Abstract

Understanding biological function at molecular level involves studying the spontaneous folding patterns of peptides to unique and biologically-active three-dimensional structures. Peptides are highly flexible in nature and fold at fast rates hence require probing at nanosecond time scale. Bioactive conformational structures of folded peptides need to be well characterized before using them in computer-aided drug design. Molecular dynamics (MD) simulations provide detailed information on the fluctuations and their affect.

Cathelicidins are a novel family of antimicrobial peptides (AMP) characterized by conserved pro-peptide sequences and a common N-terminal "cathelin" domain which have been identified in several mammalian species. Cathelicidins are cationic, frequently α-helical, amphipathic host defense peptides and provide the first-line of defense against infection by serving as "natural antibiotics". Recent evidence has indicated that the sole human cathelicidin, LL-37 so called as it starts with two leucines, possesses a broad range of chemotactic and immunostimulatory/ modulatory effects. It has the ability to neutralize bacterial lipopolysaccharide (LPS) attributed to its C-terminal end and multiple roles influencing diverse processes such as cell proliferation and migration, wound healing apoptosis and angiogenesis. LL-37 also enhances the responsiveness of hematopoietic stem cells to homing factors primarily because of its SDF-1-specific priming effect. Thus, cathelicidin antimicrobial peptides qualify as prototypes of innovative drugs that may be used to treat infection and modulate the immune response.

The present study attempts to get a deeper insight into the conformational features of cathelicidin focused at establishing structure-activity relationships and investigates the propensities of cathelicidin to adopt different conformations by using different computational protocols.

Aims and Objectives

- The aim of this study was to use computational methods to assess the conformational profiles of antimicrobial peptide cathelicidin.

- Investigate the effect of different force fields on the folding patterns of peptides to get a deeper insight into the conformational features of cathelicidin focused at establishing structure activity relationships.

- Comparison of trajectories obtained from two different molecular dynamics softwares AMBER12 and GROMACS.

- Evaluate the performance of different force fields ff96, ff99 and ff99SB in AMBER12 and gromos96 43a1 in GROMACS. Compare different force fields and their use in computational approaches for exploring the energy landscape of peptides.

- Investigate the effect of Berendsen in GROMACS and Langevin thermostat algorithm in AMBER on the folding pattern of cathelicidin.

Hence a series of 100 ns long molecular dynamics (MD) simulations were applied to cathelicidin antimicrobial peptide to assess the thermodynamics of its folding characteristics.

Introduction

Simulations enable us to characterize and predict the behavior of real systems. Simulations aimed at understanding the mechanism of peptide folding are a considerable challenge. The quality of the model used and the available computational power are the factors that determine extend to which this goal is achieved. No model can be an exact representation of the real system. It needs to be verified and compared to experimental data.

This study involved a series of molecular dynamics (MD) simulations using AMBER (Case 2006) and GROMACS (Lindahl E) programs applied to Cathelicidins– A family of antimicrobial peptides (AMP) using the sole human cathelicidin LL-37 (Benachour, Zaiou *et al*. 2009) . Structure chosen for same was linear Cathelicidin and the NMR structure (PDB ID: 2K6O).

Cathelicidin antimicrobial peptides qualify as prototypes of innovative drugs that may be used to treat infection and/or modulate the immune response. The modes of action by which AMPs kill bacteria is varied and includes disrupting membranes, interfering with metabolism, and targeting cytoplasmic components. The initial contact between the peptide and the target organism is electrostatic, as most bacterial surfaces are anionic, or hydrophobic (Bals, Weiner *et al.* 1999).

While experimental knowledge is essential for the understanding of the structure and dynamical properties of peptides in solutions, theoretical studies involving computer simulations complement the experimental data. In particular, molecular dynamics (MD) simulations provide detailed information on the fluctuations and their effect. Generalized Born/surface area (GB/SA) implicit solvent models have recently played a significant role in protein-protein docking, the NMR-derived structures and the molecular conformational changes of proteins and nucleic acids, and are routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes (Onufriev 2004). In the MD approach, Newton's or Lagrange's equations of motion are solved by numerical integration, which provides much higher efficiency of a single step compared to other sampling methods such as canonical Monte Carlo (MC) simulations (Karplus M 2002).

To investigate the propensities of a peptide to adopt different conformations, the computational methodologies available provide detailed complementary information about the intrinsic conformational features. Especially solvent dynamics (MD) simulations allow sampling of limited regions of the phase space in different peptides, proteins, and nucleic acids. These models have successfully been used to study biological systems in aqueous solvents and more recently in the context of biological membranes described by multiple layers with different dielectric constants (Caflisch and Karplus 1994).

In order to assess the folding pattern of cathelicidin and compare the results with the NMR results in SDS, simulations were performed on the extended structure. Implicit solvent models were considered as a good approximation specifically Generalized Born (GB) was selected to treat the solvent since. Different force fields ff96, ff99, ff99SB in AMBER 12 and gromos96 43a1 in GROMACS were selected for the computation of energy. Shake algorithm was used to constraint the bonds involving hydrogen atoms with a time step of 2 fs. Four 100 ns MD trajectories were run at 300K using langevin thermostat and Onufriev, Bashford and Case (OBC) implementation of the Generalized Born (GB) approximation. All calculations in the present work have been carried out using AMBER 12. The analysis of the trajectories obtained was using the PTRAJ module in AMBER. CLASICO (https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho) software was used to analyze the formation and destruction of secondary structures to be monitored during the folding process, as well as in the characterization of the group of structures that represents the folded molecule.

Chapter 1

Review of literature

1.1 Peptides

Amino acid is the basic monomeric unit of a protein and peptide bonds between amino acids are responsible for the formation of primary structure of a protein. A peptide bond is formed when the carboxy group of the first amino acid condenses with the amino group of the next to eliminate water. Polymers of amino acids connected by amide bonds (peptide bonds) between the carboxy group of one residue and the amino group of the following amino acid are termed as peptides (Buehler 2009).

Peptides are key modulators of most cellular functions such as signal transduction, immune response, and catalysis of metabolic reactions both under normal physiological and diseased conditions (Buehler 2009). Peptides are highly diversified and their biological function is governed by the nature of the amino acids present in the sequence, the length of the peptide and the spatial arrangement to a specific three-dimensional structure. Hence understanding the basic properties of the amino acids present in nature is required to relate the peptide folding with its biological function. Knowing the thermodynamics and mechanisms of peptide and protein folding remains one of the yet unconquered challenges in modern molecular biology.

1.2 Protein/Peptide Folding

Folding of peptidic agonists to their specific and precise three-dimensional conformations is needed for attaining their biological activity. All information determining the native structure is fully contained in the amino acid sequence of a protein. The self-assembly process by which the polypeptide chain having a non-structured or random coil unfolded configuration folds and attains its characteristic functional three dimensional structure is referred to as Peptide or protein folding (Leach 2001).

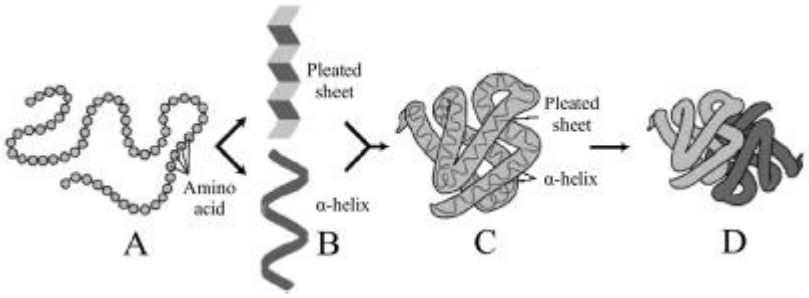


Figure 1: Hierarchical levels in protein folding (*Image courtesy: http://eduframe.net/andc/biology/Ravi-Bio/Ravi20proteinsODL.htm)*

The structure of proteins and peptides can be classified into four basic levels (primary, secondary, tertiary and quaternary) depending upon their complexity.

The primary structure refers to the linear amino acid sequence held together by covalent bonds- peptide bond. Secondary structure corresponds to the highly regular local sub-structures, with the α-helix and the β-strand being the two main types of secondary structures defined by patterns of hydrogen bonds between the main-chain peptide groups.

An α-helix is formed when hydrogen bonding occurs between the i and i + 4 residues. On the other hand a β-sheet is formed when two strands are bridged by hydrogen bonds involving alternating residues on each participating strand (Balamurugan, Gopalakrishnan *et al.* 2010). A tertiary structure represents the three dimensional (3D) structure of a single protein molecule where the α-helices and the β-sheets are folded into a compact globule. Hydrogen bonds also play an important role in the formation of tertiary structures through the interaction of side chains. Quaternary structure is a larger assembly of several protein molecules or polypeptide chains stabilized by the non-covalent interactions and disulfide bonds similar to those present in tertiary structures (Kabsch 1983).

There are three aspects associated with the study of protein folding namely
1. Thermodynamics properties.
2. Kinetics.
3. Structure prediction.

A protein consisting of n amino acids can in principle have 3^n conformations, due to the three possible combinations of the phi and psi angles. If the protein randomly searches all the available conformations, it would take longer than the lifetime of the universe, known as the Levinthal's paradox (Levinthal 1968). To investigate protein folding different computational models have been used to such as simple lattice models and atomistic models. Considering the fact that the interactions at atomic level play a crucial role in the equilibrium between folded and unfolded state, MD simulation is, a valuable tool to describe the folding of a peptide or protein as it sheds light in the governing factors of the folding pathways. The transition to the folded state never occurs directly from a fully unstructured conformation, the peptide follows a stepwise pathway through partially structured-unfolded conformation that finally triggers the transition state to the final folded hairpin structure (Daidone I 2005, Thukral L 2009)

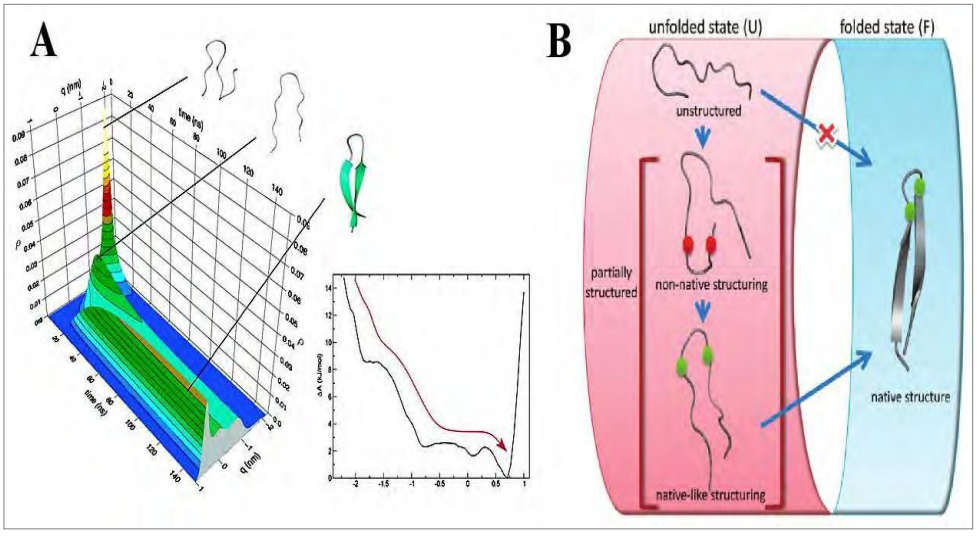


Figure 2: Representation of a folding process by MD simulations. **A**. Kinetics of the folding of a β-structure. **B.** The folded state is not reached directly from fully unstructured conformations. Instead, it is formed via partially structured-unfolded conformation to β-hairpin. (*Image courtesy:J Am Chem Soc. 2009 Dec 23;131(50):18147-52.*)

Christian Anfinsen who was awarded the nobel prize demonstrated that the correct refolding of the unfolded Ribonuclease A into its native and enzymatically active structure occur spontaneously in free solution (Taniuchi H 1969). Protein folds along certain pathways in sequential steps which simplify the folding process. Intermediates of this pathway being the stabilized folded structures obtained which define the individual steps of the pathway. Folding intermediates possesses stabilized structural elements, mainly of secondary structural origin, in combination with the unstructured regions (Baldwin 1996).
The folding kinetics behaves like a diffusive process over the free potential energy surface, exhibiting a funnel-like almost barrier less free energy profile. The free energy profile describing the main conformational transitions showed the characteristic features of a funnelled free energy landscape, with a downhill surface towards the basin (Baldwin 1999). The folding process is described by an energy landscape or a folding funnel with a vast array of down-hill routes to the native state in a more or less rugged surface (Radford 2000). Dips in the funnel wall indicate the local energy minima in which the proteins may get trapped in unfavourable intermediate states during the folding process.
The native state of a protein is described thermodynamically as the free energy minimum of all possible structures populated according to the Boltzmann distribution (Schultz 2000.) Solvation effects may enhance entropic contributions on different structures and have a statistical significance. There are various methods to calculate solvation free energy of a conformation that may be added as an additional term to the intramolecular energy. A denatured protein whether would be prone to intramolecular aggregation or would reach the native state efficiently will depend on the rate of the folding process, i.e., how fast a globular structure is reached in which the hydrophobic surfaces are minimally exposed. How a given amino acid sequence encodes a defined three-dimensional structure is however not fully understood (Onuchic 2004).

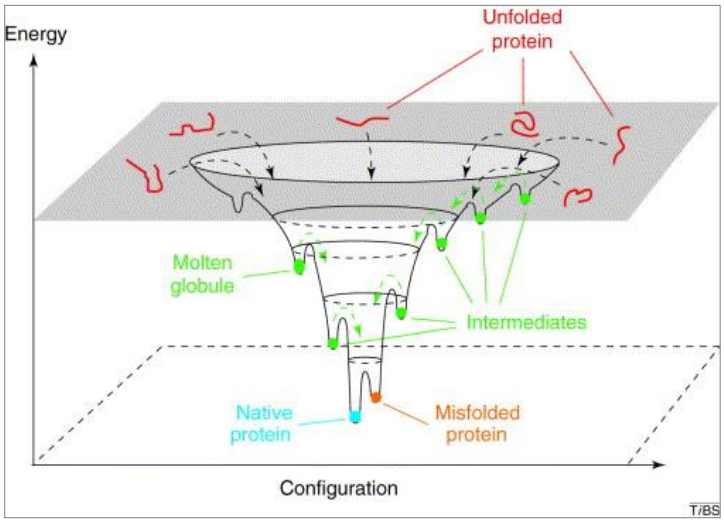


Figure 3: Schematic diagram of a folding energy landscape for a protein molecule (Funnel type. Denatured molecules at the top of the funnel might fold to the native state by a myriad of different routes, some of which involve transient intermediates (local energy minima) whereas others involve significant kinetic traps (misfolded states). For proteins that fold without populating intermediates, the surface of the funnel would be smooth. (*Image courtesy: Protein folding progress made and promises ahead. Vol 25, Issue 12, 2000, 611–618 Trends in Biochemical sciences).*

1.3 Stem Cells and Homing

One of the human body's master cells, having the ability to grow into any type of the body‘s cell types are stem cells. They are unspecialized, undifferentiated or ‘blank’ cells and divide indefinitely (symmetrical divisions) and retain this ability throughout life (Jiang Y 2002). Stem cells are a clonal, self-renewing entity being multipotent. Unlike mature differentiated cells, committed to their fate permanently, stem cells have self renewal ability i.e. generating daughter cells identical to their mother and even producing progenitor cells with more restricted potential committed to give rise to fully differentiated cells.

Homing is the term given to the ability of a cell to find their destination or "niche. Active navigation is required for migration of cells through blood, across the endothelial vasculature to different organs and to bone marrow niches.

Homing being a rapid process and is the first and essential step in clinical stem cell transplantation. It is a step wise process involving at first rolling of cells leading to binding and activation of adhesion proteins and the chemo attractants leading them to the target site through the endothelium (Beckers 1998). Specific cues that steer stem cells to their target niche and increase the efficiency of the homing is an area of intense investigation. The homing efficiency determines the success of transplantation of hematopoetic stem cells and their engraftment and repopulation in the bone marrow

CXCR4, a chemokine receptor, a member of a member of family of seven-transmembrane domain receptors coupled to heterotrimeric G proteins specifically interacts with the chemokine CXCL12 or stromal cell-derived factor (SDF-1) that is constitutively produced at high levels in the bone marrow. SDF-1 is known to be a potent chemoattractant for hematpoetic stem cells playing an essential role in the development of blood cells.

SDF-1–CXCR4 axis plays a role in stem cell homing during embryogenesis and adulthood (Hongbao Liu*. 2012). The binding of CXCR4 to SDF-1 results in the activation of multiple signal transduction pathways which ultimately trigger chemotaxis to regulate cell adhesion, survival and cell-cycle. The SDF-1-CXCR4 axis has been shown to play a pivotal role in homing; retention and mobilization of HSCs. SDF-1/CXCR4 interactions and signaling have been described as a principal axis regulating retention, migration, and mobilization of hematopoietic stem cells. The migrating cells from the mesoderm and endoderm contain CXCR4 messages while SDF-1 is expressed by embryonic ectoderm cells. (Michael Rosu-Myles 2000).

1.4 Antimicrobial Peptides

Cytolytic peptides produced by living organism that serve as host defence peptides and provide the first-line of defense against infection are called as Antimicrobial peptides. Being widely distributed in nature they represent an ancient mechanism of host defense (Ganz and Lehrer 1999). Initially were found in invertebrates (Agerberth, Lee *et al*. 1991) but now also have been reported in vertebrates, including humans. They have the ability to rapidly kill a broad range of microbes and thus play an important role in innate immunity to microbial invasion. They have a unique mode of action which has made such peptides promising candidates for the development of a new class of antibiotics. Antimicrobial peptides are promptly synthesized at low metabolic cost in living organisms of all types and are stored in large amounts. They are instantly available shortly after an infection to neutralize the invader microbes. They are known to play crucial role in the innate as well as the adaptive defense systems (Yang, Chertov *et al.* 2001). As they exhibit rapid killing, often within minutes in vitro, and a broad spectrum of activity against various targets, including Gram-positive and -negative bacteria, fungi, parasites, enveloped viruses, and tumor cells.

In higher organisms these are produced mainly by the epithelial surfaces and in phagocytic cells. AMPs differ from conventional antibiotics in their mechanism of killing and can be considered as attractive substitute and/or additional drugs. The key for the design of peptide-antibiotics (Lohner, Latal *et al*. 2001) lies in elucidation of the mechanisms of the action of antimicrobial peptides and their specific membrane damaging properties, i.e. their selectivity to a specific target.

It has been well proven that many antimicrobial peptides act directly on the membrane of the target cell (Ahmad, Asthana *et al*. 2009). Selectivity and preferential killing properties of AMPs are due to the net positive charge of these peptides causes their preferential binding to more than one negatively charged targets on bacteria, accounting for the selectivity of antimicrobial peptides (Hancock and Scott, 2000).

It has been reported that not only the nature of the peptide but also the characteristics of cell membrane, as well as the metabolic state of the target cells determine the mechanisms of action of antimicrobial peptides (Liang and Kim 1999).The advantage AMPs have as defense weapons is the non specificity of their mechanism, making it difficult for microbial pathogens to develop resistant mutants to overcome peptide intervention. In this present scenario imicrobes are increasingly developing resistance to antibiotics and hence using AMPs is advantageous and they have great a potential to be the next breakthrough and truly novel class of antibiotics. The key to the development of therapeutic agents lies in understanding how, when and where they function. They are already being used as antibiotics for topical use in healthcare or as preservatives in the food industry (Rao, Shinnar *et al*. 2000).

1.4.1 Structural parameters

The structural parameters that govern the action of AMPs are charge, conformation, hydrophobicity, hydrophobic moment, and size.

Charge

One of the indispensable features for the antimicrobial activity of AMPs is Cationicity. As the bacterial membrane is negatively charged the first step for AMPs activity is the electrostatic interaction between the positive charge of some of the amino acids such as arginine and lysine present in the sequence and this negative charge. Cationicity and AMP activity has a strong relationship and hence number of positively charged residues is an important feature. (Tossi, Bonin *et al*. 2000). Just the presence of the positive charge isn't the only issue that has to be taken into account for antimicrobial activity. The density of charge i.e. arrangement or the pattern of the charged amino acid residues along the sequence is also an important aspect for the action of the AMPs (Laugwitz, Weig *et al.* 2001).

Amphipathicity

Amphipathicity is the ability of a peptide to arrange all hydrophobic residues in one side and all the hydrophilic residues to the opposite side. Membrane is amphipathic in nature and due to this peptide amphipathicity has been revealed as an important parameter for AMP activity. When peptide interacts with the membrane bilayer and inserts itself, the hydrophobic residues of the peptide interact with the apolar membrane lipid tails and cluster together whereas the hydrophilic residues of the peptide interact with the polar head groups of the membrane and the aqueous environment. The problem with amphipathy is it is difficult to quantify the relationship between the activity and the peptide amphipathicity. Eisenberg *et al.* proposed the hydrophobic moment (MH), being the vectorial sum of individual amino acid hydrophobicities normalized to an ideal α-helix as a quantitative measure for peptide amphipathicity (Laugwitz, Weig *et al.* 2001).

Hydrophobicity

The percentage of hydrophobic residues within the peptide is an elememtal feature that indicates the degree of interaction of the peptide with the core of bilayer. As peptide hydrophobicity increases, binding to all membrane types increases thereby reducing membrane selectivity (Laugwitz, Weig *et al.* 2001). Hence AMPs usually have moderate amount of hydrphobic residues and show higher affinity towards microbial membranes.

1.5 Cathelicidin

Cathelicidins are a family of cationic, frequently α-helical, amphipathic antimicrobial peptides (AMP), which provide the first-line of defence against infection by serving as "natural antibiotics". They possess a wide spectrum of activity as they have been known to eradicate all classes of pathogens including bacteria, fungi, protozoa and virus at micromolar concentrations (Zanetti, Gennaro *et al.* 1995). Also showing a chemotactic and immunostimulatory/-modulatory effect they are capable of inducing wound healing, angiogenesis and modulating apoptosis. They were identified in early 90's by the presence of a conserved N terminal domain in bovine bone marrow-derived myeloid cells. Analysis of the cDNA sequence of bovine neutrophil bactenecin Bac5 led to the discovery of other peptides with the same proregion (Zanetti, Del Sal *et al.* 1993). Now the family has over 50 peptides. The first antimicrobial peptide of this group was isolated from pig (Agerberth, Lee *et al.* 1991) . The term cathelicidins was introduced in 1995 by Zanetti *et al* (Zanetti, Gennaro *et al.* 1995*)* and contain both a conserved cathelin domain and a C-terminal antimicrobial domain. Cathelin is an acronym for cathepsin L inhibitor. Cathelicidins are highly variable family of antimicrobial peptides with their precursor proteins sharing a common N-terminal "cathelin" domain. Cathelicidins thus have a conserved N terminal domain and a highly variable C terminal.

C terminal has a highly heterogeneous structure depending on the attached peptide (Zanetti 2004). For example a member of this family indolicidin, a cationic AMP isolated from bovine neutrophils has 13 amino acids and a high content of tryptophan residues (around 39%) (Subbalakshmi and Sitaram 1998). Another member bactenicin, a highly cationic polypeptide isolated from the large granules of bovine neutrophil composed of around 45% of proline and 23% of arginine residues and adopts an extended secondarystructure with turns and coils

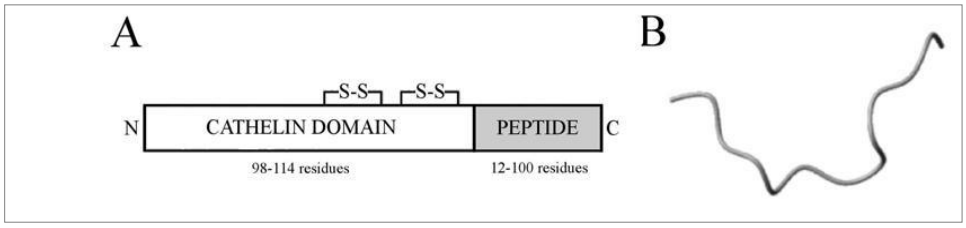


Figure 4: Structure of AMPs with unusual proportion of regular amino acids. (**A**). Cathelicidins (**B**). Indolicidin. (*Image courtesy: Zanetti 2004)*

Cathelicidin family is divided into into three groups:

1. First containing 12-18 residues with beta-hairpin structures stabilized by one or two disulfide bonds. This group also includes a 13-residue linear peptide with a high content of tryptophans called as indolicidin (Xia, Tsui *et al*. 2002, Li, Li *et al.* 2006)
2. Second group containing 23- 37 residues and has the potential to form a helical structure. LL-37 is a member of this group (Xia, Tsui *et al*. 2002, Li, Li *et al.* 2006).

3. Third group having peptides such as PR-39 are rich in pralines with 39-80 residues (Xia, Tsui *et al*. 2002, Li, Li *et al.* 2006).

Cathelicidins become active when the C-terminal domain is cleaved from the cathelin domain. In humans the sole human cathelicidin identified till date is a 37 residue long peptide having two leucines at its *N*-terminal, hence the acronym LL-37. The gene which codes for LL-37 is named CAMP located on chromosome 3 at the p21 locus (Agerberth, Gunne *et al.* 1995).

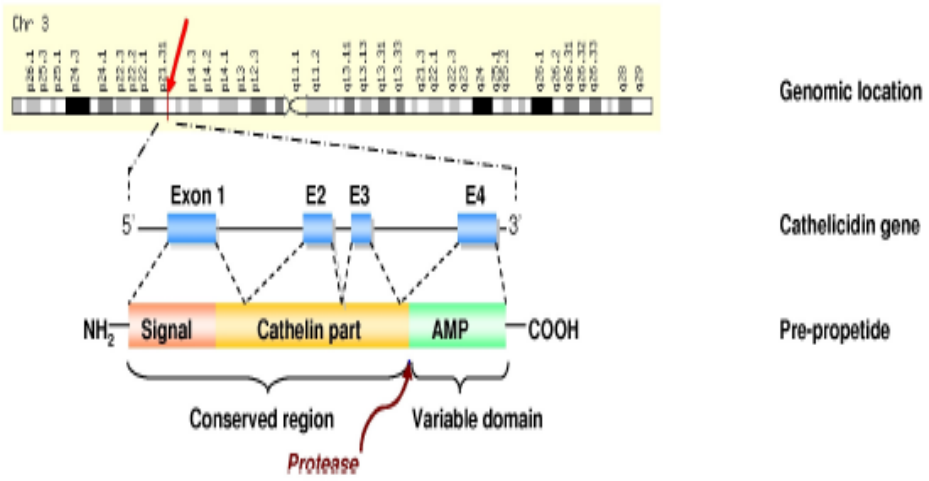


Figure 5: Location of the cathelicidin gene on the human genome and its structure. (*Image courtesy: Pharmaceuticals* ***2010****, 3, 3435-3460; doi:10.3390/ph3113435)*

Several animal models have been used to demonstrate the protective role of human host defense cathelicidin LL-37 against infection. The activity of LL-37 is regulated by proteases at two levels: generation from its precursor protein hCAP-18 and degradation to smaller fragments, which may, or may not, be biologically active. Several active fragments of LL-37 have been detected including LL-23, LL-29 in human skin (Yamasaki, Schauber *et al.* 2006). In vagina which is cleaved by, gastricsin, a prostatic protease, (Sorensen, Gram *et al*. 2003). In human sweat the LL-37 is degraded to shorter peptic fragments RK-31, KS-30 and KR-20 (Lee, Ohtake *et al*. 2005). The protease that cleaves hCAP-18 into LL-37 in neutrophils is proteinase 3, while other proteases such as kallikreins are proposed to cut LL-37 into smaller fragments. (Yamasaki, Schauber *et al*. 2006).

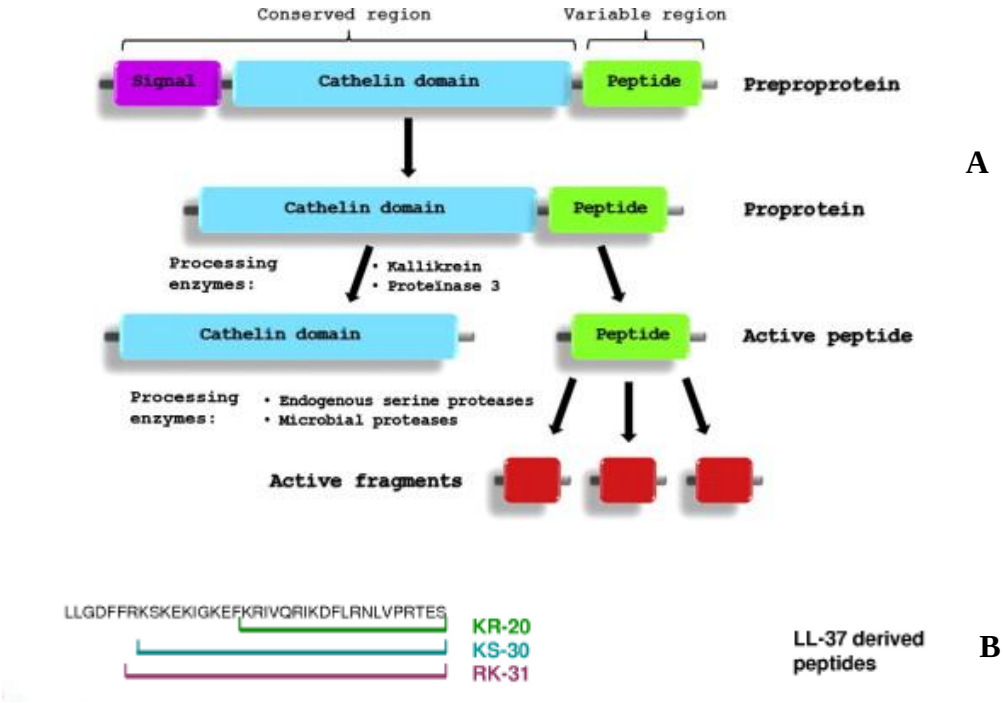


Figure 6: (**A)**. Enzymatic processing of a cathelicidin. The preproprotein comprises of a conserved cathelin domain and a variable region. The variable region codes for the antimicrobial cathelicidin peptide. Serine proteases are responsible for processing of cathelicidin.(*Image courtesy: Cell Immunol. 2012 Nov;280(1):22-35. doi: 10.1016/j.cellimm.2012.11.009. Epub 2012 Nov 29.*) **(B)**. Sequence of various fragments of LL-37.(*Image courtesy: Pharmaceuticals **2010**, 3, 3435-3460; doi:10.3390/ph3113435)*

1.6 Structure of LL-37 (PDB ID: 26KO)

LL-37 has a well-defined long amphipathic alpha helix covering residues 2–31 with the remaining C-terminal residues disordered. This structure has been elucidated using three-dimensional triple-resonance NMR spectroscopy, in SDS micelles with ^{13}C, ^{15}N-labeled LL-37. Residues 2–32 are ordered and the C terminus is mobile. There is a bend between residues Gly-14 and Glu-16. The hydrophobic packing between residues Ile-13 and Phe-17 plays an important role in causing the helical bend in the structure of LL-37. Aligned on the concave surface is a train of hydrophobic side chains: Leu-2, Phe-5, Phe-6, Leu-13, Phe-17, Ile-20, Val-21, Ile-24, Phe-27, Leu-28 and Leu-31. There is an aromatic-aromatic stacking between Phe-5 and Phe-6 also observed in the N-terminal fragment of LL-37 (Wang 2010).

In the presence of a phospholipid bilayer or short-tail phospholipid micelles, LL-37 preferentially adopts an alpha-helical conformation as shown by FTIR (Tomasinsig, Pizzirani *et al.* 2008) and NMR (Porcelli, Verardi *et al.* 2008, Wang 2010).
As the alpha-helix is amphipatic, it is expected to burry its hydrophobic residues in the hydrophobic part of the bilayer while the polar/charged residues should stay at the level of the lipid polar head groups. This topology is consistent with NMR data (Wang 2010)

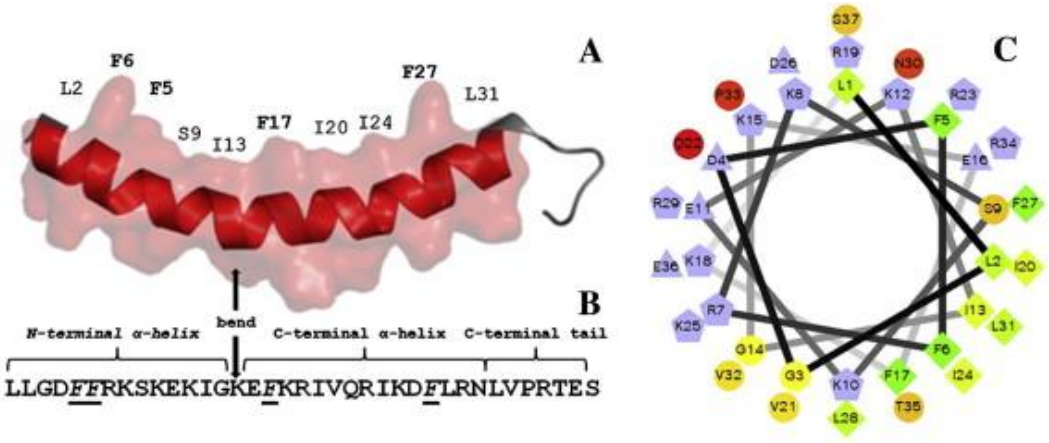


Figure 7: 3-dimensional structure of LL-37. **(A)** The three-dimensional structure of the human cathelicidin reveals a helix-bend-helix conformation, followed by a disorderd C-terminal tail. **(B)**Sequence of LL-37. **(C)** Helical wheel representation of LL-37, illustrating the amphipathic and cationic nature of LL-37. Hydrophilic residues are presented as circles, hydrophobic residues as diamonds, potentially negatively charged as triangles, and potentially positively charged as pentagons. Hydrophobicity is color coded with the most hydrophobic residue coloured green, and the amount of green is decreasing proportionally to the hydrophobicity, with zero hydrophobicity coded as yellow. Hydrophilic residues are coded red with pure red being the most hydrophilic (uncharged) residue, and the amount of red decreasing proportionally to the hydrophilicity. *Image courtesy: Cell Immunol. 2012 Nov;280(1):22-35.*

1.7 Mechanism of Action of Antimicrobial Peptides: Cathelicidin

AMPs have many mechanisms by which they interact with the membrane. The mechanism by which cathelicidins neutralize the bacterial membrane can be explained with a Carpet-like model in which peptides through a relatively diffuse manner (Bucki, Leszczynska *et al.* 2010). A localized carpet is formed on the membrane surface by the accumulation of peptides. Peptide chains get parallel oriented with respect to the membrane surface and so they become trapped in the head- group region. This accumulation creates tension between the two leaflets of the bilayer, leading to disintegration or rupture of the membrane. Transient holes get formed in the bilayer which may be an intermediate step before the membrane collapses. This model is considered to be the most general per-meabilization mechanism. Other AMPs that follow this model are cecropins (Gazit 1994) and dermaseptins (Strahilevitz 1994).

1.8 LL-37 and Stem Cell Homing

It had been shown that immunomodulatary agents like Cathelidicin and β-defensins have been shown to play a role in stem cell homing (Wan Wu 2012 April). LL-37 is known to enhance the responsiveness of hematopoetic stem cells to homing factors as it has a SDF-1-specific priming effect. It enhances chemotactic responsivenes to a low gradient of SDF-1 and selectively primes the responsiveness of HSPCs to SDF-1. Even at low levels of SDF-1 (1–2ng/ml) LL-37 has the ability to enhance of HSPCs migration of supporting the notion

that without the need of increasing SDF-1 secretion, this phenomenon of priming modulates the responsiveness of HSPCs to physiological concentrations of SDF-1. LL-37,has primarily antimicrobial functions and is harmless to mammalian cells and their property to prime SDF-1–CXCR4 receptor axis and hence play an important role in retention of HSPCs in BM. (Wan Wu 2012 April)

1.9 Molecular Modelling

Molecular modelling comprises of all theoretical methods and computational techniques utilized to model or mimic the behaviour of molecules. The techniques are used in the fields of computational chemistry, drug design, and computational biology for studying molecular systems. Molecular modelling techniques deal with the atomistic level description of the molecular systems which include treating atoms as the smallest individual unit (the Molecular mechanics approach), or explicitly modeling electrons of each atom (the quantum chemistry approach).

Most molecular modelling studies can be divided in three stages.

1. Selecting a molecule and describing the intra- and inter- molecular interactions in the system. These models enable calculation of energy of the arrangement of the atoms and molecules in the system to be calculated, and allow the determination of the change in energy values with respect to change in position
2. Second stage refers to calculating the various parameters such as an energy minimization, a molecular dynamics or Monte Carlo simulation, or a conformational search.
3. Third stage is the analysis of the calculation performed to calculate properties and also to check that it has been performed properly.

Molecular modelling methods are used to investigate the structure, dynamics, surface properties and thermodynamics biological systems. The types of biological activity that have been studied using molecular modelling include protein folding, enzyme catalysis, protein stability, conformational changes associated with biomolecular function etc (Rahman 1964).

1.10 Molecular Mechanics (MM)

When the laws of classical physics are put to use for predicting the structure and properties of molecular systems it is known as Molecular mechanics (MM). MM also known as force-field methods predict the energy of a molecule as a function of its conformation by using classical type models. This enable prediction of equilibrium geometries and transition states (Smit 1980)

The atoms in a molecule have different characteristics and show different behaviour, depending upon their environment in the structure. For example, a carbon atom in a carbonyl group is treated differently and will exhibit a different behaviour from one bonded to three hydrogens. In MM a molecule is considered to be a collection of masses

interacting with each other through harmonic forces with the atoms in a molecule been are treated as balls of varying sizes which are joined together by springs of variable strength and equilibrium distances (bonds) (Aleman 2009).
This simplification makes molecular mechanics a fast computational model which can be applied to molecules of any size. The total energy is made up of a sum of different contributions that calculate the deviations from the equilibrium values of bond lengths, angles and torsions plus non-bonded interactions.
A simple molecular mechanics energy equation is written as:

Total Energy = Stretching Energy + Bending Energy + Torsion Energy + Non-Bonded Interaction Energy

Meaning that within the molecular framework, the "total energy" of a molecule is described in terms of a sum of contributions arising from ALL DEVIATIONS from "ideal" bond distances (stretch contributions), bond angles (bend contributions) and dihedral angles (torsion contributions) summarized by:

$$E^{Total} = \sum_{i}^{bonds} E_i^{stretch} + \sum_{i}^{bondangles} E_i^{bend} + \sum_{i}^{dihedralangles} E_i^{torsion} + \sum_{ij}^{atompairs} E^{vanderwaals} + \sum_{ij}^{atompairs} E^{electrostatic}$$

Covalent Interactions — Non - covalent Interactions

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 (Corcho).
These equations together with the data (parameters) required to describe the behavior of different kinds of atoms and bonds, is called a **"FORCE FIELD".** A force field is a collection of a set of equations that define how the potential energy of a molecule varies with respect to locations of its component atoms, and a series of defined atom types (Levitt 1976).

1.11 Molecular Dynamics (MD)

Molecular dynamics (MD) is as a form of computer simulation in which the atoms and molecules are allowed to interact for a period of time by approximations of known physical attributes, resulting in the simulation of the motion (Marx 2000, Post 2006) for a system of particles. As it is **i**mpossible to find the properties analytically, MD simulation circumvents this problem by using numerical methods by representing an interface between laboratory experiments and theory, and can be called as a "virtual experiment". MD probes the relationship between molecular structure, movement and function with its laws and theories coming from various interdisciplinary fields of mathematics, physics, and chemistry, and employing algorithms from computer science (Marx 2000). First introduced by Alder and Wainwright in the late 1950's to study the interaction of hard spheres (Alder BJ 1959) it

provides the detailed information about the fluctuation and conformational changes in the protein molecule over the certain period of time.. The central idea behind the molecular dynamics is to be able to study the biological activity in context of time dependent interaction between molecules (Lindahl E).A graphical flowchart for the general molecular dynamics protocol is shown in Figure 8

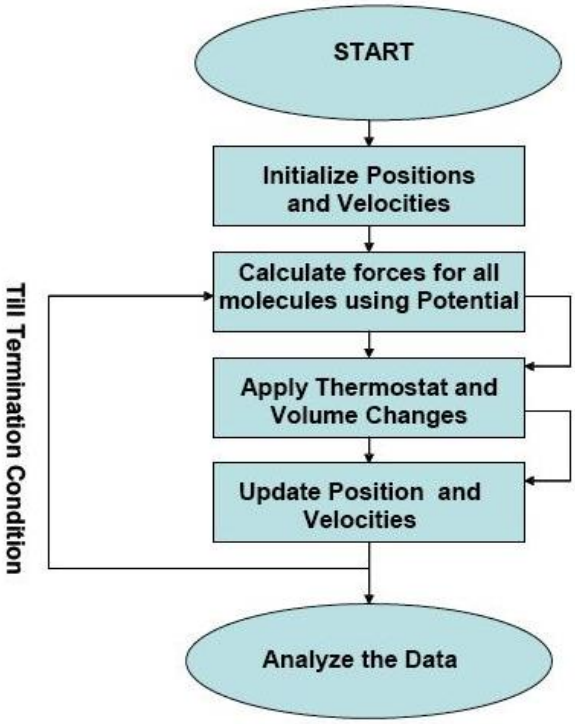


Figure 8: Flowchart showing the general molecular dynamics protocol. (*Image Courtesy: Imaghttp://www-rohan.sdsu.edu/~spydell/md/md.html. 2009. Molecular Dynamics and Visualization)*

The assumption is that the atoms in the molecule interact with each other according to the rules of the employed force field applied.

According to Newton's second law or the equation of motion, **F = ma**, where "F" is the force exerted on the particle, "m" is its mass and "a" is its acceleration.

From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Successive configurations of the system are generated by integrating Newton's Laws of motion which result in a trajectory which reveals the variation in the positions and velocities of the particles in the system vary with time. The average values of properties can be determined from the trajectory. This method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Molecular dynamics simulation hence is an important tool for understanding the physical basis of the structure and function of biological macromolecules. Proteins earlier were considered as relatively rigid structures which now has been replaced by a dynamic model in which the internal motions and resulting conformational changes play an essential role in their function (Karplus M 2002).

These conformational changes can be analysed by MD the stability of the complex molecule in the solvent can also be determined. Various thermodynamic and kinetic parameters such as energy, RMSD, Hydrogen bond analysis, Vander wall interaction electrostatic forces etc. using the Newtonian equation is a classical mechanics to describe the motion of the atom are used to study the macroscopic properties of the molecule

Classical MD simulation uses Lennard Jones potential for the description of the Vander wall forces and Coulomb's law for the electrostatic interaction between the molecules. In MD simulations molecules have assigned velocity which is related to the temperature of the system. The deviation in the bond length of the molecule is described by the potential function or potential energy. Kinetic energy is used to describe the number of conformations can occur in the protein molecule thus helps to study the protein folding.

In MD simulation the total temperature is calculated from the kinetic energy in order to study the fluctuation in the protein molecule (Karplus M 2002). The total energy (sum of kinetic and potential energy) describes the thermodynamic property of the system by analyzing the bonded energy of the system. The energy required to bend the bond from its equilibrium position, the non-bonded interaction referring to Vander Waal terms and electrostatic interactions between not covalently bonded atom pairs can all be known from the energy profile. RMSD (Root Mean Square Deviation) measure the equilibration of the system and the flexibility of the protein molecule referring to the structural stability. RMSD values are used to check the deviation in the simulated protein structure from the original structure. By analyzing the trajectory the changes in the protein structure at the different time period can be visualized. The geometry of the protein molecule can hence be described from which the folding changes in the protein structure can be predicted.

Molecular dynamics allows the study the motion of the individual atoms, something which is not possible in the laboratories experiment. Molecular dynamics is the specialized discipline of molecular modeling and computational simulation which is based on statistical mechanics. Main justification for the MD simulation being that the statistical ensemble average is equal to the time ensemble average. MD simulation has varied applications involving study of the dynamics of large macromolecules, including biological systems such as proteins, nucleic acids (DNA, RNA) and biological membranes. Dynamical events play a key role in the controlling processes which affect the functional properties of biomolecules.

1.11.1 Periodic Boundary Conditions (PBC)

Using the solvent explicitly done by soaking the molecule in a box of solvent molecules is more realistic approach in simulations but it requires additional computational effort. Periodic Boundary Conditions (PBC) are employed to model the bulk solvent so in the case of an infinite PBC, the simulation box is infinitely replicated in all directions so that a lattice is formed. practically, for the computational efficiency most MD simulations evaluate potentials using some cutoff scheme in which each particle interacts with the nearest images of the other n-1 particles (minimum-image convention) (Zhou 2007).

1.11.2 Ewald Summation Techniques

In most MD simulations, the most time consuming are the long-range Coulombic interactions. Hence Ewald summation was introduced in 1921 (Sugita Y. 1999) In order to efficiently sum the long-range interactions between infinite particles and all their infinite periodic images. 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 (Hukushima 1996). A Gaussian charge distribution is commonly used. The sum over the point charges is converted to a sum of the interactions between the charges plus the neutralizing distributions according to the equation below:

$$\mathbf{U}_{\mathbf{Ewald}} = \mathbf{U}^{\mathbf{r}} + \mathbf{U}^{\mathbf{m}} + \mathbf{U}^{\mathbf{0}}$$

This part is the real space sum $\mathbf{U}^{\mathbf{r}}$. A second charge distribution is added to the system which exactly counteracts the first neutralizing distribution. This summation is performed in the reciprocal space and is termed $\mathbf{U}^{\mathbf{m}}$. The dipole term $\mathbf{U}^{\mathbf{0}}$ includes the effect of the total dipole moment of the unit cell, the shape of the macroscopic lattice, and the dielectric constant of the surrounding medium (Zhou 2007).

1.11.3 Particle-Mesh Ewald (PME)

The Particle-Mesh Ewald method (PME) divides the potential energy into Ewald's standard direct and reciprocal sums and uses the conventional Gaussian charge distributions (Okabe 2001). 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. PME does not interpolate but rather evaluates the forces by analytically differentiating the energies, thus reducing memory requirements substantially (Zhou 2007).

1.11.4 Thermostats in MD

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

Thus various types of thermostats, such as Berendsen, Langevin and Nosé-Hoover, to control the particle motions have been proposed. A Berendsen thermostat is a proportional type of thermostat, and corrects deviations of T from the set point T_0 by multiplying the velocities by a factor to control the value of T (Berendsen 1984). Nosé-Hoover thermostat is an integral type of thermostat, and introduces additional degrees of freedom (momentum) into the Hamiltonian of a system (Nosé 1984, Hoover 1985).

The Langevin thermostats follow the Langevin equation of motion instead of Newton's equation of motion (Adelman 1976). In the Langevin equation, a frictional force added to the conservative force is proportional to the velocity, and it adjusts the kinetic energy of the particle so that the temperature matches the set temperature. The force is randomly determined from a Gaussian distribution to add kinetic energy to the particle, and its variance is the function of set temperatures and time steps. The balance of random force with frictional force therefore maintains the system temperature at a set value.

1.11.5 Use of Charges and Solvents

Normally MD simulations are carried out in vacuum conditions by setting the dielectric constant, ε =1. But consideration of solvent effects is required if the molecules to be investigated contain charges and dipoles otherwise the conformations are influenced by strong electrostatic interactions. Force fields try to maximize the attractive electrostatic interactions, which result in energetically strongly preferred but unrealistic low-energy conformations of the molecule. This phenomenon can be pre vented by using the corresponding solvent dielectric constant. For example ε = 80 in the case of water. For realistic biomolecular simulations an accurate description of the solvent environment is essential but may become computationally expensive. The Generalized Born (GB) model is one of the implicit solvent models that treat the solvent as a dielectric continuum (Onufriev 2004).

The electrostatic interactions in the implicit solvent model are rigorously described by the Poisson Equation (PE), and the electrostatic component of the solvation free-energy ΔG_{el}, can be obtained by solving PE numerically. It efficiently describes the electrostatics of molecules in a water environment, by representing the solvent implicitly as a continuum with the dielectric properties of water, and includes the charge screening effects of a salt. There are several versions of the GB model (Onufriev 2004). Implemented in AMBER 12 (Case 2006). In addition to the charges of the interacting particles, the algorithm also takes into account both the dielectric constant of the solvent (approximately 80 for water), and the smoothing function which depends on atomic radii and interatomic distances of the charged particles. Some of the useful features of GB models(Onufriev 2004) include:

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

1.12 Energy-Minimization Procedures in Simulations

Depending on the order of derivative used for locating a minimum on the potential energy surface energy minimization methods can be divided into different classes. Methods that use energy function to identify regions of low energy through a grid search method are

Zero order methods. Several first derivative techniques including the steepest descent method or the conjugate gradient method, are also there which makes use of the gradient of the energy function (Corcho).

1.12.1 Steepest Descent Method

The steepest descent moves directly down the steepest slope on the potential energy surface, with limited changes to the molecular structure. Useful for fast correction of bad starting geometries or for removing bad contacts, and is most effective when the molecular system is far from a minimum. Since the calculation does not readily converge and can oscillate, it is often recommended that a large step-size value be chosen. The Steepest Descent method computes the gradient at its current location, and then travels in the opposite direction of the gradient until it reaches a minimum in this direction. The energy is calculated for the initial geometry and after the movements of one of the atoms in a small increment in one of the directions of the coordinate system. This process is repeated for all the atoms which finally are moved to a new position downhill on the energy surface, due to the fact that every new step is at right angles to the one before it, making numerous smaller steps to proceed down along a narrow valley, and stops when a predetermined threshold condition is fulfilled (Peng 1992). The optimization process is slow near the minimum, and consequently, the steepest descent method is often used for the structures far from the minimum as a first rough and introductory run, followed by a subsequent minimization employing a more advanced algorithm such as the conjugate gradient (Corcho , Toukmaji 1996, Corcho F 2004)

1.12.2 Conjugate Gradient Method

It is a first order minimizer and differs from the steepest descent technique as it uses both the current gradient and the previous search direction to drive the minimization. It uses the minimization history to calculate the search direction, and converges faster than the steepest descent technique. The units of the gradient are kcal mol^{-1} $Å^{-1}$, due to the fact that it is the rate of change (first derivative) of the total energy with respect to atomic positions. The conjugate gradient method produces a set of directions that overcome the oscillatory behaviour of the steepest descents in narrow valleys. Successive directions are not at right angles to each other (Darden 1993) and the conjugate gradient algorithm accumulates the information about the function from one iteration to the next. The gradient is calculated for each minimization step and used as additional information for computing the new direction vector of the minimization procedure. Thus, refinement towards the direction of minimum is done at each successive step. It requires greater computational effort and larger storage than those for the steepest descent, but the conjugate gradient method for larger systems is the method of choice. The greater total computational expense and the longer time per iteration is more than compensated by the more efficient convergence to the minimum achieved in the case of conjugate gradients (Darden 1993, Zhou 2007) . There are several ways in molecular minimization to define In order to judge the quality of the actual geometry of the molecular system in non-gradient minimizers, only the increments in the

energy and the coordinates can be taken into account. However, in all gradient minimizers, the atomic gradients are used for this purpose. The best procedure in this respect is to calculate the root-mean-square gradients of the forces on each atom of a molecule (Darden 1993, Corcho F 2004). The value chosen as a maximum derivative will depend on the objective of the minimization. If a simple relaxation of a strained molecule is desired, a rough convergence criterion like a maximum derivative of 0.1 kcal $mol^{-1}Å^{-1}$ is sufficient, while for other cases 0.001 kcal $mol^{-1}Å^{-1}$ for a final minimum is sufficient (Darden 1993).

Chapter 2

Computational Procedures

2.1 AMBER 12.0

AMBER (an acronym for **Assisted Model Building with Energy Refinement**) is a collection of programs designed to carry out molecular dynamics force field simulations on biomolecules. It is also a package of molecular simulation programs (about 50) which includes source code and demos and also a family of force field that is implemented in the AMBER 12 computer program. (Case 2006, Huang, Zhang *et al*. 2006). It consists of a substructure database, a force field parameter file, and a variety of useful programs. Understanding where to begin in AMBER is primarily a problem of managing the flow of information in this program. While working with the simulation programs within the AMBER such as SANDER, PMEMD one needs to clearly understand what information is needed .The following information is needed by all the programs (Case 2006):

(i) Cartesian coordinates for each atom in the system.
X-ray crystallography, NMR spectroscopy, or model-building programs usually give this information. The program LEaP is also used to model the coordinates for each atom in the system.

(ii) "Topology" refers to the information regarding the connectivity, atom names, atom types, residue names, information about dihedral angles, bond distances and charges. We get this information from the database. It contains information about the standard amino acids as well as the N and C-terminal charged amino acids. Usually the protein database (PDB) files are used to generate coordinate information.

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

(iv) The desired parameters and procedural options are provided to the programs in three separate files. First the coordinates file, second the topology and parameters containing file, called the "topology file" and the third the command or input file.
The primary aim of this study was to study the conformational profiles of medium sized human antimicrobial peptide cathelicidin using force-field simulations within the framework of molecular mechanics. For this purpose, three additional components of the AMBER program extensively used in this study, are described below:

2.1.1 Preparatory programs in AMBER (Case 2006).

LEaP and ANTECHAMBER modules as part of the preparatory programs within the AMBER program were used to build all the peptide sequences, used in this study.
LEaP is used for basic model building and Amber coordinate and parameter/topology input file creation. It includes a molecular editor which allows for building residues and manipulating molecules. It combines the functionality of prep, link, edit, and parm from earlier versions.
ANTECHAMBER is the main program from the Antechamber suite. This program automates the process of developing force field descriptors for most organic molecules. It

starts with structures (usually in PDB format), and generates files that can be read into LEaP for use in molecular modeling. The force field description that is generated is designed to be compatible with the general Amber force fields for proteins and nucleic acids

2.1.2 Simulation programs in AMBER (Case 2006).

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

***Sander*:** SANDER, refers to Simulated Annealing with NMR-Derived Energy Restraints. This module is the central simulation program and provides facilities for energy minimization and molecular dynamics with a wide variety of options and is used for everything except Gibbs free energy calculations. This allows for NMR refinement based on NOE-derived distance restraints, torsion angle restraints, and penalty functions based on chemical shifts and NOESY volumes. Sander relaxes the structure by iteratively moving the atoms down the energy gradient until a sufficiently low average gradient is obtained. As a standard practice, structures are generally minimized before a molecular dynamics simulation. By integrating the Newtonian equations of motion, the molecular dynamics portion generates configurations of the system. MD will sample more configurational space and will allow the structure to cross over small potential energy barriers. Configurations may be saved at regular intervals during the simulation for later analysis, and basic free energy calculations using thermodynamic integration may be performed.

2.1.3 Analysis programs in AMBER (Case 2006).

Analysis modules incorporated in the AMBER program were used to process and analyse the trajectories after MD simulation, in this only PTRAJ module was relevant for the studies undertaken, and is discussed below.

PTRAJ: A general purpose utility module used to analyze MD trajectories, computing a variety of things, like RMS deviation from a reference structure, calculation of bond/angle/dihedral values, average structure, hydrogen bonding analysis, time-correlation functions, diffusional behaviour and so on.

2.2 Conformation Classification (CLASICO)
(https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho)

CLASICO is a computer program designed specifically to group thousands of structures obtained through different computational method, like MD, REMD and SA trajectories into hundreds of patterns based on backbone dihedral angles, easing the subsequent treatment of the information obtained. For a given conformation the procedure consists of assigning a letter to each of the residues of the peptide according to the values of its backbone dihedral angles, following a partition of the space (Srinivasan 1999.). The procedure is based on splitting the Ramachandran plot into different regions as shown in **Figure 9.**

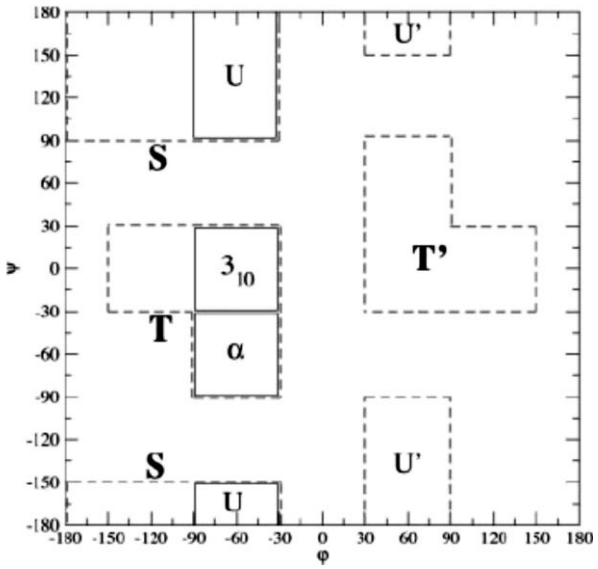


Figure 9: Conformational space which is partitioned into several regions corresponding to the secondary structure motif that each encompasses (3_{10}: 3_{10}-helix; H: α-helix; S: β-strand; T: type I β-turn, T': type I' β-turn; U: type II β-turn; U': type II' β-turn).(*Image courtesy: (Srinivasan 1999.))*

These regions are labelled S, T, T', U, U', 3_{10} and α accordingly, based on the dihedral angles of each residue, sequence of letters have been used to describe the conformations. In a second step, to each of the conformations which has already been coded by a string of letters, using a set of rules successive two- or three-letters **(Figure 9)** are considered to assign conformational motifs. These new strings formed comprising of conformational motifs are called patterns. This protocol enables identification of the different conformational patterns attained by the peptide, and even as comparison of the differences of the conformational space which has been sampled using different computational methods. Secondary structure patterns thus obtained are then numbered and plotted along the trajectory. The frequency of these different patterns can be graphically assessed by looking at the distribution of points. Following this procedure, the CLASICO algorithm (https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho), can translate the thousands of snapshots stored during the MD, REMD or SA trajectories into 11 different possible motifs shown in **Table 1** (H, 3_{10}, S, I1, I2, i1, i2, II1, II2, ii1, and ii2), and the classification of β-turns on the basis of dihedral angles (ϕ and ψ) is depicted in **Table 2**.

Table 1: Conditions for secondary structure definition of three consecutive residues, where (3_{10}: 3_{10}-helix; H: α-helix; S: β-strand; T: type I β-turn, T': type I' β-turn; U: type II β-turn; U': type II' β-turn and j: amino acid residue of peptide sequence).

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

Table 2: Definition of β-turns classified on the basis of dihedral angles

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

A flowchart showing the protocol and the commands used in CLASICO program (https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho) is depicted in Figure. The dihedral angles obtained from different trajectories are used as input in CLASICO and the secondary structure motifs of the peptides are generated

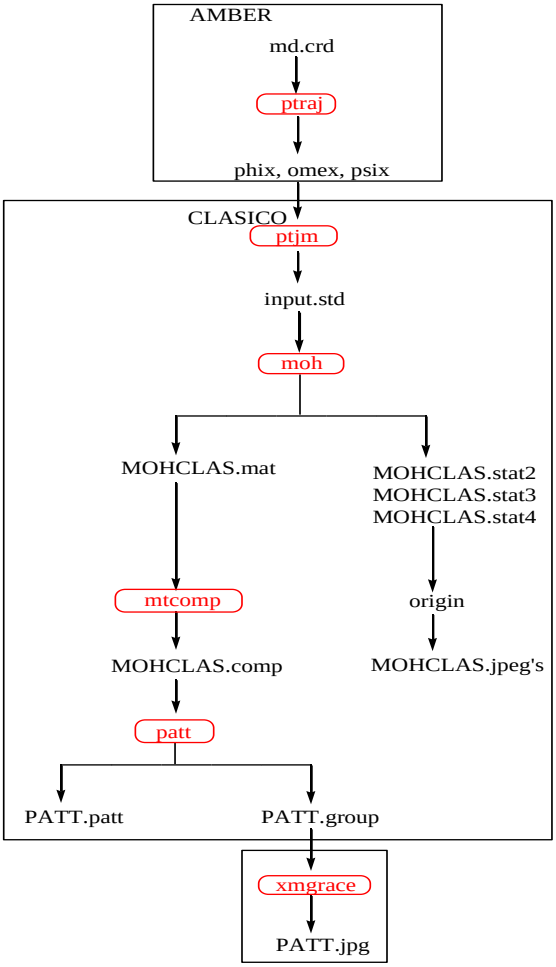


Figure 10: Flow chart for the protocol in the CLASICO program (commands are shown in red).

Programs in CLASICO **(https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho)**

ptjm

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

moh

This program classifies structures based on structural motifs. 3 consecutive residues which accomplish whatever helical condition (H, 3_{10} or PI) are considered an helix. The specific motif assignation is done in a local way depending on the condition accomplished. When different assignations can be done, the order of priority is H > 3_{10} > PI. The input.std file will be read by moh executable to create the MOHCLAS.mat, where the pattern information is encoded and *MOHCLAS.stat2*, with statistics about the presence of each motif per residue.

mtcomp

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

patt

The program patt classifies the structures in patterns based on the conformational motifs present. MOHCLAS.comp and PATT.sum are used as input files.

PATT.patt is the pattern file (identical to MOHCLAS.comp but only with non-repeated patterns) PATT.group is the file for recording the appearance of new patterns- pattern numbers are assigned to the structure.

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

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

2.3 *GROMACS*

GROMACS (**GROningen MAchine for Chemical Simulations**) is an engine primarily designed for molecular dynamics simulations and energy minimization of proteins, lipids and nucleic acids .It simulate the Newtonian equations of motion for systems with hundreds to millions of particles. GROMACS is extremely fast at calculating the nonbonded interactions due to its algorithmic and processor-specific optimization, it runs 3-10 times faster than many simulation programs (Gnanakaran 2003).

Gromacs uses the concept of periodic boundary condition and groups:

Periodic boundary condition is a classical way used in Gromacs to minimize the edge effect in system. The atoms of the system to be simulated will be placed in a space-filling box, which is surrounded by translated copies of the atom shown in **Figure 11.**

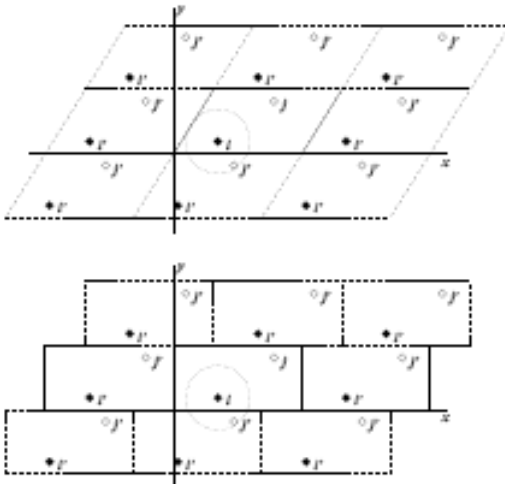


Figure 11: Periodic boundary condition in two dimensions. (*Image courtesy: (Lindahl))*

In Gromacs there are several shapes for space filling unit boxes like triclinic, cubic, and octahedron. **Group** is the second concept used by Gromacs**.** The Gromacs MD and analysis

programs use user-defined groups of atoms to show an action. Each group contain only a maximum number of 256 atoms, where each atom can have six different groups.

2.3.1 Flowchart Of Gromacs (manual 2010)

For molecular dynamics simulation of a protein Gromacs need several steps to set up a file input in the simulation. These steps can be seen in flowchart below:

1. Conversion of the PDB file
2. Generate box
3. Solvate protein
4. Energy minimization
5. Molecular dynamics simulation
6. Analysis

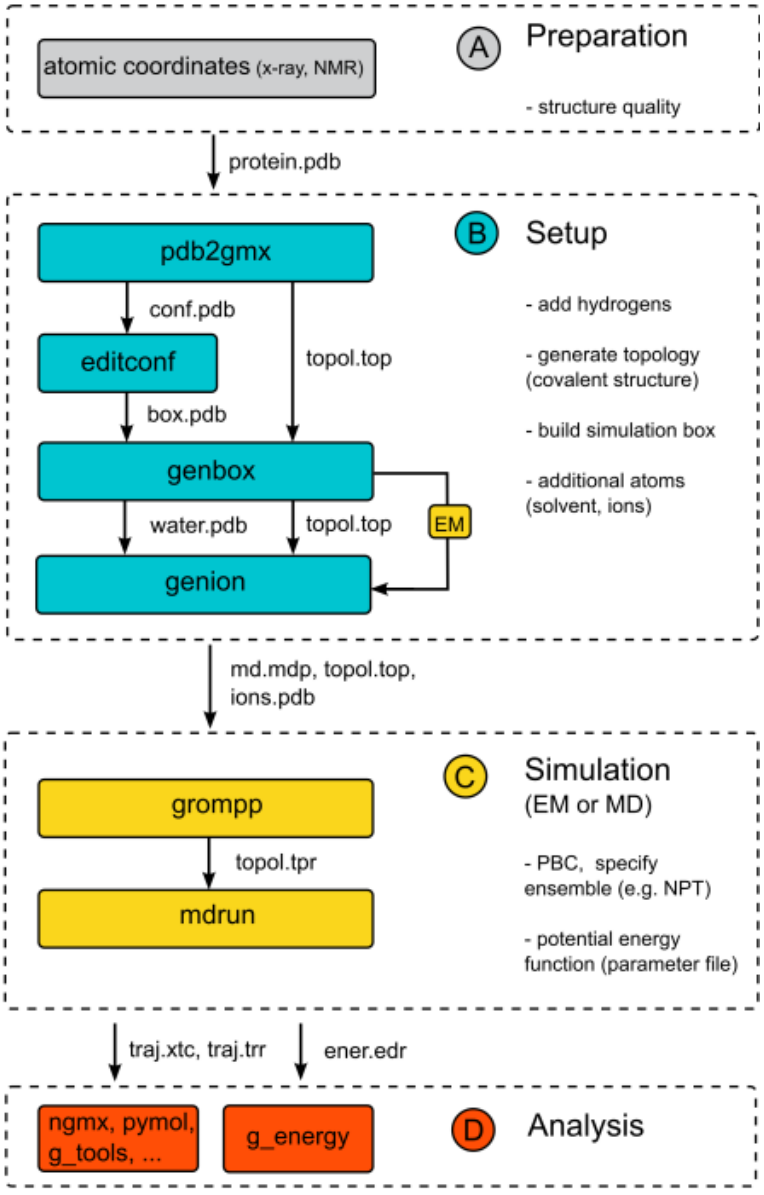


Figure 12 : Flowchart Gromacs.(*Image courtesy: (manual 2010))*

Gromacs Programs (manual 2010)

3.3.1. Pdb2gmx

Pdb2gmx is a program used to convert pdb file. Pdb2gmx can do some things such as reading pdb file, adding hydrogen to molecule structure, and generate coordinate file a topology file.

Editconf

To define a water box that will be used for simulation editconf program will be used. It not only defines the model, but also set the relative distance between edge of box and molecules. For this 3 types of boxes will be used:

- A Triclinic, shaped box
- A Cubic box, with all four side equal
- An Octahedron, which is a combination of octahedron and dodecahedron.

Grompp (GROMACS pre-processor)

Grompp is a pre-processor program. It process the coordinate file and topology (which describes the molecules) to generate an atomic-level input (*.tpr). The *.tpr file contains all the parameters for all of the atoms in the system.

Grompp have ability to read a molecular topology file, check validity of file, extract the molecular information into the atomic information and expand the topology, recognize and read the parameter file (*.tpr), coordinate file (*.gro) and the topology file (*.top).

Genbox

Genbox is another program that is used for the generation of solvent box , salvation of protein and addition of extra molecules on random position .It can also remove atom if distance between solvent and solute is less then sum of vanderwalls radii of each atom.

Genion

Genion program is used to neutralize the system by generating monoatomic ions on energetically favorable positions.

Mdrun

Mdrun is the main programming engine in GROMACS. It not only performs molecular dynamics simulation, but can also perform Brownian dynamics, Langevin dynamics, and energy minimization.
It uses *. tpr as input file and produces four output files ,the trajectory file (*.trr),which contains coordinates, velocities and optionally forces. The structure file (*.gro) that contains the coordinates and velocities of the last step. The energy file (*.edr) contains energies, the temperature, pressure, etc, and compressed trajectory file (*.xtc).

2.3.2 Analysis program in GROMACS

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

Grace stands for Graphing, Advanced Computation and Exploration of data. Grace is also known as xmgr/xmgrace. It is aplotting software used to plot 2D graphs (Grace development team, G. development, 2008.Grace user guide:"What is grace?")

Grace reads the .xvg file generated from various Gromacs programs. These are the special file format readable by grace to plot the 2D graph.

Chapter 3
Methodology

To investigate the folding pattern of cathelicidin in implicit water 100 ns long MD simulations were run starting from extended structure of cathelicidin (built using tLEaP) and the NMR structure available in PDB (PDB ID: 2K6O). AMBER 12 suite installed on Ubuntu, an operating system based on the Debian linux distribution was used to run the simulations. Different force fields AMBER ff96 and AMBER ff99 were used for the computation of energy. AMBER ff96 force field was chosen as it has been reported that ff96 set accurately reproduces the results of explicit solvent simulations when used with implicit solvent systems (Kollman 1997) . AMBER ff99 is an advancement of the same (Calero, Lago *et al*. 2004, Gnanakaran 2003) and hence as used for the sake of a comparison.

Implicit solvent simulations using GB approximation treat the solvent as dielectric continuum. Implicit solvent systems are computationally less extensive and less time consuming hence they enable the peptide to achieve equilibrium faster. Onufriev, Bashford and Case (OBC) implementation of the Generalized Born (GB) approximation was chosen to treat the solvent since it provides high computations efficiency and allows for the analytical evaluation of forces (Still 1990, Chen 2008)

3.1 MD on Linear and NMR Structure of LL-37 using GB Method and Langevin Thermostat Algorithm.

The NMR structure PDB ID: 26KO and extended structure of cathelicidin LL-37 was processed using tLEaP module of AMBER 12 and used as the starting structure.
For the linear structure amino acid sequence of LL-37 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES) was used as input where the N terminal was acetlylated and C terminal N methylated. For MD on NMR structure the PDB structure was used as input in tLEaP. Topology and coordinate files generated from tLEaP were used as input for minimization. Three different force fields AMBER ff96, AMBER ff99 and AMBER ff99SB were selected for the computation of energy. AMBER ff99SB is an advanced version with improvements over older force fields and has been recommended in literature to be used in case of peptide simulations (Case 2006, Wickstrom 2009). and the structure was then energetically minimized using SANDER module of AMBER 12 suite with a total of 1000 steps: 100 steps of steepest gradient, followed by a subsequent minimization using 900 steps of conjugate gradient algorithm until a convergence of the gradient norm was lower than 0.005 kcal mol^{-1} $Å^{-1.}$ In order to mimic the physiological conditions 0.2M salt concentration was used. The minimized structures were then heated for 100 ps in 5000 steps by gradually increasing the temperature from 0 to 300 K. Shake algorithm was used to constraint the bonds involving hydrogen atoms with a time step (dt) of 2fs. Two sets of 100 ns MD simulations using force fields AMBER ff96 and AMBER ff99 were performed at 300K using Langevin thermostat and Onufriev, Bashford and Case (OBC) implementation of the Generalized Born (GB) approximation. Dielectric constant in the peptide was set to 1 while an external dielectric constant of 80 corresponding to water was employed and a 12Å cutoff with coupled thermostat of 1.0 ps (tautp) was used. All

calculations in the present work have been carried out using AMBER 12 suite of programs(Case 2006). The analysis of the trajectories obtained, was done using the PTRAJ module in AMBER. CLASICO program was used to analyze the formation and destruction of secondary structures to be monitored during the folding process, as well as in the characterization of the group of structures that represents the folded molecule. The thermodynamic profiles of the MD trajectories were plotted using XMGRACE. Visual Molecular Dynamics (VMD) program was used to view the trajectory and Pymol used to view the structures obtained during MD.

All the input files used for carrying out this simulation run have been provided in Appendix1

AMBER MD Simulation Flow Chart

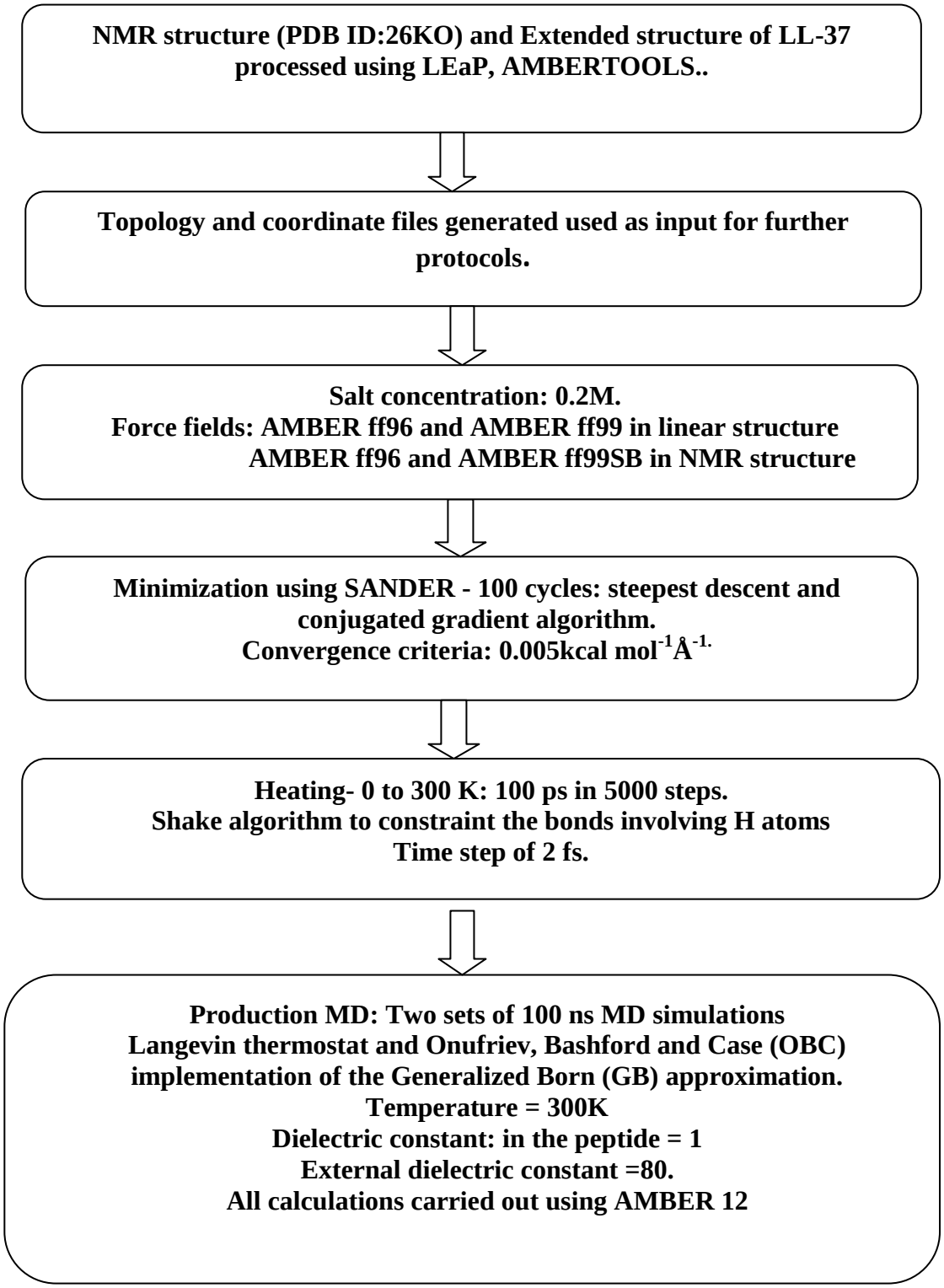


Figure 13: Amber MD simulation Flowchart.

3.2 MD Simulations using the Explicit Solvent Method and Berendsen Thermostat

To study the overall conformation of Human LL-37 Cathelicidin extensive MD simulations was performed on a fully hydrated model of structures. All simulations were performed using the GROMACS (Groningen Machine for Chemical Simulation) package (manual 2010). Molecular dynamics simulations use Newton's equations of motion to calculate trajectories of particles, starting from a defined configuration. The total force acting on each particle in the system is calculated from the interactions with other particles, as described by the force field. The force together with the prior position and velocity, determines what the new position will be after a small time step.

3.2.1. Pre -processing of the structure.

The crystal structure of the Human LL-37 Cathelicidin (amino acid residues 1-37), available in Protein data bank (PDB ID: 2K60) which is a repository for three dimensional structures of proteins and contains on a total of over 50000 structures was used as the starting point for all the molecular dynamics studies.

3.2.1(a) Solvation

Water, usually as a solvent has a fundamental influence on the structure, dynamics and thermodynamics of biological molecules, both locally and globally. One of the most important effects of the solvent is the screening of electrostatic interactions. Solvation is generally done to create an explicitly solvated system, where water molecules and optionally counterions are added to protein. All simulations were performed in explicit water using the GROMACS (Groningen Machine for Chemical Simulation) package (manual 2010) in conjunction with the GROMOS96 43a1 force field for condensed phases. For the simulation to take place in a reasonable amount of time the size of the system was restricted by placing the molecule in the periodic boundary conditions, meaning that a single unit cell is defined, which can be stacked in a space filling way to avoid the edge effects due to the walls of the simulation volume, so a octahedron unit cell is used. The minimum distance between any atom of the protein's periphery and the box wall was 1.0 nm. Thus water molecules were added to solvate the box. The Extended Simple Point Charge (SPC/E) model (H. J. C. Berendsen 1987) was used to represent the water. spc216.gro, which is a generic equilibrated 3-point solvent model, is used as the solvent configuration for SPC/E. For the simulation to take place in a reasonable amount of time the size of the system was restricted by placing the molecule in the octahedron. The minimum distance between any atom of the protein's periphery and the box wall was 1.0 nm. Thus water molecules were added to solvate the box (Swaroop Chatterjee 2008). Then the 6Cl counterions were added to neutralize the system i.e. the solvent molecules were replaced by monoatomic Cl ions at the position of the first atoms with the most favourable electrostatic potential or at random because the cathelicidin structure contain a net charge of +6e (based on its amino acid composition). The molecular dynamics simulations were performed at constant temperature and pressure.

3.2.2 Energy Minimization

Energy Minimization is considered to be an important step in Molecular Dynamics simulation. It is a process through which structure is relaxed and it also ensures that the system has no steric clashes or inappropriate geometry. The added hydrogens and broken hydrogen bond network in water leads to quite large forces and structure distortion, hence to remove these forces it is essential to perform minimization. To minimize the effects of truncating the electrostatic interactions beyond the 1.4 nm long-range cutoff, a reaction field correction was applied using a relative dielectric constant of 80.4. Explicit hydrogen atoms in the force field were replaced by dummy atoms to remove high-frequency degrees of freedom, the positions of which were constructed each step from the coordinates of the heavy atoms to which they are attached. A steepest descents minimization for 50000 steps i.e. 100ps is performed to remove any bad Vander Waals contacts

3.2.3. Protein restraint dynamics

To avoid any unnecessary distortion of the protein during the molecular dynamics simulation, equilibration was performed where all heavy atoms are restrained while the water is allowed to relax around the structure. This was done to soak the water molecules into the macromolecule.

Equilibration is done in two phases:

- The first phase is done under an *NVT* ensemble (constant Number of particles, Volume, and Temperature). This ensemble is also referred to as "isothermal-isochoric" or "canonical.")(Berendsen 1984), The timeframe for such a procedure is dependent upon the contents of the system, but in *NVT*, the temperature of the system should reach a plateau at the desired value. After energy minimization (EM) using a steepest descent algorithm, a 500 ps time period, with 250000 at a time step of 0.002ps. at 300 K and using brendsen thermostat equilibration were performed to gently relax the system(manual 2010).
- To stabilize the pressure (and thus also the density) of the system equilibration is conducted under an *NPT* ensemble (constant Number of particles, Pressure, and Temperature). This ensemble is also called the "isothermal-isobaric". After NVT equilibration, a 500 ps time period, with 250000 at a time step of 0.002ps. at 300 K and using parrinello-rahman (Lindahl , Godzi, Tolstova *et al.* 2010)and isotropic coupling with time constant tau_p 1.0 ps. (which is the period of pressure fluctuation at equilibrium) equilibration were performed.

3.2.4 Molecular Dynamics

To see the conformational changes in the peptide over a considerable time, the peptide was analyzed by a molecular dynamics simulation, for 100 ns, following the minimization and the restraint phases. Unrestrained molecular dynamics (MD) were then performed at 300 K for 100 ns of simulation to assess the stability of the structures. During the simulations the temperature and the pressure were maintained at 300 K and 1 bar by coupling to an

external heat and an isotropic pressure bath. The relaxation times were 0.1 psec and 0.5 psec, respectively. The coordinates were saved at every 2 ps and analyzed.

All the input files used for carrying out this simulation run have been provided in appendix1

Chapter 4

Results

*4.1 Analysis of MD Simulation on Linear Structure of LL-37 using GB Method and Langevin Thermostat Algorith***m.**

4.1.1 Thermodynamic Profiles

The quality of the MD trajectory was measured by plotting the thermodynamic parameters such as temperature and the total energy. The plots testify that the simulation is relatively stable over the entire MD trajectory, in both AMBERff96 and AMBERff99 force fields

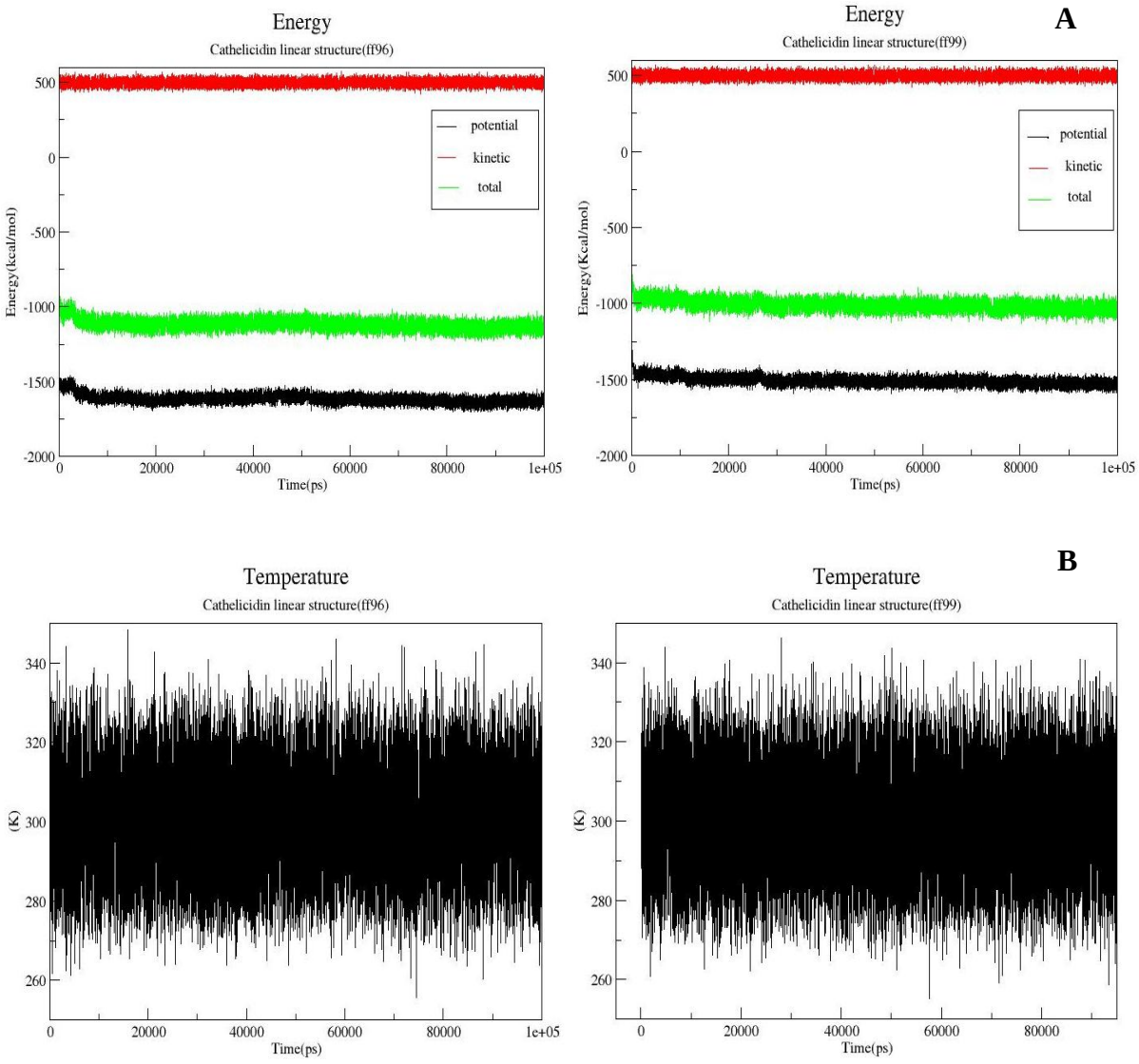


Figure 14: Themodynamic profile plots of MD trajectory of linear LL-37 under force field AMBERff96 and AMBER ff 99

The energy graphs demonstrates that there is a small decrease in total and potential energy ~100 Kcal mol–1,(**Figure 14 A)** after ~5 ns in case of MD trajectory of force field ff96 and then stabilize implying that the system has folded up to a state more stable than the starting linear structure. Whereas in force field ff99 there are not many fluctuations it gives

a stable graph throughout the trajectory. In both the cases kinetic energy is stabilized from the starting of the trajectory as expected. This is because temperature is directly proportional to the kinetic energy.
The temperature plots **(Figure 14 B)** reveal that after equilibration for 100 ps at 300K the temperature is constant at 300K throughout the trajectory in case of both force field ff96 and ff99.

4.1.2 RMSD Analysis

To have a better insight into the structural features of the cathelicidin and know about the structural stability of the conformers obtained, RMSD analysis was done with the initial structure considered as a reference structure. **Figure 15** describes the structural changes of the peptide along the trajectory during the simulation where the root-mean-square deviation (rmsd) with respect to the main chain atoms of the starting structure using PTRAJ module of AMBER12 (Case 2006).

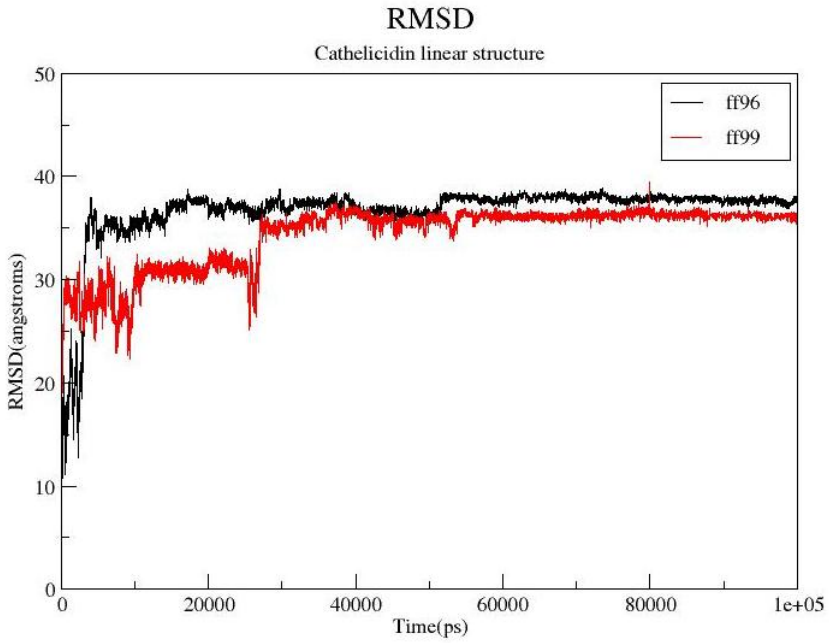


Figure 15: Root mean square deviations (backbone-backbone) of LL-37 from the starting structure in case of MD ^{implicit} force field ff99 in red and MD ^{implicit} force field ff96 in black obtained at 300K.

From the plot of RMSD vs Time as evident there was a significantly higher fluctuation from the start of the simulation until 30 ns in force field ff99 **(Figure 15 shown in red)** and upto 10 ns in case of force field ff96 **(Figure 15 shown in black)** clearly indicating larger structural deviations of the conformations from the reference structure in force field ff96.
A plateau about RMSD~38 Å in case of ff96 and around RMSD 35 Å formed from 50 ns to 100 ns and from 35 ns to 100 ns segment of simulation trajectory shows comparatively stable structures, reason for this probably being the sufficient folding during MD sampling. This data also suggest that the peptide have folded to a considerable extent after 50ns and

35ns of the trajectory respectively, and does not show any sharp changes in the conformations obtained. The RMSD plot also shows a relatively stable trajectory form 20 ns to 40 ns in case of force field ff96.

4.1.3 Pattern Profiles

To get better insight into this the sampling efficiency of the MD simulations force fields ff96 and ff99 was monitored for the different patterns of LL-37 obtained during the progress of each trajectory. Basically patterns represent the structures classified on the basis of the type of conformational motifs present (Corcho F 2004). In order to achieve this, pattern profiles were computed in each trajectory for every snapshot using the CLASICO program.

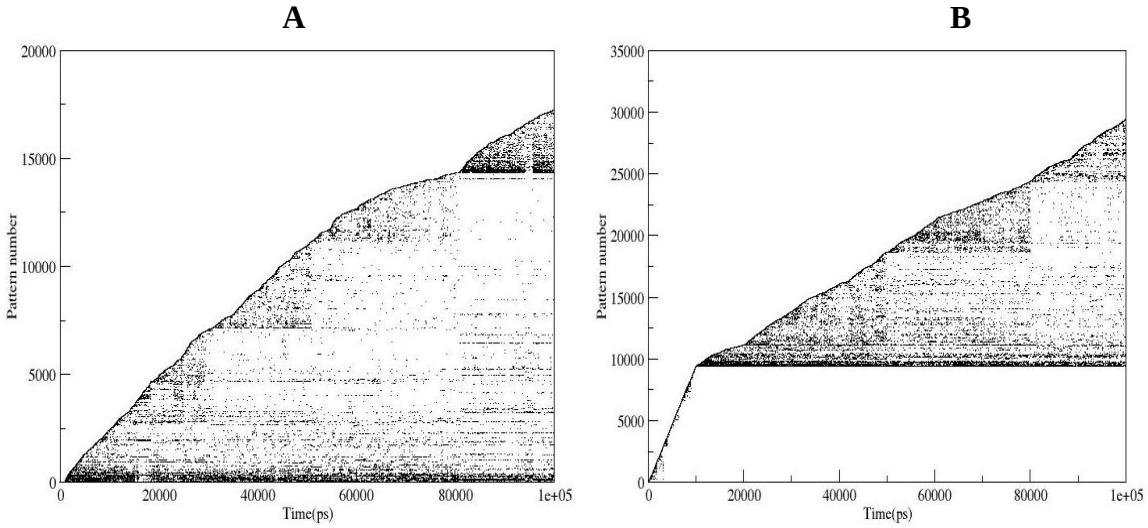


Figure 16: Evaluation of new patterns along the trajectories for LL-37 in **(A)** MD ^{implicit} force field ff96 **(B)** MD^{implicit} force field ff999SB

As depicted in **Figure 16** after 100,000 snapshots of MD implicit ff96 **(Figure 16A)** and ff99 trajectory **(Figure 16B**) 17000 and 30000 new patterns respectively were identified.
The percent efficiency of these trajectories for generating new patterns in case of MD ff96 was 17% and for ff99 was 30%. On observing **Figure 16 (B)** closely it can be inferred that upto 10ns, we obtained new patterns but post this for the entire length of the trajectory similar patterns were generated in a regular fashion. Whereas in the case of MD ff96 **(Figure 16 A)** new patterns have been identified from the very start of the trajectory going on for the entire length upto 100 ns. Despite this pattern profiles obtained from ff96 showed maximum similar patterns in majority of the conformations, clearly suggesting stability of the simulation in this part of the trajectory which is consistent with the RMSD results obtained.The dark regions in the conformational space show restrictive nature of peptide conformations to explore new patterns at certain intervals.

4.1.4 Secondary Structural Motifs

In order to characterize the structural features analysis of all trajectories was done using the CLASICO software. This software enables the monitoring of formation/destruction of secondary structures during the folding process and also the characterization of the group of structures representing the folded molecule. The protocol followed had been previously reported in literature (https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho) and it also permits identification of different conformational patterns attained and compare differences of the conformational space sampled.

To achieve this a qualitative analysis of the secondary motifs was done using CLASICO software that allows the identification of secondary motifs from each snapshot (LaFargaCPL: (https://lafarga.cpl.upc.edu/projects/clusterit. and (b) Corcho)) the sequential procedure followed by CLASSICO program is that it translates each snapshot into a string of letters following the sequence firstly it computes backbone dihedral angles for each residue and assigns it a letter in accordance with the Zimmerman partition of the Ramachandran map (Zimmerman 1977). Each string is then analysed using a three letter string following a set of **(Figure 9)** in order to assign the corresponding secondary motif. Histograms of the secondary motifs per residue for each of the simulations done on cathelicidin LL-37 are shown in **Figure 17 (A) and (B)**.

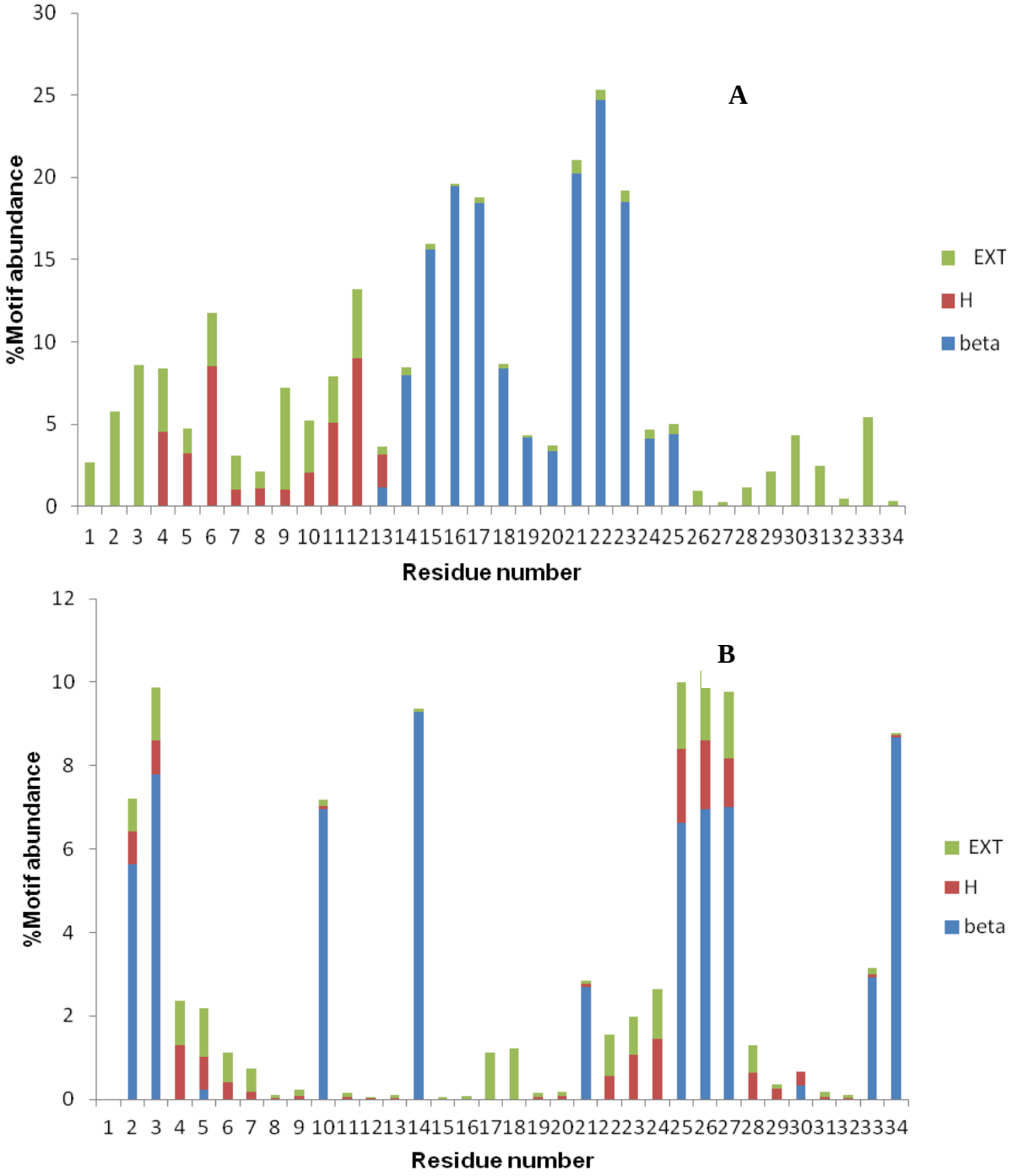
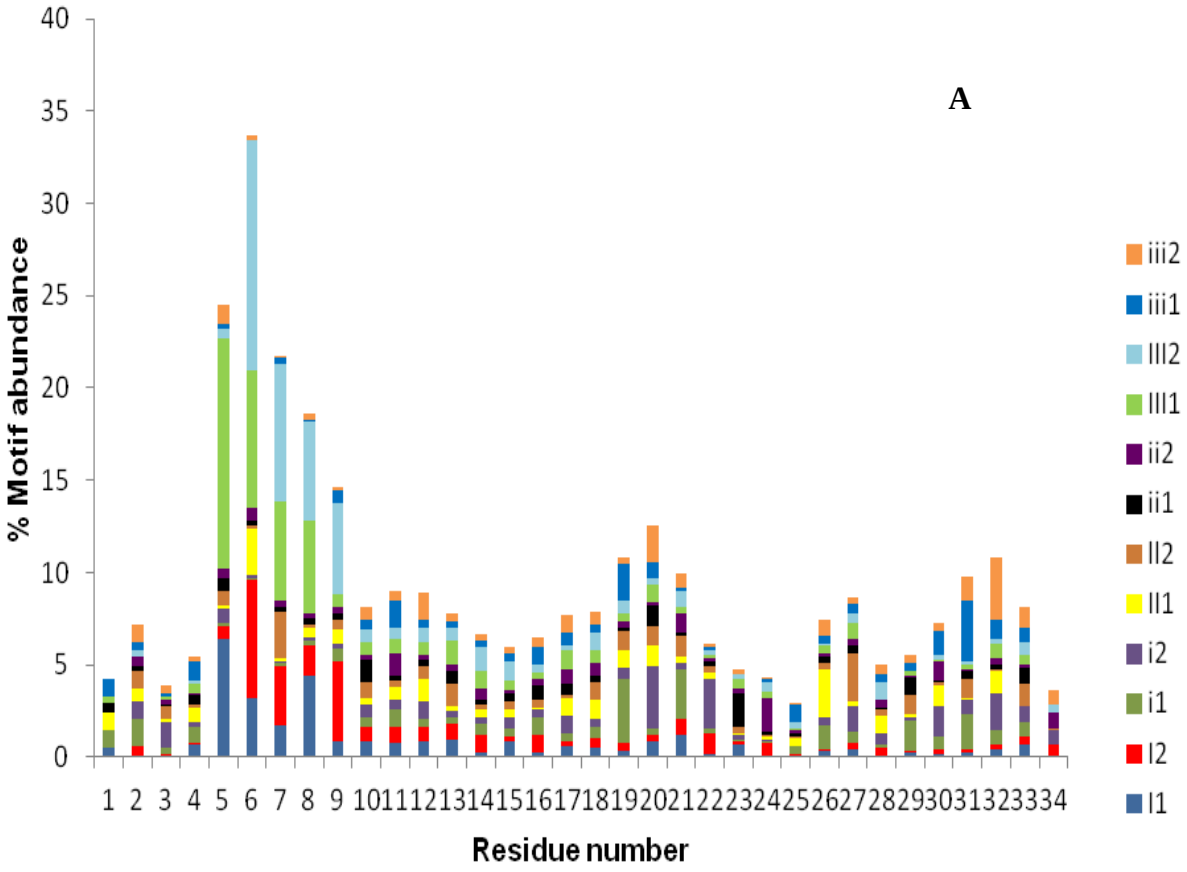


Figure 17: Motif abundance from MD trajectory linear LL-37 under **(A)** force field ff96 **(B)** force field ff99.

The figures show about 20-25% for the MD calculations using AMBER ff96 and about 10% for the MD calculation using AMBER ff99. **Figure17 (A)** illustrates the secondary structure motifs obtained during the MD simulation by using force field ff96, shows predominantly a β region between residues, with a stronger propensity Gly^{14} to Ile^{24}, a stronger propensity being between the residue Lys^{15} to Lys^{18} and Val^{21} to Arg^{23} and residues Gly^{14}, Arg^{19} and Ile^{20} shows lower propensity β region. To some extent around 4%-5% we get 3_{10}-helix between residues Asp^{4} to Phe^{6} and between Glu^{11} to Ile^{13} but it lasted for a very short duration. The N and C-terminal region of the peptide mostly remain

extended and flexible not showing any secondary structural feature. **Figure 17 (B)** shows the secondary structural motifs obtained during the MD simulation by using force field ff99, shows a predominantly helical region formed by consecutive β turns mostly around the C terminal region between residue Val^{21} and Phe^{27} with a higher propensity being between Ile^{24} and Phe^{27}. From residue Arg^{7} to Glu^{11} alpha helical region was sampled. Rest of the N terminal was extended and flexible.

The β-turn secondary motifs attained by residues of the LL-37 peptide were further classified into different types using the two-residue window of the CLASICO program. To determine the conformational motifs attained by the LL-37 peptide, the structures obtained from the MD trajectories at 300 K were analyzed using the CLASTERIT algorithm and. The statistics of all the motifs found in the peptide for each residue are shown in **Figure 18 (A)** and **(B)**.



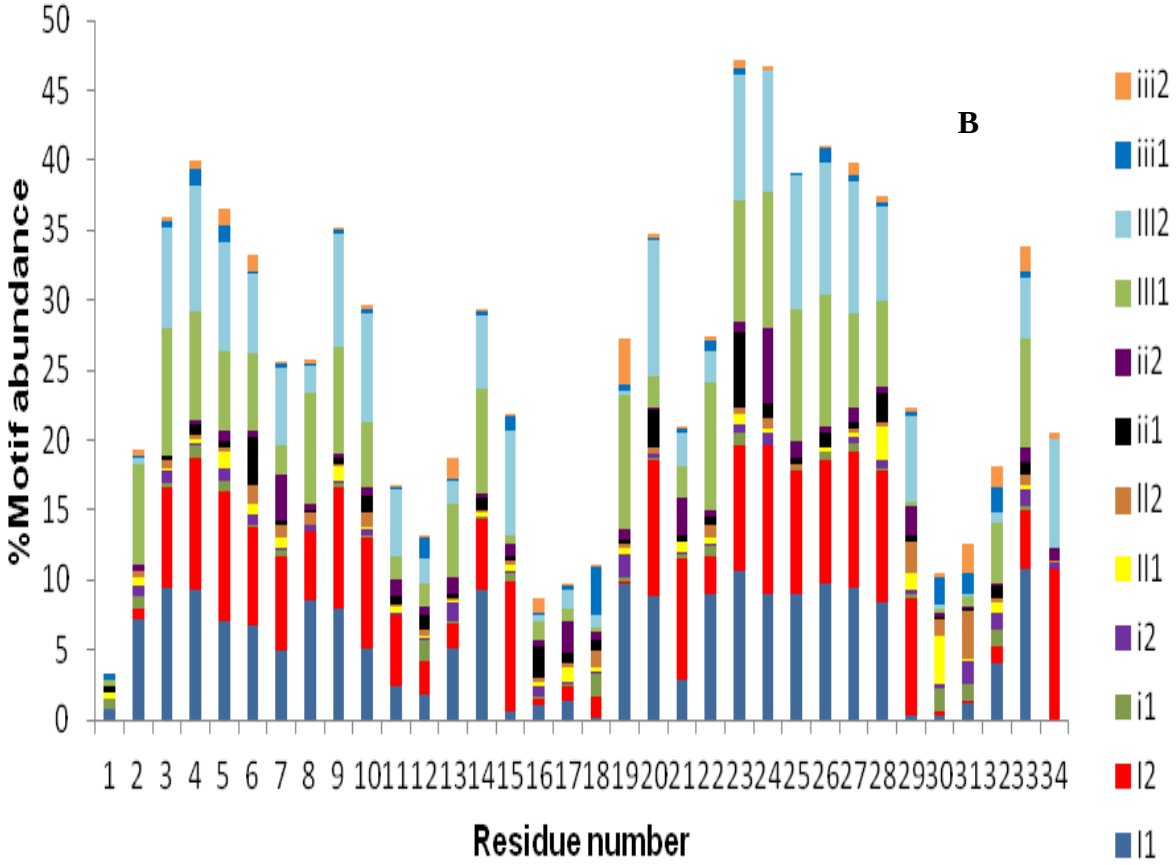


Figure 18 : Classification of β turns obtained in the MD trajectory of linear LL-37 **(A)** force field ff96 **(B)** force field ff99.

The secondary structure motifs appear almost from the beginning of the trajectory in case of ff96 with a the predominance of a β-turn type III between residues Asp4 and Ser^{9} with a lower propensity of type I in general . To some extent, β-turn type I was also observed between residues Phe^{5} and Lys^{8} in some of the conformations sampled.

Figure 18 (B) shows the profile obtained in the MD ff99 trajectory where the peptide exhibits predominantly β-turns type III with higher propensity between residues Leu^{2} and Lys^{10} and a lower propensity between residues Ile^{20} and Arg^{29}. To some extent, β-turn type I (mirror conformation of β-turn type I) was also found between in the overall trajectory. The percentage order of β-turns adopted by the MD ff96 was 15 to 35% from residue 5 to 9 and close to 15% from residue 10 to 34. In case of MD ff99 the percentage of β-turns sampled were about 30 to 40% from residue 3 to 13 and close to 50% between residues 22 to 29.

4.1.5 Hydrogen Bond Analysis

Table 3 summarizes the results obtained by intramolecular hydrogen bond analysis performed using PTRAJ module of AMBER 12 program.

Table 3: Secondary structures observed due to mainchain - mainchain hydrogen bond interactions and their percentages in different trajectories for MDff99. Secondary structures, α-helix, π-helix, β-turn, reverse turn and loop, are assigned in terms of hydrogen bond interactions between i to i+4, i to i+5, i to i+2, i+n (n >2) to i and i to i+n (n >3) residues, respectively in the peptide.

No.	Definition	2*Structure	MD^{ff99}(%)
1	(Lys^{8})CO…NH(Lys^{12})	α–helical	2.98
2	(Lys^{12})CO…NH(Glu^{16})	α –helical	0.12
3	(Lys^{15})CO…NH(Arg^{19})	α –helical	1.55
4	(Asn^{23})CO…NH(Phe^{27})	α –helical	0.35
5	(Asp^{26})CO…NH(Asn^{30})	α –helical	1.66
6	(Arg^{29})CO…NH(Pro^{33})	α –helical	6.39
7	(Phe^{27})CO…NH(Leu^{31})	α –helical	1.07
8	(Lys^{8})CO…NH(Glu^{11})	β–turn	4.08
9	(Arg^{7})CO…NH(Lys^{10})	β–turn	3.74
10	(Asp^{4})CO…NH(Arg^{7})	β–turn	4.67
11	(Lys^{12})CO…NH(Lys^{15})	β–turn	1.40
12	(Lys^{15})CO…NH(Lys^{18})	β–turn	3.20
13	(Ile^{20})CO…NH(Asn^{23})	β–turn	-
14	(Asn^{23})CO…NH(Asp^{26})	β–turn	0.56
15	(Asp^{26})CO…NH(Arg^{29})	β–turn	2.53
16	(Phe^{27})CO…NH(Asn^{30})	β–turn	4.17
17	(Arg^{29})CO…NH(Val^{32})	β–turn	5.38
18	(Phe^{5})CO…NH(Lys^{8})	β–turn	3.52
19	(Glu^{34})CO…NH(Ser^{37})	β–turn	3.29
20	(Leu^{31})CO…NH(Phe^{5})	loop	-

Table 3 shows the interaction involving the mainchain hydrogen bonds. The results obtained revealed that conformations in ff99 adopted β-turn predominantly in between the starting residues (Asp^{4}-Glu^{11}), the central residues (Lys^{12}-Lys^{18}) and the C-terminal residues (Asn^{23}–Asn^{30} and Arg^{29}–Val^{32}) showing higher percentage in residues Arg^{29} – Val^{32} , Phe^{27} – Asn^{30}. To some extent the α-helical regions were also seen between residues (Lys^{8} –Lys^{12}, Lys^{12}-Arg^{19}, Asn^{23}-Pro^{33}) with a higher percentage in residues Arg^{29}-Pro^{33}.

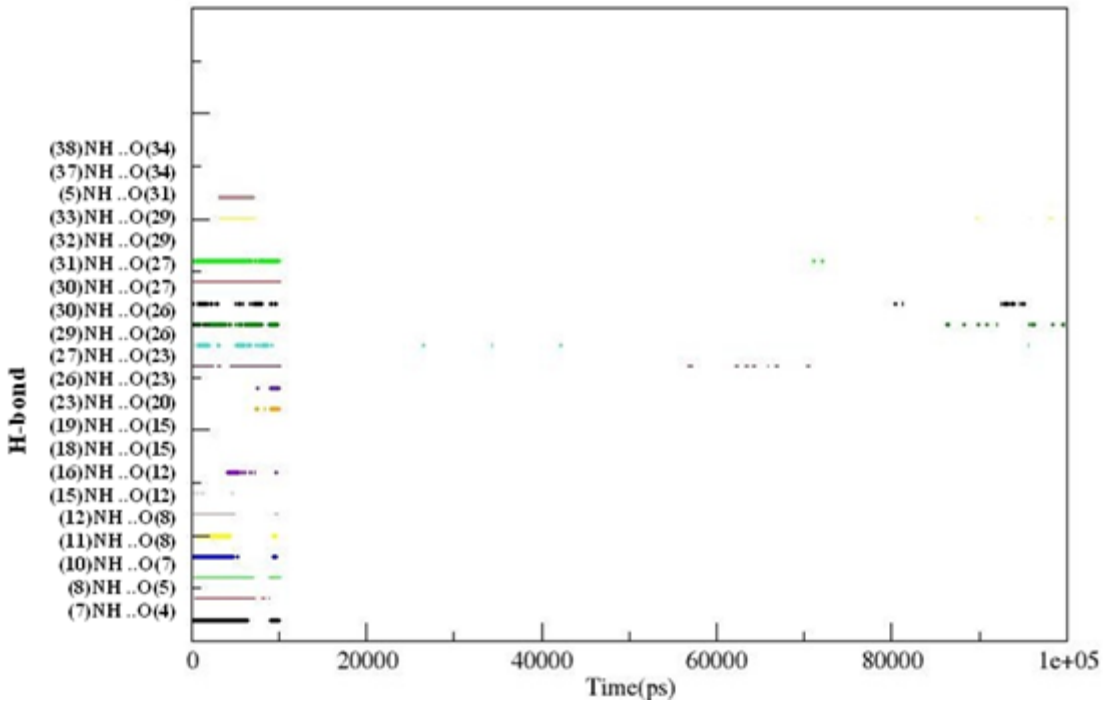


Figure 19: Progress of hydrogen bonds **(Table 3)** monitored between important residues for cathelicidin in **MD ff99** trajectories. Secondary structures, α-helix, π-helix, γ-turn, β-turn, reverse turn and loop, are assigned in terms of hydrogen bond interactions between i to i+4, i to i+5, i to i+2, i to i+3, i+n (n >2) to i and i to i+n (n >3) residues, respectively in the peptide.

Figure 19 which is the diagrammatic representation of the permanence of the difference in hydrogen bonding patterns along the dynamics for the trajectory shows that during the initial 10 ns of the trajectory a hydrogen bond formed between different residues corresponding to helicity, β-turn conformation in some of the structures. The helical region between residues 23 to 27, 26 to 30 can be observed during the time period 55 ns-70 ns, beyond 80ns, respectively. Similarly small stretch of β-turn between residues 27 to 30 were visible around 95ns but to a small extent of time interval. Overall these results suggest that the peptide has a high propensity to remain in the β-turn conformation. This shows that consecutive β-turns are involved in the formation of helix.

Table 4: Secondary structures observed due to mainchain-mainchain hydrogen bond interactions and their percentages in trajectory of linear LL-37 under ff96.

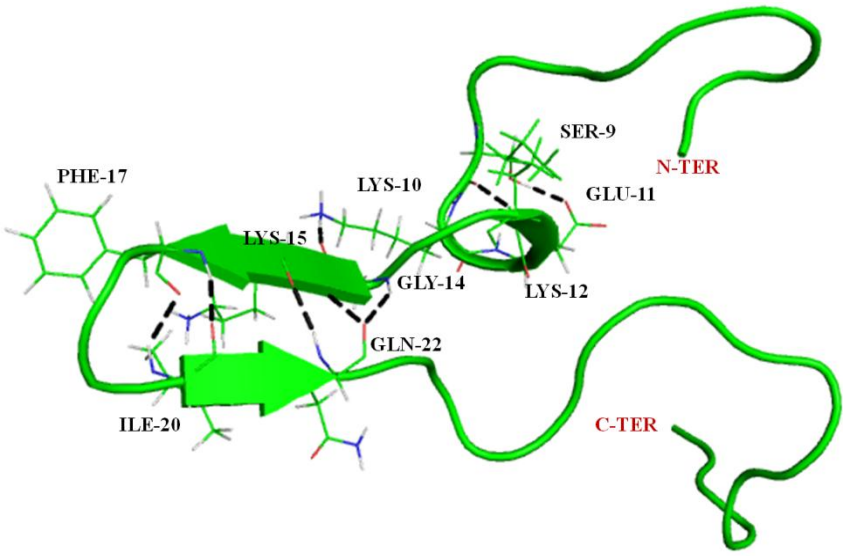
No.	Definition	2*Structure	MD^{ff96}(%)
1	(Leu^{2})CO…NH(Phe^{6})	α–helical	8.9
2	(Leu^{2})CO…NH(Arg^{7})	α–helical	7.3
3	(Gly^{3})CO…NH(Phe^{5})	α–helical	0.88
4	(Arg^{7})CO…NH(Asp^{4})	α–helical	0.33
5	(Glu^{11})CO…NH(Arg^{35})	α–helical	0.60
6	(Glu^{16})CO…NH(Asn^{23})	α–helical	-
7	(Lys^{18})CO…NH(Val^{21})	α–helical	15.0

8	(Val^{21})CO…NH(Lys^{18})	α–helical	8.2
9	(Asn^{23})CO…NH(Lys^{15})	β–turn	0.01
10	(Lys^{25})CO…NH(Ile^{13})	β–turn	7.5

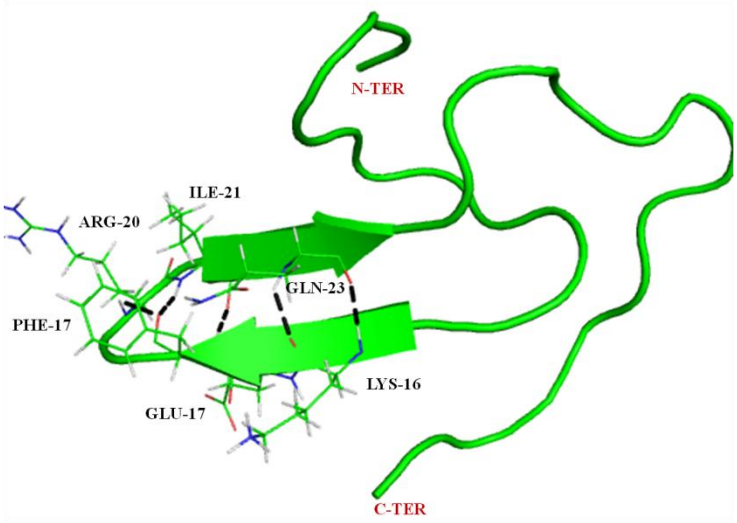
4.1.6 Structural Analysis

AMBER ff96 anti parallel β sheet conformations were sampled from Lys^{15}-Phe^{17} and Ile^{20}-Gln^{22} which are being stabilized by backbone hydrogen bonds between residue Phe^{17} and Ile^{20} **(Figure 20 A-F).** Large aromatic residues like Phe and β-branched amino acids such as Ile are favoured to be found in β strands and such anti parallel β sheets in antimicrobial peptides have been reported (L. Tomasinsig and M. Zanetti 2005). Anti parallel β sheet is the most prevalent structure between residues. With the change in orientation of the aromatic ring of Phe in the trajectory the length of the β sheet varies. Phe^{17} is a key residue involved in its formation throughout the trajectory hence we are encouraged to believe that it plays a role in the formation and stability of the β sheet conformation.

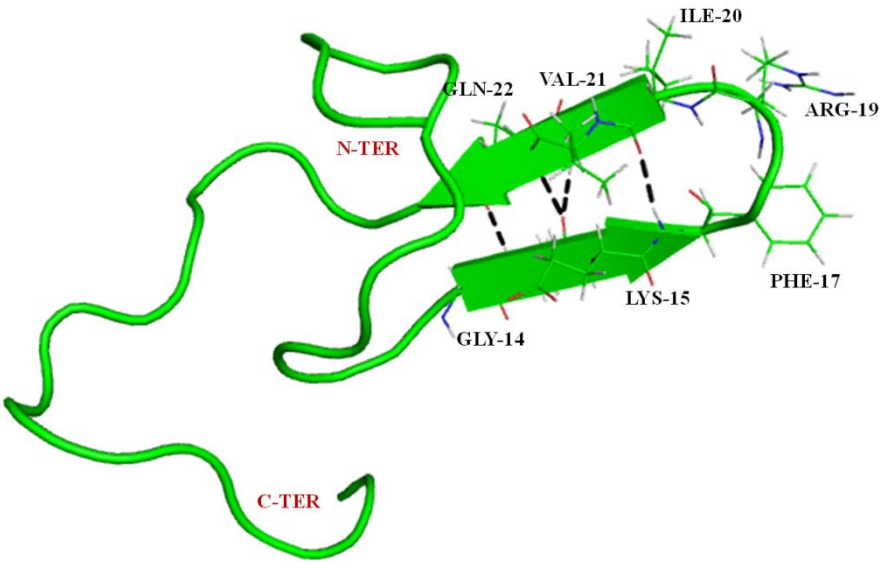
Helical conformations were sampled predominantly when ff99 was used **(Figure 21 A-F)** Mostly the helical nature of the peptide was found to be near the N and C terminal throughout the trajectory. Key residues involved being Phe^{6}, Lys ^{9} towards the N terminal and Arg^{24}, Asp^{27} and Leu^{32} towards C terminal. π-helix formation was observed from Asp^{4} –Ser^{9} around 90 ns and a -helix from Asp^{4}-Lys^{8} towards N-terminal at 100 ns. α-helical conformation was also seen at the start of the trajectory between Asp^{4} and Lys^{8}. In the 70ns snapshot α-helix towards C-terminal from residue Asp^{26} - Asn^{30} was observed. The N- and C-terminal of the peptide was found to be in disordered random coil structure.


SER-9
N-TER
PHE-17
LYS-10
GLU-11
LYS-15
GLY-14
LYS-12
GLN-22
ILE-20
C-TER

A


N-TER
ILE-21
ARG-20
GLN-23
PHE-17
LYS-16
GLU-17
C-TER

B


ILE-20
GLN-22
VAL-21
ARG-19
N-TER
PHE-17
LYS-15
GLY-14
C-TER

C

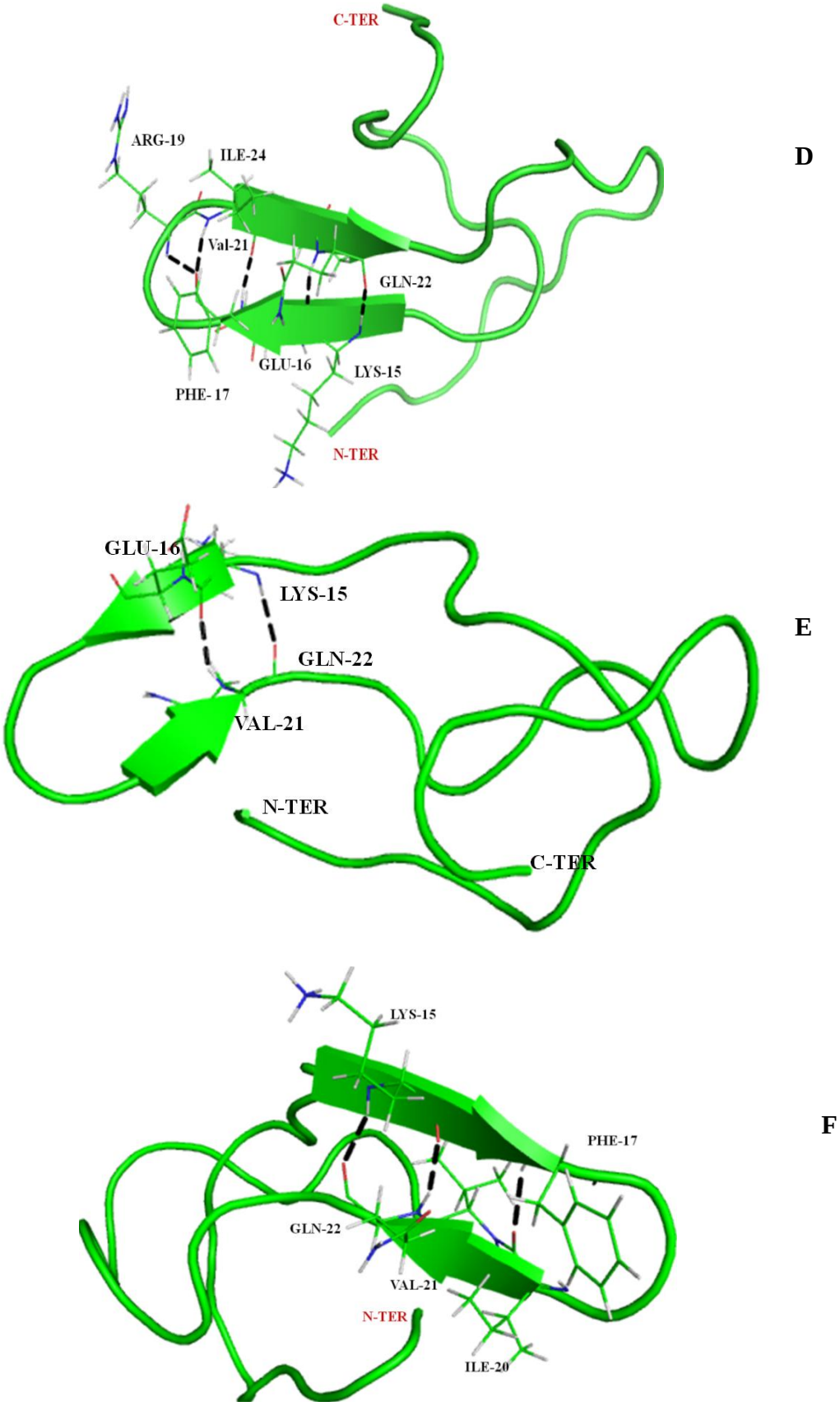
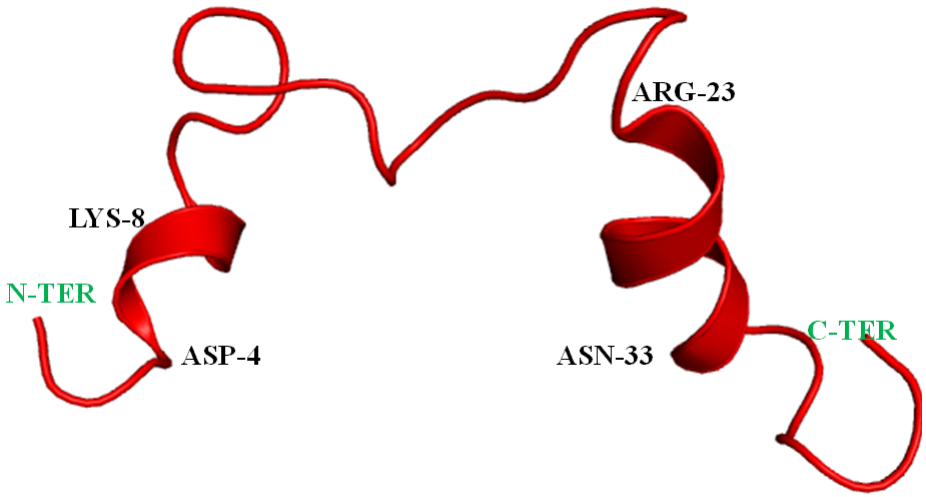
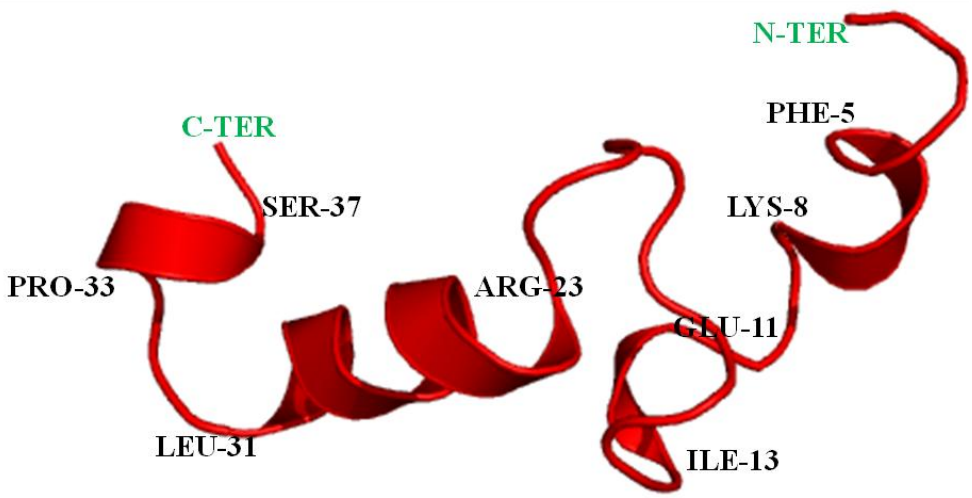


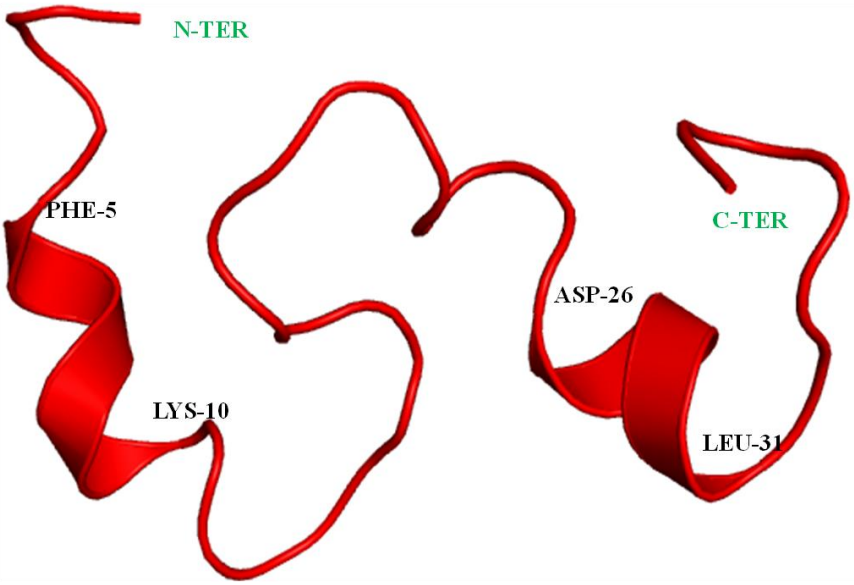
Figure 20: Average Structures of linear LL-37 obtained through out 100 ns trajectory at different time intervals **(A-F)** under force field AMBER ff96


ARG-23
LYS-8
N-TER
ASP-4
ASN-33
C-TER

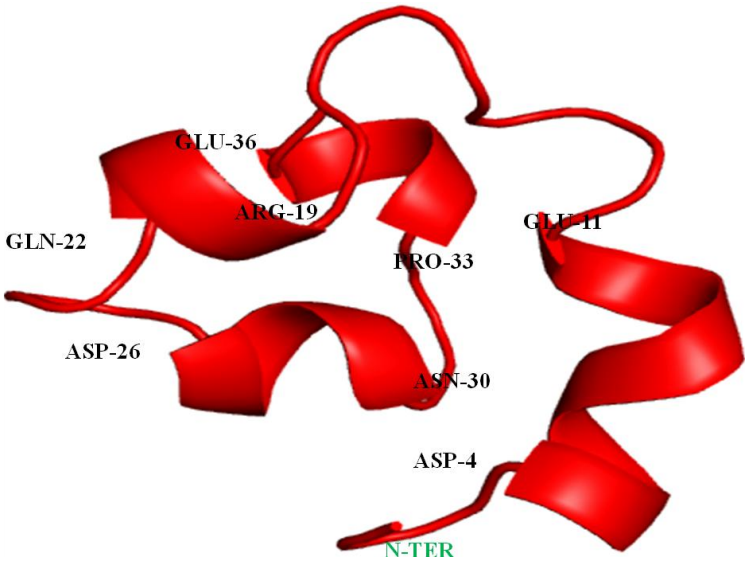
A


N-TER
PHE-5
C-TER
SER-37
LYS-8
PRO-33
ARG-23
GLU-11
LEU-31
ILE-13

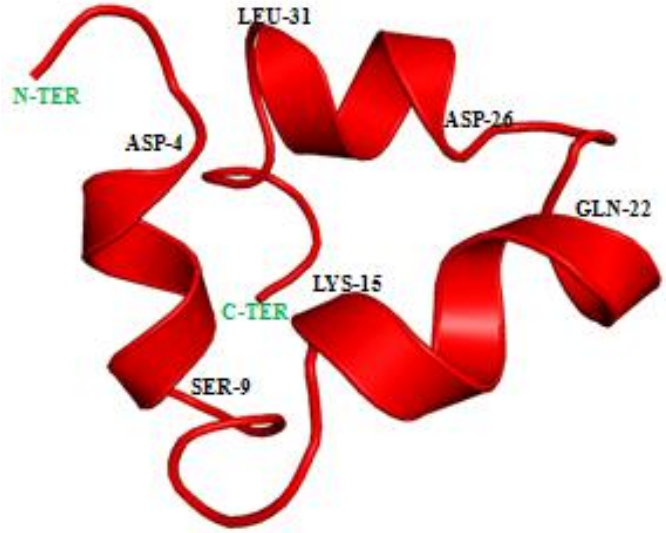
B


N-TER
PHE-5
C-TER
ASP-26
LYS-10
LEU-31

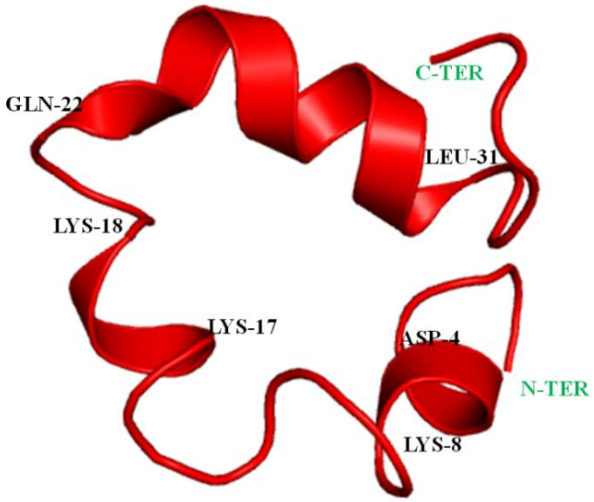
C



D



E



F

Figure 21: Average Structures of linear LL-37 obtained through out 100 ns trajectory at different time intervals **(A-F)** under force field AMBER ff99

4.2 MD on NMR Structure of LL-37 using GB Method and Langevin Thermostat Algorithm.

4.2.1 Thermodynamic Profiles

The thermodynamic parameters such as temperature and the total energy were plotted with the plotting software Grace to measure the quality of the MD trajectory **(Figure 22 A** and **B)**. The plots testify that the simulation is relatively stable over the entire MD trajectory, in both AMBERff96 and AMBERff99SB force fields. The kinetic energy is stable as expected since the temperature plot, which is directly proportional to the kinetic energy, was stable **(Figure 22 A).**

The temperature plots **(Figure 22 B)** reveal that after equilibration for 100 ps at 300K the temperature is constant at 300K throughout the trajectory in case of both force field ff96 and ff99SB and the system is relaxed**.**

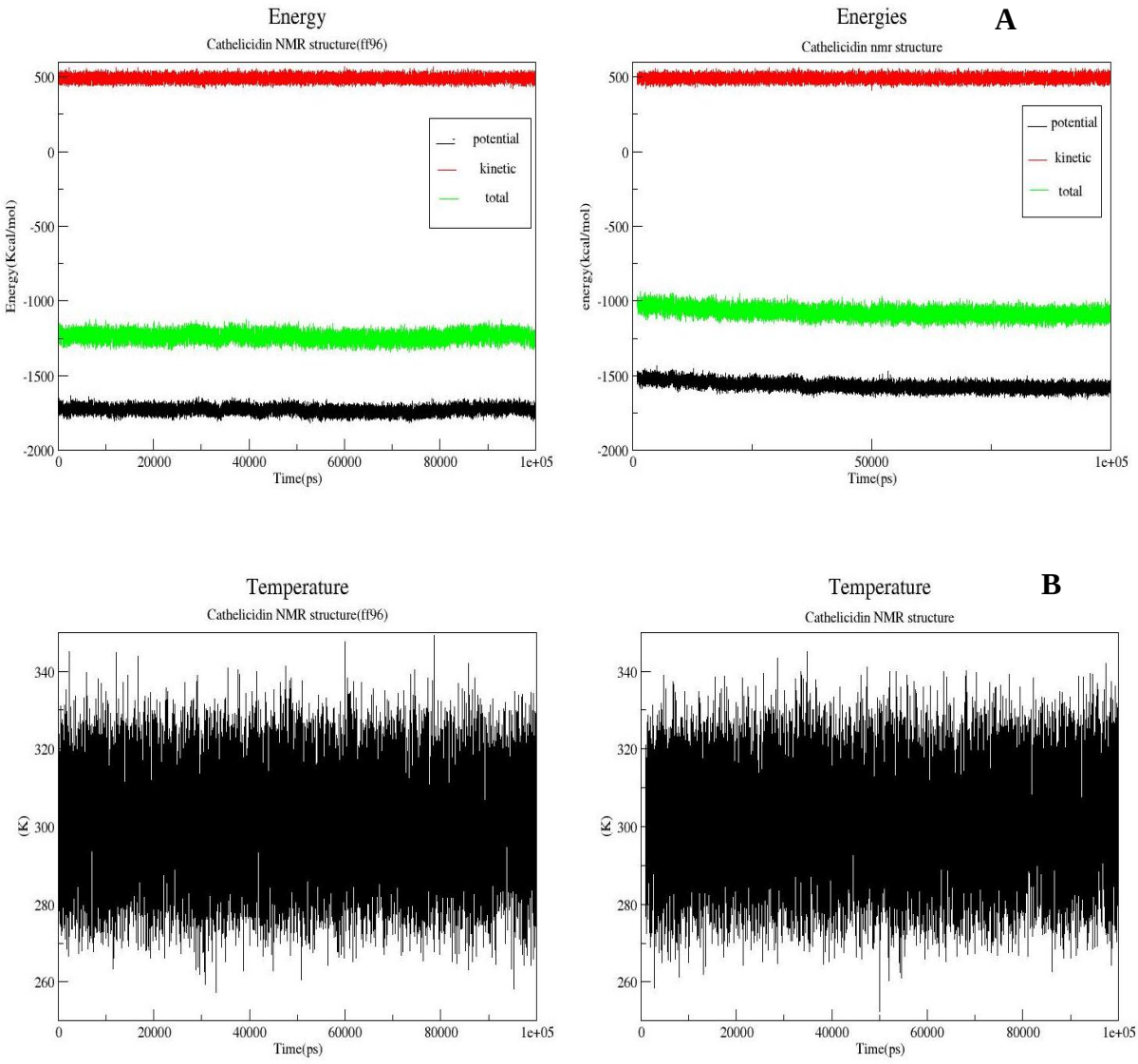


Figure 22: Themodynamic profile plots of MD trajectory of NMR structure of LL-37 under force field AMBERff96 and AMBER ff 99SB

4.2.2 RMSD Analysis

To have a better insight into the structural features of the cathelicidin and know about the structural stability of the conformers obtained RMSD analysis was done with the initial structure considered as a reference. **Figure 23** represents the root-mean-square deviations (RMSD) of the MD trajectory with respect to the backbone atoms of the starting structure using PTRAJ module of AMBER12 (Case 2006).
From the plot of RMSD vs Time as evident there was a significantly higher fluctuation from the start of the simulation until 20ns in force field ff99SB **(Figure 23 shown in red)** and in force field ff96 **(Figure 23 shown in black)** shows fluctuations throughout from the starting of the trajectory clearly indicating larger structural deviations of the conformations from the reference structure in force field ff96.

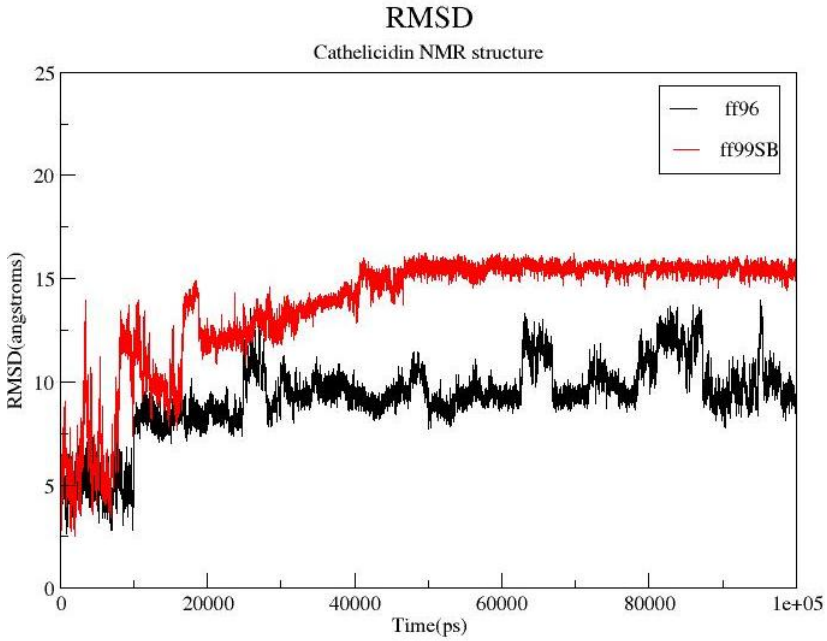


Figure 23: Root mean square deviations (backbone-backbone) of NMR structure of LL-37 from the starting structure in case of MD ^{implicit} force field ff99 SB in red and MD ^{implicit} force field ff96 in black obtained at 300K.

A plateau about RMSD~15 Å in case of ff99SB formed from 50ns to 100ns segment of simulation trajectory shows comparatively stable structures, reason for this probably being the sufficient folding during MD sampling. But in case of force field ff96 plateau is formed at different time steps dividing it into different clusters which signify that at there is not much fluctuations in the peptide in those intervals. This data also suggest that the peptide have folded to a considerable extent after 50ns of the trajectory in ff99SB, and does not show any sharp changes in the conformations obtained.

4.2.3 Pattern Profiles

In order to get a better understanding of the distribution of structures in the MD trajectory, on the basis of their RMSD patterns shown in **Figure 23.** The sampling efficiency of the

MD simulations by using force fields ff96 and ff99SB was monitored for the different patterns of LL-37 obtained during the progress of each trajectory in terms of pattern profiles using the CLASICO program and is pictorially depicted in the **Figure 24 (A) and (B).** Patterns basically represent the structures classified on the basis of the type of conformational motifs present(Corcho F 2004)

From a batch of 100000 structures, 26,000 new patterns (**Figure 24 A**) were observed for MD ff96, while 20000 (**Figure 24 B**) new patterns were obtained for MD ff99SB. The efficiency of these clusters in terms of generating new patterns is in the order: MD FF96 (26%) > MD ff99SB (20%). Despite this, MD simulation with force field ff96 showed maximum similarity of patterns in majority of the conformations, which clearly suggests that the trajectory is stable. This is consistent with the RMSD results obtained in the segment.

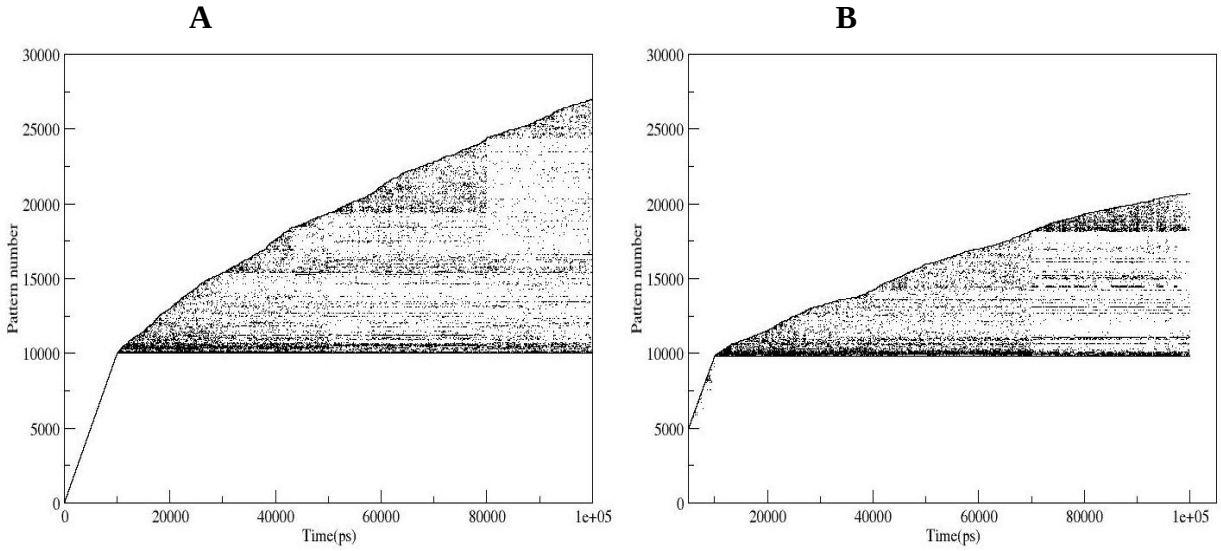


Figure 24: Evaluation of new patterns along the trajectories for NMR structure of LL-37 in **(A)** MD ^{implicit} force field ff96 **(B)** MD ^{implicit} force field ff99SB

A closer look at **Figure 24 (A) and (B)** reveals that upto 10 ns in case of ff99SB **(Figure 24 B)** new patterns were generated but from thereafter until 100ns similar patterns were generated in a regular fashion throughout the length of the trajectory.

Further characterization involved the identification of the secondary structure motifs of the conformations obtained in the MD simulation using the force field ff96 and ff99SB was done with the help of CLASICO program.

4.2.4 Secondary Structural Motifs

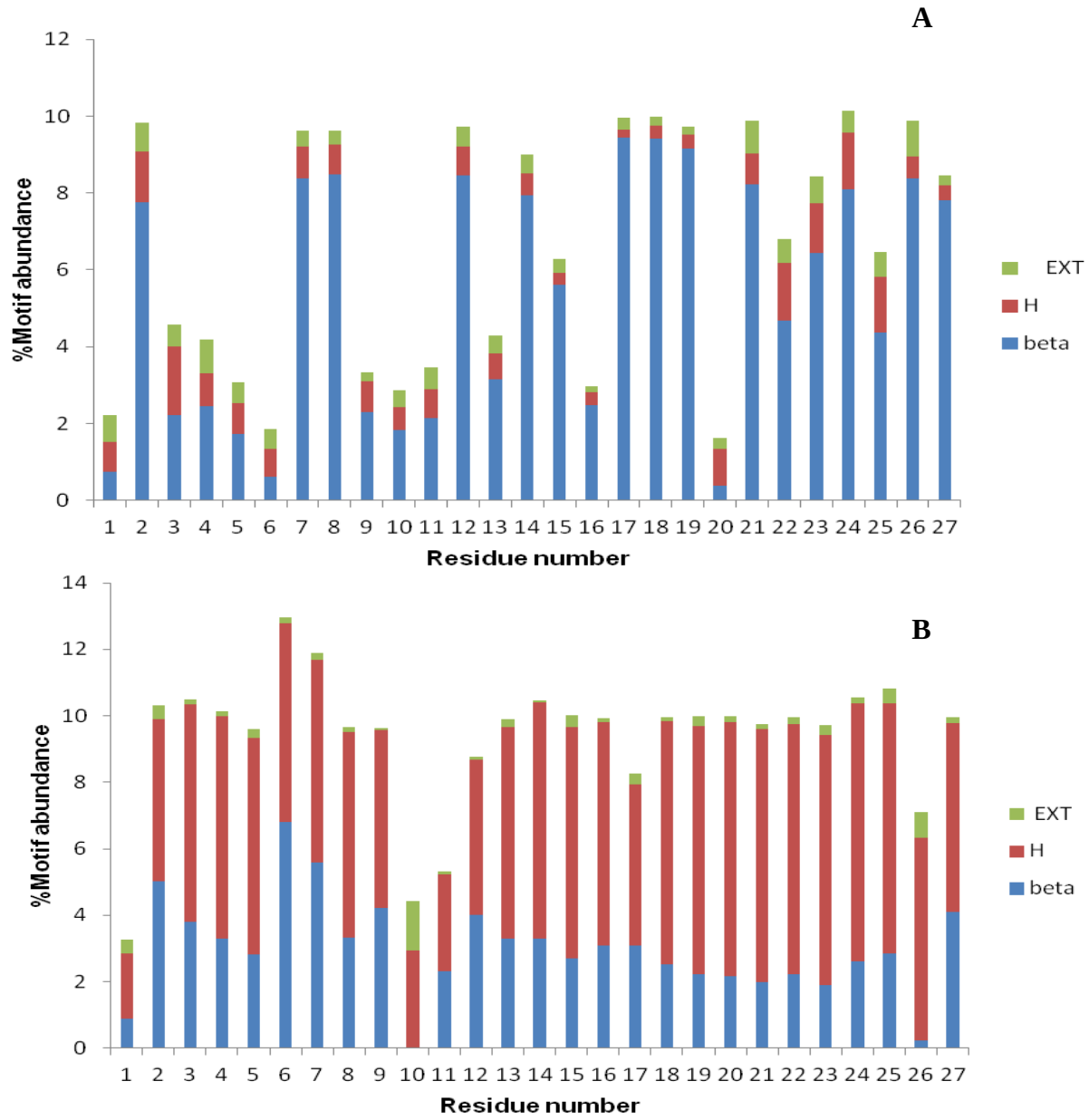
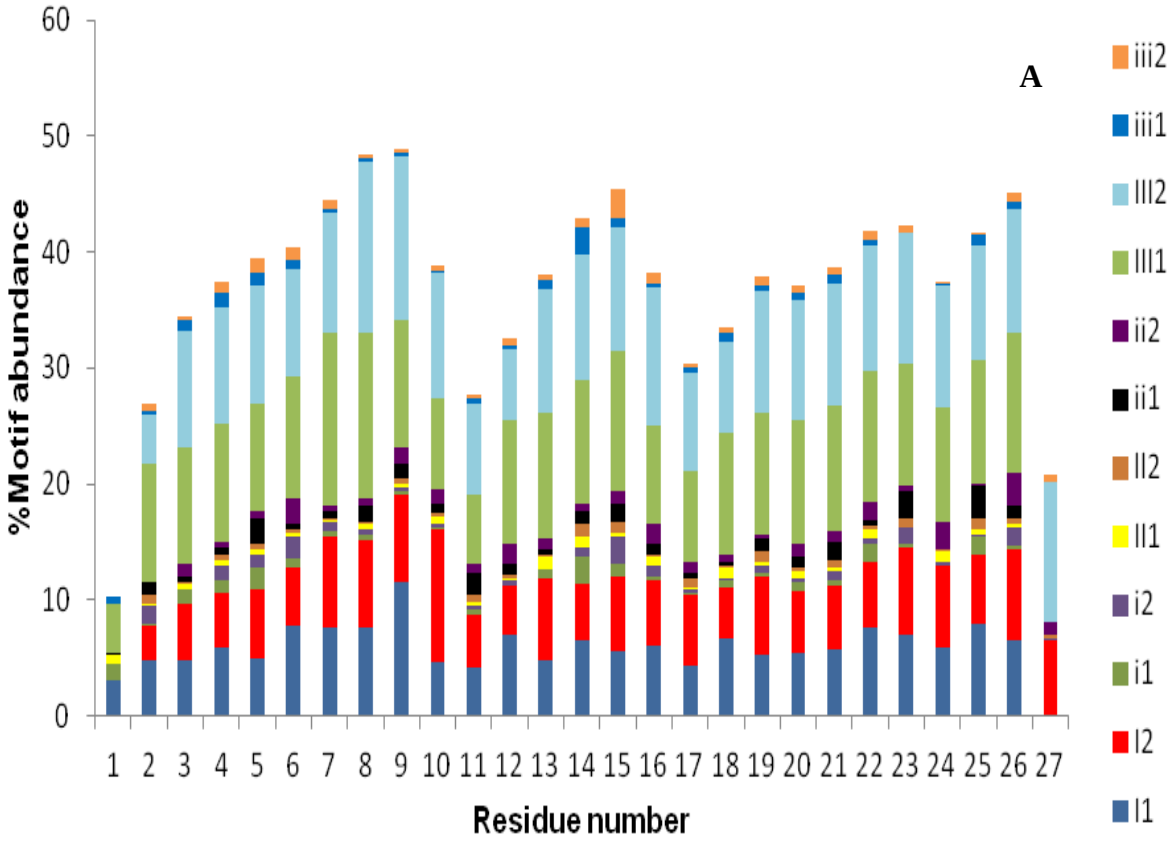


Figure 25: Motif abundance from MD trajectory NMR structure of LL-37 under **(A)** force field ff96 **(B)** force field ff99 SB.

Figure 25 (A) and **(B)** represent the statistics of the conformational motifs for each residue of the cathelicidin peptide in MD of force field ff96 and ff99SB respectively. A closer inspection of **Figure25 (A)** reveals that alpha-helical are the preferred secondary motifs acquired by most of the residues in the sampled conformations.

Figure 25 (A) show about 10-13% for the MD calculations using AMBER ff96 and about 10% for the MD calculation using AMBER ff99SB. **Figure25 (A)** depicts the secondary structure motifs obtained during the MD simulation force field ff96, shows predominantly a helical region between residues, Leu^{2} to Phe27, a stronger propensity between the residue Gly^{3} to Ser9 and Gly^{14} to Phe^{27} and residues Lys^{10} ,and Glu^{11} shows lower propensity helical region. To some extent around 3%-4% of consecutive β turns were obtained giving

rise to helical conformation. The N and C-terminal region of the peptide mostly remain extended and flexible not showing any secondary structural feature. **Figure 25 (B)** shows the secondary structural motifs obtained during the MD simulation by using force field ff99SB, shows a predominantly β-turns mostly around the C terminal region with a higher propensity being between residue Val[21] and Phe[27]. Rest of the C terminal was extended and flexible.



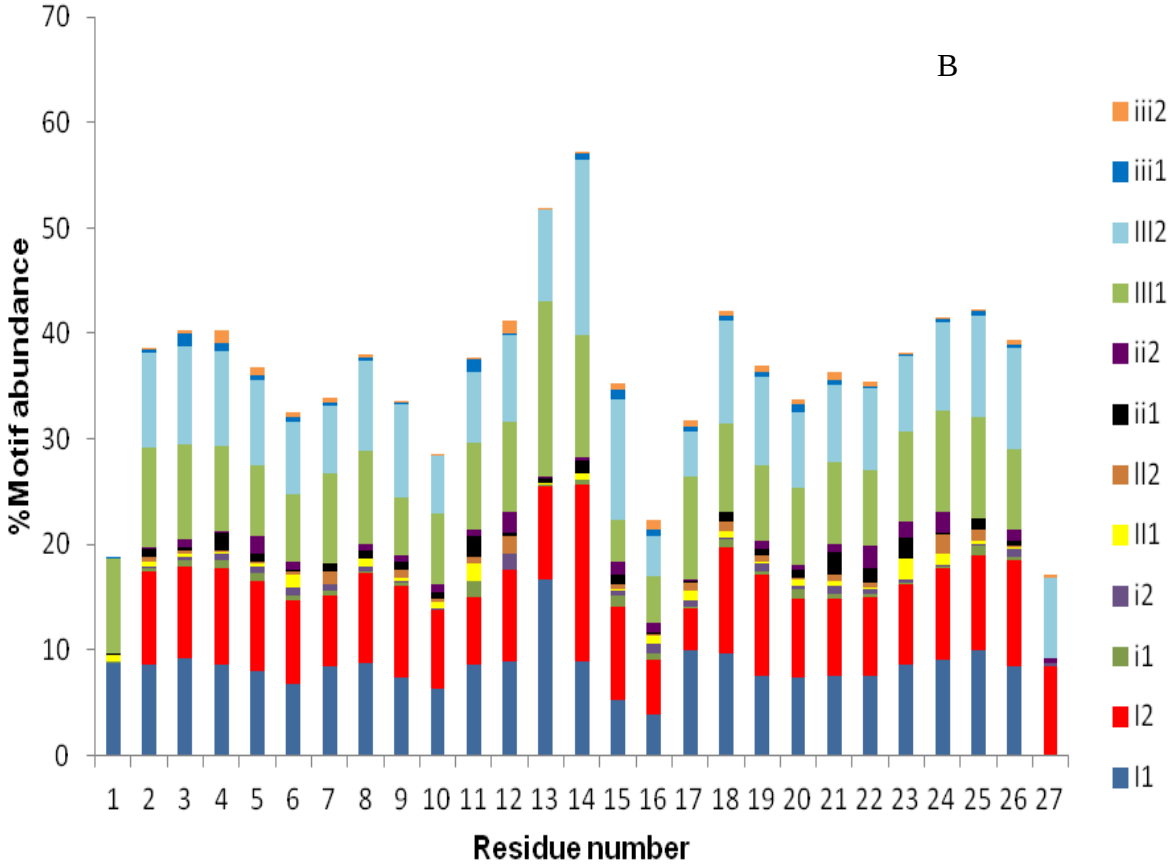


Figure 26: Classification of β turns obtained in the MD trajectory of NMR structure of LL-37 **(A)** force field ff96 **(B)** force field ff99SB

The β- turn secondary motifs which are used consecutively in the formation of α- helix attained by residues of the cathelicidin peptide were further classified into different types (**Table 1**) using the two-residue window of the CLASICO program, and are pictorially depicted in Fig. Clearly, all residues exhibited β- turn type III in the conformations from force field ff96 **(Figure 26 A).** The profile also reveals propensity of type I β- turn conformations throughout the trajectory highest being between residues Phe6 – Lys10. The conformations sampled by using ff99SB (**Figure 26 B)** exhibited almost similar profiles showing β- turn type III as their preferred secondary features throughout the trajectory. However, the extent of β- turns was more predominant in residues Glu11-Gly14 and between Lys18-Asp26 adopted β- turns type III and type I both. The percentage order of β- turns adopted by the peptide in the different protocols used is MD ff96 MD and sff99SB (50%).

4.2.5 Hydrogen Bond Analysis

Despite having large differences in RMSD values, the secondary structure analysis performed could not efficiently distinguish between the structural features of peptides under both the force field used. Therefore, the observed propensities were further

rationalized by intramolecular hydrogen bond analysis performed using PTRAJ module of AMBER 12 program, and the results are summarized in **Table 5.**

Table 5: Secondary structures observed due to mainchain- mainchain hydrogen bond interactions and their percentages in trajectory MD ff 99SB. Secondary structures α –helix, β-turn, γ -turn are assigned in terms of hydrogen bond interactions.

S. No.	Definition	2*Structure	MD^{ff99SB}(%)
1	(Gly^{3})CO…NH(Arg^{7})	α–helical	2.41
2	(Asp^{4})CO…NH(Lys^{8})	α –helical	2.06
3	(Arg^{7})CO…NH(Glu^{11})	α –helical	2.11
4	(Lys^{8})CO…NH(Lys^{12})	α –helical	0.41
5	(Lys^{15})CO…NH(Arg^{19})	α –helical	1.10
6	(Val^{21})CO…NH(Lys^{25})	α –helical	0.33
7	(Phe^{27})CO…NH(Leu^{31})	α –helical	5.2
8	(Leu^{28})CO…NH(Val^{32})	α –helical	5.25
9	(Gly^{3})CO…NH(Phe^{6})	β–turn	3.22
10	(Lys^{8})CO…NH(Glu^{11})	β–turn	1.83
11	(Ser^{9})CO…NH(Lys^{12})	β–turn	4.48
12	(Lys^{15})CO…NH(Lys^{18})	β–turn	2.71
13	(Val^{21})CO…NH(Ile^{24})	β–turn	5.6
14	(Lys^{25})CO…NH(Leu^{28})	β–turn	_
15	(Phe^{27})CO…NH(Asn^{30})	β–turn	6.52
16	(Lys^{25})CO…NH(Phe^{27})	γ-turn	0.0

The geometrical criterion used for the donor (A)-acceptor (B) distance was ≤3.5 Å and the angle HAB (<HAB) ≤ 120°.In order to obtain significant result, only the hydrogen bonds with a percentage of existence ≥1.0 was considered. The hydrogen bond between i and i + 3 residues correspond to a beta-turn, while that between i and i + 4 residues represent an alpha-helical conformation. In this case, A is donor (i.e. N-H) and B the acceptor (i.e. O of carbonyl group). **Figure 27** is a diagrammatic representation of the permanence of the difference in hydrogen bonding patterns along the dynamics for the trajectory performed using the Langevin thermostat. During the initial 10 ns of the trajectory a hydrogen bond formed between different residues corresponding to helicity, β-turn conformation in some of the structures. HBs responsible for the helicity are almost negligible beyond 15 ns whereby the β-turn between residues 15 to 18 can be observed beyond. Table shows the type of interactions along with the percentage of conformations, calculated in terms of mainchain - mainchain hydrogen bonding. Although most of the conformations sampled in each of the trajectories correspond to α-helical (residues 3 to 12, 15 to 19 and 21 to 32). These results also demonstrate the formation of β-turn predominantly in the starting residues (Gly^{3}-Phe^{6}, Lys^{8}-Lys^{12} and Lys^{12}-Gly^{14}), and the central residues (Lys^{15}-Arg^{18} and Val^{21} - Phe^{24}). Reverse turns

were also observed in MD between Lys[25] to Phe[27] residues. Hence our results suggest the predominance of the β-turn, on the basis of hydrogen bond analysis, were in agreement with those obtained from the secondary structure analysis.

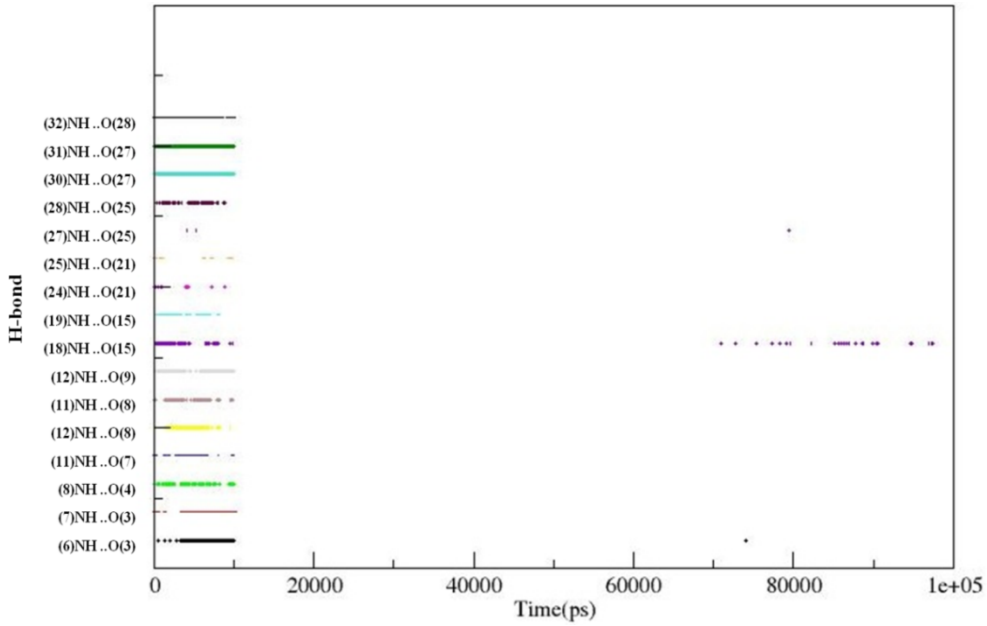


Figure 27: Progress of hydrogen bonds **(Table 5)** monitored between important residues for cathelicidin in **MD ff 99SB** trajectories.

On the other hand **Table 5** summarizes the results obtained by mainchain-mainchain hydrogen bond analysis performed using PTRAJ module of AMBER 12 program. The table shows the interaction between backbone atoms responsible in the formation of hydrogen bonds. The results obtained revealed that conformations in ff96 adopted α-helical region predominantly between the residues Gly[3] and Glu[11], the central region of the peptide from Lys[12] to Ile[20] and the C-terminal residues from Val[21] to Asp[26] and Leu[28] to Val[32] and were in agreement with those obtained from the secondary structure analysis described earlier in **Figure 25 (A) and (B).** β-turns were also seen between residues Phe[6] -Ser[9], Glu[11]-Gly[14], Glu[16]-Arg[19], Gln[22]-Lys[25] and Leu[28]-Leu[31] with a lesser propensity.

Table 6: Secondary structures observed due to mainchain-mainchain hydrogen bond interactions and their percentages in trajectory of NMR structure LL-37 under ff96. Secondary structures, α –helix, β-turn, γ -turn are assigned in terms of hydrogen bond interactions.

S. No.	Definition	2* Structure	MD^{ff96}(%)
1	(Gly^{3})CO…NH(Arg^{7})	α–helical	2.63
2	(Asp^{4})CO…NH(Lys^{8})	α –helical	7.85
3	(Phe^{6})CO…NH(Lys^{10})	α –helical	6.54
4	(Arg^{7})CO…NH(Glu^{11})	α –helical	6.67
5	(Glu^{11})CO…NH(Lys^{15})	α –helical	5.31
6	(Lys^{12})CO…NH(Glu^{16})	α –helical	3.47
7	(Ile^{13})CO…NH(Phe^{17})	α –helical	2.91
8	(Gly^{14})CO…NH(Lys^{18})	α –helical	5.67
9	(Lys^{15})CO…NH(Arg^{19})	α –helical	7.33
10	(Glu^{16})CO…NH(Ile^{20})	α –helical	7.77
11	(Phe^{17})CO…NH(Val^{21})	α –helical	5.37
12	(Val^{21})CO…NH(Lys^{25})	α –helical	8.0
13	(Gln^{22})CO…NH(Asp^{26})	α –helical	8.06
14	(Asp^{26})CO…NH(Asn^{30})	α –helical	-
15	(Leu^{28})CO…NH(Val^{32})	α –helical	9.22
16	(Phe^{6})CO…NH(Ser^{9})	β–turn	1.81
17	(Glu^{11})CO…NH(Gly^{14})	β–turn	2.93
18	(Glu^{16})CO…NH(Arg^{19})	β–turn	1.37
19	(Gln^{22})CO…NH(Lys^{25})	β–turn	0.42
20	(Asp^{26})CO…NH(Arg^{29})	β–turn	-
21	(Leu^{28})CO…NH(Leu^{31})	β–turn	1.40
22	(Arg^{7})CO…NH(Lys^{12})	π–turn	6.67
23	(Gln^{22})CO…NH(Phe^{27})	π–turn	0.57
24	(Asn^{23})CO…NH(Leu^{28})	π–turn	0.59

Figure 28 which is the diagrammatic representation of the permanence of the hydrogen bonding patterns along the dynamics for the trajectory shows that during the initial 10 ns of the trajectory a hydrogen bond formed between different residues corresponding to helicity, β-turn secondary structural features. HBs responsible for the helicity are almost negligible beyond 15 ns between many residues, whereby the helical region between residues 4 to 8, 6 to 10 and 26 to 30 can be observed during the time period 20 ns-30 ns, 80 ns-100 ns and 80 ns -90 ns, respectively. Similarly small stretch of β-turn and π-turn between residues 26 to 29 and 23 to 28, respectively were visible after 15 ns but to a small extent of time interval. Overall these results suggest that the peptide has a high propensity to remain in the α-helical conformation.

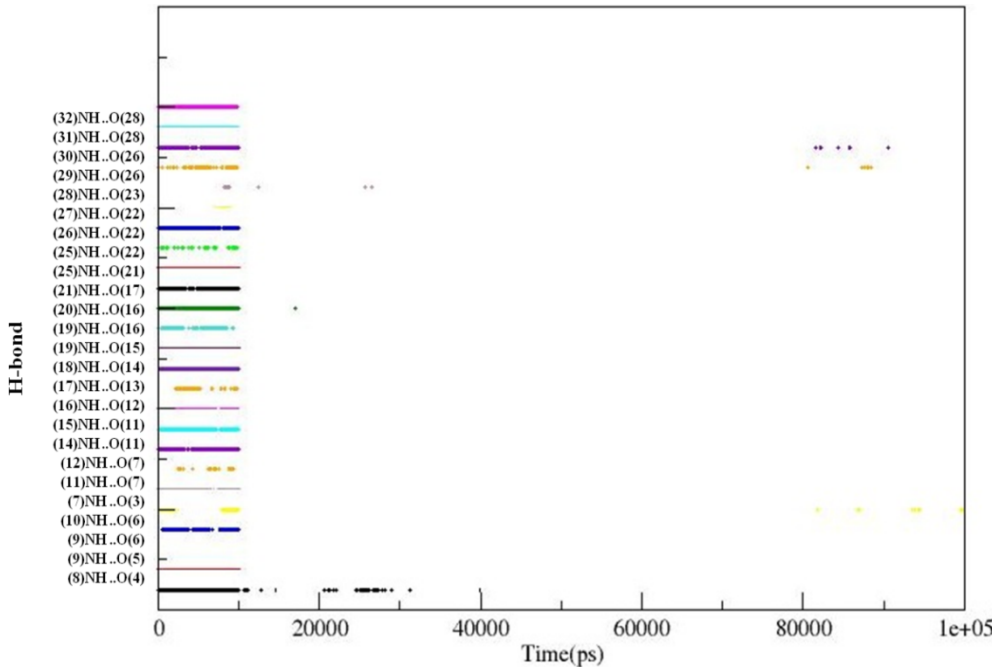
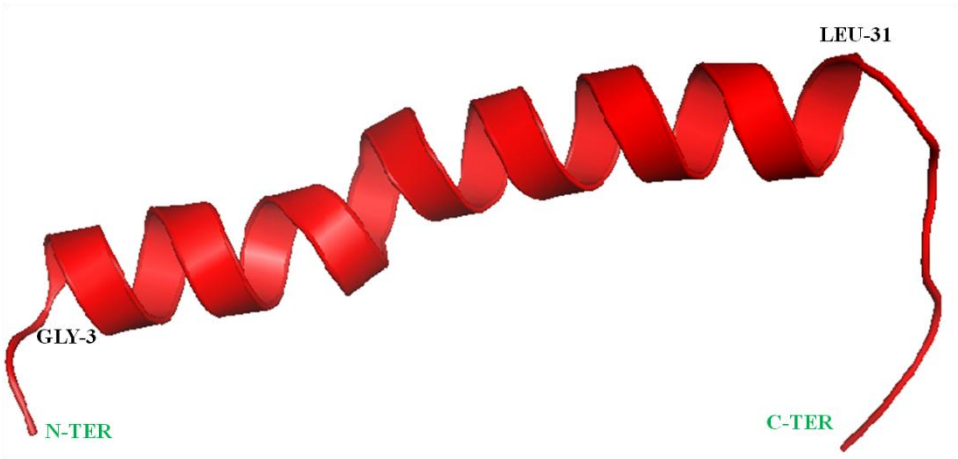


Figure 28: Progress of hydrogen bonds **(Table 6)** monitored between important residues for cathelicidin in **MD ff96** trajectories

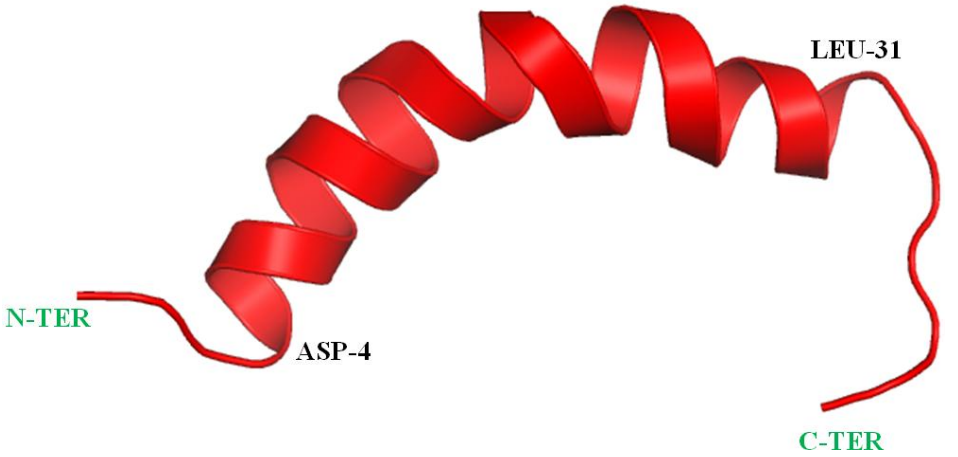
4.2.6 Structural Analysis

In order to get a clearer picture of the structural features of peptide conformations obtained using force field ff96 and ff99SB, the average structures were obtained at different time interval using PTRAJ. Under the effect of ff96, the average structure showed only alpha helix and turns and to some extent loops also, clearly supporting the results obtained from secondary structure analysis and hydrogen bond analysis. The alpha helical part remained stable for most length of the trajectory **(Figure 29 A-F).**

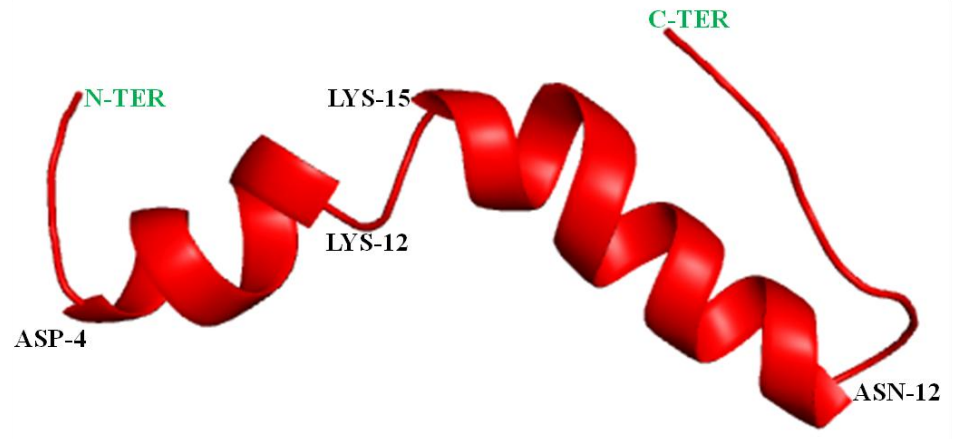
For force-field ff 99SB **Figure 30 (A-F)** the average structure exhibited different α- helical regions between residues in the N- terminus (Asp^{4}-Lys^{12}, Arg^{7}-Glu^{11}), second in the central region (Ile^{20}-Ile^{24}) and the third flanked by residues (Ile^{20}-Leu^{31} and Lys^{25}-Leu^{31}) in the C-terminus region of peptide. Also 3_{10} helix in residues (Glu^{16}-Lys^{18}) and β- turn between residues (Phe^{5}-Lys^{8}, Gly^{14}-Phe^{17} and Leu^{28}-Leu^{31}) at the N-terminal, central and C-terminal region were reported **(Figure 30).**


LEU-31
GLY-3
N-TER
C-TER

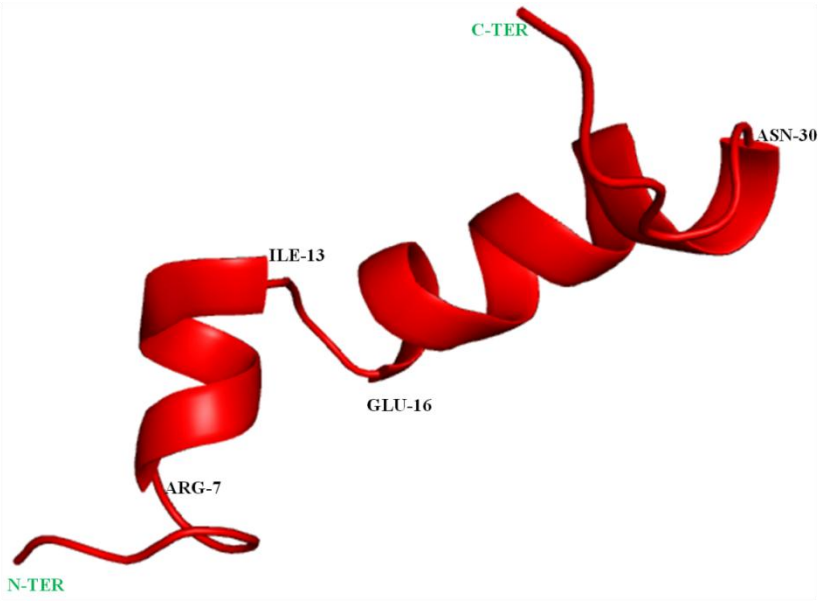
A


LEU-31
N-TER
ASP-4
C-TER

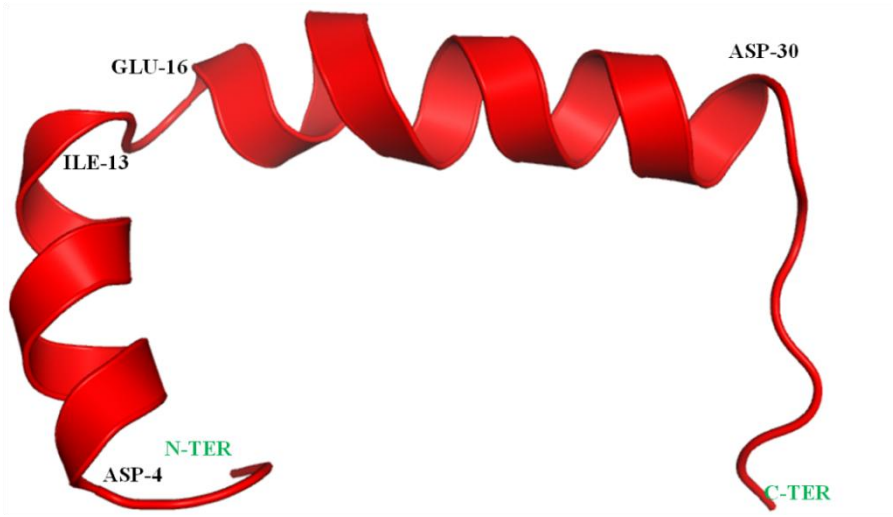
B


C-TER
N-TER
LYS-15
LYS-12
ASP-4
ASN-12

C



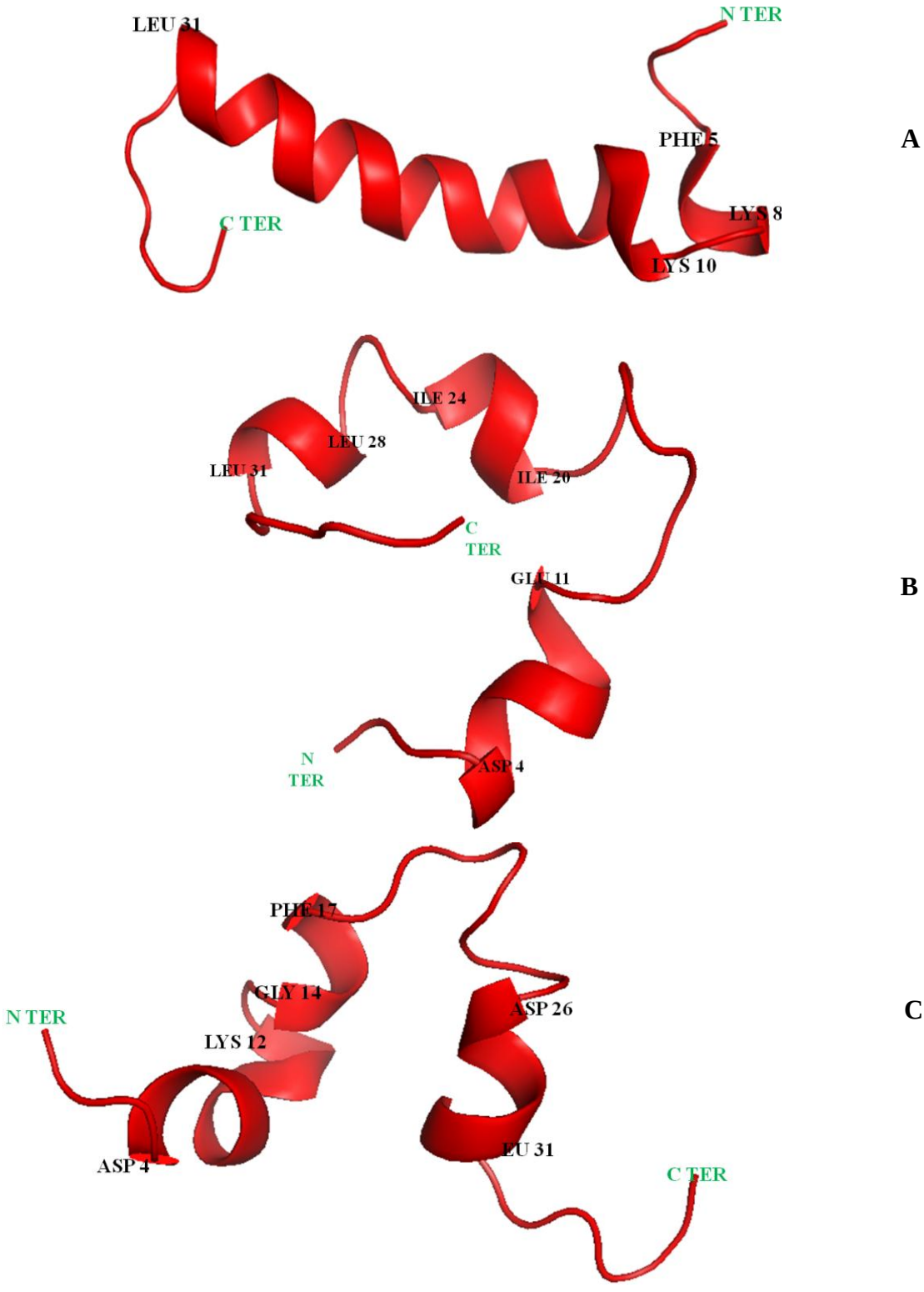
D



E

F

Figure 29: Average Structures of NMR structure LL-37 obtained through out 100 ns trajectory at different tine intervals **(A-F)** under force field AMBER


LEU 31
N TER
PHE 5
LYS 8
C TER
LYS 10
A
ILE 24
LEU 28
LEU 31
ILE 20
C
TER
GLU 11
B
N
TER
ASP 4
PHE 17
GLY 14
N TER
ASP 26
LYS 12
LEU 31
ASP 4
C TER
C

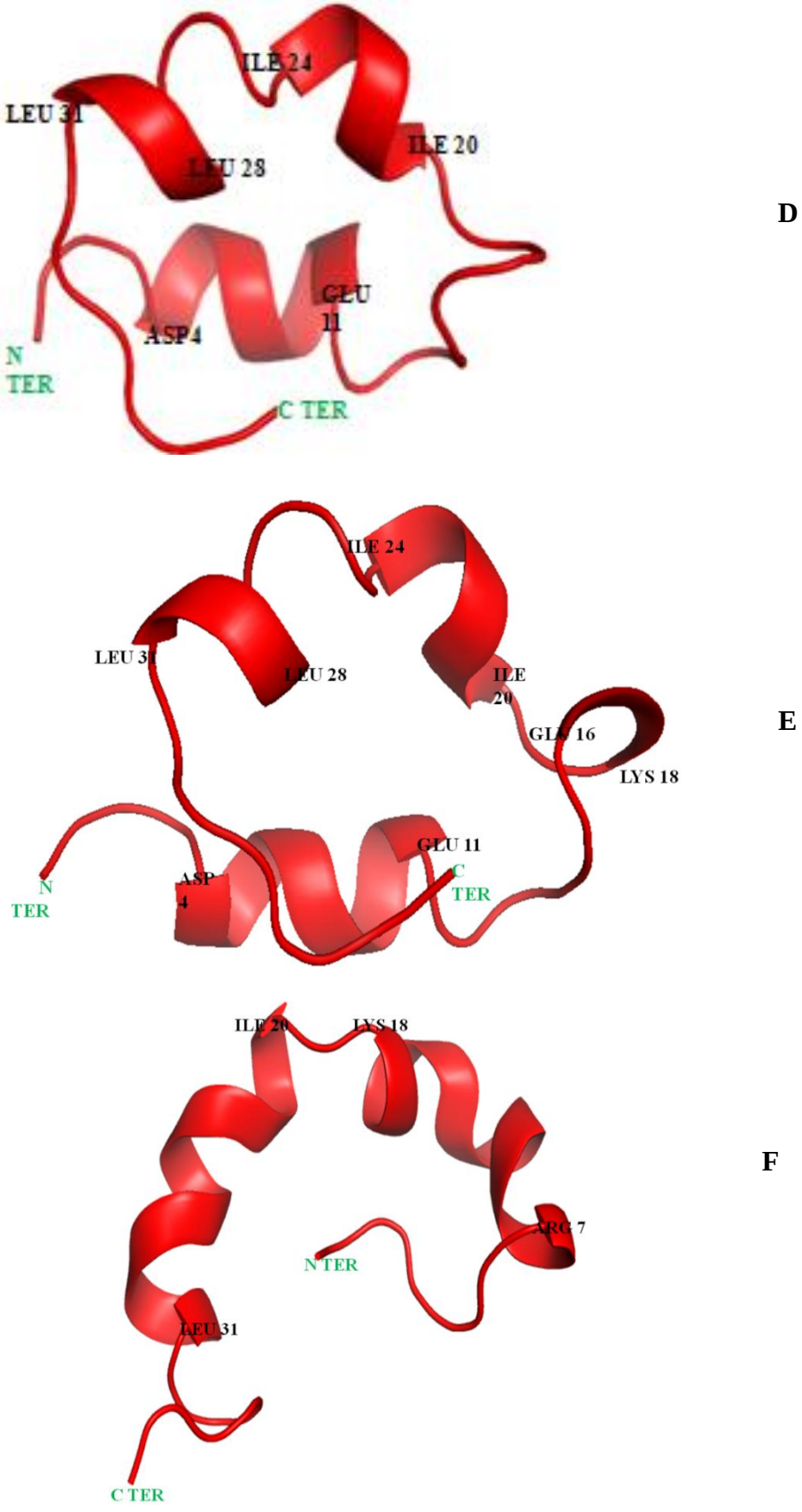


Figure 30: Average Structures of NMR structure LL-37 obtained through out 100 ns trajectory at different tine intervals **(A-F)** under force field AMBER ff 99SB

4.3 Analysis of MD Simulations using the Explicit Solvent Method and Berendsen Thermostat

The folding of the peptide LL-37 was carried out first at constant volume. This was done because the initially constructed system of peptide-water was not expected to undergo substantial density changes during the simulations. On the basis of this assumption, folding of peptide was then carried out at constant pressure. We found average pressures in peptide-water system in the range of ~ -300 to 350. The trajectories exhibit stable pressure throughout the run (**Figure 31**). Hence, no artificial pressures biasing the conformational sampling were present in the system.

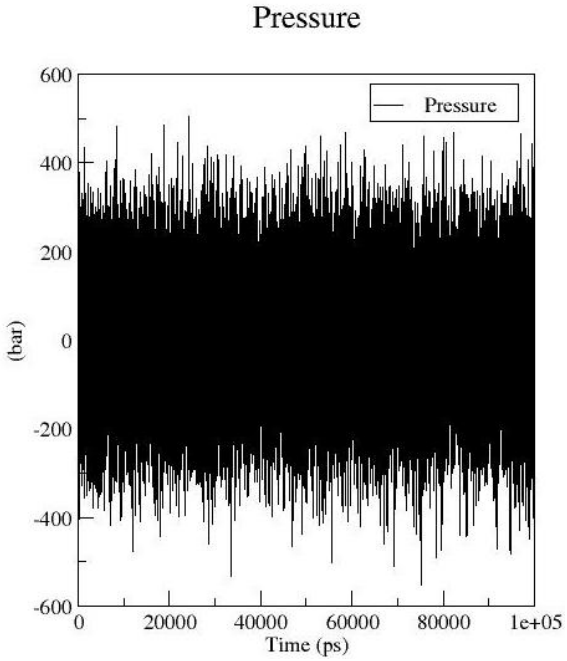


Figure 31: Constant pressure profile plot during the 100ns MD trajectory under force field gromos96 43a1.

The convergence of the peptide fold was examined by following the time evolution of thermodynamic and structural quantities of the peptide during the 100 ns simulation. Peptide total (free) energy, root mean square deviation (RMSD) of the peptide relative to the initial structure, and the solvent exposed surface area as well as the overall secondary structure was monitored.

4.3.1 Thermodynamic Profiles

The structural stability of cathelicidin was monitored in the 100-ns long MD simulation. The simulation was performed in explicit water with the number of atoms of 14,170 at 300 K initial temperature. The plots of temperature and the total energy testify that the simulation is relatively stable over the entire MD trajectory **(Figure 32).** The kinetic energy is stable as expected since the temperature plot, which is directly proportional to the kinetic energy, was

stable after equilibration at 300K. The total energy and potential energy plot also shows that, there is not much fluctuation and the system must be stable from the starting of the trajectory.

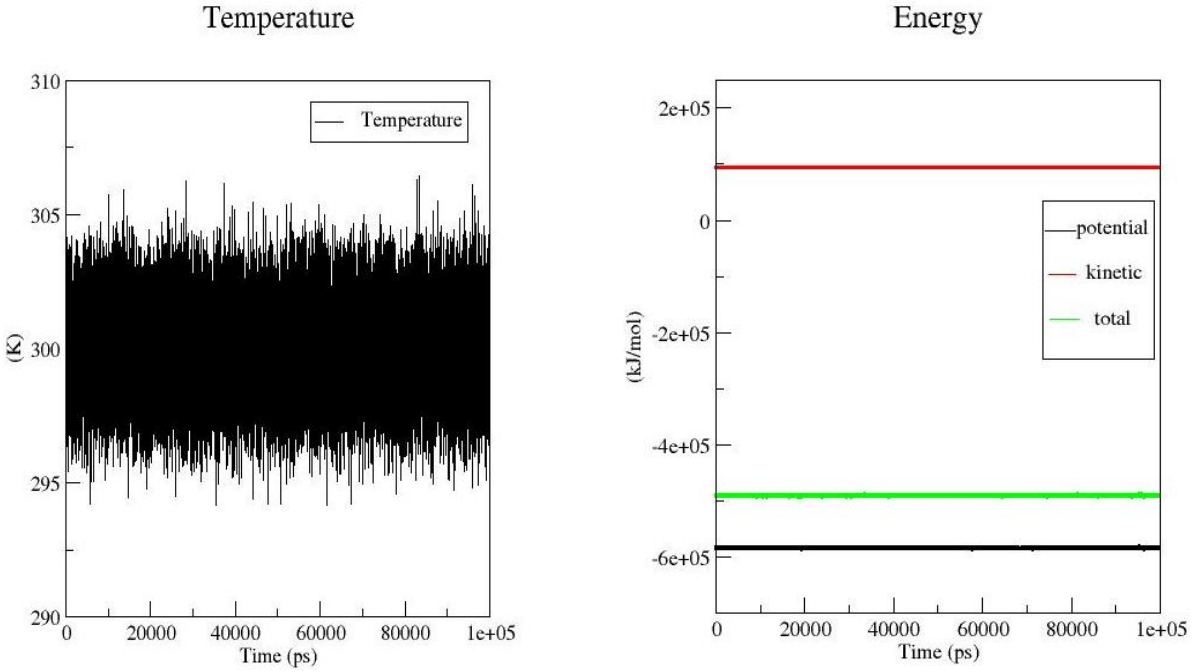


Figure 32: Themodynamic profile plots of MD trajectory of NMR structure of LL-37 under force field gromos96 43a1.

.

4.3.2 RMSD Analysis

RMSD analysis enables us to analyze one of the most important fundamental properties, that is whether the structures obtained are stable and close to the native structure known as the structural stability. Thus, to have a better insight into the structural features of the cathelicidin and to know about the structural stability of the conformers obtained RMSD analysis was done by the comparison of the backbone atoms, main chain atoms and c-alpha atoms of the MD conformers with the native NMR structure available in PDB. The plots of RMSD levels revealed five different clusters in which we got stable structures. The first from time 10 ns – 18 ns, second being from 20-30 ns then from 30 to 40 , 42 to 50 ns and from 55ns onwards till the end of the trajectory ie 100 ns we got a stable RMSD profile. A plateau about RMSD~12 Å was formed from 55 ns to 100 ns segment of simulation trajectory shows comparatively stable structures, reason for this probably being the sufficient folding during MD sampling **(Figure 32**).

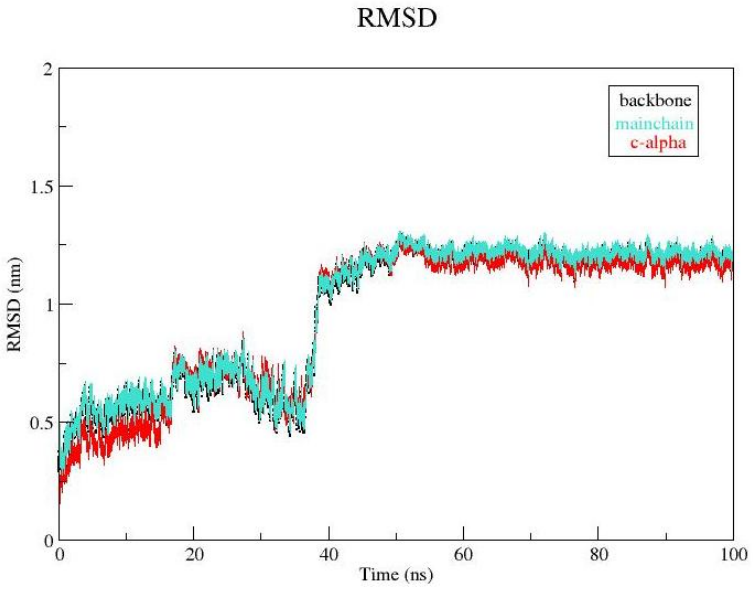


Figure 33: Root mean square deviations (backbone, main chain and C^{α}) of LL-37 from the starting structure obtained at 300K.

4.3.3 Radius of Gyration

The measure of the compactness of a protein is termed as radius of gyration. If a protein is stably folded a relatively stable value $R_{g,}$ will be maintained. If a protein unfolds, its R_g will change over time (Lemkul). The plot of radius of gyration with respect to time shows that initially the peptide shows unfolding but after 50 ns we get a stable radius of gyration value of 0.8 nm (8 Å). The unfolding pattern can be correlated with the clusters of RMSD plots. As there too we got a stable RMSD value after 50 ns and the Rg plot also shows that the peptide remains very stable in its compact, folded form from here on over the course of the trajectory at 300K.

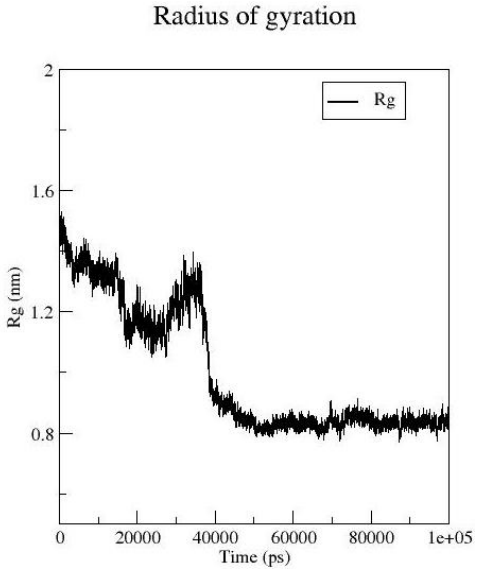


Figure 34: Plot of change in radius of gyration over the trajectory for LL-37 in explicit solvent.

4.3.4 Solvent Accessible Surface Area

The extent to which an amino acid interacts with the solvent environment and the protein core is proportional to the surface area exposed to the environment. The solvent-accessible surface area is a geometric measure of this exposure (Viacarra C 2005). Thus it is an important driving force in protein folding as it is a measure of effects like surface area minimization, burial of hydrophobic side chains, and side-chain packing density (Koehl P 1994). The graph of solvent accessible surface area clearly indicates the variation in accessible surface area as a result of conformational changes induced due to folding of the peptide throughout the trajectory. The graph of solvent accessible surface area clearly indicates the variation in accessible surface area as a result of conformational changes induced due to folding of the peptide throughout the trajectory from the starting of the trajectory upto40 ns there is fluctuations in hydrophobic and total surface area but beyond 50 ns both shows stable trajectory whereas in case of hydrophilic the trajectory is stable throughout with little or no fluctuation .Indicating that the water exposed fraction of the peptide and thus its conformation, remains constant in accord with solvent-accessible surface area calculations **(Figure 34)**

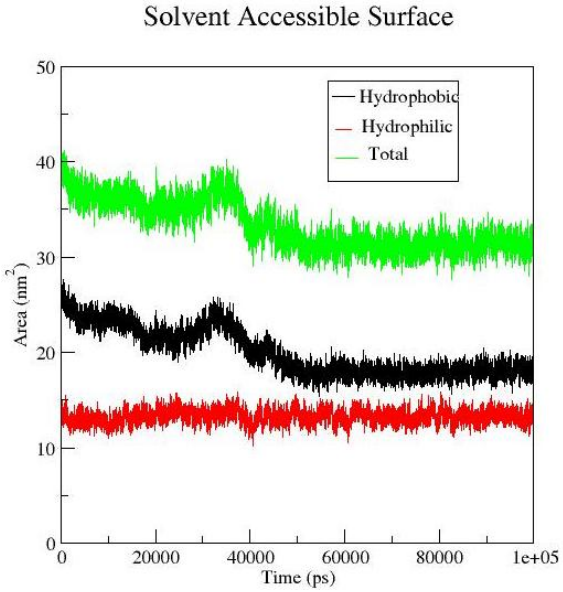
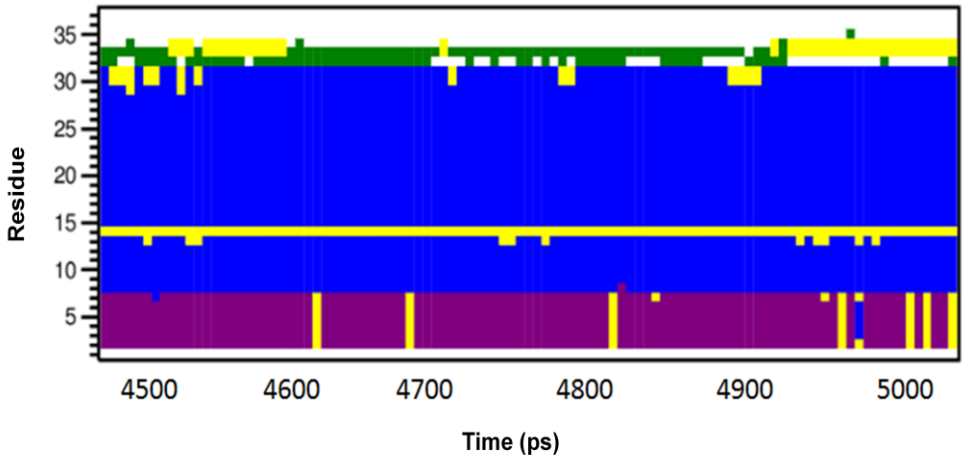


Figure 35: Plot of change in solvent accessible surface area over the trajectory for LL-37 in explicit solvent.

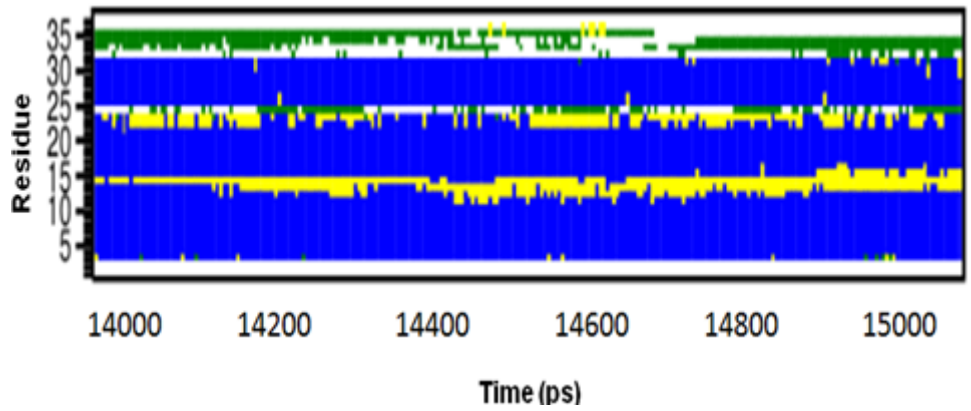
4.3.5 Secondary Structure Analysis using DSSP

To get a better understanding of the distribution of structures in the MD trajectory, on the basis of their RMSD patterns shown in **Figure 33**, the sampling efficiency of the MD simulations was monitored for the different secondary structure maps of the peptide LL-37. This was computed as a function of simulation time using DSSP (Kabsch and Sander, 1983) as illustrated in **Figure 36 (A-E).** The structure maps confirm that the secondary structure

conformations occurred throughout the trajectory. α- helical region (blue) is the most prevalent secondary structure that appeared around Gly^{3} – Lys^{12} and Lys25 – Leu^{31}. At the beginning of the trajectory around 4 ns to 5 ns we got π helix between residues Leu^{2}-Arg^{7} **(Figure 36 A).** But after 5 ns this region was classified as α-helix due to change in hydrogen bonding pattern between residues. At around 44-45 ns again π helix formed between residues Glu^{11}-Glu^{16} **(Figure 36 B).** At times in the graphs β-turn region (yellow) was also classified i.e., the backbone curvature is either above (bend) or below (turn) 70^{*} (Kabsch and Sander, 1983). Between residues Ser^{9}-Lys^{12} and Gly^{14}-Phe^{17}, around 44-45 ns and mostly beyond 55 ns β-turns were observed. This suggests that C- and N-terminal of the peptide exhibit flexibility such that end-to-end distance may vary. In most of the graphs we got bends that were responsible for the stabilization of the system. These mostly appeared in between the central residue and at the C terminal residues. From the start to the end of the trajectory α-helical conformation was maintained.



A



B

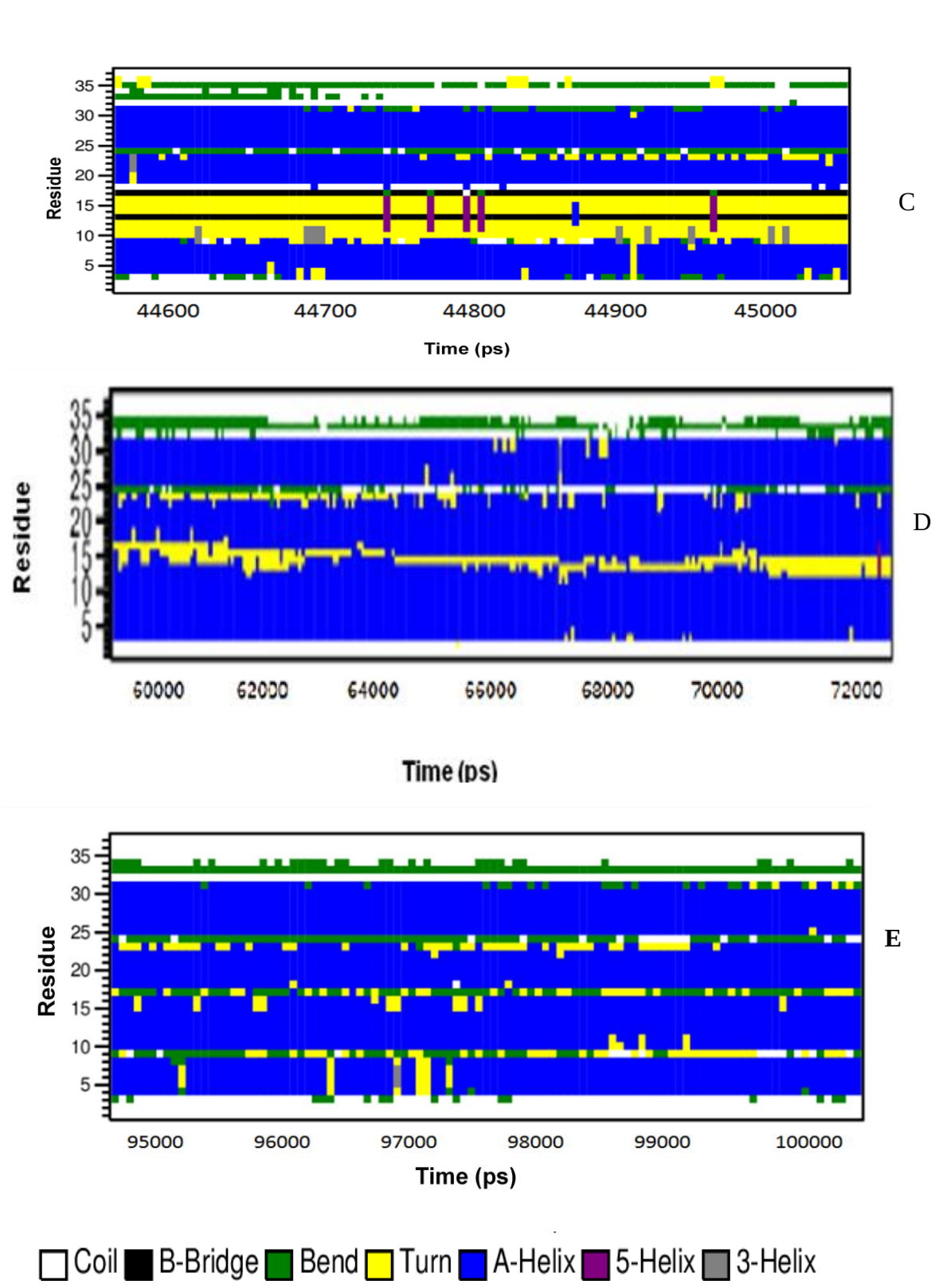
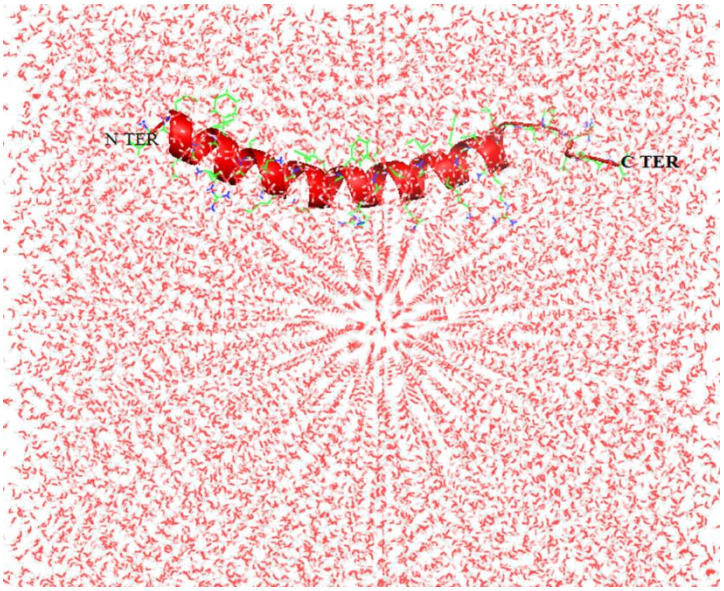


Figure 36: Peptide secondary structure per residue along the trajectories of peptide configurations. The secondary structure was computed with the program DSSP (Kabsch and Sander, 1983). **(A)** At 4-5ns **(B**) 14-15ns **(C)** 44-45 ns **(D)** 62-72ns **(E)** 95-100ns**.**

4.3.6 Structural Analysis

The peptide was enclosed in a water box and MD simulation was carried out to see the effect of water on the peptide. Whether the peptide remains stable or gets distorted which would give a clearer picture of the structural features of peptide conformations. Subsequently the average structures were obtained at different time interval using trajconv command, which are diagrammatically represented in **Figure 37**. The average structure shows only α- helix and turns, to some extent loops and π-helix. This clearly supports the results obtained from secondary structure analysis **(Figure 36).** The α-helical part remained stable for most length of the trajectory. The average structure obtained at around 10-18 ns shows a continuous helical, region between residues Leu^{2}-Leu^{31} **(Figure 37 B).** From cluster II i.e. ,at 20-30 ns the average structure exhibits two a-helical regions, one between the residues Gly^{3}-Arg^{23}, while the second between residues Lys^{25}-Leu^{31} **(Figure 37 C).** 30-40 ns trajectory also shows α-helical between residue Gly^{3}-Lys^{12} and Lys^{25}-Leu^{31} and to some extent π-helix in residues Lys^{18}-Arg^{23}. After 55 ns the peptide exhibited almost same type of conformation. There was not much change in the structure. It showed different α-helical regions between residues at the N- terminus (Asp^{4} - Lys^{8}, Lys^{10}-Glu^{16}). The second helical region flanked by residues (Lys^{25}-Leu^{31}) at the C-terminus region of peptide. Also π helix between residues (Lys^{18}-Arg^{23}) was reported. The rest of the C-terminal is flexible and extended.

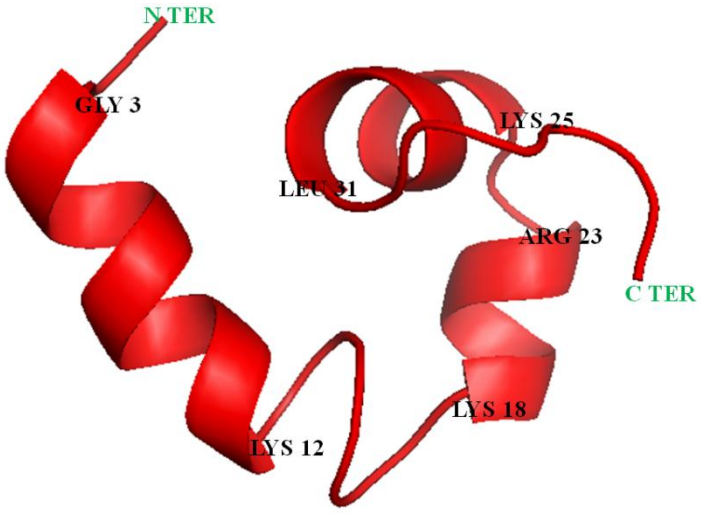
Overall, these results reveal that the peptide LL-37 has a stronger propensity to attain helical conformation at its N-terminal in the central region and extending towards the C-terminal end and thus showing a good agreement with the helical profile displayed by the NMR structure (2K6O) in the water solvent.


N TER
C TER

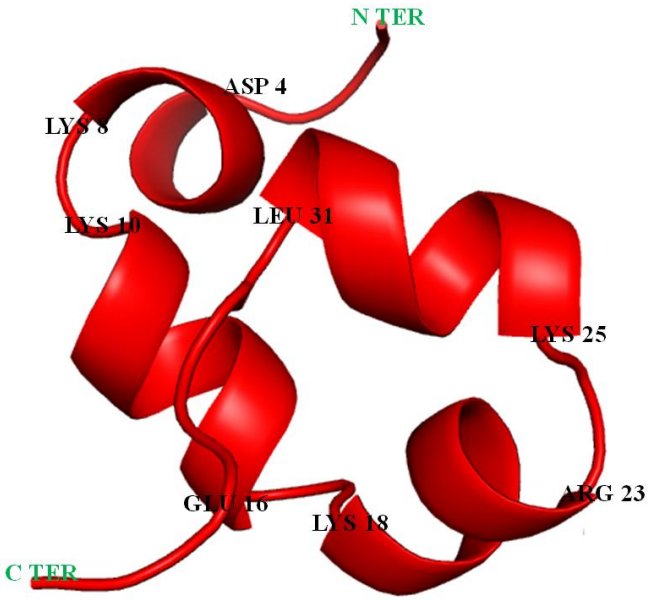
A


C-TER
N-TER
LEU 2
LEU 31

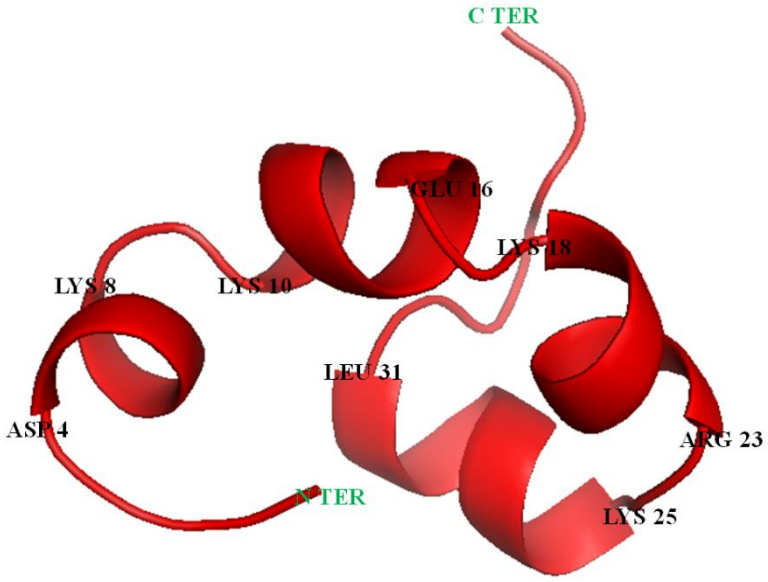
B


N TER
GLY 3
LYS 25
LEU 31
ARG 23
C TER
LYS 18
LYS 12

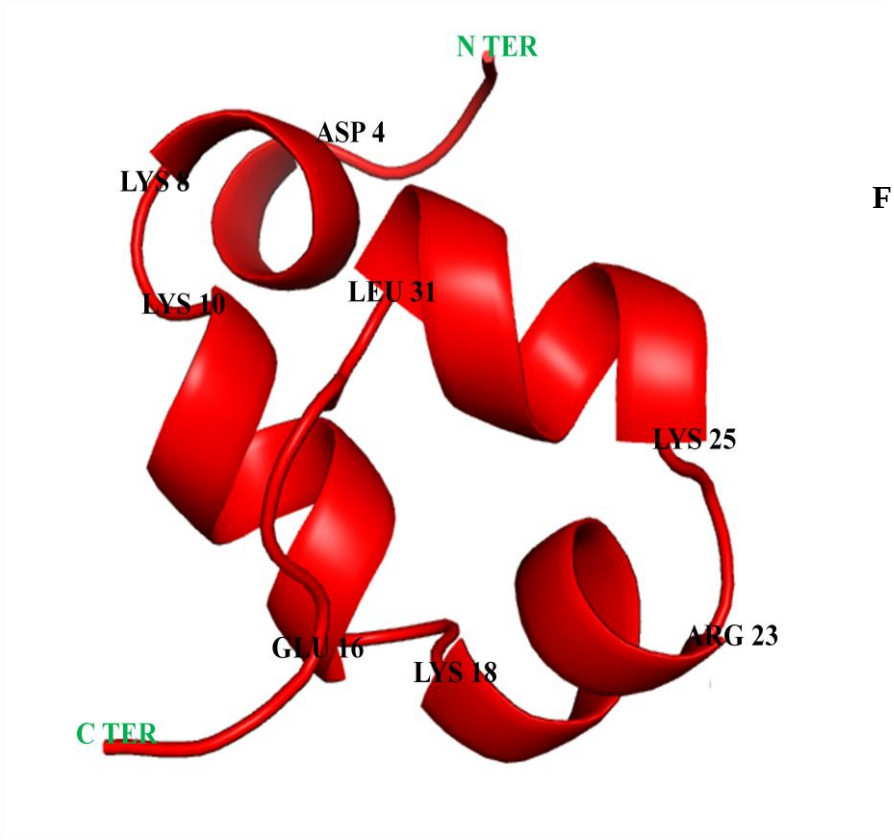
C


N TER
ASP 4
LYS 8
LYS 10
LEU 31
LYS 25
ARG 23
GLU 16
LYS 18
C TER

D


C TER
GLU 16
LYS 18
LYS 8
LYS 10
LEU 31
ASP 4
ARG 23
N TER
LYS 25

E



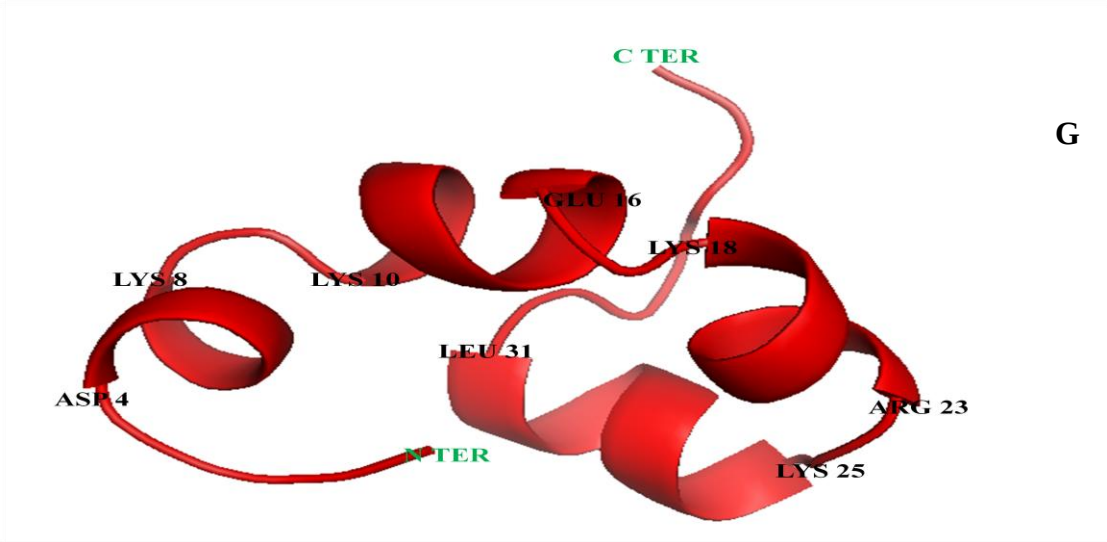


Figure 37: (A) Peptide in water box. Average structures of LL-37 at different time intervals **(B)** 5-15ns **(C)** 20-30ns **(D)** 40-50ns and **(E-G)** 55ns – 100ns

Chapter-5

Conclusion

The aim of this present work was to a get a deeper insight of the structure-activity relationships for antimicrobial peptide cathelicidin. For this purpose the conformational profiles the sole human cathelicidin LL-37 was explored by running 100 ns long molecular dynamics (MD) simulations using the OBC implementation of the GB method to describe the solvent. The results demonstrate that the LL-37 has a tendency to attain folded, unfolded and helical conformations, probably due to the low energy barrier between the states accounting for the high level of flexibility of the peptide. Analysis of the MD trajectories of LL-37 depicts the propensity of this peptide to attain α helices, β sheets and β-turns under the influence of force fields ff96, ff99 and ff99 SB. In MD trajectory of linear LL-37 under the influence ff96 force field, β sheet conformations were sampled. This has not been seen yet in case of LL-37 which otherwise is an alpha helical peptide as been reported in its NMR structure. β sheets have been reported in case of other antimicrobial peptides with a similar mechanism of action (L. Tomasinsig and M. Zanetti 2005) encouraging our belief that β sheets in case of cathelicidin is a novel structural feature having a role in its activity. On the other hand under AMBER ff99 force field the peptide attains a helical structure with the N and C terminals being extended and flexible. Hence this analysis of the trajectories obtained under ff96 and ff99 revealed an unprecedented dual conformational behaviour of this peptide with a tendency to exhibit an α-helical behaviour under ff99, in accordance with reported literature and β sheet conformation a new structural level under ff96.

MD simulation on the NMR structure of LL-37 using force fields AMBERff96 and AMBER ff99 SB different alpha-helical regions; between residues around the N- terminus, second in the central region (Ile^{20}-Ile^{24}) and the third flanked by residues around the C-terminus region of peptide. Also 3_{10} helix in the N-terminal, central and C-terminal region were reported. Even though force filed AMBER ff99SB is an advanced force field recommended to be used for peptides, ff96 in case of the NMR structure was efficient in sampling conformations throughout the trajectory thus ff96 force field, was adequate to be used in combination with implicit solvent methods as already been reported in literature (Zhou 2007) the ff96 set can reproduce accurately the results of explicit solvent simulations when used with implicit solvent models.

In case of MD simulation carried on the peptide enclosed in water box, water restricted the peptide hence less conformations were sampled.

Overall, on comparing of the results obtained in the present work using different force fields suggest that the majority of the motifs of LL-37 peptide adopt preferably folded helical conformations regardless of the computational protocol with the exception being force field ff96 in case of the extended structure.

Chapter- 6

Discussion and Future Perspective

MD simulations aim to characterize peptide folding require robust and advanced force fields so that they are able to successfully model the folded peptide starting from an extended conformation. The force fields ff96, ff96 and ff99SB in AMBER 12 have been used in computational simulation studies of peptides with appropriate GB models. The current simulation studies on the selected cathelicidin LL-37 indicate that the AMBER ff96 force field in combination with the Onufriev, Bashford and Case (OBC) implementation of the GB method is able to sample secondary structures much more efficiently even though much advanced versions of the same are available. Hence, this force field can be employed for future studies on the medium sized peptides with similar mechanism of action.

LL-37 has been known to play an important role in line of defense against local infection. Understanding the relationship between its structure and activity may assist in the development of lead compounds mimicking those of the human immune system. Results from this work contribute vital information regarding structure, function and folding dynamics of cathelicidin. The structures of LL-37 obtained from these trajectories can be used to explore and even design target molecules of these peptides applied as potential lead compounds in the field of computer-aided drug design and peptidomimetics.

The conformational profiles obtained can be studied to explore some anticancer properties of LL-37 correlating it with the secondary structure obtained.

Implicit solvent simulations have been used to simulate folding repertoire of peptides in cases where explicit solvents simulations are computationally not feasible due to longer timescales. However the next stage of this work involves an assessment of interaction of the cathelicidin model peptide with biomembranes, especially the lipid-bilayer. Since it has been known to target the lipid bilayer hence simulating it in a bilayer mimicking environment will be more accurate in terms of physiological relevance.

References

Adelman, S. A., and Doll, J.D (1976). "Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering Off Harmonic Solids, ." Journal of Chemical Physics **64**: 2375-2388.

Agerberth, B., H. Gunne, J. Odeberg, P. Kogner, H. G. Boman and G. H. Gudmundsson (1995). "FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis." Proc Natl Acad Sci U S A **92**(1): 195-199.

Agerberth, B., J. Y. Lee, T. Bergman, M. Carlquist, H. G. Boman, V. Mutt and H. Jornvall (1991). "Amino acid sequence of PR-39. Isolation from pig intestine of a new member of the family of proline-arginine-rich antibacterial peptides." Eur J Biochem **202**(3): 849-854.

Ahmad, A., N. Asthana, S. Azmi, R. M. Srivastava, B. K. Pandey, V. Yadav and J. K. Ghosh (2009). "Structure-function study of cathelicidin-derived bovine antimicrobial peptide BMAP-28: design of its cell-selective analogs by amino acid substitutions in the heptad repeat sequences." Biochim Biophys Acta **1788**(11): 2411-2420.

Alder BJ, W. T. (1959). "Studies in Molecular Dynamics. I. General Method." The Journal of Chemical Physics. **2**(31): 459–466.

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.

Balamurugan, K., R. Gopalakrishnan, S. S. Raman and V. Subramanian (2010). "Exploring the changes in the structure of alpha-helical peptides adsorbed onto a single walled carbon nanotube using classical molecular dynamics simulation." J Phys Chem B **114**(44): 14048-14058.

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

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

Bals, R., D. J. Weiner, A. D. Moscioni, R. L. Meegalla and J. M. Wilson (1999). "Augmentation of innate host defense by expression of a cathelicidin antimicrobial peptide." Infect Immun **67**(11): 6084-6089.

Benachour, H., M. Zaiou, A. Samara, B. Herbeth, M. Pfister, D. Lambert, G. Siest and S. Visvikis-Siest (2009). "Association of human cathelicidin (hCAP-18/LL-37) gene expression with cardiovascular disease risk factors." Nutr Metab Cardiovasc Dis **19**(10): 720-728.

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

Berendsen, H. J. C. (1984). "s Molecular-Dynamics with Coupling to an External Bath." Journal of Chemical Physics.

Bucki, R., K. Leszczynska, A. Namiot and W. Sokolowski (2010). "Cathelicidin LL-37: a multitask antimicrobial peptide." Arch Immunol Ther Exp (Warsz) **58**(1): 15-25.

Buehler, L. K. (2009). "An introduction to molecular interaction in biological system." http://www.whatislife.com/reader/interaction-reader.html.

Buehler, L. K. (2009). An introduction to molecular interaction in biological systems.

Caflisch, A. and M. Karplus (1994). "Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase." Proc Natl Acad Sci U S A **91**(5): 1746-1750.

Calero, S., S. Lago, W. F. van Gunsteren and X. Daura (2004). "Modelling of the complex between a 15-residue peptide from mSos2 and the N-terminal SH3 domain of Grb2 by molecular-dynamics simulation." Chem Biodivers **1**(3): 505-519.

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

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

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

Corcho F, C. J., Perez JJ. (2004). "Comparative Analysis of theConformational Profile of Substance P Using Simulated Annealing and Molecular Dynamics." J. Comput. Chem **25**: 1937–1952.

Daidone I, D. A. M., Di Nola A, Amadei (2005). "Theoretical characterization of alpha-helix and beta-hairpin folding kinetics.

." J Am Chem Soc**;127(42)**: :14825-14832.

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

Ganz, T. and R. I. Lehrer (1999). "Antibiotic peptides from higher eukaryotes: biology and applications." Mol Med Today **5**(7): 292-297.

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

Godzi, M. G., A. P. Tolstova and I. V. Oferkin (2010). "[Application of molecular dynamics simulation to the interpretation of atomic force microscopy data]." Biofizika **55**(3): 415-423.

H. J. C. Berendsen, J. R. G., and T. P. Straatsma (1987). "The missing term in effective pair potentials." Journal of Physical Chemistry **91 pp**: 6269 - 6271.

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

https://lafarga.cpl.upc.edu/projects/clusterit., a. L. C. P. I. A. a. and F. (b) Corcho, Canto, J., and Perez, J.J. 2004. " Comparative analysis of the conformational profile of substance P using simulated annealing and molecular dynamics " Journal of Computational Chemistry **25** 1937-1952.

Huang, L. L., L. Z. Zhang, Q. Shao, J. Wang, L. H. Lu, X. H. Lu, S. Y. Jiang and W. F. Shen (2006). "Molecular dynamics simulation study of the structural characteristics of water molecules confined in functionalized carbon nanotubes." J Phys Chem B **110**(51): 25761-25768.

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

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

Karplus M, M. J. (2002). "Molecular dynamics simulations of biomolecules." Nat. Struct. Biol **9**: 646–652.

Koehl P, D. M. (1994). "Polar and nonpolar atomic environments in the protein core: implications for folding and binding." Proteins
20(3): 264–278.

Kollman, P. A., Dixon, R., Cornell, W., Fox, T., Chipot, C., and Pohorille, A. (1997). " The development/application of a 'minimalist' organic/biochemical molecular mechanics force

field using a combination of ab initio calculations and experimental data." Computer Simulation of Biomolecular Systems **3**: 83–96.

Laugwitz, K. L., H. J. Weig, A. Moretti, E. Hoffmann, P. Ueblacker, I. Pragst, K. Rosport, A. Schomig and M. Ungerer (2001). "Gene transfer of heterologous G protein-coupled receptors to cardiomyocytes: differential effects on contractility." Circ Res **88**(7): 688-695.

Leach (2001). "Safe detection and retrieval of drugs concealed internally." J Clin Forensic Med.: 169-170.

Lee, P. H., T. Ohtake, M. Zaiou, M. Murakami, J. A. Rudisill, K. H. Lin and R. L. Gallo (2005). "Expression of an additional cathelicidin antimicrobial peptide protects against bacterial skin infection." Proc Natl Acad Sci U S A **102**(10): 3750-3755.

Lemkul, J. "GROMACS Tutorial Lysozyme in Water." Department of Biochemistry, Virginia Tech.

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

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

Li, X., Y. Li, H. Han, D. W. Miller and G. Wang (2006). "Solution structures of human LL-37 fragments and NMR-based identification of a minimal membrane-targeting antimicrobial and anticancer region." J Am Chem Soc **128**(17): 5776-5785.

Liang, J. F. and S. C. Kim (1999). "Not only the nature of peptide but also the characteristics of cell membrane determine the antimicrobial mechanism of a peptide." J Pept Res **53**(5): 518-522.

Lindahl E, H. B., Van Der Spoel D "GROMACS 3.0: a package for molecular simulation and trajectory analysis." Journal of Molecular Modeling. **7(8)**: 306-317.

Lindahl, E. d. " Gromacs User Manual. http://www.gromacs.org ".

Lohner, K., A. Latal, G. Degovics and P. Garidel (2001). "Packing characteristics of a model system mimicking cytoplasmic bacterial membranes." Chem Phys Lipids **111**(2): 177-192.

manual, g. (2010). "^ a b c "About Gromacs".".

Marx, D. a. H., J. 2000. Ab initio molecular dynamics: Theory and Implementation. (2000). "Ab initio molecular dynamics: Theory and Implementation." Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, Julich, NIC Series **1**: 301-449.

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

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

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

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

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

Porcelli, F., R. Verardi, L. Shi, K. A. Henzler-Wildman, A. Ramamoorthy and G. Veglia (2008). "NMR structure of the cathelicidin-derived human antimicrobial peptide LL-37 in dodecylphosphocholine micelles." Biochemistry **47**(20): 5565-5572.

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

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

Rahman (1964). "Correlations in the Motion of Atoms in Liquid Argon." Phys. Rev. 1 **136(2A):A405.**

Rao, M. N., A. E. Shinnar, L. A. Noecker, T. L. Chao, B. Feibush, B. Snyder, I. Sharkansky, A. Sarkahian, X. Zhang, S. R. Jones, W. A. Kinney and M. Zasloff (2000). "Aminosterols from the dogfish shark Squalus acanthias." J Nat Prod **63**(5): 631-635.

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

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

Sorensen, O. E., L. Gram, A. H. Johnsen, E. Andersson, S. Bangsboll, G. S. Tjabringa, P. S. Hiemstra, J. Malm, A. Egesten and N. Borregaard (2003). "Processing of seminal plasma hCAP-18 to ALL-38 by gastricsin: a novel mechanism of generating antimicrobial peptides in vagina." J Biol Chem **278**(31): 28540-28546.

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

Still, W. C., Tempczyk, A., Hawley, R.C. and Hendrickson, T. (1990). "Semianalytical treatment of solvation for molecular mechanics and dynamics." Journal of American Chemical Society **12**: 6127-6129.

Subbalakshmi, C. and N. Sitaram (1998). "Mechanism of antimicrobial action of indolicidin." FEMS Microbiol Lett **160**(1): 91-96.

Sugita Y., a. O. Y. R.-e. m. d. m. f. p. f., Chemical Physics Letters, (314): 141-151 (1999). "Replica-exchange molecular dynamics method for protein folding." Chemical Physics Letters, **314**: 141-151

Swaroop Chatterjee, P. G. D., Frank H. Stillinger, and Ruth M. Lynden-Bell (2008). "A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models." Journal of Chemical Physics **128**: 12451.

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

Thukral L, S. J., Daidone I. (2009). "Common folding mechanism of a beta-hairpin peptide via non-native turn formation revealed by unbiased molecular dynamics simulations." J Am Chem Soc. **131(50)**: 18147-18152.

Tomasinsig, L., C. Pizzirani, B. Skerlavaj, P. Pellegatti, S. Gulinelli, A. Tossi, F. Di Virgilio and M. Zanetti (2008). "The human cathelicidin LL-37 modulates the activities of the P2X7 receptor in a structure-dependent manner." J Biol Chem **283**(45): 30471-30481.

Tossi, A., I. Bonin, N. Antcheva, S. Norbedo, F. Benedetti, S. Miertus, A. C. Nair, T. Maliar, F. Dal Bello, G. Palu and D. Romeo (2000). "Aspartic protease inhibitors. An integrated approach for the design andsynthesis of diaminodiol-based peptidomimetics." Eur J Biochem **267**(6): 1715-1722.

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

Viacarra C, M. S. (2005). "Electrostatics in computational protein design." Curr Opin Chem Biol **9(6)**: 622–626.

Wang, G. (2010). "Structure, dynamics and mapping of membrane-binding residues of micelle-bound antimicrobial peptides by natural abundance (13)C NMR spectroscopy." Biochim Biophys Acta **1798**(2): 114-121.

Wickstrom, L. O., A. (2009). "Simmerling, C. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. ." Biophys. J., **97**: 853–856.

Xia, B., V. Tsui, D. A. Case, H. J. Dyson and P. E. Wright (2002). "Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water." J Biomol NMR **22**(4): 317-331.

Yamasaki, K., J. Schauber, A. Coda, H. Lin, R. A. Dorschner, N. M. Schechter, C. Bonnart, P. Descargues, A. Hovnanian and R. L. Gallo (2006). "Kallikrein-mediated proteolysis regulates the antimicrobial effects of cathelicidins in skin." FASEB J **20**(12): 2068-2080.

Yang, D., O. Chertov and J. J. Oppenheim (2001). "Participation of mammalian defensins and cathelicidins in anti-microbial immunity: receptors and activities of human defensins and cathelicidin (LL-37)." J Leukoc Biol **69**(5): 691-697.

Zanetti, M., G. Del Sal, P. Storici, C. Schneider and D. Romeo (1993). "The cDNA of the neutrophil antibiotic Bac5 predicts a pro-sequence homologous to a cysteine proteinase inhibitor that is common to other neutrophil antibiotics." J Biol Chem **268**(1): 522-526.

Zanetti, M., R. Gennaro and D. Romeo (1995). "Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain." FEBS Lett **374**(1): 1-5.

Zhou, R. (2007). "Replica exchange molecular dynamics method for protein folding simulation." Methods Mol Biol **350**: 205-223.

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

Appendix 1

1. AMBER input files

(i)tLEaP Input file

```
source leaprc.ff99SB
cathlin = sequence {ACE LEU LEU GLY ASP PHE PHE ARG LYS SER LYS
GLU LYS ILE GLY LYS GLU PHE LYS ARG ILE VAL GLN ARG ILE LYS
ASP PHE LEU ARG ASN LEU VAL PRO ARG THR GLU SER NME}
set default PBradii mbondi2
saveamberparm cathlin  cathlin.prmtop cathlin.prmcrd
quit
/
```

(ii). Minimization Input file

```
 min.in minimization  &cntrl
  imin=1, maxcyc=100,
  cut=300.0, igb=2, saltcon=0.2, gbsa=1,
  ntpr=10, ntx=1, ntb=0,
 &end
 /
```

(iii) Heat Input file

```
 heat.in Stage 1 heating of cath  0 to 300K  &cntrl
  imin=0, irest=0, ntx=1,
 nstlim=500000, dt=0.002,
ntc=2, ntf=2,
ntt=3, tautp=1.0, gamma_ln=1,
tempi=0.0, temp0=300.0,
ntpr=500, ntwx=500,
ntb=0, igb=2, cut=12,
 /
```

(iv) MD production run Input file

```
 md.in Stage 2 md of cath at 100ns &cntrl
imin=0, irest=1, ntx=5,
nstlim=50000000, dt=0.002,
ntc=2, ntf=2,
ntt=3, tautp=1.0, gamma_ln=1,
tempi=300.0, temp0=300.0,
ntpr=500, ntwx=500, ntwe=500, nscm=1000,
ntb=0, igb=2, saltcon=0.2, gbsa=1, cut=12,
```

2. GROMACS Input Files

(i). em.mdp : Parameters describing what to do, when to stop and what to save

```
constraint      = none
integrator      = steep        ; Algorithm (steep = steepest descent minimization)
emtol           = 1000.0       ; Stop minimization when the maximum force < 1000.0
kJ/mol/nm
emstep          = 0.01
dt              = 0.002
nsteps          = 50000        ; Maximum number of (minimization) steps to perform
nstlist         = 1            ; Frequency to update the neighbor list and long range
forces
ns_type         = grid         ; Method to determine neighbor list (simple, grid)
rlist           = 1.0          ; Cut-off for making neighbor list (short range forces)
coulombtype     = PME          ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0          ; Short-range electrostatic cut-off
rvdw            = 1.0          ; Short-range Van der Waals cut-off
pbc             = xyz          ; Periodic Boundary Conditions (yes/no)
~
```

(ii) md.mdp: MD Run parameters

```
integrator      = md           ; leap-frog integrator
nsteps          = 50000000     ; 2 * 50000000 = 10000 ps, 100 ns
dt              = 0.002        ; 2 fs
; Output control
nstxout         = 1000         ; save coordinates every 2 ps
nstvout         = 1000         ; save velocities every 2 ps
nstxtcout       = 1000         ; xtc compressed trajectory output every 2 ps
nstenergy       = 1000         ; save energies every 2 ps
nstlog          = 1000         ; update log file every 2 ps
; Bond parameters
continuation    = yes          ; Restarting after NPT
constraint_algorithm = lincs   ; holonomic constraints
constraints     = all-bonds    ; all bonds (even heavy atom-H bonds)
constrained
lincs_iter      = 1            ; accuracy of LINCS
lincs_order     = 4            ; also related to accuracy
; Neighborsearching
ns_type         = grid         ; search neighboring grid cells
nstlist         = 5            ; 10 fs
rlist           = 1.0          ; short-range neighborlist cutoff (in nm)
rcoulomb        = 1.0          ; short-range electrostatic cutoff (in nm)
rvdw            = 1.0          ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype     = PME          ; Particle Mesh Ewald for long-range
electrostatics
pme_order       = 4            ; cubic interpolation
fourierspacing  = 0.16         ; grid spacing for FFT
; Temperature coupling is on
tcoupl          = V-rescale    ; modified Berendsen thermostat
```

```
tc-grps                 = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                   = 0.1   0.1     ; time constant, in ps
ref_t                   = 300   300     ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                  = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype              = isotropic     ; uniform scaling of box vectors
tau_p                   = 2.0           ; time constant, in ps
ref_p                   = 1.0           ; reference pressure, in bar
compressibility         = 4.5e-5        ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc                     = xyz           ; 3-D PBC
; Dispersion correction
DispCorr                = EnerPres      ; account for cut-off vdW scheme
; Velocity generation
gen_vel                 = no            ; Velocity generation is off
/
```