REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
12-12-2014	MS Thesis	-

4. TITLE AND SUBTITLE
Molecular Dynamics Analysis of Lysozyme Proten in Ethanol-Water Mixed Solvent

5a. CONTRACT NUMBER
W911NF-11-1-0168

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
633002

6. AUTHORS
Henry Ochije

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
North Carolina A&T State University
1601 East Market Street

Greensboro, NC 27411 -0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
60143-ST-H.14

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Effect of protein-solvent interaction on the protein structure is widely studied with experimental and computational techniques. Despite this, molecular level understanding of proteins and some simple solvents is still not fully understood. This work focuses on detailed molecular dynamics simulations of solvent effect on lysozyme protein, using water, ethanol, and different concentrations of water-ethanol mixtures as solvents. The lysozyme protein structure in water, ethanol and ethanol-water mixture (0–12% ethanol) was studied using GROMACS molecular dynamics simulation (MD) code. Compared to water environment, the lysozyme structure showed remarkable

15. SUBJECT TERMS
lysozyme protein, molecular dynamics, binary solvent

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ajit Kelkar
UU	UU	UU	UU		19b. TELEPHONE NUMBER 336-285-2864

Report Title

Molecular Dynamics Analysis of Lysozyme Proten in Ethanol-Water Mixed Solvent

ABSTRACT

Effect of protein-solvent interaction on the protein structure is widely studied with experimental and computational techniques. Despite this, molecular level understanding of proteins and some simple solvents is still not fully understood. This work focuses on detailed molecular dynamics simulations of solvent effect on lysozyme protein, using water, ethanol, and different concentrations of water-ethanol mixtures as solvents. The lysozyme protein structure in water, ethanol and ethanol-water mixture (0–12% ethanol) was studied using GROMACS molecular dynamics simulation (MD) code. Compared to water environment, the lysozyme structure showed remarkable changes in water-ethanol with increasing ethanol concentration. Significant changes were observed in the protein secondary structure involving alpha helices. A study of thermodynamic and structural properties indicate that increasing ethanol concentration resulted in a systematic increase in total energy, enthalpy, root mean square deviation (RMSD), and radius of gyration of lysozyme. A polynomial interpolation approach is presented to determine these quantities for any intermediate alcohol percentage, and compared with the values obtained from a full MD simulation. Results from MD simulation were in reasonably good agreement with those from the interpolation approach. The polynomial approach eliminates the need for computationally intensive full MD analysis for the concentrations within the range of 0 - 12%

Molecular Dynamics Analysis of Lysozyme Protein in Ethanol-Water Mixed Solvent Environment

Henry Ikechukwu Ochije

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department: Nanoengineering

Major: Nanoengineering

Major Professor: Dr. Ram Mohan

Greensboro, North Carolina

2012

School of Graduate Studies
North Carolina Agricultural and Technical State University
This is to certify that the Master's Thesis of

Henry Ikechukwu Ochije

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2012

Approved by:

Dr. Ram Mohan Major Professor	Dr. Ajit Kelkar Committee Member
Dr. Goundla Srinivas Committee Member	Dr. Ajit Kelkar Department Chair

Dr.Sanjiv Sarin
Associate Vice Chancellor for
Research and Graduate Dean

Biographical Sketch

Henry Ikechukwu Ochije was born and raised up in Lagos state Nigeria. He received his Bachelor of Science degree in Physics from University of Lagos Akoka and holds Masters of Science from North Carolina Central University, Durham, He is currently a candidate for the Master of Science degree in Nanoengineering.

Dedication

I dedicate this thesis to God Almighty who has given me the strength to go this far, the completion of this thesis was by his grace; I was able to make it through this program with limited time on my side.

Acknowledgement

I want to acknowledge and thank my advisor Dr Ram Mohan and Dr. Goundla Srinivas who have mentored me and taught me what I need to know. Special thanks go to my wife Sochima Ochije and my son David Ochije for their moral support and understanding through this period. I would also want to appreciate my parents for the sacrifice of time and resources towards my academic pursuit.

Table of Contents

List of Figures

List of Tables

Abstract

Effect of protein-solvent interaction on the protein structure is widely studied with experimental and computational techniques. Despite this, molecular level understanding of proteins and some simple solvents is still not fully understood. This work focuses on detailed molecular dynamics simulations of solvent effect on lysozyme protein, using water, ethanol, and different concentrations of water-ethanol mixtures as solvents. The lysozyme protein structure in water, ethanol and ethanol-water mixture (0–12% ethanol) was studied using GROMACS molecular dynamics simulation (MD) code. Compared to water environment, the lysozyme structure showed remarkable changes in water-ethanol with increasing ethanol concentration. Significant changes were observed in the protein secondary structure involving alpha helices. A study of thermodynamic and structural properties indicate that increasing ethanol concentration resulted in a systematic increase in total energy, enthalpy, root mean square deviation (RMSD), and radius of gyration of lysozyme. A polynomial interpolation approach is presented to determine these quantities for any intermediate alcohol percentage, and compared with the values obtained from a full MD simulation. Results from MD simulation were in reasonably good agreement with those from the interpolation approach. The polynomial approach eliminates the need for computationally intensive full MD analysis for the concentrations within the range (0-12%) studied.

CHAPTER 1

Introduction

Protein molecules are very important in the formation of most biological building blocks in humans and animals. Proteins perform various functions in biological systems, the structure and the type of protein determines its function. For example, proteins like collagen has a structure that will support the cell structure due to its coiled helical shape which is long, strong and stringy (Branden & Tooze, 1991). Another example is seen with hemoglobin, a globular protein having a folded and compact shape which can maneuver through the blood vessel providing a function to supply oxygen to the body cells(Dunker, Brown, Lawson, Iakoucheva, & Obradovic, 2002). Several experimental, theoretical and computational studies focused on understanding the structure-function relationship of proteins.

Solvents influence the protein structure to a great extent. For example, a protein adapts dissimilar structure in a hydrophobic solvent compared to that in water. Despite numerous experimental and simulation studies, behavior of proteins in simple mixed solvents such as ethanol -water mixtures has not been fully understood from a molecular viewpoint. In this work, we conduct extensive molecular dynamics simulation studies of a lysozyme protein in ethanol-water mixture at various concentrations of ethanol in water to understand the solvent influence on protein structure and dynamics. For this purpose we use the well-known GROMACS molecular dynamics (MD) simulation analysis code.

Even with several perspectives and results from laboratory experiments on different kinds of proteins, it is also important to have a visual understanding of the dynamics of protein structure and changes in different solvent environments. This requires understanding of the dynamics at the molecular level which can be achieved with molecular dynamics modeling

simulation tools such as Groningen Machine for Chemical Simulation (GROMACS). Such modeling analysis will facilitate further understanding of the nature of protein and the effect of different solvent systems and conditions on protein molecules such as lysozyme.

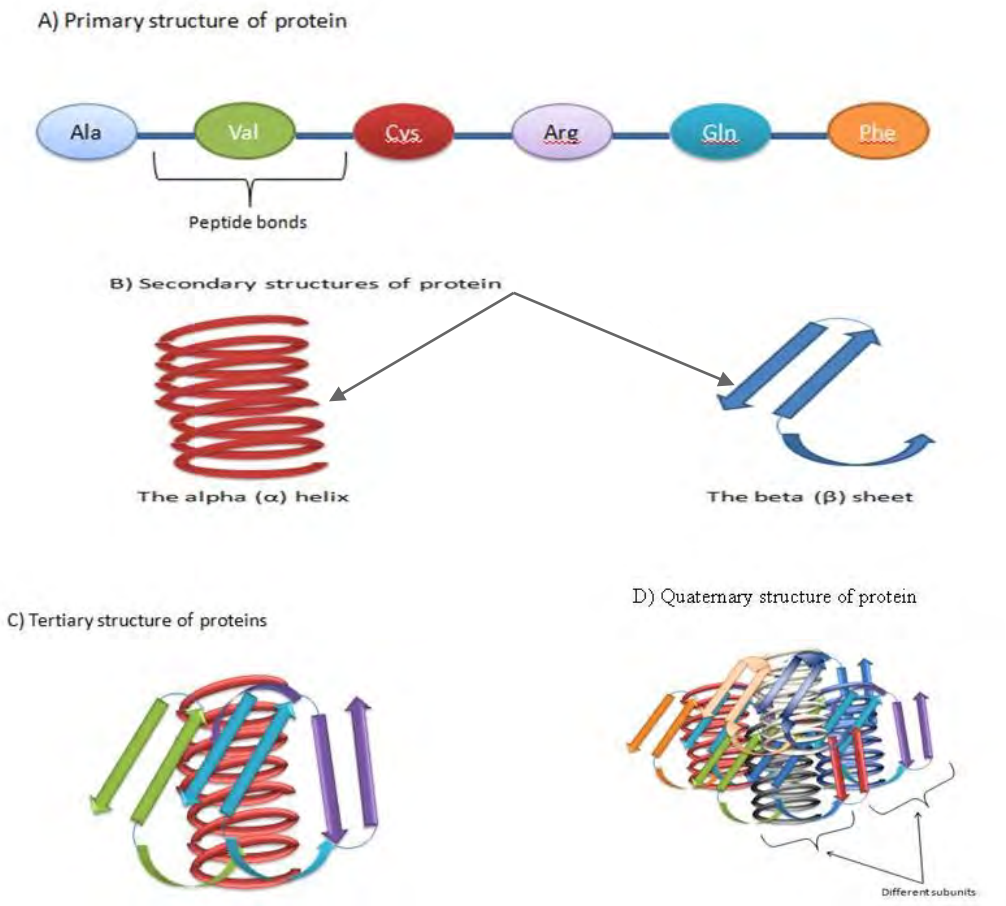

Figure 1.Schematic of the protein structures. (A) Primary structure, (B) Secondary structure, (C) Tertiary structure, (D) Quaternary structure.

Protein is a complex organic compound consisting of amino acids connected to each other by peptide bonds having a high molecular weight. There are 20 known amino acids and different combination of these amino acids forms the four main structures of a protein (figure 1). They are the primary structure, secondary structure, tertiary structure and the quaternary structure (Richardson, 1981). The primary structure is the linear amino acid sequence linked to

each other by a peptide bond, the secondary structure is the local conformation of the primary polypeptide chain stabilized by a hydrogen bond, the tertiary structure is the global folding of the protein chain by the combination of different secondary structures which is three dimensional in its form, and the quaternary structure is the arrangement of multiple folded protein forming a multi-subunit (Richardson, 1981). The interaction of protein molecule without covalent bond forming a stable oligomer forms the quaternary structure(Richardson, 1981). Figure 1 presents the schematic of the four structures of a protein. In this study we consider a protein, Lysozyme, also called muramidase or N-acetylmuramide glycanhyrolase (Rubio & Befrits, 2009) is an enzyme (which is also a protein) that functions to destroy the cell walls of bacteria. In humans it is encoded by the LYZ gene. The enzymatic activity of lysozyme is seen optimally at a pH of 5.2 and can decrease either above or below this pH value. Lysozyme contains two amino acid residues in the active site that is essential for catalysis: Glu^{35} and Asp^{52}.A value of equal pH at which equal basic and acidic concentration is present is known as the *pK* value. The *pK* value for the carboxyl side chains of these two residues are 5.9 and 4.5, respectively. For the lysozyme to be active Asp^{52} is deprotonated and Glu^{35} is protonated. When the pH is above 5.9, Glu^{35} is deprotonated and when it is below 4.5, Asp^{52} becomes protonated. This pH- activity profile of the two key amino residues decreases the activity of lysozyme and explains the ionization states of these two amino acid residues and the pH-activity profile of lysozyme (Lehninger, Nelson, & Cox, 2005).

Figure 2 presents the cartoon representation of lysozyme protein obtained from the protein data bank file 1AKI.pdb. We chose lysozyme because it has almost the same structure as that of the proteins found in humans. Secondly it is moderate in size for molecular dynamics simulation analysis compared to most other proteins.

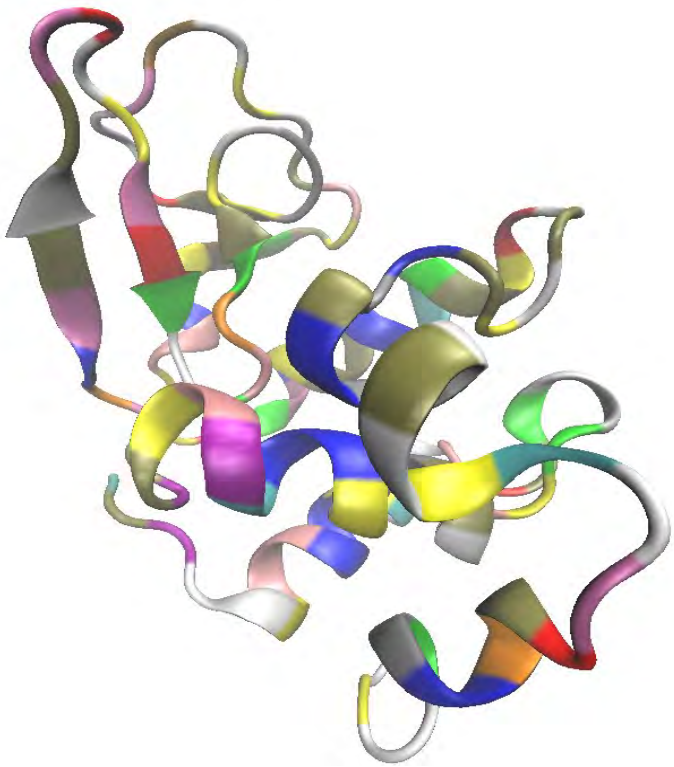

Figure 2. A cartoon depiction of lysozyme protein structure obtained from protein data bank file 1AKI.pdb.

Water plays an important role in maintaining cell membrane and enzyme activities acting as lubricant for protein movements in cells. More importantly, proteins need water to function and are their natural environment. Hence, we started with simulating and understanding the behavior of lysozyme in a water solvent environment. This is followed by the simulation and understanding of lysozyme in an ethanol-water environment at different ethanol-water ratios.

Ethanol is an organic compound that has the hydroxyl functional group (-OH) bonded to a carbon atom. The center of the bond is saturated with a single bond to three other atoms. The general formula for ethanol is given as C_2H_5OH (Burkhard & Dennison, 1951). It is colorless, volatile and flammable liquid. The physical properties of ethanol are from its hydroxyl group and its short carbon chain. The hydroxyl group can bond with hydrogen to make it less volatile and

more viscous than lower polar organic compound with almost same molecular weight for example, propane. It is miscible with water and some other organic compounds.

The main focus of this work is to study and understand the effect of different solvent environments (ethanol-water mix) on the lysozyme protein. We anticipate an understanding of the effect of ethanol on lysozyme could provide insights into similar other protein-solvent interactions as well.

1.1 Effect of water on protein and the effect of ethanol on proteins

Several experimental studies have been carried out to study the effect of ethanol on protein and the effect of water on protein with tangible results (Preedy & Peters, 1988; TIERNAN & WARD, 1986; Wang, 1954; Zaks & Klibanov, 1988). However, the effect of ethanol and ethanol-water mixtures on lysozyme protein has not been studied computationally in detail. In this work, we explore the effect of ethanol on lysozyme by carrying out detailed atomistic molecular dynamics simulation studies.

1.2 Molecular Dynamics Simulation

Molecular dynamic simulation provides an in depth understanding at a molecular level and has a potential to bridge the gap between experimental results and theory. It will also help us understand protein stability, conformational changes in different solvent environments, protein folding and its dynamics in biological system. Molecular dynamics simulation also provides both visual and dynamic effect of the environments on the trajectories of protein molecule at the atomic/molecular level. In addition, this would also provide a molecular level understanding of the biological proteins in ethanol-water solvent environments, a potential pathway to understand the biological effect of human alcohol consumption.

The remainder of this thesis is organized as follows.

- Chapter 2 provides the details of the existing literature and background on the molecular dynamic modeling methodology employed.
- Chapter 3 discusses the molecular dynamics analysis of lysozyme protein in water environment that corresponds to the natural solvent state of the protein.
- Chapter 4 discusses the molecular dynamics analysis of lysozyme protein in a full ethanol environment.
- Chapter 5 presents and discusses the molecular dynamics analysis of lysozyme protein in an ethanol – water mixture. Different ethanol percentages ranging from approximately 2% – 12% are analyzed. Each ethanol percentage requires a separate computationally intensive MD analysis. An interpolation approach for key post-analysis parameters employing the computationally obtained values from full dynamic analysis is presented. The results of the key post-analysis parameters based on the full molecular dynamics analysis is compared with the values from the interpolation analysis. Results show an excellent correlation between the values obtained from interpolation and the full molecular dynamics simulations.
- Chapter 6 summarizes the results and conclusions from the current study and potential for future directions.

CHAPTER 2

Literature Review

The structure of protein plays an important role in biological functions. Hunt *et. al*.(Hunt, Kattner, Shanks, & Wynne, 2007) studied the dynamics of water-protein interaction using ultrafast optical Kerr-effect spectroscopy which measures third-order polarizability anisotropy of liquids. This method was used to learn the dynamics behind the hydrogen bonds that can cause remarkable changes in the liquid system (Hunt et al., 2007). In this work, experiments determined the absorption band on the secondary structure of the protein as a function of methanol concentration. A rise in methanol concentration causes a decrease in the strength of the hydrogen bonds between the peptide backbone and the solvent (Hunt et al., 2007). The formation of alpha helix is associated to the solvent-peptide intermolecular hydrogen bonding, and hydration near peptide aids the dynamics of the solvation shell serving as lubricant. This plays a key role in determining the structure and activities of protein(Hunt et al., 2007).

Computationally, Ping and his coworkers studied the influence of water on protein formation using molecular dynamic simulations (Yanni, 2012). A TIP3P water model and six ring water cluster model with lysozyme was used in the investigation with a focus on the molecular dynamics modeling of water and protein. Also in addition the radial distribution function of solvent around lysozyme was evaluated. It was found that the distribution of water molecule around lysozyme is similar to that of water cluster(Yanni, 2012). They also observed that the four pairs of disulfide bonds of lysozyme was broken in lysozyme water cluster and the tertiary structure of lysozyme was destroyed as the temperature increased to 80^{o} C that also affected its function (Yanni, 2012).

In addition to temperature, as described above, protein structure can also be influenced by the solvent surrounding the protein. For example, hydrophilic solvents relatively stabilize protein structure while the hydrophobic solvents such as oil, are known to destabilize protein structure. In the present work, we aim to study lysozyme protein structure and dynamics in varying solvent conditions, in particular, ethanol-water mixtures. Wensink *et al* (Wensink, Hoffmann, van Maaren, & van der Spoel, 2003b) studied binary mixtures of alcohol and water using molecular dynamic simulation. They computed the shear viscosity using non-equilibrium molecular dynamic simulation. The diffusion constant was studied along with the rotational correlation time, and was found that mobility correlates with viscosity data, i.e. the viscosity is maximal at intermediate alcohol concentration(Wensink et al., 2003b). It was found that at maximal viscosity, mobility was minimal. They combined viscosity and diffusion calculations to compute the effective hydrodynamic radius of the particles in the mixture using Stoke - Einstein relation (Wensink, Hoffmann, van Maaren, & van der Spoel, 2003a). The analysis indicated that there is no collective diffusion of molecular clusters in the mixture and pure liquid.

The present work examines lysozyme protein structure and dynamics in various alcohol-water mixtures by performing series of molecular dynamics simulations at different alcohol-water solvent ratios. A brief background of the molecular dynamics analysis is presented next for completeness.

2.1 Molecular Dynamics Simulations

The concept of molecular dynamics is to investigate the structure of solid, liquid and gas using classical mechanics theories and Newton laws. This concept was introduced in the late 1950's by Alder and Wainwright to help in the study of the interaction of hard spheres(Astuti &

Mutiara, 2009). This led to the fist molecular dynamic simulation of protein in 1977 (Berendsen, Postma, Van Gunsteren, & Hermans, 1981; Levitt & Warshel, 1975). The idea for molecular dynamic simulation was born out of the quest to solve and understand the following

- To generate the trajectory and understand the dynamics of molecules within a short time period.
- To give a detailed understanding from the simulations and comparisons to lab experiments.
- To serve as a bridge between the experimental analysis and theory.

Molecular Dynamics provides a dynamic time integration method to compute the intermolecular motion and forces between molecules, solving the Newton's equation at the atomic/molecular level for the position and the velocities of the associated atoms in a molecular structure(Astuti & Mutiara, 2009). Molecular Dynamics computes the future position and velocity of atoms if the initial position, force and time interval is known (Astuti & Mutiara, 2009).

In recent years, computer simulation of the dynamics of molecules has become an important tool in the research field owing to the advancement in software and hardware. This chapter briefly provides an overview of the general methods involved in molecular dynamics analysis. The introduction of molecular dynamic simulation (MD) as a method has facilitated the understanding of structures of biological materials.(Pal, Weiss, Keller, & Müller-Plathe, 2005). Series of computational studies have been previously conducted on the protein behavior in vacuum and in different mediums using molecular dynamic simulations. MD took its root and made significant impact in the 1980's with the advent and introduction of some general-purpose force fields (SPC/E,TIP4P,TIP3P)(Pal et al., 2005) developed for water. Over the years,

several force fields such as AMBER(Kollman & Merz Jr, 1990), GROMOS(Schuler, Daura, & Van Gunsteren, 2001), CHARMM(Turner, Moliner, & Williams, 1999), AMBER (Chipot, Maigret, Pearlman, & Kollman, 1996) and OPLS(Chipot et al., 1996) were introduced for studying proteins DNA and other similar bio-molecules.

In addition, other molecular dynamic simulation tools are also currently available. These include the AMBER packages(Chipot et al., 1996) , which contains a suit of programs with sets of molecular and mechanical force field that can model bio-molecules, OPENMM simulation package(Eastman & Pande, 2010) which is also molecular oriented , PINY_MD(Balasubramanian & Bagchi, 2002),which is object oriented; KALYPSO (Karolewski, 2005) is another MD analysis code for metallic and bimetallic crystal oriented . GROMACS is the MD analysis code employed in the present work. Groningen machine for chemical simulations(GROMACS)(Lindahl, Hess, & Van Der Spoel, 2001) is most commonly used open software for molecular dynamics studies of materials and biological systems.

Among the simulation packages mentioned above, we chose to use GROMACS as it is one of the fastest molecular dynamics codes among the open source MD simulation and analysis codes that are currently available. It is widely used for research in biological, materials, physical and chemical computational studies. Moreover, it is fast and user friendly and gives the option to choose the parameters we wish to calculate during simulation analysis studies.

2.2 GROMACS

GROMACS was developed at the University of Groningen, Netherlands, in early 1990s. The code currently operates on various operating platforms including UNIX and LINUX operating systems and could run on a multiple CPU core. One convenient feature is, it has the

ability to convert protein data bank (pdb) structure files to GROMACS compatible structure files understood by GROMACS MD code as part of the package. GROMACS has no force field of its own but it is compatible with generalized force fields such as OPLS, GROMOS,AMBER and ENCAD (Van Der Spoel et al., 2005).

2.3 Force Fields

Force field defines the potential energy associated with various molecular motions that are accounted in molecular dynamics analysis. The sum of potential and kinetic energy function defines the Hamiltonian H in a classical molecular mechanics system (Dirac, 1950). This is dependent on the velocity (v) and the position (r) of the particles constituting the system. The momentum p is dependent on the velocity v of the particle and the Hamiltonian: H can be written as a function of position and momentum

$$\mathrm{H}(q,p) = \frac{p^2}{2m} + v(r)_i \tag{2.1}$$

Kinetic energy is defined as a function of the particles mass and momentum and the potential V defining the intermolecular interaction is dependent on the positions of the particles. This interaction between molecules is defined by force field. Of the different possible force fields, the present study employed the OPLS force field(Ponder & Case, 2003). The potential energy of a molecular mechanics system can be expressed as a function of the coordinates of the particles or atoms (r). The force $f_i(r)$ acting on individual particle/atom is the negative derivative of potential energy $V(r)$ with respect to the position of the particle.

$$f_i(r) = -\frac{\partial v_i}{\partial r_i} \tag{2.2}$$

The relationship between position and time is expressed by the following Newton's equations

$$v_i = \frac{dr_i}{dt}, \tag{2.3}$$

$$\dot{v}_i = a_i = \frac{dv_i}{dt} \tag{2.4}$$

Equation (2.4) above is the relationship between the velocity, acceleration and time where a_i is the acceleration of the particle and v_i the velocity. The relationship between the force, mass and acceleration is given by

$$f_i = m_i a_i \tag{2.5}$$

The force field and the initial configuration can be used to estimate or calculate the motion and position of the particles/atoms in a molecular system. The sum of the intermolecular interaction and intra-molecular interaction is equal to the total potential energy of the system. The interaction of the atoms of same molecule is known as the intra-molecular interaction while the interaction of atoms of different molecules is known as intermolecular interaction. In the following section, we describe these two interactions in detail.

2.4 Bonded Interaction

In a molecular dynamic system there are interactions between atoms of the same molecules. The interaction between the atoms of the same molecule sharing bonds is known as bonded interaction. This interaction could be in the form of stretching, torsion or angle bending. The bond and angle vibration are harmonic. Bond vibrations have very high frequency which

will require slow time step this is handled by introducing a rigid constraint bond to replace the bond potential, that will allow the use of bigger time steps(Comba & Hambley, 2007). The bond stretching and bond bending energy equation is based on Hooke's law("ntroduction to Molecular Mechanics,").

$$E = \sum k_b (r - r_0)^2 \tag{2.6}$$

where E is energy, K_b is interaction constant, and r_0 is the equilibrium position length between two bonded atoms("ntroduction to Molecular Mechanics,").

$$\mathrm{E} = \sum k_\theta (\theta - \theta_0)^2 \tag{2.7}$$

where E represents energy, k is constant, θ_0 is the equilibrium angle.

2.5 Intermolecular Interaction

Nonbonded interactions refer to the interaction between atoms of different molecules. Force fields are separated into different mathematical terms to help account for individual element of the potential for computational purposes (Pal, 2005). The interaction of the non bonded molecules can be modeled using the Lennard Jones potential and the columbic potential(Müller-Plathe, 1993) as described in the following equation

$$v_{nonbonded(i,j)} = v_{lernard(i,j)} + v_{coulomb(i,j)} \tag{2.8}$$

where v is the potential for non bonded interaction between atoms of different molecules i and j with the total potential is given as the sum of the coulomb potential and the Lennard Jones potential.

$$v_{coulumb(i,j)} = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2.9}$$

$$v_{lernard(i,j)} = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{2.10}$$

ε_0 is the permittivity of free space, $q_{(i,j)}$ are charges, ε is the energy parameter with reference to the depth of the potential well, σ_{ij} is the finite distance for which inter-particle potential is zero.

2.6 Periodic Boundary Condition (PBC)

This is a technique used to study bulk or infinite system within a molecular simulation. This involves replicating the simulation box to form infinite lattice. PBC addresses the problem encountered during simulation when atoms move out of simulation box; the image enters the simulation box from the opposite side to help maintain the density of molecules in the box at a constant level during simulation. Care must be taken to make sure the reference atom does not interact with its periodic image by employing the cut off radius less than or equal to that of reference atom.

We use the molecular dynamics simulation techniques and methods described in this chapter to explore lysozyme structure and its dynamics in binary mixtures of water and alcohol. For this purpose, we begin our simulation study with lysozyme protein in water that is discussed in the next chapter.

CHAPTER 3

Molecular Dynamic Simulation of Lysozyme Protein in Water

3.1 Introduction

Lysozyme protein in water environment has been extensively studied both by experimental and computational techniques. Since our intention is to study lysozyme in different solvent conditions, we begin with the well studied example, namely lysozyme in water environment. For this purpose, we analyze various thermodynamic quantities in addition to structural parameters such as radius of gyration and root mean square deviation (RMSD) for the molecular system. This initial study helps us verify the consistency of our present simulation method by comparing with the results from the existing literature. It also serves as a reference to compare and contrast the protein structure and its dynamics under other solvent environments such as the ethanol-water mixture focused in the present work. As mentioned earlier, we use GROMACS MD code for this purpose. The following sections describe the simulation details of lysozyme protein in water.

3.2 Simulation Details

Simulation of lysozyme protein in water was performed using GROMACS MD code with the lysozyme (pdb code: 1AKI) protein structure obtained from the protein data bank. The pdb structure of protein was converted to a GROMACS compatible structure. The minimization file minim.mdp, the equilibration files nvt.mdp, npt.mdp, md.mdp were all created in the same directory for the simulation. TIP3P water was used as solvent in this simulation(Jorgensen & Jenson, 1998), while computational tool VMD (Humphrey, Dalke, & Schulten, 1996) was used to visualize simulation progress and inspect protein structure visually. All computations were

performed on an in-house computation cluster, *Hermes* located on the NC A&T campus. The following steps were employed in the simulation analysis.

- Converted the lysozyme structure obtained from the protein data bank to a GROMACS compatible structure
- Using the right configuration and force field, we generated the system that was viewed using the VMD software.
- We generated output files for the minimization which will help keep the system stable at the minimum energy state. By using conjugate gradient method, the protein was minimized.
- Protein was solvated with TIP3P water molecules. The solvated protein was equilibrated using NVT and NPT ensembles to ensure the stability of the system.
- Both NVT and NPT equilibration simulations were carried out for 200ps with a 1fs time step. During this NVT and NPT equilibrations, the temperature used was 300K and the pressure used was 1 bar. The final configuration from NPT equilibration was taken as the starting structure for the production MD.
- We carried out the MD run using the equilibrated configuration as the starting configuration of production simulations.

The simulation system contained one lysozyme protein molecule of 129 residues with 1960 atoms and 12365 water molecules (37095 atoms). The system contained a total of 39055 atoms. We used the Berendsen thermostat (Lemak & Balabaev, 1994) which helps keep the average temperature of the system stable thereby regulating the velocities of the components of the system. For the MD production dynamics analysis, the system was run for 50,000,000(fifty

million) time steps with a 1 fs time step size for a total duration of 50ns at 1atm pressure and 300K temperature. All the simulation system details are presented in appendix A.

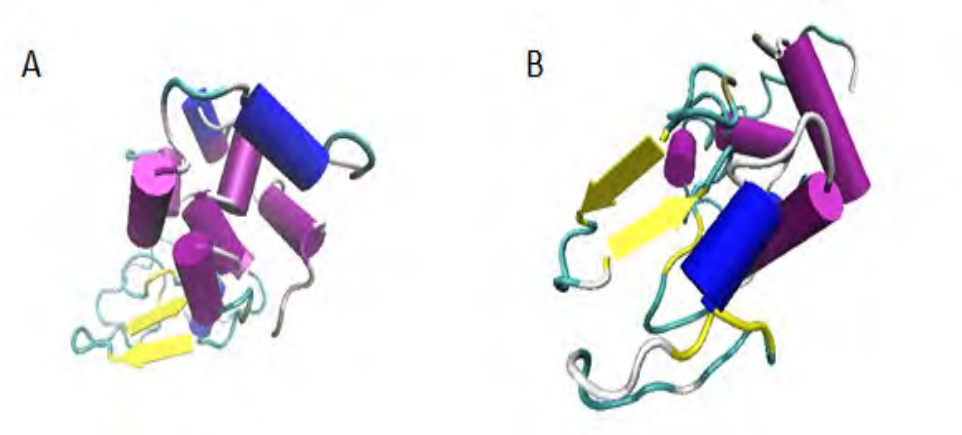

Figure 3. (A) Initial structure and (B) final structure of lysozyme protein in water.

During the production dynamics simulation of solvated protein, we recorded the trajectory by collecting data at every 100ps for post-processing. Figure 3 presents the initial structure and final structure of lysozyme protein in water (water molecules are not displayed for clarity). The recorded data included basic thermodynamic quantities such as energy, pressure, temperature, enthalpy and position and velocity of atoms present in the system. We describe these quantities along with the quantities used in analyzing the simulation data next.

3.3 Total Energy

Energy is the ability to do work. Total energy of the system is the sum of the kinetic and the potential energy. The change in kinetic energy determines the amount of work done by the system. Total energy of protein in water system plotted as a function of simulation time shown in figure 4. The total energy of the system remains near constant during the simulation, indicating the energy conservation within the simulated system. The associated enthalpy of the system is shown in figure 5.

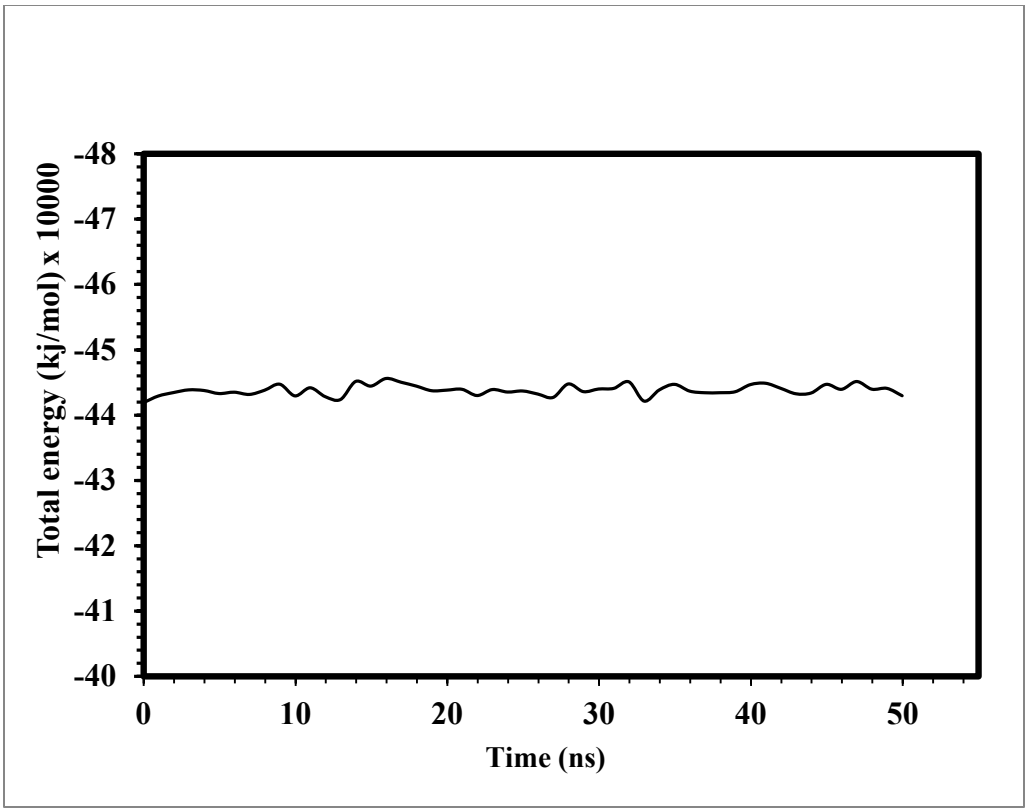

Figure 4. Total energy of protein in water.

3.4 Enthalpy

Enthalpy is a measure of the of the total energy of a thermodynamic system(Korosec, Limacher, Luthi, & Brandle, 2010). This includes the energy required to create the system known as the internal energy(U) and the product of the pressure and volume (PV) given by the equation below

$$H = U + PV \tag{3.1}$$

During the reaction process at constant pressure the heat evolved is equal to the change in enthalpy.

The change in enthalpy is given as (ΔH) and can be written as

$$\Delta H = \Delta U + P\Delta V \tag{3.2}$$

The calculated enthalpy of the system as observed in simulations is shown as a function of time in figure 5.

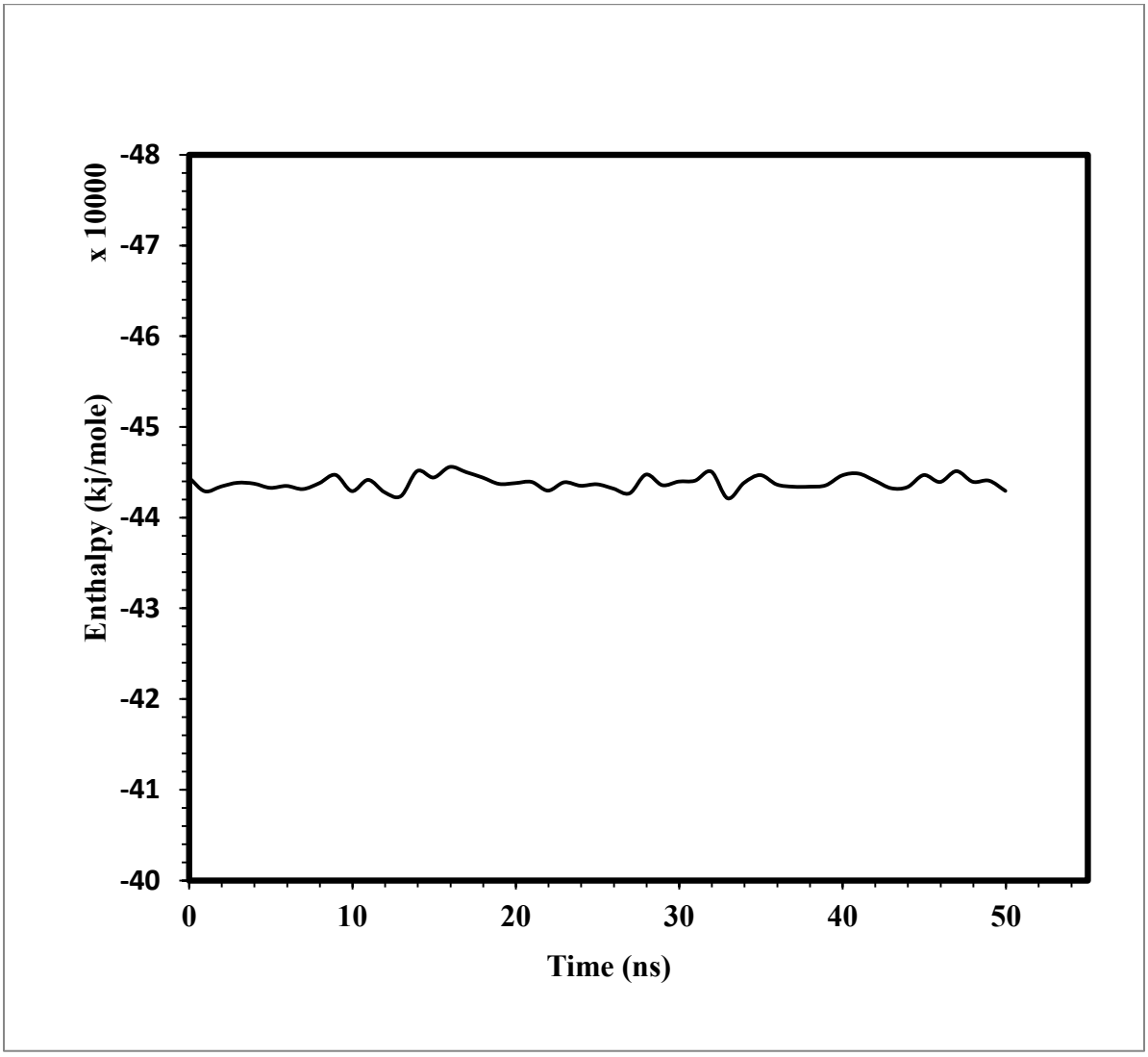

Figure 5. Enthalpy of the system in water environment.

The average enthalpy was calculated to be -443,841(KJ/mol) for protein in water environment; this is in reasonable agreement with the previous simulation studies (Wensink et al., 2003b) (Zhou).

The compactness of the protein is analyzed by plotting the radius of gyration of protein molecule during the simulation time of 50ns as shown in figure 6.

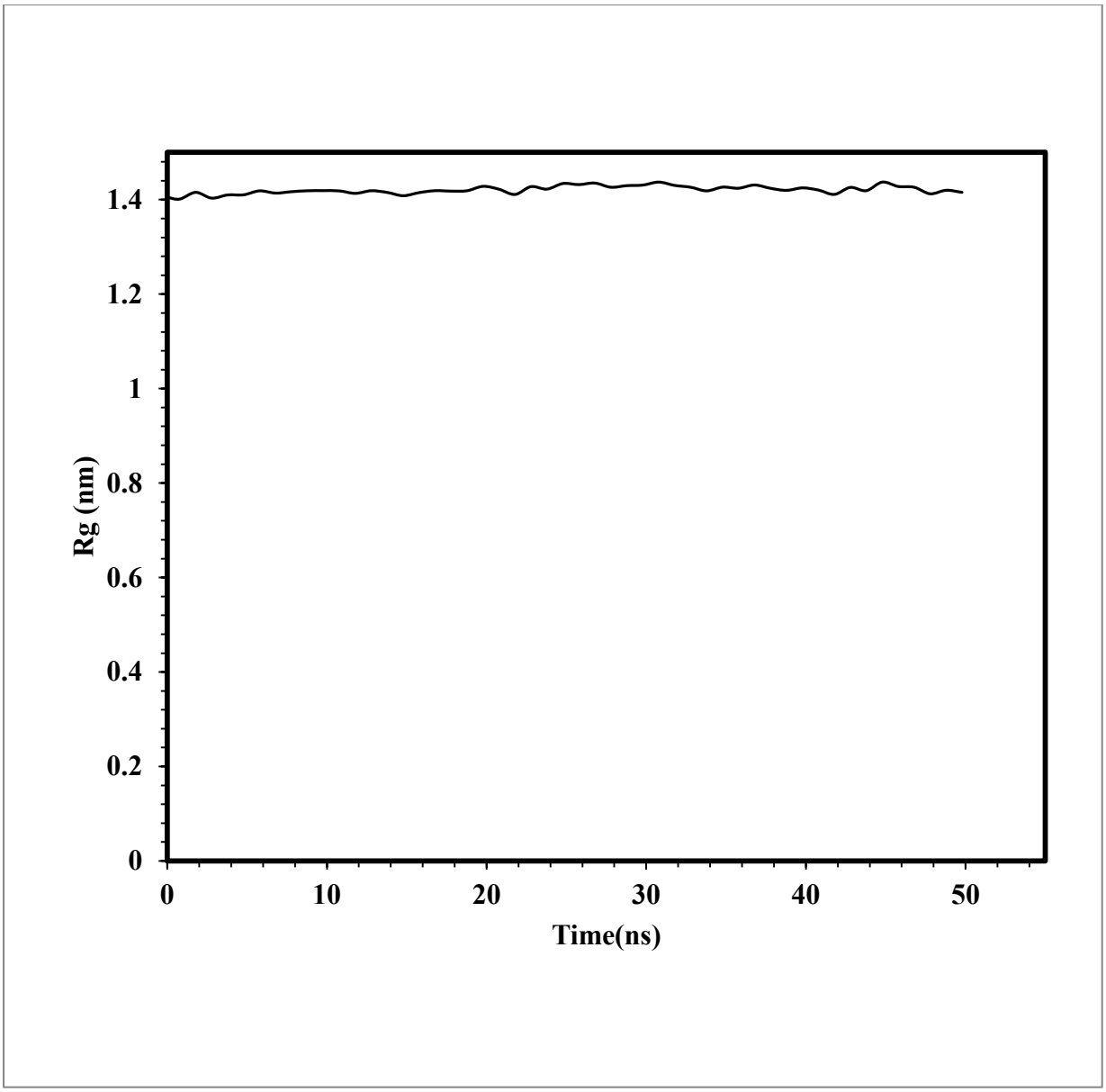

Figure 6. Radius of gyration of protein in water.

3.5 Radius of Gyration

This helps to measure and understand the compactness of the protein structure. For example, how much the structure has spread or contracted with regards to its degree of freedom. The formula used in calculating the radius of gyration is given by (Newcomer, Lewis, & Quiocho, 1981).

$$R_g = \left(\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \quad (3.3)$$

m is the mass of the atom r_i and i are the positions specific to a particular atom with the reference point being the center of mass(Lobanov, Bogatyreva, & Galzitskaya, 2008).

The R_g value for the lysozyme protein obtained from present simulation study is shown as a function of time in figure 6. As can be seen in figure 6, the R_g value fluctuates around an average value of 1.42nm throughout the simulation. This indicates the protein maintains its relative compact size during the entire dynamics simulation time. This observation concurred with the VMD visual inspection. The final structure of the protein at the end of 50ns is shown in figure 3B. As can be seen from the figure the protein is still in compact folded configuration, revealing the stability of the protein in water for the entire duration of the dynamics simulation.

3.6 RMSD

The RMSD (root-mean square deviation) is the measure of the average distance between the atoms of the back bone of superimposed proteins. The RMSD can be calculated with the equation(Carugo & Pongor, 2008)

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2} \quad (3.4)$$

RMSD could be used for quantitative comparison between the structure of the native state of protein and its partially folded state. For the present lysozyme-water system, RMSD value for the entire simulation time (50ns) is shown in figure 7. The average RMSD value is less than 1nm, indicating a relatively folded structure for the protein in this environment, consistent with the existing simulation studies(Bowman & Pande, 2009).

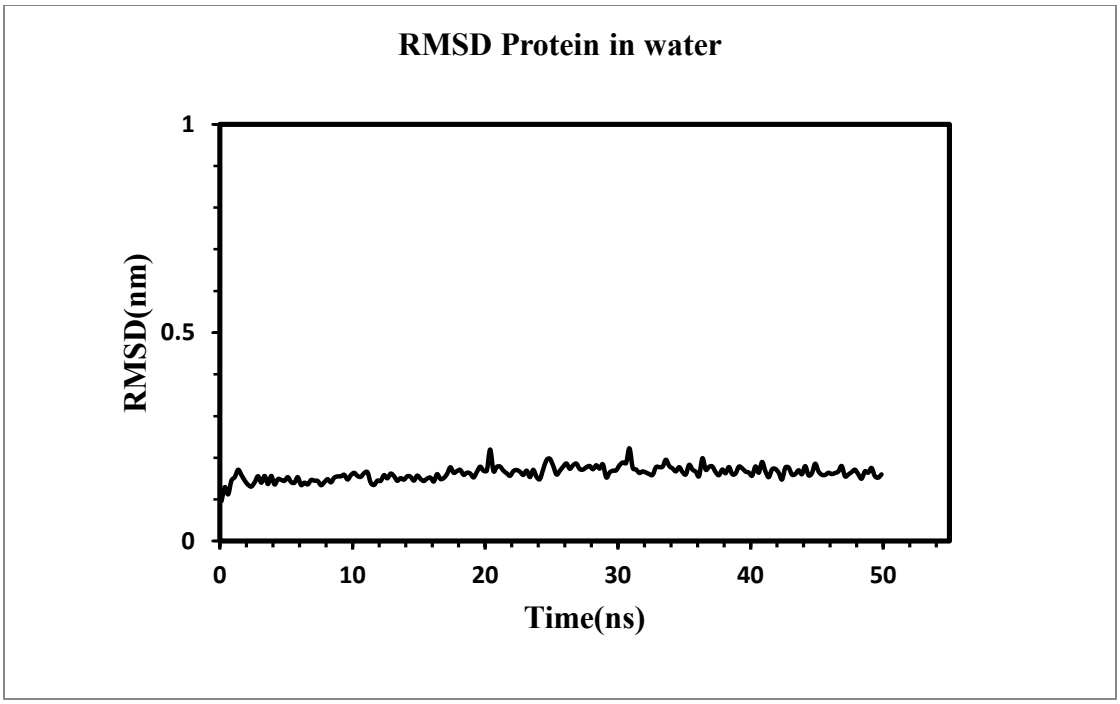

Figure 7. RMSD of lysozyme protein in water.

3.7 Summary

In summary, this chapter presented a detailed study of lysozyme protein structure and dynamics in water employing GROMACS molecular dynamics simulations. MD analysis studies involved the solvation of the protein by using appropriate number of TIP3P water molecules. The solvated structure was equilibrated in NVT and NPT ensembles for 200ps in each case. The final equilibrated structure containing a total of 39,055 atoms was used as the starting structure for the production, dynamics simulation. The dynamics analysis was conducted for 50ns with a 1fs time step at 1atm pressure and 300K temperature with the simulation data collected at every 100ps for post-production analysis.

Average values for both the total energy and enthalpy were in agreement with the previous simulation studies. This confirmed the validity of our simulation setup. The average RMSD was found to remain less than 1nm throughout the simulation, confirming the stability of

the protein in a folded state. Similarly, the average *Rg* value for the protein in the present study is 1.42nm, indicating a compact protein structure during the entire simulation.

Our simulations show that the lysozyme protein remains in a stable folded confirmation in water at 300K and 1 atm pressure conditions. This is in good agreement with earlier simulation analysis reported in the literature(Dadarlat & Post, 2001). The next chapters discuss the lysozyme protein in different solvent conditions that include ethanol, and ethanol-water mixture and compared with the data from the water environment.

CHAPTER 4

Molecular Dynamic Simulation of Protein in Ethanol

4.1 Introduction

Previous chapter discussed the effect of lysozyme in water studied via molecular dynamics simulations. The present chapter focuses on the molecular dynamics analysis of lysozyme in ethanol solvent environment. Ethanol is known to have profound effect on proteins and biomolecules and mixes well with water. The behavior of lysozyme in a 100% ethanol solvent condition is presented and discussed in this chapter. Ethanol is described as a versatile solvent (van der Waal & van Bekkum, 1997). Despite extensive experimental studies, the effect of ethanol on lysozyme protein is not clearly understood from a molecular view point. In order to gain a molecular level understanding, present work focused on the molecular dynamic simulation of lysozyme in ethanol environment. We aim to gain insights into the structure and dynamics of lysozyme protein in ethanol and compare with that from water.

As before, in the present case lysozyme protein structure (obtained from protein data bank) is solvated with 100% ethanol compared to the solvation in a full water environment discussed earlier. The associated simulation details are described next.

4.2 Simulation Details

Molecular Dynamics analysis employing GROMACS was completed following the same steps as in the case of water. Subsequent post-analysis also followed the same methodology as in the case of water.

The lysozyme-ethanol molecular system employed in the present study contained one lysozyme protein (1,960 atoms) and 2,289 ethanol molecules (20,601 atoms), resulting in a total of 22,561 atoms in the simulation system. We used the Berendsen thermostat to keep the

average temperature of the system stable and to regulate the velocities of the system components. The details of simulation system are shown in appendix B.

At the end of the simulation we analyzed and compared the thermodynamic quantities following the same approach as in the case of water environment. We visually examined protein structure with VMD for any apparent visible structural changes in the protein structure in ethanol environment before and after the simulation. The initial and final structures are shown in figure 8.

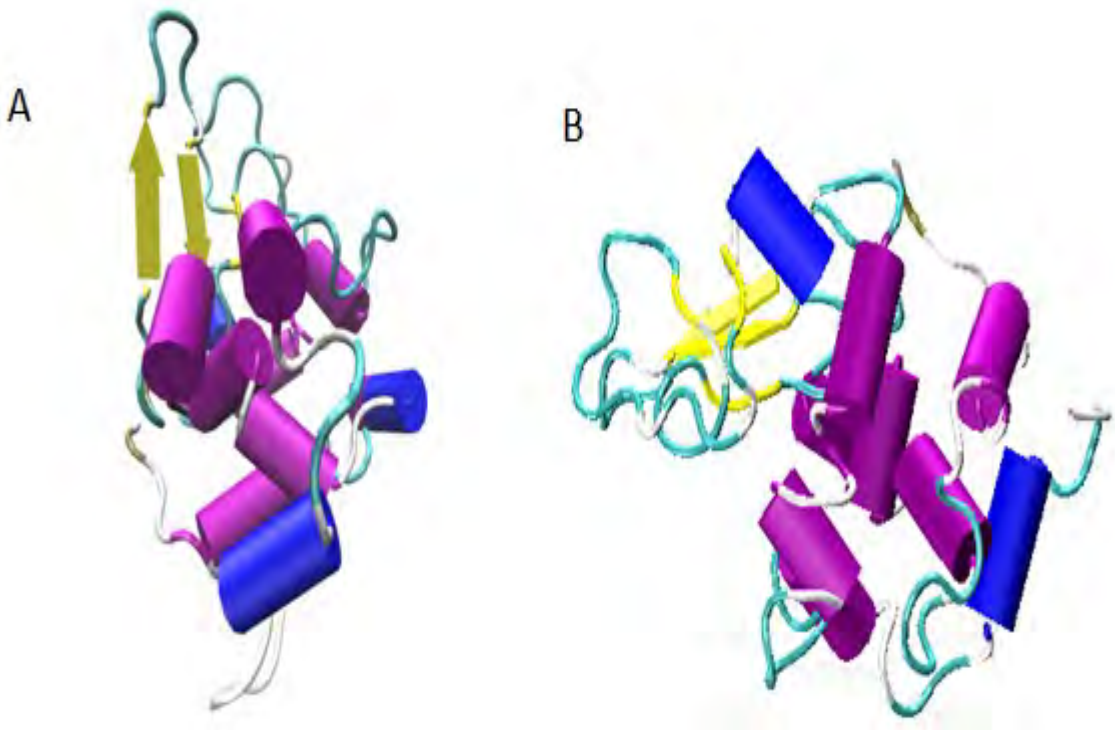

Figure 8. (A) Initial and (B) final structure of protein in ethanol.

As shown in figure 8, we observed that the protein structure showed significant changes in the size of the protein molecule. The final structure appears expanded/swollen compared to the initial structure. figure 9 presents a comparison of lysozyme in water compared to that of ethanol.

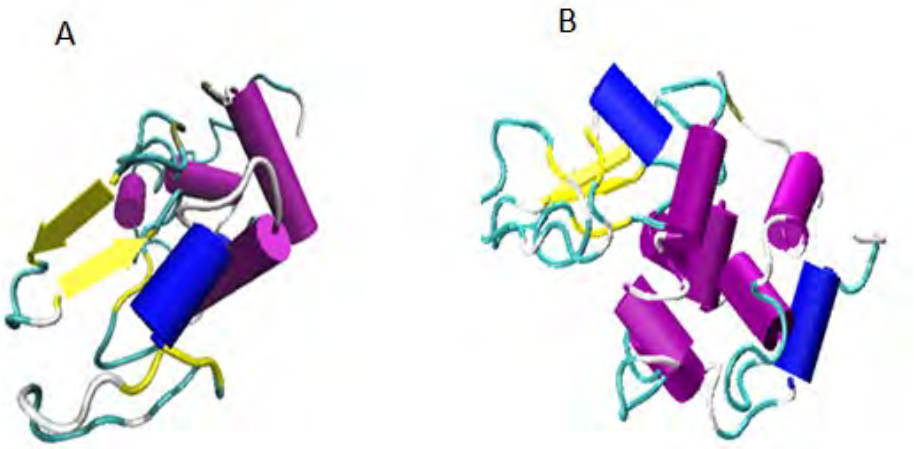

Figure 9. Comparing the final structure of protein in (A) water and (B) ethanol.

As seen from figure 9, lysozyme in ethanol appears to be swollen compared to that of water. A closer examination reveals significant difference in the protein alpha-helix content in ethanol compared to water. The alpha-helix structures were broken into relatively shorter helices in ethanol compared to water. In other words, longer helix components were no longer stabilized when solvent environment was changed to ethanol. One of the factors that stabilize protein secondary structure is the hydrogen bonding between protein and solvent. As the solvent changed from water to ethanol, the hydrogen bond network between protein and water, responsible for stabilizing alpha-helix structure was broken in case of ethanol. This led to the instability of protein secondary structure in ethanol solvent. To further confirm our findings, we calculated and compared thermodynamic quantities such as total energy and enthalpy and structural quantities such as radius of gyration (*Rg*) and RMSD. The results of these quantities are compared with the results from water environment.

The total energy of the lysozyme protein system in ethanol and water is plotted in figure 10. As shown in this figure the total energy of water was -443,901 (KJ/Mol) and that of ethanol is -59,164(KJ/Mol). This significant energy difference clearly indicates that the lysozyme protein

in water system is more stable compared to the same protein in the ethanol system. This is in accordance with the significant change in the final protein structure in ethanol and water as shown in figure 9A and 9B. To further understand the effect of ethanol on protein, we calculated and compared the enthalpy of the system as shown in figure 11 below.

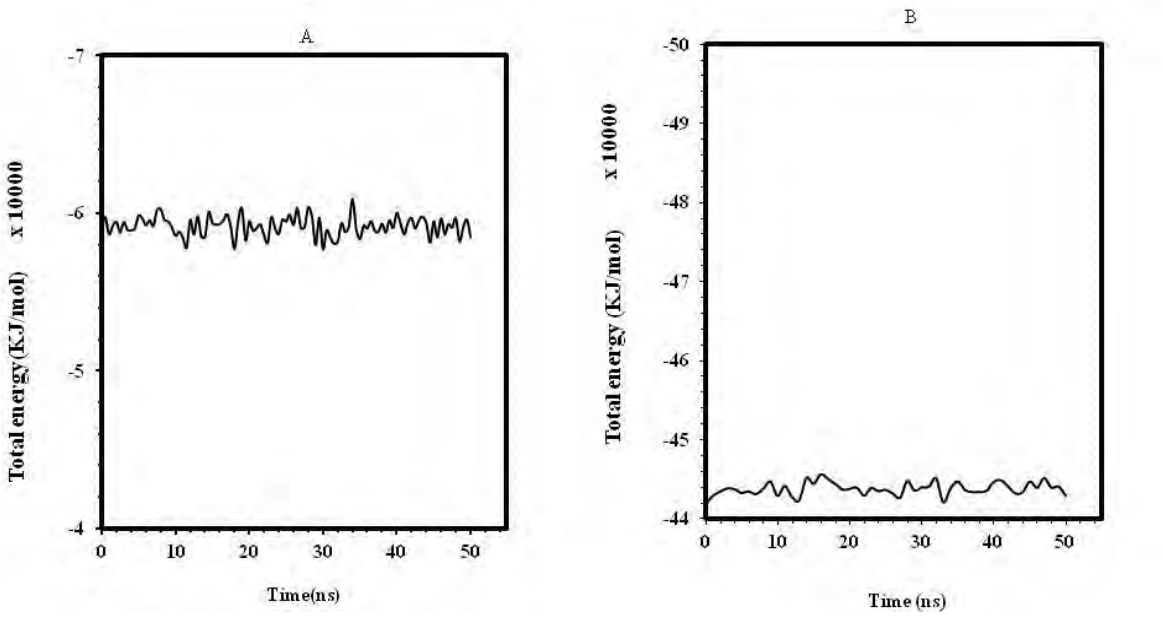

Figure 10. Total energy of protein in (A) ethanol and (B) water.

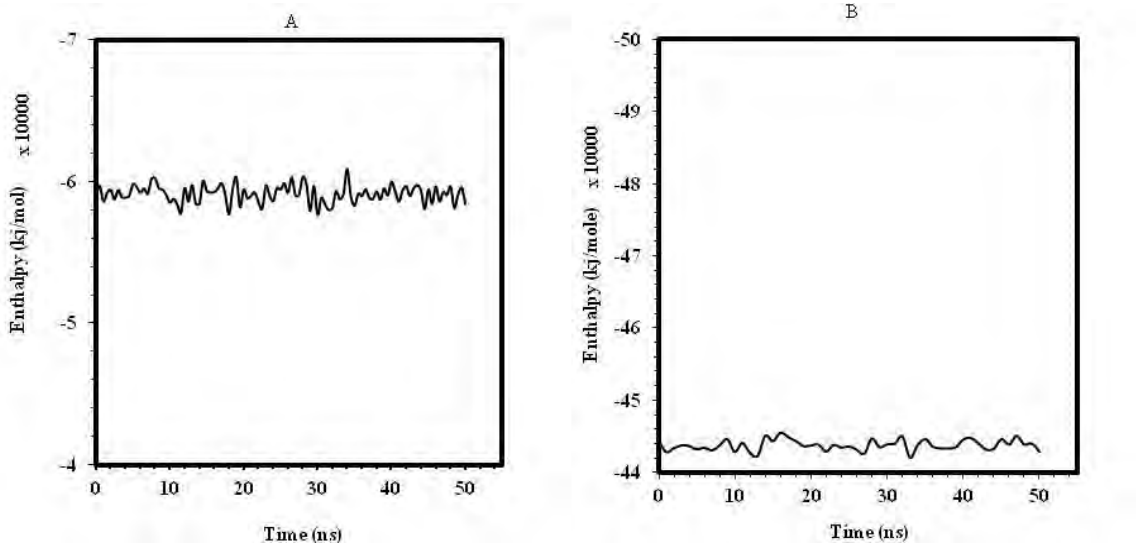

Figure 11. Enthalpy of protein in (A) ethanol and (B) water.

From figure 11, the enthalpy of protein in ethanol is approximately -59,149 (KJ/mol) and that of protein in water is approximately -443,841(KJ/mol). These results show significant increase in enthalpy. This further supports the change noticed in protein structure in figure 9. To further verify our results we analyzed the radius of gyration to check for the compactness of the protein by plotting the radius of gyration as a function of time as shown in the figure 12.

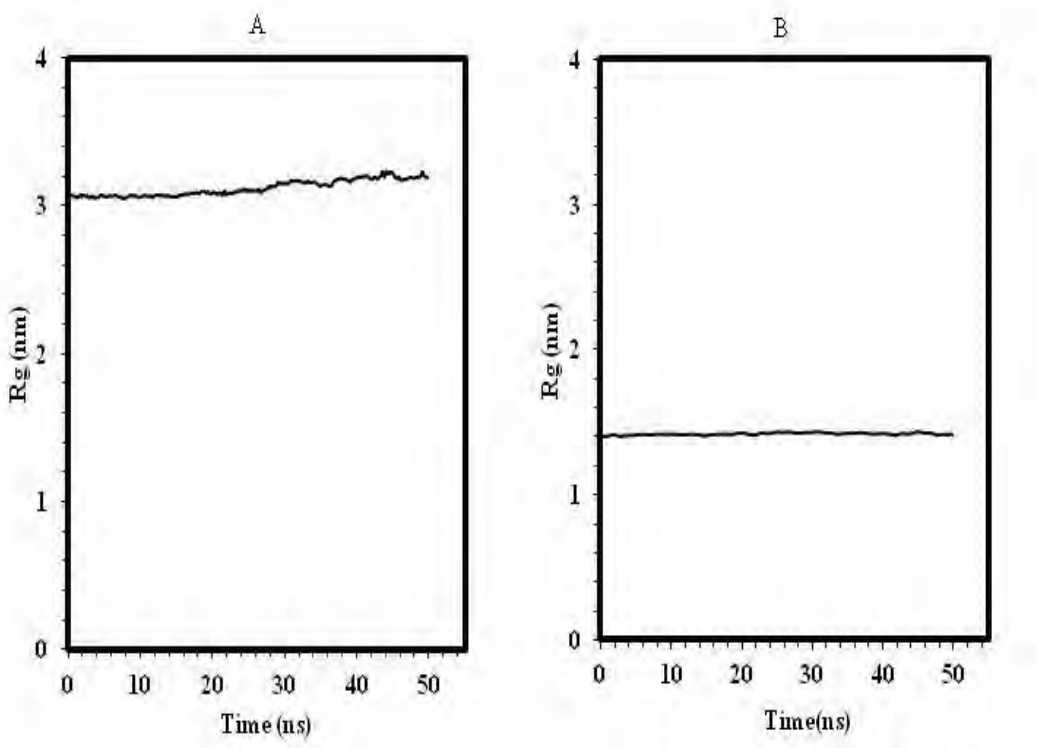

Figure 12. Radius of gyration of protein in (A) ethanol and (B) water.

In comparison, the radius of gyration of protein in water is approximately 1.4 nm as shown in figure 12B and that of ethanol in figure 12A is approximately 3.12nm. The difference between the two radiuses of gyration values from 1.4 in water to 3.12 in ethanol indicates a significant change in the compactness of the protein molecule. As a result of this change, it can be inferred that the protein molecule is swollen in ethanol compared to that in water. The structural stability of protein molecule in water and ethanol was compared by analyzing the root

mean square deviation (RMSD). Figure 13 presents the variation of the RMSD for water and ethanol during the dynamics analysis.

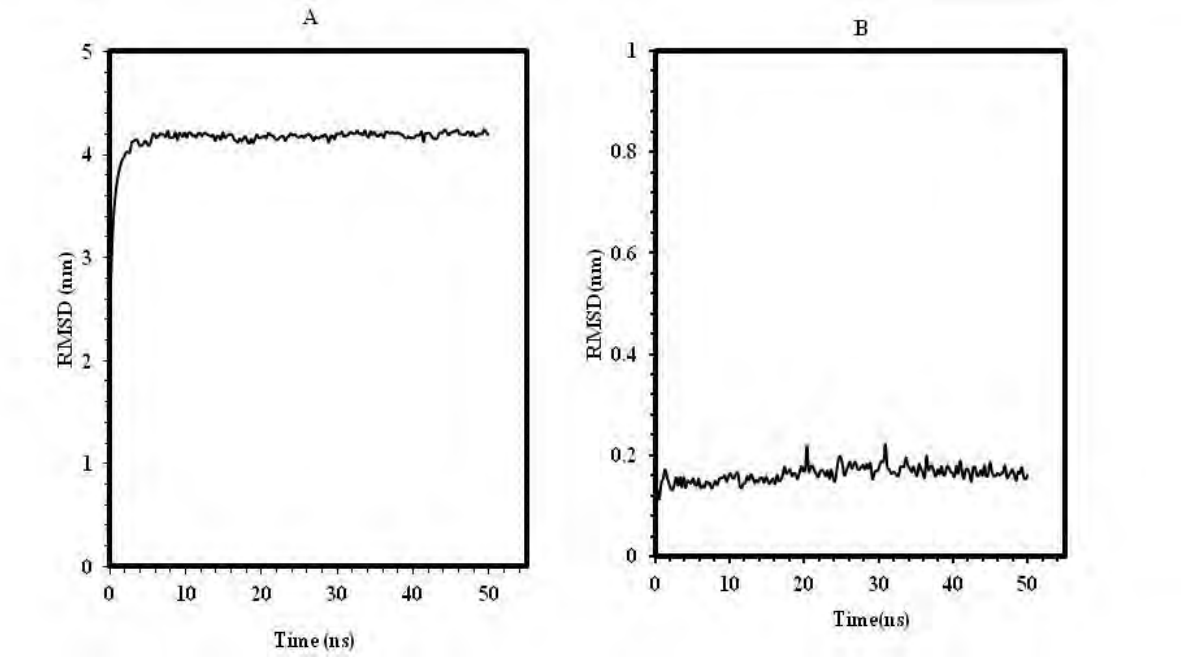

Figure 13. RMSD of protein in (A) ethanol and (B) water.

From figure 13A and figure 13B, the RMSD for protein in water and ethanol were found to be approximately 0.162 nm and 4.15nm respectively. This is a significant change in RMSD value of the lysozyme protein molecule in ethanol compared to that of water. In the folded native structure proteins typically have RMSD values of about 0.1 to 0. 2 nm (García & Onuchic, 2003) (Dokholyan, Li, Ding, & Shakhnovich, 2002). RMSD value as high as 4.15nm shows a significant deviation of the proteins from its native structure. This indicates that the lysozyme protein is more stable in water and deviated significantly from its native structure in ethanol.

4.3 Summary

In this chapter we studied lysozyme protein structure and analyzed its dynamics in 100% ethanol solvent environment. The simulation results showed increase in total energy, enthalpy, radius of gyration and root mean square deviation in the case of ethanol compared to that of

water. The significant increase in the calculated quantities affirms the changes in the stability of the system and compactness of the protein structure.

Our simulations show marked changes in protein structure and energy when solvated in ethanol compared to that in water. In order to quantify such changes in a systematic fashion, we proceed to study the effect of ethanol and water mixtures on lysozyme protein. For this purpose we use similar simulation setup as discussed in the previous chapters to create and study lysozyme protein at different ethanol concentrations in water. In the present study, low ethanol concentrations in the range of 0 – 12% in the ethanol – water mixtures are considered. Lysozyme protein behavior in the ethanol-water mixture environment at various percentages is compared to that of pure water and 100% ethanol environments, and is presented in the next chapter.

CHAPTER 5

Molecular Dynamic Simulation of Lysozyme Protein in Ethanol-Water Mixtures

5.1 Introduction

In the previous chapters we have studied the effect of pure water and pure ethanol on lysozyme protein. To further understand the effect of ethanol and water on protein in a systematic fashion and to quantify the effect, here we conducted MD simulation of protein in ethanol-water mixtures over a range of ethanol concentrations. For this purpose we set up six different ethanol concentrations of approximately 2%, 4%, 6%, 8%, 10% and 12%. This concentration range of ~2% to ~12% was selected based on the prior experimental investigations in the literature (Onori & Santucci, 1996; Wensink et al., 2003a). In this work we aim to qualitatively analyze the lysozyme protein behavior as a function of ethanol concentration and to understand the changes with the addition of low concentration of ethanol. We begin the simulation set up as before but solvating the protein with both water and ethanol in specified compositions. The simulation details are described as follows.

5.2 Simulation Details

Molecular Dynamics analysis was conducted as discussed in the previous chapters by solvating the lysozyme protein in both ethanol and water molecules in appropriate ratios so as to obtain different concentrations of ethanol. We chose approximate percentage based on the mass of ethanol molecules and water molecules for the required percentages of water and ethanol mixture. We equilibrated the lysozyme in the ethanol-water mixture system and conducted our full simulation for 50ns with a time step of 1 fs, at a pressure of 1 bar, and a temperature of 300K. At the end of the simulation process, we calculated and compared the thermodynamic and structural quantities similar to the quantities calculated in the previous chapters based on the

system information shown in appendices C and D. The initial and final structures after the MD analysis of the lysozyme protein at different ethanol-water mixture concentrations are shown in appendix E. From this figure, noticeable changes in the initial and final structure of protein at different water-ethanol concentrations is clearly seen.

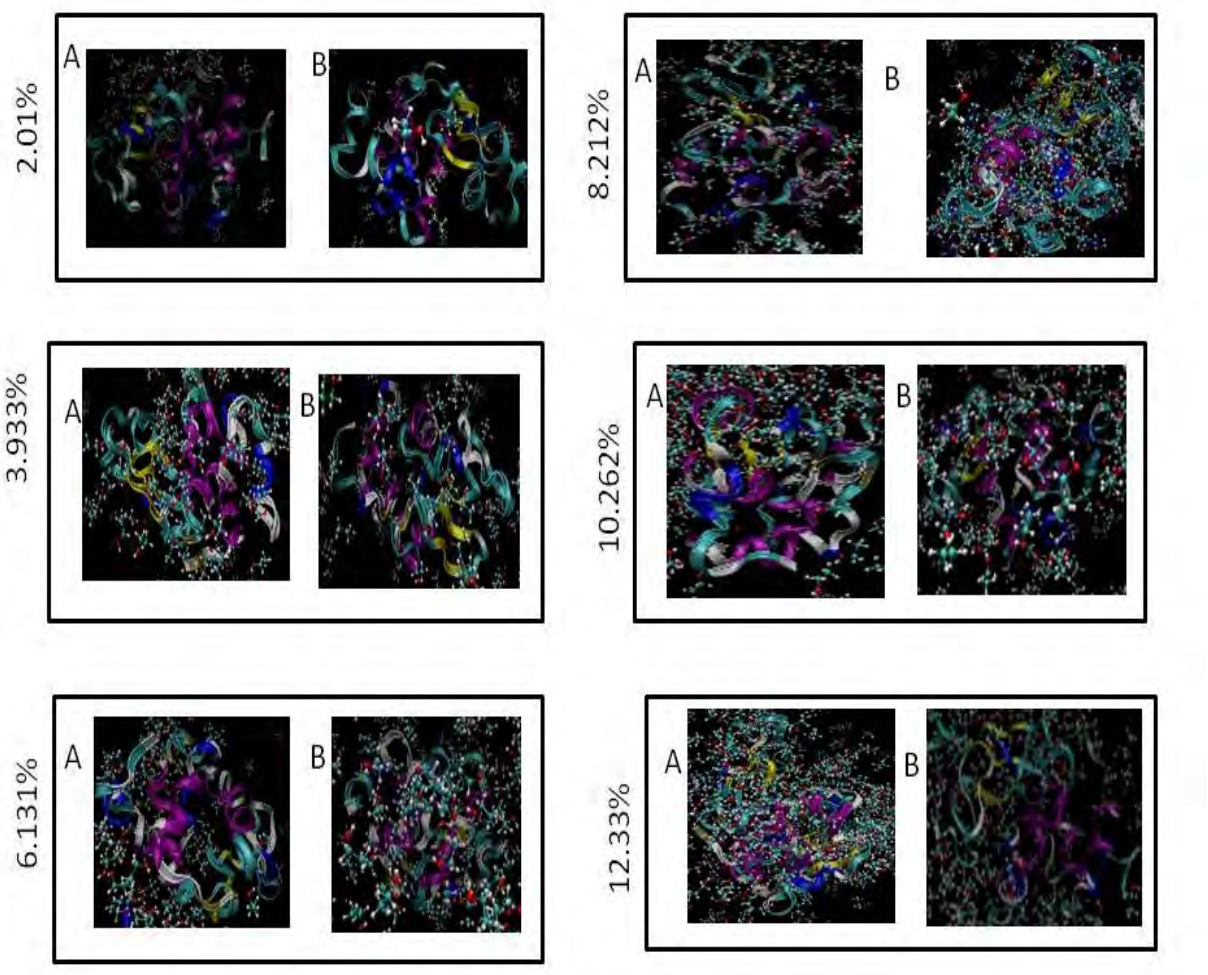

Figure 14. (A) Initial structure and (B) final structure of protein for different concentration of ethanol-water mixed solvents.

Figure 14 shows the diffusion of ethanol molecules into the lysozyme causing the protein molecule to swell as the protein gets increasingly destabilized with increase in ethanol concentration. Both lysozyme and ethanol molecules are shown in this figure. From figure 14A

and figure 14B, we notice that the protein secondary structure gets altered as ethanol progressively replaces water molecules with increasing ethanol concentration. This leads to a decrease in alpha helical content of the protein in the same direction. Together, these observations reveal that the protein increasingly gets deviated from its native folded structure as the ethanol concentration increases.

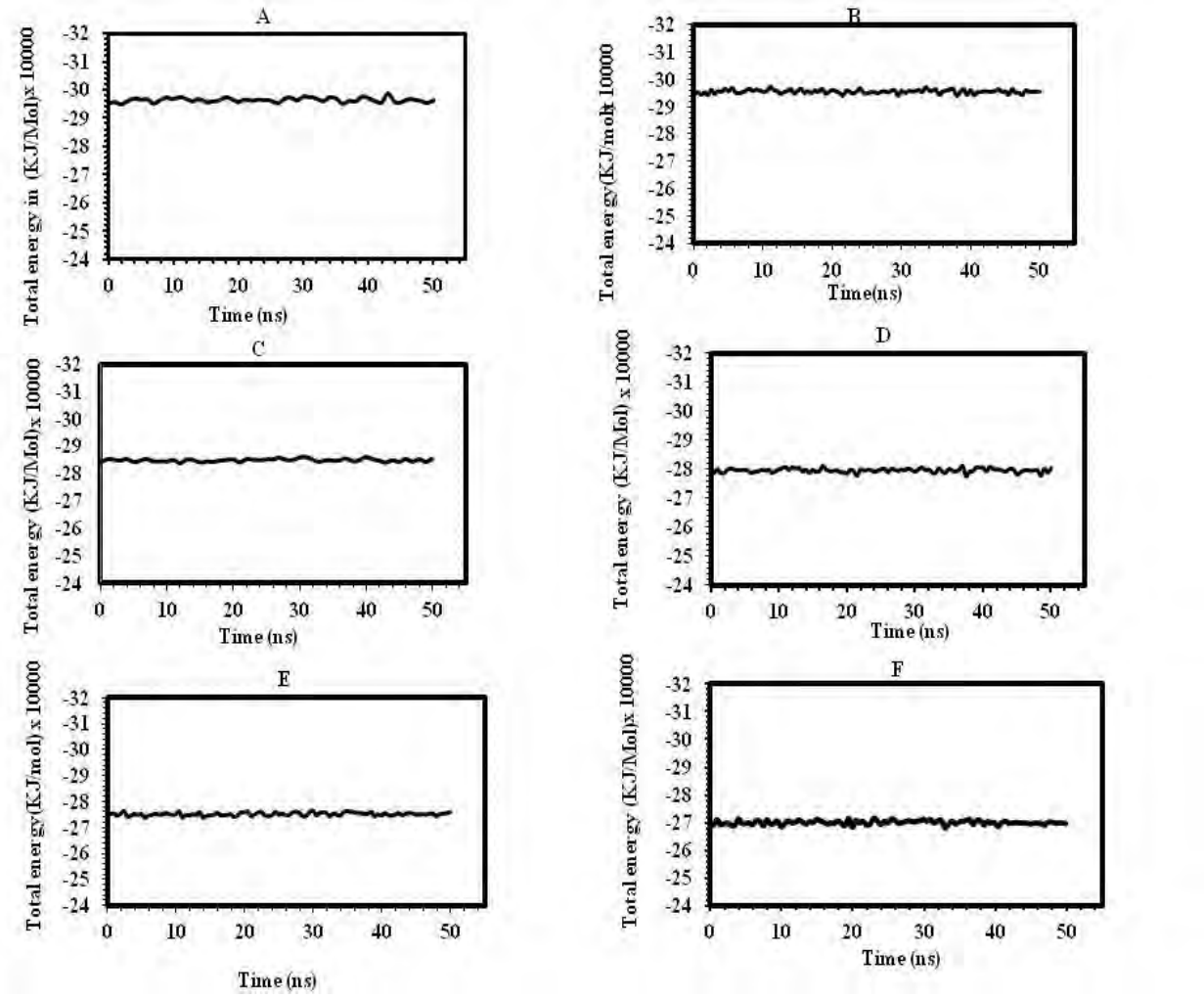

Figure 15. Total energy of protein in ethanol-water mixture for approximately (A) 2%, (B) 4%, (C) 6%, (D) 8%, (E) 10% and (F) 12% of ethanol.

To further verify our observation we calculated and compared the thermodynamic quantities starting with the total energy as presented in figure 15. An increase in the total energy

of the system with increase in the concentration of ethanol is noticed. This also indicates decrease in the stability of the system as the concentration of ethanol increases.

The enthalpy variation of the system is shown in figure 16 with an increase in the enthalpy of the protein observed as the concentration of ethanol increases.

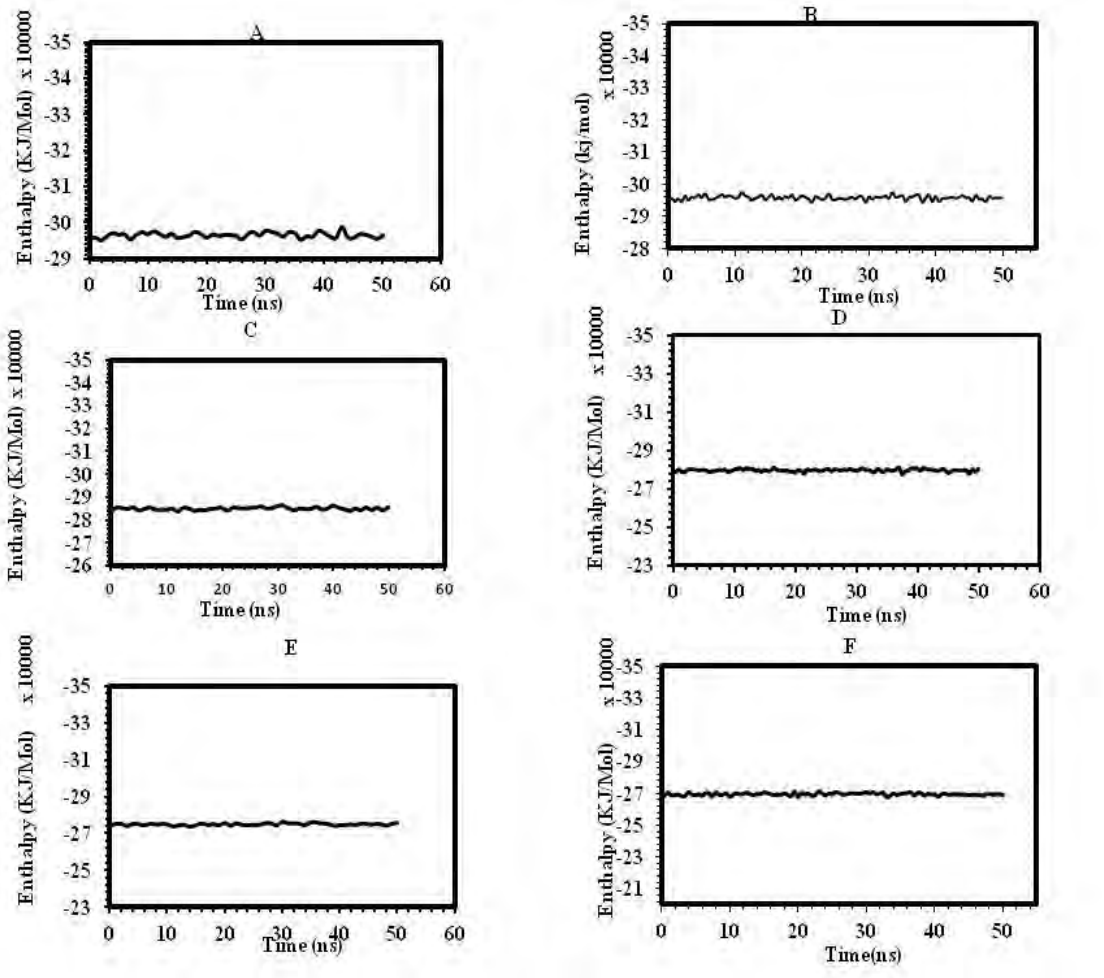

Figure 16. Enthalpy of protein in ethanol-water mixture for approximately (A) 2%, (B) 4%, (C) 6%, (D) 8%, (E) 10% and (F) 12% of ethanol.

A comparison of the radius of gyration of protein for all the six different concentrations shows an increase in the radius of gyration as the ethanol concentration increases. For the concentration of ~2% ethanol, the calculated the radius of gyration is 2.03nm and that of ~12%

ethanol is 2.62nm.In comparison, 100% ethanol showed a radius of gyration of 3.12 nm, clearly indicating an increase in the radius of gyration of protein with increase in the concentration of ethanol.

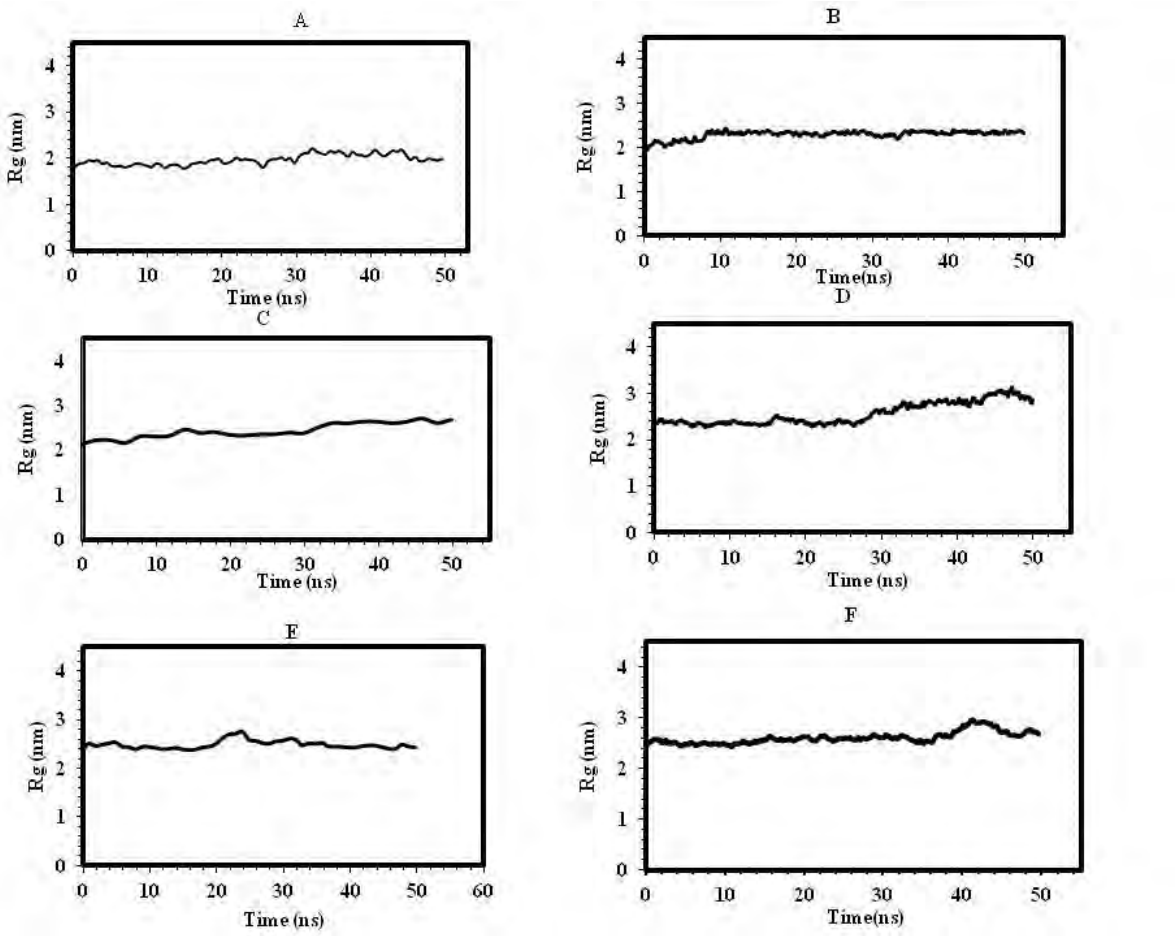

Figure 17. Radius of gyration of protein in ethanol-water mixture for approximately (A) 2%, (B) 4%, (C) 6%, (D) 8%, (E) 10% and (F) 12% of ethanol.

Similar to the case of the radius of gyration, the calculated RMSD of lysozyme protein at different ethanol concentrations is shown in figure 18. We found the lysozyme protein RMSD also increases with increase in ethanol concentration. RMSD value for ~2% ethanol was 1.93nm compared to ~12% ethanol value of 3.1nm. Present analysis at various ethanol percentages clearly shows a significant increase in RMSD with increasing ethanol concentration.

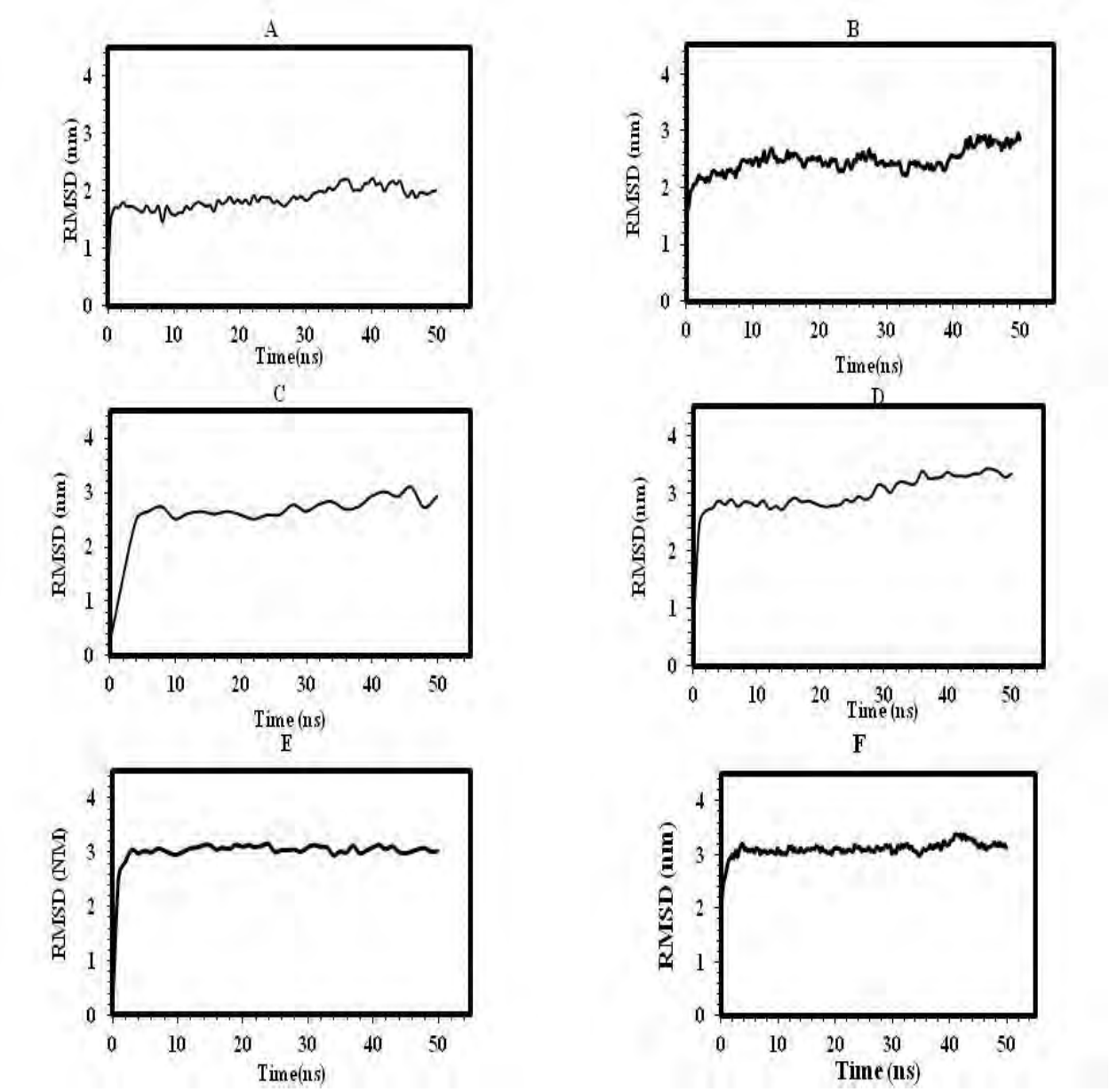

Figure 18. RMSD of protein in ethanol-water mixture for approximately (A) 2%, (B) 4%, (C) 6%, (D) 8%, (E) 10% and (F) 12% of ethanol.

Together, these results show a systematic increase in the calculated thermodynamic and structural quantities of lysozyme protein with increasing ethanol concentration in water. A plot of the average of these quantities - total energy, enthalpy, radius of gyration and RMSD - as a function of time is shown in figure 19.

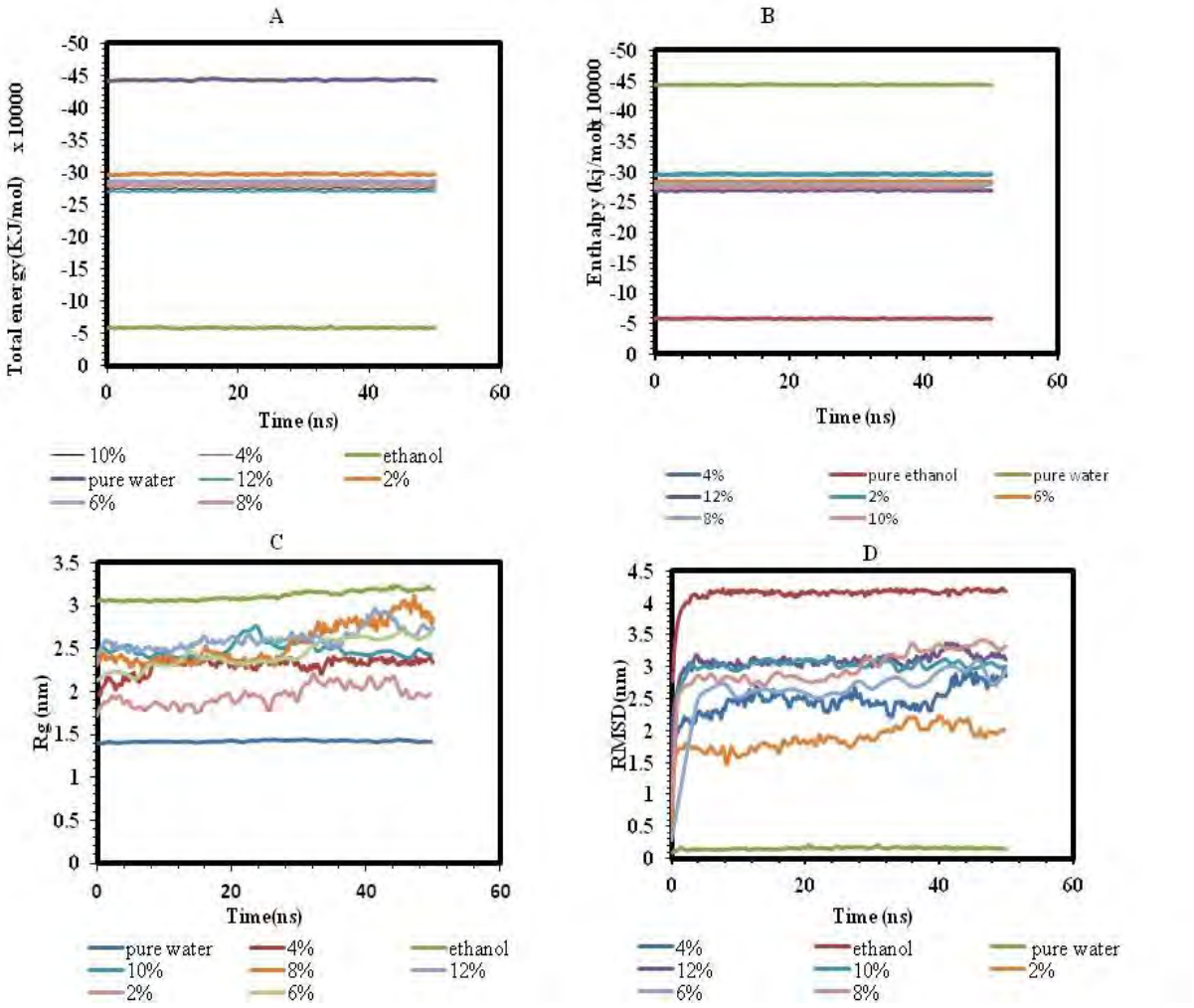

Figure 19. Average thermodynamic quantities of (A) total energy, (B) enthalpy, (C) radius of gyration and (D) RMSD as a function of time.

From figures 19A to 19D we observed a significant difference in the key parameters studied between protein in pure water and that in ethanol. In the range of ethanol-water mixture concentrations studied, we observed a trend of increase in the energy and enthalpy with increase in ethanol concentration. This increase in thermodynamic quantities further supports the observed swelling in protein structure. The dynamic variation in the key parameters over the MD analysis time duration is plotted in figure 20. The dynamic changes also showed similar behavior that of the time averaged values for the key parameters studied.

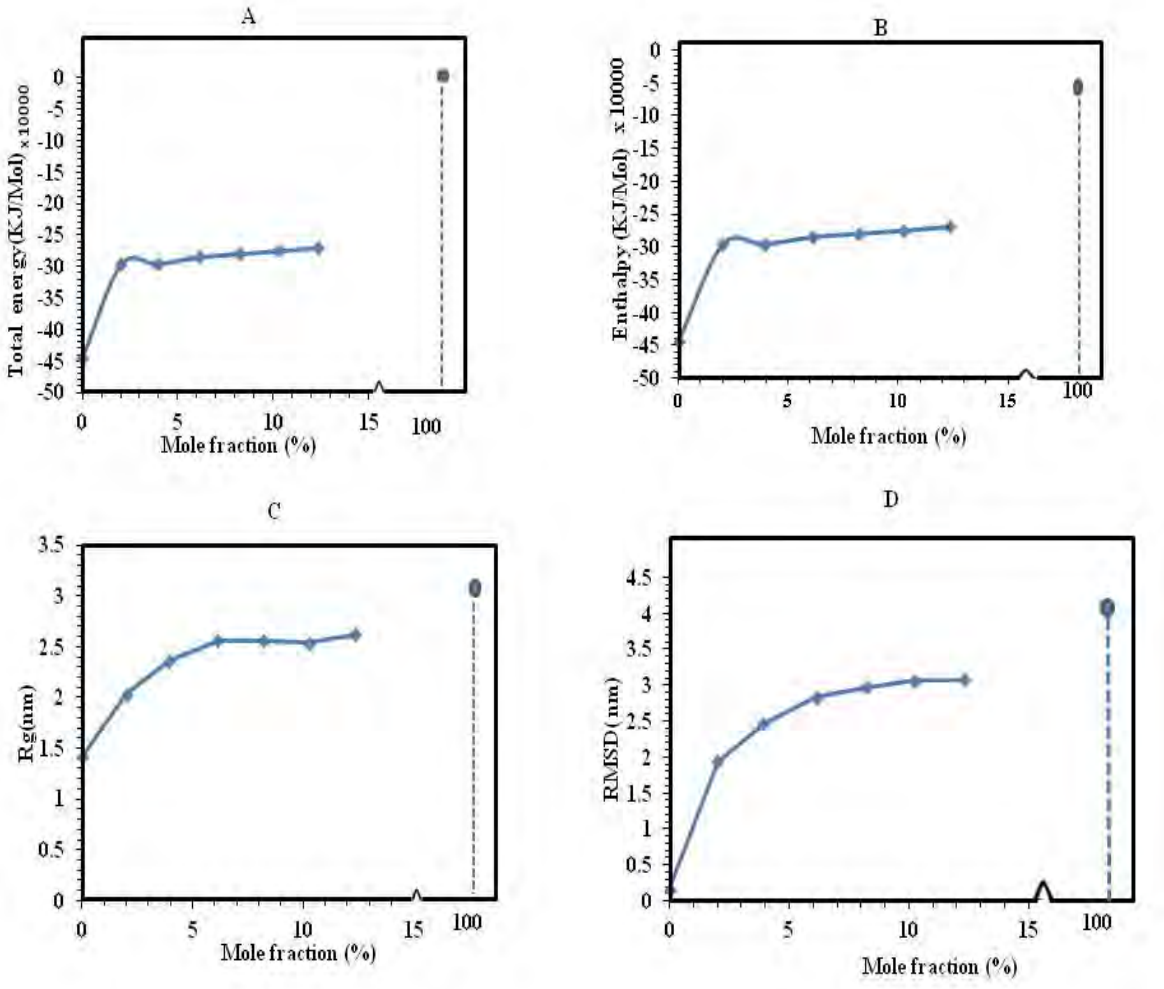

Figure 20. Time average of (A) total energy, (B) enthalpy, (C) radius of gyration and (D) RMSD as a function of mole fraction.

Figure 20 shows the average thermodynamic quantities as a function of mole fraction. From figure 20 we notice a significant increase in all the calculated quantities during the progressive passage from pure water to pure ethanol with higher absolute values at 100% ethanol. These changes further confirm the observed decrease in protein stability with increase in ethanol concentration. The summarized time average values of the thermodynamic quantities of protein in different concentration of ethanol are presented in table 1 and table 2.

Table 1

Time average values for the parameters calculated for our range of percentage.

Parameter	0%	2.01%	3.933%	6.131%
Radius of Gyration(nm)	1.42	2.03	2.36	2.42
Total energy (KJ/Mol)	-443865.80	-296529.60	-295755.50	-284809.80
Enthalpy (KJ/Mol)	-443842.0	-296515.70	-295741.40	-284796.0
RMSD(nm)	0.17	1.94	2.47	2.86

Table 2

Time average values for the parameters calculated for our range of percentages.

Parameter	8.212%	10.262%	12.333%	100%
Radius of Gyration(nm)	2.56	2.54	2.62	3.12
Total energy (KJ/Mol)	-279755.0	-275197.30	-270054.90	-54189.38
Enthalpy (KJ/Mol)	-279741.20	-275183.40	-269361.50	-59108.76
RMSD (nm)	3.00	3.08	3.10	4.17

The present computational analysis took an average of eighty four hours of computing time with thirty six processors for each concentration using GROMACS software on the multi-processor computing system at North Carolina A&T State University (*Hermes*). To obtain the thermodynamic and structural quantities for any new percentages, complete simulation will be required. Such a complete simulation would require significant computing time and resources. However, based on our present analysis, a relatively smooth variation of the time averaged values of the key parameters is clearly noticed. This could allow one to potentially interpolate the

required values from the present generated data for a different ethanol concentration within the range of ethanol concentrations studied. Based on these observations, we propose and present an interpolation methodology for the quantitative key parameters studied as an alternative way by which the need for computer simulation and/or experiments can be avoided. The interpolation approach uses our simulation data and interpolating to obtain the unknown values for another ethanol percentage that is within the range of present study (0 to 12%). The applicability of this interpolation methodology was tested as follows.

1. Select an intermediate percentage that was not used in the simulations. For this purpose, we selected ~7% ethanol concentration.
2. Using an appropriate interpolation method and exiting values in table 5.2a and table 5.2b, calculate the interpolated values of the key parameters at this intermediate range, ~7% ethanol case.
3. Subsequently, complete MD analysis was performed at this concentration level and compared to the interpolated value.

5.3 Using polynomial method to interpolate the intermediate ~7% ethanol

In order to quantitatively determine the behavior of simulated quantities as a function of ethanol concentration, we have used a polynomial fitting as explained below. The following second order polynomial equation was used to fit the simulated quantities,

$$c_2 x^2 + c_1 x + c_0 = f(x) \tag{5.1}$$

where C_0, C_1, and C_2 are constants. The resulting fits for simulation results are shown in appendix F. The values of these constants for equation 5.1 obtained by fitting to the simulation results for all thermodynamic and structural quantities discussed earlier are presented in tables 1 and 2 above.

Table 3

Table of constants for polynomial in equation 5.1 for approximately 7% ethanol.

Quantities	C_0	C_1	C_2
Total energy	-596.247	10980.613	-329718.709
Enthalpy	-596.073	10978.794	-329700.3
Radius of gyration	2.477	0.066	0.009
RMSD	3.669	-0.564	0.058

Using the coefficients C_0, C_1 and C_2 one can determine the above mentioned thermodynamic and structural quantities at any ethanol concentration (from 0-12%), without carrying out the actual simulations. In order to verify this approach, we selected an intermediate ethanol concentration that has not been used in fitting procedure. By using the polynomial equation (equation 5.1) we first theoretically determine the thermodynamics and structural quantities. In the next step, we conducted a complete simulation study as before for the selected ethanol concentration. The simulation results were then compared with that obtained from polynomial interpolation. We have selected ~7% ethanol as our test case, which was not used in the simulations discussed earlier.

The fitted interpolation of simulated quantities is shown in appendix F as a function of ethanol concentration. For the analyzed quantities: total energy, enthalpy, radius of gyration and RMSD. Result for the intermediate percentage of ethanol (~7%) chosen is also plotted in each figure, indicated by the red symbol (appendix F). Table 4 shows the interpolated values obtained for the intermediate concentration of ~7% ethanol using polynomial method.

Table 4

Interpolated values for ~7% ethanol.

Quantity	Interpolated Values
Radius of gyration(nm)	2.51
RMSD(nm)	2.86
Total energy(KJ/Mol)	-283714.93
Enthalpy (KJ/Mol)	-282313.54

5.4 Validating the Results from Interpolation Method

A full MD analysis run at approximately 7% ethanol in ethanol – water mixture was employed to compute the key parameters of energy, enthalpy, *Rg* and RMSD and compared with the corresponding interpolated value at this intermediate percentage. Actual percentage of 7.153 percent by mass of ethanol was employed. The system was set up similar to the other percentages we studied before. The details of the lysozyme-ethanol-water mixture simulation system are presented in appendix G. The calculated thermodynamic quantities and key parameters from the present simulation are presented in appendix H.

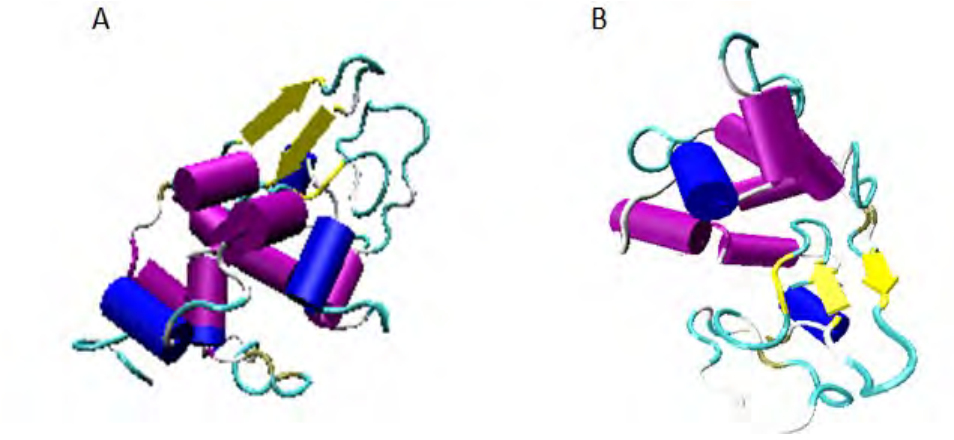

Figure 21. (A) initial and (B) final structure of protein in ~7% of ethanol concentration.

To further validate the interpolation result for 7.153%, we computed the percentage error between the interpolated values and the simulated values of 7.153 as shown in table 5. The values from the interpolated result and the actual simulation are in good agreement. The error margin found to be within the acceptable range. Such good agreement validates the interpolation method for any other percentage within our initial percentage range which is from 0 to 12 percent ethanol. This agreement can potentially avoid the need for computational or experimental procedures for other values of ethanol concentration within the concentration range studied in this work.

Table 5

Simulated results vs Interpolated results for the 7.153% concentration of ethanol in water.

Quantity	Calculated	Simulated	Percentage error
Radius of gyration(nm)	2.51	2.53	0.79
RMSD(nm)	2.86	2.84	0.63
Total energy(KJ/Mol)	-283714.93	-282421.43	0.46
Enthalpy (KJ/Mol)	-282313.54	-276895.71	1.91

5.5 Summary

In this chapter we studied the structure and dynamics of protein at different concentration of ethanol in the range of ~2% - 12% ethanol in ethanol-water mixture. We find that most of thermodynamics and structural quantities show an increase with increasing ethanol concentration. For example, total energy and enthalpies increased systematically with ethanol concentration, thereby revealing the destabilization of folded native protein structure with increasing ethanol concentration. The analysis of structural quantities such as RMSD and radius

of gyration revealed the protein structural deviation from folded state. Together, these results demonstrate a uniform trend of increase in the structural parameters with increasing ethanol concentration. Hence, we proceed to quantify this behavior as a function of ethanol concentration. For this purpose, we obtained optimal polynomial fit for each thermodynamic and structural quantity as a function of ethanol concentration. By using such polynomial expression along with the determined coefficients, we could obtain the results for any arbitrary concentration within the range of 0-12% ethanol. The results of the polynomial interpolated values of the key thermodynamic and structural parameters were compared with those obtained from the full scale MD analysis and are found to be in good agreement.

CHAPTER 6

Concluding Remarks

In this thesis we explored structure and dynamics of lysozyme protein in water, ethanol and water-ethanol binary mixtures by conducting extensive computational molecular dynamics simulation studies. In each case, we performed detailed molecular dynamic simulation and analysis on the following thermodynamic and structural quantities of lysozyme: total energy, enthalpy, radius of gyration and RMSD. MD analysis studies were carried out using GROMACS molecular dynamic simulation code. All production simulation analysis were performed based on a time step of 1fs, pressure of one bar, temperature of 300K for 50 ns of total simulation time. We carried out the simulation process for protein in water environment and performed thermodynamic and structural analysis by calculating the total energy, enthalpy, radius of gyration, and root mean square deviation (RMSD). The results from the analysis showed that protein was relatively stable in water environment, without showing significant deviations from its native folded structure. We proceed to simulate and analyze protein in pure ethanol under the same thermodynamic conditions. At the end of the analysis, we observed a significant change in protein structure between the water environment and the ethanol environment. We find the protein molecule relatively swollen in ethanol solvent compared to that in water environment. To further understand the effect of solvent on protein structure in more systematic fashion, we performed full simulations on different percentages of ethanol-water mixture (2%, 4%, 6%, 8%, 10%, and 12%) and carried out similar thermodynamic analysis as before. We observed changes in protein molecule with increase in ethanol concentration as the protein seems to increase in size based on the visual structure of protein obtained from VMD and thermodynamic analysis. We observed a trend in the variation of the thermodynamic quantities analyzed for the various

ethanol percentages studied. Based on this variation, we proceed to check the possibility of utilizing interpolation method for intermediate percentage within our range of percentages. We chose an intermediate percentage of approximately seven percent for this purpose. With polynomial interpolation method, we were able to calculate the values of the thermodynamic quantities for approximately seven percent using existing data from our previous simulation. In order to verify these results, we constructed a system of lysozyme protein in ~7% ethanol and performed a full scale molecular dynamic simulation. The results from the full simulation were compared with the interpolated results. We found both were in good agreement within the level of acceptable error.

Based on our simulation findings we conclude that the ethanol has a significant effect on lysozyme protein structure. The deviation of protein structure from its native environment suggests that the protein molecule is likely to function better in water environment compared to ethanol environment. We envisage, such molecular level insights into protein-solvent interactions can be used as guidelines in studying similar other protein-solvent interactions as well.

One of the main contributions of this work is that for subsequent percentages within a known range of percentage values, interpolation method can be successfully used to obtain the results without a need for full scale, computationally intensive MD analysis. As demonstrated in the present work, it eliminates need for the large computing time and resources involved in simulations.

6.1 Future Work

In the present work we have simulated and analyzed the dynamics of protein in ethanol-water solvent. With the knowledge acquired from this work, future studies can explore the effect

of other components on proteins and other biological molecules employing the MD analysis methodology. Along these lines, one can study the effect of anesthetic molecules on proteins as well. Despite the routine usage of anesthetic molecules in medical and surgical procedures, this procedure is not devoid of side-effects. Understanding the molecular level interaction between anesthetic molecules and proteins may provide further insights on such side effects. Similar to the study here, nature and concentration of anesthetic molecules may affect the protein structure thereby its function and provide a practical extension of the present work.

References

Astuti, A., & Mutiara, A. (2009). Performance Analysis on Molecular Dynamics Simulation of Protein Using GROMACS. *Arxiv preprint arXiv:0912.0893*.

Balasubramanian, S., & Bagchi, B. (2002). Slow orientational dynamics of water molecules at a micellar surface. *The Journal of Physical Chemistry B, 106*(14), 3668-3672.

Berendsen, H., Postma, J., Van Gunsteren, W., & Hermans, J. (1981). Interaction models for water in relation to protein hydration. *Intermolecular forces, 11*(1), 331-342.

Bowman, G. R., & Pande, V. S. (2009). The roles of entropy and kinetics in structure prediction. *PloS one, 4*(6), e5840.

Branden, C., & Tooze, J. (1991). *Introduction to protein structure* (Vol. 2): Garland New York.

Burkhard, D. G., & Dennison, D. M. (1951). The Molecular Structure of Methyl Alcohol. *Physical Review, 84*(3), 408-417.

Carugo, O., & Pongor, S. (2008). A normalized root-mean-spuare distance for comparing protein three-dimensional structures. *Protein Science, 10*(7), 1470-1473.

Chipot, C., Maigret, B., Pearlman, D. A., & Kollman, P. A. (1996). Molecular dynamics potential of mean force calculations: a study of the toluene-ammonium π-cation interactions. *Journal of the American Chemical Society, 118*(12), 2998-3005.

Comba, P., & Hambley, T. W. (2007). *Frontmatter*: Wiley Online Library.

Dadarlat, V. M., & Post, C. B. (2001). Insights into protein compressibility from molecular dynamics simulations. *The Journal of Physical Chemistry B, 105*(3), 715-724.

Dirac, P. A. M. (1950). Generalized hamiltonian dynamics. *Can. J. Math, 2*(2), 129-148.

Dokholyan, N. V., Li, L., Ding, F., & Shakhnovich, E. I. (2002). Topological determinants of protein folding. *Proceedings of the National Academy of Sciences, 99*(13), 8637-8641.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradovic, Z. (2002). Intrinsic disorder and protein function. *BIOCHEMISTRY-PENNSYLVANIA THEN WASHINGTON-, 41*(21), 6573-6582.

Eastman, P., & Pande, V. (2010). OpenMM: a hardware-independent framework for molecular simulations. *Computing in Science & Engineering, 12*(4), 34-39.

García, A. E., & Onuchic, J. N. (2003). Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proceedings of the National Academy of Sciences, 100*(24), 13898-13903.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics, 14*(1), 33-38.

Hunt, N. T., Kattner, L., Shanks, R. P., & Wynne, K. (2007). The dynamics of water-protein interaction studied by ultrafast optical Kerr-effect spectroscopy. *Journal of the American Chemical Society, 129*(11), 3168-3172.

Jorgensen, W. L., & Jenson, C. (1998). Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry, 19*(10), 1179-1186.

Karolewski, M. (2005). Kalypso: a software package for molecular dynamics simulation of atomic collisions at surfaces. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 230*(1), 402-405.

Kollman, P. A., & Merz Jr, K. M. (1990). Computer modeling of the interactions of complex molecules. *Accounts of Chemical Research, 23*(8), 246-252.

Korosec, L., Limacher, P. A., Luthi, H. P., & Brandle, M. P. (2010). Chemical Information Media in the Chemistry Lecture Hall: A Comparative Assessment of Two Online Encyclopedias. *CHIMIA International Journal for Chemistry, 64*(5), 309-314.

Lehninger, A. L., Nelson, D. L., & Cox, M. M. (2005). *Lehninger principles of biochemistry* (Vol. 1): WH Freeman & Co.

Lemak, A., & Balabaev, N. (1994). On the Berendsen thermostat. *Molecular Simulation, 13*(3), 177-187.

Levitt, M., & Warshel, A. (1975). Computer simulation of protein folding. *Nature, 253*(5494), 694.

Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling, 7*(8), 306-317.

Lobanov, M. Y., Bogatyreva, N., & Galzitskaya, O. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology, 42*(4), 623-628.

Müller-Plathe, F. (1993). YASP: A molecular simulation package. *Computer physics communications, 78*(1-2), 77-94.

Newcomer, M., Lewis, B., & Quiocho, F. (1981). The radius of gyration of L-arabinose-binding protein decreases upon binding of ligand. *Journal of Biological Chemistry, 256*(24), 13218-13222.

ntroduction to Molecular Mechanics. *www.tau.ac.il/~ephraim/MM+MD.doc*.

Onori, G., & Santucci, A. (1996). Dynamical and structural properties of water/alcohol mixtures. *Journal of molecular liquids, 69*, 161-181.

Pal, S. (2005). *School of Engineering and Science.* Max Planck Institute.

Pal, S., Weiss, H., Keller, H., & Müller-Plathe, F. (2005). Effect of nanostructure on the properties of water at the water-hydrophobic interface: a molecular dynamics simulation. *Langmuir, 21*(8), 3699-3709.

Ponder, J. W., & Case, D. A. (2003). Force fields for protein simulations. *Advances in protein chemistry, 66*, 27-85.

Preedy, V., & Peters, T. (1988). Acute effects of ethanol on protein synthesis in different muscles and muscle protein fractions of the rat. *Clinical science (London, England: 1979), 74*(5), 461.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure.

Rubio, C. A., & Befrits, R. (2009). Increased lysozyme expression in gastric biopsies with intestinal metaplasia and pseudopyloric metaplasia. *International journal of clinical and experimental medicine, 2*(3), 248.

Schuler, L. D., Daura, X., & Van Gunsteren, W. F. (2001). An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry, 22*(11), 1205-1218.

TIERNAN, J. M., & WARD, L. C. (1986). Acute effects of ethanol on protein synthesis in the rat. *Alcohol and Alcoholism, 21*(2), 171-179.

Turner, A. J., Moliner, V., & Williams, I. H. (1999). Transition-state structural refinement with GRACE and CHARMM: Flexible QM/MM modelling for lactate dehydrogenase. *Phys. Chem. Chem. Phys., 1*(6), 1323-1331.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry, 26*(16), 1701-1718.

van der Waal, J. C., & van Bekkum, H. (1997). Zeolite titanium beta: A versatile epoxidation catalyst. Solvent effects. *Journal of Molecular Catalysis A: Chemical, 124*(2), 137-146.

Wang, J. H. (1954). Theory of the self-diffusion of water in protein solutions. A new method for studying the hydration and shape of protein molecules. *Journal of the American Chemical Society, 76*(19), 4755-4763.

Wensink, E. J. W., Hoffmann, A. C., van Maaren, P. J., & van der Spoel, D. (2003a). Dynamic properties of water/alcohol mixtures studied by computer simulation. *The Journal of chemical physics, 119*(14), 7308-7317.

Wensink, E. J. W., Hoffmann, A. C., van Maaren, P. J., & van der Spoel, D. (2003b). Dynamic properties of water/alcohol mixtures studied by computer simulation. *The Journal of chemical physics, 119*, 7308.

Yanni, N. P. C. B. W. Y. W. J. L. (2012). Analysis and Simulation of Molecular Dynamics of Lysozyme in Water Cluster System. *DOI 10.1007/s12209-012-1775-9, Vol.18 No.1 2012*(DOI 10.1007/s12209-012-1775-9), 001-007.

Zaks, A., & Klibanov, A. M. (1988). The effect of water on enzyme action in organic media. *Journal of Biological Chemistry, 263*(17), 8017-8021.

Zhou, D. R.). Molecular Dynamics Analysis of Protein Dynamics Retrieved 08/24/2012, 2012, from http://www.nd.edu/~gezelter/Teaching/650-Fall-2005/exercises/gromacs/index.html

Appendix A

Table

System information for lysozyme in water environment simulation.

System Parameters	Parameter Description
Protein	Lysozyme
Force field	OPLS
Temperature	300K
Pressure	1 bar
Potential used	Lernnard Jones
Thermostat	Berendsen
NPT,NVT Equilibration	200,000 steps (200ps)
MD run	50,000,000 steps (50ns)
Protein molecule	1 molecule
Protein residue	129 residues
Protein atoms	1,960 atoms
Water molecule	12,365 atoms
Water atoms	37,095 atoms
Molecular weight of water	222,570 g/mol
Total atom of the system	39,055 atoms
System size	(7.33925*7.33925*7.33925)(nm)
Number of processors used	36 processors

Appendix B

Table

System information for ethanol environment simulation.

System Parameters	Parameter Description
Protein	Lysozyme
Force field	OPLS
Temperature	300K
Pressure	1 bar
Potential used	Lernnard Jones
Thermostat	Berendsen
Time step	1fs
NPT,NVT Equilibration	200 ps
simulation length	50,000,000 steps (50ns)
Protein molecule	1 molecule
Protein residue	129 residue
Ethanol atoms	20,601 atoms
Ethanol molecules	2,289 molecules
Molecular mass of ethanol	46.06844 g/mole
Molecular weight of ethanol	46,022.4 g/mole
Total system atom	22,561 atoms
System size	(9.96426*9.96421*9.96426)(nm)
Number of processors used	36 processors

Appendix C

Table

Lysozyme – Ethanol – Water System information for approximately 2%, 4%, 6%, 8%, 10% and 12% ethanol concentration in water.

Ethanol percentage	Protein residues	Protein atoms	Water molecules	Water atoms	Molecular weight of water(g/mol)	Ethanol molecule	Ethanol atoms
2.01	129	1960	6977	20931	125586	56	504
3.933	129	1960	6928	2078	124704	111	999
6.131	129	1960	6612	19836	119016	169	1521
8.212	129	1960	6455	19365	116190	226	2034
10.262	129	1960	6302	18906	113436	282	2538
12.333	129	1960	6140	18420	110520	338	3042

Appendix D

Table

Lysozyme – Ethanol – Water System information for approximately 2%, 4%, 6%, 8%, 10% and 12% ethanol concentration in water.

Ethanol percentage	Protein molecule	Molecular weight of ethanol(g/mol)	Total system atoms	System size (nm)	Number of processors used
2.01	1	2576	23395	$(6.13121)^3$	36
3.933	1	5113	23743	$(6.13121)^3$	36
6.131	1	7774	23317	$(6.16800)^3$	36
8.212	1	10396	23359	$(6.16800)^3$	36
10.262	1	12972	23404	$(6.16800)^3$	36
12.333	1	15548	23422	$(6.16800)^3$	36

Appendix E

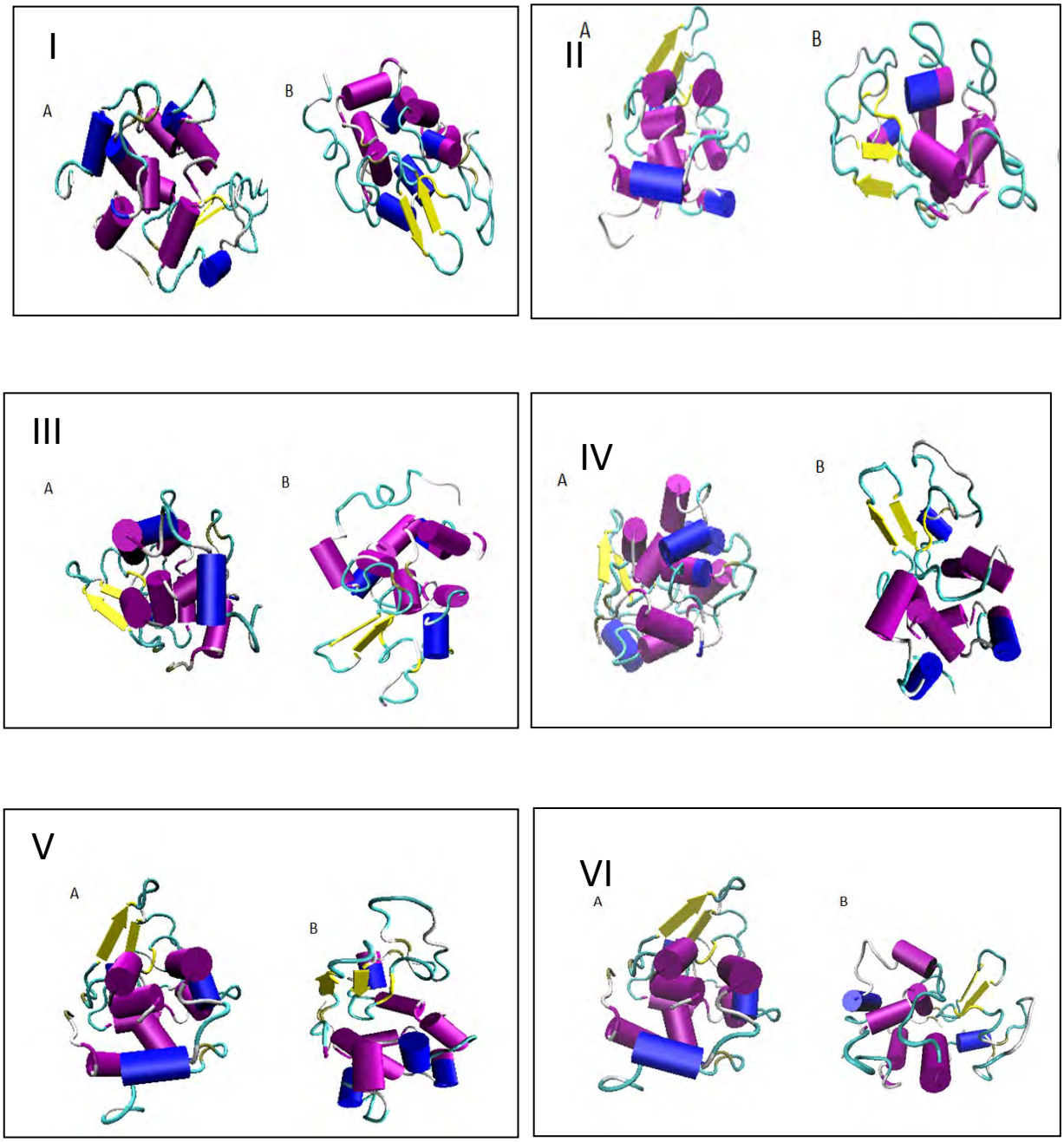

Figure (A) initial and (B) final structure of protein in ethanol-water mixture in different concentrations of ethanol approximated at I) 2%, II) 4%, III) 6%, IV) 8%, V) 10% and VI) 12%.

Appendix F

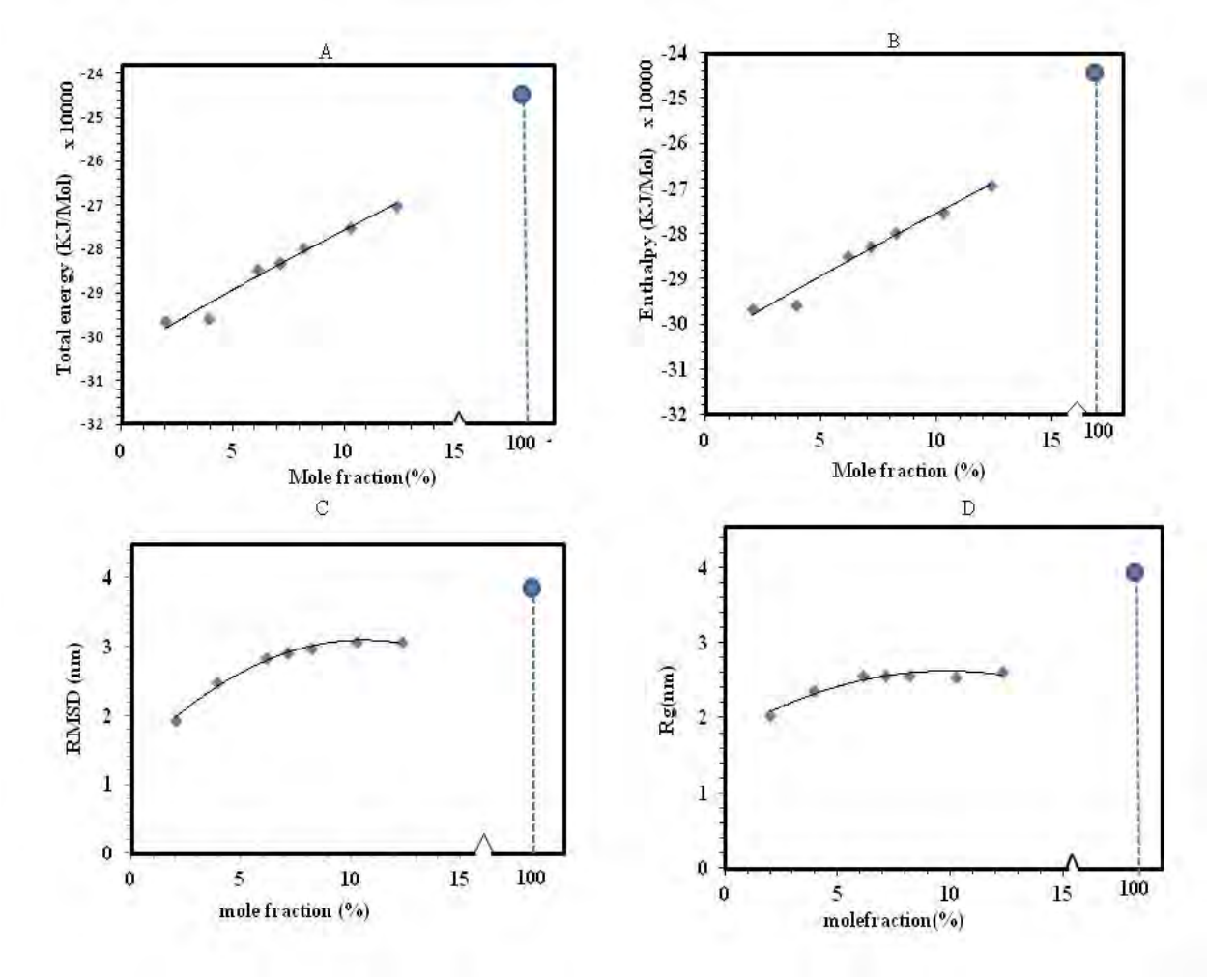

Figure Interpolated (A) total energy, (B) enthalpy, (C) radius of gyration and (D) RMSD for approximately 7% of ethanol (in each figure red symbol corresponds to ~ 7% ethanol case).

Appendix G

Table 1

Molecular system information for ~ 7% ethanol concentration.

Ethanol percentage	Protein molecule	Molecular weight ethanol(g/mol)	Total atoms	System size (nm)	Processors used
7.153	1	9062	23338	$(6.16800)^3$	36

Table 2

Molecular system information for ~ 7% ethanol concentration.

Protein residues	Protein atoms	Water molecules	Water atoms	Molecular weight of water	Ethanol molecule	Ethanol atoms
129	1960	6535	19605	117630	197	1773

Appendix H

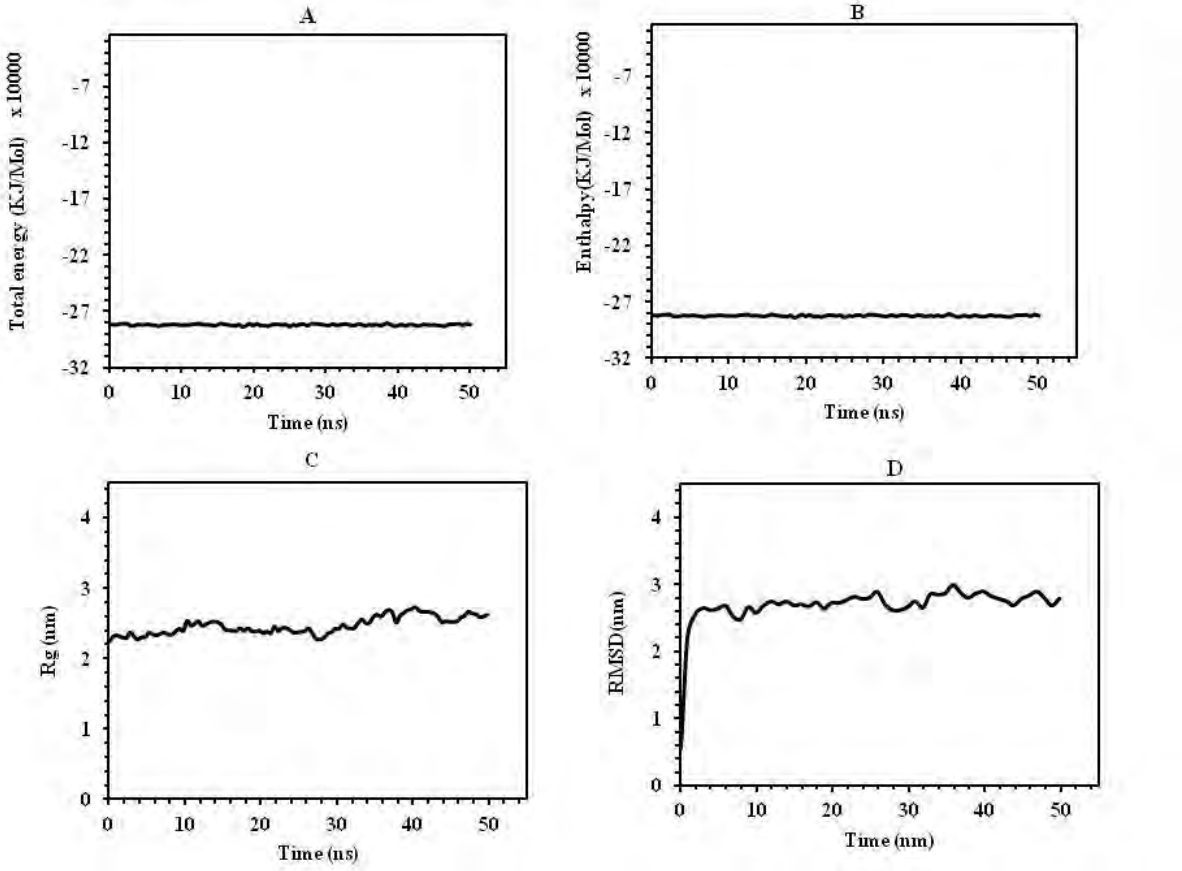

Figure (A) total energy, (B) enthalpy, (C) radius of gyration and (D) RMSD for protein in ~ 7% ethanol as a function of time.