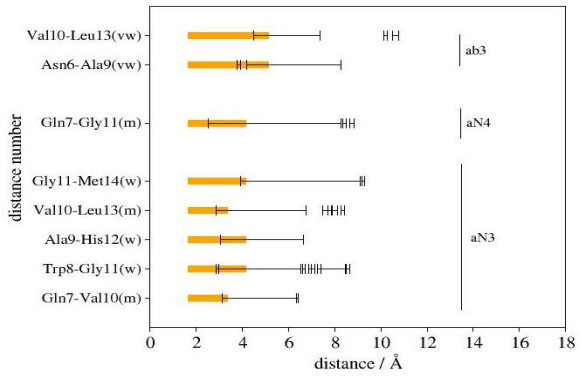
Finally, it is a good exercise to compare the results with the aggregated information provided from the NMR spectrum. Since the NOE intensities of an NMR spectrum are inversely related to the sixth power of the distance between the atoms involved, the distances computed from the NMR spectrum are directly comparable with the results of the simulations. Eight long NOE distances obtained from the NMR spectrum recorded in TFE/water, were used to compare the present results **[156]**. These signals can be classified according to its intensity as strong (s), medium (m), weak (w) and very weak (vw) **[168]**. The intensity of the signal provides a range for the corresponding atom-atom distances **[169]**. The NOEs considered in the present study include: i) C^{α}-N (aN3, Figures 4.5-4.7) between Gln^{7}-Val^{10} (m), Trp^{8}-Gly^{11} (w), Ala^{9}-His^{12} (w), Val^{10}-Leu^{13} (m) and Gly^{11}-Met^{14} (w); ii) a C^{α}-N (aN4, Figures 4.5-4.7) between Gln^{7}-Gly^{11} (m) and iii) C^{α}- C^{β} (ab3, Figures 4.5-4.7) Asn^{6}-Ala^{9} (vw) and Val^{10}-

Leu^{13} (vw) **[156]**. The distances between the corresponding atoms responsible for the different experimental NOEs were computed independently from the different calculations performed in the present study. NOE distances are reported as the average of the distance computed for each snapshot with a tolerance factor of ±1.96 standard deviations, covering a 95% of the variance assuming a normal distribution. Figures 4.5-4.7 pictorially shows the overlapping between experimental NMR and computed distances for the MD^{Lang}, MD^{Beren} and REMD calculations, respectively. On comparing the different boxes of one calculation, all three figures show the same pattern, with a full overlap between NMR results and computations for the structures of intervals 1-3 and partial overlap for the structures of interval 4, whereas for the structures of interval 5, the overlap is negligible. In the case of the structures of sets 1-3 there is partial agreement between the computed distances and the NOE derived distances. From Figure 4.3 it can be seen that these structures exhibit in common a helical structure along the segment 6-14 that correlates well with the observed distances derived from the NMR spectrum. However, there is a considerable difference in the dispersion of the distances sampled during the simulations. Specifically, Figures 4.7a and 4.7b, in comparison show an increase in the dispersion of distances Gly^{11}-Met^{14} and Gln^{7}-Gly^{11}. Moreover, this effect is increased dramatically in all distances when Figures 4.7b and 4.7c are compared. In fact, inspection of the average structures shown in Figure 4.3a, 4.3b and 4.3c suggests deviations from the standard α-helix at the C-terminus. Specifically, deviations may be in the form of a π-helix as shown at the last turn of the structure of Figure 4.3b. This observation of a π-helix at the C-terminus is also corroborated by the

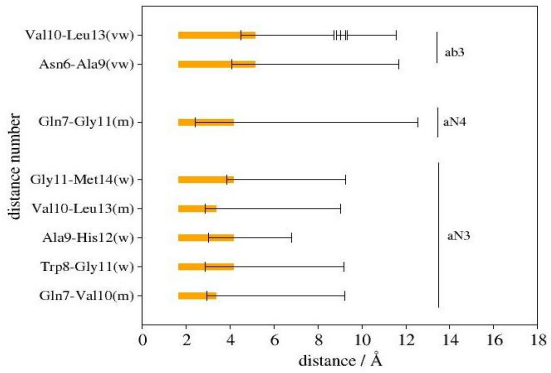
appearance of a HB between $Trp^{8}CO…NHLeu^{13}$ and $Ala^{9}CO…NHMet^{14}$, shown in the HB analysis listed in Table 4.2. In contrast, structures belonging to interval 4 do not agree with the NMR derived distances Val^{10}-Leu^{13} and Gly^{11}-Met^{14}. As can be seen in Figure 4.3d these structures do not show helicity at the C-terminus.

Dispersions of the distance for the structures classified in intervals 3, 4 and 5 are similar in the three calculations. In contrast, the distance dispersion in the two first intervals (Figures 4.5a, 4.6a and 4.7a, and 4.5b, 4.6b and 4.7b) is larger for the Langevin simulation, whereas for the REMD and Berendsen, they are similar. Thus, comparison of the results obtained with the two MD simulations, suggests that the thermostats play a role in the sampling process. Inspection of the rmsd from a 6-14 helical structure (Figure 4.2) suggests that the peptide follows more frequently the folded conformation using Berendsen's thermostat compared with Langevin's thermostat. This is due to an observed biasing effect of the former to enhance folded structures **[170]**. On comparing REMD calculations with MD using the Langevin's thermostat, it can be suggested that a better agreement of the NMR results for the former, is due to the known extremely rapid process of equilibration observed in the REMD calculations **[171]** that is reflected in the number of configurations sampled in intervals 1 and 2 as compared with the MD simulations.

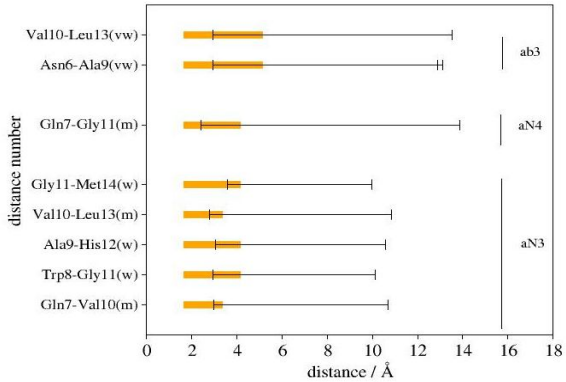
Interval 1



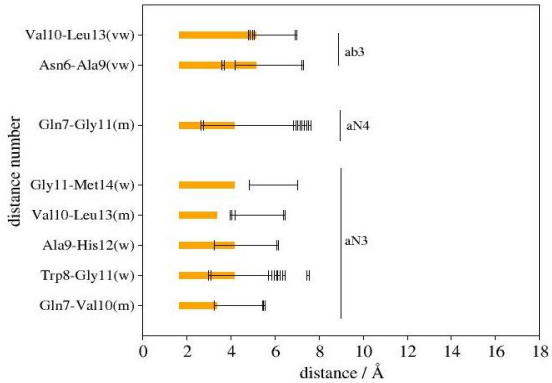
Interval 2



Interval 3



Interval 4



Interval 5

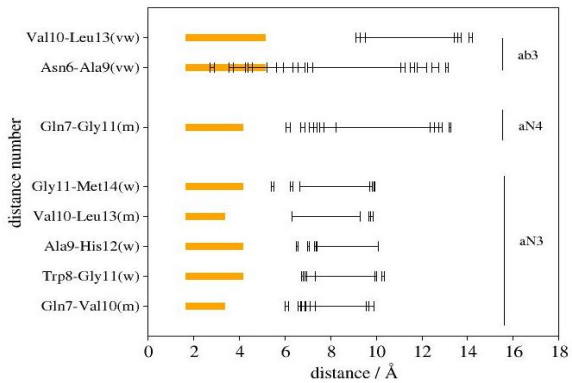
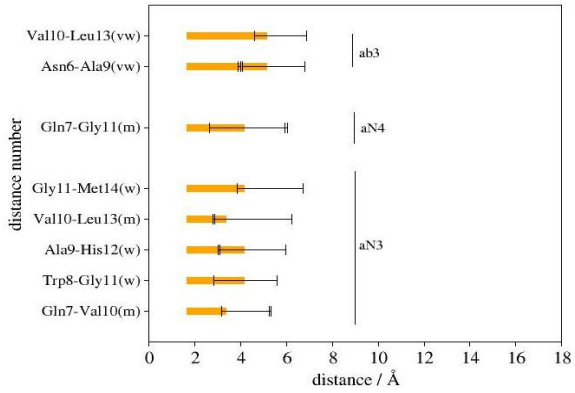
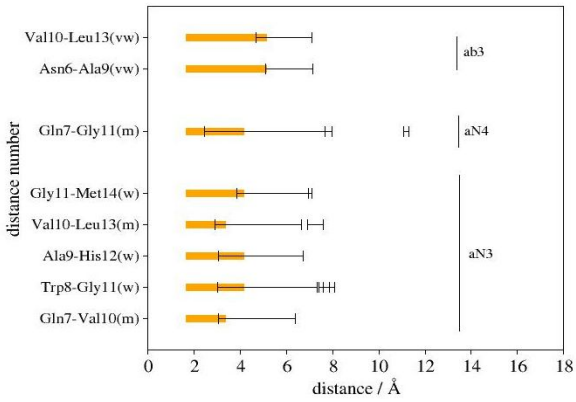


Figure 4.5 Comparison of reported NMR distances **[156]** shown in orange and average of the distance intervals (1-5) containing 95% structures during the MD^{Lang} sampling process. Different intervals are classified according to the rmsd w.r.t the reference structure: interval 1 rmsd∈[0,1]; interval 2 rmsd∈[1,2]; interval 3 rmsd∈[2,3]; interval 4 rmsd∈[3,4]; interval 5 rmsd∈[4,7].

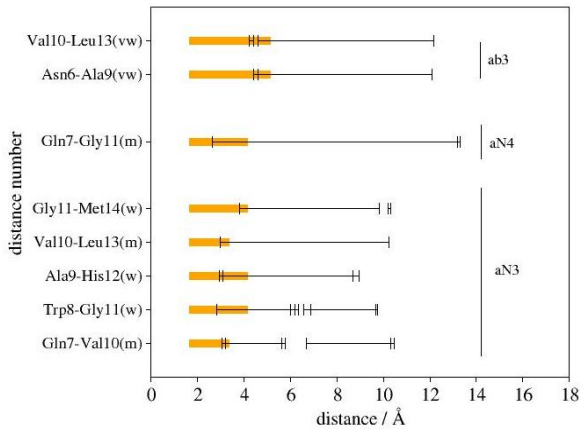
Interval 1



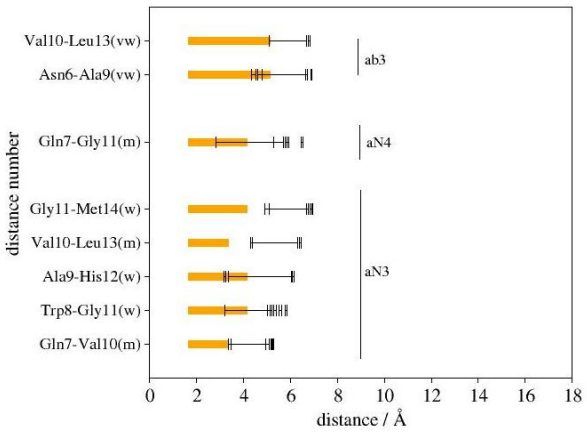
Interval 2



Interval 3



Interval 4



Interval 5

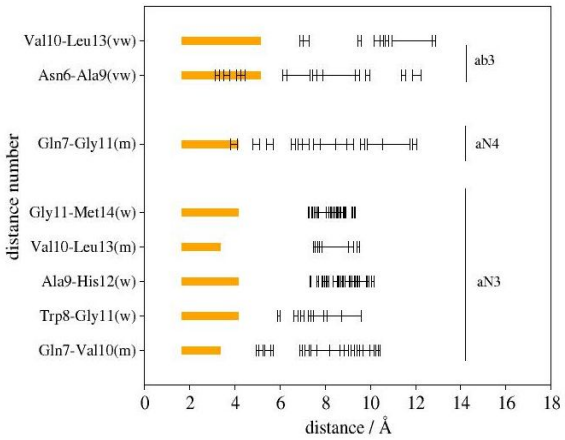
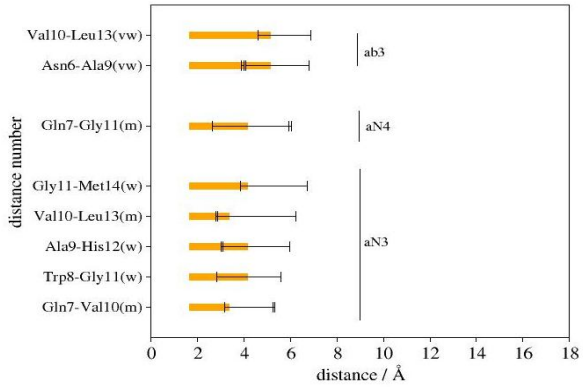
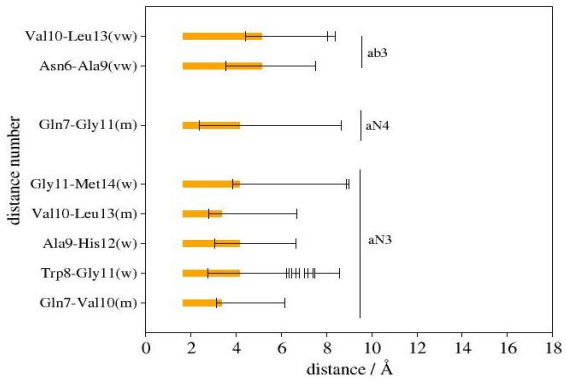


Figure 4.6 Comparison of reported NMR distances **[156]** shown in orange and the average of the distance intervals (1-5) containing 95% structures during the MD^{Beren} sampling process. Different intervals are classified according to the rmsd w.r.t the reference structure: interval 1 rmsd∈[0,1]; interval 2 rmsd∈[1,2]; interval 3 rmsd∈[2,3]; interval 4 rmsd∈[3,4]; interval 5 rmsd∈[4,7].

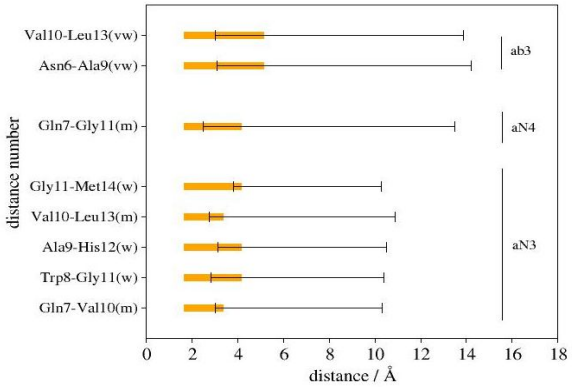
Interval 1



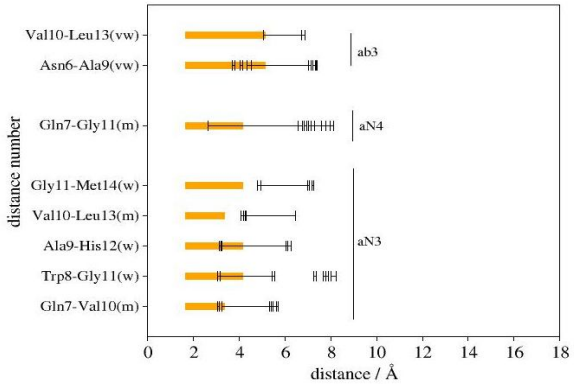
Interval 2



Interval 3



Interval 4



Interval 5

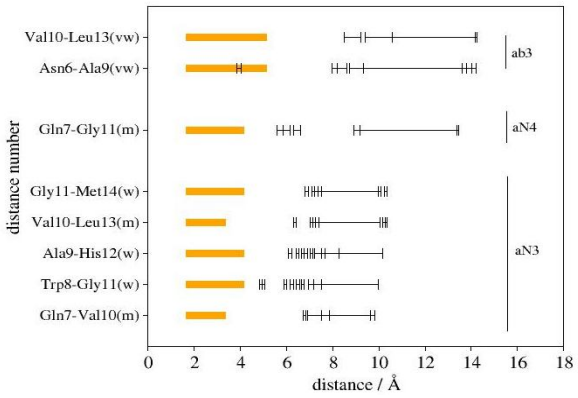


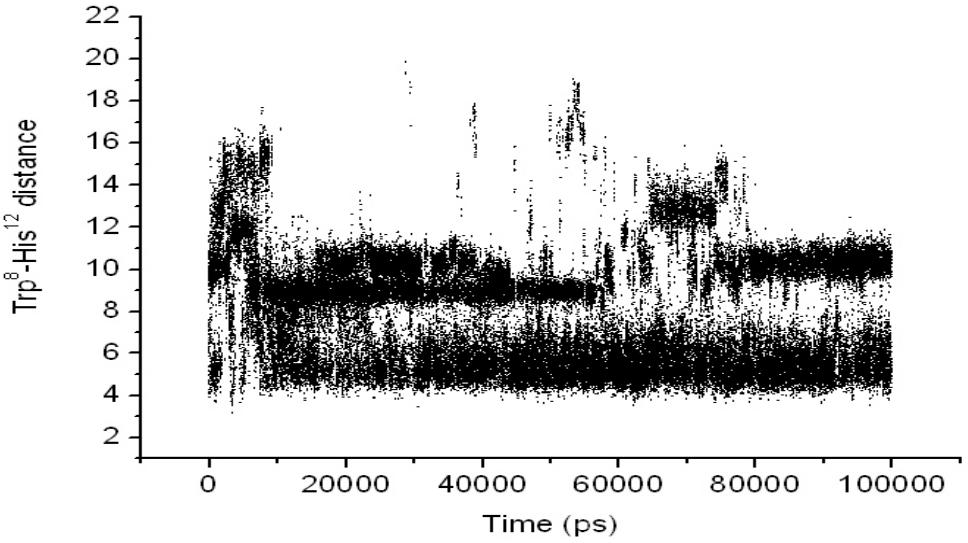
Figure 4.7 Comparison of reported NMR distances **[156]** shown in orange and the average of the distance intervals (1-5) during the REMD sampling process. Different intervals are classified according to the rmsd w.r.t the reference structure: interval 1 rmsd∈[0,1]; interval 2 rmsd∈[1,2]; interval 3 rmsd∈[2,3]; interval 4 rmsd∈[3,4]; interval 5 rmsd∈[4,7].

Structure-activity ideas, combining the known dependence of the size and sequence of the peptide on its biological activity, analyzed in view of the structural results reported in the present work, is discussed. Our results clearly show that the peptide attains a helical structure on the segment 6-14 regardless of the methodologies used. Moreover, the helical structure exhibits a tendency to unwind at the C-terminus showing a propensity to adopt a π-helix at its last turn. This behaviour is related to the presence of two glycines at strategic positions 5 and 11, known as a helix breaker residue. Gly^{5} acts as a hinge between the N-terminus and the rest of the chain, whereas Gly^{11} perturbs the helix at the C-terminus. These structural features explain the known biological data. The fragment 6-14 of bombesin is known to be the shortest fragment retaining activity. This can be correlated with the clear tendency of the peptide to adopt a helical structure in this segment. Indeed, analogs containing residues with a helical inducer like α-aminoisobutyric acid (Aib), exhibit enhanced affinity **[172]**. Moreover, it is known that residues Gln^{7}, Trp^{8} and His^{12} are important for bombesin activity. These residues may play a role in stabilizing the helical secondary structure and/or in the recognition process with the receptor through their side chains. Analysis of the trajectories reported in the present work reveal that the distance between Trp^{8} and His^{12} side chains moves about 50% between 4 Å and 9 Å. When the two side chains get closer they form a π- π stacking interaction (see Figure 4.8a), providing a structural support for the helical structure and possibly plays a role in the recognition of the receptor. When the two side chains are parted at the longest distance, the helix gets distorted at the C-terminus. The analysis of the calculations reported in the present work also shows that Gln^{7} and Trp^{8} side chains are involved in

a NH…π hydrogen bond during about 40% of the trajectories (Figure 4.8b). This interaction reinforces the structure of a helical turn between residues Gly^5 and Ala^9 stabilized by the corresponding carbonyl-amide backbone hydrogen bond (see Table 4.2). In this situation the helix is distorted, preventing the interaction between the two aromatic side chains Trp^8 and His^{12}. Accordingly, the dynamical picture emerging is that Trp^8 plays the role of a hinge, reinforcing the helical structure on the one hand through π- π interactions, and stabilizing the Gly^5-Ala^9 helical-turn on the other. The reinforcing role on the helix turn of Gln^7 can be further justified taking into account that when substituted by histidine as in neuromedin C, the peptide exhibits no significant loss of biological activity to the BB2 receptor, but two orders of magnitude loss to the neuromedin B receptor (NMBR or BB1 receptor). This could be explained if we consider that His is a poorer helical inducer than Gln and consequently, the peptide may loose part of its intrinsic tendency to form a helix, but alternatively, it can also form a hydrogen bond with the Trp side chain as Gln does. This dual behaviour could explain the dual pharmacological behaviour of the peptide with regard to receptors BB1 and gastrin releasing peptide receptor (GRPR or BB2 receptor). However, further investigation needs to be done in this direction.

The analysis of the calculations reported in the present work also shows that Gln^7 and Trp^8 side chains are involved in a NH…π hydrogen bond during about 40% of the trajectories (Figure 4.8b). This interaction reinforces the structure of a helical turn between residues Gly^5 and Ala^9 stabilized by the corresponding carbonyl-amide backbone hydrogen bond (see Table 4.2). In this situation the helix is distorted, preventing the interaction between the two aromatic side chains Trp^8 and His^{12}.

(a)



(b)

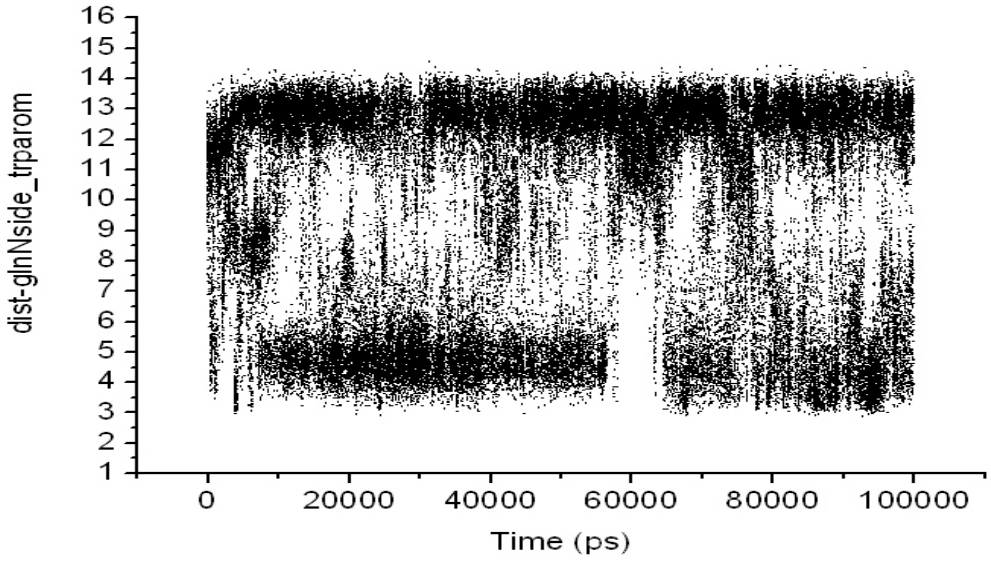


Figure 4.8 (a) Distance (obtained using ptraj module of AMBER 9 **[92]**) between the centers of the aromatic rings of Trp8 and His12 during the REMD sampling process; **(b)** Values of the distance between the amide group of the Gln7 side chain and the center of the aromatic ring of Trp8 during the RMED sampling process.

4.4 Conclusions

Two 200 ns MD trajectories (using Langevin and Berendsen thermostats, respectively) and a 100 ns REMD sampling were used to explore the conformational space of bombesin using the OBC implementation of the GB method to describe the solvent. Calculations carried out using the AMBER ff96 force field, were considered adequate to use in combination with implicit solvent methods **[44, 163-165]**. Analysis of the results clearly shows that the peptide attains a helical structure on the segment 6-14 regardless of the methodologies used. This result is clearly connected to the activity of the peptide, since it is known that the bombesin fragment 6-14 is the shortest fragment retaining activity. Moreover, the results also show the tendency of the peptide to adopt a π-helix at the C-terminus aligning the aromatic residues Trp^{8} and His^{12}. The side chain of Gln^{7} is found to form a NH-π hydrogen bond with the side chain of Trp^{8}, distorting the helix at its C-terminus, suggesting that Trp^{8} plays the role of a hinge winding and unwinding of the helical C-terminus of the peptide. With regard to the methodological aspects of the present results, although the three methodologies provide the same general conclusion, the REMD is more robust for the purpose of sampling, providing less uncertainty as measured through the agreement of the NMR results as compared with MD simulations **[172]**. MD^{Beren} shows a bias in the sampling of the folded structures in comparison to the MD^{Lang}. Moreover, taking into account that the REMD calculations were done using the Langevin thermostat, direct comparison to MD^{Lang} demonstrates the quick equilibration process being undertaken when using the REMD procedure.

CHAPTER 5

CONFORMATIONAL PROFILE STUDY OF NEUROMEDIN C

This chapter focuses on a methodological study used to assess the folding characteristics of neuromedin C (NMC). Replica exchange molecular dynamics (REMD) under implicit ($REMD^{implicit}$) and explicit ($REMD^{explicit}$) solvent conditions was used as a sampling technique. Moreover, two additional molecular dynamics (MD) trajectories performed using Langevin (MD^{Lang}) and Berendsen (MD^{Berend}) algorithms were also suitably competent to sample similar kinds of conformations. The comparison of results obtained from all the trajectories and those derived from the NMR studies of a Ni(II) complex of NMC indicates that the REMD under explicit solvent conditions is more efficient in sampling the conformations, and show good agreement with the experimental results.

5.1 Introduction

Bombesin-like peptides, are mammalian regulatory peptides that exhibit high sequence homology with bombesin and known to be widely distributed in the mammalian gastrointestinal tract, spinal cord and brain. Gastrin-releasing peptide (GRP) or its C-terminus decapeptide (GRP18-27): neuromedin C (NMC) that retains full action **[173]** and neuromedin B (NMB) are the peptides most well characterized. They were originally isolated from porcine spinal cord and canine intestinal mucosa, respectively **[173-174]** but it is now well established that they are widely distributed in the gastrointestinal tract, spinal cord and brain **[175]**. These peptides have been shown to elicit various physiological effects, including inhibition of feeding, smooth muscle contraction, exocrine and endocrine secretions, thermoregulation, blood pressure and sucrose regulations and cell growth **[176]**. Specifically, NMC acts as a

growth factor in a wide range of tumors including carcinomas of the pancreas, stomach, breast, prostate, and colon **[149]**.

Recently, NMR spectroscopy and MD simulations **[177]** in aqueous solution have been employed to determine the structure of the 27-residue GRP. Only a short helical structure was observed in the N-terminal region and the overall structure of the GRP peptide did not show any ordered secondary structure, similar to the observations made for other peptide hormones in aqueous solution **[178]**. Bombesin, on the other hand, has been reported to adopt helical conformations between residues 4 to 10 on the basis of NMR experiments in the presence of 50% trifluroethanol-water mixture **[155-156, 179]**. Additionally, circular dichroism (CD) studies of NMC, equivalent to the last 18-27 residues of GRP in the water-lipid environment suggest that the peptide in a polar medium exists as a very flexible structure, and to some extent α-helical in a non-polar medium **[180]**. Moreover, two dimensional (2D)-NMR experiments have demonstrated that NMC adopts two well defined turns after forming a complex with Ni(II) using its initial three N-terminus residues **[169]**. The first turn corresponds to the Ni(II) coordination ligands in a square planar conformation, and the second reveal the interaction between fourth residue (Tryptophan) and eighth residue (Histidine). Several experimental studies reveal that the peptides reproducing Ni(II)-protein coordination sites dramatically change their conformation upon metal ion complexation **[181]**. NMC, being shorter in size, is quite flexible to take on several different conformations making assessment of its bioactive conformation quite challenging. Therefore, despite having great biological and pathological significance

[181], the unique native conformation of NMC has not to date been clearly elucidated on the basis of available spectroscopic results.

While experimental knowledge is essential for the understanding of effects of counterions on the structure and dynamical properties of proteins in solutions, theoretical studies involving computer simulations complement the experimental data. In particular, MD simulations provide detailed information on the fluctuations and conformational changes of proteins and nucleic acids, and are routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes **[182]**. However, the sampling efficiency of MD simulations is severely hampered by the rugged energy surface of polypeptides, with small relative free energies separating native, folded protein conformations from unfolded states. In order to allow the MD simulation to escape the multi entrapment minima and explore a wider portion of the conformational space, enhanced sampling techniques have been developed. One of the most popular ones, Replica Exchange Molecular Dynamics (REMD), is a multicanonical sampling procedure based on running multiple parallel MD simulations or replicas of a system at increasing temperatures T1, T2, etc, allowing exchange of configurations between neighbouring replicas according to a certain probability of swapping **[88, 183]**. Thus, conformations are simulated at all temperatures and escape local minima with the kinetic energy provided at higher temperatures, while Boltzmann distributions are generated at all temperatures.

Recently, it was demonstrated that MD simulations employing the implicit continuous model of Generalized Born (GB) approximation in combination with AMBERff96 show faster and better efficiency to sample the conformational space **[184]**. However, the

disadvantage associated with these kinds of simulations is their inability to sample those conformations which are stabilized through intermolecular interactions with the solvent molecules, consequently enhancing more compact and bent structures. The present studies involve the conformational space search of NMC both under implicit and explicit water conditions using REMD as a sampling method. The selection of water environment for the current studies has been made on the basis of the expected direct binding of NMC to the protein receptor of the cell, as reported in the literature **[169, 157]**. Two additional MD simulations using (Langevin and Berendsen algorithms) were also performed using the Onufriev, Bashford, and Case (OBC) implicit water model **[91]**. The results obtained from MD simulations were compared with those obtained from REMD simulations. Finally, NMR results were employed to validate the present computational protocols.

5.2 Computational details

5.2.1 Replica Exchange Molecular Dynamics

An extended structure of NMC was generated using the leap module of AMBER 9.0 **[92]**. The N-terminal of peptide was protonated while the C-terminal was amidated. For implicit solvent REMD, the extended structure of the peptide was energetically minimized with a convergence criterion of 0.005 kcal $mol^{-1}Å^{-1}$. REMD was performed on the minimized structure of NMC using Onufriev, Bashford and Case (OBC) implementations based on the Generalized Born approximation **[91]**. The dielectric constant around the peptide (internal dielectric constant) was set to 1 and the external dielectric constant was set to 80, corresponding to water. Prior to the REMD simulation in an implicit solvent, standard MD simulations were performed for 5 ns at

different temperatures (200-900 K range) with a 100 K interval. In the present study, 12 replicas were used and the temperature of each replica was set at: 276, 300, 326, 354, 385, 419, 456, 497, 542, 591, 645, and 705 K. To determine the distribution of the target temperatures for the replicas, the system was equilibrated at twelve different temperatures from 200K to 900 K in 100K increments for 1 ns. From these calculations a polynomial of average energy as a function of the temperature E(T), was obtained and used to compute the temperatures of the replicas using a swapping probability of 0.2 and setting a first replica at 300 K **[87b]**. The time step was set to 0.2 fs, and the SHAKE method was used to constrain all hydrogen atoms. The temperature during MD simulations was regulated by the Langevin thermostat **[139, 185]**. Each replica was simulated simultaneously and independently at different replica temperatures. The replica exchange was performed every 2 ps for 50,000 steps during the REMD simulations.

In the case of a REMD simulation in an explicit solvent, the extended conformation of NMC was solvated in a TIP3P water box consisting of 8894 atoms **[186]**. In order to mimic the physiological conditions a 0.2 M salt concentration was used. The system was then subjected to a minimization with 5000 initial steps of steepest descent followed by 10000 steps of conjugate gradient at constant volume. The energy minimization of this new system was completed when a convergence criteria of 0.001 kcal $mol^{-1}Å^{-1}$ was fulfilled. The minimized system was heated to 300 K at constant volume in 20 ps in 50 K increments. Subsequently, the system was equilibrated for the next 100 ps at a constant pressure using periodic boundary conditions (PBC). Eight standard MD simulations were performed thereafter at constant volume for 5 ns

in the range of temperatures 200-900 K with a gap of 100 K. Simulations were performed with PBC and the particle mesh Ewald (PME) method was used for the treatment of the electrostatic interactions **[81]**. A cutoff was set to 12 Å with a grid spacing of approximately 1 Å. The van der Waals interactions were truncated at 12 Å with a switching function from 10 to 12 Å.

Since the number of replicas needed to perform a REMD is proportional to the square root of the degrees of freedom of the system **[87b]**, for REMD in the explicit water solvent, 44 replicas corresponding to the temperatures: 297.1, 300, 303, 306, 309.1, 312.2, 315.3, 318.5, 321.7, 325, 328.3, 331.7, 335.1, 338.6, 342.1, 345.7, 349.3, 353, 356.8, 360.6, 364.5, 368.4, 372.4, 376.5, 380.6, 384.8, 389.1, 393.4, 397.8, 402.3, 406.9, 411.6, 416.3, 421.1, 426, 431, 436.1, 441.3, 446.6, 452, 457.5, 463.1, 468.8, 474.6 K, were chosen with a condition to have a probability of an exchange of 20% with the mean potential energies obtained from standard MD simulations which are adjusted to a 3^{rd} degree polynomial. REMD simulation was run with a 2 fs time step and SHAKE algorithm used to constrain all bonds to the hydrogen atoms. The replica exchange was performed every 2 ps for 25,000 steps during the REMD run. Coordinates were saved after each picosecond. The PTRAJ module in the AMBER program and the CLASICO program **[135]** were used to carry out the analysis of the results obtained from the simulations. The characterization of a β-turn profile for each of the trajectories was carried out using the procedure described by Corcho and coworkers **[135]**.

5.2.2 Molecular Dynamics

Two MD simulations considering Langevin and Berendsen algorithms were performed at 300 K temperature using the GB approximation. The internal dielectric constant around the peptide was set at 1, while the external dielectric constant of 80 corresponding to water was employed. No cutoff was considered in these investigations. In order to mimic the physiological conditions a 0.2 M salt concentration was used. The SHAKE method was used to constrain the bonds involving hydrogen atoms with a time-step of 2 fs.

5.3 Results and Discussion

Figures 5.1a and 5.1b show the time series of temperature exchanges of NMC for the initial 100 steps of $REMD^{implicit}$ and 50 steps of $REMD^{explicit}$ trajectories, respectively. Clearly, a random walk in the 'temperature space' between low and high temperature was achieved in each case. Figures 5.1c and 5.1d, on other hand, shows the canonical probability distributions of the total potential energy along the $REMD^{implicit}$ and $REMD^{explicit}$ trajectories respectively, and reveal good overlaps between all neighbouring pairs of distributions suggesting good exchange of the potential energies between different replicas during the sampling process **[79]**.

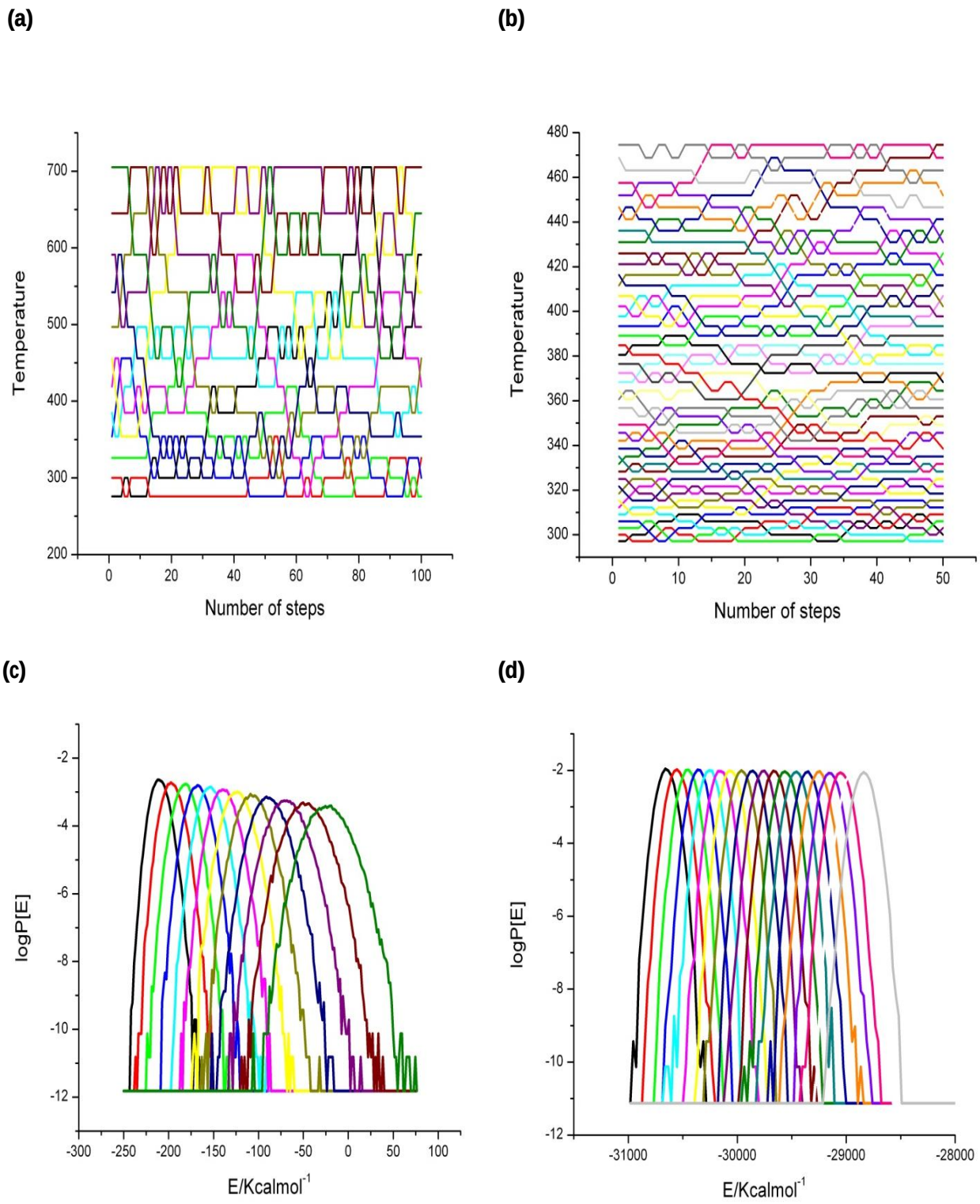
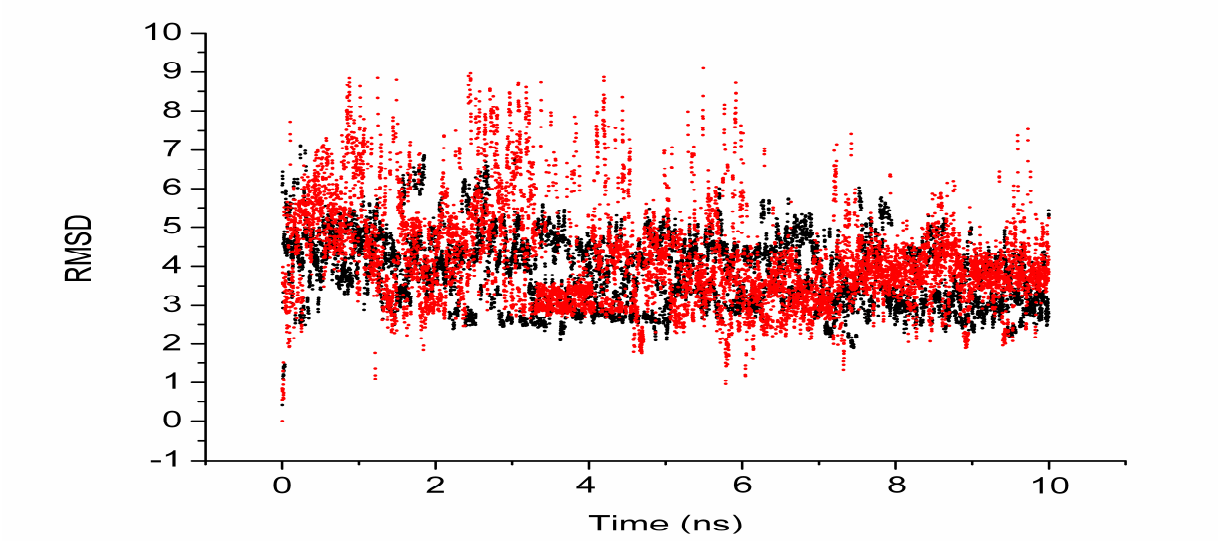


Figure 5.1 Time series of temperature exchange for **(a)** REMD^{implicit} **(b)** REMD^{explicit} and the canonical probability distributions of the total potential energy of NMC obtained from the **(c)** REMD^{implicit} at the twelve temperatures and **(d)** REMD^{explicit} at the forty four temperatures (only eighteen shown). The distributions in **(c)** and **(d)** correspond to the temperature ranges (from left to right): 276-705 K and 297.1-474.6 K, respectively.

In order to mimic the physiological temperature, the REMD trajectories (both for implicit solvent and explicit solvent) obtained only at 300 K were selected for the analysis. Figures 5.2a-b show the root-mean-square deviations (RMSD) of an initial 10 ns of REMD and MD trajectories respectively, monitored relative to the backbone atoms of starting structures. The RMSD values corresponding to the $REMD^{explicit}$ and MD^{Berend} simulations are shown in black whereas those related to $REMD^{implicit}$ and MD^{Lang} trajectories are displayed in red. Figures 5.2a-b clearly shows that RMSD values increase rapidly during the initial 50 ps suggesting initial folding steps of NMC in the trajectories. In case of $REMD^{explicit}$ these values oscillate between 3-5 Å clearly suggesting its structural equilibrium, in contrast to the higher fluctuations (3-9 Å) observed in the $REMD^{implicit}$ trajectory. RMSD comparison of MD trajectories (Figure 5.2b), on other hand, reveals that most of the sampled conformations are quite flexible and oscillate between 1-9 Å throughout the length of these simulations. RMSD values obtained from the MD^{Berend} were found to be slightly higher than the MD^{Lang} in most parts of its trajectory, implying that majority of the conformations (RMSD >1 Å, Figure 5.2b) obtained from MD trajectories are different from the starting structures. To some extent, conformations (RMSD <1 Å) closer to the starting structures were also sampled at certain intervals making their efficiency less effective in comparison to the REMD trajectories. The restricted movement of the peptide in the case of the explicit solvent REMD could probably be attributed to the physical presence of water molecules resulting in less fluctuation (Figure 5.2a, in black).

(a)



(b)

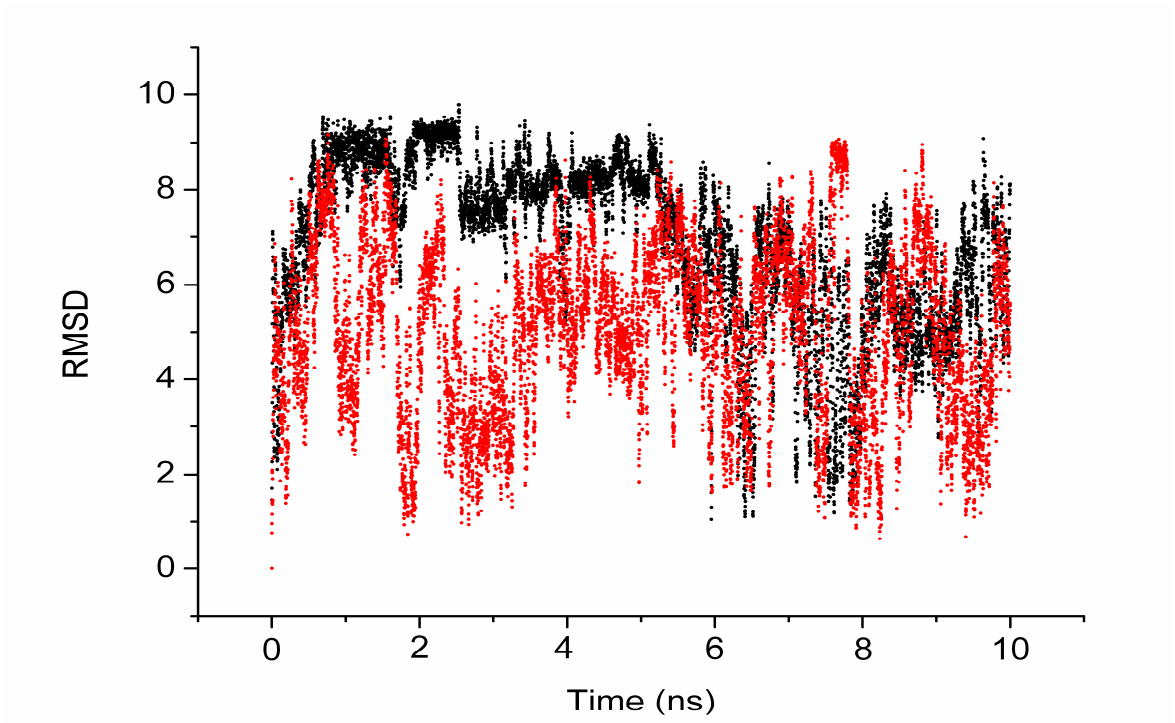


Figure 5.2 Root mean square deviations (backbone-backbone) of NMC from the starting structure in case of **(a)** $REMD^{implicit}$ in red and $REMD^{explicit}$ in black, **(b)** MD^{Lang} in red and MD^{Berend} in black, obtained at 300K.

The sampling efficiency of MD and REMD simulations were monitored for the different patterns of NMC attained during the progress of each trajectory. The patterns basically represent the structures classified on the basis of the type of conformational motifs present **[135]**. For this purpose, pattern profiles were computed for every snapshot in each of the trajectories using the CLASICO program **[135]**, and are depicted in Figure 5.3. After 100000 snapshots of $REMD^{implicit}$ (Figure 5.3a) and 50000 snapshots of $REMD^{explicit}$ (Figure 5.3b) trajectories, 3206 and 1774 new patterns respectively were obtained. Similarly, 3565 and 7981 different patterns were identified for 200000 snapshots of MD^{Lang} (Figure 5.3c) and MD^{Berend} (Figure 5.3d) trajectories respectively. The percent efficiency of these trajectories in terms of generating new patterns were $REMD^{explicit}$ (3.6%), MD^{Berend} (3.4%), $REMD^{explicit}$ (3.2%) and MD^{Lang} (1.8%). A closer inspection of Figure 5.3 reveals that new patterns are generated in a regular fashion throughout the length of the REMD trajectories (Figures 5.3a-b), whereas in case of MD (Figures 5.3c-d), the peptide conformations seems to get trapped (dark areas) in regions of the conformational space at certain intervals suggesting their restrictive nature to explore the new patterns.

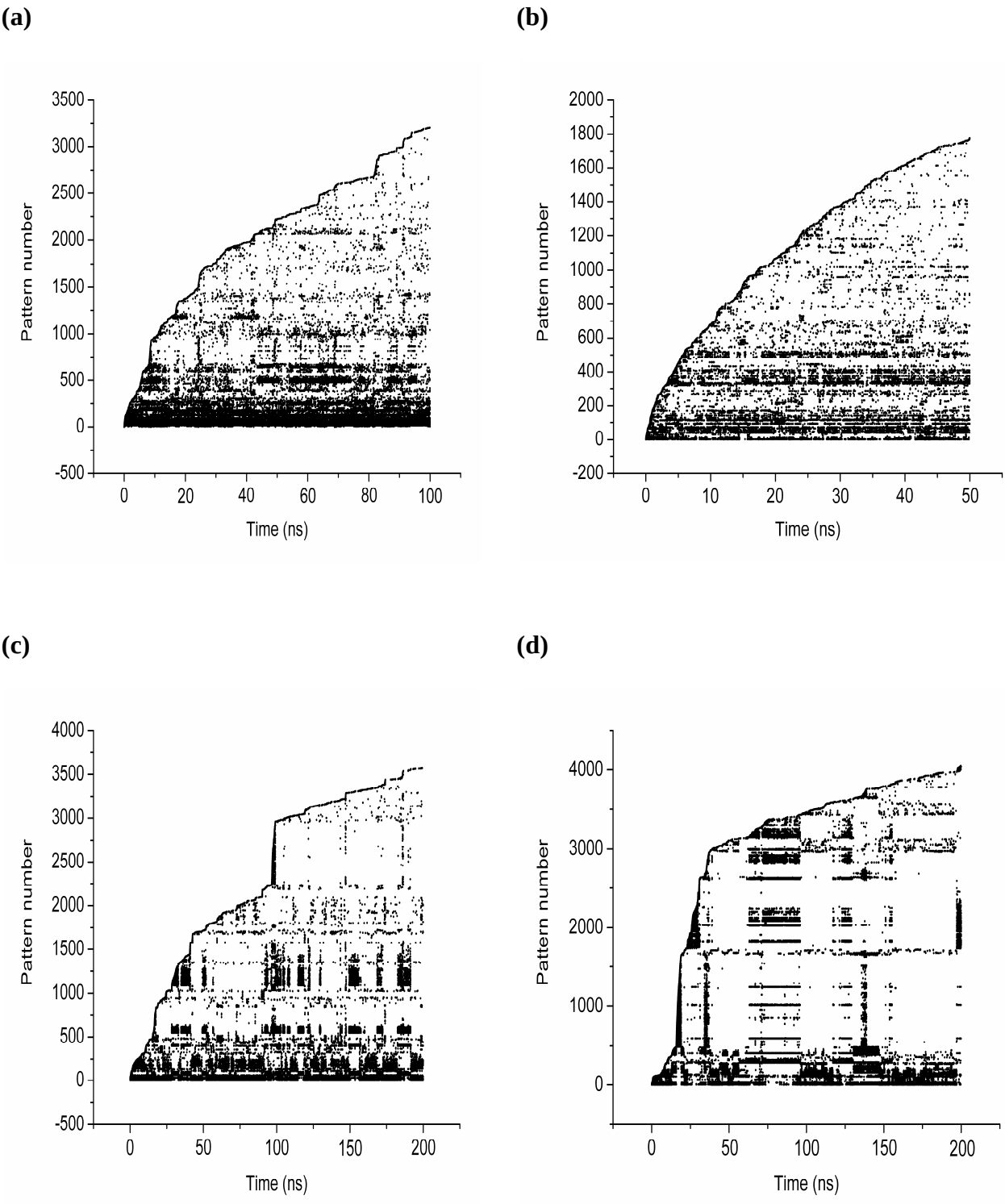


Figure 5.3 Evaluation of new patterns along the trajectories for the NMC in **(a)** $REMD^{implicit}$ **(b)** $REMD^{explicit}$ **(c)** MD^{Lang} and **(d**) MD^{Berend} trajectories obtained using the CLASICO program **[135]**. Dark areas represent trapped conformations of NMC in the configurational space.

The structures obtained from all the MD and REMD trajectories at 300 K were analyzed to determine the conformational motifs attained by the NMC peptide, using the CLASTERIT algorithm of the CLASICO program **[135]**. The statistics of all the motifs found in the peptide for each residue are depicted in Figures 5.4a–d. The secondary structure motifs appear almost from the beginning of each of the trajectories. The conformational motifs obtained in the $REMD^{implicit}$ trajectory (Figure 5.4a), shows the predominance of a β-turn type I between residues 4 and 6 with a lower propensity of type III between residues 6 and 9. To some extent, β-turn type II was also observed between residues 6 and 7 in some of the conformations sampled. Figure 5.4b shows the profile obtained in the $REMD^{explicit}$ trajectory where the peptide exhibits predominantly β-turns type I with higher propensity between residues 4 and 9 and a lower propensity between residues 2 and 3. To some extent, β-turn type i (mirror conformation of β-turn type I) was also found between residues 4 and 5. In the case of MD^{Lang} (Figure 5.4c), the peptide adopts type-I β-turns between residues 4 and 9, with a higher propensity between residues 4 and 7 and lower propensity between residues 7 and 9. Some conformations showing β-turn type III between residues 5 to 9 were also sampled in the MD^{Lang} trajectory. Finally, the structures exhibiting β-turn type I predominantly between residues 3 and 9 along with lower propensity flanked by residues 2, 3 and 7-9 were sampled in MD^{Berend} (Figure 5.4d) trajectory. The percentage order of β-turns adopted by the peptide in the different protocols used is $REMD^{explicit}$ (85%) > $REMD^{implicit}$ (48%) > MD^{Berend} (46%)> MD^{Lang} (35%) while the percentage order of helicity flanked by residues 4 to 10 is MD^{Lang} (12.2%) > MD^{Berend} (13.1%) > $REMD^{implicit}$ (10.1%) > $REMD^{explicit}$ (3.2%).

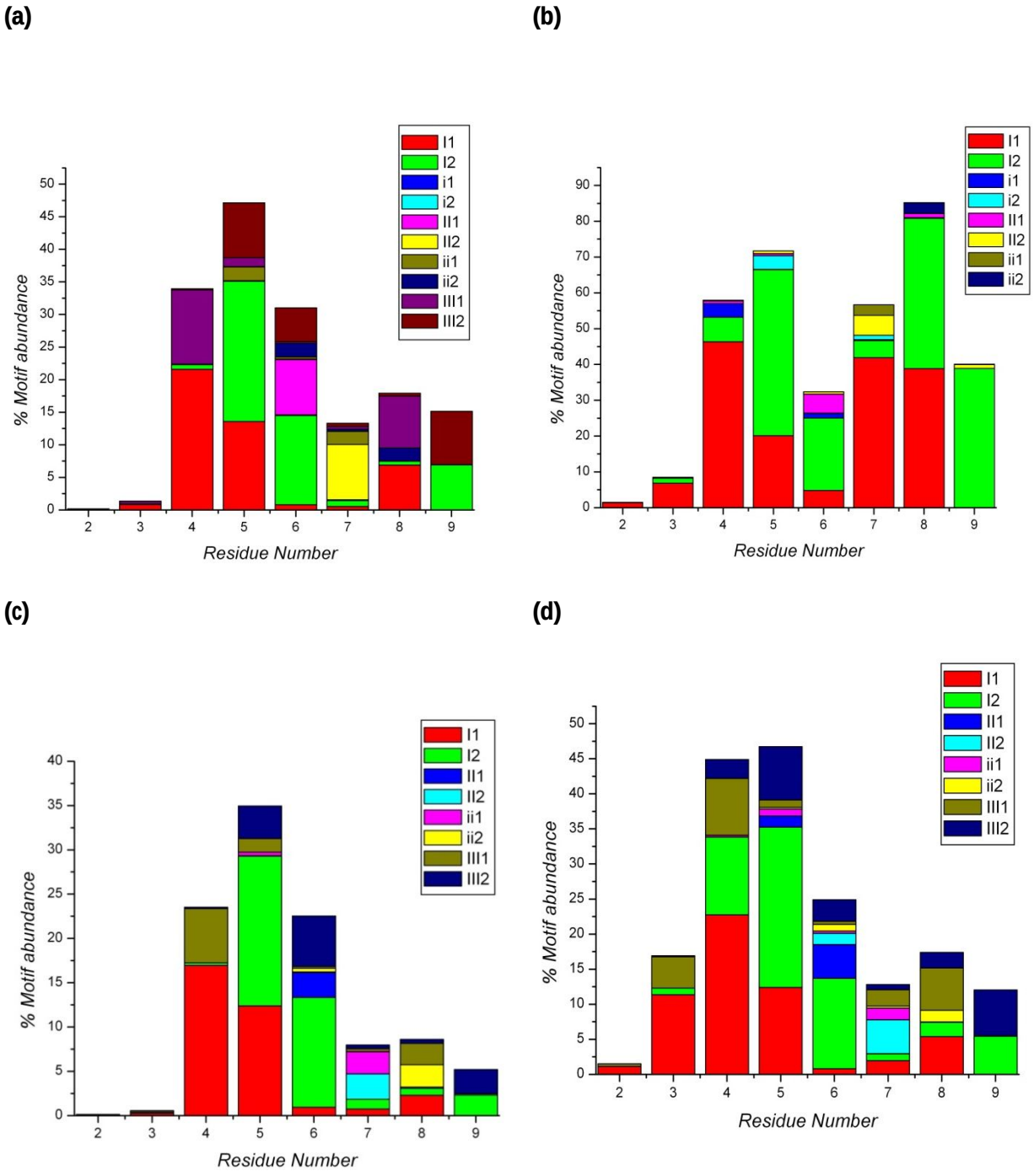


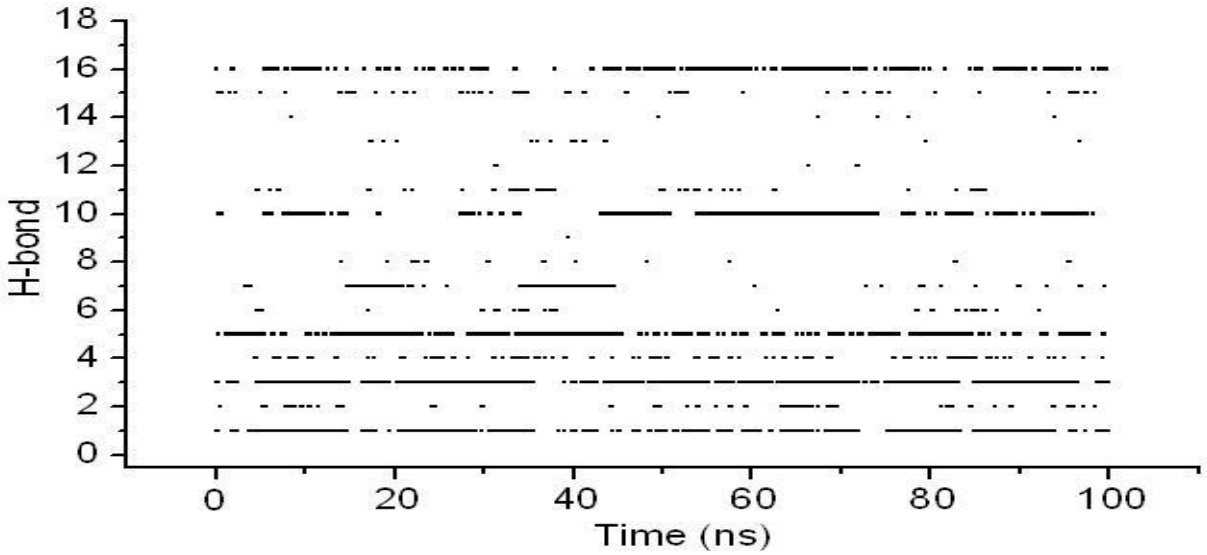
Figure 5.4 Conformational motif abundance attained by NMC peptide in **(a)** REMD^{implicit} **(b)** REMD^{explicit} **(c)** MD^{Lang} and **(d)** MD^{Berend} trajectories. Types of β-turns were defined according to table 3.2 **[135]**.

Overall these results reveal that NMC peptide adopts predominantly β-turns type I conformations, irrespective of the method and environment used for the simulations. The current results also demonstrate that the energy barrier between the folded and helical conformers is small resulting in a population of most of the trajectories with both secondary motifs. The observed conformational behaviour of NMC in the current investigations is consistent with those characterized using experimental results reported in the literature **[169, 180]**.

Hydrogen bonding is known to play an important role in determining the three-dimensional structures adopted by proteins and nucleic bases. In these macromolecules, bonding between parts of the same macromolecule cause it to fold into a specific shape, which helps to determine the molecule's physiological or biochemical role. In view of the significance of H-bonding in peptides and proteins, H-bond analysis has been performed on the conformational profile of NMC obtained from the different trajectories and secondary structures were accordingly characterized. Figures 5.5a-b and Figures 5.6a-b show the appearance of H-bond during the progress of simulation in the $REMD^{implicit}$, $REMD^{explicit}$, MD^{Lang} and MD^{Berend} trajectories, respectively. Table 5.1 illustrates the type of interactions and indicates the existence percentage of the conformations in different simulations; and identifies the type of secondary structures adopted by the conformations of NMC. The first specified percentage value, regardless of where it starts from, of each trajectory in Table 5.1 corresponds to the first H-bond distance displayed in Figures 5.5a-b and 5.6a-b. A closer inspection of Figure 5.5a reveals that NMC, in $REMD^{implicit}$ trajectory,

adopts some helical conformations stabilized by two hydrogen bonds between residues 3 and 7, and 5 and 9 (Table 5.1).

(a)



(b)

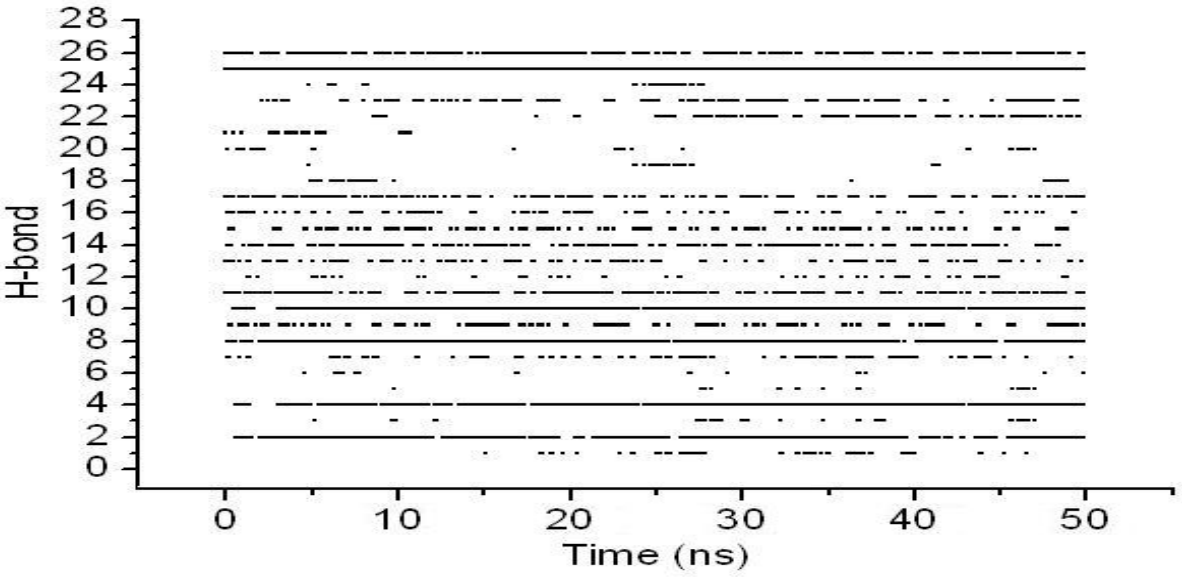
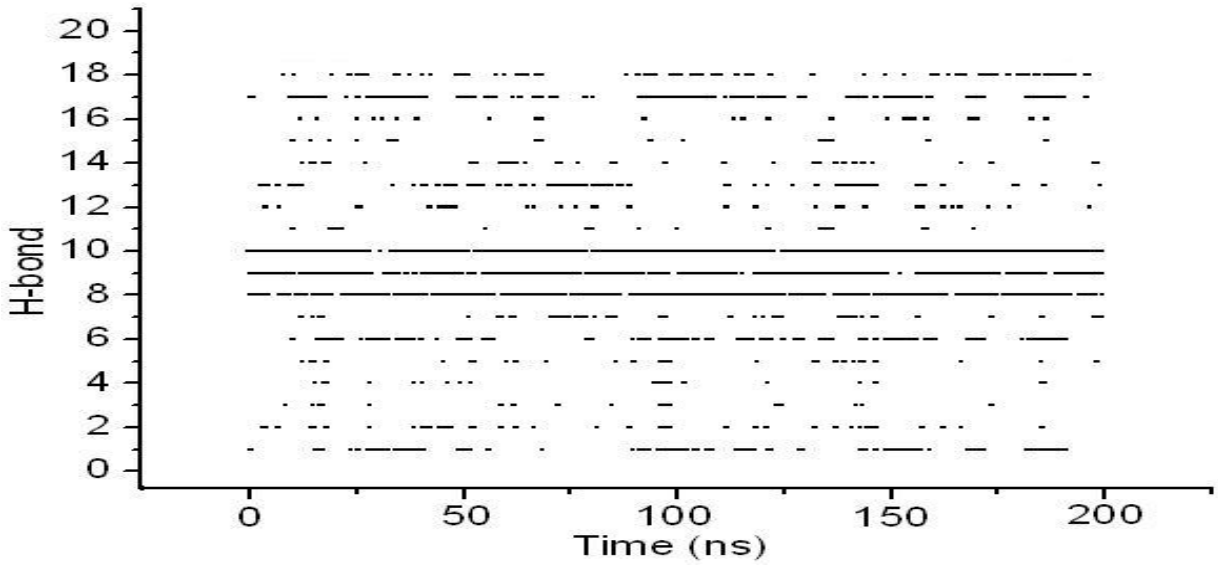


Figure 5.5 Progress of hydrogen bonds (Table 5.1) monitored between important residues for NMC in **(a)** REMD^{implicit} and **(b)** REMD^{explicit} trajectories. Secondary structures, α-helix, π-helix, γ-turn, β-turn, reverse turn and loop, are assigned in terms of hydrogen bond interactions between i to i+4, i to i+5, i to i+2, i to i+3, i+n (n >2) to i and i to i+n (n >3) residues, respectively in the peptide **[14-16]**.

(a)



(b)

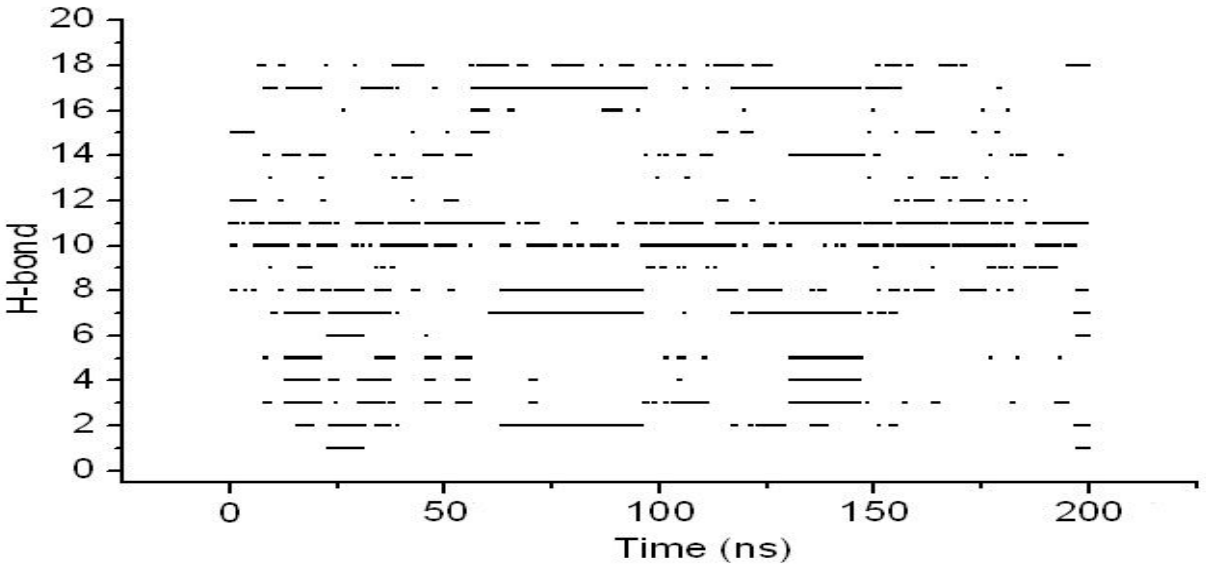


Figure 5.6 Progress of hydrogen bonds (Table 5.1) monitored between important residues for NMC in **(a)** MD^{Lang} and **(b)** MD^{Berend} trajectories. Secondary structures, α-helix, π-helix, γ-turn, β-turn, reverse turn and loop, are assigned in terms of hydrogen bond interactions between i to i+4, i to i+5, i to i+2, i to i+3, i+n (n >2) to i and i to i+n (n >3) residues, respectively in the peptide **[14-16]**.

However, most of the conformations in this trajectory exist preferably in the form of reverse turns and loops (Table 5.1). NMC conformations showing two β-turns between residues 3 and 6, and 4 and 7 were also sampled in this trajectory. $REMD^{explicit}$ (Figure 5.5b, Table 5.1), on other hand, sampled favourably those conformations which have a higher content of β, γ and reverse-turns in their structures and were flanked by residues 2 and 10. To some extent, conformations showing α-helical regions between residues 3 to 7 and 6 to 10 were also sampled in this trajectory. The secondary structures obtained in MD^{Lang} (Figure 5.5a) and MD^{Berend} (Figure 5.5b) trajectories involved α-, π-helices, β–, γ– and reverse turns, and loops (Table 5.1) more or less in equal proportions. The helical conformations sampled in these trajectories show α-helical regions between 3 to 7, 5 to 9 and 6 to 10 residues (Table 5.1). The π-helical region was flanked by residues 4 to 9 and 5 to 10 in some of the conformations in these trajectories (Table 5.1). In addition, some conformations showing side chain-main chain interactions between imidazolic-hydrogens of residue 3 and the backbone nitrogen of residue 5, observed in the rest of the simulations, were also sampled in this trajectory (Table 5.1). Similar kinds of interactions involving the imidazolic hydrogen of residue 8 and backbone nitrogen of residue 10 were observed in the rest of the trajectories excluding $REMD^{implicit}$. The number of hydrogen bonds (>1%) appearing in the investigated protocols follows the order: $REMD^{explicit}$ (26) >MD^{Lang} (18) =MD^{Berend} (18) > $REMD^{implicit}$ (16), clearly suggesting the better efficiency of $REMD^{explicit}$ to sample stabilized conformations.

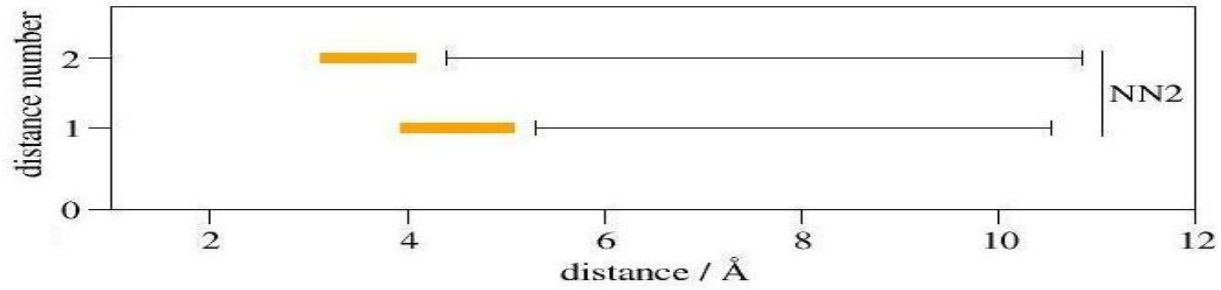
Table 5.1 Secondary structures observed due to backbone-backbone hydrogen bond interactions and their percentages in different trajectories for NMC **[14-16]**.

Donor-acceptor	2^o structure	$Remd^{explicit}$ (%)	$Remd^{implicit}$ (%)	GB^{Lang} (%)	GB^{Berend} (%)
2O...6N	α-helical	1.0	---	---	**---**
3O...7N	α helical	4.0	10.1	10.1	1.5
5O...9N	α -helical	0.8	---	3.0	9.0
6O...10N	α -helical	3.2	---	0.4	6.0
4O...9N	π-helical	---	---	0.5	2.5
5O...10N	π -helical	0.3	6.5	1.1	4.5
10O...4N	β-turn	4.2	---	---	---
10O...4N	β-turn	5.1	---	---	---
2O...5N	β-turn	10.1	---	---	---
3O...6N	β-turn	6.5	5.1		
4O...7N	β-turn	20.5	4.9	9.5	10.2
6O...9N	β-turn	7.0	---	7.2	8.7
7O...10N	β-turn	3.4	---	---	---
2O...4N	β-turn	1.5	---	---	---
3O...5N	γ-turn	8.5	---	---	6.2
4O...6N	γ-turn	5.1	---	---	---
5O...7N	γ-turn	2.2	10.2	4.0	3.8
7O...9N	γ-turn	2.4	---	---	---
8O...10N	γ-turn	2.2	---	2.5	4.1
7O...4N	Reverse turn	---	2.5	2.3	3.2
9O...5N	Reverse turn	---	3.8	2.8	2.5
9O...7N	Reverse turn	---	3.2	1.8	1.1
8O...2N	Reverse turn	1.5	---	---	---
8O...5N	Reverse turn		4.2	**---**	---
10O...7N	Reverse turn	1.3	10.8	2.3	7.1
10O...5N	Reverse turn	1.1	2.8	**---**	---
10O...2N	Reverse turn	2.0	---	**---**	---
1O...9N	Loop	4.1	5.6	**---**	---
2O...9N	Loop	---	2.8	2.9	2.5
2O...10N	Loop	---	1.5	1.1	2.5
3O...10N	Loop	1.3	4.5	0.5	---
3ND1...5N	Side chain interaction	14.0	7.8	3.4	3.5
8ND1...10N	Side chain interaction	1.3	**---**	1.2	2.4

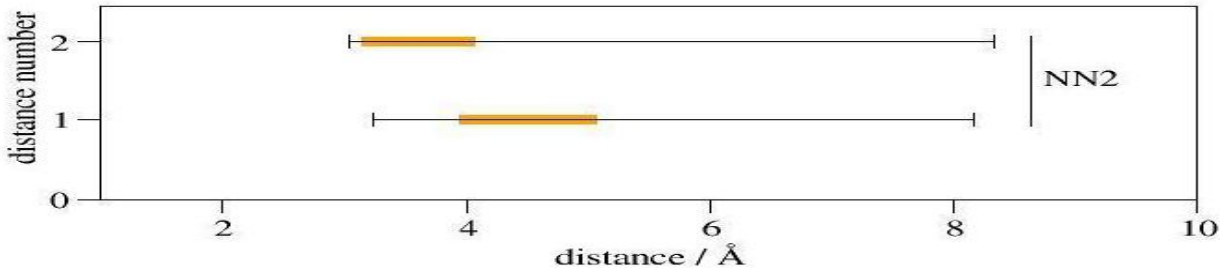
Since the computational analysis described above only provides the information about the average structure of NMC peptide, a direct comparison of the performance of the different protocols used to describe the peptide folding process with the reported NMR experiments **[169]** was thought to be an accurate approach. Since the

intensities of NMR NOE's has an inverse relation with the six power of distance between atoms **[187]**, it was thought worthwhile to perform a comparison of those with the distances measured in the computations. Due to the non-availability of experimental NMR distances for NMC, it was thought equally important to use those reported for the NMC-Ni(II) complex **[169]** as these distances only considers residues 4 to 9 of NMC excluding those participating in the complexation with Ni(II). Since reported biological activity of bombesin and GRP peptides resides in their C-terminal residues, the utilization of NMR NOE's, which includes middle and C-terminal residues of NMC (uncomplexed segment of NMC), with the present computational results was also thought to be another significant reason for this comparison. Accordingly, the CLASTERIT algorithm of the CLASICO program **[135]** was used to compute average distances, corresponding to the NMR NOE's **[169]**, between atoms for each of the different trajectories investigated in the present studies. Only 95% of the structures for each of the trajectories were considered for these calculations and they were assumed to exhibit a normal distribution. The results obtained from both MD and REMD trajectories were compared with NMR results by computing the overlaps between reported NMR distances and those computed from the present studies, as depicted in Figures 5.7 and 5.8. These overlapping results involve the comparison of both long distances (i to i+2 type interactions) and short distances (i to i+1 type interactions) between the atoms.

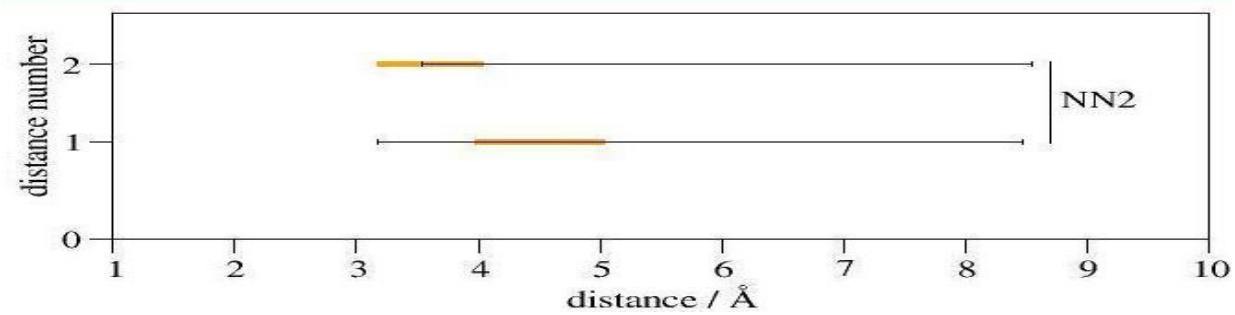
(a)



(b)



(c)



(d)

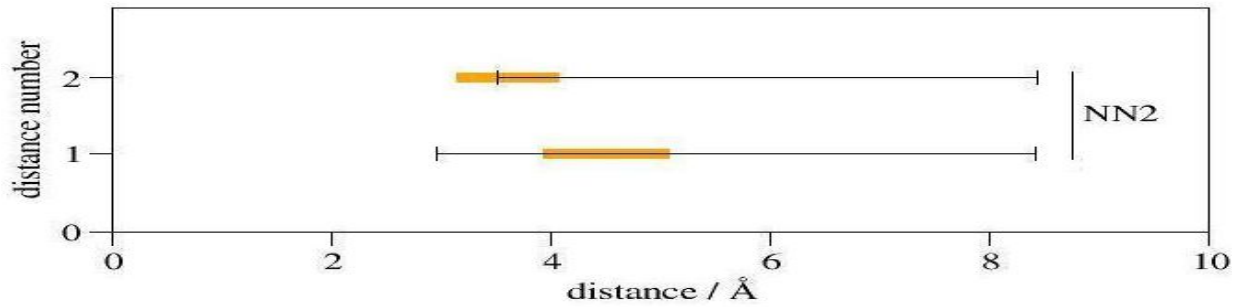


Figure 5.7 Comparison of NMR derived long distances (i to i+2) obtained from Gasmi et al. **[169]**, shown in orange and the average with the distance interval containing 95% of the structures for **(a)** $REMD^{implicit}$ **(b)** $REMD^{explicit}$ **(c)** MD^{Lang} and **(d)** MD^{Berend} trajectories.

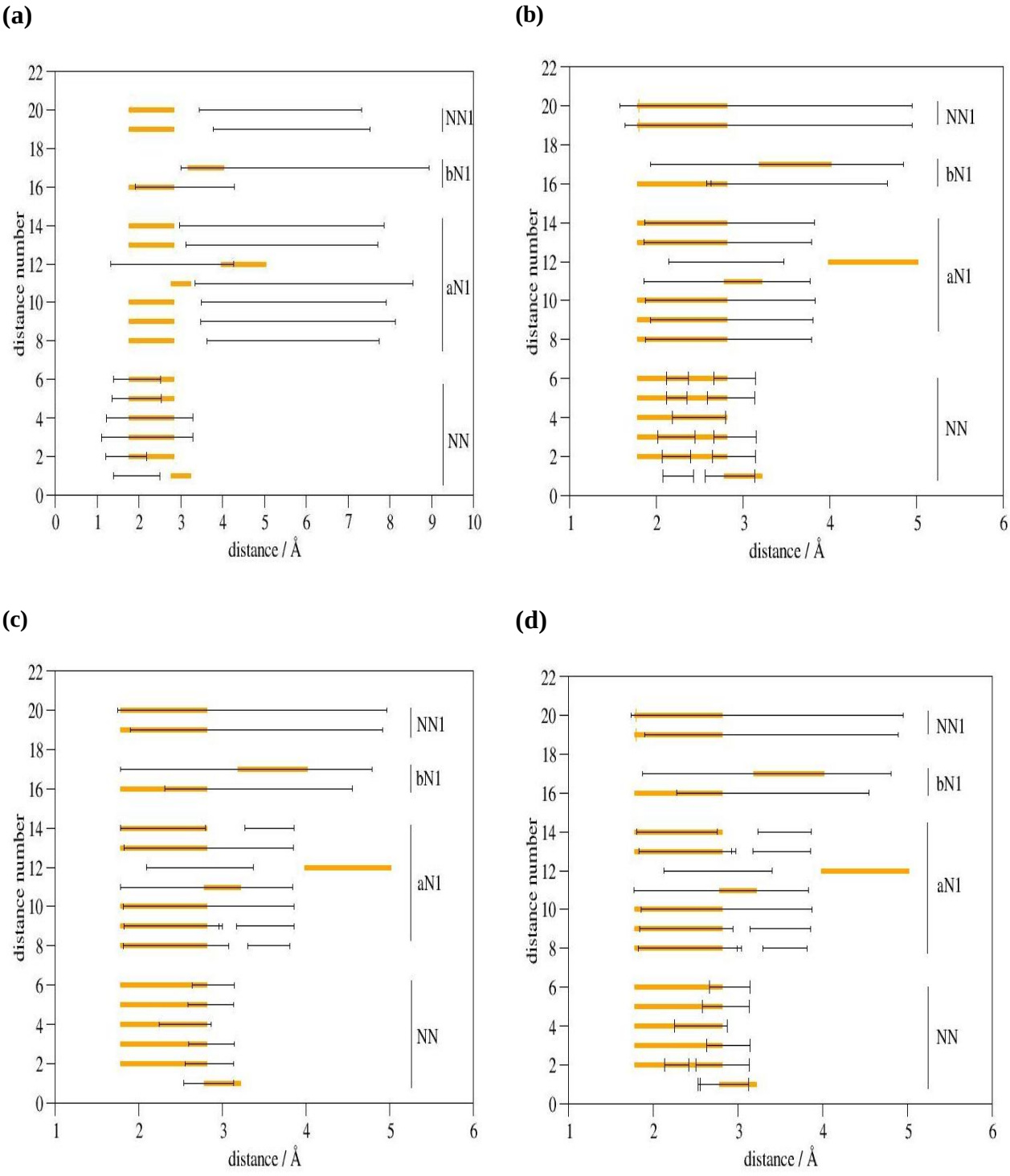


Figure 5.8 Comparison of NMR derived short distances (i and i to i+1) obtained from Gasmi et al. **[169]**, shown in orange and the average with the distance interval containing 95% of the structures for **(a)** $REMD^{implicit}$ **(b)** $REMD^{explicit}$ **(c)** MD^{Lang} and **(d)** MD^{Berend} trajectories.

Specifically, Figure 5.7a reveals that distances calculated from $REMD^{implicit}$ trajectory have no agreement with the NMR results clearly suggesting the absence of NMR structures in the trajectory. $REMD^{explicit}$ (Figure 5.7b) , MD^{Lang} (Figure 5.7c) and MD^{Berend} (Figure 5.7d) trajectories, on other hand, shows good overlapping of the calculated distances with the NMR results although the extent of overlap in MD structures was comparatively low. Similarly, a closer inspection of Figure 5.8 shows that of the 17 reported short NMR distances, there is a good agreement in 9 for those computed from the trajectory $REMD^{implicit}$, 16 for the trajectory $REMD^{explicit}$, 16 for the trajectory MD^{Lang}, and 16 for trajectory MD^{Berend}, respectively. The elongation of computed distances suggests that structures are in a dynamic equilibrium between ordered and unordered forms during the simulations.

5.4 Conclusions

The results obtained reveal that the peptide has a tendency to attain folded, unfolded and helical conformations, probably due to the low energy barrier between them, and therefore accounts for the higher flexibility of NMC. $REMD^{explicit}$ on the other hand samples preferably folded conformations with a higher content of turns (β and γ) and does not promote helicity to a considerable extent. Overall, comparison of the results obtained in the present work using different computational protocols suggests that the majority of the motifs of the NMC peptide adopt preferably folded conformations (turns) as compared to the helical form. Moreover, the four simulations performed produce different patterns of the configurational space sampled (Figure 5.3). Thus, both $REMD^{explicit}$ and $REMD^{implicit}$ simulations samples the space with new conformations appearing in a regular fashion, whereas MD trajectories (MD^{Lang} and

MD^{Berend}) were observed to get trapped in different regions of the conformational space. The sampling results obtained from the different trajectories were compared with the reported experimental 2D-NMR results performed on NMC in the presence of Ni(II) metal. In particular, $REMD^{explicit}$ provides a better agreement to the long and short range distances. Our results suggest that REMD under explicit solvent conditions is better than the rest of the simulations in terms of their efficiency, and compared favorably with the experimental NMR data. Moreover, these results compliment the work of Sarkar and co-workers **[169]**, where they have reported the role of NMC as a Cu(II) and Ni(II) transporter in the central nervous system on the basis of NMR experiments. Hence, the present work provides comprehensive information about the conformational preferences of NMC, which could assist to better understand its native conformation and pave the way towards developing new antagonists.

CHAPTER 6

A CONFORMATIONAL PROFILE STUDY OF NEUROMEDIN B USING DIFFERENT SAMPLING TECHNIQUES

In this chapter, molecular dynamics (MD) replica exchange molecular dynamics (REMD) methodologies were employed to explore the conformational profile of neuromedin B (NMB), a structural analogue of bombesin and NMC. In addition, the simulated annealing (SA) method has been performed and the results were analyzed by cluster analysis. Comparison of the results obtained from these simulations indicates that the peptide has a tendency to attain both β-turns and α-helical conformations regardless of the simulation protocols used. Moreover, the REMD results compared favorably with the corresponding NMR data of NMB reported in the literature.

6.1 Introduction

Neuromedin B (NMB) **[174]**, which belongs to the ranatensin subfamily of bombesin-like peptides **[188]**, was first isolated from porcine spinal cord, that exhibits various biological functions in the central nervous system and gastrointestinal tracts including thermoregulation **[62]**, stimulation of the secretion of the gastrointestinal hormones **[189]**, the regulation of smooth muscle contraction **[190]**, and the ability to function as a growth factor in small cell lung cancer cells and murine 3T3 cells **[191]**. Bombesin-like peptides share similar amino acid sequences in their amidated C-terminal regions and these amino acids are thought to play an important role in the binding of their receptors, which are responsible for their related pharmacological effects. NMB exerts its effect presumably by binding to the cell surfaces **[192]**. There are three

known mammalian bombesin receptor subtypes *viz*., the neuromedin B receptor (NMB-R or BB1) **[193]**, the gastrin-releasing peptide receptor (GRP-R, or BB2) **[194]**, and bombesin receptor subtype 3 (BRS-3, or BB3) **[195]**. Despite sharing a sequence of 50% similarity, these three subtypes are pharmacologically distinct. NMB binds to the NMB-R with the highest affinity, GRP-R with a lower affinity and BRS-3 the lowest affinity **[196]**.

Diverse spectroscopic studies of NMB having the sequence GNLWATGHFM-NH_2 in NMR **[197]**, Infrared (IR) **[198]**, Circular Dichroism (CD) and Fluorescence spectroscopy **[199]** are widely reported in the literature. A recent study of the structure activity relationship (SAR) of bombesin (Glp-Gln-Arg-Leu-Gly-Asn-Gln-Trp-Ala-Val-Gly-His-Leu-Met-NH_2) using an alanine scan, a technique used to determine the contribution of specific residues to a protein's function by mutating the residues to alanine, **[200]** suggested that Trp^4, His^8 and Leu^{12} residues corresponding to Trp^4, His^{12} and Phe^9 respectively in NMB, are important for the binding to the NMB receptors **[201]**. Erne and Schwyzer **[198]** suggested that in the phospholipids bilayer NMB adopts α-helical conformation in the C-terminal region based on IR studies on NMB. Recent studies have demonstrated that small peptides are able to exist in a dynamic equilibrium between folded and unfolded structures, depending on the solvent polarity and their interaction with the membrane phase **[202]**. In aqueous solutions small peptides are known to adopt many conformations since the hydrogen bond formation between the polar backbone carbonyl and the amide groups and water solvent effectively competes with an intramolecular hydrogen-bond formation **[202-203]**. Eugenia et al. **[199]** suggested that, based on circular dichroism,

fluorescence and molecular dynamics studies, NMB adopts an α-helical structure in an apolar environment. However, in aqueous solution NMB adopts unordered and very flexible structures. In vacuum 50% **[199]** of the structures of NMB are helix-like, with a right-handed chirality beginning from the tryptophan residue through to the C terminus and was found to be independent of the initial conformation. Moreover, two-dimensional (2D) nuclear magnetic resonance (NMR) studies of NMB suggest that the peptide adopts a relaxed helical conformation from Trp^4 to Met^{10} in a 50% aqueous trifluoroethanol (TFE) solution, and in 150 mM sodium dodecyl sulfate (SDS) micelles. Several reports also suggested that there might be a conformational change to a β-turn type structure upon binding to the receptor **[204]**.

Despite being remarkably vital, spectroscopic methods alone cannot provide all the structural details necessary to fully understand the conformational profile of the peptides in solution due to the flexibility of these molecules. Computational studies on the other hand, can provide detailed complementary information about the intrinsic conformational features of the peptide. The methodologies available nowadays to investigate the propensities of a peptide to adopt different conformations are solid enough to provide a reasonable picture of the conformational features of a peptide and the way the solvent affects them. Despite having its imperative role as a growth factor in small lung cancer cells, no bioactive conformation of NMB has been elucidated on the basis of available experimental and computational techniques. Accordingly for the purposes of this study, only a short molecular dynamics (MD) simulation of 300 ps on NMB under implicit solvent conditions has been reported

[199]. Previous studies **[205]** have demonstrated that long enough MD simulations represent an adequate tool to obtain the folded conformation of a peptide in solution. Although, an accurate description of the solvent environment is essential for realistic biomolecular simulations, it is always found to be computationally expensive. Recently, Generalized Born surface area implicit solvent models have been used in bimolecular simulations. This methodology has become popular, especially in molecular dynamics applications due to its relative simplicity and computational efficiency, compared to the more standard numerical solution of the Poisson–Boltzmann (PB) equation **[205]**. Recent modifications to the standard GB implementations extend its applicability to the entire range from low- to high dielectric environments and thus play an imperative role to reproduce the environment induced by different explicit solvents **[206]**.

In this study we investigate the folding of NMB in implicit water using longer MD and REMD methodologies, starting from an extended configuration. Both implicit solvent simulations were carried out using Generalized Born (GB) approximation since the GB models treat the solvent as a dielectric continuum, as reported in literature **[3, 5]**. Specifically, the Onufriev, Bashford, and Case (OBC) implicit water model **[91]** has been employed for the current investigations as this solvent model in combination with AMBERff96 is reported to generate a better extent of the helices and β-sheet conformations in peptides **[184]**. The significance of the REMD simulations for the current investigation enhances the sampling of conformations, owing to their frequent switching of simulation temperatures between a number of concurrent simulations. Simulated annealing (SA) as an additional method has also been employed to

sample the low energy conformations attained by NMB peptide in implicit water, using the Onufriev, Bashford, and Case (OBC) continuum solvent model **[91]**.

6.2 Computational Methods

6.2.1 Replica Exchange Molecular Dynamics (REMD)

The linear sequence of NMB (GNLWATGHFM-NH_2) was constructed using the leap module of AMBER 9 **[92]** with the N-terminal protonated and C-terminal amidated. The extended structure of NMB was minimized when a convergence criterion of 0.005 kcal mol^{-1} $Å^{-1}$ was reached. REMD was then performed on the minimized structure using the Generalized Born implicit solvent model (solvent dielectric constant 78.5, surface tension 0.005 cal/$mol^{-1}Å^2$ **[207]**) was used to model the effects of solvation **[208]**. The internal dielectric constant around the peptide was set to 1. The SHAKE algorithm with a relative geometric tolerance of 10^{-5} was used to constrain all bond lengths to their equilibrium distances. Prior to the REMD simulations, standard MD simulations were performed for 5 ns at a range of different temperatures (200-900 K), with a temperature difference of 100 K. In the present study, twelve replicas were used and the temperature of each replica was set at: 277, 300, 326, 354, 385, 419, 457, 498, 544, 595, 651, and 713 K, with a time step of 0.2 fs. The temperature during the MD simulations was regulated by the Langevin thermostat **[185]**. Each replica was simulated simultaneously and independently at different replica temperatures. The replica exchange was performed every 2 ps for 50,000 steps during the REMD simulations.

6.2.2 Molecular Dynamics (MD)

MD trajectory was undertaken using the Generalized Born (GB) approximation at 300 K employing the Langevin coupling algorithm. Internal dielectric constant around the peptide was set to 1, while the external dielectric constant of 80 corresponding to water was employed. In order to mimic the physiological conditions a 0.2 M salt concentration was used. SHAKE was used on all bonds involving hydrogen atoms with a time-step of 2 fs.

6.2.3 Simulated Annealing (SA)

The extended conformation of NMB peptide was energy minimized using the steepest descent method followed by a conjugate gradient method until a convergence of less than 0.001 kcal mol^{-1} $Å^{-1}$ between successive steps was achieved using the SANDER module of AMBER 9 **[92]**. The SA calculation was performed under implicit solvent conditions using the GB-OBC continuum solvent model **[91]**. For this purpose, all electrostatic calculations throughout this study were done with the relative permittivity of 80. The minimized starting structure was heated up to 900 K at a rate of 100 K ps^{-1}. This means that the structure was first heated to 200 K, allowed to equilibrate and then reheated to 300 K, and this heating process was repeated until a temperature of 900 K was reached. The use of this high temperature was to provide the molecules with sufficient kinetic energy to enable them to cross energy barriers between different conformations, as quickly as possible. At this point the structure was slowly cooled from 900 K down to 200 K at a rate of 50 K ps^{-1}. In this technique the system was cooled down at regular time intervals, by decreasing the simulation temperature from 900 K to 200 K in intervals

of 50 K. As the temperature approaches 200 K the molecule is trapped in the nearest local minimum conformation. At the end of the annealing cycle, the geometry of the structure was minimized at 200 K, in order to remove the internal strain of the molecule. Information regarding the coordinates and minimized energy data at 200 K is saved separately on a data file, which completes a single cycle of simulated annealing. Subsequently the optimized structure was used as the starting conformation for the next cycle of SA. In this case, 8000 cycles of iterative simulated annealing resulted in a library of 8000 structures being accumulated, were ranked according to their energy values. The primary objective of a conformational analysis is the identification of low energy structures, and this forms an important part of understanding the relationship between the structure and the biological activity of a molecule. The biological activity of a drug molecule depends on a single unique conformation hidden amongst all the low energy conformations **[209]**. The search for this so-called bioactive conformation for sets of compounds is one of the major tasks in medicinal chemistry. Only the bioactive conformation can bind to the specific macromolecular environment at the active site of the receptor protein **[210]**. An understanding of the manipulation of the conformational structures of peptides using highly restricted segments ultimately leads to the design of bioactive peptides to fit the three dimensional receptor site requirements. In the identification of low energy structures, the SA strategy employed is widely used in the characterization of low energy conformations **[209]**, and the following protocols were used. Firstly, the structures were rank-ordered by energy every 100 cycles and checked for uniqueness. The efficiency of this process was monitored according to the equation:

$$\lambda(N) = \frac{\xi(N)100}{N\xi(100)} \qquad 6.1$$

The efficiency parameter, γ, was computed every 100 cycles of SA, which is defined as the number of unique conformations, ξ, found after N cycles of SA, ξ (N), divided by N, and adjusted by a coefficient so that the efficiency parameter is unity after the first 100 cycles performed, which completes the criterion of the iterative process **[211]**. The procedure was terminated in all cases when the calculated efficiency of the process, γ was at least 10% below the starting value. The evolution of this parameter was monitored along the conformational profile, for the peptide.

6.3 Results and discussion

6.3.1 MD and REMD

Thermodynamic profiles shown in Figures 6.1a-b were used to monitor the quality of the MD trajectory, and reveals that the simulation is relatively stable over the 200 ns of MD trajectory. The potential energy fluctuating around -225 kcal mol^{-1} is shown in red, while the kinetic energy oscillating around -100 kcal mol^{-1} is shown in black in Figure 6.1a. On other hand, the time series of temperature exchanges (Figures 6.1c) of NMB for the initial 100 steps of REMD trajectory revealed that a random walk in the 'temperature space' between low and high temperature was realized in each case.

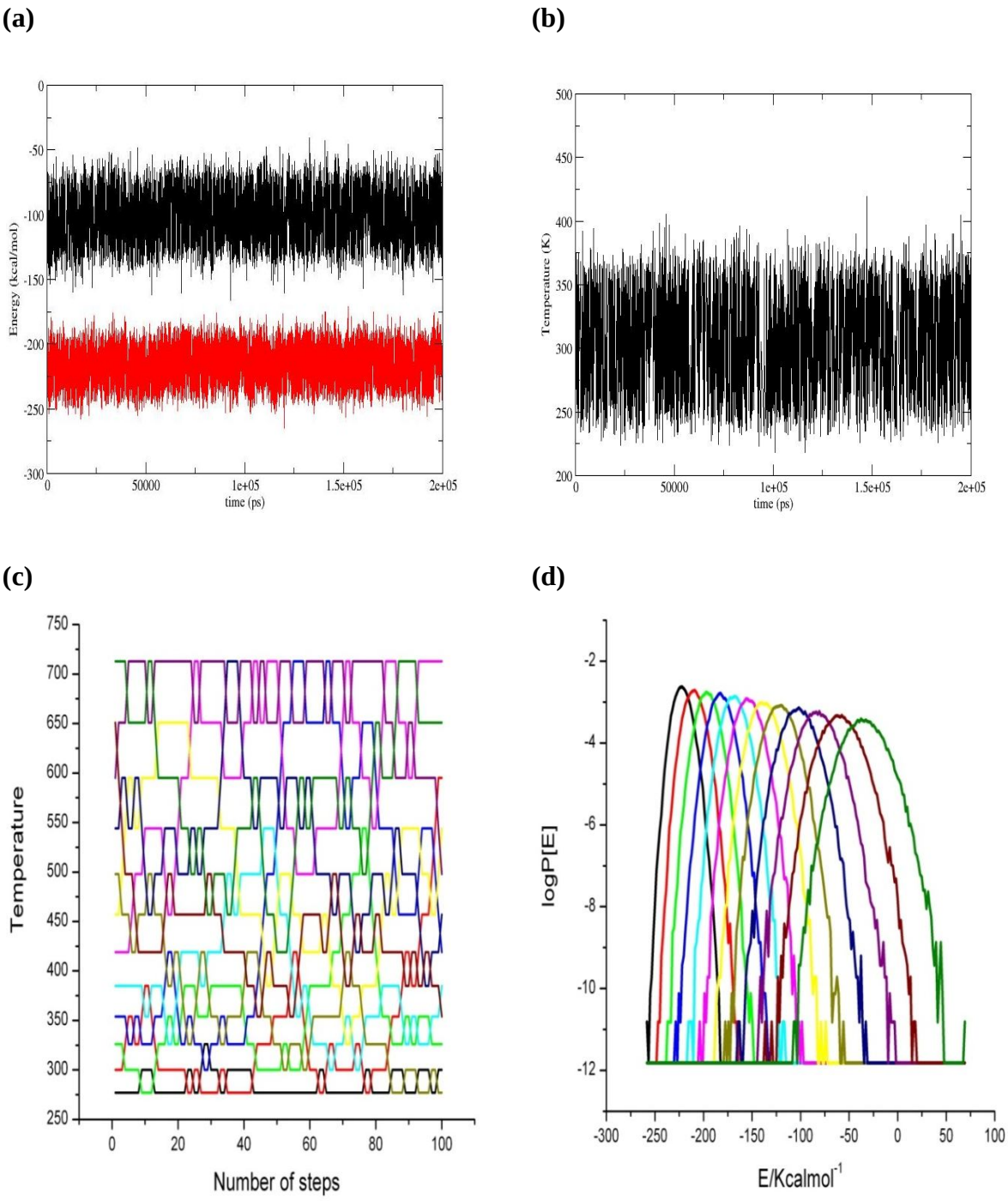
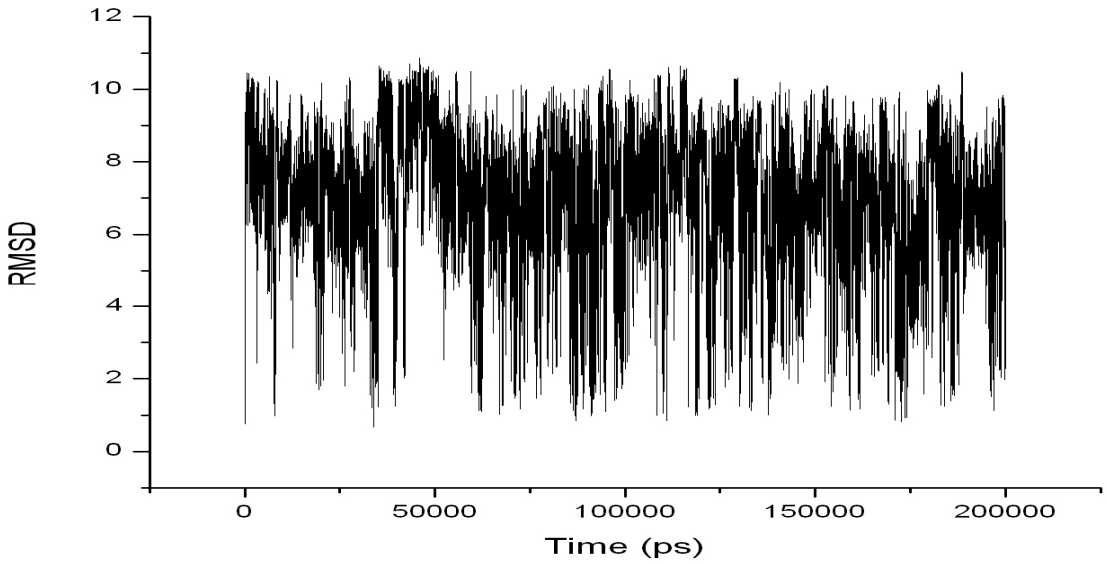


Figure 6.1 Trends of **(a)** energies (potential energy in red and kinetic energy in black) and **(b)** temperature during MD run. Time series of temperature exchange **(c)** for initial 100 steps of REMD, and the canonical probability distributions of the total potential energy **(d)** of NMB obtained from REMD simulation at twelve temperatures.

The canonical probability distributions of the total potential energy depicted in Figures 6.1d along the REMD trajectory reveal the overlap between all the neighbouring pairs of distributions, suggesting their efficient performance.

Figures 6.2a and 6.2b show the root-mean-square deviations (RMSD) of 200 ns of MD and 100 ns of REMD trajectories respectively, monitored relative to the backbone atoms of starting structures. Specifically the RMSD (Figure 6.2a) shows considerably higher fluctuations (between 1-10 Á) clearly indicating a rapid equilibrium between folded and unfolded structures. These results also suggest that the conformations closer (RMSD ~ 1 Á) to the starting structures are regularly sampled at different intervals of the MD trajectory. RMSD comparison of REMD trajectory (Figure 6.2b), on other hand, reveals that most of the sampled conformations are different from the starting structure and oscillate mostly between 3-8 Á during the course of the trajectory. Overall, these results imply that the conformational energy barrier between different conformations is probably small and account for the high flexibility of the NMB peptide observed in both the MD and REMD trajectories.

(a)



(b)

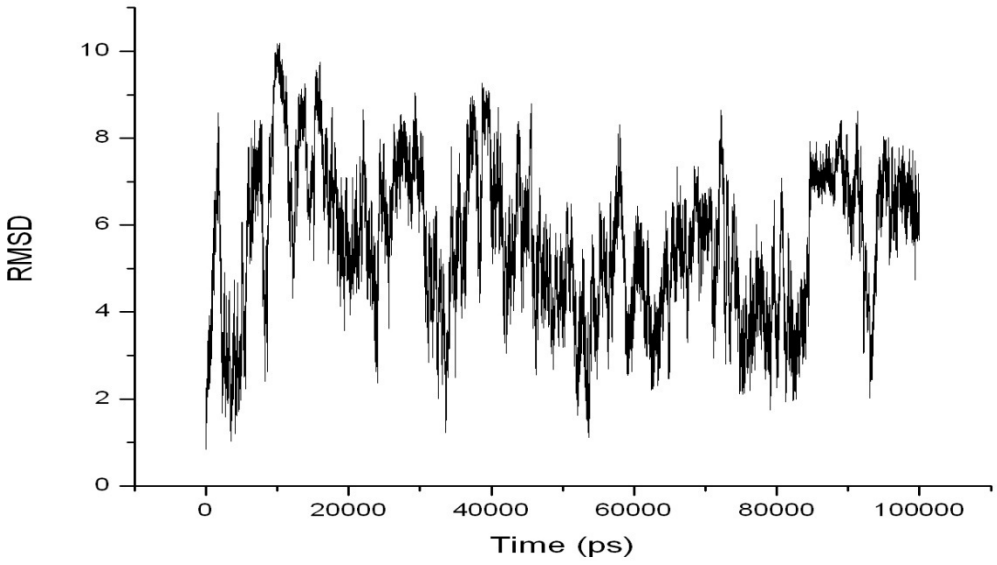
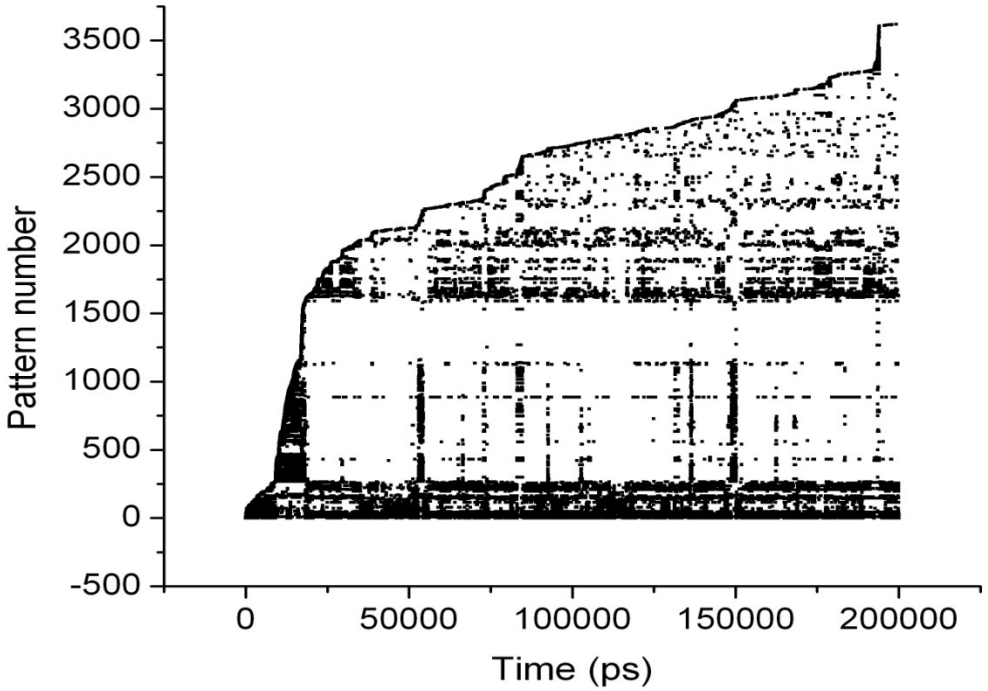


Figure 6.2 Root mean square deviations (backbone-backbone) of NMB from the starting structure in case of **(a)** MD and **(b)** REMD trajectories.

The sampling efficiency of MD and REMD trajectories were monitored by establishing different conformational patterns attained by NMB during the progress of the simulations. For this purpose, the CLASICO program **[135]** was used to compute the pattern profile for every snapshot of MD and REMD trajectories, and is depicted in Figures 6.3a and 6.3b respectively. Accordingly, 105439 (52.7%) patterns (Figure 6.3a) were obtained for 200000 snapshots of MD whereas 68753 (68.7%) patterns (Figure 6.3b) were identified for 100000 snapshots of REMD trajectory. These plots provide a broad estimation of the performance of the different protocols in sampling new patterns. A closer inspection of Figure 6.3a reveals that new patterns were sampled in a steady manner during the initial 10 ns of the trajectory suggesting probably the initial folding process of the peptide. A sharp increase in the number of new patterns was observed after 10 ns and remained in the same rate for the next 5 ns after which new patterns were sampled in a slow but constant manner until the end of the trajectory. However, the peptide conformations seem to get trapped (dark areas) in regions of the conformational space at certain intervals clearly suggesting its restrictive nature to explore the new patterns. In the case of REMD (Figure 6.3b), new patterns were sampled from the start of the simulation and progressed in a uniform fashion during the evolution of the trajectory. Moreover, the presence of a less dark region in the plot (Figure 6.3b) reveals that new patterns were explored with less restriction, clearly suggesting a better sampling performance of REMD over MD. Moreover, convergence seems to be attained in the case of REMD after 100 ns as the appearance of new patterns were almost negligible at the end of the trajectory.

(a)



(b)

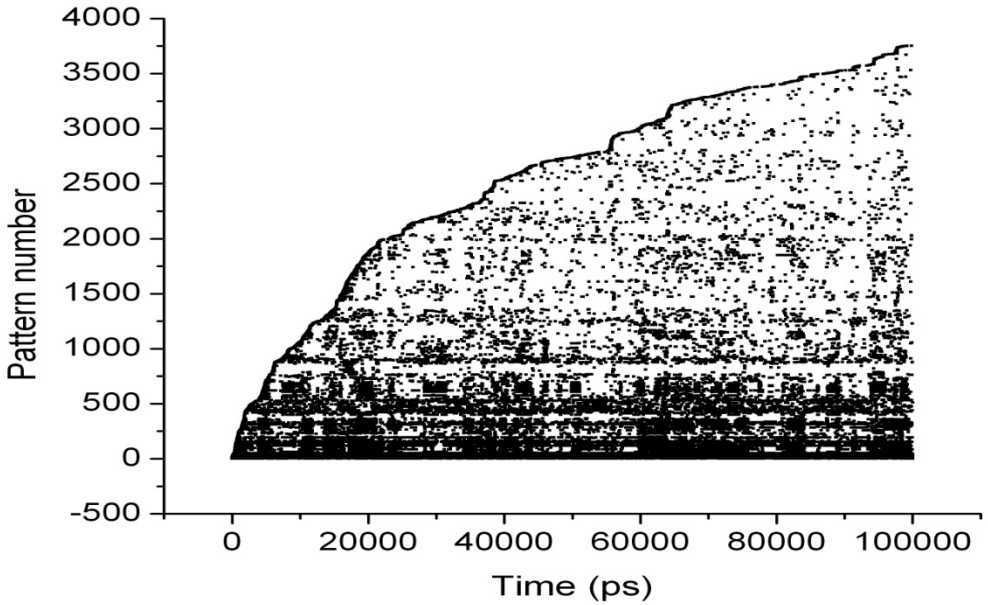


Figure 6.3 Evaluation of new patterns for the NMB in **(a)** MD and **(b)** REMD trajectories.

Secondary structure analysis was performed for every snapshot of the MD and REMD trajectories using the CLASICO program. Figures 6.4a-b represents the statistics of the conformational motifs for each residue of the NMB peptide in MD and REMD trajectories, respectively. Figure 6.4a shows the classification of secondary structures obtained in MD trajectory where the peptide exhibits predominantly β-turns (~25%) between residues 3 to 9 with a stronger propensity between residues 5 and 6. Additionally, an α-helical region (4-5%) flanked by residues 3 to 6 were also observed in some of the sampled structures (Figure 6.4a) with the complete absence of 3_{10}-helical conformations. To some extent β-strands (2-3%) in a region between residues 2 to 6 were also found in some of the conformations. The REMD protocol (Figure 6.4b) on other hand, was more efficient at inducing β-turns (~33%) in the sampled conformations flanked by residues 2 to 9 with a higher percentage between residues 5 and 6. The second major conformational motif attained by the structures sampled in the REMD trajectory was the α-helical region (8-10%) flanked by residues 3 to 9 with a higher propensity between residues 3 to 6. All the α-helical conformations were observed to be right-handed, and to a minor extent conformations exhibiting the 3_{10}-helical region between residues 4 to 6 were also obtained (Figure 6.4b). Structures with β-strands between residues 2 to 6 were also observed in REMD trajectory.

The β-turn motifs attained by residues of the NMB peptide were further classified into different types using a two-residue window of the CLASICO program **[135]**, and are depicted in the Figures 6.4c-d. The motifs obtained in the MD trajectory (Figure 6.4c), shows the predominance of a β-turn type I between residues 4 and 9 with a higher

propensity of the type II between residues 6 and 7. To some extent β-turn type III (3_{10}-α helix) was also observed between residues 3 to 9, with a higher percentage between residues 5 and 6, and a lower percentage between residues 3 to 4 and 7 to 9. In addition, β-turn type ii (mirror conformation of β-turn type II) was also observed between residues 5 to 8 (Figure 6.4c). The conformations obtained from the REMD trajectory attain preferably a β-turn type I between residues 3 to 9 with a stronger propensity between residues 3 to 6 (Figure 6.4d). Structures displaying β-turn type III (3_{10}-α helix) flanked by residues 3 to 9, with a stronger propensity between residues 5 and 6 were also sampled in the REMD trajectory. As mentioned before, the CLASICO program does not include the first and last residues of the peptide in the secondary structure calculations and this account for the absence of any of secondary structural features in Figures 6.4a-d.

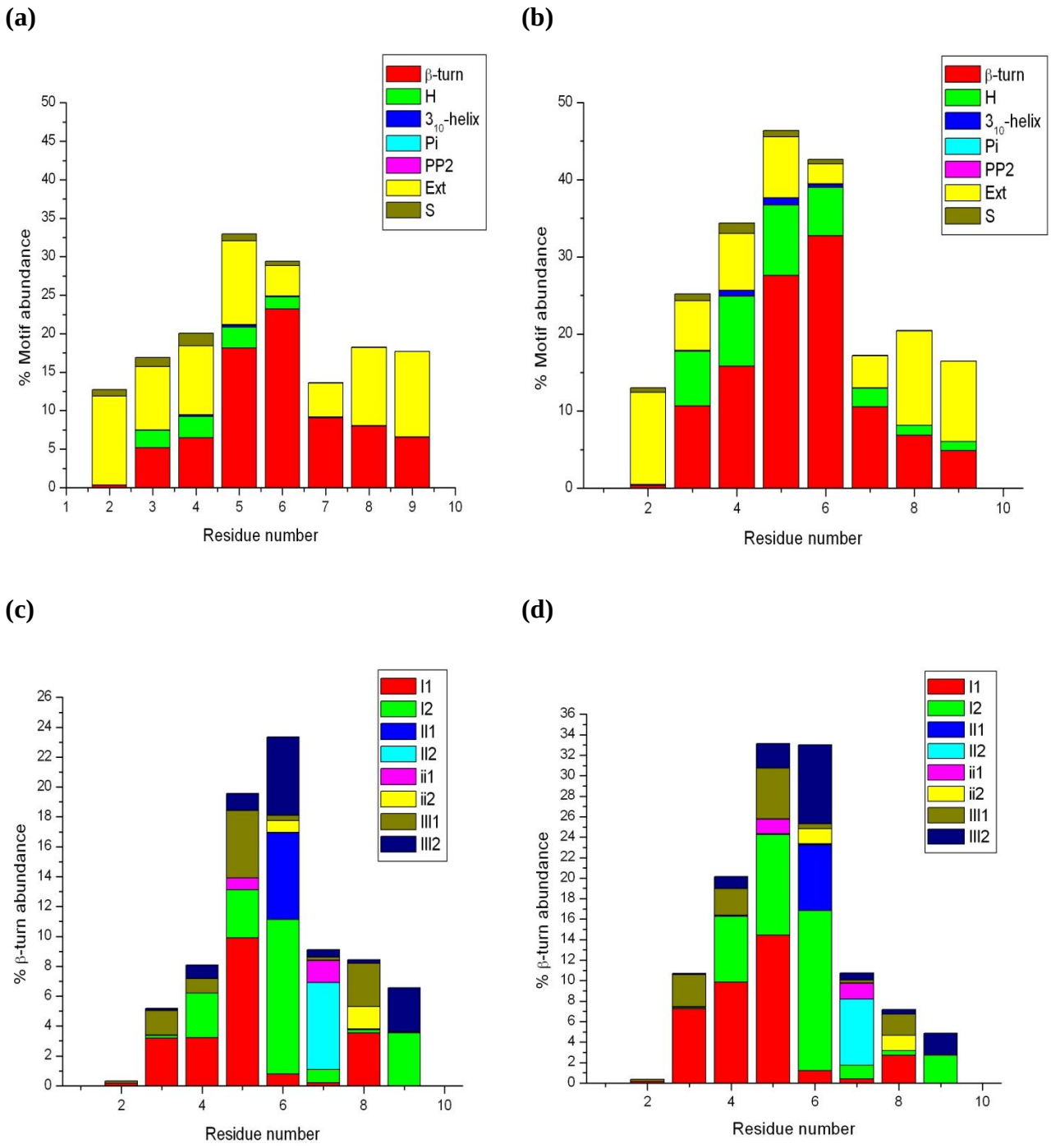


Figure 6.4 Motif abundance for NMB in **(a)** MD and **(b)** REMD trajectories. Conformational motifs are labeled: H (α-helix), PI (π-helix), PP2 (polyproline II), Ext (extended), S (β-strand) **[135]**. Type of β-turns attained by NMB peptide in **(c)** MD and **(d)** REMD trajectories.

Since the computational analysis described above provides an estimation of the average structure of NMB, it was considered worthwhile to compare the results of the different protocols with the reported NMR experiments **[197]**.

The average distances corresponding to the NMR NOE's **[197]** were computed using the CLASTERIT algorithm **[135]** of the CLASICO program. The overlaps between the distances obtained from NMR experiments and those computed from the present studies are depicted in Figures 6.5a-d. These overlapping results compare both long distances (i to i+2 and i to i+ 3 type interactions) and short distances (i to i+1 type interactions) between the atoms. Specifically, Figure 6.5a reveals that only 12 long distances (LD) in the case of MD are in agreement with the corresponding NMR distances **[197]**, and clearly suggests the absence of NMR structure in this simulation, as, only the LD accounts for the secondary structures. A good agreement of short distances (SD) in the case of MD (Figure 6.5c) does not make any contribution in the helicity. On the other hand, all computed LD (Figure 6.5b) and SD (Figure 6.5d) from the REMD trajectory corresponds to the NMR distances clearly revealing the presence of NMR structure. However, elongation of the computed distances (Figures 6.5b and 6.5d) clearly reveals the rapid exchange between NMR and the unordered structures in this segment of the trajectory. Overall these results reveal that the peptide is in a rapid equilibrium between ordered and unordered conformations and suggests low conformational energy barrier between them accounting for the higher flexibility of NMB. Moreover, REMD method is more efficient to induce the helicity and β-turns in the peptide and was also successful in the sampling of the NMR structures which were completely absent in the MD trajectory.

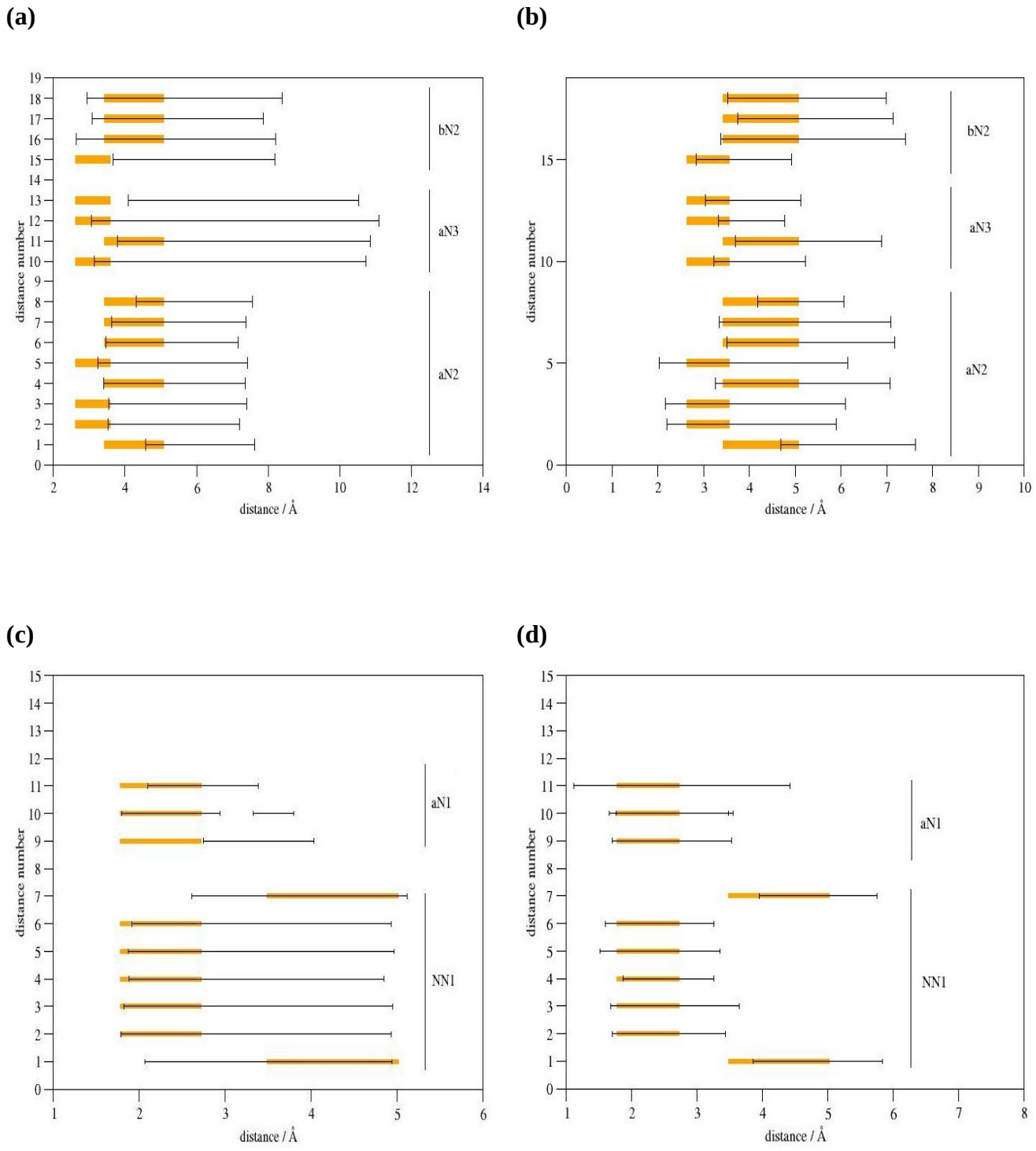


Figure 6.5 Comparison of NMR derived long distances (LD) obtained from Lee et al. **[197]**. shown in orange and the computed average distances in a interval containing 95% of the structures for **(a)** MD and **(b)** REMD trajectories. Similar comparison of short distances (SD) for **(c)** MD and **(d)** REMD trajectories.

6.3.2 Simulated Annealing (SA)

The conformational searches for NMB presented in this study were performed with the simulated annealing (SA) protocol in an iterative fashion, as a sampling technique. The sampling procedure was stopped after 8000 cycles of an iterative SA process for which the sampling efficiency λ (see Equation 6.1) was found to be less than 10% of the starting value. The evolution of this parameter λ was computed using Equation 6.1 and monitored along the conformational profile, as shown in Figure 6.6.

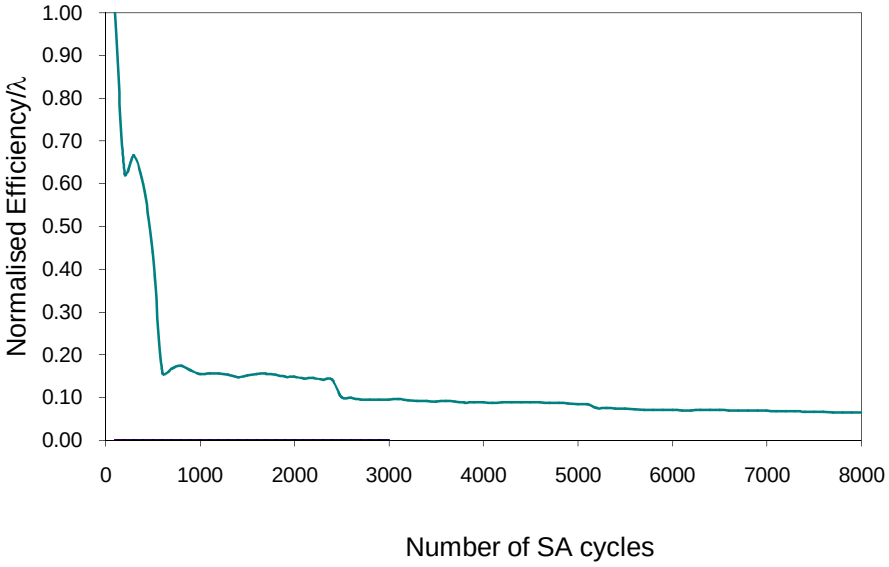


Figure 6.6 Normalised efficiency values for the NMB peptide search.

The shape of the above figure suggests that the search procedure employed was acceptable both in quality and computer time. Low levels of performance were reached in finding new low energy conformations (λ = 0.1) which is the expected result for most peptide analogues **[212]**. After completion of 8000 cycles of SA, the

resulting 8000 structures were stored in a library. The criterion described in the Methods Section 6.3.2 was used to compute the total number of unique conformations for the NMB peptide. Of the total 5521 unique conformations, only 205 low energy structures (<5 kcal mol^{-1}) were observed (Table 6.1).

Table 6.1 Summary of conformational analysis obtained from SA

NMB sequence	Total number of unique conformations	Number of classes obtained from cluster analysis	Number of unique conformations within 5 kcal mol^{-1}
H-GNLWATGHFM-NH_2	5521	10	205

To describe the preferred conformational domains exhibited by the peptide, the low energy structures were clustered into groups according to the values of the root mean square deviation (RMSD) of the distances between the backbone atoms of every structure.

Furthermore the RMSD's for the unique conformations were calculated using the Kleiweg clustering method **[213]** to perform the hierarchical cluster analysis. The clustering can be visually represented by constructing a dendrogram, which indicates the relationship between the items in the data set (i.e. RMSD) and is graphically represented in Figure 6.7. The dendrogram enables us to identify how many clusters there are at any stage and what the corresponding members of the clusters are. It is a useful tool to show the underlying structure of the data and for suggesting the appropriate number of clusters to choose. A line drawn horizontally across the dendrogram enables one to read off how many clusters there are at any particular distance measured, as shown in Figure 6.7.

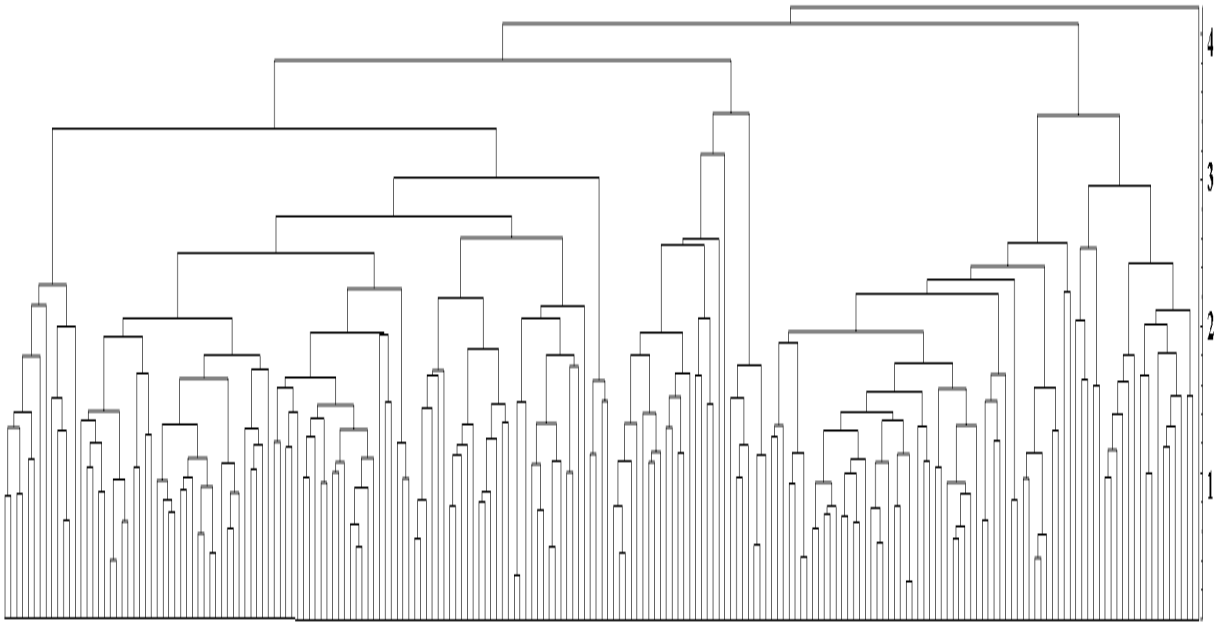


Figure 6.7 Dendrogram showing different clusters for NMB classified according to their RMSD (shown along y-axis) using the Kleiweg clustering method **[213]**.

Since the objective of cluster analysis is to determine the representative structures of the conformational space explored, a carefully selected cut-off value is important as this method will avoid choosing subclusters. Accordingly the RMSD value of 0.8 Å was chosen as a cut-off value. Since another goal of this work was to get a better understanding of the structural motifs of the NMB peptide, further conformational analysis for this peptide was carried out on the low energy structures. The 205 unique conformations were thereby classified into ten clusters (D1 to D10), summarized in Table 6.2. For each of the ten different classes of clusters, a representative of the cluster was chosen on the basis of having the lowest relative energy (designated by E_r min, Table 6.2).

Table 6.2 Cluster analysis for H-GNLWATGHFM-NH_2.

Cluster name	percentage of structures in the cluster	Structure representative of the cluster	E_r min /Kcal mol^{-1}
D1	0.5	39	4.2
D2	10.2	142	3.9
D3	3.4	140	2.8
D4	41.6	108	4.7
D5	0.5	201	3.8
D6	21.9	87	1.5
D7	6.3	110	2.9
D8	9.2	116	1.9
D9	3.9	193	4.0
D10	2.4	123	4.4

The five most abundant clusters represented by D2, D4, D6, D7 and D8 in Table 6.2 corresponds to 10.2%, 41.6%, 21.9%, 6.3% and 9.2% of the total number of structures respectively, clearly suggesting that the bulk of the structures are restricted to small number of clusters. Inspection of Table 6.2 reveals that 89% of the structures are represented by the most abundant clusters.

All the structures in each of the five most abundant clusters were analyzed to determine the conformational motifs attained by the NMB peptide, using CLASTERIT algorithm of the CLASICO program **[135]**. The statistics of all the motifs found in clusters D2, D4, D6, D7 and D8 are depicted in Figures 6.8a–e respectively.

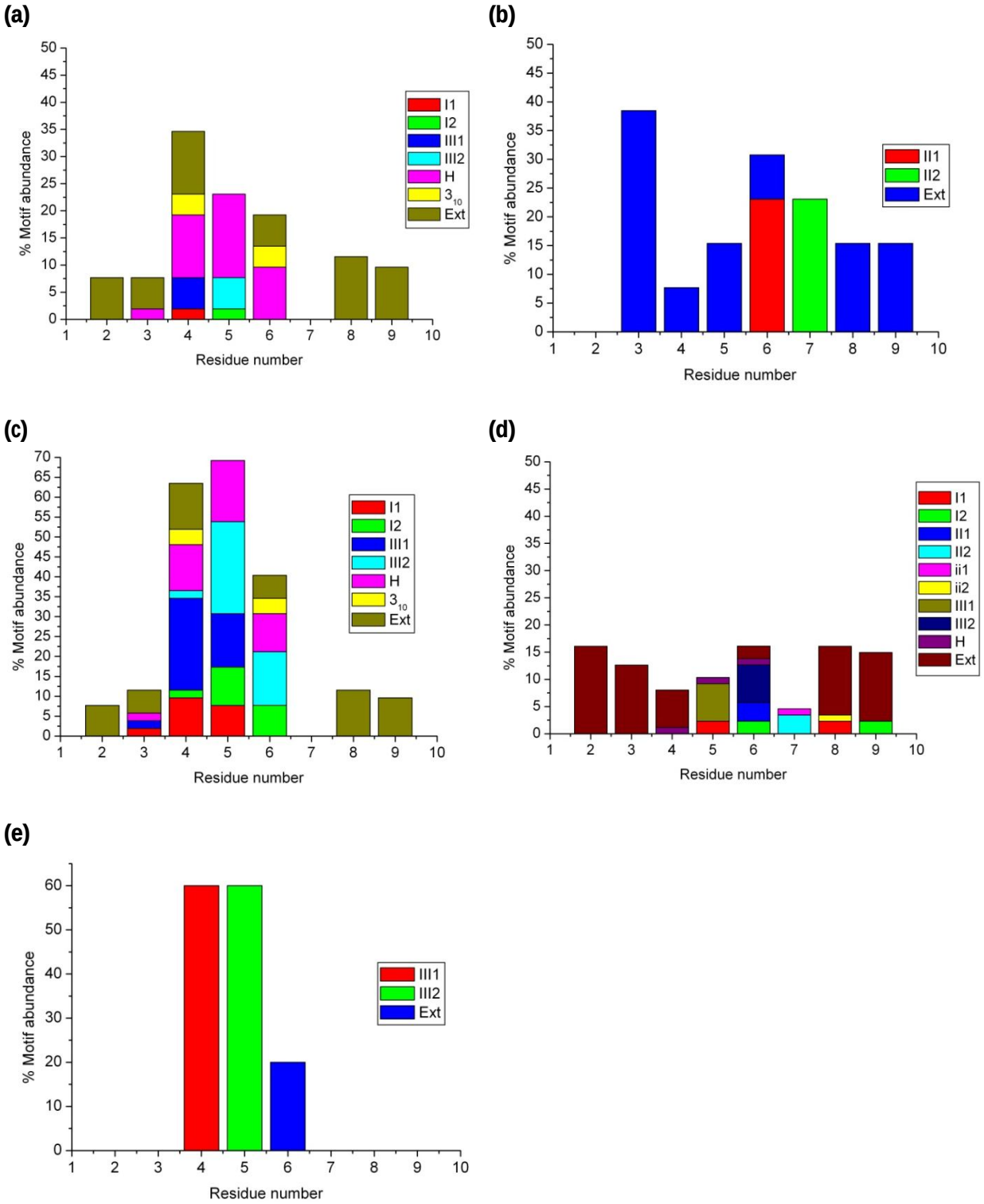


Figure 6.8 Conformational motif abundance attained by NMB peptides in **(a)** cluster D2 **(b)** cluster D4 **(c)** cluster D6 **(d)** cluster D7 and **(e)** cluster D8. Conformational motifs are labeled: H (α-helix), 3_{10} (3_{10}-helix), Ext (extended), as defined in Table 3.1 and Table 3.2 **[135]**.

A closer inspection of Figure 6.8a reveals that most of the structures in cluster D2 (Table 6.2) predominantly exhibit an α-helical region between residues 3 to 6, while most of the residues (2-4, 8-9, Figure 6.8a) prefer to stay in the extended form. The structures of the most populated cluster D4, on other hand, were observed displaying a β-turn type II between residues 6 and 7 while the rest showed only an extended region (Figure 6.8 b). The conformations of the second most abundant cluster D6 (Figure 6.8c) displayed predominantly an α-helical region between residues 3 to 6 with a stronger propensity between residues 4-5. To some extent the β-turn type III (3_{10}-α helical) and type I between residues 3-6 were also observed in some of the structures. However residue 7 did not show any secondary structural feature as most of the structures in clusters D7 were extended, while some displaying α-helical region between residues 4-6 and β-turns of type II and type I between residues 6-7 and 5-6 respectively, were also part of the cluster (Figure 6.8d). Almost 60 % of the structures in cluster D5 did not display any ordered conformation except a β-turn type III (3_{10}-α helical) between residues 4 and 5 (Figure 6.8d).

6.4 Conclusions

Our results suggest that the peptide adopts different folded and unfolded conformations regardless of the protocols used. REMD under GB conditions sample the new patterns in a uniform fashion and appears to have easily reached the convergence (Figure 6.3b), whereas conformations within the MD simulation seems to get trapped in certain regions of the conformational space making it less efficient. Moreover, the results obtained from REMD and MD computational protocols were compared with the available NMR results of NMB in literature. The comparison indicates that REMD shows good agreement with the reported NMR results. MD results, on the other hand, do not correspond with the reported NMR NOEs, clearly indicating the absence of NMR derived structures in this simulation. Moreover, the results obtained from SA is also in agreement with the corresponding REMD results clearly suggesting the probable existence of both turns and helicity in the NMB peptide, and thus may be responsible for binding of NMB at its receptor site. Hence, the present work provides comprehensive information about the conformational preferences of NMB explored using three different techniques which could be helpful to better understand its native conformation for future investigations.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

The present work was aimed at getting a deeper understanding of the structure-activity relationships for three related neuropeptides: bombesin, and its two mammalian analogues neuromedin B (NMB) and neuromedin C (NMC). For this purpose the conformational profiles of these peptides were explored using different computational techniques. Methods employed included: simulated annealing (SA), molecular dynamics (MD) simulations and replica exchange molecular dynamics (REMD), using different computational protocols. The performance of these protocols was assessed by comparison with the available NOE measurements obtained from NMR experiments published in literature. Analysis of the results obtained for bombesin revealed an unprecedented dual conformational behaviour of this peptide with a clear tendency to exhibit an α-helix between residues 6-14, in accordance with the reported literature, and a helix turn observed between residues Gly^5 and Ala^9. This information could be extremely useful to better understand the selective affinity of this peptide to the neuromedin B receptor (BB1) and gastrin releasing peptide receptor (BB2). The computational studies carried out on NMC reveal the propensity of this peptide to attain β and γ-turns preferentially over the helical conformations, which fully supports the experimental results. In relation to NMB the computational studies suggest the tendency of this peptide to adopt both turns and helices in the low energy states, and points to the possibility of assessing two different conformations, one for the recognition and the second for its binding ability with the receptor. Overall, these results contribute imperative information regarding

the structure, function and folding properties of neuropeptides, by complementing and strengthening the available experimental NMR data.

Recommendations

The current simulation studies on the selected neuropeptides (bombesin, NMB and NMC) indicate that the AMBER ff96 force field in combination with the Onufriev, Bashford and Case (OBC) implementation of the GB method reproduces 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. Since the studies carried out on NMB suggests that the results from the simulated annealing are equivalent to those obtained from MD simulations, comparisons of the two procedures need to be performed by comparing a Maxwell-Boltzmann distribution using the SA results with a converged MD simulation. In particular, the structures of peptides obtained from these simulations can be used to design agonist 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 modelling: scientific and technological boundaries, Pure & Appl. Chem. 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. 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.

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

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

6. Emsley, J. 1980. Very Strong Hydrogen Bonds, Chemical Society Reviews, (9): 91–124.

7. Campbell, N.A., Brad, W., Robin J., and Heyden. 2006. Biology: Exploring Life. Boston, Massachusetts: Pearson Prentice Hall.

8. Burley, S.K. and Petsko, G.A. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization, Science, (229): 23-28.

9. Gerald, R., Grimsley, K.L., Shaw, L.R., Fee, R.W., Alston, B.M.P. Huyghues-Despointes, R.L., Thurlkill, J.M., Scholtz, C., and Nick P. 1999. Increasing protein stability by altering long-range coulombic interactions, Protein Science, (8): 1843–1849.

10. 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.

11. Thomasson, W.A. 2008. Unravelling the mystery of protein folding. http:/www.faseb.org/opa. Breakthroughs in science. Accessed on 23 May 2008.

12. 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.

13. 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.

14. IUPAC-IUB Commission on Biochemical Nomenclature .1970. Abbreviations and symbols for the description of the conformation of polypeptide chains, Journal of Biological Chemistry, (245): 6489–6497.

15. Taylor, H.S. 1942. Large molecules through atomic spectacles, Proceedings of the American Philosophical Society, (85): 1–12.

16. Kabsch, K., and Sander, C. 1983. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure, Biopolymers, (12): 2577–2637.

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

18. 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.

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

20. Baldwin, R.L. 1996. On-pathway versus off-pathway folding intermediates, Fold Des, (1): 1-8.

21. Baldwin, R.L. and Rose, G.D. 1999 Is protein folding hierarchic? II. Folding intermediates and transition states, Trends in Biochemical Sciences, (24): 77-83.

22. Privalov, P.L. 1996. Intermediate states in protein folding, Journal of Molecular Biology, (258): 707-725.

23. Daggett, V. and Fersht, A.R. 2003. Is there a unifying mechanism for protein folding? Trends in Biochemical Sciences, (28): 18-25.

24. Kim, P.S. and Baldwin, R.L. 1990. Intermediates in the folding reactions of small proteins, Annual Review of Biochemistry, (59): 631-660.

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

26. Baldwin, R.L. 1989. How does protein folding get started? Trends in Biochemical Sciences, (14): 291-294.

27. Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins, Journal of the American Chemical Society, (84): 4240-4247.

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

29. Otzen, D.E., Itzhaki, L.S., elMasry, 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.

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

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

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

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

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

35. 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.

36. 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.

37. 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.

38. 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.

39. (a) Pitera, J., and Swope, W. 2003. Understanding folding and design: replica-exchange simulations of the trp-cage miniprotein. Proceedings of the National Academy of Sciences USA, (100):7587–7592. (b) Gallicchio, E., Zhang, L.Y., Levy, R.M. 2004. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the agbnp implicit solvent model. Proteins, (56): 310–321.

40. (a) Duan, Y., and Kollman, P.A. 1998. 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.

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

42. 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.

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

44. Rodriguez, A., Mokoena, P., Corcho, F., Bisetty, K., and Perez, J.J. 2010. Computational study of the free energy landscape of the miniprotein CLN025 in explicit and implicit solvent, Journal of Physical Chemistry B**,** in press.

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

46. Tache, Y., Melchiorri, P., and Negri, L. 1988. Bombesin-like peptides in health and disease, Annals of the New York Academy of Sciences, (547): 174-182.

47. Albers, H.E., Liou, S.Y., Stopa, E.G., and Zoeller, R.T. 1991. Interaction of colocalized neuropeptides: functional significance in the circadian timing system, Journal of Neuroscience, (11): 846-851.

48. Brown, M.R., Carver, K. and Fisher, L.A. 1988. Bombesin: central nervous system actions to affect the autonomic nervous system, Annals of the New York Academy of Sciences. (547): 174.

49. Hill, D.J., and McDonald, T.J. 1992. Mitogenic action of gastrin-releasing polypeptide on isolated epiphyeal growth plate chondrocytes from the ovine fetils, Endocrinology, (130): 2811-2819.

50. Sunday. M.E., Hua, J., Reyes, B., Masui, H., and Torday, J.S. 1993. Anti-bombesin monoclonal antibodies modulate fetal mouse lung growth and maturation in utero and in organ cultures, The Anatomical Record, (236): 25-32.

51. Severi, C., Jensen, R.T., Erspamer, V., D'Arpino, L., Coy, D.H., Torsoli, A., and Delie F.G. 1991. Different receptors mediate the action of bombesin- related peptides on gastric smooth muscle cells, The American Journal of Physiology*,* (260): G683-690.

52. Jensen, R.T., Coy, D.H., Saeed, Z.A., Heinz-Erian, P., Mantey, S. and Gardner, J.D. 1988. Interaction of bombesin and related peptides with receptors on pancreatic acini, Annals of the New York Academy of Sciences, (547): 138-149.

53. Ruff, M., Schiffmann, E., Terranova V., and Pert, C.B., 1985. Neuropeptides are chemoattractants for human tumor cells and monocytes: a possible mechanism for metastasis, Clinical Immunology and Immunopathology, (37): 387-395.

54. DelaFuente, M., DelRio, M., and Hernanz, A. 1993. Stimulation of natural killer and antibody-dependent: cellular cytotoxicity activities in mouse leukocytes by bombesin, gastrin-releasing peptide and neuromedin C: involvement of cyclic AMP, inositol 1,4,5-trisphosphate and protein kinase C, Journal of Neuroimmunology, (48): 143-150.

55. Battey, J., and Wada, E. 1991. Two distinct receptors for mammalian bombesin-like peptides, Trends in Neurosciences, (14): 524-528.

56. Battey, J.F., Way, J.M., Corjay, M.H., Shapira, H., Kusano, K., Harkins, R., Wu, J.M., Slattery, T., Mann, E., and Feldman, R.I. 1991. Molecular cloning of the bombesin/gastrin-releasing peptide receptor from Swiss 3T3 cells, Proceedings of the National Academy of Sciences USA, (88): 395-399.

57. Von Schrenck. T., Heinz-Erian, P., Moran,T., Mantey, S.A., Gardner, J.D., and Jensen, R.T. 1989. Neuromedin B receptor in esophagus: evidence for subtypes of bombesin receptors, American Journal of Physiology (256): G747-758.

58. Benya, R.V., Wada, E., Battey, J.F., Fahti, Z., Wang, L.H., Mantey, S.A., Coy, D.H., and Jensen, R.T. 1992. Neuromedin B receptors retain functional expression when transfected into BALB 3T3 fibroblasts: analysis of binding, kinetics, stoichiometry, modulation by guanine nucleotide binding proteins, signal transduction and comparison with natively expressed receptors, Molecular Pharmacology. (42): 1058-1068.

59. Wang, L.H., Battey, J.F., Wada, E., Lin, J.T., Mantey, S., Coy, D.H., and Jensen, R.T. 1992. Activation of neuromedin B-preferring bombesin receptors on rat glioblastoma C-6 cells increases cellular Ca^{2+} and phosphoinositides, Biochemical Journal, (286): 641-648.

60. Rozenguri, E. 1988. Bombesin-induction of cell proliferation in 3T3 cells. Specific receptors and early signaling events, Annals of the New York Academy of Sciences. (547): 277-292.

61. Jensen, R.T., and Coy, D.H. 1991. Progress in the development of potent bombesin receptor antagonists, Trends in Pharmacological Sciences. (12): 13-19.

62. Marki, W., Brown, M., and Rivier. J.E. 1981. Bombesin analogs: Effects on thermoregulation and glucose metabolism, Peptides Supplement, (2):169-177.

63. Rivier, J.E., and Brown, M.R. 1978. Bombesin, bombesin analogues and related peptides: Effects on thermo regulation, Biochemistry, (17): 1766-1771.

64. Mukai, H., Tanaka, H., Goto, K., Kawai, K., Yamashita, K., and Munekata, E. 1991. Structure-activity relationships of mammalian bombesin-like neuropeptides in the contraction of rat uterus, Neuropeptides, (19): 243-250.

65. Jensen, R.T. 1994. Receptors on pancreatic acinar cells, in: Physiology of the Gastrointestinal Tract, 3rd edn., eds. L.R. Johnson, E.D. Jacobsen, J. Christensen, D.H. Alpers and J.H. Walsh, Raven Press, New York, p. 1377.

66. Smit, W.M.A. and Bode, J.H.G. 1980. Absolute infrared intensities of methane. Dipole moment derivatives and bond charge parameters, Journal of Physical Chemistry, (84): 198- 202.

67. Galabov, B. 1981. A model for parametric analysis of the intensities in the infrared spectra, Journal of Chemical Physics, (74): 1599-1605.

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

69. 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.

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

71. 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.

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

73. 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.

74. 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.

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

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

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

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

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

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

81. 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.

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

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

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

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

86. 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.

87. (a) Filizola, M., Centeno, N.B., Perez, J.J. 1997. Computational study of the conformational domains of peptide T, Journal of Peptide Science, (3): 85-92. (b) Sanbonmatsu, K.Y., and Garcia, A.E. 2002. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics, Proteins: Structure, Function, and Bioinformatics, (46): 225-234.

88. Kirkpatrick, S., Gelatt Jr., C.D., and Vecchi M.P. 1983. Optimization by simulated annealing, Science, (22): 671-681.

89. Xiaomei, L. 2007. Protein folding based on simulated annealing algorithm, ICNC, (4): 256-259, Third International Conference on Natural Computation.

90. Leach, A.R. 1996. Molecular Modelling: Principles and Applications, Addison Longman Limited, England, ISBN 0-470-84310-1, 123-140.

91. 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.

92. 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.

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

94. 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.

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

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

97. 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.

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

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

100. 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.

101. 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.

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

103. 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.

104. 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.

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

106. 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.

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

108. 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.

109. (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.

110. 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.

111. Dehle, F.J., Finley, J.W., Stephens, P.J., and Frisch, M.J. 1995. Functionalist, Journal of Physical Chemistry, (99): 16883-16902.

112. (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.

113. Vijayan, M. 1996. Form and function: X-ray in structural biology, X-ray Crystallography and NMR, (70): 889-898.

114. Honda, S., Akiba, T., Kato, Y.S., Sawada, Y., Sekijima, M., Ishimura, M., Ooishi, A., Watanabe, H., Odahara, T., and Harata, K. 2008. Crystal structure of a ten-amino acid protein, Journal of the American Chemical Society, (130): 15327-15331.

115. Theriault, B.Y. 1991. Structural determination of the vasoactive intestinal peptide by two-dimensional H-NMR spectroscopy, Biopolymers, (3): 459-464.

116. Inooka, H., Endo, S., Kitada, C., Mizuta, E., and Fujino, M. 1992. Pituitary adenylate cyclase activating polypeptide (PACAP) with 27 residues. Conformation determined by 1H NMR and CD spectroscopies and distance geometry in 25% methanol solution, International Journal of Peptide and Protein Research, (40): 456-464.

117. Corcho, F.J., Salvatella, X., Canto, J., Giralt, E., and Perez, J.J. 2007. Structural analysis of substance P using molecular dynamics and NMR spectroscopy, Journal of Peptide Science, (13): 728-741.

118. Wray, V., Kakoschke, C., Nokihara, K., and Naruse, S. 1993. Solution structure of pituitary adenylate cyclase activating polypeptide by nuclear magnetic resonance spectroscopy, Biochemistry, (32): 5832-5841.

119. Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., and Grigorieff, N. 2008. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction, Proceedings of the National Academy of Sciences, (105): 1867-1872.

120. (a) Kandasamy, S.K. and Larson, R.G. 2006. Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch, Biophysical Journal, (90): 2326-2343. (b) Lacape’re, J.J., Pebay-Peyroula E., Neumann, J.M., and Etchebest, C. 2007. Determining membrane protein structures: still a challenge! Trends in Biochemical Sciences, (32): 259-270.

121. Nair, K.S., Jaleel, A., Asmann, Y.W., Short, K.R., and Raghavakaimal, S. 2004. Proteomic research: potential opportunities for clinical and physiological investigators, American Journal of Endocrinology and Metabolism, (286): 863-874.

122. Phillips, D.C. Biomolecular stereodynamics II (ed. Sarma, R.H.) 497 (Adenine press, Guilderland, New York; 1981).

123. Ichiye, T., and Karplus, M. 1983. Fluorescence depolarization of tryptophan in proteins: a molecular dynamics study, Biochemistry, (22): 2884-2893.

124. Dobson, C.M., and Karplus, M. 1986. Internal motion of proteins: Nuclear magnetic resonance measurements and dynamic simulations, Methods in Enzymology, (131): 362-389.

125. Smith, J., Cusack, S., Pezzeca, U., Brooks, B.R., and Karplus, M.J. 1986. Inelastic neutron scattering analysis of low frequency motion in proteins: a normal mode study of the pancreatic trypsin inhibitor, Chemical Physics, (85); 3636-3654.

126. (a) Brűnger, A.T., Brooks, C.L III., and Karplus, M. 1985. Active site dynamics of ribonuclease A, Proc. Natl. acad. Sci. USA, (82); 8458-8462. (b) Frauenfelder, H. et al. 1987. Thermal expansion of a protein, Biochemistry, (26): 254-261.

127. Brűnger, A.T., Kuriyan, J., and Karplus, M. 1987. Crystallographic R-factor Refinement by Molecular Dynamics, Science, (235): 458-460.

128. Nilsson, L., Clore, G.M., Gronenborn, A.M., Brűnger, A.T., and Karplus, M.J. 1986. Structure refinement of oligonucleotides by molecular dynamics with NOE interproton distance restraints: application to 5'd(CGTACG) 2, Molecular Biology, (188): 455-457.

129. Brooks, B.R. et al. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations, Journal of Computational Chemistry, (4): 187-217.

130. Weiner, P.W., and kollman, P.A. 1981. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions, Journal of Computational Chemistry, (2): 287-303.

131. Scott, W.R.P et al. 1999. The GROMOS Biomolecular Simulation Program Package, Journal of Physical Chemistry A, (103): 3596-3607.

132. (a) Tuckerman, M.E., and Martyna, G.J. 2000. Understanding modern molecular dynamics: Techniques and applications, Journal of Physical Chemistry, (104): 159-178. (b) van Gunsteren, W.F., Weiner, P.K., and Wilkinson, A.J. 1983. Theoretical and Experimental Applications Vol. 2 (ESCOM, Leiden. (c) Becker, O.M., MacKerell, A.D. Jr, Roux, B., and Watanabe, M. 2001. Computational Biochemistry and Biophysics (Marcel Dekker, New York).

133. Vitkup, D., Ringe, D., Petsko, G.A., and Karplus, M. 2000. Solvent mobility and the protein 'glass' transition, Nature Structure & Molecular Biology, 34-38.

134. Hagen, S.J., Hofrichter, J., and Eatson, W.A. 1995. Protein Reaction kinetics in a room-temperature glass, Science, (269): 959-962.

135. (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.

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

137. Buck, M. 1998. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins, Quarterly Review of Biophysics, (31): 297-355.

138. Perez, J.J., Villar, H.O., and Arteca, G.A. 1994. Distribution of conformational energy minima in molecules with multiple torsional degrees of freedom, Journal of Physical Chemistry, (98): 2318-2324.

139. 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.

140. Anastasi, A., Erspamer, V., and Bucci, M. 1971. Isolation and structure of bombesin and alytesin, 2 analogous active peptides from skin of European amphibians bombina and alytes, Experientia, (27): 166-167.

141. Erspamer, V., Improta, G., Melchiorri, P., and Sopranri, N. 1974. Evidence of cholecystokinin release by bombesin in dog, British Journal of Pharmacology, (52): 227-232.

142. Bruzzone, R. 1989. Mechanism of action of bombesin on amylase secretion - evidence for a Ca-2+-independent pathway, European Journal of Biochemistry, (179): 323-331.

143. Erspamer, V., and Melchiorri, P. 1980. Active polypeptides - from amphibian skin to gastrointestinal-tract and brain of mammals, Trends in Pharmacological Sciences, (1): 391-395.

144. Mayer, E.A., Reeve Jr., J. R., Khawaja, S., Chew, P., Elashoff, J., Clark, B., and Walsh, J.H. 1986. Potency of natural and synthetic canine gastrin-releasing decapeptide on canine antral muscle, American Journal of Physiology, (250): G581-G587.

145. Herrmann, C., Cuber, J.C., Dakka, T., Bernard, C., and Chayvialle, J.A. 1991. Bombesin potentiates taurocholic acid-induced neurotensin release in rats, Endocrinology, (128): 2853-2857.

146. Alexander, R.W., Upp, J.R., Poston, G.J., Gupta, V., Townsend Jr., C.M., and Thompson, J.C. 1988. Effects of bombesin on growth of human small cell lung-carcinoma in vivo, Cancer Research, (48): 1439-1441.

147. Sunday, M.E., Choi, N., Spindel, E.R., Chin, W.W., and Mark, E.J. 1991. Gastrin-releasing peptide gene-expression in small-cell and large cell undifferentiated lung carcinomas, Human Pathology, (22): 1030-1039.

148. Mahmoud, S., Staley, J., Taylor, J., Bogden, A., Moreau, J. –P., Coy, D., Avis, I., Cuttitta, F., Mulshine, J.L., and Moody, T.W. 1991. Bombesin analogs inhibit growth of small-cell lung-cancer in vitro and in vivo, Cancer Research, (51): 1798-1802.

149. Qin, Y., Ertl, T., Cai, R-Z., Halmos, G., and Schally, A.V. 1994. Inhibitory effect of bombesin receptor antagonist rc-3095 on the growth of human pancreatic-cancer cells in-vivo and in-vitro, Cancer Research, (54): 1035-1041.

150. Merali, Z., McIntosh, J., and Anisman, H., 1999. Role of bombesin-related peptides in the control of food intake, Neuropeptides, (33): 376-386.

151. Yamada, K., Wada, E., Santo-Yamada, Y., and Wada, K. 2002. Bombesin and its family of peptides: prospects for the treatment of obesity, European Journal of Pharmacology, (440): 281-290.

152. Carver, J.A. The conformation of bombesin in solution as determined by two-dimensional h-1-nmr techniques, European Journal of Biochemistry, (168): 193-199.

153. Carver, J.A. A two-dimensional H-1-NMR study of the solution conformation of gastrin releasing peptide, Biochemical and Biophysical Research Communications, (150): 552-560.

154. DiBello, C., Gozzini, L., Tonellato, M., Corradini, M.G., D'Auria, G., Paolillo, L., and Trivellone, E. 1988. 500 MHz NMR characterization of synthetic bombesin and related peptides in DMSO-D6 by two-dimensional techniques, FEBS Letters, (237): 85 -90.

155. Carver, J.A., and Collins, J.G. 1990. NMR identification of a partial helical conformation for bombesin in solution, European Journal of Biochemistry, (187): 645-650.

156. DÌaz, M.D., Fioroni, M., Burger, K., and Berger, S. 2002. Evidence of complete hydrophobic coating of bombesin by trifluoroethanol in aqueous solution: An NMR spectroscopic and molecular dynamics study, Chemistry- A European Journal, (80): 1663-1669.

157. Erne, D., and Schwyzer, R. 1987. Membrane-structure of bombesin studied by infrared-spectroscopy - prediction of membrane interactions of gastrin-releasing peptide, neuromedin-B, and neuromedin-C, Biochemistry, (26): 6316-6319.

158. Cavatorta, P., Farruggia, G., Masotti, L., Sartor, G., and Szabo, A.G. 1986. Conformational flexibility of the hormonal peptide bombesin and its interaction with lipids, Biochemical and Biophysical Research Communications, (141): 99-105.

159. Luo, P., and Baldwin, R.L. 1997. Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water, Biochemistry, (36): 8413-8421.

160. Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics, Journal of American Chemical Society, (112): 6127-6129.

161. Chen, J., Brooks III, C.L., and Khandogin, J. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations, Current Opinion in Structural Biology, (18); 140–148.

162. Kollman, P.A., Dixon, R., Cornell, W., Fox, T., Chipot, C., and Pohorille, A. 1997. The development/application of a 'minimalist' organic/biochemical molecular mechanics force field using a combination of ab initio calculations and experimental data, Computer Simulation of Biomolecular Systems, (3): 83–96.

163. Zhou, R. 2003. Free energy landscape of protein folding in water: explicit vs. implicit solvent, Proteins: Structure, Function, and Bioinformatics, (53): 148-161.

164. Roe, D.R., Okur, A., Wickstrom, L., Hornak, V., and Simmerling, C. 2007. Secondary Structure Bias in Generalized Born Solvent Models: Comparison of Conformational Ensembles and Free Energy of Solvent Polarization from Explicit and Implicit Solvation, Journal of Physical Chemistry B, (111): 1846-1857.

165. Shell, M.S., Ritterson, R., and Dill, K.A. 2008. A Test on Peptide Stability of AMBER Force Fields with Implicit Solvation, Journal of Physical Chemistry B, (112): 6878–6886.

166. Singh, U.C., and Kollman, P.A. 1984. An approach to computing electrostatic charges for molecules, Journal of Computational Chemistry, (5); 129-145.

167. Zimmerman, S.S., Pottle, M.S., Nemethy, G., and Scheraga, H.A. 1977. Conformational-analysis of 20 naturally occurring amino-acid residues using ECEPP, Macromolecules, (10): 1-9.

168. Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., and Wuthrich. K. 1998. Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids, Journal of molecular biology, (280): 933-952.

169. Gasmi, G., Singer, A., Forman-Kay, J., Sarkar. B. 1997. NMR structure of neuromedin C, a neurotransmitter with an amino terminal Cu^{II}-, Ni^{II}-binding (ATCUN) motif, Journal of Peptide Research, (49): 500-509.

170. Rosta, E., Buchete, N.-V., and Hummer, G. 2009. Thermostat Artifacts in Replica Exchange Molecular Dynamics Simulations, Journal of Chemical Theory and Computation, (5): 1393–1399.

171. Cooke, B., and Schmidler, S.C. 2008. Preserving the Boltzmann ensemble in replica-exchange molecular Dynamics, Journal of Chemical Physics (129): 164112-164129.

172. Prasad, S., Mathur, A., Gupta, N., Jaggi, M., Singh, A.T., Rajendran, P., Sanna, V. K., and Mukherjee, R. 2007. Bombesin analogs containing *α*-amino-isobutyric acid with potent anticancer activity, Journal of Peptide Science, (13): 54-62.

173. Jensen, R.T., Battey, J.F., Spindel, E.R., Benya, R.V. 2008. International Union of Pharmacology. LXVIII. Mammalian bombesin receptors: nomenclature, distribution, pharmacology, signaling, and functions in normal and disease states, Pharmacological Reviews, (60): 1–42.

174. Minamino, N., Kangawa, K., and Matsuo, H. 1983. Neuromedin B: a novel bombesin like peptide identified in porcine spinal cord, Biochemical and Biophysical Reserach Communications, (114): 541–548.

175. Spindel, E.R., Giladi, E., Segerson, T.P., Nagalla, S. 1993. Bombesin-like peptides: of ligands and receptors, Recent Progress in Hormone Research, (48): 365-391.

176. Gonzalez, N., Moody, T.W., Igarashi, H., Ito, T., and Jensen, R.T. 2008. Bombesin-related peptides and their receptors: recent advances in their role in physiology and disease states, Current Opinion in Endocrinology, Diabetes & Obesity, (15): 58-64.

177. Shin, C., Mok, K.H., Han, J.H., Ahn, J-H., and Lim, Y. 2006. Conformational analysis in solution of gastrin releasing peptide, Biochemical and Biophysical Reserach Communications, (350): 120-124.

178. Sankararamakrishnan, R. 2006. Recognition of GPCRs by Peptide Ligands and Membrane Compartments theory: Structural Studies of Endogenous Peptide Hormones in Membrane Environment, Bioscience Reports, (26): 131-158.

179. Gante, J. 1994**.** Peptidomimetics - Tailored Enzyme Inhibitors, Angewandte Chemie International Edition, (33): 1699-1720.

180. Polverini, E., Neyroz, P., Fariselli, P., Casadio, R., and Masotti, L. 1995. The Effect of Membranes on the Conformation of Neuromedin C, Biochemical and Biophysical Research Communications, (214): 663-668.

181. (a) Bal, W., Jerzowka-Bojczuk, M., and Kasprzak, K.S. 1997. Ni(II) specifically cleaves the C-terminal tail of the major variant of histone H_2A and forms an oxidative damage-2 mediating complex with the cleaved-off octapeptide, Chemical Research in Toxicology, (10): 906–914. (b) Lu, Y., Berry, S.M., and Pfister, T.D. 2001. Engineering Novel Metalloproteins: Design of Metal-Binding Sites into Native Protein Scaffolds, Chemical Review, (101): 3047–3080.

182. (a) 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. (b) 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 (c) 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. (d) 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.

183. (a) Pitera, J., and Swope, W. 2003. Understanding folding and design: replica-exchange simulations of the trp-cage miniprotein, Proceedings of the National Academy of Sciences. U.S.A., (100):7587–7592. (b) Gallicchio, E., Zhang, L.Y., Levy, R.M. 2004. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the agbnp implicit solvent model, Proteins, (56): 310–321.

184. Terada, T., and Shimizu, K. 2008. A comparison of generalized Born methods in folding simulations, Chemical Physics Letters, (460): 295–299.

185. (a) Wu, X., and Brooks, B.R. 2003. Self-guided Langevin dynamics simulation method, Chemical Physics Letters, (381): 512-518. (b) Andersen, H.C. 1980. Molecular dynamics simulations at constant pressure and/or temperature, Journal of Chemical Physics, (72): 2384-2393.

186. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Imprey, R.W., and Klein, M.L. 1983. Comparison of simple potential functions for simulating liquid water, Journal of Chemical Physics, (79): 926–935.

187. Havel, T.F. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonances, Review Article, Progress in Biophysics and Molecular Biology, (56): 43-78.

188. Erspamer, V. 1980. Comprehensive Endocrinology (Glass, G.B.J., Ed.). 343-361, Raven Press, New York.

189. Ghatei, M.A., Jung, R.T., Stevenson, J.C., Hillyard, C.J., Adrian, T.E., Lee, Y.C., Christofides, N.D., Sarson, D.L., Mashiter, K., MacIntyre, I., and Bloom, S.R. 1982. Bombesin - Action on gut hormones and calcium in man, The Journal of Clinical Endocrinology and Metabolism, (54): 980-985.

190. Erspamer, V. 1988. Discovery, isolation and characterization of bombesin- like peptides, Annals of the New York Academy of Science, (547): 3-9.

191. (a) Cuttitta, F., Carney, D.N., Mulshine, J., Moody, T.W., Fedorko, J., Fischler, A., and Minna, J.D. 1985. Bombesin-like peptides can function as autocrine growth factors in human small-cell lung cancer, Nature, (316): 823-826. (b) Corps, A.N., Rees, L.H., and Brown, K.D. 1985. A peptide that inhibits the mitogenic stimulation of Swiss 3T3 cells by bombesin or vasopressin, Biochemical Journal, (231): 781-784. (c) Moody, T.W., Carney, D.N., Cuttita, F., Quattrocchi, K., and Minna, J.D. 1985. High affinity receptors for bombesin/GRP-like peptides on human small cell lung cancer, Life Sciences, (37): 105-113.

192. Ohki-Hamazaki, H. 2000. Neuromedin B, Progress in Neurobiology. (62): 297-312.

193. Wada, E., Way, J., Shapira, H., Kusano, K., Lebacq-Verheyden, A.M., Coy, D., Jensen, R., and Battey, J. 1991. cDNA cloning, characterization, and brain region-specific expression of a neuromedin-B-preferring bombesin receptor, Neuron, (6): 421-430.

194. Corjay, M.H., Dobrzanski, D.J., Way, J.M., Viallet, J., Shapira, H., Worland, P., Sausville, E.A., and Battey, J.F.J. 1991. Two distinct homhcsin receptor subtypes are expressed and functional in human lung carcinoma cells, Journal of Biological Chemistry, (266): 18771-18779.

195. Gorbulev, V., Akhundova, A., Biichner, H., and Fahrenholz. F. 1992. Molecular cloning of a new bombesin receptor subtype expressed in uterus during pregnancy, European Journal of Biochemistry, (208): 405-410.

196. Mantey, S., Weber, C., Sainz, E., Akeson, M., Ryan, R., Pradhan, T., Searles, R., Spindel, E., Battey, J.F., Coy, D.H., and Jensen, R.T. 1997. Discovery of a high affinity radioligand for the human orphan receptor, bombesin receptor subtype 3, which demonstrates it has a unique pharmacology compared to other mammalian bombesin receptors, The Journal of Biological Chemistry, (272): 26062-26071.

197. Lee, S., and Kim. Y. 1999. Solution structure of neuromedin B by ^{1}H nuclear magnetic resonance spectroscopy, FEBS letters, (460): 263-269.

198. Erneand, D., and Schwyzer, R. 1987. Membrane structure of bombesin studied by infrared spectroscopy. Prediction of membrane interactions of gastrin-releasing peptide, neuromedin B, and neuromedin C, Biochemistry, (26): 6316-6319.

199. Polverini, E., Casadio, R., Neyroz, P., and Masotti, L. 1998. Conformational changes of Neuromedin B and Delta sleep inducing peptide induced by their interaction with lipid membranes as revealed by spectroscopic techniques and molecular dynamics simulations, Archives of Biochemistry and Biophysics, (349): 225-235.

200. Horwell, D.C., Howson, W., Naylor, D., Osborne, S., Pinnock, R.D., Ratclii, G.S., and Suman-Chauman, N. 1996. Alanine scan and N-methyl amide derivatives of Ac-bombesin[7-14]. Development of a proposed binding conformation at the neuromedin B (NMB) and gastrin releasing peptide (GRP) receptors, International Journal of Peptide and Protein Research, (48): 522-531.

201. Sainz, E., Akeson, M., Mantey, S.A., Jensen, R.T., and Battey, J.F. 1998. Four Amino Acid Residues Are Critical for High Affinity Binding of Neuromedin B to the Neuromedin B Receptor, Journal of Biological Chemistry, (273): 15927-15932.

202. (a) Erne, D., Sargent, D.F., and Schwyzer, R. 1985. Preferred conformation, orientation and accumulation of dynorphin A(1–13), Biochemistry, (24): 4261-4263. (b) Kaiser, E.T., and Kezdy. F.J. 1987. Peptides with affinity for membranes, Annual Review of Biophysics and Biophysical Chemistry, (16): 561–581. (c) Dyson, H.J., and Wright. P.E. 1991. Peptide conformation and protein folding, Annual Review of Biophysics and Biophysical Chemistry, (20): 519-538.

203. Zhong, L., and Jr. Johnson, W.C. 1992. Environment Affects Amino Acid Preference for Secondary Structure, Proceedings of the National Academy of Sciences USA, (89): 4462-4465.

204. (a) Coy, D.H., Heinz-Erian, P., Jiang, N.-Y., Sasaki, Y., Taylor, J., Moureau, J.-P., Wolfrey, W.T., Gardner, J.D., and Jensen, R.T. 1988. Probing peptide backbone function in bombesin: a reduced peptide bond analogue with potent and specific receptor antagonist activity, Journal of Biological Chemistry, (263): 5056-5060. (b) Rivier, J.E., and Brown, M.R. 1978. Bombesin, bombesin analogs, and related peptides: Effects on thermoregulation, Biochemistry, (17): 1766-1771. (c) Schwyzer, R. 1986. Molecular mechanism of opioid receptor selection, Biochemistry, (25): 6335-6342.

205. (a) Dominy, B., and Brooks, C. 1999. Development of a Generalized Born Model Parameterization for Proteins and Nucleic Acids, Journal of Physical Chemistry B, (103): 3765-3773 (b) Calimet, N., Schaefer, M., and Simonson, T. 2001. Continuum solvent molecular dynamics study of flexibility in interleukin-8, Journal of Molecular Graphics and Modelling (19): 136-145.

206. (a) M. Feig, W. Im, C.L. Brooks 3rd, 2004. Implicit Solvation Based on Generalized Born Theory in Different Dielectric Environment, Journal of Chemical Physics, (120): 903-911. (b) G. Sigalov, P. Scheffel, A. Onufriev, 2005. Incorporating variable dielectric environments into the generalized Born model, Journal of Chemical Physics, (122): 094511-094515.

207. Sitkoff, D., Sharp, K.A., and Honig, B. 1994. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models, Journal of Physical Chemistry, (98): 1978-1988.

208. Tsui, V., and Case. D.A. 2001. Theory and applications of the generalized Born solvation model in macromolecular simulation, Biopolymers, (56): 275-291.

209. Ghose, A.K., Crippen, G.M., Revankar, G.R., Smee, D.F., McKernan, P.A., and Robins, R.K. 1989. Analysis of the in Vitro Antiviral Activity of Certain Ribonucleosides

against Parainfluenza Virus Using a Novel Computer Aided Receptor Modeling Procedure, Journal of Medicinal Chemistry, (32): 746-756.

210. Jörgensen, W.L. 1991. Rusting of the lock and key model for protein-ligand binding, Science, (254): 954-955.

211. Corcho, F.J., Filizola, M. and Perez, J.J. 1999. Assessment of the bioactive conformation of the farnesyltransferase protein binding recognition motif by computational methods, Journal of Biomolecular Structure & Dynamics, (5): 1043-1052.

212. Filizola, M., Centeno, N.B., Farina, M.C. and Perez, J.J. 1998. Conformational analysis of the highly potent bradykinin antagonist Hoe-140 by means of two different computational methods, Journal of Biomolecular Structure & Dynamics, (15): 639-652.

213. (a) Kleiweg, P., Nerbonne, J., and Bosveld, L. 2004. Geographical Projection of cluster composites. In: A. Blackwell, K. Marriott and A. Shimojima (Eds.). (b) Jain, Anil, K., Murty, M.N., and Flynn, P.J. 1999. Data Clustering: A Review. ACM Computing Surveys 31(3): 264-323.

Publications arising from this work:

1. Computational Study of Conformational Profile of Bombesin. **Parul Sharma,** Parvesh Singh, Krishna Bisetty, Francesc J. Corcho, Juan J. Perez. Journal of Molecular Graphics and Modelling, **2010**, 29, 581-590.

2. Conformational space search of Neuromedin C using replica exchange molecular dynamics and molecular dynamics. **Parul Sharma**, Parvesh Singh, Krishna Bisetty, Alex Rodriguez and Juan J. Perez. Journal of Peptide Science, **2011**, 17, 174-183.

3. A Computational study of Neuromedin B. **Parul Sharma**, Parvesh Singh, Krishna Bisetty and Juan. J. Perez. (Under preparation).