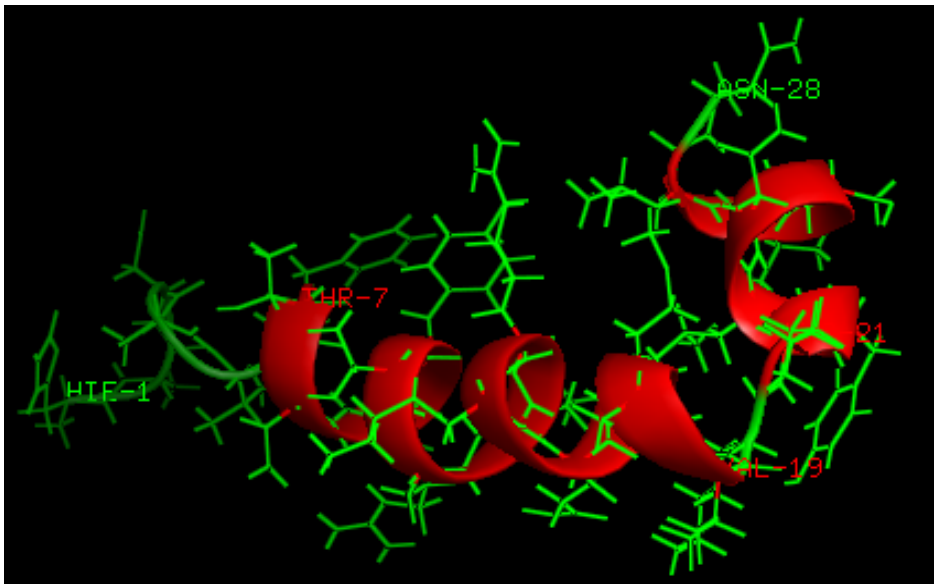
COMPUTATIONAL STUDIES OF THE FOLDING PATTERNS OF SMALL AND MEDIUM-SIZE POLYPEPTIDES

M P MOKOENA



DURBAN UNIVERSITY OF TECHNOLOGY

February 2010

COMPUTATIONAL STUDIES OF THE FOLDING PATTERNS OF SMALL AND MEDIUM-SIZE POLYPEPTIDES

PAUL MOKOENA

Submitted in partial fulfilment for the degree of Doctor of Technology: Biotechnology

February 2010

DECLARATION

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

……..

ABSTRACT

This study involved a series of molecular dynamics (MD) simulations applied to case studies of small and medium-size polypeptides to assess the thermodynamics of their folding characteristics. Peptide folding is a complex and vital phenomenon taking place in all living systems. Bioactive conformational structures of folded peptides need to be well characterized before using them in computer-aided drug design. The computational procedure was validated on the 10-residue long chignolin-like synthetic mini-protein (CLN025). For this peptide, replica exchange molecular dynamics (REMD) calculations were carried out in explicit and implicit solvents using the generalized Born (GB)/surface area (SA) approximation with different sets of force field parameters. Following this validation procedure, case studies of the folding conformations of peptides of different lengths including the 5-residue met-enkephalin, the 27-residue pituitary adenylate-activating polypeptide 27(PACAP27) and the 28-residue vasoactive intestinal peptide (VIP) were undertaken. The latter two peptides are multifunctional hormones that mediate diverse biological functions, such as the cell cycle, cardiac muscle relaxation, immune response, septic shock, bone metabolism, and endocrine function.

Results obtained indicate that when explicit water, methanol and DMSO solvents were used, it appeared that methanol (MeOH) and dimethylsulphoxide (DMSO) afforded met-enkephalin the ability to form more intra-hydrogen bonds than water, producing type I and type III β-turn structures; thus enhancing the helical conformation of the peptide. MD trajectories of longer polypeptides (VIP and PACAP27) were also populated with type I and type III β-turns, which occurred consecutively; with α- and 3_{10}-helices occurring from the middle of each peptide towards the C-terminal. Characterization of implicit solvent results, reveal that these simulations have been able to reproduce the same type of conformers obtained by experimental NMR studies published in literature, which structurally resemble the native conformation of the bioactive peptides. These conformational structures will be applied as lead agents in computer-aided drug design. One of the major achievements of this study is the ability to optimize and validate the force field parameter sets to describe the thermodynamic properties of peptide systems in an unbiased manner, a non-trivial task for even the smallest of peptides. These findings re-affirm the notion that computational methods have matured enough to model dynamic biological phenomena such as peptide folding, a feat previously thought to be impossible.

ACKNOWLEDGEMENTS

I would like to give thanks to the omniscient, almighty God for His wisdom, revelation and strength, and to my family for allowing me time to complete this research project. Their support and encouragement throughout this research project is highly appreciated.

I would also like to sincerely thank my supervisors; Prof K. Bisetty (Department of Chemistry, DUT), Prof J.J. Perez and Dr F.J. Corcho (Molecular Engineering Laboratory, UPC, Barcelona) for their time and mentorship during this work, especially, their valuable insight in supervising the project. Special thanks also to Alex Rodriguez and Josep Canto from Barcelona for their assistance with technical analyses. Finally, I wish to thank my fellow researchers, Dr Parvesh Singh and Parul Sharma for sharing thoughts and their emotional support during the project.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ALPB : Analytical Linearized Poisson-Boltzman

BBB : Blood-brain barrier

CA : Cluster analysis

CLN025 : Chignolin-like peptide 025

CNS : Central nervous System

crd : Coordinates file

DMSO : Dimethyl sulfoxide

E_{bend} : Angle bending energy

E_{elec} : Electrostatic energy

E_{str} :Bond-stretching energy

E_{tors} :Torsional angle energy

E_{tot} :Total energy

E_{vdw} : Van der Waals energy

ff : Force field

ff99 : Force field for proteins and general organic molecules developed by reparametizing ff94.

ff99SP : Modification of ff99 by changing the φ torsion angles.

GB : Generalized Born

GB^{HCT} : The pairwise generalized Born model developed by Hawkins, Cramer and Truhlar

GB^{OBC} : The generalized Born model developed by Onufriev, Bashford and Case

HB : Hydrogen bond

IR spectra : Infra-red spectra

IT : Information technology

Leu : Leucine

MC : Monte-Carlo simulations

MD : Molecular Dynamics simulations

NMR : Nuclear Magnetic Resonance spectroscopy

NOEs : Nuclear Overhauser Effect Intensities

Met : Methionine

PACAP27 : pituitary adenylate cyaclase-activating polypeptide

PCA : Principal component analysis

PDB : Protein Data Bank

REMD : Replica exchange molecular dynamics

RMSD : Root mean square deviation

TFE : Trifluoroethanol

top : Topology file

VIP : Vasoactive intestinal peptide

2D : two-dimensional

3D : three-dimensional

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Molecular Modelling of Peptides

Most cellular functions such as signal transduction, immune response, and catalysis of metabolic reactions are mediated by functional peptides and proteins. Peptides and proteins are implicated in regulating cellular activities both under normal and disease conditions. With regard to enzymes, natural ligands are the substrates which have to bind at the active site before being chemically processed into products. Catalytic reactions can be suppressed by competitive inhibitors which also have the affinity for the same binding site as the substrate. On the contrary, receptors (e.g. cell surface proteins) bind ligands without chemically modifying them. Instead, the binding induces a conformational change in the receptor protein that can trigger a chemical reaction of a substrate bound somewhere else on the same protein or affect the binding affinity of a second molecule that interacts with the receptor. Ligands that activate receptors are called agonists, while competitive inhibitors of these ligands are called antagonists **[1]**.

Peptidic agonists need to be folded to their specific and precise three-dimensional conformations prior to attaining their biological activity. Since many peptides are biological hormones, they can be synthesized in one cellular compartment and have to be transported to the target sites, where they initiate the cascading signal by interacting with specific receptors **[2, 3, 4]**. However, understanding the thermodynamics and mechanisms of peptide and protein folding remains one of the

current challenges in modern molecular biology **[2, 3]**. Despite numerous contributions from different researchers, protein folding is still not adequately understood **[4]**. 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 **[5, 6, 7, 83]**.

Understanding the process of protein folding is also of great importance in biotechnology, since the expression of recombinant proteins often results in inclusion bodies consisting of misfolded and consequently inactive proteins. In addition, the three-dimensional (3D) structures of proteins are linked to their biological functions. Therefore, determination of accurate protein structure is important in elucidating disease conditions, understanding biological functions of proteins, and in peptide-based drug design **[2, 4]**. Moreover, to obtain a good ligand (drug) for a protein or DNA surface receptor, one has to study the structure and function of natural agonists and antagonist, or the surface topology of the binding site on the macromolecule. Using structural information for drug discovery is referred to as rational drug design and it makes use of the concept of chemical similarity **[1]**.

With the successful deciphering of the human genome, the emphasis is now on determining the structure and functions of the proteins coded for by the different genes. Since the pioneering work whereby the structure of myoglobin was determined by using X-ray crystallography **[8]**, 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 **[38, 67, 69, 72, 78]**. More recently, cryomicroscopy has been used to decipher the features of some protein complexes **[9]**. Proteomics, which is the study of a collection of proteins within a cell or entire organism, has advanced beyond merely making catalogues of newly found proteins and now also focuses on the comprehensive and detailed characterization of these proteins. This is not an easy task because proteins can undergo many post-translational modifications that affect structure and function. Understanding factors regulating peptide folding is one of the key areas of research **[10]**.

The contribution of these techniques is reflected by the increasing number of structures available 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 in BLUE. 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 **[11, 12, 13]**. Prediction of the 3D structure of a protein using first principles is still in its infancy. Although important advances have been achieved in the past regarding the process of protein folding, the tools available still need upgrading to produce enough accurate structures. Areas for improvement include the development of more robust force fields, and the use of appropriate solvent models and solute charges. Computational methods such as molecular dynamics (MD), replica exchange molecular dynamics (REMD) and simulated annealing (SA) have contributed a lot 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 **[4]**.

Peptides are of interest to structural biologists and computational chemists because they present a more tractable system than proteins whilst they have the unique properties of proteins **[2]**. Therefore, peptides offer an opportunity to bridge the gap between theoretical and experimental understanding of protein folding. With the advent of computational chemistry and structural biology, increase in computational power, and the development and implementation of simulation algorithms, it has recently become possible to simulate folding of peptides and small proteins starting from their extended conformations **[2, 4, 14]**. There is still a paucity of accurate structures of peptides and proteins determined by computational methods from first principles due to the fact that simulations in explicit solvents are computationally demanding, whilst they are quite desirable. It has been observed that simulations with the solvent explicitly expressed tend to give superior information to implicit simulations in many instances **[4]**. As a result, choices in modes of computational simulations are informed by time available for the simulation and whether the algorithm used in the force fields would allow for an exhaustive exploration of the energy landscape, thus resulting in a trajectory that is populated by structures with conformational features comparable to the average experimental NMR structure for a specific peptide.

Peptides and proteins evolved to play a major role in co-ordinating the activities of living systems. To perform this critical role, they are endowed with a great deal of

flexibility. Since peptides and proteins are both composed of sequences of amino acids, there are similarities in their physical behaviour, with most proteins characterized by being long heteropolymers of amino acids. However, peptides exhibit some differential features in comparison to proteins. Peptides are known to be highly flexible molecules, without a definite structure in solution at room temperature, exhibiting a complex conformational profile that is solvent dependent **[4]**. Therefore, in NMR studies, peptides are characterized as exhibiting a random coil type of structure in water, whereas in structuring solvents (e.g methanol and trifluoroacetic acid) specific structural features are identifiable. Likewise, peptides exhibit different conformations in crystal forms owing to the flexible nature associated at a microscopic level with a dynamic exchange between conformations in a complex potential energy surface that has multiple local minima **[4, 15]**. Traditional experimental methods reflect an average structure that corresponds to a blurred image since it is the result of the superposition of different co-existing structures of the compound that is being investigated. However, there is a newly developed ROESY NMR technique that allows one to obtain much better 3D structural information for smaller peptides **[84]**. On the other hand, computational methods enable the exploration of the different possible structures, and by taking into account their potential energies, the conformation of the biologically active structure can be identified with ease **[15]**.

In the pharmaceutical drug industry, it has become critical to substantially describe the structures of potential drug candidates. Where this information was found lacking, it has resulted in drug candidates not making it beyond phase I of the clinical

trial stage. Drug interactions resulting in phenomena such as side effects, toxicity, etc. depend very much on the detailed structure of the drug candidate compounds **[16]**. Since the 3D structure of peptides is informed by the amino acid sequence and modulated by the solvent environment, computational methods represent a useful complementary approach to understand the conformational profile of peptides, with a possibility of assigning more appropriate functions to newly discovered peptides and proteins **[16]**.

When the structures of peptides and proteins are well described, it becomes much easier to assign their functions, as well as to study their mechanism of action in interactions with their receptors and their interactions with other biological molecules such as DNA, RNA, and other proteins in living systems. This will contribute a lot in understanding the role of peptide and protein folding in biology, and their application as lead compounds in computer-aided drug design studies.

1.2 Aims and Objectives

The aims of this study were to:

- use computational methods to assess the conformational profiles of peptides.
- investigate the effect of solvent environment on the folding patterns of peptides.
- compare different force fields and their use in computational approaches for exploring the energy landscape of peptides.

The **objectives** of this study were to:

- investigate the performance of different implicit solvent models on the conformational features of a model ten-residue chignoline-like peptide (CLN025) using the explicit solvent calculation as reference. Sampling of the configurational phase space was carried out using replica exchange molecular dynamics method (REMD).
- study the conformational features of a five-residue peptide (met-enkephalin) using MD.
- study the conformational profiles of a 27-ersidue peptide (PACAP27) and a 28-residue peptide (VIP) by MD using an implicit solvent models.

CHAPTER 2

THEORETICAL METHODS AND CONCEPTS

Molecular modelling can be described as a collective term that refers to theoretical methods and computational techniques to model or mimic the behaviour of molecules. Alternatively, it can be described as a science of representing molecular structures numerically and simulating their behaviour with the equations of quantum mechanics and classical physics **[17]**. These techniques are used in the fields of computational chemistry, structural biology and materials science for studying molecular systems ranging from small chemical systems to macromolecules and material assemblies **[18, 19]**.

The theoretical methods used in the molecular modelling arena include automatic generation, analysis of 3D databases, building of protein models by techniques based on sequence homology, diversity analysis, and docking of ligands to their receptors **[20]**. Molecular modelling is a field concerned with the application of different strategies to model or deduce information of a system at the atomic level. This field includes all methodologies used in computational chemistry, such as computation of the energy of a molecular system, energy minimization, Monte Carlo (MC) methods or molecular dynamics (MD) **[20]**. In this regard, molecular modelling tools have been used to obtain the three-dimensional structure and the conformational flexibility of biological molecules **[21]**. Combined or hybrid QM/MM techniques have been designed that enable the modeling of reactive biomolecular systems at a reasonable computational effort while providing the necessary accuracy [85].

Identification of biomolecular moieties involved in the interaction with a specific receptor allows for the understanding of the molecular mechanism responsible for its specific biological activity. Subsequently, this knowledge is used in designing new biologically active molecules that can be successfully used as drugs. However, simulation accuracy is limited to the accuracy of the constructed models. It is therefore advisable to compare computational simulations with experimental results to confirm the accuracy of models and to modify the models if necessary, in order to obtain better representations of the system **[20]**.

2.1 Molecular Mechanics

In molecular mechanics, the electrons and protons of the atoms are not explicitly included in the calculations. Molecular mechanics considers a molecule to be a collection of masses interacting with each other through harmonic forces. Thus, the atoms in molecules are treated as balls of different sizes joined together by springs of variable strength and equilibrium distances (bonds). This simplification allows for using molecular mechanics as a fast computational model that can be applied to molecules of any size. The total energy of a system is minimized with respect to the atomic coordinates, and is made up of a sum of different contributions that calculate the deviations from equilibrium values of bond lengths, angles and torsions plus non-bonded interactions [**20, 22, 23]**:

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

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

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

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

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

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

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

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

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

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

Where ε is the dielectric constant, and Q_1 and Q_2 are the atomic charges of interacting atoms and r_{ij} is the interatomic distance. Modest level quantum mechanical methods are adequate in accuracy and efficiency for describing intramolecular energy surfaces. Thus, the equilibrium values of bond lengths and bond angles are derived from the force constants used in the potential energy function defined in the force field, and define a set known as force field parameters **[20, 22]**. Deviations from any of the equilibrium values inevitably result in increasing total energy of the molecule. As a result, the total energy of the system is taken as a measure of intramolecular strain relative to a hypothetical molecule with equilibrium values **[20, 22]**. 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 **[20, 24, 25]**.

2.2 Use of Charges and Solvents

Molecular mechanics calculations are 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 conformations

are influenced by strong electrostatic interactions. Force fields try to maximise 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, such as ε=80 for water.

Equation (4) shows that 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 far away from each other, simulating the effect of displacement of solvent molecules in the path of a ligand molecule approaching a macromolecular surface.

The analytic generalized Born (GB) model is a method for computing charge-charge interactions. It efficiently describes electrostatics of molecules in a water environment. It represents the solvent (water) implicitly as a continuum with the dielectric properties of water, and includes the charge screening effects of salt. There are several versions of the GB model implemented in AMBER 9.0 **[26]**. In addition to the charges of the interacting particles, the algorithm also takes into account both the dielectric constant of the solvent, and the smoothing function which depends on atomic radii and interatomic distances of the charged particles. Some of the useful features of GB models include:

- The computational cost associated with the use of these models in MD simulations is much smaller than the cost of representing water explicitly.

- The model describes an instantaneous solvent dielectric response which eliminates the need for lengthy equilibration of water necessary in explicit water simulations.
- Due to the absence of viscocity associated with explicit water environment the molecules of interest can explore the conformational space much faster.
- Estimating energies of solvated structures is much easier than with explicit water models.

2.3 Energy-Minimization Procedures

Energy minimization methods can be divided into different classes depending on the order of derivative used for locating a minimum on the energy surface. Zero order methods are those that only use the energy function to identify regions of low energy through a grid search method. Within first-derivative techniques, there are several techniques including the steepest descent method or the conjugate gradient method, which make 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 first-derivative methods have been used **[20]**.

2.3.1 Steepest Descent Method

In the steepest descent method, the minimizer calculates numerically, the first derivative of the energy function to find the minimum. The energy is calculated for the

initial geometry and again after one of the atoms has moved in a small increment in one of the directions of the coordinate system. This process is repeated for all 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, and therefore takes numerous tiny steps to proceed down a long narrow valley. The procedure 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 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 method **[13, 20, 27, 28]**.

2.3.2 Conjugate Gradient Method

In the steepest descent method every successive step is at right angles to the previous one, making the method to have many tiny steps as it proceeds down a long narrow valley due to being forced to make a right-angled turn at every point even if this is not the best toute to find the minimum. The conjugate gradient method, on the other hand, produces a set of directions that overcome the oscillatory behaviour of steepest descents in narrow valleys. Successive directions are not at right angles to each other **[27]**, but 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 the storage requirements are greater than

those for steepest descent, but the conjugate gradient is the method of choice for larger systems. The greater total computational expanse and the longer time per iteration are more than compensated by the more efficient convergence to the minimum achieved in the case of conjugate gradients **[27, 28]**.

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. 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 **[20, 28]**.

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 convergence to a maximum derivative less than 0.001 kcal $mol^{-1}Å^{-1}$ is required to find a final minimum **[28]**. Energy barriers are overcome by using replica exchange dynamics whereby trajectories swop temperatures and the higher temperatures enable peptides to attain higher energies and escape local minima trappings.

2.4 Molecular Dynamics (MD)

Molecular dynamics (MD) is described as a form of computer simulation in which atoms and molecules are allowed to interact for a period of time by approximations of known physical attributes, giving a view of the motion of the atoms **[29]**. Alternatively, MD is the simulation of motion for a system of particles **[30]**. 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. Molecular dynamics is a multidisciplinary method whose laws and theories stem from mathematics, physics, and chemistry, and it employs algorithms from computer science and information technology. 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. **[30, 31]**.

Molecular dynamics is concerned with simulating the motion of molecules to gain a deeper understanding of chemical reactions, fluid flow, phase transitions, droplet formation, and other physical phenomena that arise from molecular interactions **[29, 30, 31, 32]**. These studies include not only the motion of many molecules as it's the case in a fluid, but also the motion of a single large molecule consisting of hundreds or thousands of atoms, such as a protein. Motion is inherent to all chemical processes. Simple vibrations, like bond stretching and angle bending, give rise to infra-red (IR) spectra. Chemical reactions, hormone-receptor binding, and other complex processes

are associated with many kinds of intra- and intermolecular motions. This work uses simple classical Newtonian mechanics, most notably Newton's law:

$$\mathbf{F}_i = m_i\mathbf{a}_i \qquad \text{.....(5)}$$

For each atom *i* in a system constituted by N atoms. Here, m_i is the atom mass, $\mathbf{a}_i = d^2\mathbf{r}_i/dt^2$ its acceleration, and $\mathbf{F}_i$ the force acting upon it, due to the interactions with other atoms. Therefore, in contrast with the other methods, molecular dynamics is a deterministic technique: given an initial set of positions and velocities, the subsequent time evolution is in principle completely determined.

As one of its many and varied applications, MD allows studying the dynamics of large macromolecules, including biological systems such as proteins, nucleic acids (DNA, RNA), and biological membranes. Dynamical events may play a key role in controlling processes, which affect functional properties of biomolecules. Drug design is commonly used in the pharmaceutical industry to test properties of a molecule at the computer level without the need to synthesize it, which is far more expensive **[33]**.

2.4.1 Periodic Boundary Conditions

A more realistic approach in simulations is to use the solvent explicitly. This is done by soaking the molecule in a box of solvent molecules. This method has the disadvantage of requiring additional computational effort. Periodic Boundary Conditions (PBC) are normally employed to model the bulk solvent. In PBC, the simulation box is

infinitely replicated in all directions to form a lattice. In practice, most molecular dynamics (MD) simulations evaluate potentials using some cutoff scheme for computational efficiency. In these cutoff schemes, each particle interacts with the nearest images of the other N-1 particles (minimum-image convention). The use of cutoff methods, however, has been shown to introduce significant errors and artificial behaviour in simulations **[20, 28]**.

2.4.2 Ewald Summation Techniques

In most MD simulations, the long-range interactions (Coulomb interactions) are the most time consuming. Ewald summation was introduced in 1921 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, each of which converges much more rapidly and a constant term:

$$U_{Ewald} = U^{r} + U^{m} + U^{0} \quad \text{.....(6)}$$

This is done by considering each charge to be surrounded by a neutralising charge distribution of equal magnitude but of opposite sign. The sum over point charges is converted to a sum of the interactions between the charges plus the neutralising distributions. This part is the real space sum U^{r}. A second charge distribution is added to the system which exactly counteracts the first neutralising distribution. This

summation is performed in the reciprocal space and is termed U^m. The dipole term includes the effects of the total dipole moment of the unit cell, the shape of the macroscopic lattice, and the dielectric constant of the surrounding medium [**20, 28]**.

2.4.3 Particle-Mesh Ewald

The Particle-Mesh Ewald method (PME) divides the potential energy into Ewald's standard direct and reciprocal sums and uses the conventional Gaussian charges distributions. 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. Furthermore, PME does not interpolate but rather evaluates the forces by analytically differentiating the energies, thus reducing memory requirements substantially **[20, 28]**.

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. By doing so, energy barriers on the potential energy surface might be overcome, allowing for the exploration of new conformational space.

Replica exchange molecular dynamics is a computational method that couples MD trajectories with a temperature exchange Monte Carlo (MC) process for efficient

sampling of the conformational space **[34]**. 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 high energy barriers **[35]**. Each replica evolves independently by MD and only 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 **[5, 35]**. The direction of this exchange of neighbouring replicas is chosen at random **[35]**.

2.6 Simulated Annealing (SA)

Annealing is the process in which the temperature of a molten substance is slowly reduced until the material crystallises to give a large single crystal, which corresponds to the global minimum of the free energy. Simulated annealing (SA) is a computational method that mimics this process in order to find the optimal solutions to problems which have a large number of possible solutions **[36]**. It is a heuristic global optimization algorithm, which has been successfully applied to solve many difficult optimization problems **[37]**. 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 high temperatures, the system is able to occupy high energy regions of the conformational space and to surmount high energy barriers. As the temperature falls, the lower energy states become more attainable in accordance with the Boltzman

distribution. At absolute zero the system should occupy the lowest energy state, often referred to as the global minimum energy conformation **[37]**. For successful SA, the procedure should be repeated many times and statistical methods should be used to demonstrate that the chances of getting new low energy structures are getting smaller.

2.7 Structure Classification

The physical, chemical and biological properties of a molecule often depend critically upon the 3D structures, or their conformations. Conformational analysis is the study of the conformations of a molecule and their influence on its properties. A key component of a conformational analysis is the conformational search, the objective of which is to identify the preferred conformations of a molecule, and these conformations determine its behaviour. This requires the location of conformations that are at minimum points on the energy surface. Molecular dynamics and MC methods can generate an ensemble of states that includes structures not at energy minima. On the other hand, conformational search algorithms can generate conformations that are very similar, if not identical. Under such circumstances it is desirable to be able to select from the data set a smaller, representative set of conformations for subsequent analysis. This can be done using cluster analysis (CA), which groups together similar objects, from which the representatives can be extracted. CA requires a measure of the similarity or dissimilarity between pairs of objects. Comparison of conformations could be based on the root mean square deviation (RMSD) or the distance between two conformations measured in terms of their torsion angles, as was the case in this work. All clustering methods first join the two structures that are closest, to which other similar conformations are added

to eventually form a cluster of structures. In addition linkage algorithms are then used to join closely related clusters. These linkage methods are hierarchical clustering methods because there is a specific order in which the clusters are formed and amalgamated **[28]**.

An example of a non-hierarchical clustering method is the Jarvis-Patrick algorithm. It uses a 'nearest neighbour' approach. The nearest neighbours of each conformation are the other conformations that are the shortest distance away. Two conformations are considered to be in the same cluster in the Jarvis-Patrick method if they satisfy the following criteria:

- They are in each others's list of *m* nearest neighbours.
- They have *p* (where $p<m$) nearest neighbours in common.

Conformations can thus be placed in clusters and clusters fused together without any hierarchical relationships. This method can also be extended to take into account not only the number of nearest neighbours but also the position of each conformation within the neighbouring list **[28]**.

CHAPTER 3

VALIDATION STUDY: SIMULATIONS OF PEPTIDE FOLDING USING A TEN-RESIDUE MINIPROTEIN (CLN025)

3.1 Introduction

In chemical terms, proteins and peptides are linear heteropolymers of amino acids. The term “protein” generally refers to natural amino acids joined in a particular sequence and having a defined 3D structure and a specific function, whereas the term “peptide” is generally used for short oligomers that often lack a definite conformation. Moreover, a protein has the same chemical structure both in solution and in crystal form, whereas a peptide can adopt one structural conformation when in solution and a different one when it is crystallized. Peptides and proteins are composed of twenty different naturally occurring amino acids, although certain peculiar amino acids can be found in peptides isolated from lesser evolved species. However, this repertoire can be further expanded with synthetic amino acids. Today peptides can easily be synthesized using solid phase synthesis methods **[38]**.

Determination of the precise conformation of polypeptides is vital to designing *de novo* drugs (peptidomimetics) using computer-aided drug design methods **[2]**. To achieve this goal, molecular dynamics (MD) simulations of peptide folding can be used to get information about the conformational features of peptides as a complementary technique to experimental methods like NMR or X-ray crystallography. However, computational methods rely heavily on the accuracy of force fields representing the

physical properties of peptides **[39]**. Accordingly, a continuous effort in improving them and in running benchmark calculations is necessary. Indeed, the early AMBER force field set of parameters (ff94) developed in the early nineties was found to overemphasize helical conformations at the expense of extended conformations **[39]**. Improvements carried out on this force field gave rise to the current second generation of force fields (ff96, ff99and modifications of ff99). Thus, force field ff96 was developed using quantum mechanics calculations in order to improve the peptide backbone φ-ψ torsional parameters of force field ff94. With subsequent improvements, the force fields ff96 and ff99 sets were found to inaccurately represent glycine and exhibited a strong α-helical bias. Consequently, the ff99 set was later modified in different attempts to re-parametize the backbone torsional terms using calculations on alanine and glycine tetrapeptides. One of these attempts resulted in the formation of force field ff99SB, which is rather unbiased and provides a better balance between extended and helical conformations. Following a different approach the force field ff03 set was developed by refitting charges in force field ff99 altogether using quantum calculations with a continuum dielectric constant to better represent the effects of the solvent **[39]**.

As mentioned earlier, there is a need to continuously benchmark the performance of force fields by subjecting them to different conditions, especially when implicit descriptions of the solvent are used. This work focused on testing the performance of force fields ff94, ff96 and ff99SB (modified ff99) in implicit solvent for studying peptide folding. In order to perform this study a small chignolin-like peptide (CLN025) with a well characterized 3-D structure was considered. CLN025 (with

sequence: Tyr^{1}-Tyr^{2}-Asp^{3}-Pro^{4}-Glu^{5}-Thr^{6}-Gly^{7}-Thr^{8}-Trp^{9}-Tyr^{10}) is a synthetic ten-residue polypeptide with a stable topology in an aqueous solution whose crystal structure is also known **[38]**. Calculations in this study were done using the replica exchange molecular dynamics (REMD) methodology to avoid getting trapped in one of the many minima in the energy landscape and thereby avoided failing to reach the actual global minimum which often results in obtaining fewer conformations of the peptide populating the trajectory **[5, 35]**.

Therefore, to test the stability and validity of the force fields used in the AMBER package, the REMD protocol was used instead of standard molecular dynamics to study folding patterns of CLN025. The water solvent was treated explicitly using the force field ff99SB parameters with the addition of counterions (this was the reference calculation, informed by the understanding that results from explicit solvent simulations are closer to experimental data), and implicitly using the analytical linearized Poisson-Boltzmann (ALPB) approach **[40]** and the generalized Born approximation modified by Onufriev, Bashford and Case **[41]**. The trajectories were run using a Langevin thermostat.

3.2 Computational Protocols

All the REMD calculations were carried out with the AMBER 9.0 programme **[26]**. The ten residue peptide (CLN025) was modelled implicitly and using an explicit solvent by considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms. The remaining hydrogen atoms were considered as part of the carbon atoms to which they are covalently attached **[5, 35]**.

The force field parameters were taken from the all-atom version of AMBER9.0 **[26]** and the dielectric constant around the peptide was set to ε=1.0. The temperature during the simulations was controlled by the constraint method. The unit timestep was set to 0.5 fs, and a regular MD simulation run of 1.0 ns was done for each replica, starting from an extended conformation of the peptide (CLN025). In the current study 12 replicas (M=12) were used with temperatures: 274, 300, 328, 358, 389, 423, 459, 497, 538, 582, 630, and 682 K. These temperatures are spread in a manner that allows for optimal temperature distribution. Prior to data collection, regular canonical MD simulations were made for 5 ns at each temperature, and then a replica-exchange simulation of 5 ns for thermalization.

The replica exchange was attempted every 10 fs, and the data was stored for later analyses. Following preliminary MD runs for 10 ns per replica, each trajectory was run for a length of 100 ns simulation time. Integrity of the REMD trajectories was monitored by plotting probability distribution of their potential energy. Further analysis of the results included the plotting of the structural patterns and the secondary structure motifs populating the different trajectories. Superposition of REMD structures to the Crystal and NMR structures of CLN025 was evaluated by determining the RMSD value.

3.3 Results and Discussion

The ten-amino acid protein CLN025 was studied with REMD using both implicit and explicit solvent models. The explicit solvent model which was run using force field ff99SB and TIP3P for water as per AMBER 9.0 **[26]**, and this was used as a reference

calculation. Implicit solvent calculations were run using force fields ff94, ff96 and ff99SB with the ALPB method **[40]** to simulate solvent effects. Comparisons were made between the different simulation models and their results were compared with experimental data published in literature **[38].**

Temperature distribution was chosen in such a way that the probability for exchange between neighbouring replicas is P (exchange) = 0.2. Computed P (exchange) was > 0.18 for all temperatures in all cases. The number of replicas was set to 40 in the explicit solvent case and 12 in the implicit solvent case. The temperatures distributions for the explicit and implicit REMD trajectories are presented in Table 3.1.

Table 3.1: Temperature distributions for the different REMD trajectories

Explicit Solvent REMD	**Implicit Solvent REMD**		
ff96SB	**ff94**	**ff96**	**ff99SB**
Temperatures (K)			
296.8	*275.7*	275.3	275.5
300.0	*300.0*	300.0	300.0
303.2	*326.2*	326.7	326.5
306.5	*354.5*	355.6	355.2
309.8	*385.1*	386.9	386.3
313.2	*418.2*	420.8	420.0
316.6	*453.9*	457.5	456.5
320.1	*492.6*	497.3	496.1
323.6	*534.5*	540.5	539.0
327.2	*580.0*	587.5	585.7
330.8	*629.5*	638.8	636.5
334.5	*683.6*	694.9	692.0

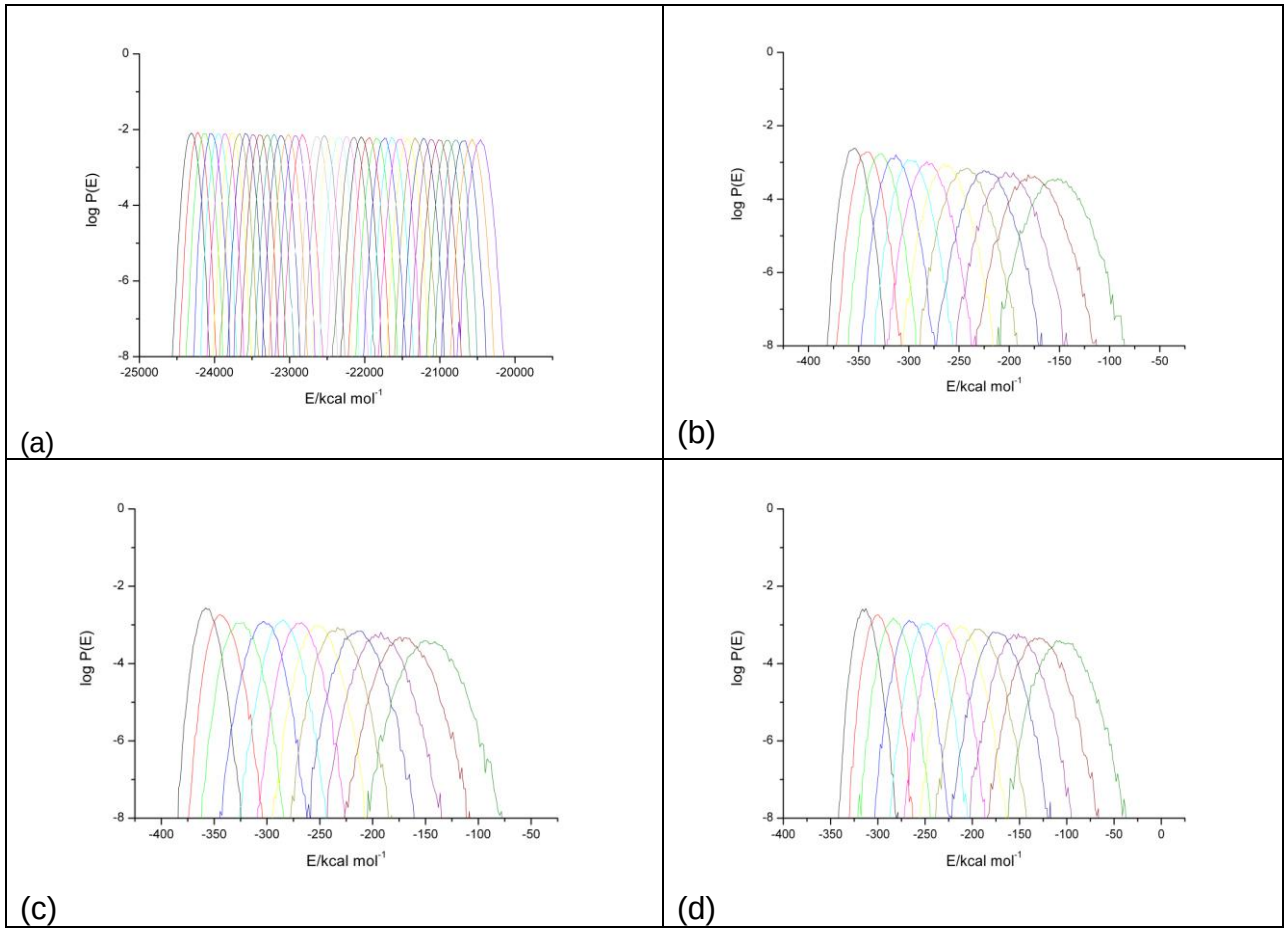


Figure 3.1: Probability distribution of the potential energy for CLN025 REMD trajectories using different force fields and solvent conditions. (a) with explicit ff99SB; (b) with implicit (ALPB) ff94; (c) with implicit (ALPB) ff96; (d) with implicit (ALPB) ff99SB.

In order to verify if the number of exchanges between replicas was sufficient, the canonical probability distribution functions of the potential energies were plotted for all the explicit and implicit REMD trajectories, as shown in Figure 3.1. It can be observed that there is enough overlap between the distributions at different temperatures to assure the condition of free walk in the temperature space, ensuring the robustness of the process followed.

The dihedrals from explicit and implicit REMD trajectories were used to obtain the conformational patterns that populate the different trajectories. These different simulations whose patterns are presented in Figures 3.2a-d, show efficient sampling of the conformational space. The implicit solvent trajectories using force fields ff94 (Figure 3.2c) and ff99SB (Figure 3.2d) sample the conformational space in a steady pace like the explicit solvent trajectory (Figure 3.2a). The trajectory ran using force field ff94 started off in a steady pace but soon levelled off. Nevertheless, all these trajectories show efficient sampling of the conformational space.

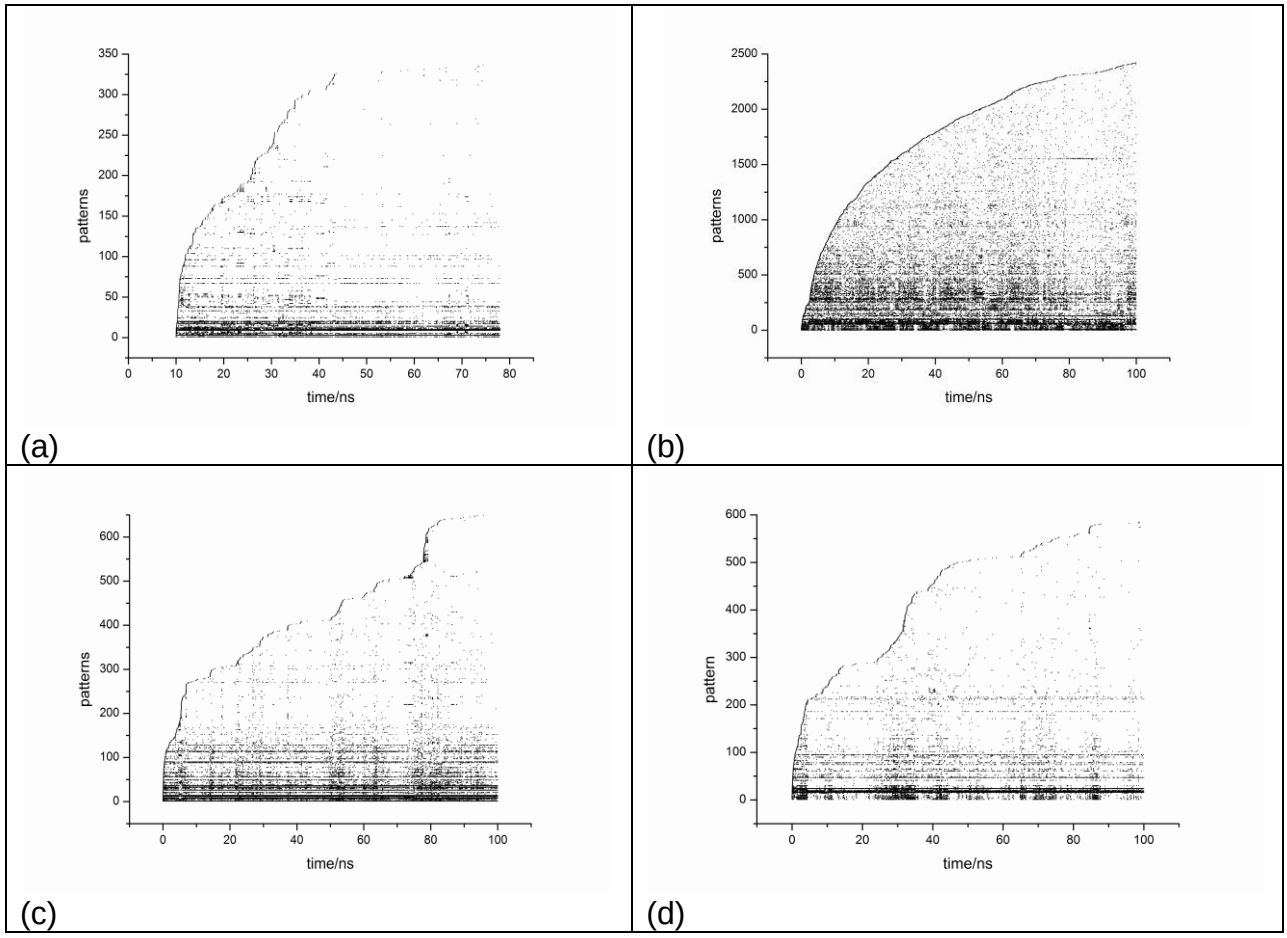


Figure 3.2: Evolution of new structural patterns obtained for the different REMD trajectories computed using explicit and implicit water solvent at 300K. (a) using explicit solvent (b) using implicit ff94; (c) using implicit ff96; and (d) using implicit ff99SB. A higher pattern number means more structures are sampled.

The trajectories at 300 K were compared with the 20 structures derived from NMR studies **[38]**, through backbone superposition. Of the twenty possible plots showing the RMSD of the configurations sampled during the MD trajectory with the different structures derived from NMR experiments, Figure 3.3 shows the structures with the lowest RMSD values in explicit solvent calculations. In Figure 3.3a, the calculated RMSD (in explicit solvent simulation) suggests that the system basically moves between two different conformations: one with an RMSD value around 1.75 Å and the other with a value around 0.9 Å. In the implicit solvent simulation using force field ff94 (Figure 3.3b), the system displays a large deviation from experimental structures. Plots corresponding to force fields ff96 (figure 3.3c) and ff99SB (Figure 3.3d) in implicit solvent suggest that the system samples only one of the two minima.

Patterns common between explicit and implicit solvent REMD calculations were computed, and the results are presented in Table 3.2. The pattern superposition was calculated based on two procedures: (1) number of patterns that implicit solvent REMD has equal to the explicit solvent REMD taking into account the total number of patterns appearing in the last part of the simulation and (2) number of configurations in the ensemble sampled by the implicit solvent REMD with a pattern corresponding to one configuration in the explicit solvent ensemble taking into account the total number of configurations sampled by the explicit solvent REMD (Table 3.2).

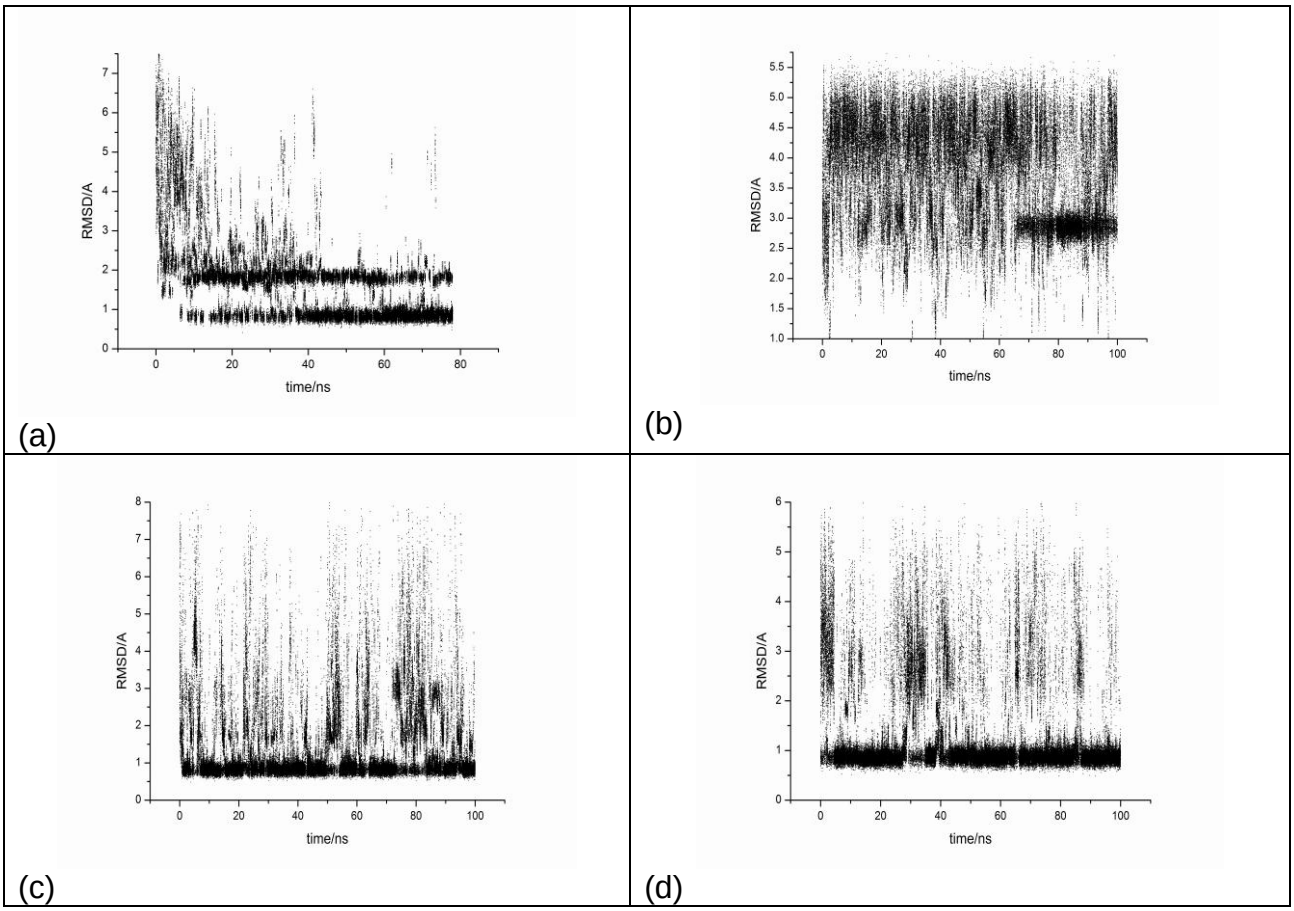


Figure 3.3: Minimum RMSD of backbone superposition between explicit REMD trajectory and CLN025 NMR structure. (a) with explicit ff99SB; (b) with implicit (ALPB) ff94; (c) with implicit (ALPB) ff96; (d) with implicit (ALPB) ff99SB.

Table 3.2: Pattern sharing in CLN025 trajectories obtained using different force fields using procedure 1 and 2 ad defined in page 30

Force field	**ff94**	**ff96**	**ff99SB**
Procedure 1	42.8 %	61.3 %	67.3 %
Procedure 2	9.6%	73,4%	78.5%

Force field ff94 REMD calculations in implicit solvent shares a lot of patterns with explicit solvent simulations (around 40%) but the sampling of these patterns is quite different. The replacement of explicit with implicit solvent model (maintaining the force field) makes inaccessible a number of patterns that may

be obtained due to direct interaction with water molecules. Implicit solvent REMD simulations with force fields ff96 and ff99SB share a similar amount of patterns with the reference explicit solvent simulation.

Table 3.3: Percentage occupancy of hydrogen bonds in the explicit solvent trajectory

DONOR	ACCEPTOR	% occ.	Lifetime
Asp3.O	**Gly7.N**	**79.74**	**7.7**
Thr8.O	**Asp3.N**	**48.61**	**9.9**
Tyr1.O	**Tyr10.N**	**46.65**	**10.5**
Asp3.OD2	Thr6.OG1	43.3	7.6
Asp3.OD1	**Thr6.OG1**	**42.45**	**7.2**
Asp3.O	**Thr6.N**	**38.66**	**2.3**
Asp3.OD2	Glu5.N	35.94	3.5
Asp3.OD1	**Glu5.N**	**35.27**	**3.5**
Asp3.OD2	Thr6.N	34.5	3.7
Asp3.OD1	Thr6.N	34.06	3.7
Tyr1.O	Trp9.N	33.27	5.2
Thr6.OG1	**Thr8.OG1**	**32.64**	**5.4**
Asp3.O	**Thr8.N**	**31.59**	**3.0**
Gly7.O	Asp3.N	36.71	2.1
Tyr2.O	Trp9.NE1	10.52	1.3
Glu5.OE1	water mol.	187.43	2.5
Glu5.OE2	water mol.	187.30	2.5
Tyr10.O	water mol.	175.39	2.7
Asp3.OD1	water mol.	140.62	2.7
Asp3.OD2	water mol.	137.29	2.7
Glu5.O	water mol.	111.77	2.1
Pro4.O	water mol.	108.32	2.1
Thr6.O	water mol.	94.01	2.0
Tyr2.O	water mol.	90.24	2.6
Trp9.O	water mol.	79.97	2.2
Thr8.OG1	water mol.	70.72	1.9
Gly7.O	water mol.	68.48	2.1
Tyr1.OH	water mol.	66.32	1.5
Tyr2.OH	water mol.	62.90	1.5
Tyr10.OH	water mol.	61.82	1.6
water mol.	Tyr2.OH	58.27	3.2
water mol.	Tyr1.OH	55.27	2.9
water mol.	Tyr2.N	52.07	2.7
water mol.	Tyr10.OH	49.49	3.1
Thr8.O	water mol.	49.20	2.2

Table 3.4: Percentage occupancy of hydrogen bonds in the implicit ff94 solvent trajectory

DONOR	ACCEPTOR	% occ.	Lifetime
Asp3.O	**Thr6.N**	**44.55**	**3**
Asp3.O	**Gly7.N**	**38.15**	**3.1**
Asp3.O	Thr6.OG1	37.99	5.5
Glu5.O	Thr8.N	37.52	3.7
Glu5.O	Thr8.OG1	35.3	4.6
Thr6.O	Trp9.N	34.61	3
Pro4.O	Gly7.N	34.26	2
Tyr2.O	Glu5.N	29.54	2.5
Tyr2.O	Thr6.N	23.46	3.4
Tyr2.O	Thr6.OG1	20.71	4.8
Asp3.OD1	Thr6.OG1	19.83	5.1
Thr6.O	Tyr10.N	19.03	3.1
Asp3.OD2	**Thr6.OG1**	**18.72**	**4.1**
Gly7.O	Tyr10.N	17.28	1.4
Asp3.OD1	Thr6.N	15.85	4.6
Thr6.OG1	**Thr8.OG1**	**15.24**	**4.7**
Thr8.OG1	Asp3.N	12.39	4.9
Asp3.OD2	Thr6.N	11.98	4.1
Asp3.O	Trp9.N	10.54	3.4

Table 3.5: Percentage occupancy of hydrogen bonds in the implicit ff96 solvent trajectory

DONOR	ACCEPTOR	% occ.	Lifetime
Asp3.O	**Gly7.N**	**77.8**	**7.5**
Tyr1.O	**Tyr10.N**	**60.08**	**9.7**
Thr8.O	**Asp3.N**	**58.38**	**6.1**
Asp3.OD1	**Thr6.OG1**	**45.01**	**9.68**
Asp3.OD2	Thr6.OG1	44.54	9.79
Asp3.OD2	Thr6.N	42.45	6.8
Asp3.OD1	Thr6.N	42.27	6.6
Asp3.O	**Thr8.N**	**37.41**	**2.4**
Thr6.OG1	**Thr8.OG1**	**26.22**	**3.2**
Gly7.O	Thr8.OG1	10.91	1.6
Asp3.OD2	Glu5.N	10.26	1.3
Asp3.OD1	**Glu5.N**	**10.14**	**1.3**
Thr8.OG1	Trp9.N	10.06	6.54

Table 3.6: Percentage occupancy of hydrogen bonds in the implicit ff99 solvent trajectory

DONOR	ACCEPTOR	% occ.	Lifetime
Asp3.O	**Gly7.N**	**72.9**	**4.8**
Tyr1.O	**Tyr10.N**	**67.45**	**11.7**
Thr8.O	**Asp3.N**	**63.03**	**5.7**
Asp3.OD2	Thr6.OG1	46.68	9.98
Asp3.OD1	**Thr6.OG1**	**43.17**	**7.8**
Asp3.OD2	Thr6.N	40.73	4.6
Asp3.OD1	Thr6.N	37.42	4.4
Thr6.OG1	**Thr8.OG1**	**29.98**	**2.8**
Asp3.OD2	Glu5.N	25.75	1.8
Asp3.O	**Thr8.N**	**24.82**	**1.6**
Asp3.OD1	**Glu5.N**	**24.09**	**1.7**
Asp3.O	**Thr6.N**	**22.3**	**1.6**
Thr6.OG1	Thr8.N	14.14	1.4
Thr6.OG1	Tyr10.OH	11.7	2.1
Thr6.O	Thr8.OG1	11.41	1.7

The number of hydrogen bonds with more than a 10% of occupancy located in each simulation was calculated and presented in the Tables 3.3 to 3.6, with bold typed entries corresponding to hydrogen bonds obtained from X-ray or NMR experimental studies **[38]**.

All the experimentally determined hydrogen bonds were also found in the explicit solvent study (Table 3.3), as well as in the implicit REMD simulations with force fields ff96 and ff99SB (Tables 3.5 to 3.6). Moreover, important hydrogen bonds exist between solvent and the peptide. On the other hand, implicit solvent force field ff94 (Table 3.4) results did not reproduce hydrogen bond patterns similar to the experimental data **[38]**.

As part of the statistical analysis used in this work, principal component analysis (PCA) of the explicit solvent trajectory was performed. To achieve this, the trajectory was initially superimposed onto the crystal structure and then, the covariance matrix of the backbone movements was calculated and diagonalized to obtain the set of principal components and their eigenvalues. Thereafter, the coordinates of all the structures were projected onto the principal components with higher eigenvalues (the principal component with highest eigenvalue was denoted as PC1, the one with the second highest eigenvalue as PC2, etc) (Figure 3.4). These projections were transformed into 2D potentials of mean force (also known as 2D free energy surfaces (FES)) by applying the formula $\Delta W_i = -RTln(P_i)$, where ΔW_i is the potential of mean force in a given area and P_i is the probability of finding the system in this area (Figure 3.5 to 3.6).

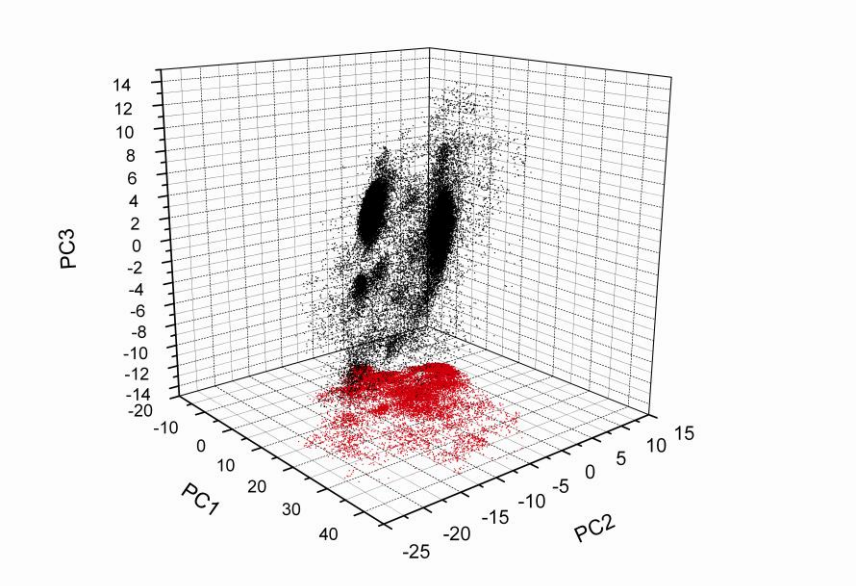


Figure 3.4: Projection of the explicit water 300 K REMD trajectory onto its three principal components. The two main black spots on the z-axis represent two low energy structures (minima).

PCA of explicit solvent REMD trajectory confirms the existence of two minima in the free energy surface (Figures 3.4 to 3.6). The experimental results in literature **[38]**

seem to approximate just one of the two minima (Figure 3.7). The absence of the second minimum in NMR and X-ray structures (Figure 3.7) **[38]** can be attributed to the fact that the average NMR structure does not have all the dynamic features of the peptide which the REMD simulation was able to explore due to its superior ability to sample the energy landscape.

The implicit solvent REMD simulations were also projected on their principal components and plotted over the 300 K free energy surfaces as shown in Figures 3.8 to 10). It can be observed that the experimental structures **[38]** were present in the implicit REMD trajectories obtained using force fields ff96 and ff99SB (Figures 3.9 and 10, respectively). These Figures suggest that the simulations were able to reproduce both minima as the reference explicit solvent REMD simulation obtained using force field ff99SB, but showed preference for the one corresponding to the experimental structures **[38]**.

However, the implicit solvent simulation using ff94 (Figure 3.8) did not reproduce either the explicit solvent simulation or the experimental structures **[38]**. This indicates the inability of this force field to adequately sample the conformational energy landscape**.** This explains why this force field needed to be modified or improved, hence the development of force fields ff96 and ff99 (and modified versions of ff99).

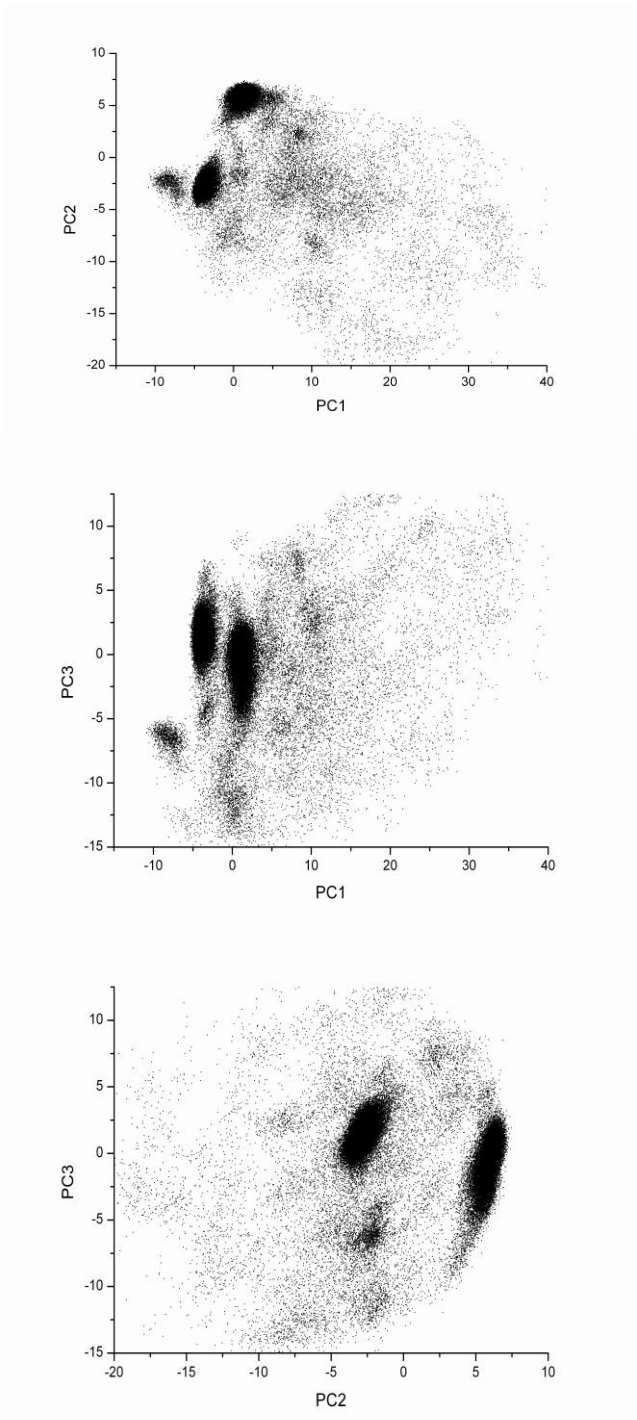
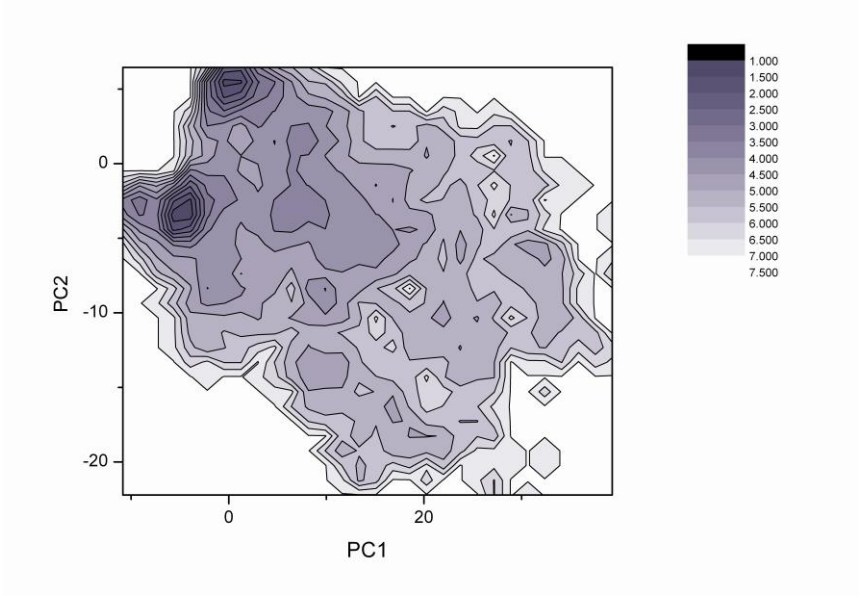
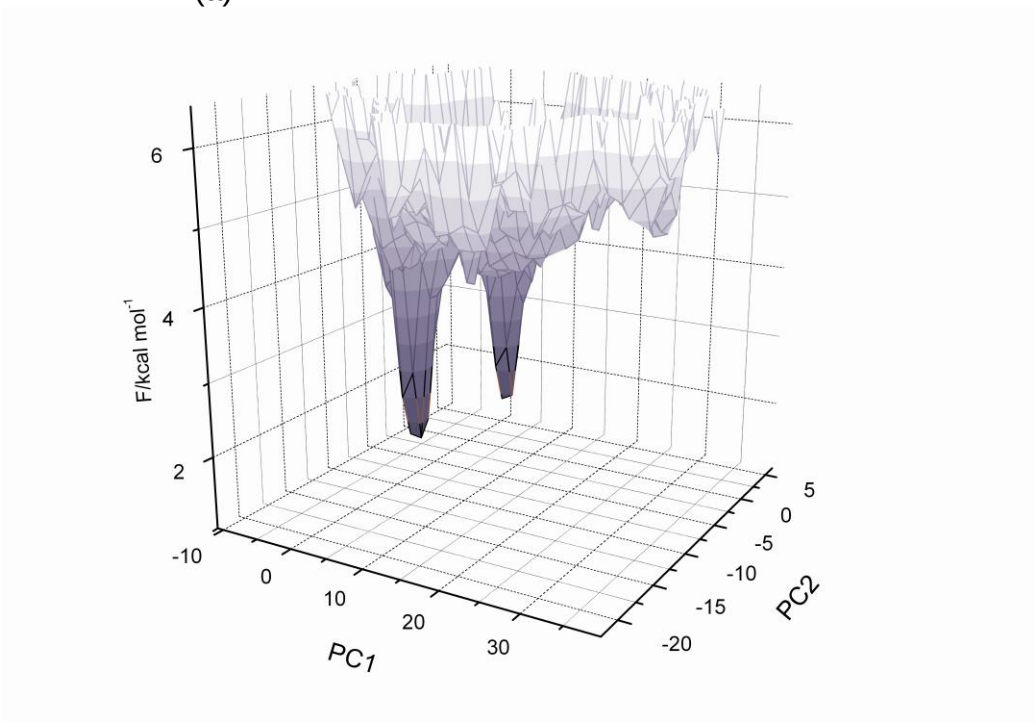


Figure 3.5: 2D projections of the last 30 ns of explicit water 300 K REMD trajectory on its three main principal components. In all three cases two energy minima are visible (bold black spots).



(a)



(b)

Figure 3.6: 2D and 3D plots of the free energy surfaces at 300 K projected on the two first principal components of the explicit REMD trajectory, showing two energy minima. (a) 2D plot; and (b) 3D plot.

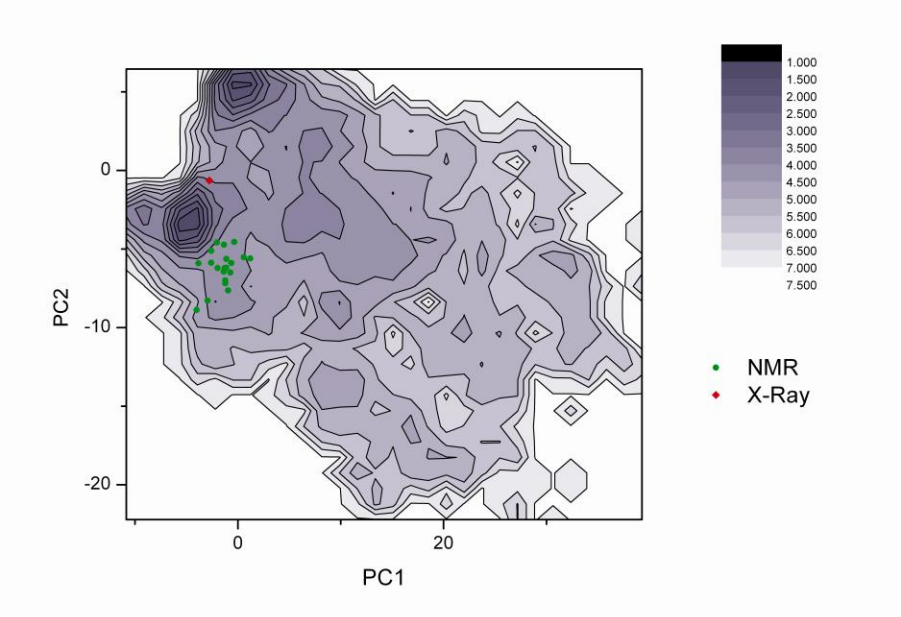


Figure 3.7: Projection of the experimental NMR and X-ray structures from literature **[38]**.

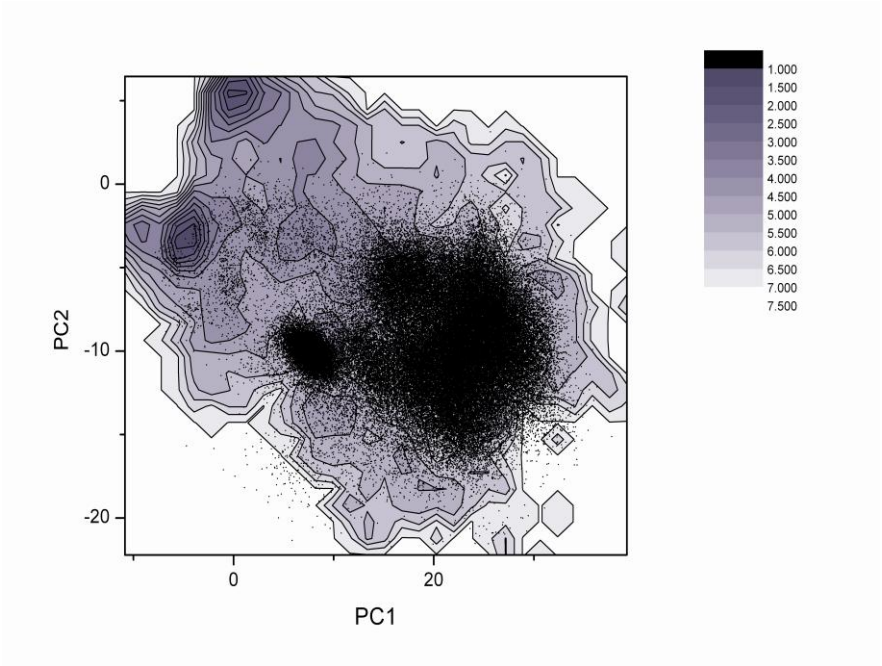


Figure 3.8: Projection of implicit solvent 300 K REMD trajectory using ff94, showing unpopulated minima in the top-left corner.

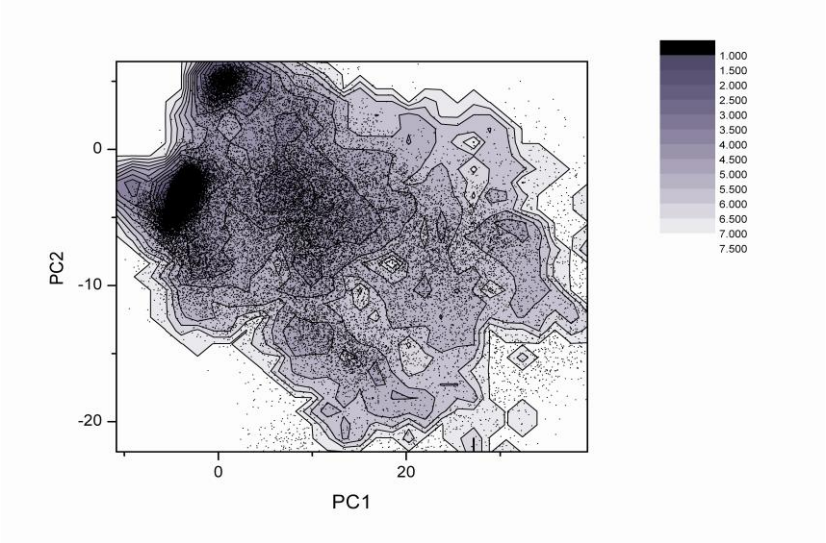


Figure 3.9: Projection of implicit solvent 300 K REMD trajectory using ff96, showing the two minima in the top-left corner densely populated with structures.

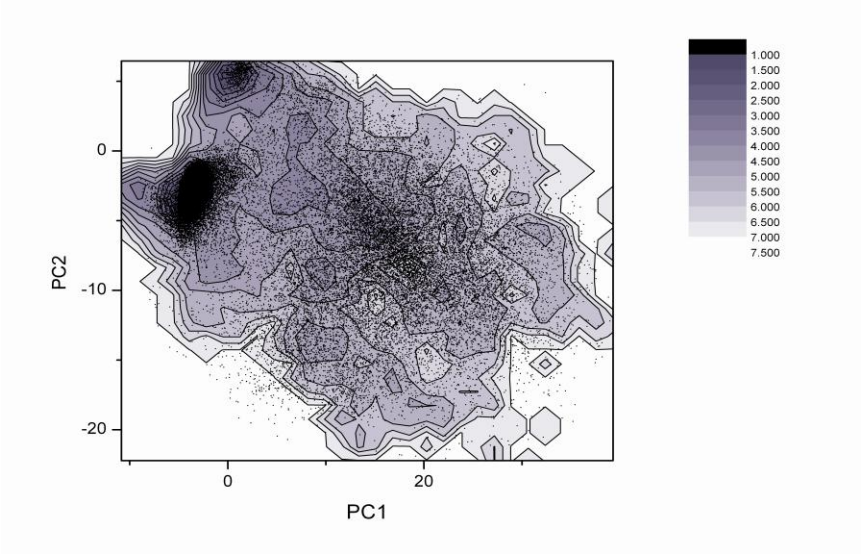


Figure 3.10: Projection of implicit solvent 300 K REMD trajectory using ff99SB, showing that one of the two minima in the top-left corner is densely populated with structures (in black).

To obtain a deeper understanding of the structural features underlying the behaviour of these two minima obtained from the explicit and implicit REMD trajectories (Figures 3.4 to 3.6, 3.9 and 3.10), cluster analysis was performed, and histograms were constructed. To achieve this, a set of structures was equidistantly chosen as a

representative of the entire explicit solvent trajectory. A distance matrix was constructed taking the backbone RMSD value as distance criteria. Thereafter, minimization of the variance hierarchic clustering was performed. Subsequently, all the related elements were grouped in 4 major clusters, and the results are presented in Figure 3.11, where it can be observed that cluster 1 and 3 are more populated compared to clusters 2 and 4. The structures of each of these clusters were projected on the two main principal components leading to a clear assignment of each of them with a different minimum (Figure 3.12). A closer examination of the different hydrogen bond interactions dominant in these two different clusters would lead to understanding the differences between the two minima.

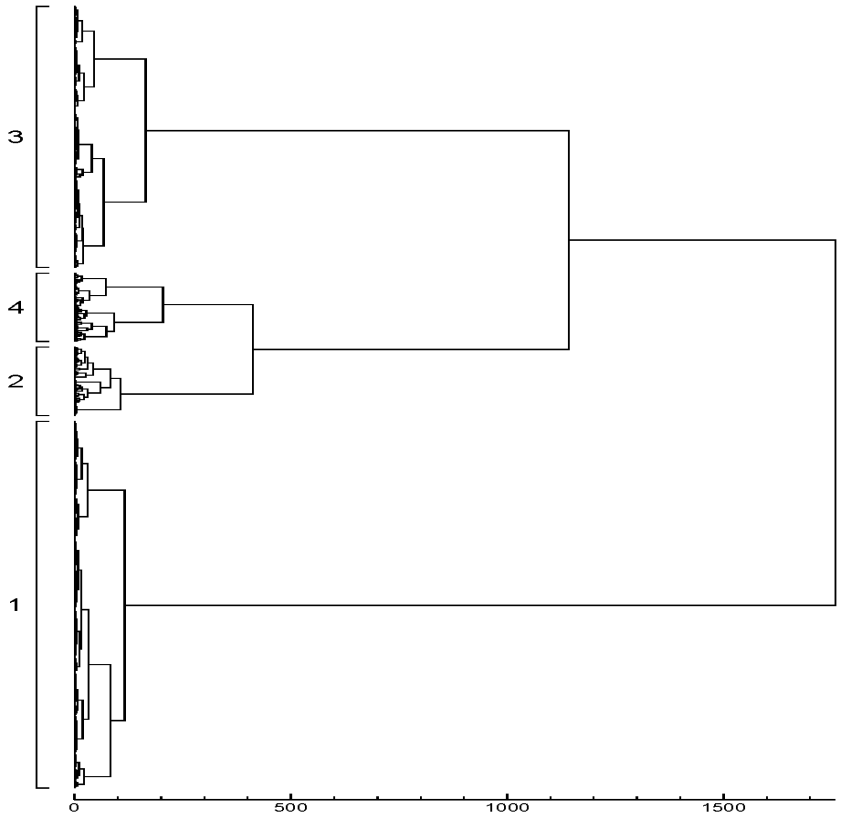


Figure 3.11: Hierarchic diagram for CLN025 explicit REMD trajectory, showing four major types of structure clusters.

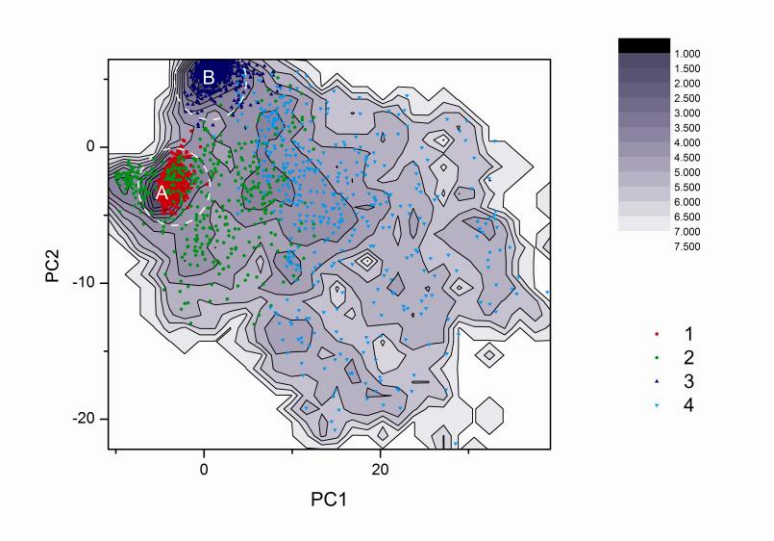


Figure 3.12: Projection of elements of the clusters on the free energy surface of explicit REMD for CLN025. The two minina in the top-left corner designated as A (red) and B (blue) are densely populated.

Table 3.7 provides a statistical hydrogen bond analysis performed for each cluster of the explicit REMD trajectory. Qualitative displays of characteristic hydrogen bond interactions dominant in cluster 1 and cluster 3 are presented in Figures 3.13 and 14**.**

Table3.7: Hydrogen bond percentage occupancy of structural clusters for CLN025

		Cl_1	***Cl_3***	***Total***
DONOR	***ACCEPTOR***	***% occ.***	***% occ.***	***% occ.***
Thr8.O	**Asp3.N**	**97.36**	**----**	**48.61**
Thr8.O	water mol.	3.79	100.18	49.20
Tyr1.O	Tyr10.N	97.62	----	46.65
Tyr1.O	Trp9.N	----	97.67	33.27
Thr6.OG1	Thr8.OG1	66.54	----	32.64
Thr6.OG1	water mol.	32.74	66.57	48.55
Trp9.O	water mol.	127.76	17.64	79.97
Gly7.O	water mol.	112.18	8.19	68.48
Gly7.O	Asp3.N	----	89.89	36.71

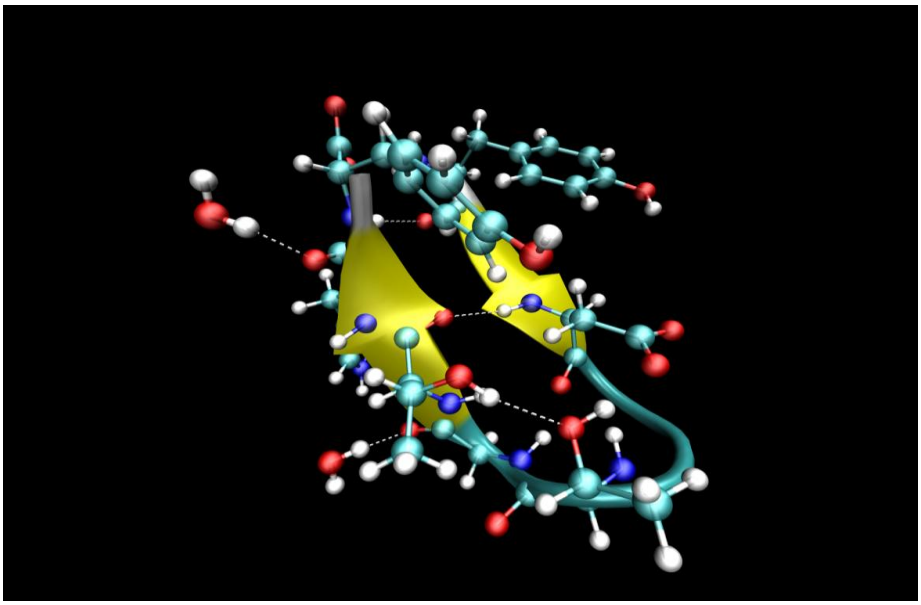


Figure 3.13: Cluster 1 representative structure of CLN025 with characteristic interactions displayed.

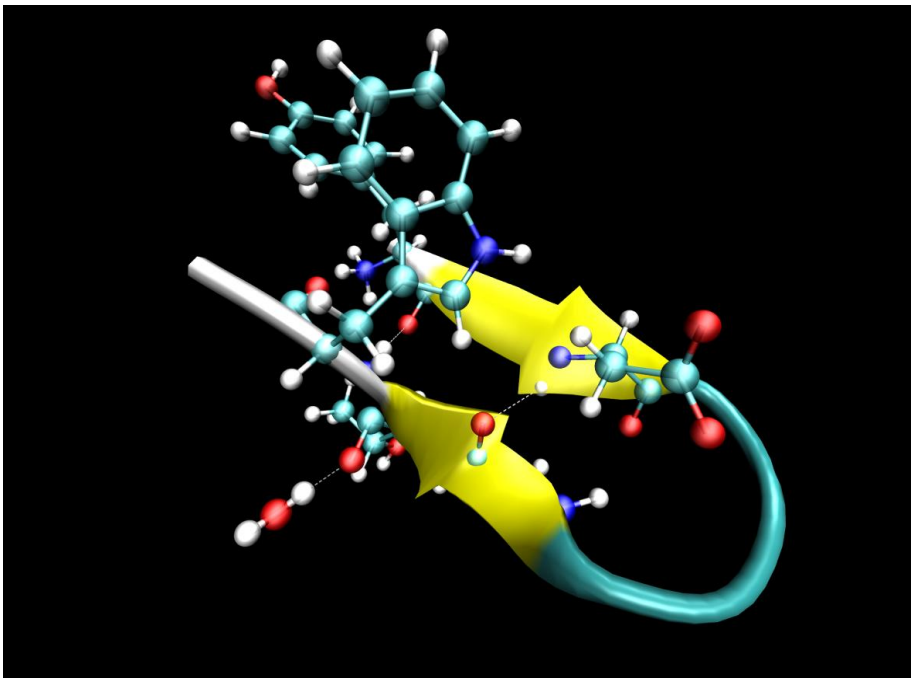
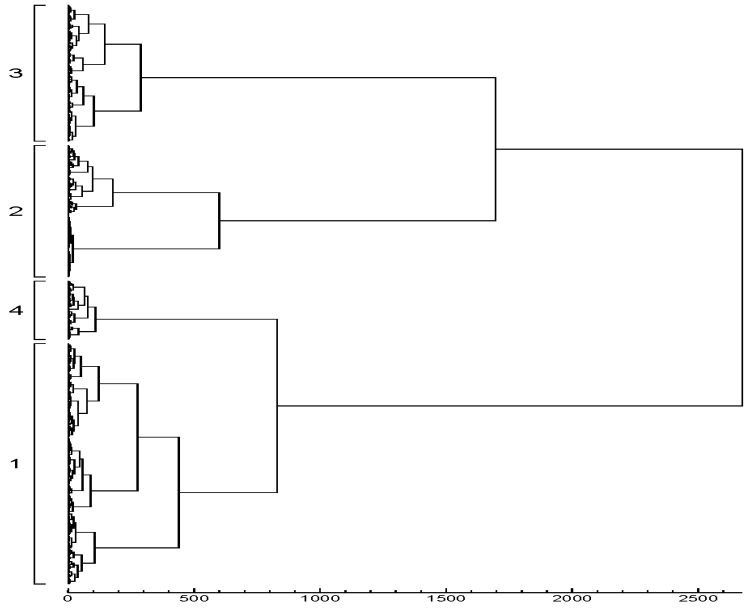
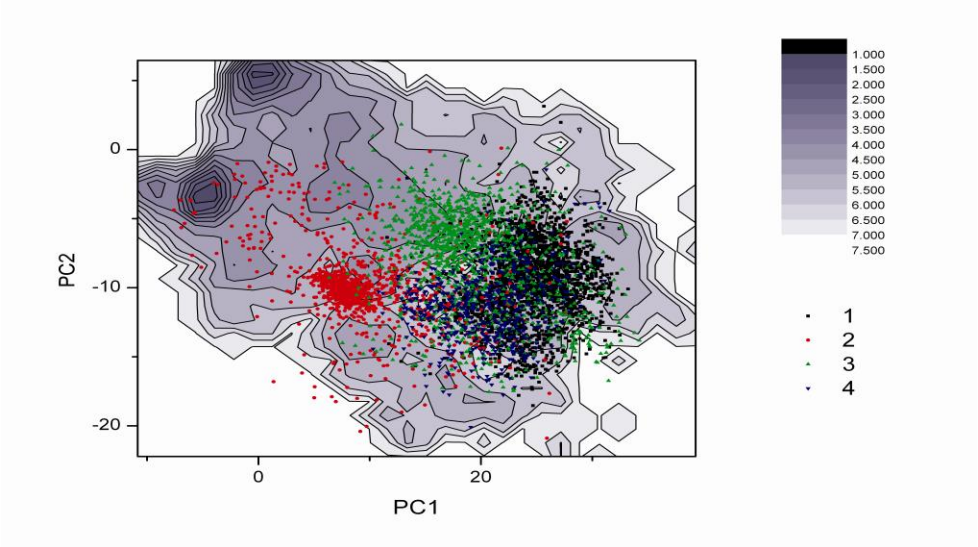


Figure 3.14: Cluster 3 representative structure of CLN025 with characteristic interactions displayed.

Further, clustering of the implicit solvent REMD trajectory structures was also carried out and the results are shown in Figures 3.15 to17 along with projections on the FES.

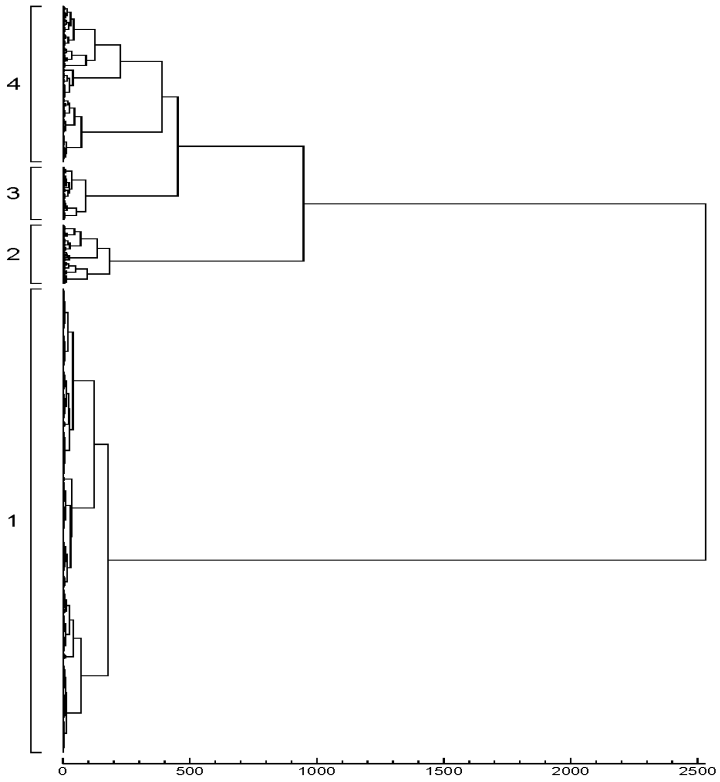


(a)

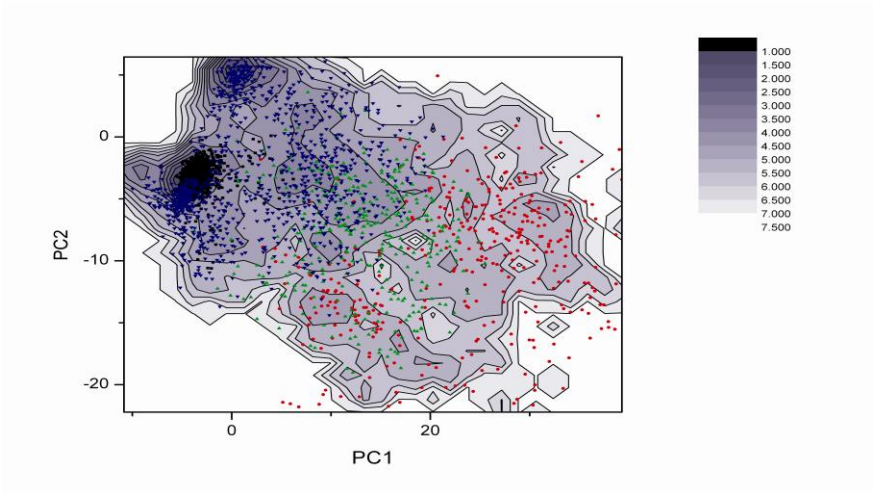


(b)

Figure 3.15: (a) Cluster histogram and (b) projection on the free energy surface of implicit solvent REMD simulation using ff94. The two minima in the top-left corner are not populated.

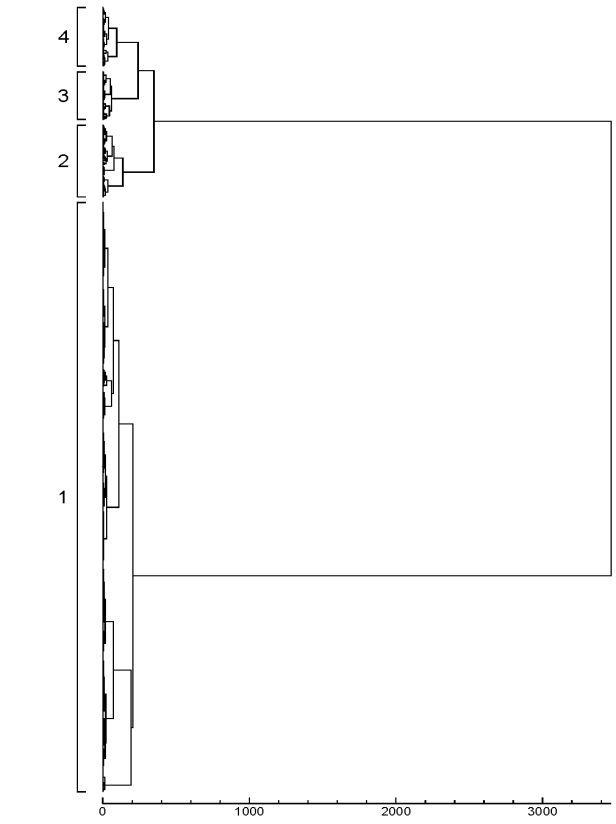


(a)

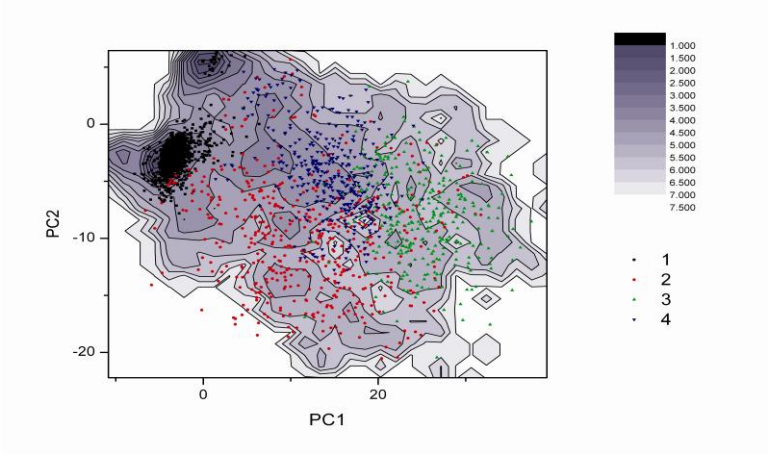


(b)

Figure 3.16: (a) Cluster histogram and (b) projection on the free energy surface of implicit solvent REMD simulation using ff96, showing one of the two minima well populated.



(a)



(b)

Figure 3.17: (a) Cluster histogram and (b) projection on the free energy surface of implicit solvent REMD simulation using ff99SB, showing one of the two minima to be well populated.

The clustering of the backbone using RMSD as the distance was able to reproduce the two minima found in free energy surfaces (FES). The characteristic interactions can be described for both minima. The difference between these minima came up from competition between hydrogen bond interactions. The competition can exist between intramolecular hydrogen bonding (in cluster 1 Tyr1.O is interacting with Tyr10.N while in cluster 3 it interacts with Trp9 (refer to Figure 3.18 for the numbering of residues) or between intramolecular and intermolecular hydrogen bonding (in cluster 1 Thr8.O interacts with Asp3.N while in cluster 3 Thr8.O interacts with a water molecule and Asp3.N interacts with Gly7.O).The participation of water molecules in these interactions explains why implicit solvent REMD simulations (Figures 3.15 to 17) were not able to reproduce the two minima found in explicit solvent REMD simulation in blue (Figure 3.12).

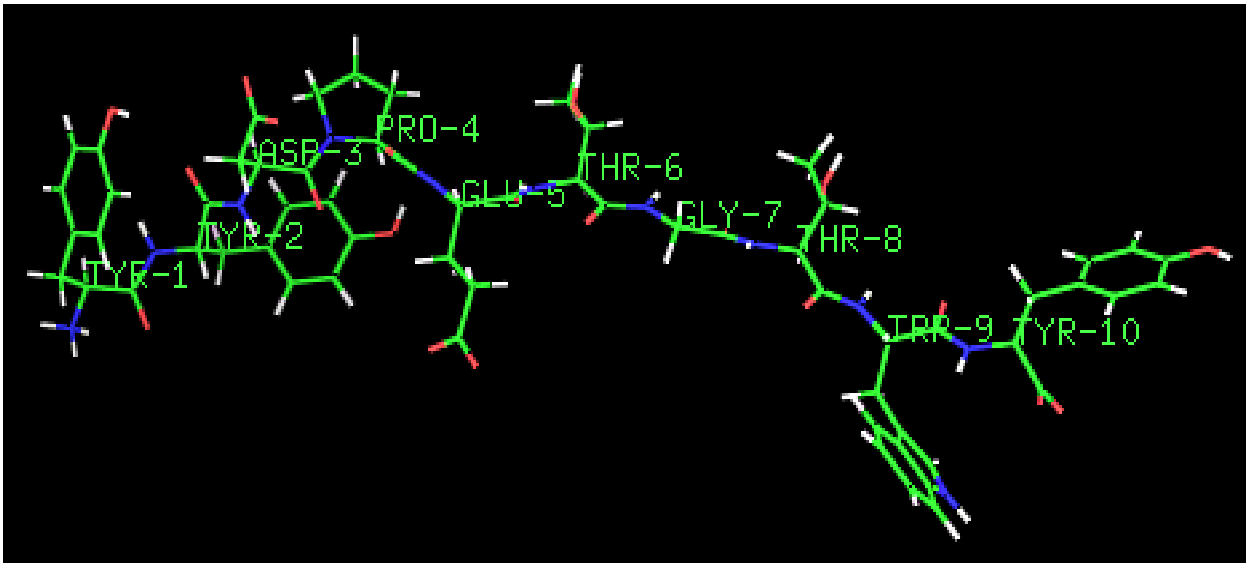


Figure 3.18: Structure of CLN025 in its extended conformation, showing all ten residue side chains.

3.4 Conclusions

This study revealed that implicit solvent REMD simulation using force field ff94 does not reproduce the explicit REMD simulation and the experimentally observable peptide conformations. On the contrary, the implicit solvent REMD simulations using force fields ff96 and modified ff99 (ff99SB) were able to reproduce the protein secondary structures, with trends consistent with explicit solvent REMD simulations and experimentally determined structures. Implicit REMD trajectories validate our findings that force fields ff96 and modified ff99 can be used to reproduce protein folding dynamics, and that force field ff94 is not suitable for this purpose.

CHAPTER 4

CASE STUDY I: MD SIMULATIONS OF METENKEPHALIN

4.1 Introduction

Enkephalins are the smallest neuropeptides with pain killing or opiate activity. They contain the five-residue homologous Tyr-Gly-Gly-Phe-X amino acid sequence. Neuropetides are small protein molecules composed of 3-100 amino acid residues, that have been localized to cell populations of the central and peripheral neurons **[42]**. Enkephalins are normal constituents of the brain and the gut and loose connective tissue **[43]**. They are also produced in some parts of the spinal cord that transmit pain impulses. In the spinal cord, enkephalins inhibit painful sensations by interacting with specific receptor sites on the sensory nerve endings. Nerve endings of the central nervous system (CNS) and the adrenal medulla release these naturally occurring morphine-like substances. Enkephalins are known to bind to opiate receptors and release controlled levels of pain and are therefore used in the treatment of pain **[44, 45]**.

Met-enkephalin (Tyr^1-Gly^2-Gly^3-Phe^4-Met^5) and Leu-enkephalin (Tyr^1-Gly^2-Gly^3-Phe^4-Leu^5) are homologous peptides differing in the fifth residue only. They are endogenous agonist for the receptors that are stimulated by opiate alkaloids. These peptides have multiple effects on the CNS, including the neuroendocrine hypothalamus **[46]**. Leu-enkephalin, in particular, controls gonadal function **[47]**. Met-enkephalin is involved in phenomena associated with modulated pain perception, regulation of memory and emotional conditions, food and liquid consumption and regulation of immunological system **[46, 48]**. It also has an impact in digestive system motility, gastric

as well as in pancreatic secretion and metabolism of carbohydrates **[48]**.

The 3D structures of the enkephalins were obtained using X-ray crystallography, circular dichroism and NMR spectroscopy **[49, 50, 51]**. The X-ray diffraction crystal structure of Met-enkephalin was found to consist of dimers forming antiparallel beta-sheets. Leu-enkephalin, on the other hand, crystallizes as a monohydrate that is isomorphous with the Met-enkephalin structure with respect to the beta-sheet, but different with respect to the tyrosine and phenylalanine side-chain conformations and water content. Previous NMR studies of the enkephalins indicated that these small peptides have no preferred structure in aqueous solution, but they may have structural preferences in other solvents like dimethyl sulfoxide (DMSO), though only the zwitterioinic form of the peptide forms a clear salt bridge **[52]**. Recent developments in NMR methodology have enabled determination of structures of peptides **[84]**.

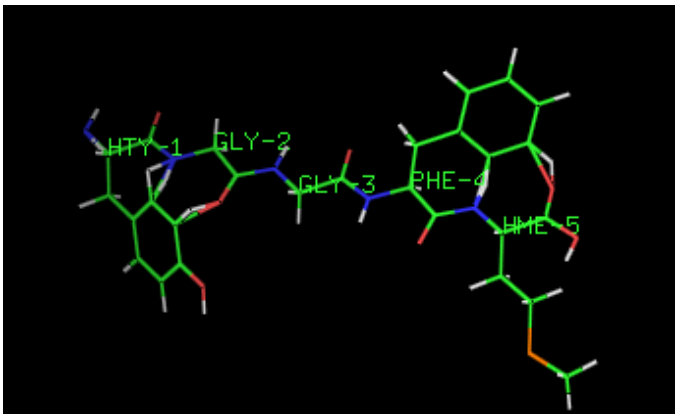


Figure 4.1: The structure of Met-enkephalin in its extended conformation.

For pharmaceutical applications of these neuropeptides there have been several attempts to describe their structure when bound to their native receptors **[53, 54]**. There

is sufficient consensus that the glycine residues adopt β-turns, with Aib analogues of the enkephalins displaying a 3_{10} helical conformation **[55]**. These pentapeptides (Met- and Leu-enkephalin) have substantial potential for the treatment of the pain sensation **[56]**. However, despite this promise, few opioid peptides have shown true clinical viability for the alleviation of centrally mediated pain. The leading factor for the observed inactivity of these peptides is the blood-brain barrier (BBB). The BBB acts as a metabolic and transport barrier, preventing the delivery of substances to the central nervous system (CNS) **[57]**.

Despite the existence of the NMR structures of the enkephalins, there is still a lot to be investigated about the conformational features of these pentapeptides in order to discover alternative analgesics to the conventional morphine, and to find novel compounds to deal with pain not alleviated by current therapeutics **[58]**. Although these enkephalins are very short, exploration of their dynamical properties can shine light regarding behaviour of longer polypeptides. Accordingly, the aim of this particular study was to establish whether molecular dynamics (MD) protocols can be successfully used to mimic peptide conformational features of Met-enkephalin, providing results similar to those obtained by experimental means such as NMR spectroscopy.

4.2 Computational Protocols

All calculations were carried out at the molecular mechanics level using the AMBER 8.0 **[26]** package and the force field ff96 parameter set. The details of the various steps of the simulation process in this project are provided in APPENDIX I.

Molecular Dynamics (MD)

The zwitterionic and the neutral forms of Met-enkephalin were studied using the standard ff96 force field libraries with the AMBER 8.0 programme **[26]**. The MD simulations of this peptide were carried out in water, dimethyl sulphoxide (DMSO), and methanol. To model the solvent molecules, pre-equilibrated boxes of water, DMSO and methanol were obtained from the AMBER 8.0 database **[26]**. In each of these three simulations, the extended conformation of the peptide was initially minimized within the SANDER module, using a dielectric constant of ε=1.0, and a cutoff of 10 Å. About 1500 cycles of steepest descent followed by the conjugated gradient method were applied until the rms distance between two consecutive structures was smaller than 0.001kcalmol^{-1}Å^{-1}. Thereafter the peptide was soaked in a rectangular box of the respective solvent molecules. The system was minimized by fixing the coordinates of the peptide and allowing the water to move. A constant pressure equilibration was carried out for 100 ps with a time step of 2 fs and a cutoff of 10 Å, and removing the restrictions on the peptide. Using a periodic boundary conditions and the Langevin thermostat, the temperature of the system was maintained at 300K as it is the same temperature used in NMR studies. This temperature control method uses Langevin dynamics with a collision frequency of unity. This was used in conjunction with the SHAKE algorithm to constrain the stretching of bonds involving hydrogen atoms. Thereafter, the Particle Mesh Ewald (PME) conditions with a tolerance for the Ewald sum of 10^{-5} were used in the production run from 200 ps onwards. In each case

trajectories were computed for 40 ns and snapshots were recorded every picosecond for further analysis.

4.3 Results and Discussion

The energy profile and the temperature profile of the systems were routinely monitored and the plots of typical trajectories are presented in Figures 4.2 and 4.3. Both the kinetic energy and potential energy started off higher but immediately stabilized throughout the 40 ns trajectories (Figure 4.2), suggesting that the trajectories were running normally without unwarranted upsets in the system and that subsequent

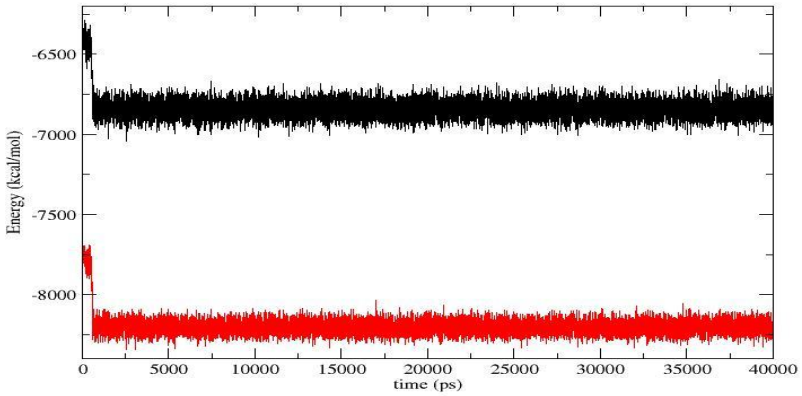


Figure 4.2: Typical energy plot for Met-enkephalin trajectories. Total energyand the kinetic energy are represented in black and red, respectively.

analyses were made based on reliable molecular dynamics trajectories. By the same token, the temperature profile of the systems all stabilized around 300 K throughout the sumulation (Figure 4.3). Further, the co-ordinate files of the systems were used to obtain the dihedrals of the peptide. The Ramachandran plots of secondary structural elements of the peptide in different solvent models are presented in Figure 4.4 to 4.6.

In terms of protein secondary structure, β-turns are regions of the polypeptide having a hydrogen bond from one main chain carbonyl oxygen to the main chain N-H group three residues along the chain (i.e. an i to i+3 interaction). These β-turns are divided into Types I, II and III based on the phi and psi angles of the amino acid residues at positions i+1 and i+2.

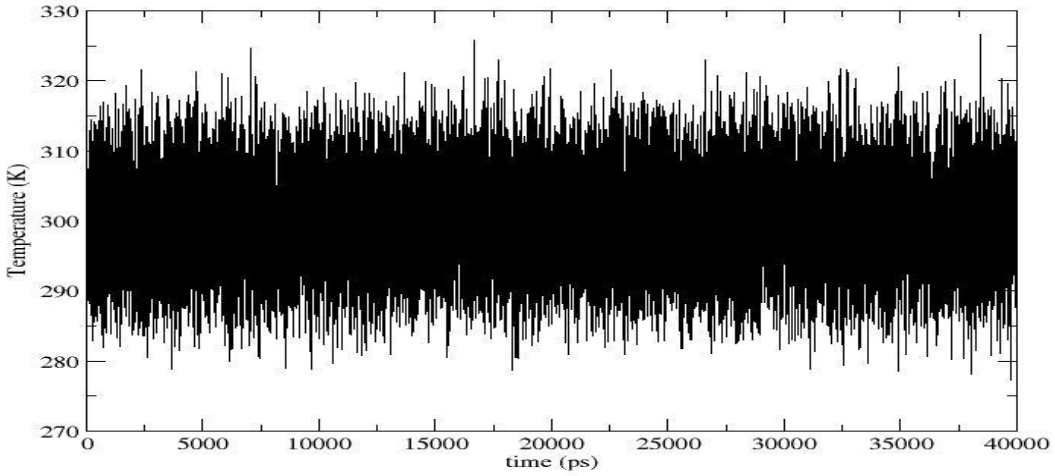


Figure 4.3: Typical temperature plot for Met-enkephalin trajectories.

Types I and II are the most common β-turns and they differ in the orientation of the peptide bond between residues (i+1) and (i+2). Hence the torsion angles for the residues (i+1) and (i+2) in these two types of β-turn lie in distinct regions of the Ramachandran plot. Specifically, the (i+1) residue of type I lies in the right handed α-helix region while its (i+2) residue lies towards the β-sheet region of the Ramachandran plot. On the other hand, (i+1) residue of type II β-turn lies in the β-sheet region while its corresponding (i+2) residue lies in the left handed α-helix region of the Ramachandran plot. Type III β-turn is simply a single turn of the 3_{10}-helix and coincides with the right handed α-helix region.

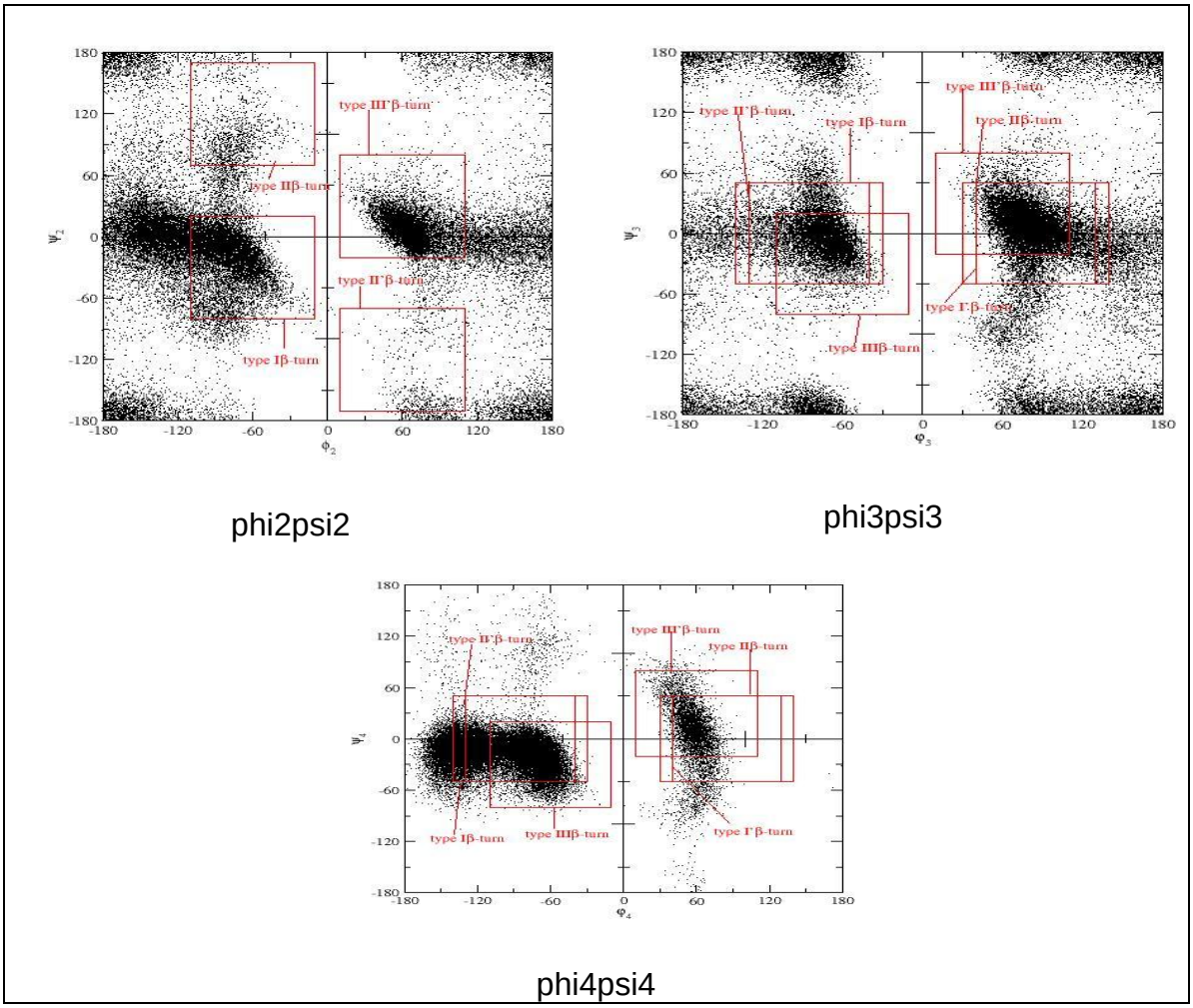


Figure 4.4: Ramachandran plots dihedrals obtained for Met-enkephalin simulations in a water box.

Figures 4.4 to 4.6 depict the fact that the trajectories were populated with Type I, Type II and Type III conformers and their mirror images. Specifically, in Figure 4.4, it can be observed that water enabled the Type I, II and III conformers to form to almost the same extent as their mirror images. The mirror images of the β-turns occupy unfavourable regions of the Ramachandran plot, and their presence in this trajectory ran using water as an implicit solvent means that the structure of the peptide (Met-enkephalin) is not stable enough in this particular solvent and has a tendency to

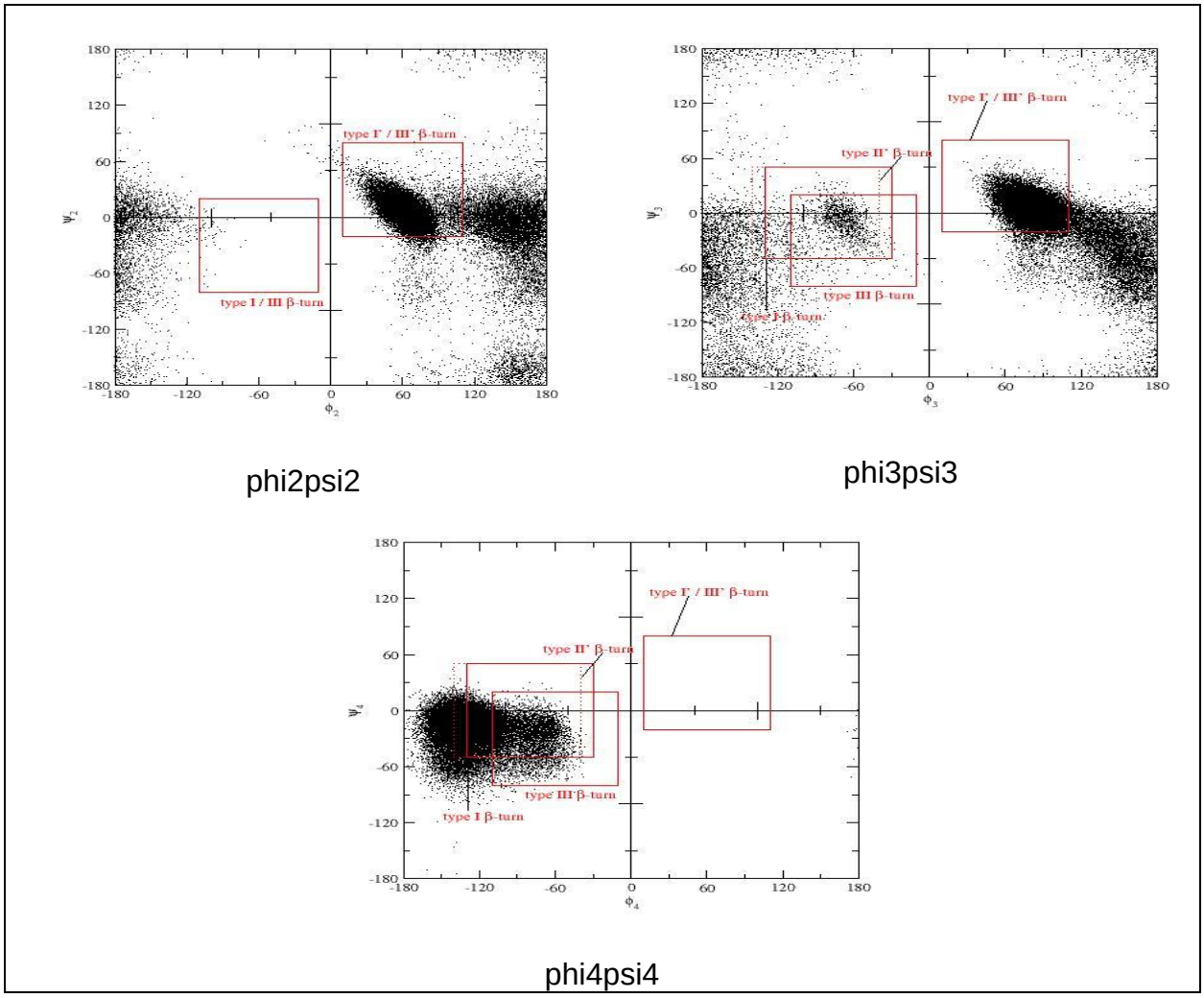


Figure 4.5: Ramachandran plots dihedrals obtained for Met-enkephalin simulations in a methanol box.

oscillate between structured and unstructured conformations. Methanol, on the other hand (Figure 4.5) supported the formation of the Type I and III β-turns and their mirror images, suggesting that the trajectory samples favourably structures belonging to the right handed α-helix and the $\alpha 3_{10}$-helix regions. Moreover, Figure 4.6 indicates that DMSO trajectory favoured the formation of the Type I, II and III β-turns over their mirror images, predicting that the trajectory is populated with more stable right handed α-helix and $\alpha 3_{10}$-helix secondary structures.

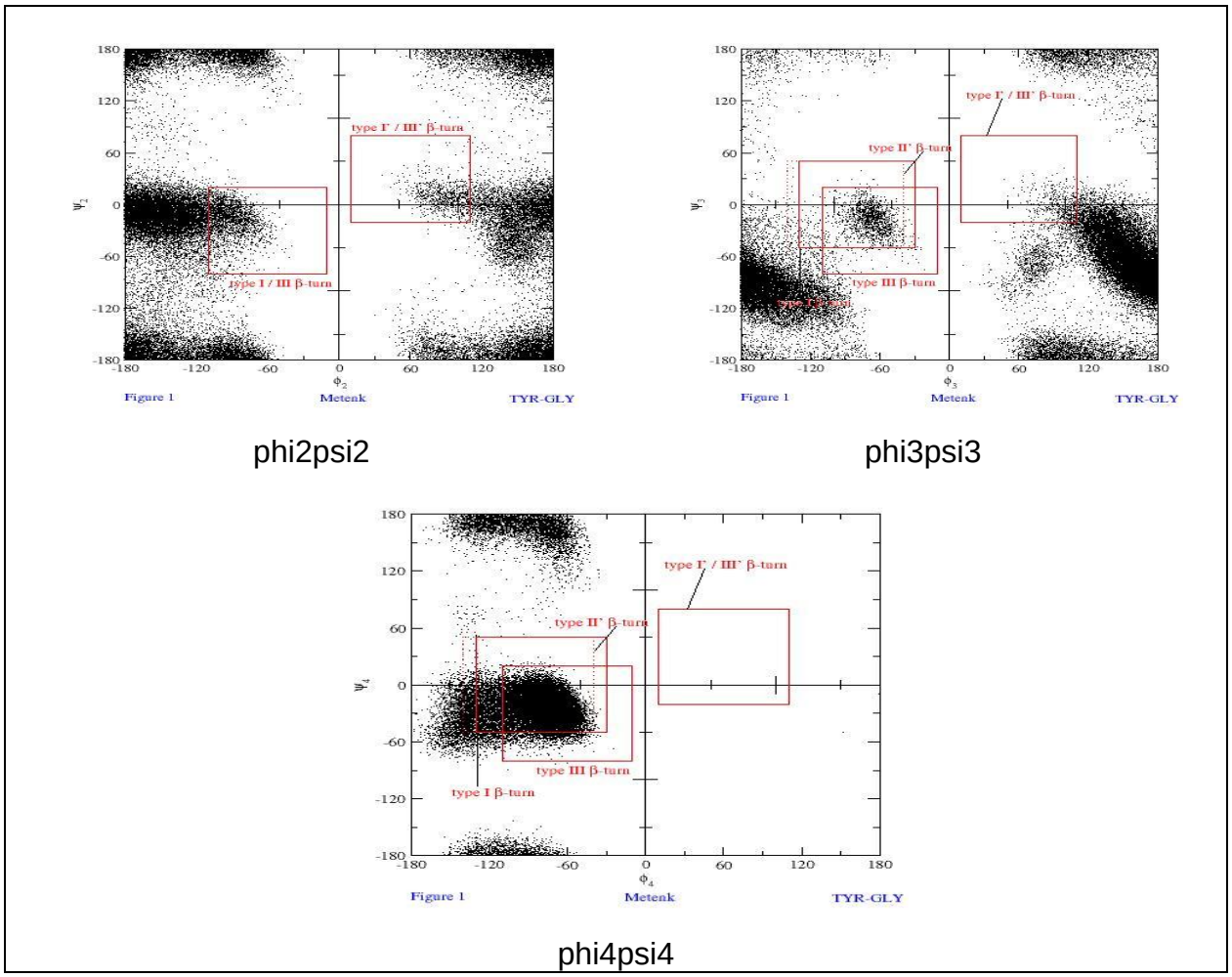


Figure 4.6: Ramachandran plots dihedrals obtained for Met-enkephalin simulations in a DMSO box.

Small peptides such as Met-enkephalin are an important class of molecules to study using computational methods in an attempt to solve protein folding problems. Moreover, since it is extremely difficult to obtain experimental information about short peptides due to different conformations they adopt whilst searching the conformational space, the molecular dynamics method is used to sample the energy landscape of the peptide. The current results demonstrate that the Met-enkephalin consists of

conformers having Type I, Type II and Type III β-turns, especially when simulated using DMSO as an implicit solvent. Due to the short length of Met-enkephalin it is not possible for this pentapeptide to assume an overall helical secondary structure; instead the peptide trajectories are populated with helical and non-helical conformers, suggesting that the energy barrier between these types of conformers is small. The conformations of Met-enkephalin as obtained using MD simulation are consistent with those characterized using experimental NMR as reported in literature **[51]**.

The Ramachandran plots (Figures4.4 to 4.6) further show that this peptide sampled the conformational space belonging to the α-helix region of the phi/psi combination, i.e right-handed α-helix. However, the glycine residues and the phenylalanine residues also sample the left-handed helix region (first quadrant) of the Ramachandran plot, which is to be expected from the chemical composition of these residues. The regions sampled by the individual omega, phi and psi angles were plotted separately (not shown) and are consistent with the conformations depicted in the Ramachandran plots from trajectories obtained for the different solvents. To account for the observed conformations, the hydrogen bond interactions associated with the different MD trajectories need to be examined.

Hydrogen Bond Analysis

Hydrogen bond (HB) analysis performed for Met-enkephalin in zwitterion and neutral form in water, DMSO and MeOH solvent environments, and the results are presented in Table 4.1 and 4.2. HB geometrical criterion was used, whereby: A··B distance is equal or less than 3.2 Å and angle is determined by HAB (∠ HAB) is equal or less than 36°. In this case, A is donor (A-H, i.e. N-H) and B is acceptor (B, i.e. O of carbonyl group). In order to obtain significant results, only the hydrogen bond with a percentage of existence equal or greater than 1.0% is considered during the simulation time.

The general structuring degree refers to backbone-backbone hydrogen bond. One overall indicator is the percentage of structures collected that have one or more backbone-backbone hydrogen bond. Another indicator is the average of the number of backbone-backbone hydrogen bond interactions per structure (ps) collected over the duration of the MD trajectories.

Table 4.1: Percentage of structures with backbone-backbone hydrogen bonds in zwitter-ionic and neutral Met-enkephalin in the presence of water, DMSO, and MeOH solvents

Met-enkephalin	Water	DMSO	MeOH
zwitter-ionic form	36 %	84 %	87 %
Neutral form	-	54 %	53 %

Table 4.2: Average number of backbone-backbone hydrogen bonds per structure (ps) collected for zwitter-ionic and neutral Met-enkephalin in the presence of water, DMSO, and MeOH solvents

Met-enkephalin	Water	DMSO	MeOH
zwitter-ionic form	0.5	1.4	1.5
Neutral form	-	0.6	0.7

From Table 4.1 it can be observed that the zwitterionic form of Met-enkephalin simulated in MeOH has the highest percentage of structures collected that have one or more backbone-backbone hydrogen bond (87%). This percentage decreased to 36%

when water was used as a solvent, but increased to 84% in the case of DMSO. The highest percentage of structures collected that have one or more backbone-backbone hydrogen bond in the neutral form of Met-enkephalin was observed in DMSO (54%). In MeOH this percentage dropped by a percentage point to 53%, whereas no structures with backbone-backbone hydrogen bonding were observed in water. The results for the average number of backbone-backbone hydrogen bonds per structure (ps) collected (Table 4.2) constitute a similar pattern with those for the percentage of backbone-backbone hydrogen bonds (Table 4.1). However, one would have expected DMSO to have higher percentage and number since it is a better structuring solvent than methanol.

The backbone-backbone hydrogen bond with equal or more than the 1% of existence that yielded defined secondary structure conformations throughout the simulation time are presented in the Tables 4.3 and 4.4 for the zwitterionic and the neutral forms of Met-enkephalin (refer to Figure 4.1 for residue numbering). Specifically, Tables 4.3 and 4.4 reflect the type of interaction and indicate the existence percentage of the conformations in the different solvents; and identify the types of secondary structures formed. Among other observations, these results demonstrate that reverse turns are present in the zwitterionic Met-enkephalin in DMSO, whereas they were absent in the neutral Met-enkephalin in the same solvent (DMSO). On the contrary, no β-turns were observed in the zwitterionic Met-enkephalin whilst they were found in the neutral Met-enkephalin in the presence of DMSO. However, in the presence of

methanol more β-turns are formed. According to these results, water did not favour the formation of reverse turns in the zwitter-ionic form of Met-enkephalin.

Table 4.3: Structural conformations formed by backbone-backbone hydrogen bond interactions and their percentages in different solvents for the zwitter-ionic form of Met-enkephalin (Residues are numbered as in Figure 4.1)

	Definition	**2^0 Structure**	**Water**	**DMSO**	**MeOH**
1	**CO(Tyr^1)···NH(Gly^3)**	**γ-turn**	3.1	-	3.4
2	CO(Gly^2)···NH(Phe^4)	**γ-turn**	6.8	1.4	1.4
3	CO(Gly^3)···NH(Met^5)	**γ-turn**	4.4	-	-
4	CO(Tyr^1)···NH(Phe^4)	**β-turn**	9.3	-	**26.9**
5	CO(Gly^2)···NH(Met^5)	**β-turn**	4,0	-	-
6	CO(Tyr^1)···NH(Met^5)	**C-13**	7.6	-	**31.8**
	Reverse turns				
1	CO(Met^5)···NH(Gly^3)	**C-11**	-	**16.5**	-
2	CO(Met^5)···NH(Gly^2)	**C-14**	1.8	**20.3**	-
3	CO(Met^5)···NH(Tyr^1)	**C-17**	7.7	**17.5**	**39.3**
	Terminal turns with O^{2-} of Met^5-term				
1	COXT(Met^5)···NH(Gly^3)	**C-11 term**	-	**48.6**	-
2	COXT(Met^5)···NH(Gly^2)	**C-14 term**	1.3	**15.2**	
3	COXT(Met^5)···NH(Tyr^1)	**C-17 term**	5.8	**15.5**	**51.2**
	Total characteristic hb backbone-backbone		***10***	***7***	***6***

Table 4.4: Structural conformations formed by backbone-backbone hydrogen bond interactions and their percentages in different solvents for the neutral form of Met-enkephalin (Residues are numbered according to Figure 4.1)

	Definition	Description	DMSO	MeOH
1	**CO(Tyr^1)···NH(Gly^3)**	**γ-turn**	6.1	7.6
2	CO(Gly^2)···NH(Phe^4)	**γ-turn**	6.2	6.9
3	CO(Gly^3)···NH(Met^5)	**γ-turn**	-	4.6
4	CO(Tyr^1)···NH(Phe^4)	**β-turn**	2.6	**17.6**
5	CO(Gly^2)···NH(Met^5)	**β-turn**	**40.4**	**14.5**
6	CO(Tyr^1)···NH(Met^5)	**C-13**	-	**11.4**
	Reverse turns			
1	CO(Met^5)···NH(Gly^3)	**C-11**	-	-
2	CO(Met^5)···NH(Gly^2)	**C-14**	-	1.2
3	CO(Phe^4)···NH(Tyr^1)	**C-14**	-	1.2
4	CO(Met^5)···NH(Tyr^1)	**C-17**	-	1.5
	Terminals turns with -OH of Met^5-term			
1	CO(Gly^2)···OH(Met^5)	**C-10**	-	1.4
2	CO(Tyr^1)···OH(Met^5)	**C-13**	-	1.3
	Total characteristic hb backbone-backbone		***4***	***11***

A higher level of classification of secondary structure conformations involves combining two or more such structures to form motifs. The two main types of secondary structure motifs are defined as follows:

2_7 Ribbon ≡ two consecutive γ-turns;

and 3_{10} helix≡ two consecutive β-turn.

Using the above definition of motifs, for Met-enkephalin in the zwitterionic form, none of the defined secondary motifs were formed at significant percentages. This means that the percentage motifs abundance in all solvents was less than 1%. On the other hand, secondary structure motifs were observed in the neutral form of Met-enkephalin. In methanol, there was 1.8% presence of motifs over all the simulation time, characterized by the simultaneous presence of $CO(Tyr^1)\cdots NH(Phe^4)$ and $CO(Gly^2)\cdots NH(Met^5)$ forming a 2_7 ribbon.

The existence of competition-cooperation interactions and folded bonds for backbone acceptors and donors was also investigated in the different simulations. In this regard, the percentage of simultaneous presence of a backbone donor or acceptor, in more than one backbone-backbone hydrogen bond interaction, over the entire simulation was determined for both the zwitterionic and neutral form of Met-enkephalin, and the results are presented in Table 4.5 and 4.6. These interactions are considered to be present if their existence percentage is equal or greater than 1%.

Table 4.5: Percentage of simultaneous presence of a backbone donor or acceptor in more than one backbone-backbone hydrogen bond interaction for the zwitter-ionic form of Met-enkephalin in different solvents

Acceptor/Donor	**Water**
$CO(Tyr^1)$	**4.2 % bifurcated:** $CO(Tyr^1)\cdots NH(Phe^4)$ & $CO(Tyr^1)\cdots NH(Met^5)$
	DMSO
$CO(Met^5)$	**1.2 % trifurcated:** $CO(Met^5)\cdots NH(Gly^3)$ & $CO(Met^5)\cdots NH(Gly^2)$ & $CO(Met^5)\cdots NH(Tyr^1)$
	13.1 % bifurcated: $CO(Met^5)\cdots NH(Gly^3)$ & $CO(Met^5)\cdots NH(Gly^2)$
	2.4 % bifurcated: $CO(Met^5)\cdots NH(Gly^3)$ & $CO(Met^5)\cdots NH(Tyr^1)$
	4.1 % bifurcated: $CO(Met^5)\cdots NH(Gly^2)$ & $CO(Met^5)\cdots NH(Tyr^1)$
$COXT(Met^5)$	**1.0 % trifurcated:** $COXT(Met^5)\cdots NH(Gly^3)$ & $COXT(Met^5)\cdots NH(Tyr^1)$ & $COXT(Met^5)\cdots NH(Gly^2)$
	6.7 % bifurcated: $COXT(Met^5)\cdots NH(Gly^3)$ & $COXT(Met^5)\cdots NH(Tyr^1)$
	3.4 % bifurcated: $COXT(Met^5)\cdots NH(Gly^3)$ & $COXT(Met^5)\cdots NH(Gly^2)$
	4.0 % bifurcated: $COXT(Met^5)\cdots NH(Tyr^1)$ & $COXT(Met^5)\cdots NH(Gly^2)$
$NH(Gly^2)$	**1.5 % bifurcated:** $CO(Met^5)\cdots NH(Gly^2)$ & $COXT(Met^5)\cdots NH(Gly^2)$
NH(Tyr1)	**6.1 % bifurcated:** $CO(Met^5)\cdots NH(Tyr^1)$ & $COXT(Met^5)\cdots NH(Tyr^1)$
	MeOH
$CO(Tyr^1)$	**1.2 % bifurcated:** $CO(Tyr^1)\cdots NH(Gly^3)$ & $CO(Tyr^1)\cdots NH(Met^5)$
	16.7 % bifurcated: $CO(Tyr^1)\cdots NH(Phe^4)$ & $CO(Tyr^1)\cdots NH(Met^5)$
$NH(Tyr^1)$	**6.0 % bifurcated:** $CO(Met^5)\cdots NH(Tyr^1)$ & $COXT(Met^5)\cdots NH(Tyr^1)$

Table 4.6: Percentage of simultaneous presence of a backbone donor or acceptor in more than one backbone-backbone hydrogen bond interaction for the neutral form of Met-enkephalin in different solvents

Acceptor/Donor	**DMSO**
$CO(Tyr^1)$	**7.6 % bifurcated:** $CO(Tyr^1)\cdots NH(Phe^4)$ & $CO(Tyr^1)\cdots NH(Met^5)$

The simultaneous presence of a backbone donor or acceptor in more than one backbone-backbone hydrogen bond interaction was detected in the zwitterionic form of

Met-enkephalin in different solvents (Table 4.5) and only once this scenario was observed in the neutral form of Met-enkephalin as shown in Table 4.6. This means that the pentapeptide exists relatively more abundantly in the zwitterionic form than the neutral form in solution.

Finally, the backbone-side chain hydrogen bond interactions were determined for both the zwitterionic and the neutral Met-enkephalin in the different solvents. For all the three solvents, no significant data (< 1.0%) was obtained for the existence of backbone-side chain hydrogen bonding for the zwitterionic form of Met-enkephalin. For the neutral form of Met-enkephalin, 1.2% existence of the SH(Met^5)···NH(Met^5) bond was detected in the MeOH solvent only. This is an intra-residue bond, as it occurs within Methionine and does not involve another residue.

4.4 Conclusions

Molecular dynamics simulations of Met-ekephalin show that the DMSO and methanol trajectories of the pentapeptide favoured the formation of a right handed α-helix and 3_{10}-helix secondary structures, and that these structures are not energetically stable enough in water. Moreover, the zwitterionic form of Met-enkephalin has a very high percentage of backbone-backbone hydrogen bond interactions in DMSO and MeOH (84% and 87%, respectively), but very low in water (36%). This means that the peptide has a high tendency to form intra-hydrogen bonds in the presence of DMSO and MeOH which typically yield α-helices and β-sheets structures for peptides. However, water did stabilize these interactions most probably due to its polarity. Thus,

both DMSO and MeOH act as structuring solvents in this case. The majority of the percentage of simultaneous presence of a backbone donor or acceptor, in more than one backbone-backbone hydrogen bond was detected in the zwiterionic Met-enkephalin. Only one such interaction was observed in the neutral Met-enkephalin. The secondary structures characterized in this study using MD simulations follow a similar trend with experimental NMR results reported in literature **[35,36,65]**.

CHAPTER 5

CASE STUDY II: MD STUDIES OF VIP AND PACAP27: EVALUATION OF THE GB MODELS AND FORCE FIELDS

5.1 Introduction

Vasoactive Intestinal Peptide (VIP) having a sequence His^{1}-Ser^{2}-Asp^{3}-Ala^{4}-Val^{5}-Phe^{6}-Thr^{7}-Asp^{8}-Asn^{9}-Tyr^{10}-Thr^{11}-Arg^{12}-Leu^{13}-Arg^{14}-Lys^{15}-Gln^{16}-Met^{17}-Ala^{18}-Val^{19}-Lys^{20}-Lys^{21}-Tyr^{22}-Leu^{23}-Asn^{24}-Ser^{25}-Ile^{26}-Leu^{27}-Asn^{28}-NH_2) is closely related to Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP), both of which are hormones belonging to the glucagon/secretin familiy that exhibit neuroprotective action. However, VIP is a 28 residue-long peptide whereas PACAP presents two amidated forms that exhibit the same pharmacological profile: a 38 residue peptide and a 27 N-terminal residue fragment, known as PACAP27. The sequence, His^{1}-Ser^{2}-Asp^{3}-Gly^{4}-Ile^{5}-Phe^{6}-Thr^{7}-Asp^{8}-Ser^{9}-Tyr^{10}-Ser^{11}-Arg^{12}-Tyr^{13}-Arg^{14}-Lys^{15}-Gln^{16}-Met^{17}-Ala^{18}-Val^{19}-Lys^{20}-Lys^{21}-Tyr^{22}-Leu^{23}-Ala^{24}-Ala^{25}-Val^{26}-Leu^{27}-NH_2, shares a high sequence homology with that of VIP.

At least three receptors of the Group II secretin family of the G-protein coupled receptors (GPCR): $VPAC_1$, $VPAC_2$ and PAC_1, have been identified to mediate their action with different degree of affinity to VIP and PACAP. Thus, whereas the PACAP-specific PAC_1 receptor displays a much higher affinity for the two forms of PACAP (IC50, 1nM) than for VIP (IC50, 1000 nM), the $VPAC_1$ and $VPAC_2$ receptors recognize PACAP38, PACAP27 and VIP with similar affinity in the nanomolar range **[59]**. The spectrum of functionalities exhibited by these peptides has prompted several authors to

highlight its potential as suitable biological targets for the treatment of inflammatory and neurodegenerative diseases **[59, 60, 61, 62, 63]**. However, despite the extensive investigation carried out in previous years there is a scarce number of ligands with agonistic or antagonistic profile available at present, thus hampering a deeper understanding of their role in the signaling processes and in the development of new therapeutic agents.

Considering the fact that PACAP27 and VIP have significant sequence homology, their selectivity for the three receptors characterized thus far can be ascribed to subtle differences in their sequence and conformational profile **[64, 65]**. Structure–activity relationships (SAR) studies carried out on intact VIP, its fragments, and synthetic analogues suggest that the integrity of VIP is necessary to exert its full biological activity **[66]**. Specifically, the C-terminal fragment of the peptide has been shown to be required for high affinity antagonism, whereas the N-terminus (residues 1–11) can be considered the locus of its biological activity. Moreover, His1 has been identified to be critical for receptor activation **[66]**, since the fragment VIP (2–28) exhibits an antagonistic profile. With regard to PACAP27 and PACAP38, residues 1-6 have been shown to be necessary for agonistic activity, since fragments lacking this segment exhibit an antagonistic profile. Finally, it has been suggested that the selectivity for the different receptors exhibited by PACAP and VIP resides principally in the features of their respective fourth and fifth residues **[64, 65]**.

With regard to the structure exhibited by these peptides in solution, two-dimensional (2D) nuclear magnetic resonance (NMR) studies on VIP in a 40%

trifluoroethanol (TFE) water mixture **[67]** suggests that the peptide consists of two amphipatic helices between residues 7–15 and 19–27 respectively connected by a region with an undefined structure. On the other hand, 2D NMR studies conducted on some VIP fragments **[68]** suggest that the N-terminal region (residues 1–6) exhibits a folded conformation. With regard to the structure of PACAP27 two different conformations have been proposed in solution. In the presence of TFE, the peptide exhibits a disordered N-terminal domain followed by an α-helix spanning residues 9 to 26 with a discontinuity between residues 20 and 21. The N-terminal end of the polypeptide has a free amine group, whereas the C-terminal end of the polypeptide has a free carboxyl group. In contrast, in a 25% methanol water mixture the conformation adopted by PACAP27 is composed of three distinct regions: a type II β-turn at residues 9-12, and two α-helices spanning residues 12-20 and 22-24 **[69]**. As a result, minor conformational changes between PACAP27 and VIP are likely to contribute to the selective recognition of the receptors.

One of the broader goals of this study was to describe the conformational features of these peptides using computational methods focused mainly on establishing the conformational features of the N-terminus of VIP using simulated annealing **[70]**. Accordingly VIP and PACAP27 were subjected to a molecular dynamics simulation with the aim to explore their conformational features, and to compare the results available from NMR experiments.

In order to provide a deeper insight into the structural features of these peptides, their folding process in solution was explored using molecular dynamics simulations

(MD) starting form their extended conformations. Long enough MD simulations have been proven to be an adequate tool to obtain the folded conformation of a peptide in solution, providing a wealth of information about the conformational profile of peptides **[2, 4]**. For instance, previous efforts of our research group provided interesting results on the conformational features of substance P (another helical peptide), starting from an extended conformation using MD, trajectories in implicit and explicit solvents in the presence of water and methanol molecules **[71, 72]**.

This study was carried out in two different steps. First, a more methodological study was applied to VIP, which served as a model to determine the best computational protocol that fits better with the experimental results available. Secondly, these results were used to run a MD simulation on PACAP27. Initially, a comparison between the conformational profile obtained from the analysis of four different 200 ns MD simulations with the available NMR results was made using VIP as a model peptide. Specifically, MD simulations were carried out using two implicit models of the solvent based on the Generalized Born (GB) approximation: Onufriev, Bashford and Case (GB^{OBC}) **[41]** and the Hawkins, Cramer and Truhlar (GB^{HCT}) **[73, 74]** implementations. In addition, the performance of the two different force fields was tested: AMBER ff99 and ff99SP. The former is a reparametization of ff96 **[75]**, whereas the latter is a modification of the force field ff99 as suggested by Sorin and Pande **[76]**. They removed the phi (φ) barriers in the AMBER force field ff99 by employing the original AMBER force field ff94 φ torsion potential with the purpose of better reproducing the experimental thermodynamics and kinetic data for the helical conformation.

An advantage of performing MD simulations using an implicit solvent model instead of explicitly representing solvent molecules is to provide a faster sampling of the conformational space. The delimitation is that simulations of this kind do not sample the conformations stabilized through interactions with solvent molecules, as a consequence more compact structures are obtained. Bearing this in mind, these simulations can be considered appropriate for comparison of conformational profiles obtained from experimental results of NMR studies performed in solvents such as water and TFE. However, it should be taken into account that TFE can be considered as a structuring solvent, known to induce and stabilize α-helical conformations in peptides, in good correlation with the intrinsic propensities of their sequences **[77]**.

5.2 Computational Protocols

VIP and PACAP27 were constructed using the leap module in AMBER 8.0 **[26]**. The N-termini were protonated (NH_3^+) and the C-terminal was amidated ($CONH_2$) for both VIP and PACAP27. Each initial structure in its extended conformation was energetically minimized with a convergence criterion of 0.005 $kcal.mol^{-1}$ $Å^{-1}$, the non-bonded cutoff of 99 Å and the direct sum of non-bonded list was updated every 10 steps. After energy minimization, structures were equilibrated for 10 ps by gradually increasing the temperature from 0 to 300 K.

Two different force fields were used to run the calculations: ff99 and ff99SP **[76]**. The salt concentration was set to 0.2 M in order to mimic physiological conditions. The dielectric constant around the peptide (internal dielectric constant) was ε=1 and the

external dielectric constant of ε=78.5, corresponding to water. More precisely, the solvent was represented with two continuum generalized Born models: the pairwise GB model introduced by Hawkins, Cramer and Truhlar (GB^{HCT}) **[73, 74]** and the model presented by Onufriev, Bashford and Case (GB^{OBC}) **[41]**. A total of 6 trajectories of 200 ns were run. Four trajectories for the peptide VIP were run with the two different force fields ff99 and ff99SP separately using the GB implementations GB^{HCT} and GB^{OBC}. Accordingly, two trajectories for PACAP 27 were run using ff99 and ff99SP using the GB^{OBC} model. After minimization the structure was equilibrated at a temperature of 300 K, with the SHAKE algorithm used to constrain hydrogen bonds. MD runs were started under GB conditions at constant temperature using weak coupling at 300 K. Bonds involving hydrogen atoms were constrained with no cutoff for non-bonded interactions. Snapshots were recorded every picosecond for later analysis. Bond distances and dihedral angles were obtained from the MD trajectories using relevant *ptraj* module of AMBER and analyzed using in-house programs. The bond distances from the trajectories were compared to intensities of Nuclear Overhauser Effects (NOEs) obtained from NMR experiments **[67, 78].**

Classification of structures (CLASICO)

Dihedral angles for the backbone were obtained from the MD trajectories using the *ptraj* module of AMBER. Subsequently, a modified version of the CLASICO program described in Corcho *et al*. **[71]** was used to study the percentage of conformational motifs present for the 200 000 structures obtained in each of the MD trajectories. The procedure is based on a division of the Ramachandran plot into different regions,

labelled with a one-letter code. From the dihedral angles of residues, conformations can be labelled by a string of letters. In a second step, using a two- or three-letter window, strings were classified into conformational patterns following a set of rules **[71]**. The new strings of conformational motifs are called patterns. This procedure permits the identification of different conformational patterns attained by the peptide, as well as comparing differences in conformational spaces sampled using different computational methods.

5.3 Results and Discussion

As described in the Computational Protocols section, four 200 ns MD trajectories of VIP were computed using the extended conformation as starting structure and using two different force fields (ff99 and ff99SP) and two different GB implementations (GB^{HCT} and GB^{OBC}). Trajectories were labelled as VIP-ff99-GB^{HCT}, VIP-ff99SP-GB^{HCT}, VIP-ff99-GB^{OBC} and VIP-ff99SP-GB^{OBC}, respectively.

For every snapshot in each trajectory, conformational patterns were assigned using an in-house programme called CLUSTERIT **[79]**. Therefore, after 200 000 snapshots for the four different trajectories were run, designated: VIP-ff99-GB^{HCT}, VIP-ff99SP-GB^{HCT}, VIP-ff99-GB^{OBC} and VIP-ff99SP-GB^{OBC}, 95 503, 119 508, 178 934 and 191 229 different patterns were identified. A representation of the evolution of the new patterns sampled along the trajectory for the different trajectories is shown in Figure 5.1. These plots provide an indication of the efficiency achieved in exploring the conformational space of these peptides with the different sets of conditions used. The

simulations using the GB^{HCT} approximation seem to sample the conformational space by hopping, meaning that the peptide gets trapped in lower regions (one of the many local minima) of the conformational space at certain intervals, whereas the GB^{OBC} approximation appears to be less restrictive in the sampling of the conformational space, evidenced by new patterns explored at a steady pace. With regard to force fields used, the ff99SP yields a higher number of patterns.

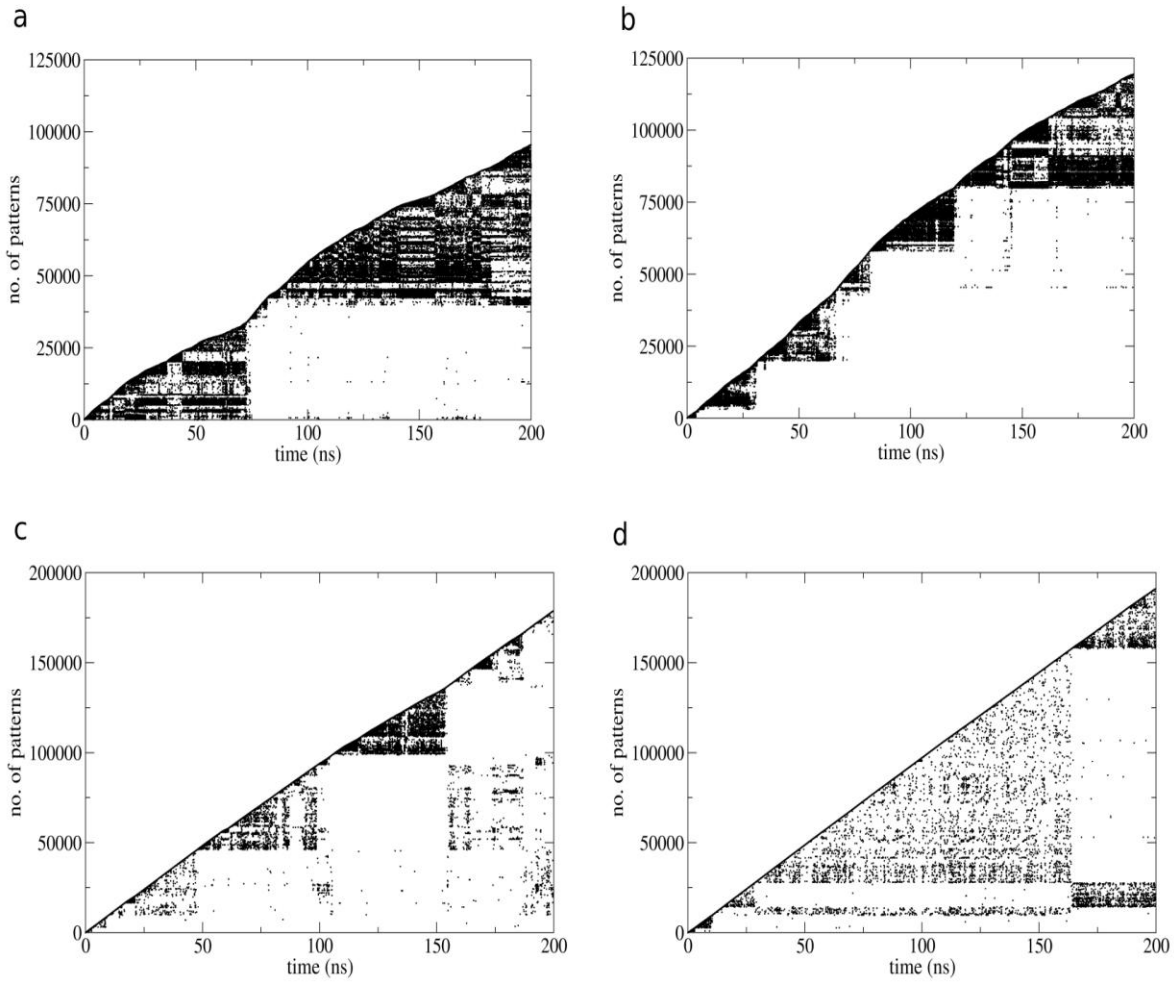


Figure 5.1: The structures obtained in the four VIP trajectories and classified using the CLASICO programme. (a) VIP-ff99-GB^{HCT} (b) VIP-ff99SP-GB^{HCT} (c) VIP-ff99-GB^{OBC} (d) VIP-ff99SP-GB^{OBC}. Each new combination of conformational motifs is a pattern, and the evolution of new patterns can be followed along the trajectories reported in the present work.

Figures 5.2a-2d show the conformational motifs sampled along each of the four trajectories using a three residue window. Specifically, Figure 5.2a shows the profile obtained in the VIP-ff99-GB^{HCT} trajectory, where the peptide exhibits a predominently helical propensity between residues 15-18 and 20-25 with low propensity for helicity between residues 2 and 4. In contrast, the region between residues 5 and 14 exhibits an undefined conformation. In the VIP-ff99SP-GB^{HCT} trajectory (Figure 5.2b), the peptide adopts α- and 3_{10}-helix structures from residues 10 and 26. The segment from residue 2 to 8 exhibits a low tendency for helical conformation, with residue 9 adopting an extended conformation. In trajectory ff99-GB^{OBC} (see Figure 5.2c) the peptide exhibits a higher degree of helicity than the two preceding trajectories computed using the GB^{HCT} approximation. Finally, the trajectory ff99SP-GB^{OBC} presents a similar conformational profile as ff99-GB^{OBC} with a higher degree of helicity in the profile of the peptide, as can be seen in Figure 5.2d.

Time evolution of conformational motifs of the peptide is depicted in Figures 5.3a to 3d for each of the four trajectories of VIP. As can be seen the percentage of secondary structures sampled in the trajectories follows the order: VIP-ff99-GB^{HCT} >VIP-ff99SP-GB^{HCT} >VIP-ff99-GB^{OBC} >VIPff99SP-GB^{OBC}. Moreover, the secondary structural motifs appear almost from the beginning of the trajectory with the transient existence of conformations with 3_{10}- and α-helical motifs.

Although secondary structure analyses provide an overall description of the average structure, the performance of the different protocols in describing the peptide folding process can be more accurately assessed by making direct comparisons with

NMR experiments **[67]**. Since NOE's are inversely related to the sixth power of the distance between the atoms involved, they offer a direct comparison with distances measured in the computer simulations. Accordingly, distances between atoms responsible for the different NOE's reported in literature **[67]** were calculated for each of the different MD trajectories in this study.

Assuming a normal distribution, distance intervals computed from the MD studies were defined by the average value of the distance ±1,96 σ, covering a 95% of the variance. Assessment of the similarity between the solution structure reported in the NMR experiments and those computed in the MD trajectories, the percentage of overlap between the NOE distance interval and the MD distance interval was calculated.

In order to avoid the effect on the distance produced by the folding process of the peptide, the overlap was computed for the total length of each trajectory (200 ns) and for the last 50 ns. The overlap was separately computed for the long distances (ld) (corresponds to NOE signals of atoms two residues away) and short distances (sd) (neighbouring atoms). Table 5.1 lists the values of the computed overlap between MD and experimental NMR results. Figure 5.4 presents a qualitative comparison between the distance intervals attained by the atoms involved in the different NOE signals of VIP inferred from the NMR results reported in literature **[67]** and the distances computed

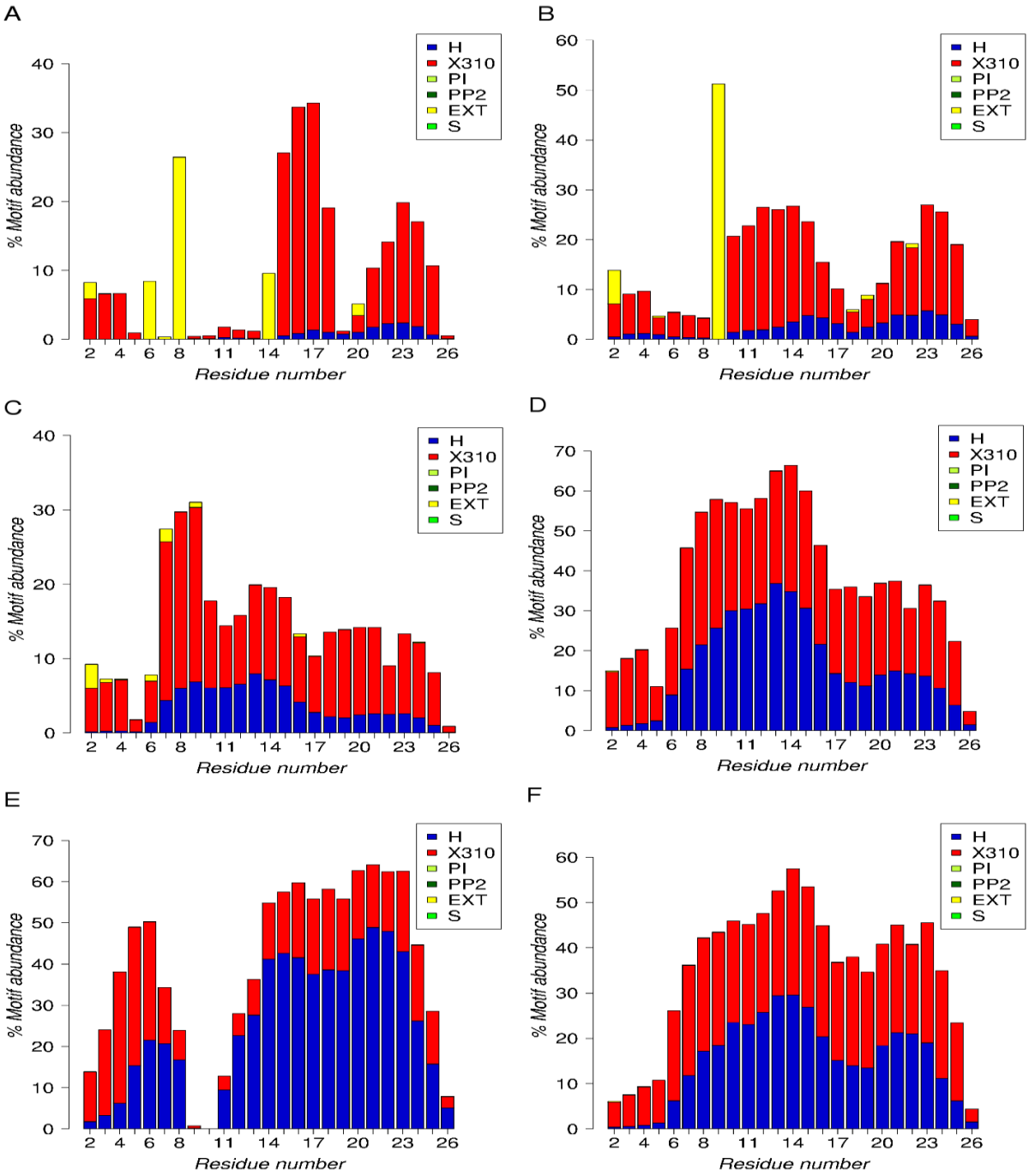


Figure 5.2: Percentage of motif abundance for the MD trajectories of VIP: (a) VIP-ff99-GB^{HCT} (b) VIP-ff99SP-GB^{HCT} (c) VIP-ff99-GB^{OBC} (d) VIP-ff99SP-GB^{OBC}. (e) PACAP27-ff99-GB^{OBC} and (f) PACAP27-ff99SP-GB^{OBC}. Conformational motifs are labelled: H (α-helix), X310 (3_{10}-helix), PI (π-helix), PP2 (polyproline II), EXT (extended conformation), S (β-strand). Only H, X310 and EXT are exhibited by the structures in the MD trajectories obtained in the present study. PI, PP2 and S are absent.

from the MD simulations. Analysis of Figure 5.4 provides a qualitative idea of the overlap between the two sets of distances for each of the four different trajectories: suggesting that an overlap of 8, 10, 14 and 14 out of 14 distances, for the trajectories VIP-ff99-GB^{HCT}, VIP-ff99SP-GB^{HCT}, VIP-ff99-GB^{OBC} and VIP-ff99SP-GB^{OBC}, respectively. This can be further analyzed in a more quantitative basis from the analysis of Table 5.1, where it can be observed that GB^{OBC} provides more reliable results. With regard to comparing the effect of the force field, it can be concluded that ff99SP exhibits the higher agreement with the NMR structure considering the tolerance associated to the different atom-atom distances.

After the analysis of the results obtained for VIP, two 200ns MD trajectories for PACAP27, starting from its extended conformation were run using the GB^{OBC} implementation with two different force fields ff99 and ff99SP. The trajectories were labelled PACAP27-ff99-GB^{OBC} and PACAP27-ff99SP-GB^{OBC}, respectively. Figures 5.2e and 5.2f show the conformational motifs characterized in the two trajectories using a three-residue window. The profile obtained describes the peptide as exhibiting a helix at the N-terminus (residues 2-8) and the C-terminus (residues 11-26), whereas residues 9-10 do not present a defined structure. On the other hand, the results obtained using ff99SP show a partially helical structure at the N-terminus (residues 2-5) and two consecutive helices at residues 6-18 and 20-25. The tendency for a helical structure is low at residue 19. These results compare well with the structure of PACAP27 obtained

Table 5.1: Percentage of overlap between the NOE's and MD interval for the VIP and PACAP27 trajectories. The NOE's interval has been taken from Thierault *et al*., **[69]**, Wray *et al*., **[78]** for VIP and PACAP27, respectively. The overlap was computed for the total length of each trajectory (200 ns) and for the last 50 ns. The overlap was separately computed for the long distances (ld) (corresponds to NOE signals of atoms two residues away) and short distances (sd).

VIP-ff99-GB^{HCT}					
0-200			**150-200**		
total	**ld**	**sd**	**total**	**ld**	**Sd**
51.3	26.6	54.1	53.4	26.8	56.7
VIP-ff99SP-GB^{HCT}					
0-200			**150-200**		
total	**ld**	**sd**	**total**	**ld**	**Sd**
51.4	34.9	53.4	53.1	38.7	54.8
VIP-ff99-GB^{OBC}					
0-200			**150-200**		
total	**ld**	**sd**	**total**	**ld**	**Sd**
53.6	37.5	55.6	54.8	41.2	56.4
VIP-ff99SP-GB^{OBC}					
0-200			**150-200**		
total	**ld**	**sd**	**total**	**ld**	**Sd**
55.9	43.8	57.4	56.6	47.3	57.7
PACAP27-ff99-GB^{OBC}					
0-200			**150-200**		
total	**ld**	**sd**	**total**	**ld**	**Sd**
21.5	11.7	28.0	21.7	12.0	28.0
PACAP27-ff99SP-GB^{OBC}					
0-200			**150-200**		
total	**ld**	**sd**	**total**	**ld**	**Sd**
20.8	11.4	27.0	21.3	11.9	27.5

by NMR spectroscopy in the presence of trifluoroethanol (TFE) where the peptide exhibits a disordered N-terminal domain followed by an α-helix expanding residues 9-26 with a discontinuity between residues 20 and 21. Time evolution of the conformational motifs is depicted pictorially in figures 5.3e and 5.3f for the two trajectories of PACAP27. In this case of PACAP27, the trajectory performed using the ff99 force field yielded a higher degree of helical structure than that performed using the ff99SP force field.

Comparison with reported NMR data, from a study carried out by Wray *et al*., **[78],** was carried out in a similar manner as for VIP. Table 5.1 shows the percentage overlap between the distances computed from the MD trajectories and those deduced from the NOE's reported in literature, and Figure 5.4e and 4f show pictorially the overlap between the two sets of distances. The results obtained indicate that of the 35 distances computed, 14 exhibit good overlap in the trajectory PACAP27-ff99-GB^{OBC}, whereas trajectory PACAP27-ff99SP-GB^{OBC} exhibits an overlap of 29 distances. In this case where the experimental results are of better quality than in VIP, the force field ff99SP trajectory of PACAP27 produced a larger degree of resemblance to the NMR structural features of PACAP27 in a TFE-water mixed solvent. In the present study a series of MD trajectories of VIP and PACAP27 were carried out under an implicit model of the solvent. The aim of this study was two-fold: firstly, to characterize the folded conformation of these two peptides and secondly, to determine which of the Generalized Born models is the most accurate in reproducing the folding structure of the peptide in solution that is similar to the NMR structures available in the literature. The analytic GB model efficiently describes electrostatics of molecules in water environment.

It represents the solvent implicitly as a continuum with the dielectric properties of water, and includes the charge screening effects of salt.

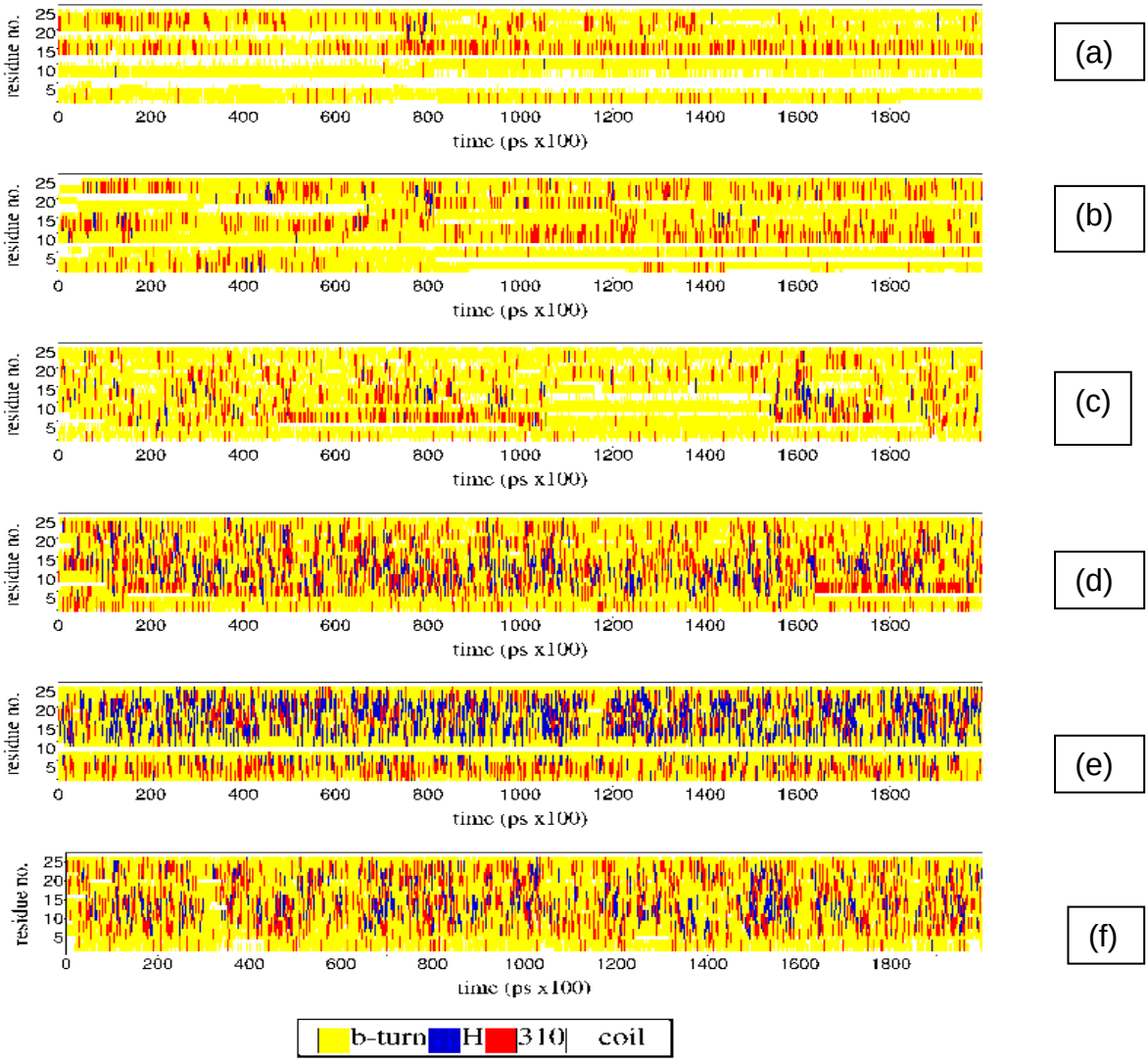


Figure 5.3: Time evolution of secondary motifs for the VIP and PACAP27 MD trajectories: (a) VIP-ff99-GB^{HCT} (b) VIP-ff99SP-GB^{HCT} (c) VIP-ff99-GB^{OBC} (d) VIP-ff99SP-GB^{OBC}. (e) PACAP27-ff99-GB^{OBC} and (f) PACAP27-ff99SP-GB^{OBC}. Conformational motifs are labelled and coloured: β-turn in yellow (considering the different types of β-turns), H in blue (α-helix), 310 in red (3_{10}-helix), and coil in white (being random coil).

The GB^{OBC} implementation uses effective Born radii that has been shown to greatly increase the accuracy of the atom-pair terms. This study has shown that the GB approximation published by Onufriev *et al*., (GB^{OBC}) **[41]** presents a higher agreement in respect to the NMR structure of VIP available in the literature than the GB^{HCT} approximation. Thus, by using the GB^{OBC}, a conformation of VIP was obtained that exhibits a partially helical N-terminus (residues 2-6) and the central and C-terminal parts of the VIP molecule were characterized as two α-helices as it has been described by NMR spectroscopic studies. Furthermore, two different force fields (ff99 and ff99SP) were analyzed for VIP. The latter represent a modification of the original force field parameters that resulted in a significant improvement over the original AMBER force field in studies of the helix-coil transition in polyalanine based peptides. The two force fields give comparable sets of conformations, although ff99SP gives a larger proportion of helicity. Regarding the relative abundance of α- and 3_{10}-helices, ff99SP gives similar amounts of both whereas ff99 gives a larger proportion of 3_{10}-helices.

Regarding PACAP27 the set of conformations obtained using ff99 exhibit helicity at the N-terminus (residues 2 to 8) and the C-terminus (residues 11-26) whereas residues 9-10 do not present a defined structure. On the other hand, the results using ff99SP show more agreement with the experimentally determined structure in TFE as it exhibits partially helical structure at the N-terminus (residues 2-5) and two consecutive helices at residues 6 to 18 and 20 to 25. The helicity is lower at residue 19. This result compares well with the structure of PACAP27 obtained by NMR spectroscopy in the

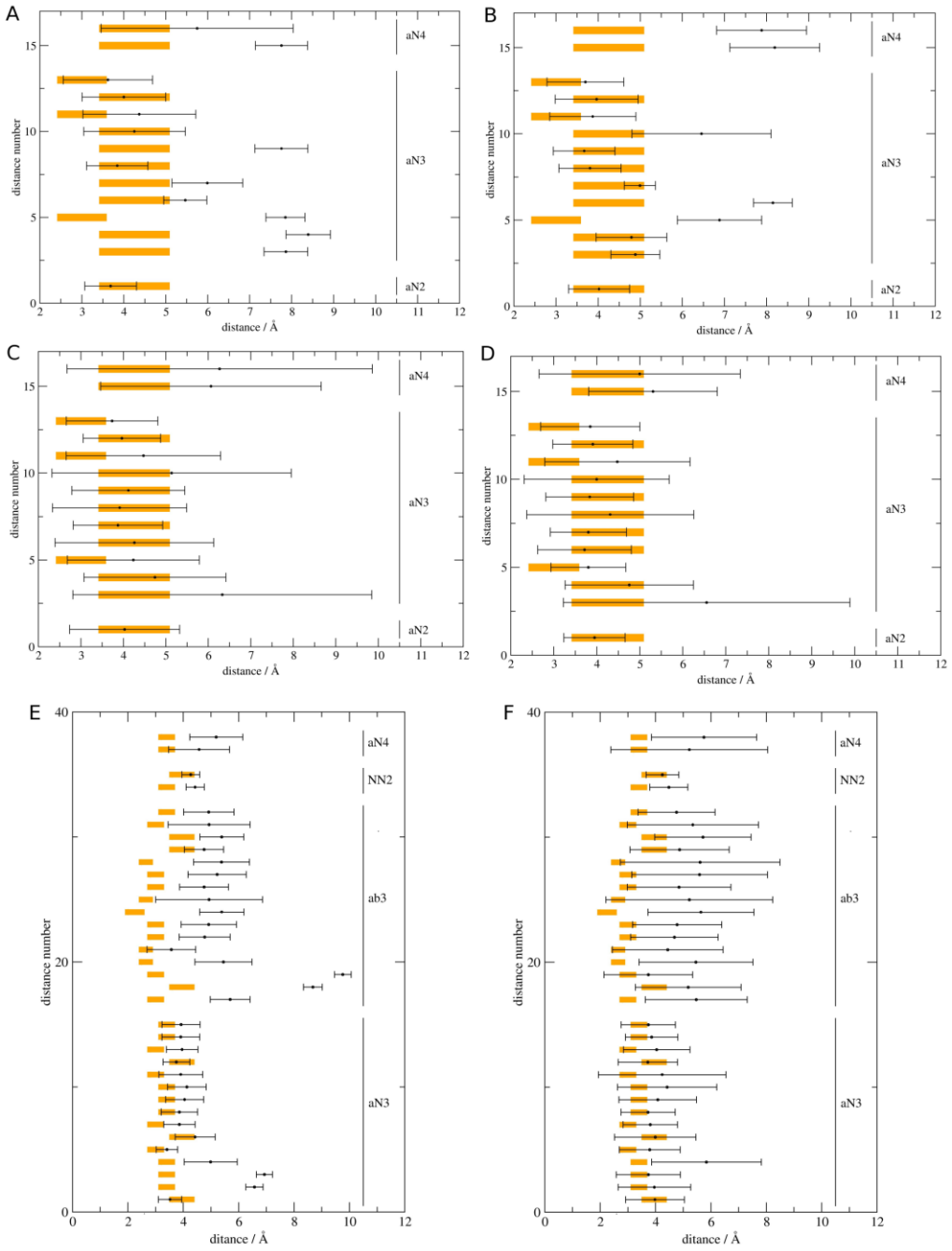


Figure 5.4: Comparison of NMR derived distances shown in orange and the MD average with the distance interval containing 95% of the structures for: a) VIP-ff99-GB^{HCT}, b) VIP-ff99SP-GB^{HCT}, c) VIP-ff99-GB^{OBC}, d) VIP-ff99SP-GB^{OBC}, e) PACAP27-ff99-GB^{OBC} and f) PACAP27-ff99SP-GB^{OBC}.

presence of trifluoroethanol (TFE) where it exhibits a disordered N-terminal domain followed by an α-helix expanding residues 9 to 26 with a discontinuity between residues 20 and 21. Time evolution of the conformational motifs is depicted pictorially in Figures 5.2e and 5.2f for the two trajectories of PACAP27.

The results provide evidence that, with the protocols used in this study, both peptides were able to fold within a 200-ns trajectory. In contrast, trajectories of this length performed with these peptides using the extended conformation as initial structure and soaked in a box with explicit solvent were not long enough for the peptide to fold (data not shown). However, whereas the four protocols yielded folded structures, there are subtle differences among them. Specifically, as shown in Figure 5.1 the four protocols do not sample the conformational space in the same manner. Only the GB^{OBC} model with the ff99SP force field seems to sample the space with a constant yield of new configurations whereas, in the case of the other three protocols the sampling seems to get stuck in different regions of the space. To understand the relevance of the information provided by the simulations, these results were compared with those available from NMR experiments in TFE. Comparison can in turn provide hints to understand the differential performance of the four protocols used.

MD simulations were performed using a continuous model of the solvent with two different implementations: GB^{OBC} and GB^{HCT}, using a permittivity ε=1 for the solute and ε=80 for the solvent. Although diverse reports support that inclusion of the solvent effects using second-generation continuous models is a reliable procedure way to include solvent effects on the energetics of peptides folding **[80]**, however, there are still

some problems regarding the best form to be used due to several factors including the effects at the boundary between the solute and the solvent; the dependence of the Born radii on the external dielectric constant; or the dependence of the implementation to the different force fields used **[88, 81]**. However, it is known that the procedure is expected to yield a peptide conformational profile biased to enhance compact conformations with structures expected to be stabilized by peptide-solvent weak bonds shown to be not too favorable. Furthermore, the external dielectric constant affects the electrostatic interactions between charges in the solute, compensating the compactness of the structures **[40]** and consequently, peptides with sequences of a majority of hydrophobic residues will not be as affected as sequences with charged residues.

In any case, it could be argued that the GB approach bias the conformational profile providing a list of the different accessible conformations that are intrinsic to its primary sequence. On the other hand, a similar argument has also been given in the literature for explaining the tendency of peptides to be more structured in experiments carried out in TFE/water mixtures **[77]**. Although the reason for this behaviour is not completely known, there is evidence supporting that this is due to the tendency of the solvent to form large clusters surrounding the peptide, enhancing in this way intramolecular hydrogen bonding of the amide group **[82]** resulting in a more defined structure in solution. Accordingly it would be reasonable to expect similar conformational bias in the two procedures, giving support for a meaningful qualitative comparison. For this purpose, a comparison of the experimental NOEs available for VIP with those computed from the different simulations permits us to draw some conclusions. Inspection of Figure 5.4 suggests that the GB^{OBC} model with the ff99SP

force field provides the best match with the experimental results for the different conditions tested. This is quantitatively confirmed looking at the results listed in Table I. Indeed, although the performance of the four protocols is similar, providing an average overlap of 57% for the long distances this percentage increases from a 27% to 59%. Based on these results PACAP27 was studied using only the GB^{OBC} implementation.

The results obtained in the present work clearly suggest that the peptides attain partially helical structures regardless of the force field. However, the subtle changes in the secondary structure induced by the f99SP force field favours the intramolecular interactions arising from the peptide sequence, similarly as the structuring solvents do experimentally. Therefore, we conclude that the combination ff99SP and GB^{OBC} appears to be a good method for predicting the behaviour of the peptides in a solvent that stabilizes secondary structure. However, further studies will be carried out with the ALPB approximation **[40]** to evaluate if it performs any better than GB^{OBC}. Following this reasoning, it was decided to adopt in the present work a computational approach aimed at characterizing the intrinsic conformational features of VIP and PACAP 27, which could provide information about the equilibrium conformation of peptides starting from an extended one.

On the other hand, different force fields perform better than others when combined with implicit solvent models. In this situation in the present work we decided to test two different Generalized Born models: the GB^{HCT} and the GB^{OBC} in combination with two different force fields: the ff99 developed from the refinement of many torsional parameters from ff94 and ff96, and a modification of it called ff99SP to avoid a strong

helical bias observed in the original force field based on studies of the helix-coil transition in polyalanine **[39]**. The conclusion of this work regarding the better performance of the GB^{OBC} model was also found in a recent work aimed at comparing the performance of different combinations of GB models and force fields in their ability to properly simulate folding of small-medium size peptides of known structure **[39]**.

One of the goals of the present contribution was to investigate the limitations of different computational procedures using an implicit model for the solvent in providing a picture of the conformational profile exhibited by a medium-sized peptide in solution without any prior knowledge. The four protocols produce different patterns of the configurational space sampled (see Figure 5.1). Thus, the combination GB^{OBC}-ff99SP samples the space with new conformations appearing at a constant rate whereas the three other simulations (GB^{OBC}-ff99, GB^{HCT}-ff99SP, and GB^{HCT}-ff99) get trapped in different areas of the conformational space.

5.4 Conclusions

In summary, the GB^{OBC} implementation is better than GB^{HCT} in terms of the efficiency of the conformational space exploration. Regarding the differences that the use of ff99 and ff99SP has produced, it can be concluded that the ff99SP fits better with the NMR experimental results in a TFE/water mix. This solvent condition induces the structuring of the peptide and thus, ff99SP produces a higher degree of secondary structure compared to what should be expected in water, thereby compensating for the need of a structuring solvent.

CHAPTER 6

CASE STUDY III: MD STUDY OF VIP: EFFECT OF THE FORCE FIELD

6.1 Introduction

Molecular dynamics (MD) simulations of even the smallest peptides is not a trivial task. The many degrees of freedom inherent in peptides make their simulation studies to be computationally demanding. Thus, to adequately sample the conformational space of peptides, not only the chosen generalized Born (GB) model is critical but the force fields employed must be equally robust. Due to the increasing interest in applying computational methods in drug discovery, development and refinement of simulation protocols is essential. Specifically, robust computational methods can be used to gain insight about peculiar features of known peptides and proteins and can also be used as a predictive tool in the study of unknown and lesser known peptides and proteins. Thus, using these techniques, the dynamic structural features of peptides and protein segments can be elucidated, resulting in the assignment of functions and determination of their mechanisms of action.

This study was carried out to evaluate the effect of force fields in MD simulations of the folding patterns of helical peptides. VIP was used as a model peptide. In the previous study (Chapter 5), ff99SP proved to be reliable in reproducing the folding patterns of VIP that were comparable to experimental results. In this particular study the performance of ff99SP was compared to ff96 and its modified version ff96n (i.e. ff96 with igb =7). In particular, the ability of these force fields to produce helical structures starting with an extended structure of VIP was evaluated.

6.2 Computational Protocols

VIP was constructed using the leap module in AMBER 8.0 **[26]**. The N-termini were protonated (NH_3^+) and the C-terminal was amidated ($CONH_2$). The initial structure in extended conformation was energetically minimized with a convergence criterion of 0.005 kcal.mol^{-1} Å^{-1}, the non-bonded cutoff was set to 99 Å, and the direct sum of non-bonded list updated every 10 steps. After energy minimization, the structure was equilibrated for 10 ps by gradually increasing the temperature from 0 to 300 K.

Since replica exchange molecular dynamics (REMD) simulations are known to sample the conformational energy space of peptides and yielding data that is consistent with experimentally determined data, a single REMD trajectory of VIP, using ff96, was used as a reference calculation. For this REMD reference trajectory, eight replicas with target temperatures allocated according to the procedure described below were run in parallel for 150 ns for VIP. The exchange frequency was set to 2 ps. The number of replicas necessary to carry out these calculations was obtained by calculating the closest square root of the number of atoms of the system that yields a positive number. In order to determine the temperature distribution of the replicas, the system was equilibrated at eight different temperatures: 200, 300, 400, 500, 600, 700, 800, and 900 K, providing information about the average energy as a function of temperature, E(T). This information was used to derive a set of seven coupled non-linear equations using a probability of acceptance criterion of 0.2 between two neighbouring replicas. The system of equations was solved iteratively by varying the temperature of each replica, subject to the condition that the temperature of the third replica was set to 300 K.

Solution to these equations yielded a set of target temperatures for the different replicas.

Table 6.1: Study design to investigate the effect of force fields

Simulation	Force field	Implicit solvent
MD: VIP	ff96	Water
	ff96 (igb=7)	Water
	ff99SP	Water
REMD: VIP	ff96	Water

Three different force fields shown in Table 6.1 were used to run 200 ns calculations: ff96, ff96n (igb =7) and ff99SP. Salt concentration was set to 0.2 M in order to mimic physiological conditions. The dielectric constant around the peptide (internal dielectric constant) was set to ε=1 and the external dielectric constant of ε=78.5, corresponding to water. More precisely, the solvent was represented with the continuum generalized Born model presented by Onufriev, Bashford and Case (GB^{OBC}) **[41]**. This GB model was chosen after it was found to be effective in simulating folding patterns of VIP and PACA27 (see Chapter 5). A total of three trajectories of 200 ns were run. One REMD trajectory using force field ff96 was used as a reference calculation. Dihedral angles were obtained from the MD trajectories using relevant *ptraj* scripts in the AMBER suite **[26]** and analyzed using our in-house programs.

6.3 Results and Discussion

Three 200 ns MD trajectories of VIP were computed using the extended conformation as starting structure and using three different force fields (ff96, ff96n (igb=7) and ff99SP) under an implicit water solvent. Trajectories were labelled as VIP-ff996, VIP-ff96n, and VIP-ff99SP, respectively. The reference REMD trajectory was labelled VIP-REMDff96.

For every snapshot in each trajectory, conformational patterns were assigned following the procedure described in the Computational Protocols section in Chapter 5. Representation of the evolution of the new patterns sampled along the different trajectories is shown in Figure 6.1. These plots provide an indication of how efficient the different force fields were in exploring the conformational space of VIP. The simulations using ff96 and ff96n seem to sample the conformational space by hopping, meaning that the peptide gets trapped in lower regions (one of the many local minima) of the conformational space at certain intervals (Figure 6.1a and b). The VIP-ff96 trajectory produced different patterns until it equilibrated, while the VIP-ff96n trajectory followed the same pattern but produced bigger clusters of identical structures. It can thus be inferred that by modifying force field ff96 to force field ff96n, certain structural conformations are stabilized. On the other hand, the use of ff99SP force field appears to be unrestrained in the sampling of the conformational space, hence the linear plot, evidenced by new patterns explored at a steady pace by the VIP-ff99SP trajectory depicted in Figure 6.1c. This trajectory produced patterns in a manner similar to the reference REMD trajectory ran using ff96 shown in Figure 6.1d.

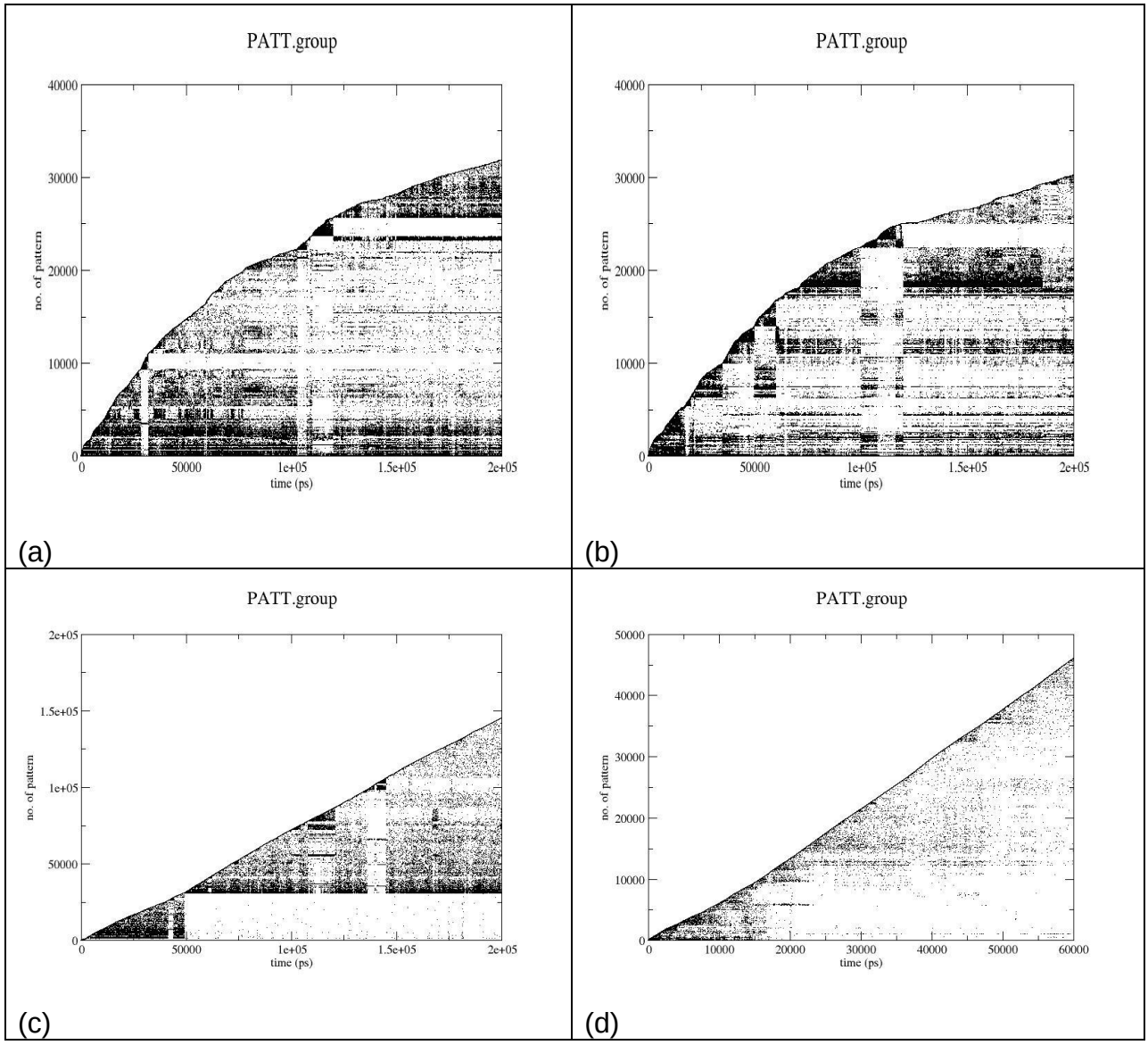


Figure 6.1: Patterns of structures obtained using the CLASICO programme for (a) VIP-ff96, (b) VIP-ff96n, (c) VIP-ff99SP, and (d) VIP-REMDff96. Each new combination of conformational motifs is a pattern.

Figures 6.2a to 2d show the conformational motifs sampled along each of the four trajectories using a three residue window. Specifically, Figures 6.2a and 6.2b show the profile obtained in the VIP-ff96 and VIP-ff96n trajectories, respectively. In both

cases, no helical structures were obtained. Instead, the peptide exhibits predominantly an extended conformation with low propensity for a β-strand. On the other hand, Figure 6.2c showing the profile for the VIP-ff99SP trajectory, exhibiting exclusively both the α-helix and the α-3_{10} helix structures. The reference calculation presented in Figure 6.2c exhibits a predominantly helical structure with a low propensity for an extended conformation in the first few residues in the N-terminal and the last few residues forming the C-terminal.

Figures 6.3a-d show the conformational motifs sampled along each of the four trajectories using a three residue window, taking the β-turns into account. Figures 6.3a and 3b both exhibit few residues with β-turn conformations. Only two residues in VIP-ff96 trajectory Figure 6.3a) have at least 20% β-turns and only four residues in the VIP-ff96n trajectory (Figure 6.3b) have a similar percentage of β-turns. On the contrary, Figure 6.3c (representing the VIP-ff99SP trajectory) has an average of 60% β-turns spanning the entire peptide, showing the tendency of this trajectory to form a helical structure. The reference calculation exhibits β-turns from the residue 2 to the C-terminal, with the middle region displaying as high as 25 % β-turns. The α-helix in this case occupies a higher percentage, with low percentage of the 3_{10} helix in few residues.

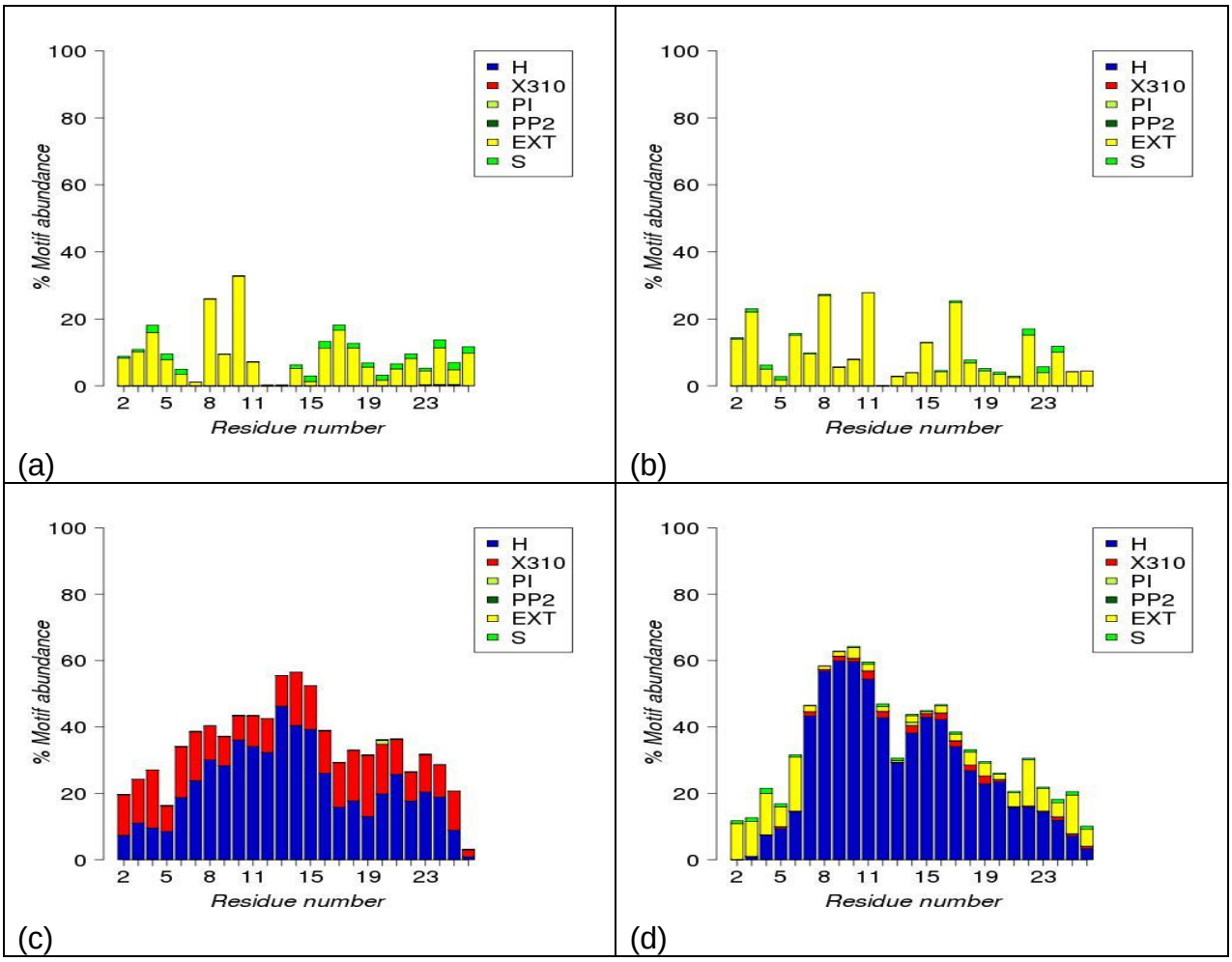


Figure 6.2: Motif abundance for the MD trajectories of VIP excluding β-turns: (a) VIP-ff96, (b) VIP-ff96n, (c) VIP-ff99SP, and (d) VIP-REMDff96. Conformational motifs are labelled: H (α-helix), X310 (3_{10}-helix), PI (π-helix), PP2 (polyproline II), EXT (extended conformation), S (β-strand). Only H, X310 and EXT are exhibited by the structures in the MD trajectories obtained in the present study, being PI, PP2 and S absent.

These results indicate that standard MD simulations using both ff96 and ff96n are not able to reproduce the folded VIP molecule as observed in experiments **[78]**. This suggests that in these simulations the peptide is not able to sample structures in the entire energy landscape but gets trapped in local minima, thus failing to reach the global minimum (no blue bars). On the contrary, standard MD simulations using ff99SP are able to sample folded helical structures of the VIP molecule, as was the case in Chapter

5. Although these simulations gave higher percentage of β-turns to α-helix and α-3_{10} helix structures, these β-turns occur consecutively and would yield a helical molecule even if they were the only conformation populating the trajectory.

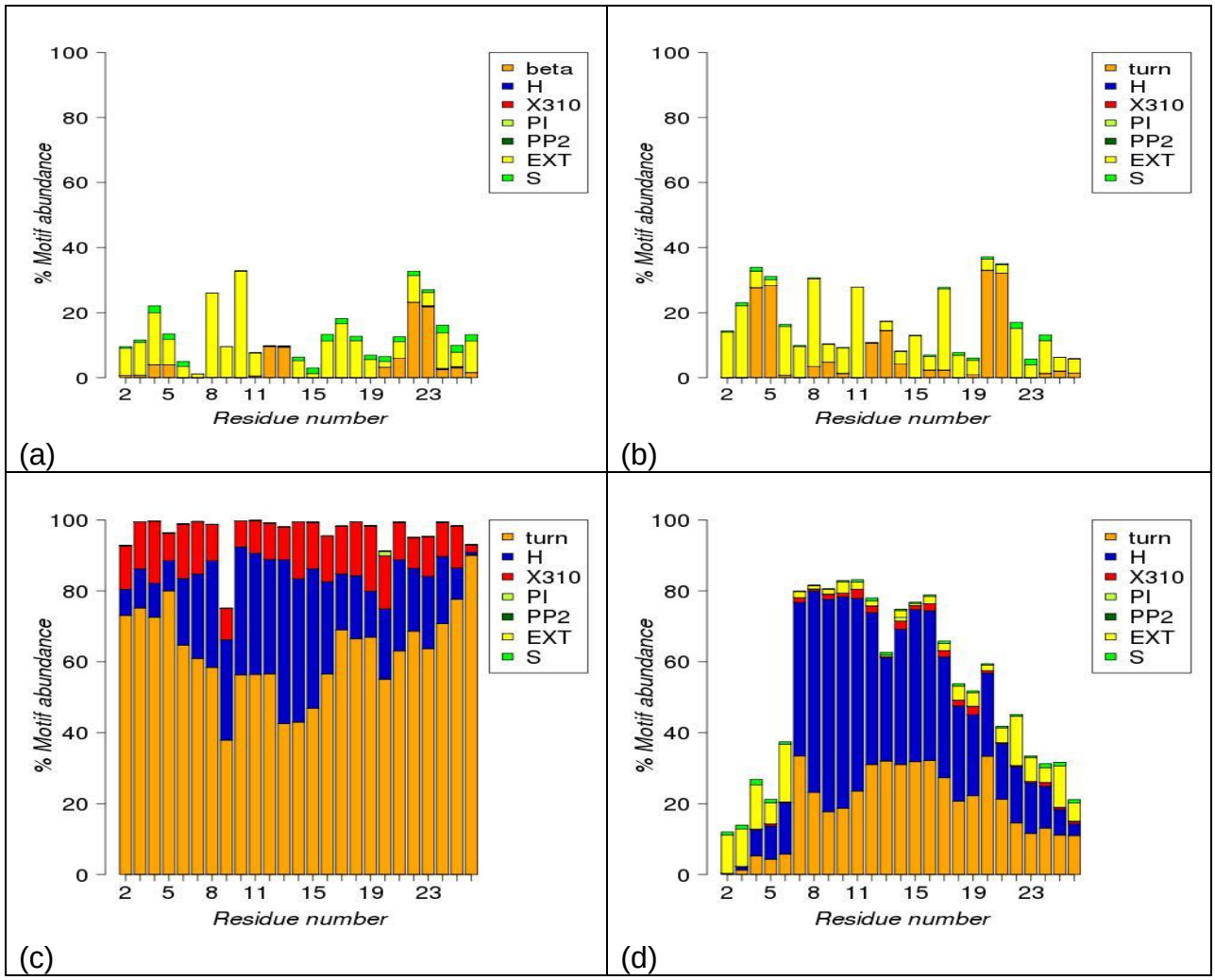


Figure 6.3: Percentage of motif abundance for the MD trajectories of VIP including β-turns: (a) VIP-ff96, (b) VIP-ff96n, (c) VIP-ff99SP, and (d) VIP-REMDff96. Conformational motifs are labelled: turn (β-turn), H (α-helix), X310 (3_{10}-helix), PI (π-helix), PP2 (polyproline II), EXT (extended conformation), S (β-strand). Only H, X310, EXT and turns were exhibited by the structures in the MD trajectories obtained in the present study, being PI, PP2 and S absent.

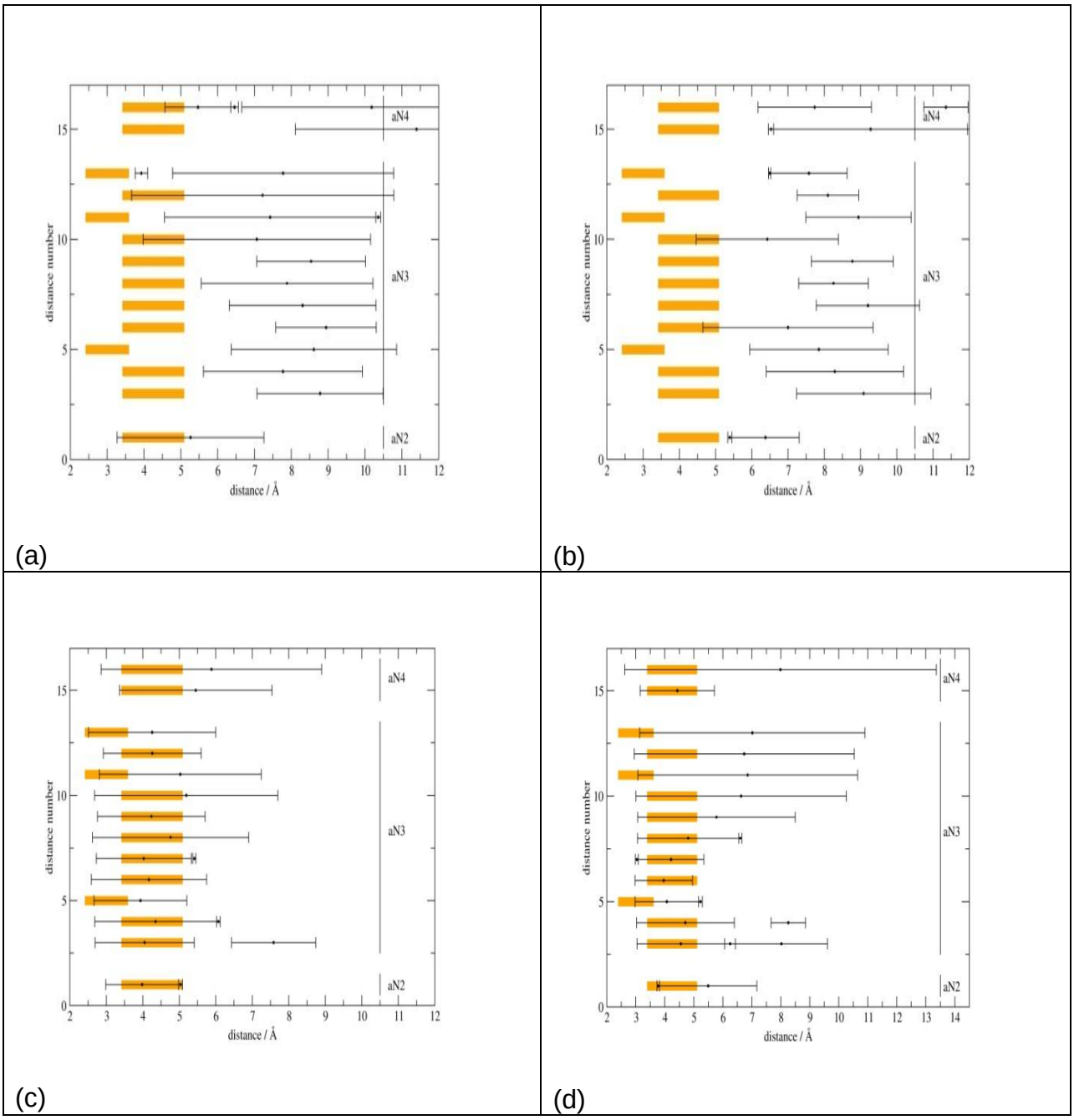


Figure 6.4: Comparison of selected NMR derived distances shown in orange and the MD average with the distance interval containing 95% of the structures obtained in the last 50 ns for: (a) VIP-ff96, (b) VIP-ff96n, (c) VIP-ff99SP, and (d) VIP-REMDff96.

The reference REMD calculations presented in Figures 6.2d, and 6.3d were able to re-produce helical structures similar to the experimentally observed results **[78]**

despite using ff96 which did not produce helical structures when used in standard MD simulations (Figures 6.2a, 6.2b, 6.3a and 6.3b). This shows that helical structures can be produced in simulations using a lesser adapted force field provided the computational method has the ability to avoid trappings in the energy landscape **[35]** (as it is the case with REMD calculations). This shows that helical structures can be produced in simulations using a lesser adapted force field provided the computational method has the ability to avoid trappings in the energy landscape **[35]** (as it is the case with REMD calculations).

The integrity of the results obtained and discussed thus far can further be verified qualitatively, by comparing the selected distances obtained from the different MD trajectories with corresponding experimental NMR distances published in literature. Accordingly the MD distances were plotted and the NMR NOEs were overlayed to determine the extent of overlap as shown in Figure 6.4. These results indicate that of the 14 distances, there was overlap in 4 distances (28%) for VIP-ff96, 2 (1.4%) for VIP-ff96n, 14 (100%) for both VIP-ff99SP and VIP-REMDff96 trajectories. It can be seen that the distance overlap in VIP-ff99SP was better than in VIP-REMDff96, though there was overlap in all the 14 distances. It can thus be deduced that distances obtained from the MD simulations of peptide folding in the VIP-ff96 and VIP-ff96n trajectories are far from experimental NMR values (Figures 6.4 a and b, respectively), suggesting that the force fields ff96 and ff96n are not ideal to be used in standard MD simulations of peptide folding.

On the contrary, distances obtained using force field ff99SP in standard MD simulations of peptide folding (Figure 6.4c), were found to be qualitatively even better than the reference trajectory (Figure 6.4d). This indicates that the force field ff99SP is ideal to be used in standard MD simulations of peptide folding depicted in Figure 6.5. Moreover, in view of the reference calculation, force field ff96 produced distances that were overlapping with the NMR experimental distances. The difference can be ascribed to the use of REMD in sampling the conformational space of the peptide.

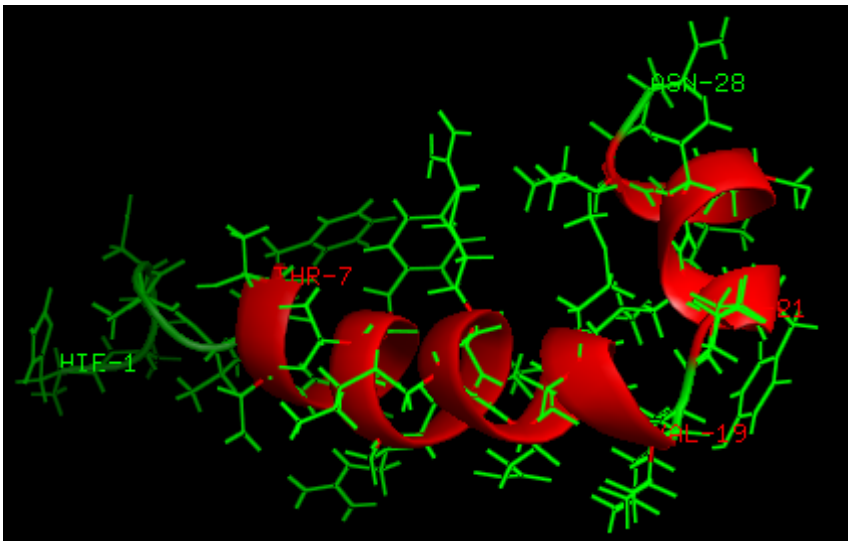


Figure 6.5: Structure of folded VIP molecule after 200 ns of standard MD obtained using Forcefield ff99SP.

6.4 Conclusions

Standard MD simulations of peptide folding require robust and advanced force fields to successfully model the folded peptide starting from an extended conformation. MD simulations using force field ff99SP are successful in reproducing the folding patterns of peptides, On the other hand, standard MD simulations using force fields ff96 and ff96n faild to adequately sample the entire energy landscape and are being trapped

in local energy minima. However, when force field ff96 is used in REMD simulations, it is able to sample helical structures by virtue of being associated with a computational method that is able to sample the energy landscape effectively. Hence, both force fields ff96 and ff99SP have a role in computational simulation studies of the dynamic properties of biological molecules such as peptides, provided they are used with appropriate GB models and computational methods.

CHAPTER 7

CASE STUDY IV: MD AND REMD STUDIES OF VIP: EFFECT OF THE DIELECTRIC CONSTANTS

7.1 Introduction

Molecular dynamics simulation of peptide folding has been traditionally performed using the Berendsen thermostat to regulate temperature of the trajectories **[26]**. Recently, some researchers have been using the Langevin thermostat (NTT=3) instead of the conventional Berendsen thermostat values in simulations of biological systems such as peptides and proteins **[26].** The Langevin thermostat was implemented in the latter versions of the AMBER programme, and works by simulating random collisions, as a molecule in solvent might feel, instead of simply scaling the velocities as in the case of the Berendsen thermostat. This method is said to equilibrate the temperature much more effectively and as a result, may enable a faster exploration of the conformational phase space. It would be interesting to determine if this method yields folded structures of flexible molecules like peptides.

In the present study molecular dynamics of VIP were performed using a Langevin thermostat under different solvent conditions. The water solvent was treated implicitly using the analytical linearized Poisson-Boltzman (ALPB) approach **[40]**, and the trajectories were run using a Langevin Dynamics. Standard MD simulations suffer from the tendency of being trapped in one of the many local minima in the rough energy landscape **[4]**. On the other hand REMD simulations efficiently hop out of local minima enabling the exploration of the entire energy landscape. This enables the peptide to

reach the global minimum, which is computationally desirable. Hence, both MD and REMD simulation protocols were used to study the folding patterns of VIP, in order to evaluate the effects of the dielectric constants of implicit solvents in the production of helical structures.

7.2 Computational Protocols

MD simulations and REMD simulations of VIP

The peptide (VIP) in its extended conformation was used as the starting structure using the Leap node of AMBER 8.0 **[26]**. VIP structure was minimized to relieve bond strains. This was followed by the equilibration step before the production dynamics step. The temperature was regulated using the AMBER 8.0 Langevin thermostat (with NTT=3) instead of the Berendsen thermostat (NTT=1) **[26]**. The temperature of the system was maintained at 300 K. This temperature control method uses Langevin dynamics with a collision frequency given by GAMMA_LN.

This temperature control method is significantly more efficient in equilibrating the system temperature than the Berendsen temperature coupling scheme (NTT=1) that was recommended for older versions of AMBER. The biggest problem with the Berendsen method is that the algorithm simply ensures that the kinetic energy is appropriate for the desired temperature; it does nothing to ensure that the temperature is even over all parts of the molecule. This can lead to the phenomenon of hot solvent-cold solute. To avoid this, elaborate temperature scaling techniques for slowly heating the molecule over the course of the simulation were recommended. The Langevin

system is much more efficient at equilibrating the temperature and is now the recommended choice for equilibrating temperature in AMBER **[26]**.

Five trajectories of 200 ns in length were run using the force field ff96 under implicit conditions of the solvent. Of these five trajectories, two standard MD trajectories were run using the GB analytical linearized Poisson Boltzman model (one trajectory with implicit water solvent, and one with implicit TFE-water solvent) and two REMD trajectories were also run using the ALPB model (one trajectory with implicit water solvent, and one with implicit TFE-water solvent). As the fifth trajectory, the REMD-ff96 trajectory was used as a reference calculation, since REMD simulations provide results that a closely comparable to experimental NMR results. However, this reference trajectory was run using the GB model of Onufriev, Bashford and Case **[41]**, since it was found to perform better in the initial study of VIP (see Chapter 5). At the completion of the trajectories, dihedral angles were extracted and analyzed for the % abundance of 3-residue conformational motifs using the clusterit.7.0 version **[71, 79]**. Accordingly, in this study different solvent dielectric constants were compared for their efficiency in simulating the folding patterns of VIP.

For the REMD trajectories including the reference calculation, eight replicas with target temperatures allocated according to the procedure described below were run in parallel for 150 ns for VIP (due to efficiency in sampling the conformational space, REMD trajectories achieve the folded structures in a shorter time). The exchange frequency was set to 2 ps. The number of replicas necessary to carry out these calculations was obtained by calculating the closest square root of the number of atoms

of the system that yields a positive number. In order to determine the temperature distribution of the replicas, the system was equilibrated at eight different temperatures: 200, 300, 400, 500, 600, 700, 800, and 900 K, providing information about the average energy as a function of temperature. This information was used to derive a set of seven coupled non-linear equations using a probability of acceptance of 0.2 between two neighbour replicas. The system of equations was solved iteratively by varying the temperature of each replica, subject to the condition that the temperature of the third replica was set to 300 K. Solution to these equations yielded a set of target temperatures for the different replicas. Figure 7.1 provides the summary of how the calculations were set up.

Table 7.1: Study design to investigate the effect of dielectric constants

Simulation	GB Model	Force field	Implicit solvent
MD: VIP	ALPB	ff96	Water
	ALPB	ff96	TFE-water
REMD: VIP	ALPB	ff96	Water
	ALPB	ff96	TFE-water
REF CALC	OBC	ff96	Water

7.3 Results and Discussion

The present study was designed to perform MD and REMD simulations with a goal to assess the conformation of VIP using a Langevin thermostat under the ALPB model and to compare the effects of the different solvent conditions (solvent dielectric

constants) in obtaining the folded structures of VIP. In all the cases, the 200 ns trajectories started with the extended conformation of VIP. Moreover, the simulations were carried out using implicit solvents since both the dynamic nature of peptides and expressing the solvents explicitly are computationally intensive entities **[4]**. The roughness of the potential energy surface of a peptide requires the use of robust algorithms **[39]** for its exploration to the extent that it justifies the use of implicit models for solvent over explicit solvation.

A total of five 200 ns MD trajectories of VIP using the extended structure as the starting conformation were calculated using the ff96 force field parameter set and the ALPB model **[39]**. The trajectories were labelled as follows: VIP-ff96-ALPB, VIP-ff96-ALPBTFE, REMD-ff96ALPB, REMD-ff96ALPB-TFE and REMD-ff96OBC, respectively.

After 200 ns the patterns characterized along the sampling process in each of the different trajectories were plotted as depicted in Figure 7.1. These plots provide a broad overview of the performance of the different trajectories in sampling new patterns as function of time. The two VIP trajectories of standard MD with ff96 seem to sample the conformational space in a staggered manner, indicating the stages when the peptide is trapped momentarily in some regions of the conformational space (Figure 7.1a and b).

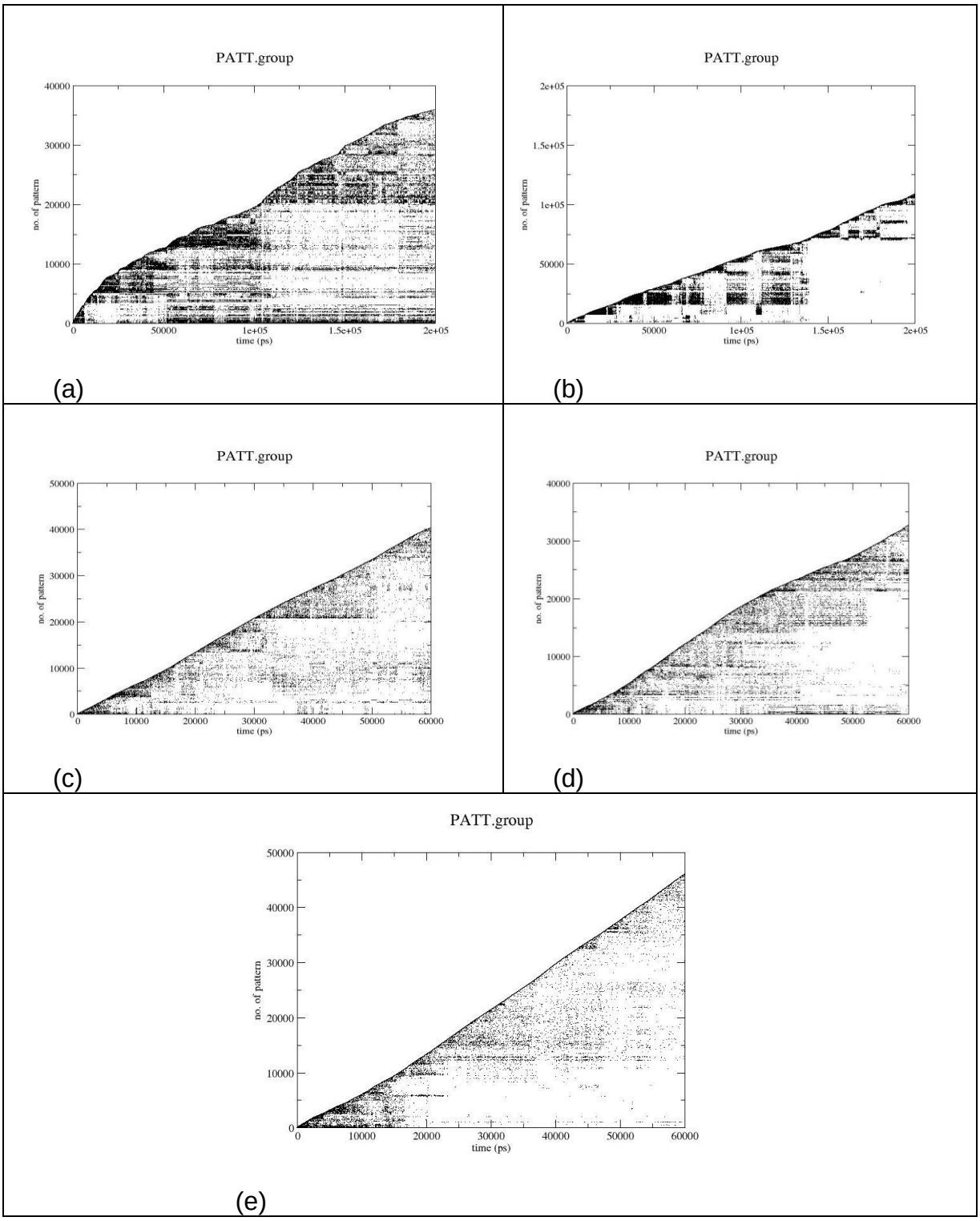


Figure 7.1: Pattern plots from MD and REMD studies of VIP using ff96 under different conditions. (a) VIP-ff96-ALPB, (b) VIP-ff96-ALPBTFE, (c) REMD-ff96ALPB, (d) REMD-ff96ALPB-TFE and (e) REMD-ff96OBC. Each new combination of conformational motifs is a pattern, and the evolution of new patterns can be followed along the trajectories.

However, in the two REMD trajectories computed with ff96 new patterns are sampled in a steady pace, indicating a lesser degree of restriction in sampling the conformational space (Figure 7.1c and d). The trajectories representing REMD-96ALPB and REMD-96ALPB-TFE (Figure 7.1c and d, respectively) sample the conformational space in a similar manner as the reference trajectory shown in Figure7.1e, by continuously sapling new patterns.

Figure 7.2a-e pictorially illustrates the conformational motifs sampled along each of the five trajectories using a three-residue window. These motifs exclude β-turns. Specifically, Figures 7.2a-d show the profiles obtained in trajectories VIP-ff96-ALPB, VIP-ff96-ALPBTFE, REMD-ff96-ALPB, and REMD-ff96-ALPBTFE. Figure 7.2e shows the conformational motifs obtained for the reference trajectory, i.e. REMD-ff96-OBC. The standard MD trajectory with TFE-water dielectric constant (Figure 7.2b) was able to sample α-helix structures (specifically from residue 12-26) as opposed to the trajectory ran using the dielectric constant of water (Figure 7.2a), which exhibited only extended conformation and β-strand structures that individually and together signify a non-helical VIP molecule. On the other hand, the two REMD trajectories (one using the dielectric constant of water and the other with the dielectric constant of TFE-water mixture) were able to sample helical structures (Figures 7.2c and d). Both these REMD trajectories showed a pattern of conformations similar to the reference trajectory (Figure 7.2e).

Figure 7.3a-e pictorially illustrates the conformational motifs (including β-turns) sampled along each of the five trajectories using a three-residue window. Accordingly, Figures 7.3a-d show the profiles obtained in trajectories VIP-ff96-ALPB, VIP-ff96-

ALPBTFE, REMD-ff96-ALPB, and REMD-96-ALPBTFE. Figure 7.3e shows the conformational motifs obtained for the reference trajectory, i.e. REMD-ff96-OBC.

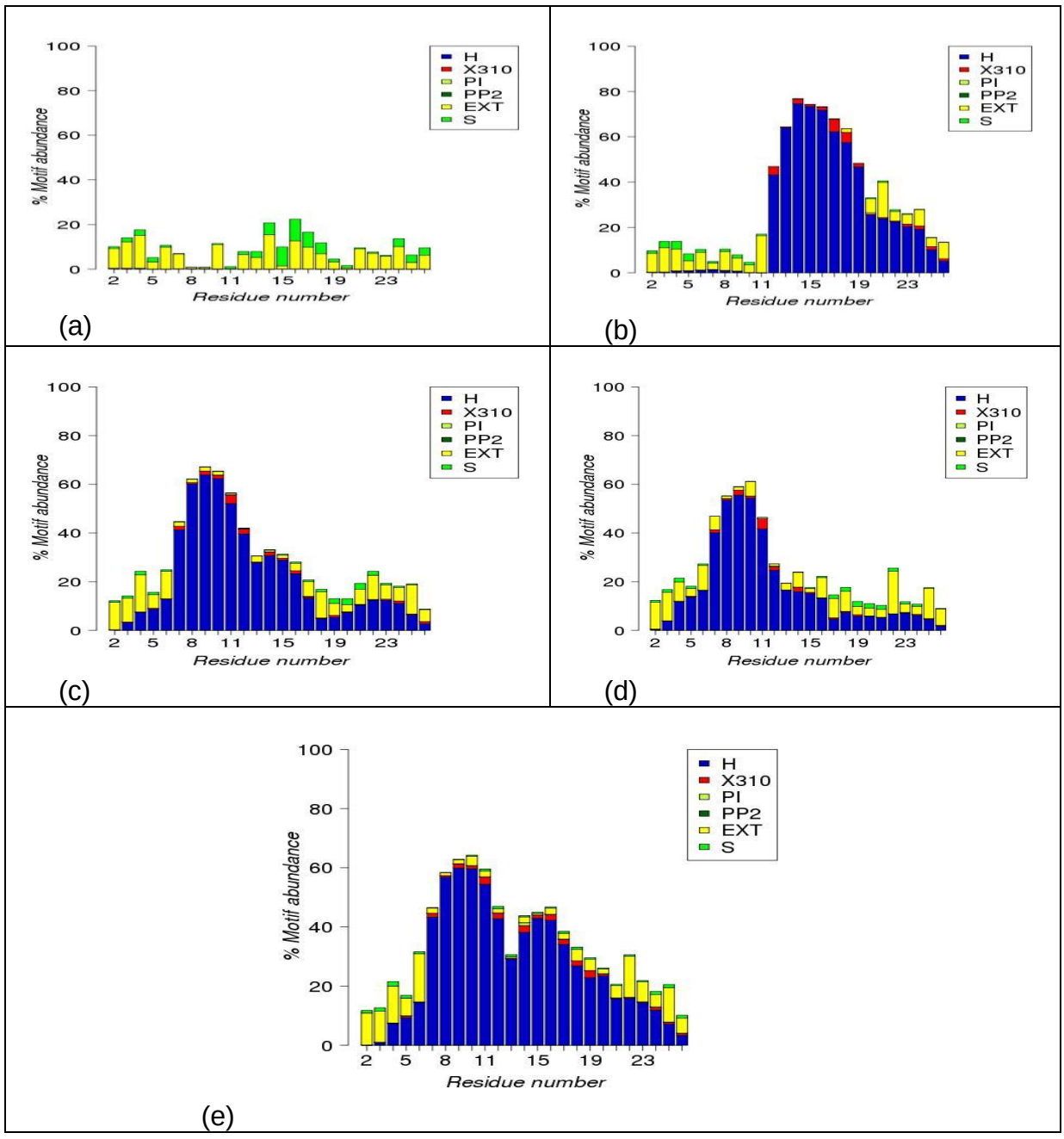


Figure 7.2: Percentage of motif abundance for the MD and REMD trajectories of VIP, excluding β-turns. (a) VIP-ff96-ALPB, (b) VIP-ff96-ALPBTFE, (c) REMD-ff96-ALPB, (d) REMD-ff96-ALPBTFE and (e) REMD-ff96-OBC. Motifs are labelled: H (α-helix), X310 (3_{10}-helix), PI (π-helix), PP2 (polyproline II), EXT (extended conformation), S (β-strand). H, X310, S, and EXT are exhibited by the structures in the MD trajectories obtained in the present study, being PI and PP2 are absent.

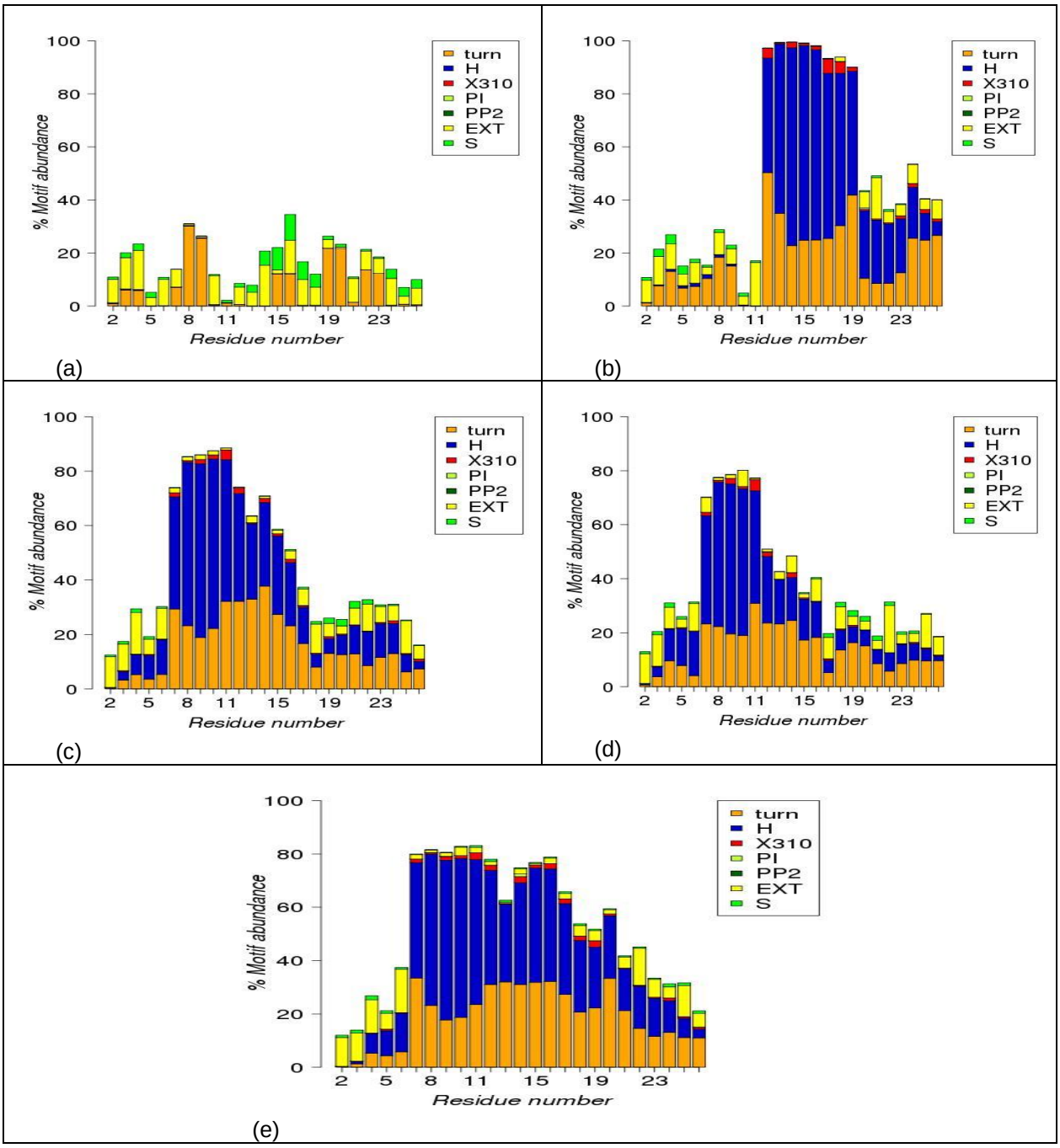


Figure7.3: Percentage of motif abundance for the MD and REMD trajectories of VIP, including β-turns. (a) VIP-ff96-ALPB, (b) VIP-ff96-ALPBTFE, (c) REMD-ff96-ALPB, (d) REMD-ff96-ALPBTFE and (e) REMD-ff96-OBC. Motifs are labelled: H (α-helix), X310 (3_{10}-helix), PI (π-helix), PP2 (polyproline II), EXT (extended conformation), S (β-strand). H, X310, S, and EXT are exhibited by the structures in the MD trajectories obtained in the present study, being PI and PP2 are absent.

Consistent with the results of conformations obtained in plots excluding β-turns, the standard MD trajectory with the TFE-water dielectric constant (Figure 7.3b) was able to sample β-turns in many residues as opposed to the lower percentage of β-turns observed in the trajectory ran using the dielectric constant of water (Figure 7.a). However, the two REMD trajectories (one using the dielectric constant of water (Figure 7.3c) and the other with the dielectric constant of TFE-water mixture (Figure 7.3d) were populated with higher percentages of β-turn structures compared to the MD trajectory as shown in Figure 7.3b. In this case, the REMD trajectory ran using the dielectric constant of water (Figure 7.c) showed a pattern of β-turn structures that is comparable to the reference trajectory (Figure 7.2e) slightly better than the trajectory ran using the TFE-water trajectory (Figure 7.3d). In general, both Figure 7.3c and 7.3d show a trend consistent with the pattern displayed by the reference trajectory. This observation suggests that the need for a structuring solvent such as TFE-water is of lesser importance in implicit REMD simulations in peptide folding studies.

The integrity of the results obtained and discussed thus far can be further verified qualitatively by comparing the selected distances from the different trajectories with corresponding experimental NMR distances published in literature. Accordingly, the MD distances were plotted and the NMR NOEs overlayed on them to determine the extent of overlap, and the results are presented in Figure 7.4a-d. Considering the selcted 14 distances, it can be observed that the standard MD trajectory ran using the dielectric constant of water had MD distance overlap with only two NMR NOEs (Figure 7.4a), while the trajectory ran using the dielectric constant of the TFE-water mixture had 8 MD distances overlapping NMR distances (Figure 7.4b). When looking at the REMD

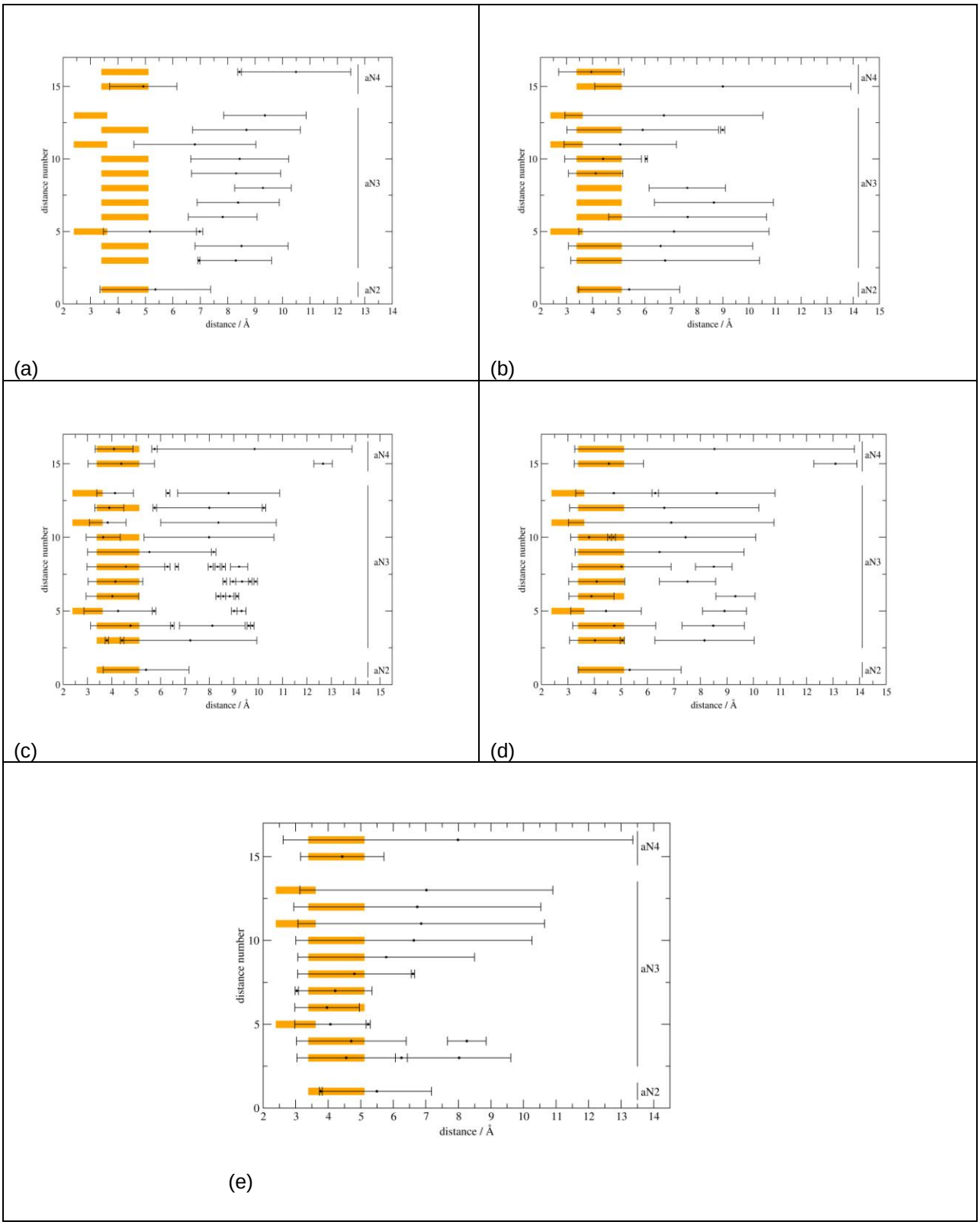


Figure 7.4: Comparison of selected NMR derived distances shown in orange and the MD average with the distance interval containing 95% of the structures for: (a) VIP-ff96-ALPB, (b) VIP-ff96-ALPBTFE, (c) REMD-ff96-ALPB, (d) REMD-ff96-ALPBTFE and (e) REMD-ff96-OBC.

trajectories, it can be observed that for the trajectory using the dielectric constant of water to to represent the solvent, there was good overlap of 13 MD distances with NMR distances (Figure 7.4c). The same observation was the case for the REMD trajectory ran using the dielectric constant of the TFE-water mixture (Figure 7.4d). These results suggest that for standard MD simulations of peptide folding, the use of a structure-stabilizing solvent like TFE-water mixture is of great importance.

However, for REMD it did not matter whether the diectric constant of water or TFE-water was used; these REMD simulations produced a similar result. There was also an overlap of 13 MD distances with NMR distances in the case of the reference trajectory, where GB^{OBC} was used instead of ALPB model, but the dielectric constant of water was used as an implicit solvent. Finally, it can be deduced that REMD simulations of peptide folding are efficient irrespective of the dielectric constant used, while standard MD simulations perform better when the dielectric constant of TFE-water is used instead of the dielectric constant of water.

7.4 Conclusions

This study has demonstrated that in the absence of a method that efficiently samples the conformational energy space of peptide systems (as in the case with standard MD simulations) the presence of a structuring solvent proves to be all important in producing helical structures. However, in the case of REMD trajectories, the presence of a structuring solvent does not seem to be that important as both β-turns and helices can be produced in abundance irrespective of the presence or absence of the TFE-water structuring solvent. Moreover, for standard MD simuations of peptide

folding there is better overlap of selected MD derived distances with NMR distances when the dielectric constant of TFE-water mixture was used instead of that of water. The REMD trajectories were able to produce a similar overlap of the selected RMD derived distances with NMR distances irrerespective of whether the dielectric constant of water or the TFE-water mixture was used, in keeping with previous results.

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

Molecular dynamics simulations of helical peptides have been performed during this study. The folding conformational features of different peptides were characterized using computational tools. Since all peptides used were known to be helical from an NMR spectroscopic point of view, it was of interest to demonstrate if computational methods would be able to reproduce structures identical to those found by experimentalists. Before MD and REMD studies were conducted, a validation study of computational methods was undertaken using explicit and water implicit solvents to run folding simulations of a ten-residue mini-protein. This validation study established that explicit REMD using ff96 was adequate to reproduce the experimentally observable α-helical structures of this mini-protein. Implicit REMD trajectories revealed that implicit force field ff96 and ff99 can be used to reproduce protein folding dynamics and that force field ff94 is not suitable for this purpose. When the force field ff96 was applied to the 5-residue peptide met-enkephalin, it was able to reproduce both type I and Type II β-turns intrinsic to this short peptide in implicit solvents using water and DMSO.

Furthermore, MD simulations protocols were performed using the different force field parameters to ascertain which models and force fields yield results that are comparable to experimental data. The generalized Born model of Onufriev, Bashford and Case using the implicit water solvent and ff99SP were found to reproduce the helicity predicted by experiments for the two model peptides, VIP and PACAP27. Our investigation using VIP revealed that structuring solvents such as TFE-water played a

crucial role in producing helical structures in implicit standard MD simulations. However, REMD simulations of VIP were able to produce helical structures both in the presence and absence of a TFE-water structuring solvent. The ability of REMD to efficiently sample the energy conformational space of peptides offsets the need for a structuring solvent.

Moreover, these implicit solvent simulations can be used to simulate folding repertoire of peptides in cases where explicit solvents simulations are computationally unfeasible due to longer timescales. Results from the present studies contribute vital information regarding structure, function and folding dynamics of peptides, by complementing and reinforcing experimental NMR work. In particular, the structures of peptides obtained from these simulations can be used to design analogues of these peptides that can be applied as potential lead compounds in computer-aided drug design. However, the next stage of this work involves an assessment of interaction of the model peptides with biomembranes, especially the lipid-bilayer (See APPENDIX II for details).

REFERENCES

1. Buehler, L.K. 2009. http://www.whatislife.com/reader/interaction-reader.html. An introduction to molecular interaction in biological systems. [Accessed on 9 Nov 2009].

2. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, and García AE. 2003. Peptide folding simulations. Current Opinion in Structural Biology, 13:168-174.

3. 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.

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

5. Cecchini, M., Rao, F., Seeber, M., and Caflisch, A. 2004. Replica-exchange molecular dynamics of amyloid peptide aggregation. The Journal of chemical physics, 121 (21): 10748-10756.

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

7. 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.

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

9. 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. PNAS, 105(6): 1867-1872.

10. 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.

11. 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.
12. 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(6): 259-270.

13. Mostofian, B and Frank, F. 2004. Basic principles in molecular modelling. Bachelor Thesis, University of Heidelberg.

14. Seibert, M.M, Patricksson, A., Hess, B., and van der Spoel, D. 2005. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. Journal of Molecular Biology, (354): 173-183.

15. Chin, W., Dognon, J.P., Piuzzi, F., Tardivel, B., Dimicoli, I., and Mons, M. 2005. Intrinsic folding of small peptide chains: Spectroscopic evidence for the formation of β-turns in gas phase. Journal of the American Chemical society, 127(2): 707-712.

16. http://www.ifpma.org/fileadmin/templates/ifpmaissues/pdfs/IFPMA-PIP-Nov2007-Final-EN.pdf. The pharmaceutical innovation platform. International Federation of Pharmaceutical Manufacturers and Associations. [Accessed on 9 Nov 2009].

17. Richon, A.B. 1994. An introduction to molecular modelling. http://www.netsci.org/Science/Compchem/feature01.html (accessed on 5 October 2009). Originally appeared in Mathematech, 1, 83.

18. *http://www.en.wikipedia.org/wiki/Molecular_modelling. Accessed on 7 October 2009.*

19. Pensak, D.A. 1989. Molecular modelling: scientific and technological boundaries. Pure and Applied Chemistry, 61(3): 601-603.

20. Corcho, F. An introduction to molecular modelling and computer-aided drug design. http://www.tesisenxarxa.net/TESIS_UPC/AVAILABLE/TDX-0323104-111352//Chapter1.pdf. [Accessed on 23 June 2007].

21. Kastowsky, M., Gutberlet, T., and Bradaczek, H. 1992. Molecular modelling of the three-dimensional structure and conformational flexibility of bacterial lipopolysaccharide. Journal of Bacteriology, 174 (14): 4798-4806.

22. 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.

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

24. 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.

25. Wallin, S and Shakhnivich, E. 2008. Understanding ensemble protein folding at atomic detail. Journal of Physics: Condensed Matter, (20) 283101 (11pp).

26. Case, D., Darden, T., Cheatham, T., Simmerling, C., Wang, J.,Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D.A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak,V., Cui, G., Beroza, P., Schafmeister, C., Cadwell, J. W.,Ross, W. S., Kollman, P. A. 2004. AMBER 8. University of California: San Francisco.

27. Hinchliffe A. 2003. Molecular modelling for beginners: Molecular Mechanics. Wiley & Sons, England, 63-78

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

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

30. 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, 301-449.

31. Lybrand, T.P. 1990. “Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods.” In: *Reviews in Computational Chemistry*, Vol. 1. K.B. Lipkowitz and D.B. Boyd (Eds.). VCH Publishers, Inc., New York, USA, pp. 295-320.

32. Brooks III, C.L., Karplus, M., Pettitt, B.M. 1985. “Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics.” In: *Advances in Chemical Physics*, Vol. LXXI. I.Prigonine and S.A. Rice (Eds.). John Wiley and Sons, New York, 1985, pp. 35-44.

33. http://www.fisica.uniud.it/~ercolessi/md/md/node8.html. 2009. Today's role of molecular dynamics. Accessed on 7 October 2009.

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

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

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

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

38. 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 (46): 15327-15331.

39. Shell, M. S., Ritterson, R., Dill, K. A. A test on peptide stability of AMBER force fields with implicit solvation. Journal of Physical chemistry B, 112: 6878–6886.

40. Sigalov, G., Scheffel, P., Onufriev, A. 2005. Incorporating variable dielectric environments into the generalized Born model. Journal of Chemical Physics, 122: 1-15.

41. Onufriev, A., Bashford, D., Case, D.A. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins, 55(2):383-94.

42. Salio C., Lossi L., Ferrini F., and Merighi M. 2006. Neuropeptides as synaptic transmitters. Cell Tissue Research, (326): 583-598.

43. Ackerman P.W., Spetea M., Nylander I., Ploj K., Ahmed M., and Kreicbergs A. 2001. An opioid system in connective tissue: A study of Acheilles tendon in the rat. Journal of Histochemistry and Cytochemistry, (49): 1387-1396.

44. Furst S. 1999. Transmitters involved in antinociception in the spinal cord. Brain Research Bulletin, 48(2): 129-141.

45. Ossipov M.H., Lai J., King T., Vanderah T.W., Malan Jr. T.P., Hruby V.J., Porreca F. 2004. Antinoceptive and nociceptive actions of opioids. Journal of Neurobiology, (61): 126-148.

46. Zhang N., Hodge D., Rogers T.J., and Oppenheim J.J. 2003. Ca^{2+}-indepepndent protein kinase Cs mediate heterologous desensitization of leucocyte chemokine receptors by opioid receptors. The Journal of Biological Chemistry, 278(15): 12729-12736.

47. Pestarino M., Mandich A., and Massari A. 1995. Evidence for enkephalin-like antigenic determinants in the gonads of the icefish *Chionodraco hamatus.* Polar Biology, (15): 125-130.

48. Owczarek D., Garlicka M., Pierzchała-Koziec K., Skulina D., Szulewski P. 2003. Met-enkephalin plasma concentration and content in liver tissue in patients with primary biliary cirrhosis. Przegl Lek., 60(7):461-466.

49. Ishida T., Kenmotsu M., Mino Y., Inoue M., Fujiwara T., Tomita K., Kimura T., and Sakakibara S. 1984. X-ray diffraction studies of enkephalins. Crystal structure of [(4'-bromo) Phe4,Leu5]enkephalin. Biochemical Journal, 218(3):677-689.

50. Jalkanen K.J. 2003. Energetics, structures, vibrational frequencies, vibrational absorption, vibrational circular dichroism and Raman intensities of Leu-enkephalin. Journal of Physics: Condensed Matter, (15): S1823-S1851.

51. Spirtes M.A., Schwartz R.W., Mattice W.L., and Coy D.H. 1978. Circulardichroism and absorption study of the structure of methionine-enkephalin in solution. Biochemical and Biophysical Research Communications, 81(2): 602-609.

52. van der Spoel D., Berendsen H. 1997. Molecular dynamics simulations of Leu-enkephalin in water and DMSO. Biophysical Journal, 72(5): 2032-2041.

53. Taylor J.W. and Osapay G. 1990. Determining the functional conformations of biologically active peptides. Accounts of Chemical Research, 23(10): 338-344.

54. Tóth F., Horváth G., Szikszay M., Farkas J., Tóth G, Borsodi A., and Benyhe S. 2004. Pharmacological and functional biochemical properties of D-Ala^2-D -Nle^5-enkephalin-Arg-Phe. Regulatory Peptides, 122(2): 139-146.

55. Sudha T.S. and Balaram P. 2009. Stabilization of β-turn conformations in enkephalins. 1982. Published online in International Journal of Peptide and Protein Research, 21(4): 381-388.

56. Lu Y., McNearney T., Wilson S., Westlund K. 2004. Endogenous pain modulation: Intra-articular treatment with enkephalin-encoding herpes simplex virus-1 produces a sustained hypoalgesia in a CFA-induced arthritis model. Journal of Pain, 5(3): S21.

57. Egleton R.D., Mitchell S.A., Huber J.D. Palian M.M., Polt R., and Davis T.P. 2001. Improved blood-brain barrier penetration and enhanced analgesia of an opioid peptide by glycosylation. Pharmacology and Experimental Therapeutics, 299(3): 967-972.

58. Gentilucci L., Federico Squassabia F., and Artali R. 2007. Re-discussion of the importance of ionic interactions in stabilizing ligand-opioid receptor complex and in activating signal transduction. Current Drug Targets, (8): 185-196.

59. Harmar, A.J., Arimura, A., Gozes, I., Journot, L.; Laburthe,. M., Pisegna, J. R., Rawlings, S. R., Robberecht, P., Said, S. I., Sreedharan,. P., Wank, S. A., Waschek, J. A. 1998.Nomenclature of receptors for vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide. Pharmacological Reviews, 50: 265-270.

60. Brenneman, D. E. 2007. Neuroprotection: A comparative view of vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide**.** Peptides, 28, 1720–1726.

61. Delgado, M., Pozo, D., Ganea, D. 2004. The significance of vasoactive intestinal peptide in immunomodulation. Pharmacological Reviews, 56, 249–290.

62. Gonzalez-Rey, E., Varela, N., Chorny, A., Delgado, M. 2007. Therapeutical approaches of vasoactive intestinal peptide as a pleiotropic immunomodulator. Current Pharmaceutical Design, 13, 1113–1139.

63. Vaudry, D., Gonzalez, B. J., Basille, M., Yon, L., Fournier, A., Vaudry, H. 2000. Pituitary adenylate cyclase-activating polypeptide and its receptors: from structure to functions. Pharmacological Reviews, 52: 269-324.

64. Onoue, S., Matsumoto, A., Nagano, Y., Ohshima, K., Ohmori, Y., Yamada, S., Kimura, R., Yajima, T., Kashimoto, K. 2004. Alpha-helical structure in the C-terminus of vasoactive intestinal peptide: functional and structural consequences. European Journal of Pharmacology, 485(1-3): 307-316.

65. Onoue, S., Misaka, S., Yamada, S. 2008. Structure-activity relationship of vasoactive intestinal peptide (VIP): potent agonists and potential clinical applications. Naunyn Schmiedebergs Archives of Pharmacology, 377: 579-590.

66. Chakder, S. and Rattan, S. 1993. The entire vasoactive intestinal polypeptide molecule is required for the activation of the vasoactive intestinal polypeptide receptor: functional and binding studies on opossum internal anal sphincter smooth muscle. Journal of Pharmacology and Experimental Therapeutics, 266: 392-399.

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

68. Goossens, J.F., Cotelle, P., Chavatte, P., Hénichart, J.P. 1996. NMR study of five N-terminalpeptide fragments of the vasoactive intestinal peptide: crucial role of aromatic residues. Peptide Research, 9: 322- 326.

69. Inooka, H., Endo, S., Kitada, C., Mizuta, E., 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.

70. Filizola, M., Cartenì-Farina, M., Perez, J.J. 1997. Conformational study of vasoactive intestinal peptide by computational methods. J. Pept. Res., 50:55-64.

71. Corcho, F., Canto, J., Perez, J. 2004. Comparative analysis of the conformational profile of substance P using simulated annealing and molecular dynamics. Journal of Computational Chemistry, 25: 1937-1952.

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

73. Hawkins, G.D., Cramer, .C.J., Truhlar, D.G. 1996. Parametized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. Journal of Physical Chemistry, 100:19824-19839.

74. Hawkins, G., Cramer, C., Truhlar, D. 1995. Pairwise solute descreening of solute charges from a dielectric medium. Chemical and Physics Letters, 246(1-2):122-129.

75. Wang, J., Cieplak, P., Kollman, P. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry, 21(12):1049-1074.

76. Sorin, E.J., Pande, V.S. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophysics Journal, 88(4):2472-93.

77. Luidens, M. K., Figge, J., Breese, K., Vajda, S. 1996. Predicted and trifluoroethanol-induced alpha-helicity of polypeptides. Biopolymers, 39: 367-76.

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

79. LaFargaCPL: CLASTERIT: Project Info. Available at: https://lafarga.cpl.upc.edu/projects/clusterit [Accessed June 12, 2008].

80. Feig, M.; Brooks III, C. L. 2004. Recent advances in the development and application of implicit solvent models in biomolecule simulations. Current Opinion in Structural Biology, 14: 217-224.

81. Chan, J., Brooks III, C. L., Khandogin, J. 2008. Recent advances in implicit solvent based methods for biomolecular simulations. Current Opinion in Structural Biology, 18(2): 140-148.

82. Guo, H., and Karplus, M. 1994. Solvent influence on the stability of the peptide hydrogen bond: A supramolecular cooperative effect. Journal of Physical Chemistry, *98*: 7104–7105.

83. 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.

84. Petzold, K., Olofsson, A., Arnqvist, A., Grobner, G., and Schleucher, J. 2009. Semiconstant-Time P,H-COSY NMR: Analysis of Complex Mixtures of Phospholipids Originating from *Helicobacter pylori*. Journal of the American Chemical Society, *131* (40), 14150-14151.

85. Senn, H.M., and Thiel, W. QM/MM Methods for Biomolecular stystems. Angewandte Chemie International Edition, 48(7): 1198-1229.

APPENDIX I

A1.1 COMPUTATIONAL DETAILS: MOLECULAR DYNAMICS PROTOCOL

In the current study molecular dynamics simulations were used to explore the folding patterns and dynamics of small and medium-size peptides. The simulations were carried out using mainly the AMBER 8 package. Amber is a collection of molecular simulation programmes, specifically for biological molecules. The acronym AMBER stands for Assisted Model Building with Energy Refinement. **Sander** is the main amber programme, used for structure minimization, equilibration, standard molecular dynamics simulations, and is also used for replica-exchange dynamics, thermodynamic integration, and potential mean force calculations. **LEaP** is an X-windows-based programme that provides for basic model building and Amber coordinate and parameter or topology creation. It contains a molecular editor to enable residue building and manipulation of molecules. **Ptraj** is an Amber component used to analyse MD trajectories, including hydrogen bonding analysis, dihedral angles, etc.

The MD protocol could be categorized into:

- model structure building
- structure energy minimization
- equilibration (short MD)
- production molecular dynamics

In the following sub-sections, pituaitary adenylate cyaclase-activating polypeptide 27 (PACAP27) will be used as an example of a peptide:

A1.1.1 Model structure development

The initial step in all modelling projects is developing the initial model structure. The structure is constructed with the aid of the **LEaP** module. The various different residues and their names are defined in library files that xleap loads on starting up the programme. The names used for the residues in the pdb files must match those defined in the default xleap library files or in user defined library files. The outputs of this process are the coordinate (crd) and topology (top) files that are used for the next step, i.e. structure energy minimization in **Sander**.

```
source leaprc.ff99f (Specifies the force field)
pacap27 = sequence {NHID SER ASP GLY ILE PHE THR ASP SER TYR SER ARG
TYR ARG LYS GLN MET ALA VAL LYS LYS TYR LEU ALA ALA VAL LEU NHE}
(amino acid sequence of the peptide)
saveamberparm pacap27 pacap27.top pacap27.crd (coordinate and parameter files)
& quit
```

- **crd:** the coordinates file contains the details of the initial set of co-ordinates. The data content of this file is not static and changes or is update during simulations.
- **top:** the topology or parameter file. It defines how atoms are connected and specifies the force field parameters chosen for the simulations. The contents of this file are static and do not change during simulations.

A1.1.2 Structure minimization before MD

The initial structure is energetically minimized with the Sander module to remove any strains in bonds or angles. Minimization with Sander is carried out using the script file:

```
#!/bin/bash
#$ -S /bin/bash
#$ -j y
#$ -cwd
#$ -pe lam-7.0.6_icc 2
# Set up the correct parallel environment
export WDIR="/usuaris/people2/paul/md99pOBC"
export AMBERHOME="/software/amber8"
export LAM_HOME="/software/lam-7.0.6_icc"
#####################################################################
echo "Got $NSLOTS processors."
echo "machines file: ${TMPDIR}/machines";
cat  $TMPDIR/machines; echo
$LAM_HOME/bin/mpirun -np $NSLOTS $AMBERHOME/exe/sander -O -i
$WDIR/min.in
-c $WDIR/pacap27.crd  -p $WDIR/pacap27.top  -ref $WDIR/pacap27.crd
-o $WDIR/pacap27.min.out -r $WDIR/pacap27.rst
```

The parameters in the input file **min.in** are specified as follows:

```
  imin   = 1, ntmin  = 2,
  maxcyc = 200000,
  drms   = 0.005,
  irest  = 0, ntx    = 1,
  ntb    = 1, ntp    = 0,
  ntpr   = 100,
  nsnb   = 10,
  cut    = 99.0,
  scee   = 1.2,
  ntb    = 0,
  ntx    = 1,
$ GENERALIZED BORN CONDITIONS
  igb=1, saltcon=0.2, gbsa=1,
 &end
```

A1.1.3 Equilibration step

After the minimization step, the next stage is the equilibration protocol, which is a short MD run (i.e. 10 ps) to allow the system to heat up from 0 to 300 K. In order to ensure that this happens without fluctuations in the peptide structure, restraints are applied.

The following script was used for the equilibration step:

```
!/bin/bash
#$ -S /bin/bash
#$ -j y
#$ -cwd
#$ -pe lam-7.0.6_icc 2
# Set up the correct parallel environment
export WDIR="/usuaris/people2/paul/md99pOBC"
export AMBERHOME="/software/amber8"
export LAM_HOME="/software/lam-7.0.6_icc"
#####################################################################
echo "Got $NSLOTS processors."
echo "machines file: ${TMPDIR}/machines";
cat  $TMPDIR/machines; echo

$LAM_HOME/bin/mpirun -np $NSLOTS $AMBERHOME/exe/sander -O -i
$WDIR/equil.in \
-c $WDIR/1000.rst  -p $WDIR/pacap27.top  -ref $WDIR/1000.rst \
-o $WDIR/out1a10ps -x $WDIR/crdde1a10ps -e $WDIR/ende1a10ps \
-inf $WDIR/min1001.inf -r $WDIR/1001.rst
```

The input file **"equil.in"** for the equilibration step is as follows:

```
  "equil.in"
  imin=0,
  ntc=2, ntf=2, cut=99.0,
  igb=2,
  saltcon=0.2, gbsa=1,(use salt concentration of 0.2 M)
  ntpr=50, ntwx=500, ntwr=500, ntwe=500, nsnb=10,
  nstlim = 5000, dt=0.002, (do simulation for 10 ps)
  ntt=1,
  tempi=0.0, temp0=300.0, (increase temperature from 0 to 300 K)
   tautp=1.0,
  ntx=1, irest=0, ntb=0,
  nscm = 1000,
&end
```

A1.1.4 MD run using an implicit solvent

The production MD simulation step was run using water as an implicit solvent and using the new crd file created from the minimization step as the input co-ordinate file for the MD run. To run a molecular dynamics simulation for 490 ps with sander the following script is used:

```
!/bin/bash
#$ -S /bin/bash
#$ -j y
#$ -cwd
#$ -pe lam-7.0.6_icc 2
# Set up the correct parallel environment
export WDIR="/usuaris/people2/paul/md99pOBC"
export AMBERHOME="/software/amber8"
export LAM_HOME="/software/lam-7.0.6_icc"
######################################################################
echo "Got $NSLOTS processors."
echo "machines file: ${TMPDIR}/machines";
cat  $TMPDIR/machines; echo

$LAM_HOME/bin/mpirun -np $NSLOTS $AMBERHOME/exe/sander -O -i
$WDIR/gbmd490.in \
-c $WDIR/1001.rst  -p $WDIR/pacap27.top  -ref $WDIR/1001.rst \
-o $WDIR/out10a500ps -x $WDIR/crdde10a500ps -e $WDIR/ende10a500ps \
-inf $WDIR/min1002.inf -r $WDIR/1002.rst
```

The input file "gb**md490.in"** is parametized as follows:

```
"gbmd490.in"
MD with GB at 300 K
&cntrl
   ntc=2, ntf=2,
   cut=99.0,
   igb=2,
   saltcon=0.2, gbsa=1,
   ntpr=50, ntwx=500, ntwe=500, nsnb=10,
   nstlim = 245000, dt=0.002, (do simulation for 490 ps)
   ntt=1,
   tempi=300.0, temp0=300.0, (maintain a temperature of 300K)
   tautp=2.0,
   ntx=5, irest=1, ntb=0,
   nscm = 1000, &end
```

Taking the equilibration step into account, the simulation is now 500 ps long. From this point onward the simulations were done in steps of 500 ps until 200 000 ps was reached. The following script was used for this purpose:

```
!/bin/bash
#$ -S /bin/bash
#$ -j y
#$ -cwd
#$ -pe lam-7.0.6_icc 2
# Set up the correct parallel environment
export WDIR="/usuaris/people2/paul/md99pOBC"
export AMBERHOME="/software/amber8"
export LAM_HOME="/software/lam-7.0.6_icc"
######################################################################
echo "Got $NSLOTS processors."
echo "machines file: ${TMPDIR}/machines";
cat  $TMPDIR/machines; echo

$LAM_HOME/bin/mpirun -np $NSLOTS $AMBERHOME/exe/sander -O -i
$WDIR/gbmd.in \
-c $WDIR/1002.rst  -p $WDIR/pacap27.top  -ref $WDIR/1002.rst \
-o $WDIR/out500a1000ps -x $WDIR/crdde500a1000ps -e
$WDIR/ende500a1000ps \
-inf $WDIR/min1003.inf -r $WDIR/1003.rst

$LAM_HOME/bin/mpirun -np $NSLOTS $AMBERHOME/exe/sander -O -i
$WDIR/gbmd.in \
-c $WDIR/1003.rst  -p $WDIR/pacap27.top  -ref $WDIR/1003.rst \
-o $WDIR/out1000a1500ps -x $WDIR/crdde1000a1500ps –e
$WDIR/ende1000a1500ps \
-inf $WDIR/min1004.inf -r $WDIR/1004.rst

etc.
```

The input file "**gbmd.in**" is as follows:

```
"gbmd.in"
MD with GB at 300 K
&cntrl
   ntc=2, ntf=2,
   cut=99.0,
   igb=2,
   saltcon=0.2, (use salt concentration of 0.2 M)
```

```
  gbsa=1,
  ntpr=50, ntwx=500, ntwe=500, nsnb=10,
  nstlim = 250000, dt=0.002, (do simulation for 500 ps)
  ntt=1,
  tempi=300.0, temp0=300.0, (maintain temperature at 300 K)
 tautp=2.0,
  ntx=5, irest=1, ntb=0,
  nscm = 1000,
&end
```