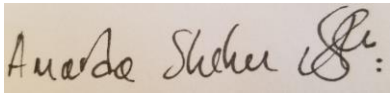
POLYMERS AND BIOMOLECULES IN SOLVENTS:
A MOLECULAR DYNAMICS STUDY

by

Gideon Kwadzo Gogovi
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Computational Sciences and Informatics

Committee:

Dr. Jason Kinser, Committee Chair

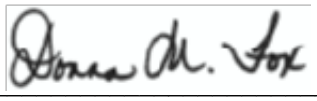


Dr. Amarda Shehu, Committee Member

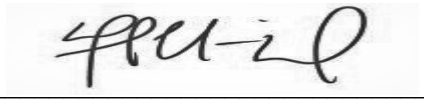
Dr. Hamdi Kavak, Committee Member

Dr. Dimitrios A. Papaconstantopoulos, Committee Member

Dr. Jason Kinser, Department Chair



Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science



Dr. Fernando R. Miralles-Wilhelm, Dean, College of Science

Date: 11/20/2020

Fall Semester 2020
George Mason University
Fairfax, VA

Polymers and Biomolecules in Solvents: A Molecular Dynamics Study

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Gideon Kwadzo Gogovi
Master of Science
West Virginia University, 2016
Bachelor of Science
University of Cape Coast, 2010

Director: Dr. Jason Kinser, Associate Professor
Department of Computational and Data Sciences

Fall 2020
George Mason University
Fairfax, VA

Dedication

To my uncle Walter C.K Dogli, my mother Gladys Kutiame and my late grandmother, Awusi Klokpa

Acknowledgments

To God, the Father, the giver and maker of vision. Thank you for a dream come true. I would like to express my heartfelt appreciation and gratitude to my advisor Dr. Jason M. Kinser for his support. His inexhaustible supply of help, advice and most importantly patience has been invaluable in the completion of this dissertation. I would also like to thank Dr. Estela Blaisten who has been instrumental in my development as a competent researcher. In addition, I wish to thank Dr. Amarda Shehu for her advice, constant support and most importantly, for believing in me right from the beginning of this program. To Dr. Hamdi Kavak and Dr. Dimitrios Papaconstantopoulos, I say thank you very much for agreeing to be members of my dissertation committee when I needed you most.

I wish to thank my friends, Ajay S. Kulkarni and Dr. Swabir Silayi, who provided many light-hearted moments during our time together as graduate students. I should especially acknowledge all the faculty and staff of the Department of Computational and Data Sciences, George Mason University. A special thanks to Natalie V Lapidot-Croitoru and Karen Underwood who have been a fountain of motivation and great friends throughout the years.

I would also like to acknowledge funding from the Department of Computational and Data Science, the Commonwealth of Virginia 4-VA grant and from the George Mason University (MDR Seed Grant No. 215001) that supported my study.

Finally, I would like to thank my uncle Mr. Walter C.K Dogli for all of the time, and effort on getting my dream in education come true. Without his help I would not have managed to get to where I am today. God bless you Mr. Walter C.K Dogli.

Table of Contents

List of Tables

List of Figures

Abstract

POLYMERS AND BIOMOLECULES IN SOLVENTS: A MOLECULAR DYNAMICS STUDY
Gideon Kwadzo Gogovi, PhD
George Mason University, 2020
Dissertation Director: Dr. Jason Kinser

This study, which is in three parts, uses all-atom Molecular Dynamics (MD) simulations to investigate the behavior of polymers and biomolecules in solvents to enhance the understanding of solvent effects on protein-solvent and polymer-solvent interactions. In the first, the structural, energetic, and dynamical properties of polyacrylamide (PAM) oligomers of different lengths solvated in pure glycerol, glycerol–water, and pure water are investigated. We predict that the oligomers' globular structure is obtained only when the modeling strategy considers the solvent as a continuous background. Meanwhile, for all-atom modeled solvents, the glycerol solutions display a strong tendency of trapping the oligomers in instantaneous elongated random coiled structures that remain locked-in over tens of nanoseconds. In pure water, the oligomers acquire considerably shorter random coiled structures of increased flexibility. The generalized amber force field is modified by including restrained electrostatic potential atomic charges for glycerol and PAM. Three PAM oligomer lengths monomers are considered in detail by monitoring the structural properties and energetics for several nanoseconds. The density and radial distribution function of glycerol solutions are calculated when modeled with the modified atomic charges and shows very good agreement with experimental results at temperatures around 300 K. Glycerol has multiple

applications, including its use in gel formation for PAM gel electrophoresis. Our findings are relevant for the design of sensors based on microfluidics and tailored pharmaceutical buffer solutions.

The second part presents a solvation effect of the structure and dynamics of a C-terminal domain of Rift Valley Fever Virus (RVFV) L protein exploration by MD using both explicit and implicit water. The force field parameters of explicit waters were taken from the TIP3P, TIP4P, SPC/E, SPCE/Fw, and OPC water models.

The generalized Born (GB) model was employed for the implicit solvent simulation. The results from the study led to the conclusion that the structural conduct and preference of this protein are highly sensitive to the accommodating environment. Also, structural characterization and clustering of the atomic trajectories enable a better understanding of the structural and dynamical behavior of the peptide along time.

In the third part of this investigation, the structural dynamics and energetic properties of the C-terminal domain of the RVFV L protein in glycerol and its aqueous solutions at different concentrations by molecular weight are presented. Secondary structure analysis was also performed to examine the extent of conformational drift for the individual α-helices and β-sheets. It is reasonable to predict from the results that the helices and sheets are maintained only when the modeling strategy considers solvents with less glycerol concentration and also, the solvent-peptide becomes more cohesive with decreasing glycerol concentrations.

Chapter 1: Introduction

Proteins or polypepetides comprise of amino acids that are transported to the ribosome by Transfer RNAs (tRNAs). Each amino acid contains an amino group ($-NH_2$) and carboxyl group ($-COOH$), and these two moieties are connected directly to the central C_α. A specific side chain ($-R$), called the R group, is also connected to the C_α to make each amino acid different. There exist 20 naturally occurring amino acids in organisms. For the tRNA to obtain its cognate amino acid, the different tRNAs are specifically recognized by their connected aminoacyl-tRNA synthetases (AARS) that catalyze the reaction to attach amino acids to their cognate tRNA's 3' end, resulting in a "charged" state. The transport proteins' elongation factor thermo unstable (EF-Tu) binds to aminoacylated, or "charged", tRNA and moves it to a free A-site on the ribosome and ensures the association of the correct tRNA anticodon with the mRNA codon with a good presicion. The ribosome uses the mRNA as the template and each tRNA anticodon recognizes the codon. The tRNA provides the amino acids one after the other until the protein synthesis is complete. During the synthesis, the amino and carboxyl groups of two amino acids can form the peptide bond ($-NH-CO-$) through a condensation reaction.

The process is repeated such that many amino acids are connected one after another, and they form a sequence. The sequence of amino acid residues in a protein is called the *primary structure*. The polypeptides are mainly connected through hydrogen bonds and form the local *secondary structures* like α - helices and β - sheets. The secondary structure and regular geometry of the segments of a protein depend upon the following factors; the bond length and bond angles of the peptide bond, the coplanar arrangement of the amide groups atoms, the hydrogen bonds between $N-H$ and $C=O$ to maintain the maximum stability, and the range of the distance in the hydrogen bonds. Segments of the peptide

chains are most likely to be held in their coiled form due to intramolecular forces. The two best-known coils are the α-helix and β-sheet. The directions of the hydrogen bonds are almost parallel to the axis of the helix. The left- and right-handed α-helices correspond to the turns of a left-handed screw and a right-handed screw. The helices do not need to have an integral number of residues per turn. In many proteins, the α-helix repeats after exactly 18 residues, which amounts to five turns. Also, each carbonyl oxygen is bonded to the amide proton on the fourth residue up the helix by a hydrogen bond. A β-sheet (β-structure or pleated sheet) has a twist and the whole structure follows this twist, resulting in an unsteady arrangement for the structural elements in the outer layers. β-Sheets can also curl and form a complete hydrogen-bonded cylinder (usually referred to as a "barrel"). A β-sheet may be antiparallel, with chains that run in the opposite direction, as if it is folded back and forth upon itself, or it may be made from segments of a chain that are looped back to run in the same direction, thus creating a parallel sheet.

The side chains of amino acids may interact in a more complex manner to form a *tertiary structure*. This is the overall folding of the protein molecules, in contrast to the secondary structure, which is the local folding (e.g., α-helix, β-sheets). The tertiary structure makes the protein compact and globular in shape. It may be divided into units called domains with a simple domain containing 100–150 amino acid residues and is about 25 Å in diameter. Under appropriate conditions, a domain can sometimes be isolated as a fragment by limited proteolysis. In the native protein molecule, each fragment has the same conformation. A fragment is also stable and can be refolded without apparent external cause from the unfolded state under native conditions.

The tertiary structure may further fold into a *Quaternary structure*. While tertiary structure refers to the topology of one polypeptide chain of a protein, the quaternary structure refers to the topology of several polypeptide chains aggregated together. The aggregation can be separated by using an external force such as ultra centrifuge. This shows that the interpeptide chain attraction is neither strong nor weak. It is not strong because it can be easily separated and it is not weak because it sticks together to form an assembly.

Understanding the cellular processes of living organisms and how biomolecules fold into specific structures to carry out these processes to result in a functioning organism is a major objective of biophysics. Figure 1.1 below depicts the protein structure characterization

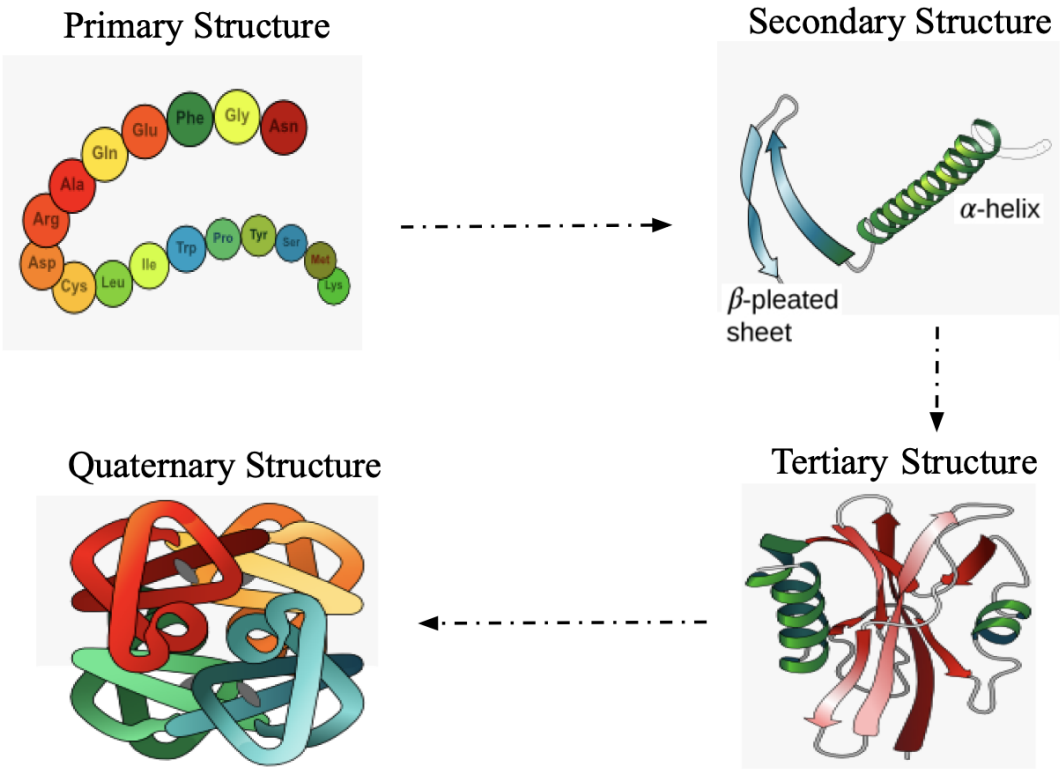


Figure 1.1: Protein Structure levels.

Polymers, also referred to as macromolecules, are very large molecules with high molecular mass and are formed by joining of repeating structural units. The repeating structural units are derived from some simple and reactive molecules called monomers and are linked to each other by covalent bonds through polymerization. Polymers form a very important class of materials without which life would be very difficult. The macroscopic physical properties of polymer containing materials are often dependent on the average microscopic conformation of the constituent polymer molecules. Polymers such as proteins carry out the most difficult tasks in living cells by interacting with specific molecules. This requires

that they fold to a specific, globular conformation that is only marginally more stable than the large ensemble of the unfolded state.

The conformation of a polymer chain in solution is coupled to the local structure of the surrounding solvent and can undergo large changes in response to variations in solvent density and temperature [4]. The correct assessment of the solvent's effect on local conformational dynamics might be critically important for the interpretation of experiments such as NMR spectroscopy, fluorescence anisotropy, high-frequency dielectric relaxation, etc., measuring characteristic relaxation times associated with short-range orientational motions in polymers.

Electron paramagnetic resonance (EPR) measurements showed that motions in the local neighborhoods of two spin-labeled amino acids decreased dramatically with decreasing solvent dielectric constant, a trend with is consistent with changes in the electrostatic force between charged residues of the protein. In general, at the lower dielectric constant many atoms of the protein move more slowly, and many of the slowest residues are near the exterior [5]. This suggests that the dielectric constant and other characteristics of the solvent play an important role in determining the properties of proteins and other biomolecules in organic media. Simulation of polymer-solvent systems is a useful tool to investigate the behavior of polymers in different environments [6].

While there is a wealth of experimental methods that are capable of determining information regarding the interfacial structure of molecules, it is often necessary to employ computational methods to support the conclusions arrived at by experimental procedures. This necessity is compounded by the fact that while the majority of experimental techniques determine information based on an ensemble average of molecules at the surface, computational techniques are capable of studying systems using discrete molecules. This makes the two techniques complementary to each other. However, even with a relatively small number of molecules in a solution interacting with each other, these calculations can become relatively long, requiring great deals of computational time to complete. While

these problems are well known and methods to overcome these difficulties have been developed, these techniques still face considerable challenges [7]. As such, many methods have been developed that can reduce the amount of time required to study any system.

Molecular mechanics is one such method. It relies upon empirically parameterized equations that are capable of modeling physical forces in the system of interest. Within the realm of molecular mechanics, there exist a variety of sub-methodologies, such as Monte Carlo, as well as molecular dynamics, which employs Newtonian motion to study the forces and energy that exist in a system of more than two particles. Molecular dynamics simulation consists of the step-by-step numerical solution of the classical equations of motion in Hamilton's dynamics or the equivalent Newton's dynamics. Given an initial state and having knowledge of the forces of nature, the idea behind molecular simulations is that one should be able to compute the behavior of the system [8].

The general objective of this dissertation is to improve the understanding of the structural and dynamical behavior of polymers and proteins in dense solvents with varying concentrations. This work employs molecular dynamics to study structural and dynamic properties of polymers of different lengths and a biomolecule in glycerol and its aqueous solutions. Specifically, polyacrylamide (PAM) and a Rift Valley Fever Virus (RVFV) L protein domain is studied in glycerol, water and aqueous glycerol solutions of different proportions by molecular weights. The solvent component of the dissertation employed high-density solvents that are also structural protectant of proteins. Glycerol is known to be a good protectant of secondary structures of proteins [9, 10] and favors the secondary structural formation and inhibits aggregation of other proteins such as creatine kinase [11]. The dissertation is organized in six (6) chapters.

Chapter 2 contains a detailed overview of the computational method, molecular dynamics, and also reviews some of the structural and dynamic properties of the molecules studied in this dissertation. The Assisted Model Building with Energy Refinement (AMBER) molecular dynamics software and the force field or model potential used for the simulations are also introduced in this chapter. I discuss the potential equations that AMBER utilizes

for molecular dynamics which reveal the force field parameters that must be developed for the molecules. Chapter 3 presents a molecular dynamics exploration of the structural and thermodynamic properties of polyacrylamide. Different lengths of polyacrylamide in glycerol, a mixture of glycerol-water (90-10% by molecular weight) and in water is studied and presented in this chapter.

The behavior of a C-terminal domain of Rift Valley Fever Virus (RVFV) L protein in different water models using molecular dynamics simulation is presented in Chapter 4. Here, five different water models, TIP3P, TIP4P, SPC/E, SPCE/Fw and OPC were selected for this study. In addition to these water models an implicit solvent simulation was also performed as a comparative study. In Chapter 5, an extended study this C-terminal domain of the Rift valley fever virus L protein in aqueous glycerol solutions is presented. Ten solvents, pure glycerol, and mixtures of glycerol:water (at 90:10%, 80:20%, 70:30%, 60:40%, 50:50%, 40:60%, 30:70%, 20:80%, 10:90%) concentrations in molecular weights are considered for this study. Finally, Chapter 6 contains concluding remarks from the previous chapters and a proposed future study.

Chapter 2: Theory and Computational Methods

2.1 Molecular Dynamics Simulation and Atomistic Empirical Force Field

This chapter presents a detailed overview of molecular dynamics (MD), the Assisted Model Building with Energy Refinement (AMBER) molecular dynamics software, and the AMBER force field or model potential used for the simulations in this dissertation. The chapter also presents some of the structural and dynamic properties of molecules studied in this work.

2.1.1 Molecular Dynamics Algorithm

Molecular dynamics simulation consists of the numerical solution of the classical equations of motion in Hamilton's dynamics or the equivalent Newton's dynamics. Molecular dynamics has proven to be a valuable tool for understanding the mechanism and evolution of several time-dependent processes, to reproduce well thermodynamic properties, and is of paramount importance for the determination of the structure of systems under different realistic thermodynamic conditions [12].

For a simple atomic system, the Newton second-order ordinary differential equations may be written as

$$\mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = m_i \frac{d^2\mathbf{r}_i}{dt^2} \tag{2.1}$$

$$\mathbf{F}_i = \sum_{j \neq i}^{N} \mathbf{f}_{ij} \tag{2.2}$$

with

$$\mathbf{f}_{ij} = -\frac{dU(r_{ij})}{dr_{ij}} \cdot \frac{\mathbf{r}_{ij}}{r_{ij}} \tag{2.3}$$

Here $\mathbf{r}_i$ are the position vectors and $\mathbf{F}_i$ are the forces acting upon the N particles in the system. According to Newton's third law, $\mathbf{f}_{ij} = -\mathbf{f}_{ji}$. A system composed of atoms with coordinates $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$ and potential energy $U(r^N)$, we introduce the atomic momenta $\mathbf{p}^N = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N)$, in terms of which the kinetic energy may be written $K(\mathbf{p}^N) = \sum_{i=1}^{N} |\mathbf{p}_i|^2/2m_i$. Then the energy, or Hamiltonian, may be written as a sum of kinetic and potential terms $\mathcal{H}(q,p) = K(p)+U(q)$. Write the classical equation of motion as;

$$\begin{aligned} \frac{d\mathbf{p}}{dt} &= -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} = \mathbf{F}_i \\ \frac{d\mathbf{q}}{dt} &= +\frac{\partial \mathcal{H}}{\partial \mathbf{p}} = \mathbf{p}_i/m_i \end{aligned} \tag{2.4}$$

2.1.2 Time Integration of the Equations of Motion

The most time-consuming component of an MD calculation is the evaluation of the forces. Several methods are available for the numerical integration of the equations of motion. To do this, the second-order differential equation of motion is rewritten as two coupled first order differential equations, as seen in Equation (2.4). Equation (2.5) demonstrates this for the y-component:

$$\begin{aligned} F_y &= m\ddot{y} \\ \dot{y} &= v \\ \dot{v} &= \frac{1}{m} F_y \end{aligned} \tag{2.5}$$

where v is the velocity and $\dot{v} = \ddot{y}$.

The simplest method for solving such a system of differential equations is Euler's method.

One just steps forward from t_n to t_{n+1} using the derivative information from Equation (2.1).

$$\begin{aligned} y_{n+1} &= y_n + \Delta t \cdot \dot{y} \\ v_{n+1} &= v_n + \Delta t \cdot \dot{v} \end{aligned} \tag{2.6}$$

Though this is simple, it is not used as it is only first-order accurate and does not produce better results. The method can be derived from the Taylor Series, which also shows that it is first-order accurate. There are a wide range of another method called Runge-Kutta methods with different accuracy. Fourth-Order Runge-Kutta is the most known with very good accuracy. The Runge-Kutta calculates the derivative several times along the interval but taking the final step from the start of the interval. The step and the k-coefficients are shown in Equation (2.7) for advancing y.

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6} h \left(k_1 + 2k_2 + 2k_3 + k_4 \right), \\ t_{n+1} &= t_n + h \\ k_1 &= f(t_n, y_n), \\ k_2 &= f\left(t_n + \frac{h}{2}, y_n + h\frac{k_1}{2} \right), \\ k_3 &= f\left(t_n + \frac{h}{2}, y_n + h\frac{k_2}{2} \right), \\ k_4 &= f\left(t_n + h, y_n + hk_3 \right), \end{aligned} \tag{2.7}$$

where y_{n+1} is the Runge-Kutta approximation of $y(t_{n+1})$, and the next value (y_{n+1}) is determined by the present value (y_n) plus the weighted average of four increments, where each increment is the product of the size of the interval, h, and an estimated slope specified by function f on the right-hand side of the differential equation.

The Velocity Verlet algorithm is yet another method that is no more complex than the

first-order Euler method but yields second-order results. The step in time in the y-direction is given by Equations

$$\begin{aligned} y_{n+1} &= y_n + v_n \Delta t + \frac{1}{2} a_n \Delta t^2 \\ v_{n+1} &= v_n + \frac{a_n + a_{n+1}}{2} \Delta t \end{aligned} \tag{2.8}$$

where a_n is the acceleration in the y-direction for the n^{th} time step. The need for acceleration at time step $n+1$ for the velocity is extra work since it requires recalculating all of the forces on each atom. However, these forces are conservative and only depend on the atomic positions. The next step also involves the computation of the accelerations so, the positions, y, are updated first. Then, the next step's accelerations are computed from the new positions and then, the velocities may be updated. At this point, the next step's accelerations are already computed so no additional force calculations are done for this second-order method.

The interest here is in the behavior of systems that involve large numbers of atoms. Due to the computing and time constraints for simulating such systems, it is preferable to run simulations of a representative system of a few molecules in isolation. Periodic boundary conditions help approximate the larger systems and this eliminates the constraints.

Periodic boundary conditions help us to deal with these constraints. A rectangular computational box is filled with solvent molecules or solution with molecules of interest and equilibrated so that the box attains a density and composition desired. This box is treated as if its image is repeated in all directions. Atoms in the box appear in all the other boxes and behave identically to its peer in the other imaged boxes. This makes for an infinite number of atoms which cannot be simulated directly. A cutoff distance is set so that when the atoms move further than the value of this distance, direct interaction is not computed between them. Moreover, the potential energy function describing the forces between atoms are usually shifted so they are zero at the cutoff distance. This has no effect on the force calculations and takes care of the functions' discontinuity. Finally, a method for computing long-range interactions is selected. The Particle Mesh Ewald method is an example of how

long-range electrostatics can be computed very quickly in Fourier space.

In the canonical ensemble (NVT), sometimes called constant temperature molecular dynamics, number of atoms N, volume V and temperature T are held constant or conserved. The energy of endothermic and exothermic processes is exchanged with a thermostat in NVT. Controlling the temperature of a system in simulations is important. A variety of thermostat algorithms are available to remove or add energy from the boundaries of an molecular dynamics simulation. Popular methods to control temperature include velocity rescaling, the Nosé–Hoover thermostat, the Berendsen thermostat, the Andersen thermostat and Langevin dynamics.

Isothermal–isobaric ensemble (NPT) has the number of atoms N, pressure P and temperature T are held constant or conserved. In addition to a thermostat as mentioned in the NVT case, a barostat is also needed in NPT for pressure control. The simplest of these is Berendsen pressure coupling which is just an exponential relaxation method. This method is not time-reversible and it has been argued that this method does not yield a correct thermodynamic ensemble, especially in the simulation of biological membranes. However, it is efficient so it is usually used to equilibrate a computational box at the beginning of a simulation. This has higher computational requirements but is recommended if pressure coupling continues through data collection.

Microcanonical ensemble (NVE) simulations conserve the number of atoms N, the volume V, and the total energy, E, constant. The temperature and the volume are not adjusted and the energy here is supposed to remain constant as conservation laws dictate. This type of simulation works well after a simulation box has been well equilibrated with NPT or NVT ensembles.

Generally, two groups of properties, static and dynamic, can be determined with molecular dynamics. Static properties are properties that can be calculated directly with single configurations. Simple properties are found directly from the trajectory data and are usually calculated during the simulation. These properties include potential and kinetic energy, and as a result, the total energy and temperature.

The radius of gyration of the individual molecules indicates the compactness of these molecules, such as polymers and biomolecules. Others include the end-to-end distance, the distance between the centers of mass of the two end end-monomers of a coiled polymer chain. This is usually measured as the distance between the center of mass of the two end monomers in a polymer or the two end residues in a protein or a peptide. The radial distribution function, $g(r)$, which is key to identifying crystal structures is used to identify the distribution of atoms or molecules in a simulation box.

Most of these dynamical properties are usually calculated based on other static properties. For example, the self-diffusion coefficient which can be computed with a single particle, and autocorrelation functions. For this, one needs a trajectory of configurations, which is exactly what molecular simulations produce.

2.1.3 AMBER Model Potential and Parameter Sets

This dissertation employs the Gaff force field parameters for the polymers and molecular solvents and the Assisted Model Building with Energy Refinement (Amber) ff14SB force field for the proteins structural study. These force fields are known to model well polymers, molecular solvents [13], and proteins [14] with the appropriate parameters. The functional form of the AMBER force field uses the following Potential function

$$\begin{aligned} V(r^N) = & \sum_{\text{bonds}} k_b(l - l_0)^2 \\ & + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{B_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \end{aligned} \tag{2.9}$$

The terms k_b, l_0, k_θ, θ_0, V_n, γ, A_{ij}, B_{ij} are parameters to be specified based on the various Amber force fields. The first term on the right of equation (2.9) is the harmonic term for bond stretching. It represents the energy between covalently bonded atoms with force constant k_b, instantaneous bond lengths, l and l_0 is the value for the bond length at equilibrium that is considered a parameter. The term summing over angles represents the energy due to the bending of two contiguous bonds. Angle bending terms are parameterized by a force constant k_θ and the equilibrium angle value θ_0 in degrees. The third term, summing over torsion angles, models the energy for twisting three contiguous bonds with the force constant V_n, the multiplicity N, a phase shift γ, and the torsion angle ω. The term (double summation over i and j) represents the non-bonded energy between all atom pairs, which can be broken down into van der Waals and electrostatic energies. The van der Waals potentials take into account repulsion between atoms at small separations accounting for the excluded volume between atoms and also weak attraction at larger distances. The common form of this potential for a pair of atoms i and j is given by a Lennard-Jones function:

$$V_{i,j} = 4\epsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right]$$

Here, the distance, $r_{i,j}$, is the distance separating the two atoms, $\epsilon_{i,j}$ is the depth of the potential well for the interaction of atoms i and j, and $\sigma_{i,j}$ is the distance where the model potential is exactly zero. The electrostatic or Coulomb potential describes the interactions between pairs of partial charges. q_i and q_j are the partial charges on the atoms i and j and ϵ_0 is the dielectric constant of vacuum.

2.1.4 The Generalized Born Model

The Generalized Born (GB) implicit solvent is a fast but approximate method for calculating molecular electrostatics in solvent by a Poisson Boltzmann equation which models water as a dielectric continuum. It is based on modeling solutes as a set of spheres whose internal

dielectric constant differs from the external solvent. The functional form of the Generalized Born model uses equation (2.10) below:

$$G_s = -\frac{1}{8\pi\epsilon_0}\left(1-\frac{1}{\epsilon}\right)\sum_{i,j}^{N}\frac{q_i q_j}{f_{GB}} \tag{2.10}$$

where

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \qquad \text{and} \qquad D = \left(\frac{r_{ij}}{2a_{ij}}\right)^2, \qquad a_{ij} = \sqrt{a_i a_j}$$

and ϵ_0 is the permittivity of vacuum, ϵ is the dielectric constant of the solvent being modeled, q_i is the electrostatic charge on particle i, r_{ij} is the distance between particles i and j and a_i is a quantity, with the dimension of length, called the effective Born radius of an atom which characterizes its degree of burial inside the solute [15]. Qualitatively it can be thought of as the distance from an atom to the surface of the molecule.

The AMBER software package, a suite of biomolecular simulation programs [16] is utilized for all simulations in this study. The basic AMBER information workflow is shown in Figure 2.1 below. Preparatory programs in AMBER include the following. *LEaP* which is the primary program use to create a new system or modify an existing one. *pdb4amber* generally helps in preparing pdb-format files coming from other places (such as rcsb.org) to be compatible with *LEaP*. The one is *parmed*. This provides a simple way to extract information about the parameters defined in a parameter-topology file *antechamber* is the main program to develop force fields for drug-like molecules or modified amino acids using the general Amber force field (GAFF).

Simulated Annealing with Nuclear Magnetic Resonance (NMR) - Derived Energy Restraints (SANDER) is one of the central simulation utilities in AMBER and provides tools for energy minimization and molecular dynamics with a wide variety of options. The minimization is achieved by relaxing the structure through iteratively moving the atoms downhill

along the energy gradient directions until such gradients are basically zero.

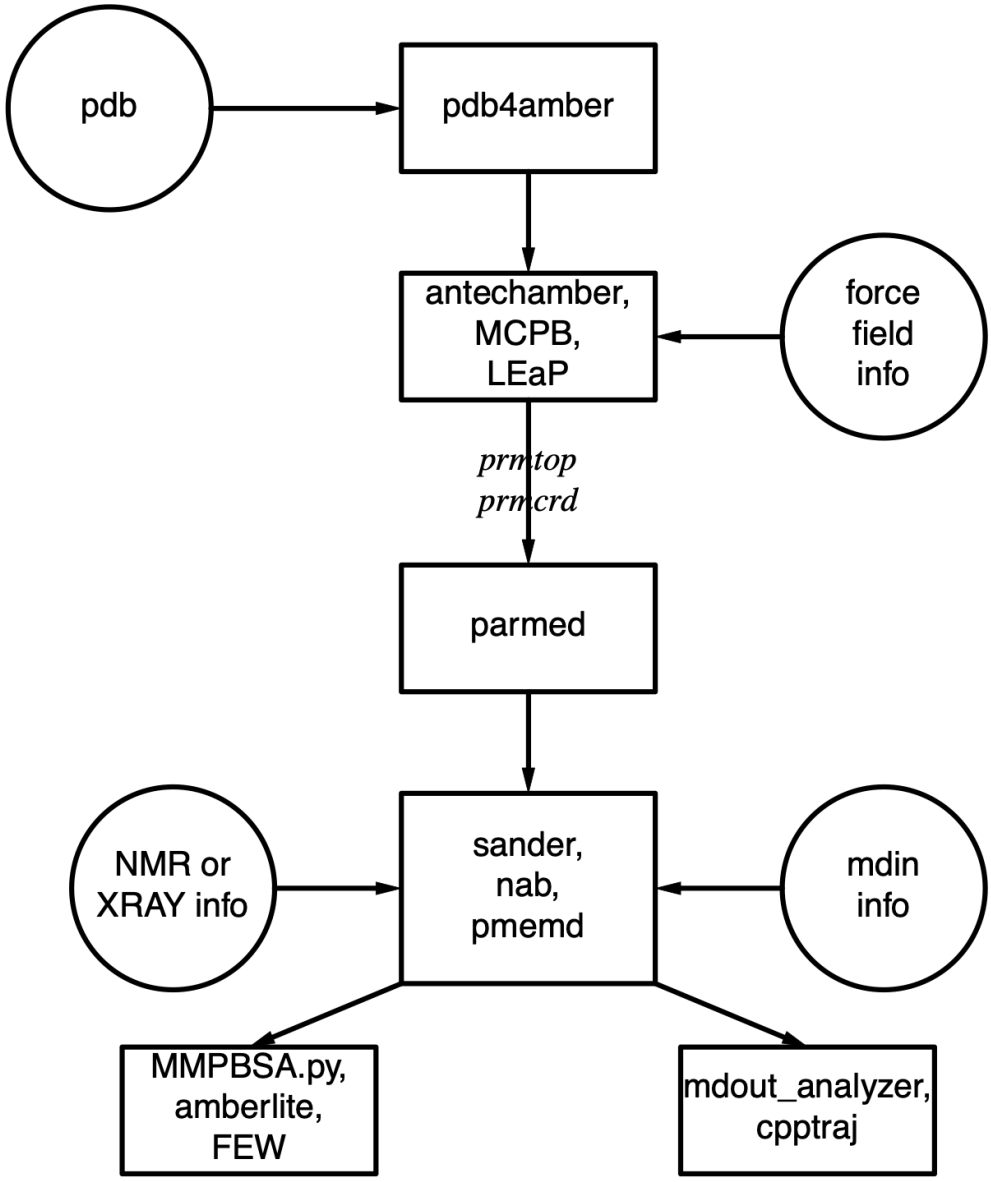


Figure 2.1: Basic information flow in Amber.
(Source: AMBER 18 manual)

The molecular dynamics portion generates time trajectories of the system configuration by integrating the Hamilton equations of motion of all atoms in the system. During

the simulation, configurations are saved at regular intervals for post-simulation analysis, and basic free energy calculations using thermodynamic integration can also be performed. CPPTRAJ is a tool that provides utilities for numerical analysis of the simulation results [17]. It is the main trajectory analysis utility for carrying out superpositions, extractions of coordinates, calculation of bond, angle, dihedral values, atomic positional fluctuations, correlation functions, analysis of hydrogen bonds, etc.

Particle Mesh Ewald Molecular Dynamics (PMEMD) is another primary molecular dynamics engine within the AMBER Software suite. This engine aims at improving performance in the most frequently used methods of SANDER. The code has since been expanded into multiple integrated programs, offering massively parallel CPU and highly functioning well or as expected GPU [18–20] capabilities for common particle simulations as well as sophisticated CPU implementations of advanced models for electronic polarization. PMEMD supports Particle Mesh Ewald simulations, Generalized Born simulations, Isotropic Periodic Sums, Analytical Linearized Poisson-Boltzmann, (ALPB) solvent, and even gas phase simulations using the AMBER Force fields. PMEMD accepts SANDER input files (*mdin*, *prmtop*, *inpcrd*, $refc$). For visualizations of structures, both Chimera [2], an extensible program for interactive visualization and analysis of molecular structures, and Visual molecular dynamics (VMD) are used. The plots are made with Gnuplot. It is a free, command-line utility for visualizing equations and discrete data. I have developed a couple of codes in the Fortran programming language for pre- and post- processing work. There are compiled on a personal computer, ARGO, and the Department of Computational and Data Science workstations.

Chapter 3: Structure and Dynamics of Polyacrylamide in Glycerol Solutions

3.1 Introduction

Polyacrylamide (PAM) is a thermoresponsive, biocompatible, and water-soluble polymer that can be tailored to meet a broad range of commercial applications, most of them based on its well-above room temperature glass transition temperature of 400 K [21,22]. The polymer is synthesized either as a simple linear chain or as a cross-linked structure. PAM increases the viscosity of water and belongs to the super water-absorbent polymers (SAP) family. When hydrated, PAM forms a soft gel used in gel electrophoresis for protein separation. Indeed, PAM is hydrophilic and can form aqueous solutions of very high concentrations [23]. Because of their gel-like properties, these aqueous solutions are employed as flocculants in the removal of suspended particles from sewage and industrial effluents such as paper mill wastewater. Through the highly reactive amide NH_2 groups, the polymer can be chemically modified to produce cationic or anionic polymers, which are particularly useful in mineral-processing and metallurgical operations for the separation of metals from residues [24]. PAM increases the viscosity of other fluid, and may be the cause of unexpected turbulent behavior in the flow of otherwise viscoelastic fluids at low Reynolds number [25].

The past couple of decades have seen a dramatic increase in computational power and high-performance computer algorithms, among which MD simulations have emerged as valuable tools for studying macromolecules and large molecular systems at the atomic scale. Along the simulations it is possible to follow in time, the interfacial dynamics of complex 3D molecular structures both localized around particular macromolecules or of their interactions with other surrounding molecules, which are yet not possible to be observed

experimentally [26]. For example, Wang et al. [27] used a coarse-grained MD for investigating the effect on linear polyacrylamide (PAM) structure by surfactant molecules in aqueous solutions. The study revealed that PAM in water curled into a cluster-like structure in the absence of surfactant molecules, while it stretched out into a beaded-necklace structure at the hydrophilic interface created by the surfactant molecules. In another study, Wu et al. [28] used an all-atom MD under the COMPASS force field for demonstrating that the addition of linear PAM at the interface of water and a foam system increased the foam stability. More recently, de Oliveira and co-authors conducted extensive all-atom MD simulations of N-propylacrylamide solvated in water [29] and were able to verify the effect of copolymerization with acrylamide on the lower critical transition temperature (LCST). In a combined experimental-computational work, Asadujjaman et al. [30] simulated with all-atom MD a 40-PAM chain in several solvents including alcohols, water, and their mixture with the objective of tuning the upper critical transition temperature (UCST) depending upon the alcohol-water relative concentration.

The atomic-scale behavior of a large number of liquids has been the subject of several decades of discovery that spans from the times that MD became a clearly useful method for studying the dynamics and structure of systems in the fluid phases [31] to the current research of more complex liquids. For example, in [29] several alcohols were MD simulated, both pure and mixed with water. In another study, a new force field was developed for ethyl acetate and its mixture with water [32]. Propane-1,2,3-triol or glycerol is a sugar alcohol with three hydroxyl groups that was MD stimulated in aqueous solutions [33]. These authors provide a good review of MD simulations of glycerol prior to their study. Worth noting is the first all-atom MD simulation using the AMBER force field by Chelli and collaborators [34], which was later reparameterized [35] with different atomic charges and Lennard-Jones parameters that reproduced the glycerol diffusion coefficient better than in earlier works. The CHARMM force-field was proposed to model glycerol [36, 37]. However, the published simulations give MD values for the diffusion coefficient about half the experimental values.

More recently, the AMBER force field of Ref. [35] with modifications for the glycerol bending and torsion angles constants [38] gave rise to reasonable MD-simulated densities of glycerol, although diffusivities were low compared to experiments. Reference [38] analyzed in detail several glycerol-water mixtures. Jahn et al. [39] conducted a comparative study of MD calculated densities and thermodynamic properties including the AMBER [34, 35], the CHARMM [36], and three versions of the OPLS force fields [40], which indicated that the AMBER force field gives reasonable thermodynamic properties. Over the years, the AMBER package has had periodical improvements to their force fields, including the GAFF force field that currently allows for the inclusion of two types of atomic charges. In this work, the latter is adopted, as it is explained in section 3.2.

This study is partly motivated by an experiment [25] that investigates the flow of a viscoelastic fluid (90:10 glycerol:water) with polyacrylamide inside along microchannels at low Reynolds numbers. In this study, it was observed that downstream the flow become unstable, consistent with features of elastic turbulence. In their experimental setup, the polymers flow initially around cylindrical obstacles; their interpretation was that the polyacrylamide is elongated at the initial times when flowing along the cylinders, while the turbulent flow structures might be associated with sudden coiling-stretching events due to fluctuations in the stream-wise velocity gradients. A recent fluid dynamics continuum simulation [41] corroborated that vortices can be produced in viscoelastic fluids if there is a spatiotemporal localization of energy on the neighborhood of flowing polymers modeled by the FENE-P dumbbells [42].

This chapter presents extensive all-atom MD simulations of PAM oligomers of various molecular weights solvated in pure glycerol, pure water, and mixed glycerol-water (90:10) by weight. The goal is to investigate the structural fate of the solvated oligomers in the three chosen solvents. The structural changes of a selected set of PAM oligomers containing 10, 20, and 30 monomers are analyzed when solvated in explicit solvents, modeled at the all-atom level. To achieve this, a new parametrization of the GAFF was carried out for pure glycerol and the mixed glycerol-water solution. Results concerning the structural properties

of the three selected PAM oligomers in the explicit solvents are summarized in section 3.3, including radius of gyration, hydrodynamic radius, end-to-end distance, Z orientation order parameter, and interaction energies between the oligomers and the solvents. Summarizing observations are cast in the discussion of Sec. IV. We predict that PAM oligomers of all sizes at ambient conditions remain as randomly coiled oligomers of variable elongations in the three all-atom explicit solvents considered. We additionally predict that conformations of PAM oligomers considered here are basically trapped in either of the glycerol-based solvents, and their structure is maintained without visible changes over tens of nanoseconds.

3.2 Methods

Three solvents are considered, pure glycerol, pure water, and their mixture at 90:10 concentration in molecular weight. The solutes in each of these solvents are polyacrylamide oligomers of various lengths, n-PAM, formed by n acrylamide monomers ($-CH_2CHCONH_2-$). The backbone of this polymer has only two carbon atoms, one of them bonded to the amide $NH_2-C=O$. The focus of this study is on the effect that three all-atom, explicitly modeled solvents (glycerol, water, 90:10 glycerol-water) have on a subset of n-PAM of increasing molecular weight modeled with our GAFF determined parameters: 10-PAM (712.82 u), 20-PAM (1423.62 u), and 30-PAM (2134.42 u). For the latter, the all-atom GAFF charges for the glycerol solvent is also generated.

3.2.1 Parameters for the GAFF force field of n-PAM and glycerol solvent

Within the GAFF force field, the atomic charges are generated for each n-PAM as a full molecule at the B3LYP 6-31G* density functional theory level, including the polarizable continuum model (PCM) [43], based on the Merz-Singh-Kollman population analysis [44, 45], and using Gaussian09 [46]. The atomic charges are later ported into the AMBER Tools18 [16] to generate their corresponding RESP charges. Additionally, for the n-PAM oligomers, an alternative set of atomic charges was generated at the BCC (additive bond

charge corrections) approach [47, 48]. The latter is based on semiempirical AM1 quantum approaches and are calculated within the AMBERTOOLS 18 routines. Files of the used n-PAM geometries and the corresponding RESP and BCC atomic charges are provided in Appendix A.

Concerning the glycerol solvent all-atom MD simulations that are part of the study, the GAFF force field is also used and the two types of atomic charges: RESP and BCC. They are generated for these molecular liquids as well. Again, Appendix A contains the RESP and BCC atomic charges for the glycerol solvent. For water in the solvent component, the SPC/E model [49] is adopted for the study.

3.2.2 Methodology associated with the MD All-atom simulation of n-PAM in explicit solvent

Glycerol is liquid between 291 and 563 K, and it is often used as a mixture with water in a large variety of relative concentrations. Before solvating n-PAM in glycerol, we validate simulations of the pure glycerol solvent and its 90:10 mixture with water. The simulations of pure glycerol contain 2000 glycerol molecules while for the 90:10 glycerol:water mixed system the simulations involved 1800 glycerol molecules and 1018 water molecules. Meanwhile, the pure water system contains 9500 water molecules. These systems were equilibrated with NPT-MD at T = 298 K and 1.01325 bar via the Berendsen thermostat and barostat [50] along 40 ns long trajectories using a 2 fs time step, a 20 Å cutoff, and periodic boundary conditions (PBC). Ewald sums are used in all calculations for the long-range electrostatics within the particle mesh implementation (PME). This equilibration stage is followed by an NVE-MD production run along 10 ns for each system at the equilibrium densities that maintained a temperature of 298 ± 2K. As described in the following sections, the NVE simulations are used for calculating the pure solvents radial distribution functions and their self-diffusion coefficients from:

$$D = \frac{1}{6t}\frac{1}{m}\sum_{k=1}^{m}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r_i}(t) - \mathbf{r_i}(t_{0k}))^2 + D_{PBC} \tag{3.1}$$

where $\mathbf{r_i}$ is the position of the ith molecule center of mass at time t and N is the number of molecules in the solvent. Each NVE run is split into m time series, each series starting from a reference position $\mathbf{r_i}(t_{0k})$ and their average is taken as indicated in Eq 3.1. The last term is the correction due to the PBC [51], $D_{PBC} = \frac{2.837297 k_B T}{6\pi\eta L}$, with k_B being Boltzmann's constant, T temperature, L computational box length, and η solvent viscosity. Values for the solvent viscosities are taken from experiments at 298 K: $\eta_{glycerol} = 945\,\mathrm{mPa\,s}$ [52, 53], $\eta_{90:10}$= 163.6 mPa s [53] and $\eta_{water} = 0.8937\,\mathrm{mPa\,s}$ [54].

The next step is the preparation of systems with one n-PAM, n=10, 20, 30, each solvated into the three different solvents. Each of the 18 systems is brought to equilibrium with NPT-MD at 298 K and 1.013 25 bar along 100 ns. Equilibrium densities of the solvated systems are within the standard deviation of the solvent equilibrated densities for the three PAM sizes considered. These simulations are followed at the equilibrium densities by NVE-MD production runs along 80 ns and temperatures around 298 K. It is from these NVE-MD simulations that the solvated n-PAM structural and dynamic properties are calculated, including their diffusion coefficients in the solution. For the diffusion coefficients, the center of mass of the full n-PAM is followed in time within the solution and calculated from Eq. (3.1) as well.

Along the MD trajectories, the energetics, several structural and dynamic properties of the n-PAM are examined such as the hydrodynamic radius of defined by Equation (3.2). This property is determined most commonly for polymers where the subparticles would then be the units of the polymer.

$$\frac{1}{R_{hyd}} = \frac{1}{2N^2}\left\langle \sum_{i\neq j}\frac{1}{r_{ij}} \right\rangle \tag{3.2}$$

where r_{ij} is the distance between subparticles i and j, and N is the number of subparticles [55]. The theoretical hydrodynamic radius R_{hyd} is important in the study of the dynamic properties of polymers moving in a solvent as it is often similar in magnitude to the radius of gyration, R_g [56] which is another structural property evaluated. Another useful property that measures the orientation of monomers in the polymer is Z-order parameter. This is defined as;

$$Z = \frac{3}{2}\left(\frac{1}{n}\sum_{i=1}^{n-1}\cos^2\alpha_i - \frac{1}{3}\right) \tag{3.3}$$

The monomers in the respective oligomer are in a random orientation when $Z \approx 0$ and in a straight orientation when $Z \approx 1$. Other structural properties calculated are the radius of gyration, which measures the dimension of the oligomer, the end-to-end distance. It is the distance between the centers mass of the two end monomers. We present average values of the hydrodynamic radius R_{hyd}, the radius of gyration R_g, end-to-end distance R_{e-e} and Z-order parameter Z at their corresponding temperatures determined from the NVE simulation at their respective densities (ρ). These averages were calculated for the last 10 ns of the production run of the NVE MD simulation. The average radius of gyration, R_g of PAM-10 and PAM-20 over 10 ns are relatively larger in the water solvent systems than in glycerol and glycerol-water systems. The corresponding hydrodynamic radius and end-to-end distances follow a similar trend even though with a much larger variance in the end-to-end distance. The time-series study of these properties however revealed how the oligomers like to stay elongated with a twist in the glycerol and glycerol-water system but showing the tendency to form cluster-like structures in the water solvent.

3.3 Results

3.3.1 Properties of all-atom MD simulated solvents

Pure solvent simulations are performed for quantifying the appropriate behavior of the solvents at 298 K. From the NPT-MD simulations, the solvent systems attain equilibrium densities of 1.258 ± 0.002 g/cm^3 for glycerol, 1.227 ± 0.002 g/cm^3 for mixed glycerol:water with RESP atomic charges. With BCC atomic charges, the attained equilibrium densities are 1.265 ± 0.002 g/cm^3 for glycerol and 1.253 ± 0.002 g/cm^3 for the glycerol:water. The obtained SPC/E water equilibrium density is 0.998 ± 0.002 g/cm^3. These values are in excellent agreement with experimental values at 298 K of 1.257 91 g/cm^3 and 1.253 31 g/cm^3 for the two glycerol solvents [38]. The agreement is also excellent with the previous SPC/E water simulation at room temperature of 0.998 g/cm^3 [57].

At these equilibrium densities and 298 K, the radial distribution function, rdf or g(r), of each solvent is calculated and agrees very well with experimental results. Figure 3.1a shows the rdf of glycerol between atom pairs in different molecules. The inset in this figure provides the label given to the glycerol molecule atoms entering in the considered pairs between molecules. The calculated peak positions for the six atom pairs depicted are 1.89, 1.89, 2.85, 2.85, 2.85, 2.85 Å (RESP case) 1.83, 1.83, 2.79, 2.79, 2.79, 2.79 Å (BCC case), which are in excellent agreement with the experimental values of 1.77 ± 0.61, 1.80 ± 0.63, 2.73 pm0.87, 2.76 ± 0.78, 2.76 ± 0.80, 2.76 ± 0.90 Å [58]. As expected, the BCC-based calculation (dashed lines) yields peaks at slightly shorter distances than the RESP case due to the system higher density. For example, the O – H peak is at 1.89 Å for the RESP case and at 1.83 Å for the BCC case. Otherwise, the glycerol liquid structure is unchanged if the glycerol atomic charges are RESP- or BCC-modeled. Figure 3.1b depicts the rdf of the O – H pairs between glycerol-water molecules, showing that the hydrogen bonds formed with glycerol oxygens are slightly longer than the bonds formed with water oxygens. This figure includes the rdf from distances between the glycerol molecules centers of mass and

the water oxygens (red), which shows that the two liquids form a hydrogen-bonded network (first peak) with the second peak associated with a first coordination shell of glycerol-water molecules at 2.8 Å and a second coordination shell at 3.5 Å. The SPC/E water rdf is known to reproduce excellently the structure of liquid water [49,59], and our results are consistent with previous simulations.

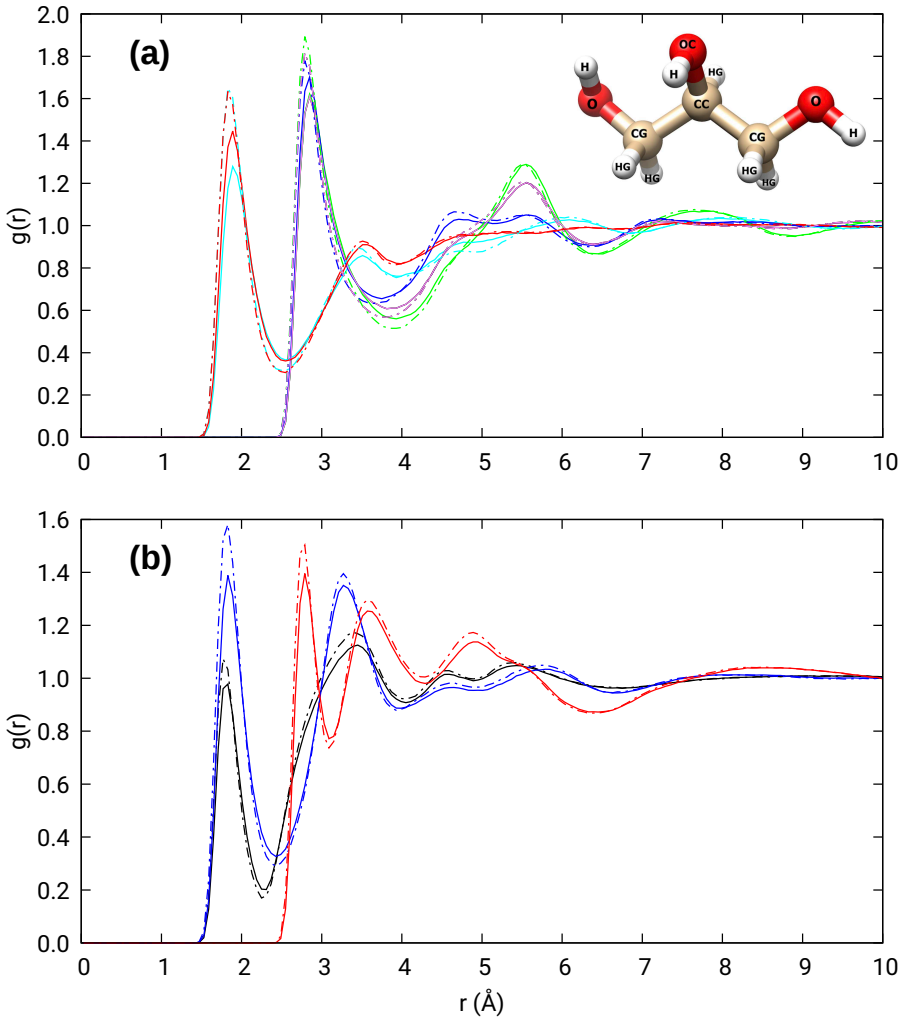


Figure 3.1: Radial distribution function of glycerol and glycerol:water, at 298 K. **a)** Glycerol at equilibrium density of 1.258 g/cm^3 (RESP, solid lines) and 1.265 g/cm^3 (BCC, dashed lines) with pairs O–H (red), OC–O (black), OC–OC (green), OC–H (cyan), O–O (blue), O–OC (violet), with atoms identification as shown in the inset. **b)** Glycerol:water at equilibrium density of 1.227 g/cm^3 (RESP, solid lines) and 1.241 g/cm^3 (BCC, dashed lines) with pairs $H_{glycerol}$ - O_{water} (black), $O_{glycerol}$ - H_{water} (blue), $CM_{glycerol}$ - O_{water} (red, CM=center of mass).

The self-diffusion coefficients of glycerol and water are estimated from Eq. (3.1), considering 40 different time origins, each of 0.5 ns of NVE time evolution. The PBC corrected self-diffusion coefficient from the simulations at 298 ± 1 K are $(3.35 \pm 0.03) \times 10^{-7}$ cm^2/s for the RESP case and $(1.93 \pm 0.02) \times 10^{-7}$ cm^2/s for the BCC case. These diffusivities compare well with the experimental value of 1.7×10^{-7} cm^2/s obtained from the Taylor dispersion method [60]. However, this experimental value is larger than the diffusion coefficient obtained from the NMR pulsed magnetic field gradient [61] or modulated gradient spin echo method [62]. The corrected self-diffusion coefficient of water from the simulation at 300 ± 1 K is 2.860×10^{-5} cm^2/s, which compares well with the experiment at 298 K of 2.299×10^{-5} cm^2/s [63].

The density of glycerol as a function of temperature is investigated as well. Figure 3.2 shows this behavior for the two types of charges as compared with experiments. The agreement of the two force fields is reasonable around 300 K. At higher temperatures, both force fields yield a lower density than the experimental value. We assess that the force field is modeling well the glycerol solutions in the temperature range adequate for the goals of this work.

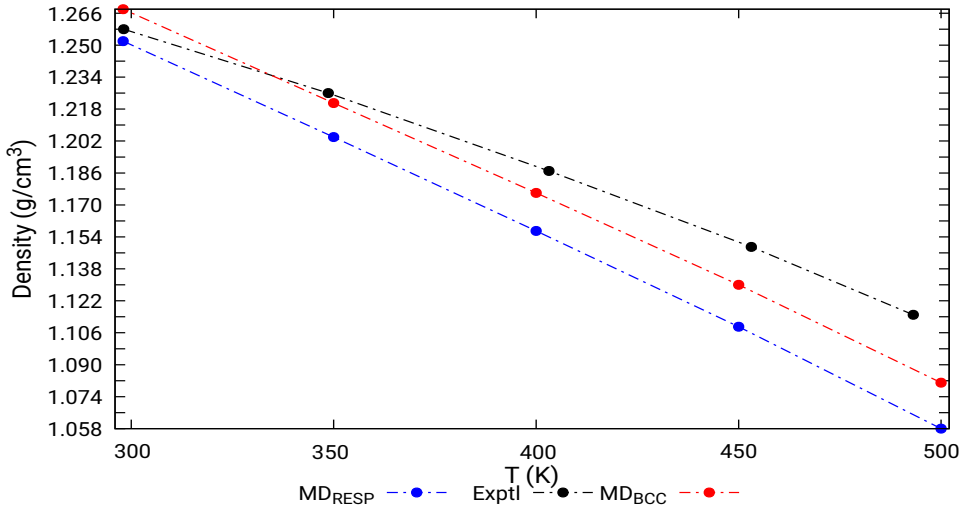


Figure 3.2: Density of glycerol as a function of temperature compared to experiments [1].

3.3.2 Properties of n-PAM in explicit solvents

All simulations are started from initial n-PAM geometries and force field atomic charges with initial R_g values of 5.1 Å, 6.8 Å, and 9.4 Å for $n = 10$, 20, and 30, respectively. These solutes in the glycerol solutions have dilution compositions by mass of 0.38%, 0.77%, 1.16%, while in water the dilution is 0.42%, 0.84%, 1.26% for n-PAM of increasing n. An energetics evaluation of the oligomers in the various solvents yields results cast in Table 3.1. Reported in the table are oligomer potential energy averages per monomer over the last 10 ns of MD NVE runs showing for the RESP case an increasing stabilization with increasing oligomer size. However, the BCC case yields E_{PAM}/monomer basically equal for all the solvated oligomers in the three solvents. The reported errors correspond to the standard deviation. The interaction energy E_{int}/atom between each n-PAM and the solvents represents the balance between the total potential energy of the system and the sum of the individually separated potential energies of the solvent and the oligomer:

$$E_{\text{Int}} = E_{\text{Sys}} - (E_{\text{solvent}} + E_{\text{polymer}}) \tag{3.4}$$

As observed in Table 3.1, the trend of the interaction energy in each of the three solvents is for the system to become less cohesive as the size of the oligomer increases. These energies present large fluctuations in the mixed glycerol-water system attributed to the mobility of the water molecules in the neighborhood of the oligomers that results in frequent changes of the solution surrounding the solutes. Comparison between the interaction energies of n-PAM in the different solvents depends on the system sizes, with a decreasing PAM stabilizing propensity as the number of water molecules increase in the solvent.

Table 3.1: Energetic evaluations: Interaction Energy (E_{int}) kJ mol^{-1}/monomer and Potential energy (PE_{tot}) kJ mol^{-1}/monomer for RESP and BCC atomic charges are MD-NVE averages at 298 K and the equilibrated density of the various solutions.

Solvent	PAM	E_{Int} RESP	E_{Int} BCC	PE_{tot} RESP	PE_{tot} BCC
Glycerol	10	-23205 ± 77	-19433 ± 69	-113.75 ± 2.91	-136.41 ± 4.26
	20	-11597 ± 41	-9773 ± 34	-186.94 ± 2.19	-148.22 ± 2.36
	30	-7739 ± 29	-6543 ± 22	-194.84 ± 2.08	-143.21 ± 1.85
Glycerol-Water	10	-11428 ± 1205	-17761 ± 1300	-119.50 ± 3.49	-138.96 ± 4.13
	20	-7458 ± 360	-9159 ± 470	-184.52 ± 2.19	-140.74 ± 2.76
	30	-3831 ± 381	-5961 ± 423	-188.47 ± 2.011	-143.84 ± 1.99
Water	10	-43658 ± 53	-43652 ± 42	-118.37 ± 3.70	-134.79 ± 3.46
	20	-21735 ± 20	-21759 ± 19	-186.96 ± 3.49	-142.47 ± 3.50
	30	-14511 ± 18	-14499 ± 16	-193.19 ± 2.94	-143.16 ± 3.54

3.3.3 Structural and Dynamic Properties of the MD simulated n-PAM

Analysis of the various MD trajectories indicates that the structural fingerprints of the solvated oligomers are acquired during the NPT MD equilibration runs. On the other hand, along the subsequent 40 ns NVE runs, those fingerprints are basically locked-in and do not change substantially with time. Calculated polymer structural properties include gyration and hydrodynamic radii, end-to-end distance, and orientational order parameter. These properties are very similar between the different solvents, although there are evident fluctuations. Indeed, analysis of these properties time evolution reveals that the random coiled conformations remain as such for longer periods of time in the glycerol and glycerol-water systems than in water. These properties are not always distributed normally; therefore, the calculation of averages has little meaning. Figure 3.3 illustrates the R_g distribution of the three oligomers modeled with RESP (solid line) and BCC (dashed line) when solvated in the three solvents considered.

Table 3.2: Diffusion Coefficients D_C ($\times 10^{-8} cm^2/s$) of PAM at 298 K for RESP and BCC atomic charges

Solvent	PAM	D_C RESP	D_C BCC
Glycerol	10	0.412 ± 0.001	2.967 ± 0.002
	20	0.296 ± 0.001	1.830 ± 0.001
	30	0.199 ± 0.001	1.587 ± 0.001
Glycerol-Water	10	1.716 ± 0.001	8.581 ± 0.004
	20	0.978 ± 0.001	6.187 ± 0.003
	30	0.927 ± 0.001	5.407 ± 0.002
Water	10	459.9 ± 1.5	432.0 ± 1.2
	20	306.1 ± 1.4	286.4 ± 1.4
	30	236.7 ± 0.5	268.7 ± 0.6

The n-PAM oligomers are massive compared to the solvent molecules. Despite this fact, we additionally evaluated the oligomers' diffusivity in each system at 298 ± 1K. Therefore, the motion of the oligomer center of mass is followed in time. Using Eq. 3.1, our estimates of n-PAM diffusion coefficients in glycerol are presented in Table 3.2 for the two force fields. In the mixed glycerol-water solvent, the diffusivity increases approximately by a factor of four for the three oligomers with n =10, 20, 30, respectively. These estimates indicate that the oligomers diffuse about ten times less than the glycerol molecules where they are solvated. Finally, within water as a solvent, the oligomers diffusion coefficients increase by about 200 times and the difference between the two atomic charges is strongly reduced.

PAM of high molecular weights is part of the family of thermoresponsive polymers and we investigated alternative simulations for evaluating the glycerol solvated oligomers sensitivity induced by temperature changes. In one of them, the runs are initiated from fully stretched-out chains, equilibrated first at a high temperature of 500 K, and finally cooled down in steps of 10 K until reaching 298 K. This cooling process was carried out with MD NPT. At each cooling step of the ladder, the system was run for 20 ns trajectories, while the high temperature and final temperature runs were 40 ns long. The in silico cooling rate is 0.8 K/ns. Another alternative explored consisted of starting the simulation at 298 K from extended oligomers and let the system run for hundreds of nanoseconds. In yet another

alternative setup, the solvated extended oligomers were heated up in 5 K steps from 298 K to about 350 K and cooled back down.

What was learned from these attempts is the enormous variability of conformations with different R_g that these n-PAM oligomers can take when they become trapped in the solvent. For example, in the RESP case and pure glycerol, PAM-30 ended these processes with $Rg \approx 9 - 10$Å, which contrasts with other attempts in which the final $Rg \approx 12 - 14$Å. Not only these outcomes for the PAM-30 radii are significantly different between them, but also the distributions are peaked at different values than what is shown in Fig. 3.3. A commonality across simulations, however, is that the RESP case gives rise to very compact Rg distributions when compared to the BCC equivalents. In the mixed glycerol-water system, a similar situation is encountered. We conjecture that the size of these oligomers is short for having a well defined lower critical transition temperature (LCST) in which the polymer changes abruptly from coiled conformation to collapsed globule. Therefore, when cooling between $350 - 300$K, the chain can be in any intermediate configuration and by lowering the temperature, the glycerol solvents trap the instantaneous geometry and the fate of the chain is to be locked-in an instantaneous structure. When solvated in water, these oligomers are flexible and they do not acquire a locked-in R_g at 298 K.

The end-to-end distance, R_{e-e} is another interesting structural property commonly associated with flexibility. In this respect and in the glycerol solvents, we observe a marked difference between the RESP and BCC cases. Compounding the observations, we assert that the glycerol molecules have the ability to keep the oligomer ends fairly localized for extended periods of 50 ns or more and promote global motions of the full macromolecule more than its ends. For all the trials and temperatures, the order parameter Z kept almost zero. Therefore, the oligomer structures, either random coiled or globular, have random distributions of angles with respect to the oligomer orientation vector.

Finally, efforts were carried out for identifying solvent molecules near the n-PAM. From several radial distribution functions, on average, we assess that glycerol molecules may approach PAM-30, at most, up to 2.5 Å forming a broad coordination shell of about 5.3 Å. As

water is added, a small amount of water clumps in a thin shell at 2.7 Å.

Table 3.3: Radius of gyration R_g, Hydrodynamic Radius R_{hyd}, End-to-end distance R_{e-e}, Z-order parameter Z, determined from the simulation at the respective densities

Solvent		Glycerol		Glycerol-Water		Water	
	PAM	RESP	BCC	RESP	BCC	RESP	BCC
R_g (Å)	10	6.60 ± 0.08	5.60 ± 0.27	6.47 ± 0.30	6.27 ± 0.24	6.39 ± 0.54	6.52 ± 0.46
	20	9.51 ± 0.15	8.12 ± 0.11	9.196 ± 0.52	10.60 ± 0.61	9.41 ± 1.60	9.22 ± 1.53
	30	9.48 ± 0.08	12.63 ± 0.28	10.05 ± 0.26	11.60 ± 0.50	9.80 ± 0.97	11.23 ± 2.46
R_{hyd} (Å)	10	10.48 ± 0.06	9.97 ± 0.17	9.74 ± 0.10	9.71 ± 0.23	10.38 ± 0.31	10.47 ± 0.24
	20	14.69 ± 0.11	14.16 ± 0.10	14.42 ± 0.08	14.47 ± 0.38	14.61 ± 0.95	15.07 ± 0.57
	30	16.95 ± 0.07	18.83 ± 0.14	17.12 ± 0.19	18.36 ± 0.21	16.74 ± 0.79	17.45 ± 1.31
R_{e-e} (Å)	10	15.05 ± 0.62	10.60 ± 2.82	15.74 ± 2.23	15.29 ± 1.57	14.02 ± 4.11	15.31 ± 3.34
	20	27.24 ± 0.48	11.26 ± 1.90	13.05 ± 1.76	15.37 ± 3.10	11.65 ± 1.25	17.06 ± 2.47
	30	18.95 ± 1.42	20.37 ± 1.49	9.14 ± 0.87	9.09 ± 0.98	18.10 ± 1.22	20.91 ± 5.81
Z	10	0.068 ± 0.053	-0.137 ± 0.069	-0.078 ± 0.051	-0.100 ± 0.103	-0.099 ± 0.118	-0.088 ± 0.139
	20	-0.063 ± 0.031	-0.117 ± 0.099	0.005 ± 0.036	-0.025 ± 0.055	-0.037 ± 0.105	-0.032 ± 0.095
	30	0.017 ± 0.024	-0.029 ± 0.044	-0.017 ± 0.041	-0.023 ± 0.096	-0.031 ± 0.069	-0.025 ± 0.072

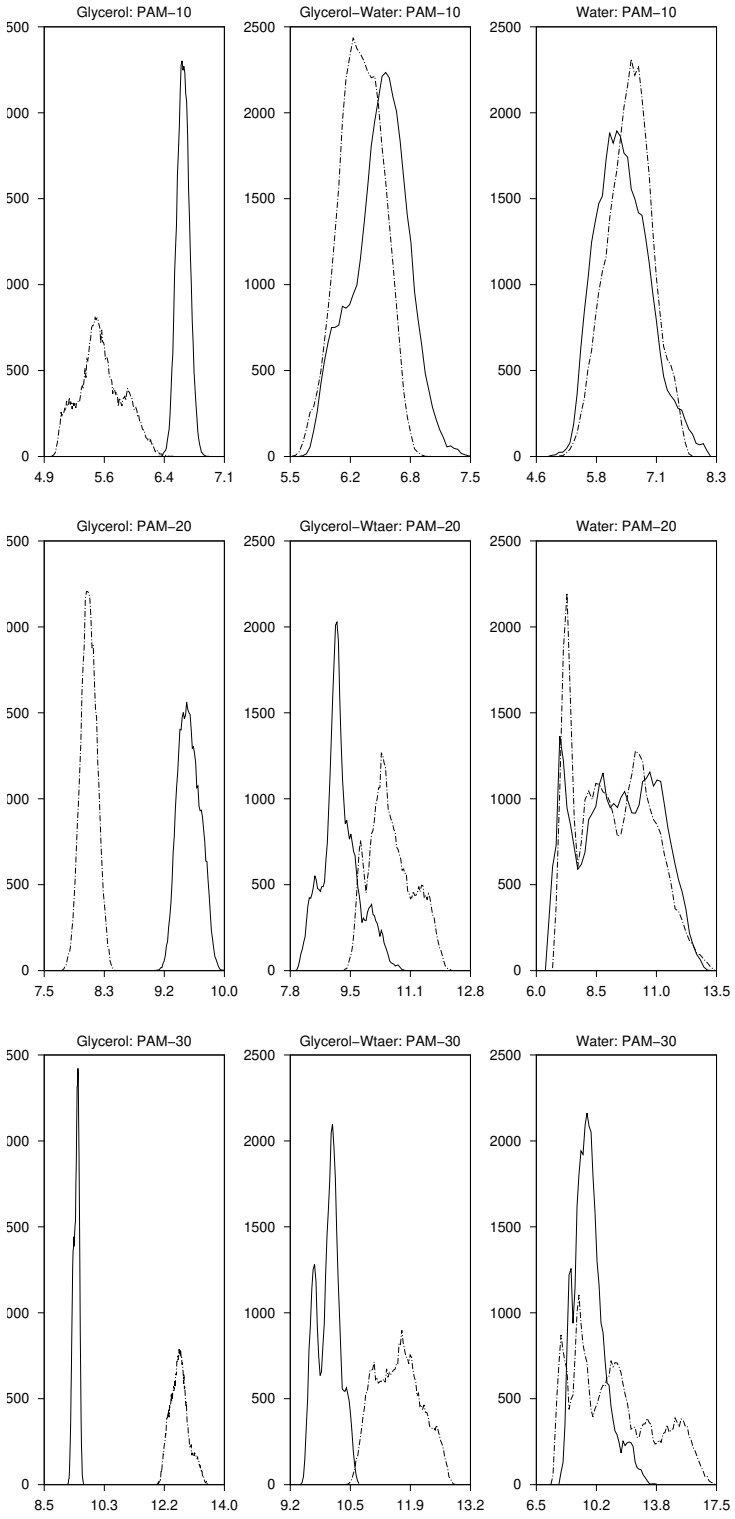


Figure 3.3: Histograms of the radius of gyration

3.3.4 The Power Law and Solvent Quality Measurement

The renormalization group findings of the 1980's determined that when a system has a mass M that does not occupy in $3D$ space its whole volume L^3, a fractal dimension, $1/\alpha$, exponent less than 3 can be defined for characterizing what fraction of the $3D$ space is actually occupied by mass[64]. The relation is given by;

$$M = \frac{1}{k} L^{1/\alpha} \tag{3.5}$$

In the polymer field, this scaling is presented reversed and associated to the radius of gyration, R_g, that identifies a volume in $3D$ as $4\pi R_g^3$. The scaling law is transformed and written in terms of the radius of gyration and molecular weight M_w as

$$R_g = k M_w^{\alpha} \tag{3.6}$$

In equation (3.5), if $\alpha = 0.5$, the fractal dimension is 2, so the space occupied in $3D$ by the polymer is equivalent to the surface of a circle. A "random coiled" model has exactly that fractal dimension. If however, $\alpha = 0.6$, the fractal dimension is 1.6666, less than 2 and identifies a polymer that fills the $3D$ space much sparsely than the random coiled polymer. When α is less than 0.5, for example, $\alpha = 0.4$, this identifies a polymer that fills the space with a fractal dimension of 2.5, which is less than 3 but much closer to filling the space compactly. Therefore, the effect of solvents on an associated volume that would characterize the compactness of a polymer in solution, is nicely given by this scaling law (power-law), equation (3.6). A plot of radius of gyration (R_g) for PAM dependence on molecular weight (M_w), and linear fits yield the coefficient α in $R_g \propto M_w^s$ equal to 0.73, 0.55 and 0.45 for the glycerol, glycerol-water and water as solvents respectively. This result shows that glycerol behaves as a "good solvent" [65] for this studied polymer as compared to the glycerol-water and the water. The oligomers in water fills the space much closer to filling the space compactly with a fractal dimension of ≈ 2.5. Figure 3.4 illustrates the molecular structure of

several PAM$-n$ sizes. These cluster-like structures are not spherical; in all cases, the ratio R_g/R_{hyd} fluctuates between 0.43 and 0.57, while a value of 0.77 is accepted to characterize spherical proteins. Therefore, these compact structures are prolate globules.

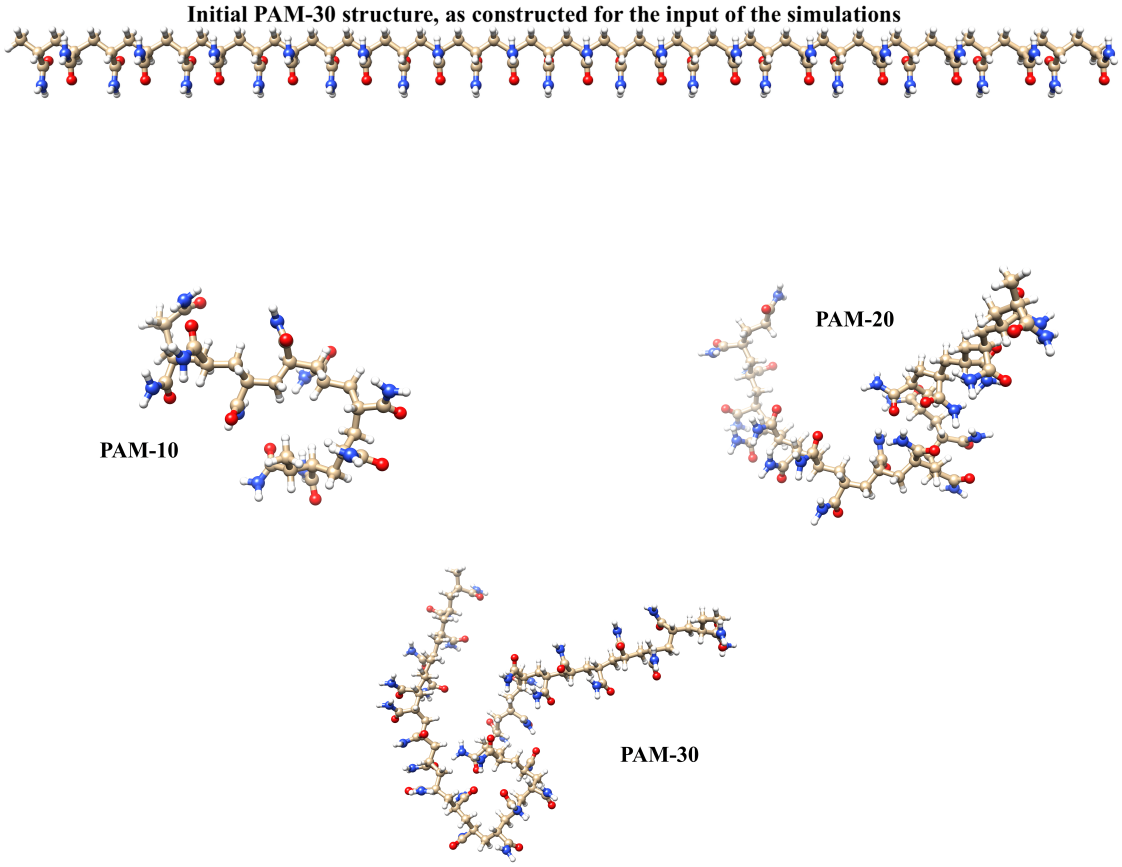


Figure 3.4: PAM-10, -20, and -30 structures in the solvent after 80 ns of NVE MD at 298 K.

3.4 Conclusions

This work presented dynamical modeling of the structure and energetics of PAM oligomers in glycerol solvents and in water with the goal of elucidating the sensitivity of the PAM structure to viscous non-Newtonian liquids such as glycerol and its mixture with water at high concentration. In order to achieve this goal, we have modified the atomic charges of the GAFF and introduced a set of RESP atomic charges that effectively permit a strong

temporal localization of the polymer chain in these glycerol solvents. We have also modified the established GAFF atomic charges of the glycerol molecules by the RESP values, obtaining a very good representation of the glycerol liquid at ambient temperature. It is the combination of these two modeling strategies that enable the PAM oligomers to acquire locked-in, swollen, and elongated structures in the glycerol solvent while they are more flexible and prone to remain as less elongated random coils in the glycerol-water solvent. In water, the oligomers are very flexible changing frequently their random coil structure without fully collapsing into a compact globular. Both the glycerol-water and the pure water solvents behave as θ solvents for the PAM-10, PAM-20, and PAM-30 oligomers considered in this work. In contrast, the BCC atomic charges case fails to provide a clear and distinct behavior of the solvated oligomers in the different solvents. This is the first simulation of its type for PAM oligomers solvated in glycerol solvents. Our protocol and simulation strategy is portable and will be useful when applied to other modeling environments.

Chapter 4: Structural Exploration of Rift Valley Fever Virus L protein Domain in Implicit and Explicit Solvents

4.1 Introduction

The Rift valley fever virus (RVFV) is an arbovirus in the Bunyavirales order, Phenuiviridae family, and Phlebovirus genus. It was first discovered in 1931 in the Great Rift Valley of Kenya, East Africa [66]. Since that time, is has caused periodic outbreaks in human and livestock populations throughout Africa and has even spread into the Arabian Peninsula. The virus is vectored by mosquitoes and, as such, outbreaks tend to follow periods of heavy rainfall that increase mosquito populations significantly [66]. The virus infects ruminants and pseudoruminants, leading to abortions in pregnant animals and high mortality among young animals. It is a negative-sense RNA virus that contains three segments of viral RNA, the S, M, and L segments and can also be transmitted to humans causing febrile illness with the possibility for the severe disease [67]. The structure of the RVFV L protein, which is made up of a sequence of 2092 amino acids has flexible termini of about 200 amino acids each and a high proportion of helical regions [68]. The structure of the C-terminal, 117 amino acid-long domain of the RVFV L protein as modeled using X-ray crystallography shows high similarity to the influenza virus PB2 cap-binding domain and the putative non-functional cap-binding domain of reptarenaviruses [69].

The interest in investigating the behavior of peptides, polypeptides, or proteins in solvent environments has grown rapidly in recent years and now constitutes a wide literature composed of thousands of research articles. Notably, Guo and Mei [70] studied the solvation effect on the structure and folding dynamics of a small peptide, Nonstructural protein 4B (NS4B) H2 in both pure water and water/2,2,2-Trifluoroethanol (TFE) cosolvent in both

explicit and implicit solvents. In this study, the force field parameters for water is taken from the TIP3P water model, and those for TFE are generated from the general AMBER force field (GAFF).

The distribution of solvent molecules around the peptide indicated that folding is triggered by the aggregation of TFE on the peptide surface but in pure water it undergoes a large structural deformation. Exploration of the effects of different pH on the structural characteristics of α-syn12 dimer using temperature replica exchange molecular dynamics (T-REMD) simulations in explicit solvent shows that the free energy surfaces contain ten highly populated regions at physiological pH, while there are only three highly populated regions contained at acidic pH [70].

A study of the free energy of unfolding in vacuum using the end-to-end distance of peptide a reaction coordinate [71] by simulating deca-alanine (Ala_{10}) was also performed using the 104-atom compact helical model used by Park et al. [72]. Although sufficient in vacuum, the study showed that end-to-end distance is incapable of capturing the full complexity of deca-alanine folding in water. Instead, the a-helical content was used as a second reaction coordinate and this led to the deduction of a more descriptive free energy landscape.

The amphiphilic peptide of the triacylglycerol lipase which plays a critical role in guarding the gate for ligand access was also studied by Nellas et al., [73] by comparing the conformations of this peptide at several water-oil interfaces and in protein environments using atomistic simulations with explicit solvents. In the oil-containing solvents, this peptide was found to able to retain a folded structure. However, when the peptide is immersed in a low-polarity solvent environment, it exhibits a "coalesced" helix structure, which has both α- and 3-10 helix components.

The structural stability and preference of a protein are highly sensitive to its accommodating environment. Solution pH is one of the most important environmental factors that affects the structure and dynamics of proteins [74, 75]. This work sought to study the behavior of the functional cap-binding domain in RVFV in the presence of different water

models. Solvents influence protein structure to a great extent.

In this work, we conduct extensive molecular dynamics simulation studies of RVFV L protein domain in five different water models to understand the solvent influence on protein structure and dynamics. We use the well-known AMBER 18 [16] with the ff14SB force field [14]. Even with several perspectives and results from laboratory experiments on different proteins, it is important to have a visual understanding of the dynamics of protein structure and changes in different solvent environments. This requires an understanding of the dynamics at the molecular level which can be achieved with molecular dynamics simulation.

The effect of solvents on the protein structure due to protein-solvent interaction has been widely studied with experimental and computational techniques. However, the molecular level understanding of protein structural behaviors in the presence of some simple solvents is still not fully understood. This work focuses on detailed molecular dynamics simulations of solvents effect on a functional cap-binding domain of Rift valley fever virus (RVFV) L protein in different water models and also in implicit solvent in order to well understand the structural and dynamic behaviors of the L protein domain. In order to achieve this, several structural and dynamic properties are presented and conclusions are drawn.

The chapter is organized as follows. Section 4.2 describes the methods used, including the computational setups and details of the simulations, including the force fields used. Section 4.3 contains our results and discussions from the simulation process. Conclusions of this paper are presented in section 4.4.

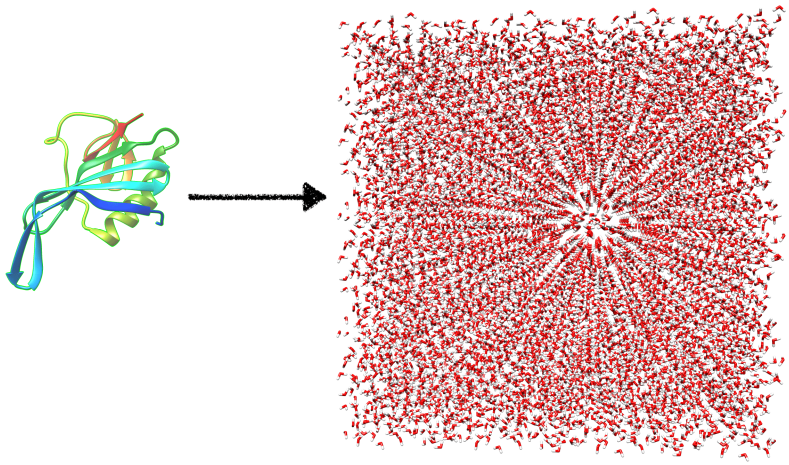


Figure 4.1: Initial conformation of RVFV L protein peptide placed in a cubic periodic box using chimera.

4.2 Methods

4.2.1 Details of the simulations

Molecular dynamics (MD) simulations of a functional cap-binding domain in Rift valley fever virus L protein in implicit and explicit solvents were carried out using the AMBER 18 package with the ff14SB force field [14]. Five different water models; TIP3P, TIP4P, OPC, SPC/E and SPCE/Fw [49, 76–78] were adopted as the aqueous media for the simulations with 10521 molecules for the explicit solvent. Before solvating the peptide in the different water models, an NVT MD simulation for 1 ns was performed to thermalize the solvents in periodic boxes of size 72.5 Å per side with cutoff distance of a 12 Å and a time step of 1 fs after an energy minimization. This was followed by an equlibration with NPT simulation for 40 ns with a time step of 2 fs using velocity rescale thermostat [79] and a pressure of 1 bar with the Berendsen barostat [50] for the pressure correction. Another 1 ns NVT was performed after the NPT. The densities of the solvents were obtained from the NPT simulations (see section 4.3). Finally, NVE MD production runs along 20 ns for each solvent box at the respective equilibrium densities that maintained a temperature of $T = 298 \pm$

2 K was performed. As described below, the NVE simulations are used for calculation of the solvents energies and their self-diffusion coefficients from Equation (3.1) with viscosity value, $\eta = 0.8937\,\mathrm{mPa\,s}$, taken from experiment at 298 K [54].

The next step is the preparation of systems with the RVFV L protein peptide, solvated in each of the different water models. The starting coordinates of the peptide were taken from the X-ray crystallographic structure (PDB ID: 6QHG) [69]. This peptide is made up of 117 amino acids making up a total of 1849 atoms including the hydrogen atoms. For the explicit solvent simulations, the visualization and analysis package, chimera [2] was used to place the RVFV L protein peptide in a cubic periodic box of size 72.5 Å per side. Energy minimization was performed using steepest descent method followed by conjugate gradient to remove possible clashes between atoms that maybe too close. The cutoff distance was increased to 16 Å after the introduction of the peptide. Position restraints were used on heavy atoms during annealing, when the system was gradually heated from $T = 0\,\mathrm{K}$ to $T = 293.15\,\mathrm{K}$ in 50 ps with periodic boundary conditions.

The systems were thermalized again for 100 ns at a constant volume and temperature of $T = 293.15\,\mathrm{K}$ before equilibration with NVT ensemble for another 100 ns at the same temperature via the Langevin thermostat with a collision frequency of 5 ps and a time step of 1 fs. Ewald sums are used in all calculations for the long-range electrostatics within the particle mesh implementation (PME) [80]. Finally, using the generated ff14SB force field parameters and the AMBER package, another simulation of 300 ns NVT Molecular Dynamics was performed for the peptide using the generalized Born model for implicit solvent model [81] at $T = 293.15\,\mathrm{K}$ with a dielectric constant, $\varepsilon = 78.5$. Energy minimization was performed using steepest descent method followed by conjugate gradient to relax the system. Position restraints were again used on heavy atoms during annealing, when the system was gradually heated from $T = 0\,\mathrm{K}$ to $T = 293.15\,\mathrm{K}$ in 50 ps.

Along the MD simulations, the energetics and several structural properties of the peptide

such as the end-to-end distance R_{ee}, radius of gyration

$$R_g^2 = \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{cm})^2 / N \tag{4.1}$$

and the hydrodynamic radius

$$\frac{1}{R_{hyd}} = \frac{1}{N^2} \sum_{i=1}^{N-1} \sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{4.2}$$

are examined where $\mathbf{r}_i$ are atomic position vectors referred to the peptide center of mass, $\mathbf{r}_{cm}$ is the center of mass position vector, r_{ij} are distances between atoms i and j and N is the number of atoms in the peptide. Other properties examined are the root-mean-square deviation, $RMSD$ with reference to the starting structure coordinates, and the solvent accessible surface, $SASA$.

The $SASA$ is one measure of protein behavior that is governed by the interactions or non-interactions of hydrophobic and hydrophilic amino acids with water [82]. The solvent molecules create a surface tension near the protein-solvent interface which affects protein dynamics and structure. Because of this, a good solvent model is expected to reproduce the SASA. This is used as a metric of comparison between the different water models used in the study. The Linear Combinations of Pairwise Overlaps (LCPO) method [83] is used for approximating the SASA. LCPO calculates the SASA of each atom by estimating the overlap between the atom and neighboring atoms. The more a protein atom is overlapped by other protein atoms, the less the atom is exposed to the solvent. LCPO defines the SASA of an atom with four terms:

$$A_i = P_1 S_i + P_2 \sum_{j \in N(i)} A_{ij} + P_3 \sum_{\substack{j,k \in N(i) \\ k \in N(j) \\ k \neq j}} A_{jk} + P_4 \sum_{j \in N(i)} A_{ij} \sum_{\substack{k \in N(i) \\ k \in N(j) \\ k \neq j}} A_{jk} \tag{4.3}$$

where the overlap between spheres i and j is

$$A_{ij} = \pi R_i [2R_i - r_{ij} - (1/r_{ij})(R_i^2 - R_j^2)]$$

The parameters $P1$, $P2$, $P3$ and $P4$ are parameterized for different atom types. The first term involves the surface area of the atom before overlap, $S_i = 4\pi R_i^2$ where R is the atomic radius (ie. vdW radius plus probe radius of 1.4 Å). The second term estimates the total overlaps of all neighboring ($j \in N(i)$ means any atom j for which $r_{ij} < R_i + R_j$) atoms with atom i. The third term is the sum of overlaps of i's neighbors with each other. The more i's neighbors overlap each other, the more they over subtracted surface area in the second term. The fourth term is a further correction for multiple overlaps. Each overlap of j with i is weighted by how much j is overlapped with all mutual neighbors k.

Along with the potential energy of the peptide calculated, the interaction energy between the solvent and solute molecule is calculated with the formula

$$E_{\text{Int}} = E_{\text{Sys}} - (E_{\text{solvent}} + E_{\text{peptide}}) \tag{4.4}$$

where E_{Sys} is the energy of the whole system, E_{solvent} is the energy of the solvent component of the system and E_{peptide} is the energy of the peptide. The results of these simulations are given in section 4.3 below.

4.3 Results and Discussion

For the RVFV L protein peptide, we investigated the behavior of sequence of 117 amino acids in the C-terminal explicitly in five water models and also in an implicit solvent environment using the generalized Born (GB) model by setting the dielectric constant to match experimental values. All methods used in the simulations are described in detail in Section 4.2, while the results obtained from each of the various steps in this workflow are presented below.

Table 4.1 presents a comparison of some properties calculated from the MD simulations to experiment and other simulations. The properties obtained from the MD simulations of the water molecules are in good agreement with other simulations.

Table 4.1: Solvent properties compared to experiment and other simulations: Potential energy PE, Density ρ, Self-diffusion coefficient D, and Temperature T.

Property	TIP3P	TIP4P	SPC/E	SPC/Fw	OPC
This work: PE kJ/mol	-40.08 ± 0.03	-31.09 ± 0.03	-46.78 ± 0.03	-45.32 ± 0.03	-38.57 ± 0.03
Other works: PE* kJ/mol	-39.8 ± 0.08 [57]	-41.8 [84]	-45.4 ± 0.03 [57]	–	–
This work: ρ g/cm^3	0.985 ± 0.004	0.993 ± 0.004	0.998 ± 0.004	0.992 ± 0.003	0.996 ± 0.004
Other works: ρ* g/cm^3	0.998 [57]	1.001 [85]	0.998 [57]	1.012± 0.016 [78]	0.997 ± 0.001 [77]
This work: D ($\times 10^{-5}$cm^2/s)	5.8798 ± 0.0009	3.7466 ± 0.0007	2.7159 ± 0.0006	3.2230 ± 0.0004	2.3527 ± 0.0004
Other works: D* ($\times 10^{-5}$cm^2/s)	5.9 [57] ± 0.09	3.9 [85]	2.8 ± 0.06 [57]	2.32 ± 0.05 [78]	2.3 ± 0.02 [77]
This work: T K	298.16 ± 2	298.04 ± 1	298.13 ± 2	298.02 ± 1	298.51 ± 1
Other works: T* K	299.2 ± 1 [57]	298.15 [84]	298.2 ± 1 [57]	301 ± 1 [78]	298.16 [77]

Expt. [76, 86, 87]: PE=−41.5 kJ/mol, ρ = 0.997 g/cm^3, D = 2.3×10^{-5}cm^2/s, 298.15 K

4.3.1 Dictionary of Secondary Structures of Proteins (DSSP) Analysis

The α-helices, β-sheets, and turns are the common secondary structures in proteins with the common element being the presence of characteristic hydrogen bonds. Because their backbone ϕ and ψ angles repeat, helices are classified as repetitive secondary structure. Conversely, if the backbone dihedral angle pairs are the same for each residue, the resulting conformation will assume a helical conformation about some axis in space. Helices were often designated by the number of residues per helical turn and the number of atoms in one hydrogen-bonded ring [88]. Helices are the most abundant form of a secondary structure

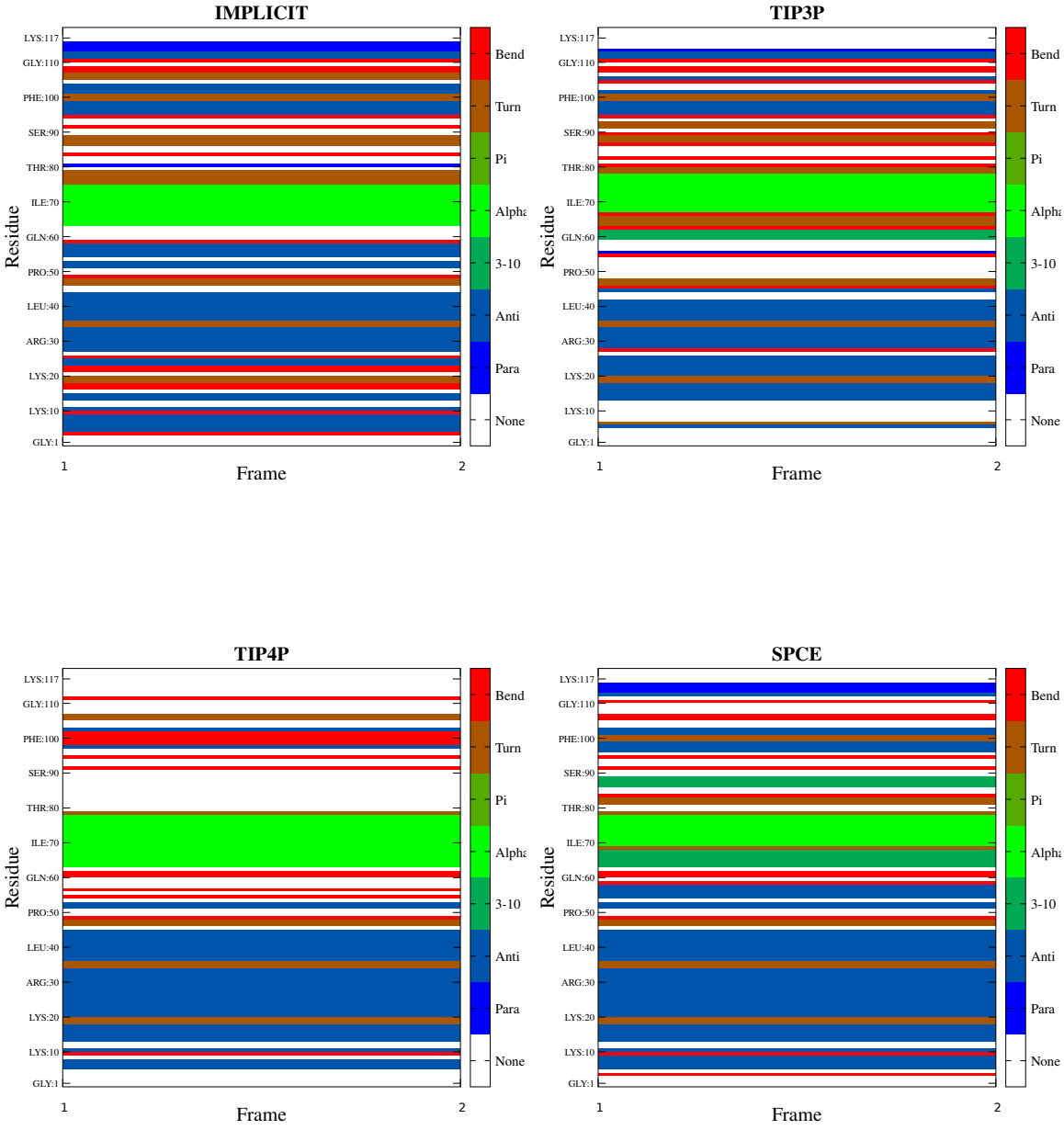
containing approximately 32-38% of the residues in globular proteins [3].

The α helix is the most abundant helical conformation found in globular proteins accounting for 32-38% of all residues [3, 89]. The 3-10 helix is not a common secondary structural element in proteins. Only 3.4% of the residues are involved in 3-10 helices in the Kabsch and Sander database [3], and nearly all those in helical segments containing 1-3 hydrogen bonds. α-helices sometimes begin or end with a single turn of a 3-10 helix. The π-helix is an extremely rare secondary structural element in proteins. Hydrogen bonds within a π-helix display a repeating pattern in which the backbone C=O of residue i hydrogen bonds to the backbone HN of residue $i + 5$. One turn of π-helix is also sometimes found at the ends of regular α-helices.

The β-sheets are another major structural element in globular proteins containing 20-28% of all residues [3, 89]. β-sheets are found in two forms designated as Parallel or Antiparallel based on the relative directions of two interacting beta strands. The basic unit of a beta sheet is a β strand with approximate backbone dihedral angles $\phi = -120$ and $\psi = +120$ producing a translation of 3.2 to 3.4 Å/residue for residues in antiparallel and parallel strands, respectively. Antiparallel β-sheets are thought to be intrinsically more stable than parallel sheets due to the more optimal orientation of the interstrand hydrogen bonds. The hydrogen bonds in a parallel β sheet are not perpendicular to the individual strands resulting in components parallel to the strand [90].

Turns are another classical secondary structures with approximately one-third of all residues in globular proteins. Turns are located primarily on the protein surface and accordingly contain polar and charged residues. The behavior of these secondary structures in the RVFV L protein domain is investigated. The appearance, disappearance or the movement of the helices and sheets in the from a residue to another within the protein domain in the final structures from the simulations are analyzed. The results of this analysis is presented in FIG. 4.2. These results also show how the SPC/E and the implicit solvent simulations produced structures that are not too different from the initial structure. Even though the other water models looks like they also try to maintain the structure of

the peptide, there is enough evidence from FIG. 4.2 that to some extent destabilizes the protein.



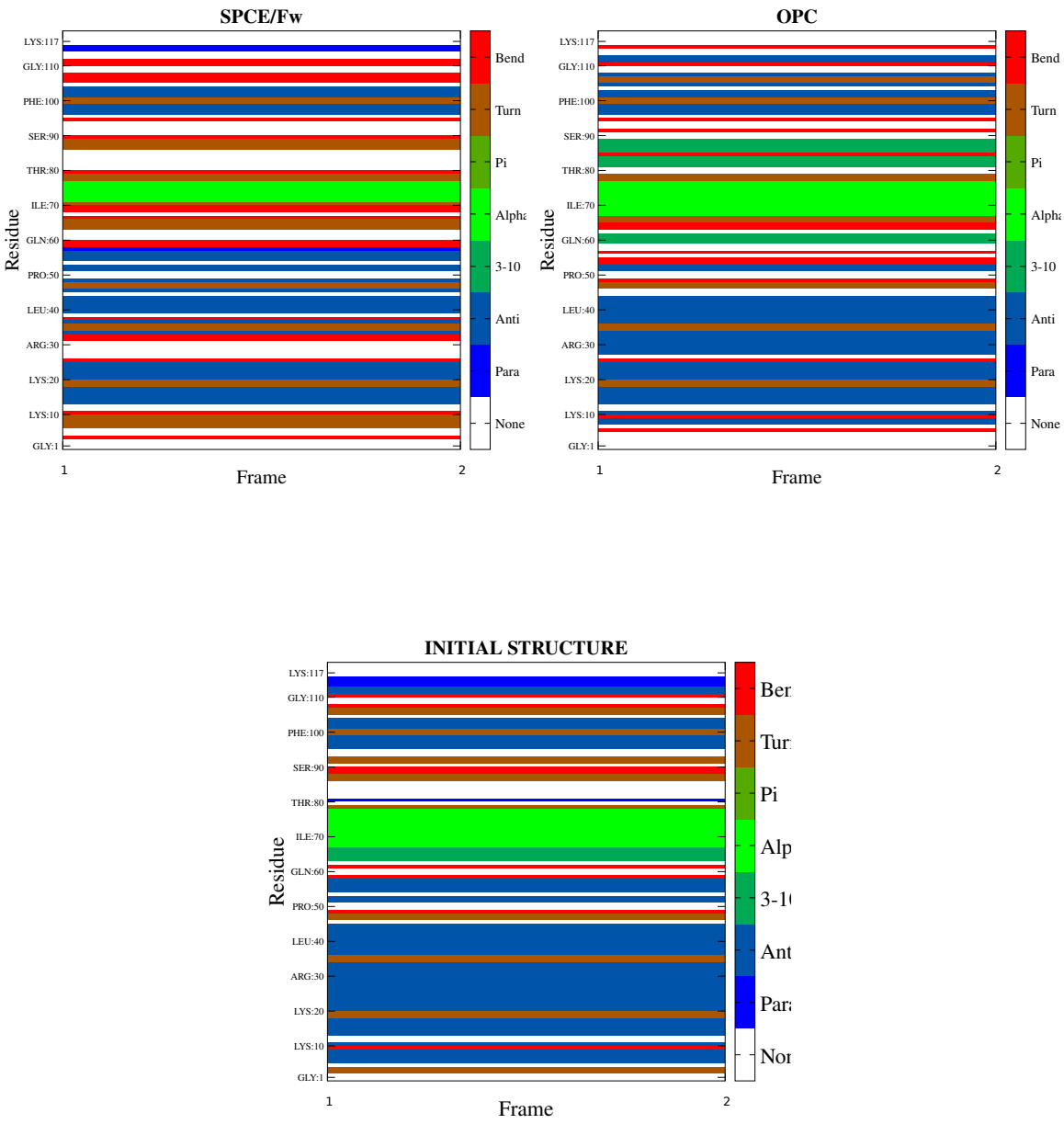


Figure 4.2: DSSP Analysis of the RVFV L protein domain

4.3.2 Structural properties and energetics

In Table 4.2, we present the average values of some structural properties of the peptide over the 100 ns of the NVT simulations. It can be observed from the SASA calculated from the different solvent models that, in simulations where TIP3P, OPC and SPCE/Fw water models were used, the protein surface area increased drastically by 23.75%, 25.23% and 21.74%, respectively whereas, the SPC/E and TIP4P presented a much smaller in percentage increase of 3.98%, and 10.92% respectively (see Table 4.2) for the SASA. This signifies that these water models do not allow the protein to clench tight, but rather make the protein bigger/larger than its size in the stable state. The implicit solvent simulation however produces an average SASA value that showed a decrease of 4.96% with respect to the initial structure.

A possible measure of protein size is the radius of gyration, R_g, calculated with Equation (4.1). An approximation of the Stokes radius measurable from size-exclusion chromatography is the hydrodynamic radius, R_{hyd}, Equation (4.2). While R_g is slightly more dependent on the structure of the protein of interest than R_{hyd}, their ratio R_g/R_{hyd} provides information on the molecular shape. The characteristic R_g/R_{hyd} value of a globular protein is $\approx$ 0.77 or $(3/5)^{1/2}$ [91]. When molecules deviate from globular to nonspherical or elongated structures, then R_g/R_{hyd} tends toward values away from 0.77. The R_g/R_{hyd} values Table 4.2 shows that the RVFV L protein domain is strongly not spherical, with a ratio ranging from 0.491 to 0.522. Correlation between the R_{hyd} of folded or unfolded proteins and the number of residues indicates that the 117 residues domain in the different water models are consistent with denatured proteins through the empirical equation $R_{hyd} = (2.21 \pm 1.07)117^{0.5\pm0.02}$ [92].

Also worth mentioning is that a plot of the ratio of the radius of gyration to the hydrodynamic radius of the peptide in water against the interaction energy, shows the relatively smaller R_g/R_{hyd} ratios having lower peptide-solvent interaction energy than the larger ratios. This suggests that the solvent stabilizes better those structures leading to smaller

ratios. Another useful property is the end-to-end distance R_{ee} defined as the distance between the two end residues of the polypeptide chain. This describes the flexibility of the protein domain. Table 4.2 also shows the RMSD between the starting structure (6QHG) which also corresponds to the structure with the lowest potential energy and the other structures from the different simulations. These RMSDs are of considerable importance, indicating that the models have significant structural differences between each structure and the initial structure across the different water models. The structures obtained from the implicit solvent simulation however showed an average RMSD of 0.4 nm which is also considerably large. This was expected the since initial structure is considered to be the most stable.

Table 4.2: Property and energetics evaluation of RVFV peptide in the different water models at $T = 293.15$ K: Root-mean-square deviation, $RMSD$, Radius of gyration, Rg, Hydrodynamics radius R_{hyd}, End-to-end distance R_{ee}, Solvent-accessible surface area, $SASA$, Potential energy PE, and Interaction energy, E_{int}.

Model	$RMSD$ (nm)	R_g (nm)	R_{hyd} (nm)	R_g/R_{hyd}	R_{ee} (nm)	$SASA$ (nm^2)	PE (kJ/mol)	E_{int} (kJ/mol)
TIP3P	0.63 ± 0.04	1.64 ± 0.02	3.14 ± 0.02	0.522	2.54 ± 0.38	833.7 ± 26.1	-5415 ± 259	-428885 ± 703
TIP4P	0.37 ± 0.02	1.51 ± 0.02	3.02 ± 0.02	0.500	2.90 ± 0.39	747.3 ± 21.6	-5787 ± 345	-443791 ± 815
SPC/E	0.33 ± 0.01	1.45 ± 0.01	2.94 ± 0.01	0.493	2.45 ± 0.09	700.5 ± 15.2	-6400 ± 262	-498772 ± 790
SPC/Fw	0.82 ± 0.03	1.65 ± 0.02	3.19 ± 0.02	0.517	2.39 ± 0.19	820.2 ± 20.0	-5396 ± 246	-484335 ± 760
OPC	0.62 ± 0.03	1.65 ± 0.02	3.18 ± 0.02	0.519	2.36 ± 0.32	843.7 ± 18.7	-5278 ± 297	-549610 ± 738
Implicit Solvent	0.33 ± 0.02	1.43 ± 0.02	2.91 ± 0.02	0.491	2.46 ± 0.26	640.3 ± 20.9	-6753 ± 297	
6QHG	**0.00**	**1.51**	**2.91**	**0.519**	**2.28**	**673.71**	**-8227.469**	

4.3.3 Cluster Analysis of MD trajectory of the RVFV L protein peptide

One of the popular Clustering techniques in Machine Learning is the hierarchical clustering algorithms which are either top-down or bottom-up strategic ordering. Bottom-up algorithms treat each "information" as a singleton cluster at the outset and then successively merge or agglomerate pairs of clusters until all clusters have been merged into a single cluster that contains all information. Bottom-up hierarchical clustering is therefore called hierarchical agglomerative clustering (HAC) and does not require a prespecified number of clusters. Top-down clustering requires a method for splitting a cluster and proceeds by splitting clusters recursively until individual documents (trajectories) are reached. The

monotonicity of the merge operation in HAC is one fundamental assumption. In this work, the bottom-up hierarchical clustering approach was employed to cluster the MD trajectories of the RVFV L protein domain. Hierarchical agglomerative clustering on the back bone atoms of the peptide using average-linkage and stopping when either 5 clusters are reached or the minimum euclidean distance between clusters is 3.0 Å was used. The results from the clustering analysis is presented in an RMSD time series plot in FIG. 4.3. Under the stopping criteria used, it was found that across all the explicit solvents and the implicit solvent simulations, a cluster of 5 was produced. The fractional cluster proportion within each solvent environment however varies. FIG. 4.3 again shows how the peptide atomic position changes along the 100 ns run with respect to the position.

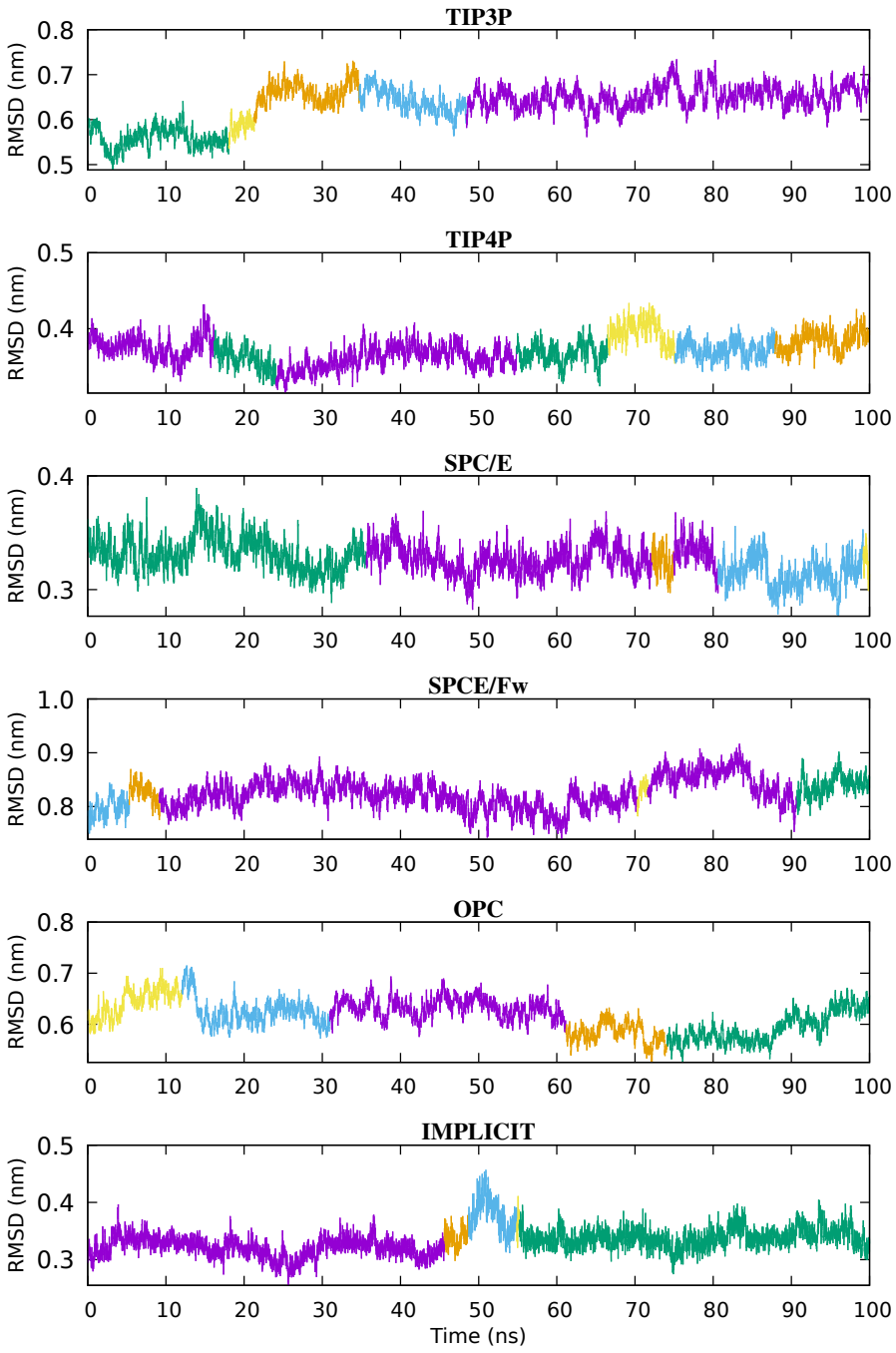


Figure 4.3: Cluster distribution along the MD trajectory of the RVFV domain from the hierarchical agglomerative clustering. RMSD values as a function of time over the trajectory which are colored based on their cluster memberships along the 100 ns MD NVT runs at 293.15 K.

4.4 Conclusions

Using molecular dynamics, we studied the structural and dynamic behavior of RVFV L protein C-terminal domain in both explicit and implicit solvent. Hierarchical agglomerative clustering analysis was also performed on the atomic trajectories obtained from the MD simulation. It was found that under the clustering criteria used, 5 clusters were obtained in all the water models and the implicit solvent environment. The structures obtained from the simulations are far from being globular as evidenced by the ratio R_g/R_{hyd} in the range [0.491 0.522]. Nonetheless, in the SPC/E water model and the implicit solvent, the protein domain like to remain compact with the α-helices and β-sheets somewhat in place likes the initial structure (see FIG. 4.2 and FIG. 4.4).

Based on energetics, the more stable structures were the structures obtained from implicit solvent simulation with the next more stable being structures from the SPC/E water model simulation. The RVFV L protein domain structures produced from the OPC water models turn to have the lowest interaction energy showing a strong attraction between the peptide and the water molecules. The interaction energy also shows how the models obtained from the TIP3P water model simulation demonstrate the most repulsiveness as compared to the other systems. The structural characterization of the atomic trajectories enabled a better understanding of the structural and dynamical behavior of the RVFV domain under the conditions in the water models studied along time.

Considering the fact that information on RVFV L protein or its domains being sparse, the findings presented in this study on the structural behavior of the RVFV L-protein domain will facilitate protein-solvent and protein-protein interaction studies as well as the journey toward drug discovery. In an ongoing research on this and other RVFV L-protein domains, exploration of the structural behaviors in dense solvents such as glycerol and its aqueous solutions is being investigated.

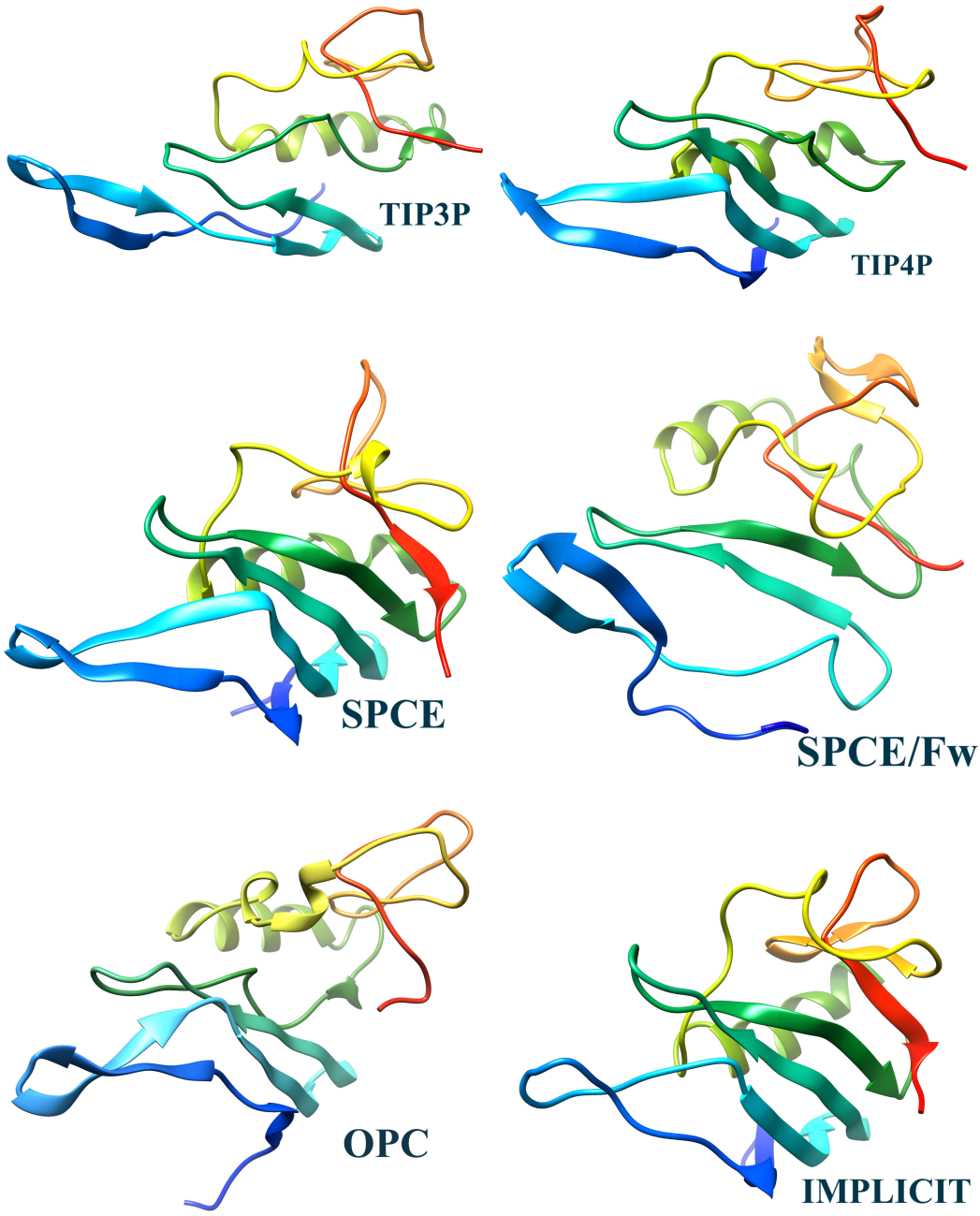


Figure 4.4: Atomic structure of RVFV L protein C-terminal domain from the various environments studied. N and C-termini are colored with blue and red, respectively.

Chapter 5: Computing the Structural Dynamics of RVFV L Protein Domain in Aqueous Glycerol Solutions

5.1 Introduction

Many biological and biotechnological processes are controlled by protein-protein and protein-solvent interactions. In order to better understand, predict and optimize such processes, it is valuable to understand how solvents of additives such as salts, sugars, polyols and denaturants affect proteins-solvent interactions. This can be done through an all-atom modeling of the components using Molecular Dynamics (MD) for analysis of the system thermodynamic and energetic properties. In the course of the MD simulations, it is possible to observe the time evolution of the interfacial dynamics of complex molecular structures; either localized around particular macromolecules or interacting with molecular liquids. Such details are not yet observable experimentally [93].

Even with such limitations, the interest in experimentally investigating the behavior of peptides or proteins in solvent environments has grown rapidly in recent years. A number of experimental techniques including densimetry, neutron scattering, and dielectric relaxation [94–96], have been employed for a range of cosolvents in order to determine whether proteins are preferentially solvated by a specific solvent or by its cosolvent. It has been shown that the structural stability and biochemical activity of proteins can dramatically be affected by the addition of cosolvents to aqueous protein solutions [97, 98]. It has also been shown that the relative abundance of each solvent in the solvation shell of proteins in solvent mixtures has a critical impact on their properties [99].

Computationally, local solvation preferences can be quantified over the entire protein

surface from extended molecular dynamics simulations [100]. Though some cosolvents denature proteins, others preserve the structure of the protein. Solvents such as urea are denaturants and polyols like glycerol and sugars are protectants [9, 10]. Ou W. *et al.* [11] studied the effects of glycerol in the refolding, reactivation, unfolding, and inactivation of guanidine-denatured creatine kinase by observing the fluorescence emission spectra, the circular dichroism spectra, and by recovery and inactivation of enzymatic activity and aggregation. Their results showed that low concentrations of glycerol ($< 25\%$) improve the refolding yields of creatine kinase but high glycerol concentrations decrease its recovery. Glycerol also favors the secondary structural formation and inhibits aggregation of creatine kinase as proline does.

In another study by Rariy and Klibanov [101], unfolded and reduced hen egg-white lysozyme was refolded and reoxidized in glycerol containing varying amounts of water. A densimetric investigation of the interactions between the solvent components in the glycerol-water mixtures (between $10 - 40$ by vol % glycerol) and seven proteins carried out in the acid pH region showed that all the proteins were consistently preferentially hydrated in all cases. This was expected since such thermodynamically unfavorable interactions - addition of the proteins to the mixed solvent resulted in an increase in glycerol chemical potential, tend to minimize the surface of contact between proteins and glycerol and in this way, stabilize the native structure of globular proteins [94].

M. Farnum and C. Zukoski [102], in a related study, investigated glycerol and ionic strength effect on the solubility and strength of interactions of bovine pancreatic trypsin inhibitor in a related study. The two variables in their study were found to have opposite effects on the intermolecular forces. Attractions increased with NaCl whereas repulsions increased with glycerol concentration. The bovine pancreatic trypsin inhibitor follows the same general phase behavior as other globular macromolecules where a robust correlation between protein solution second virial coefficient and solubility has been developed.

Even with the multiple perspectives and results from laboratory experiments on different proteins in hand, it is important to have a visual understanding of the dynamics of protein

structure and changes in different, higher density solvent environments. This requires a molecular-level understanding of the dynamics of proteins in these environments. Despite the many experimental advancements, protein dynamic properties still remain comparably less well understood. In fact, a comprehensive molecular picture of protein (de)stabilization by cosolvents has so far remained elusive. Specifically, a molecular-level understanding of protein structural behavior in the presence of some simple solvents is still lacking. To this end, we conduct an extensive, all-atom molecular dynamics simulation of the RVFV L protein fragment in glycerol and its aqueous solutions. We focus on the structural fate of RVFV L protein fragment as solutes in glycerol and its aqueous solutions in different concentrations by molecular weight.

This chapter is arranged in four sections. Following this section is Section 5.2, where the computational approaches and methods are discussed. In Section 5.3, we elaborate on the results obtained from the MD simulations. Section 5.4 provides the conclusions from the study.

5.2 Methods

In our simulations, ten different solvents, pure glycerol and its aqueous mixtures at 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90 of glycerol:water concentrations in molecular weights are considered. The solute in each of these solvents is a C-terminal domain from the RVFV L protein. This fragment from the L protein is composed of 117 amino-acids (1706-1822aa). The system sizes, simulation time, and the periodic box sizes are listed in the Table 5.1.

Table 5.1: Summary of the simulations performed

$x_1 : x_2$	# atoms	Time (ns)	box size/sidex (Å)
100:00	43849	100×7	71.421
90:10	44230	100×5	71.527
80:20	44611	100×9	71.580
70:30	44992	100×5	71.961
60:40	45373	100×5	72.502
50:50	45754	100×10	73.061
40:60	46135	100×5	73.788
30:70	46516	100×4	74.225
20:80	46897	100×8	74.845
10:90	47278	100×7	75.470

To generate the atomic charges, we use the general AMBER Force Field (GAFF) [16, 103], and do this for a glycerol molecule parameterized to reproduce the B3LYP 6-31G* charges. This includes the polarizable continuum model (PCM) [43], which is based on the Merz-Singh-Kollman population analysis [44, 45], and is done using Gaussian09 [46]. The atomic charges are then ported into the AMBER Tools18 [16] to generate the corresponding RESP values which are employed in this study for the glycerol component. Glycerol is a liquid between 291 K and 563 K and it is often used mixed with water in a large variety of relative concentrations. As a check, before solvating the protein domain in glycerol and its aqueous solutions, we validate simulations of the pure glycerol solvent and the aqueous mixtures.

The simulations of pure glycerol contain 3000 glycerol molecules while for the 90:10 glycerol:water mixed systems the simulations involved 2700-1527 glycerol-water molecules, and the 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, 10:90 have 2400-3054, 2100-4581, 1800-6108, 1500-7635, 1200-9162, 900-10689, 600-12216, 300-13743 glycerol-water molecules respectively. These systems were equilibrated with NPT-MD at 298.15 K and 1.013 25 bar via the Langevin thermostat with a collision frequency of $5\,\mathrm{ps}^{-1}$ and a time step of 1 fs along a minimum of 20 ns long trajectories using a 1 fs time step and a 14 Å cutoff with periodic boundary conditions (PBC). The periodic box sizes are indicated in Table 5.1. Ewald sums are used in all calculations for the long-range electrostatics within the particle

mesh implementation (PME). Prior to the NPT-MD simulations, all the solvent boxes are thermalized with NVT-MD simulations for 10 ns after energy minimization. We run a total of 60000 steps of minimization; 50000 of steepest descent followed by 10000 of conjugate gradient to relax the systems. For the water component, the SPC/E water model [49] is used.

Once we achieve the equilibrium density, the 100:00 simulation is followed by an NVE-MD production run for 10 ns at temperatures around 298 K. In this way, we are able to calculate the diffusion coefficients of 100% glycerol. The center of mass of the glycerol molecules is tracked in time within the solution and determined from Eq. 5.1.

$$D = \frac{1}{6t}\frac{1}{m}\sum_{k=1}^{m}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r_i}(t) - \mathbf{r_i}(t_{0k}))^2 + D_{PBC} \tag{5.1}$$

where $\mathbf{r_i}$ is the position of the ith molecule center of mass at time t and N is the number of molecules in the solvent. Each NVE run is split into m time series. Each of the runs starts from a reference position $\mathbf{r_i}(t_{0k})$, and their average is taken as indicated in Eq 5.1. The last term is the correction due to the periodic boundary conditions [51], $D_{PBC} = \frac{2.837297 k_B T}{6\pi\eta L}$, with k_B being Boltzmann's constant, T temperature, L computational box length, and η solvent viscosity. The value for the η is taken from experiment at 298 K: $\eta_{glycerol} = 945$ mPa s [52].

The next step is the preparation of systems with the domain solvated into each of the ten different solvents, as demonstrated in Figure 5.1 with a focus on the peptide. The systems were relaxed with another energy minimization and followed with a long NVT-MD simulation at 298.15 K and the volume that gave the respective equilibrium densities for hundreds of ns. It is from the last 200 ns of these NVT-MD simulations that the solvated RVFV peptides structural and dynamic properties are calculated. The behavior of the secondary structure elements was calculated from the secondary structure analysis according to the Kabsch and Sander procedure [3].

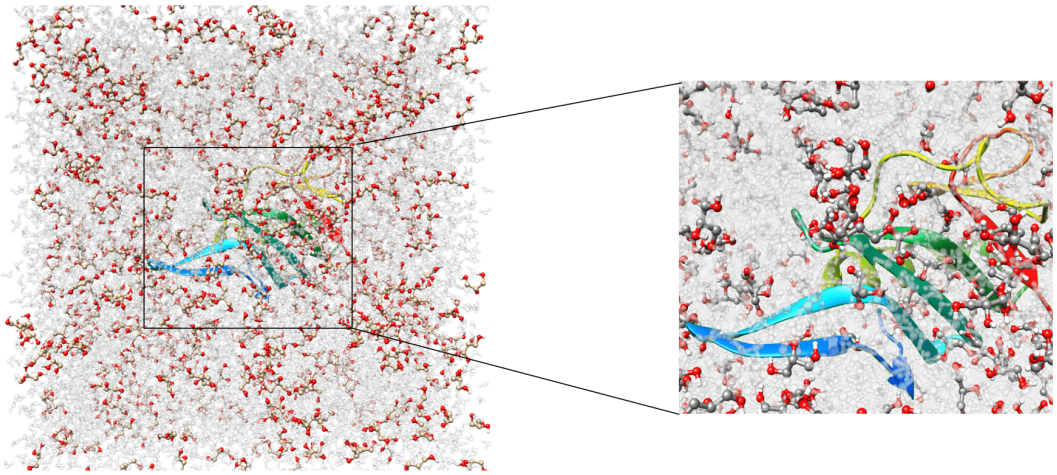


Figure 5.1: Initial conformation of the system. RVFV L protein peptide placed in a cubic periodic box using chimera [2] package.

5.3 Results and Discussion

5.3.1 Properties of all-atom MD simulated solvents

The glycerol solution simulations are performed for quantifying the appropriate behavior of the solvents at 298.15 K. From the NPT-MD simulations, the solvent systems attain equilibrium densities presented in Table 5.2. These densities as a function of glycerol concentration are in very good agreement with experimental glycerol solution densities at 298.15 K.

At these equilibrium densities and temperature of 298.15 K, the radial distribution function, rdf or g(r), of each solvent is calculated and compared with experiment and this also shows a very good agreement with experimental results. Figure 5.2 shows the rdf of glycerol between atom pairs in different molecules. Figure 5.3 provides the label given to the glycerol molecule atoms entering in the considered pairs between molecules. The calculated peak positions for the six atom pairs depicted are 1.85, 1.84, 2.80, 2.80, 2.80, 2.80 Å, which are in excellent agreement within standard deviation with experimental values of 1.77 ± 0.61, 1.80 ± 0.63, 2.73 ± 0.87, 2.76 ± 0.78, 2.76 ± 0.80, 2.76 ± 0.90 Å [58]. Figure 5.2 depicts the rdf of the atom pairs in the glycerol molecules.

The self-diffusion coefficients of the pure glycerol is estimated from Eq. (5.1) by considering 40 different time origins, each of which is a 0.5 ns NVE-MD time evolution simulation. The periodic boundary conditions corrected self-diffusion coefficient from the simulation at 298 ± 1 K are $(1.93 \pm 0.02) \times 10^{-7}$ cm^2/s. The calculated diffusion coefficient compares well with the experimental value of 1.7×10^{-7} cm^2/s obtained from the Taylor dispersion method [60] and another simulation with a value in the order ($\times 10^{-7}$) using the AMBER force fields [39] at ≈ 298.51 K. However, this experimental value is larger than the diffusion coefficient obtained from the NMR pulsed magnetic field gradient [61] or modulated gradient spin echo method [62]. Our calculated densities from the solvents composed of $50\% - 90\%$ water produced higher deviation from the experimentally measured densities. Still, these calculated densities are much better than what has been reported in another simulation study [104]. We can therefore conclude that the selected Force Field is modeling the glycerol solutions at 298.15 K adequately for the goals of this work.

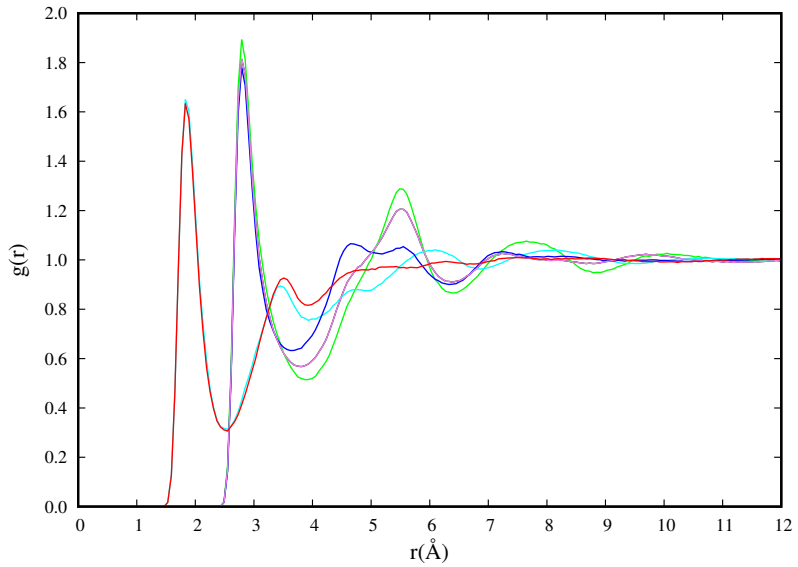


Figure 5.2: Radial distribution function of glycerol at 298 K and equilibrium density 1.259 g/cm^3 with pairs O−H (red), OC−O (black), OC−OC (green), OC−H (cyan), O−O (blue), O−OC (violet), with atoms identification as shown in Fig 5.3.

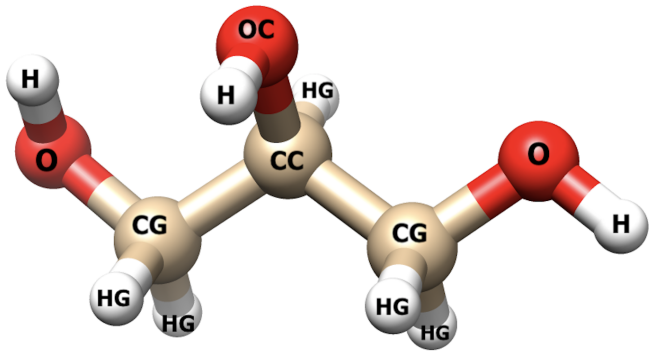


Figure 5.3: Atoms identification in the glycerol molecule.

Table 5.2: Densities, ρ (g/cm^3) of {glycerol (1) + water (2)} mixtures at 298.15 g/cm^3 and atmospheric pressure.

$x_1 : x_2$	Calc. ρ	Expt. [1]
100 : 00	1.259 ± 0.002	1.25791
90 : 10	1.253 ± 0.003	1.25331
80 : 20	1.250 ± 0.002	1.24648
70 : 30	1.230 ± 0.001	1.23632
60 : 40	1.202 ± 0.002	1.22565
50 : 50	1.174 ± 0.002	1.21375
40 : 60	1.146 ± 0.001	1.19845
30 : 70	1.119 ± 0.002	1.18300
20 : 80	1.091 ± 0.002	1.14286
10 : 90	1.063 ± 0.002	1.09524

.

5.3.2 Energetic Evaluation of RVFV L protein domain

Energetics evaluation of the peptide in the various solvents yields results provided in Table 5.3. The average potential energy of the peptide over the last 200 ns of the simulation runs shows a decreasing potential energy as glycerol concentration decreases. This indicates an increasing stabilization with decreasing glycerol concentration. The interaction energy between the peptide and the solvents represents the balance between total potential energy of the system and the sum of individually separated potential energies of the solvent and the peptide: $E_{\text{int}} = E_{\text{full-system}} - (E_{\text{solvent}} + E_{\text{peptide}})$. As observed in Table 5.3, the

solvent-peptide interaction energy in each of the solvents becomes more cohesive as the concentration of glycerol decreases. This suggests that the water-dominant solvents stabilize the structure better, leading to a smaller radius of gyration distribution, as shown in Figure 5.7. Therefore, it can be concluded that the interaction energies of the peptide in aqueous glycerol solution depends on the glycerol concentrations, with the peptide stabilizing propensity increasing as the number of water molecules increases in the solvent. The large fluctuations in the energies of systems with higher glycerol concentrations are attributed to the mobility of the few water molecules in the neighborhood of the peptide that results in frequent changes in the solution surrounding the solutes.

Table 5.3: Energetics evaluation at $T = 298.15\,\mathrm{K}$: Interaction energy, E_{int} and Potential energy PE of RVFV peptide in the solvents.

$x_1 : x_2$	E_{int} ($\mathrm{kJ\,mol^{-1}}$)	PE ($\mathrm{kJ\,mol^{-1}}$)
100:00	-87045 ± 1465	-835 ± 155
90:10	-205431 ± 6910	-1272 ± 179
80:20	-395309 ± 2395	-1705 ± 185
70:30	-462843 ± 3344	-2541 ± 176
60:40	-513576 ± 1553	-2714 ± 273
50:50	-552467 ± 876	-2252 ± 215
40:60	-589103 ± 920	-2751 ± 213
30:70	-621278 ± 900	-3078 ± 246
20:80	-659920 ± 824	-3702 ± 231
10:90	-694648 ± 808	-4397 ± 219
6QHG		**-8227.469**

5.3.3 Properties of RVFV L protein domain in glycerol solutions

Several structural properties of the peptide and the solvents are monitored during the MD simulations. An analysis of the various MD trajectories indicates that the structural impressions of the solvated peptide are acquired during the simulations. The calculated peptide structural properties include root-mean-square deviation, radii of gyration and hydrodynamic, end-to-end distance, and solvent-accessible surface area. The average values of these properties, presented in Table 5.4, are very similar within a particular property for the

peptide across different solvent concentrations.

The root-mean-square deviation (RMSD) of Cα atoms from the initial structure is used to measure the overall conformational drift of a protein over the course of a simulation for the last 200 ns (Figure 5.4). Examining the RMSD as a function of time for all residues in the protein together, it can be seen that there is a relative rise between $\approx 0.4\,\text{nm} - 0.9\,\text{nm}$ with fluctuations along the last 200 ns of each simulation. A large increase occurs in the Cα positional RMSD value in the water-dominant solvents indicating rapid divergence away from the starting conformation. This instability of peptides in water had been observed in other simulation studies [105, 106]. Similar to other studies[107], our MD simulations have shown that the RMSD is increased by adding co-solvents to the glycerol solutions. This is significant as it is an indication that solvents constrain Cα movement with respect to the initial structure. Solvent-accessible surface area (SASA), the surface area of a bio-molecule that is accessible to a solvent is also calculated. We found that in general, the SASA decreased in systems with lower glycerol concentration indicating that as the water molecules increase, the peptide surface exposed to solvent reduces. Under the conditions studied, the peptide have SASA sizes between 963.39 ± 6.46 nm^2 and 758.29 ± 39.41 nm^2.

Table 5.4: Structural property evaluation of RVFV peptide in the solvents at 298.15 K: Root-mean-square deviation, $RMSD$, Radius of gyration, Rg, Hydrodynamics radius R_{hyd}, End-to-end distance R_{e-e}, and Solvent-accessible surface area, $SASA$.

$x_1 : x_2$	100:00	90:10	80:20	70:30	60:40	50:50	40:60	30:70	20:80	10:90	**6QHG**
$RMSD$ (nm)	0.41 ± 0.003	0.45 ± 0.01	0.48 ± 0.01	0.59 ± 0.02	0.64 ± 0.03	0.63 ± 0.02	0.88 ± 0.07	0.48 ± 0.02	0.76 ± 0.02	0.68 ± 0.03	**0.00**
R_g (nm)	1.67 ± 0.004	1.70 ± 0.01	1.66 ± 0.01	1.76 ± 0.01	1.72 ± 0.02	1.61 ± 0.01	1.81 ± 0.04	1.61 ± 0.01	1.54 ± 0.01	1.64 ± 0.02	**1.51**
R_{hyd} (nm)	3.35 ± 0.01	3.38 ± 0.01	3.32 ± 0.01	3.37 ± 0.01	3.34 ± 0.01	3.23 ± 0.01	3.40 ± 0.02	3.19 ± 0.02	3.08 ± 0.01	3.16 ± 0.02	**2.91**
R_g/R_{hyd}	0.499	0.503	0.500	0.522	0.515	0.498	0.532	0.505	0.500	0.519	**0.519**
R_{e-e} (nm)	2.81 ± 0.03	2.69 ± 0.05	2.34 ± 0.21	3.19 ± 0.15	2.70 ± 0.30	3.62 ± 0.31	3.08 ± 0.54	2.24 ± 0.09	2.51 ± 0.10	1.95 ± 0.37	**2.28**
$SASA$ (nm^2)	946.36 ± 30.53	963.39 ± 6.46	929.85 ± 36.53	961.55 ± 27.82	930.23 ± 56.27	837.14 ± 40.42	891.02 ± 76.75	796.71 ± 32.75	759.11 ± 66.75	758.29 ± 39.41	**673.71**

In contrast to natively folded proteins, intrinsically disordered proteins generally lack well-defined 3D structures. Consequently, they explore a large number of distinct conformations, and their conformational properties are thus best described in statistical terms. One useful and informative way of representing this large conformational ensemble is through a distribution of the radius of gyration, R_g, calculated with the Equation $R_g^2 = \sum_{i=1}^{N}(\mathbf{r}_i -$

$\mathbf{r}_{cm})^2/N$ and the hydrodynamic radius, R_{hyd}, calculated from $\frac{1}{R_{hyd}} = \frac{1}{N^2}\sum_{i=1}^{N-1}\sum_{j>i}^{N}\frac{1}{r_{ij}}$. The hydrodynamic radius is an approximation of the Stokes radius measurable through size-exclusion chromatography. In both radii $\mathbf{r}_i$ are atomic position vectors relative to the peptide center of mass, $\mathbf{r}_{cm}$ is the center of mass position vector, r_{ij} are distances between atoms i and j and N is the number of atoms in the peptide. The ensemble averages give a rough measure of how compact a protein is and may be compared to the values for other proteins of similar lengths. We calculated average values of several properties including R_g and R_{hyd} (see Table 5.4).

A common feature observed across the glycerol concentrations along the simulations is that the peptide gives rise to very compact R_g distributions (see Figure 5.7) when compared between the solvent concentrations. We surmise that the peptide prefers to remain compact across the glycerol concentrations along the simulation and at temperature 298.15 K. The solvent molecules trap the instantaneous geometry, and the fate of the protein is to be locked in an instantaneous structure. However, when simulated in the water-dominant solvents, the peptide exhibits a little flexibility, as shown by the range of R_g distribution presented in Figure 5.7.

Because both R_g and R_{hyd} probe the compactness of a disordered protein, and because they may contain complementary information about the distribution of states [108] there have been several studies on the relationship between the R_g and R_{hyd} for disordered proteins and polymers [91, 108–110]. In line with theoretical expectations, it was found that the ratio R_g/R_{hyd} depends substantially on the compaction of the protein chain, so that compact states have ratios ≈ 0.77 or $(3/5)^{1/2}$ [91]. When molecules deviate from globular to nonspherical or elongated/extended structures, then R_g/R_{hyd} tends toward values away from $(3/5)^{1/2}$. Because the relative level of compactness of the chain, when quantified by R_g, also depends on the chain length, the ratio R_g/R_{hyd} also depends on the number of residues (N) of the protein. Recently, these two effects were combined into a single, physically-motivated and empirically parameterized equation that enables one to calculate

R_{hyd} for a configuration of an intrinsically disordered protein from its R_g [111]:

$$\frac{R_g}{R_{hyd}}(N, R_g) = \frac{\alpha_1(R_g - \alpha_2 N^{0.33})}{N^{0.60} - N^{0.33}} + \alpha_3$$

where α_1, α_2 and α_3 are parameters that are fitted to maximize agreement between the model and hydrodynamic calculations. The R_g/R_{hyd} values calculated from the simulations in this study from the different solvent concentrations shows that the RVFV L protein domain maintained its none spherical shape, with a R_g/R_{hyd} ratio in the range $0.498 \leq R_g/R_{hyd} \leq 0.532$ as shown in Table 5.4. Another useful property is the end-to-end distance R_{e-e} defined as the distance between the two end residues of the peptide chain. This describes the flexibility of the protein domain.

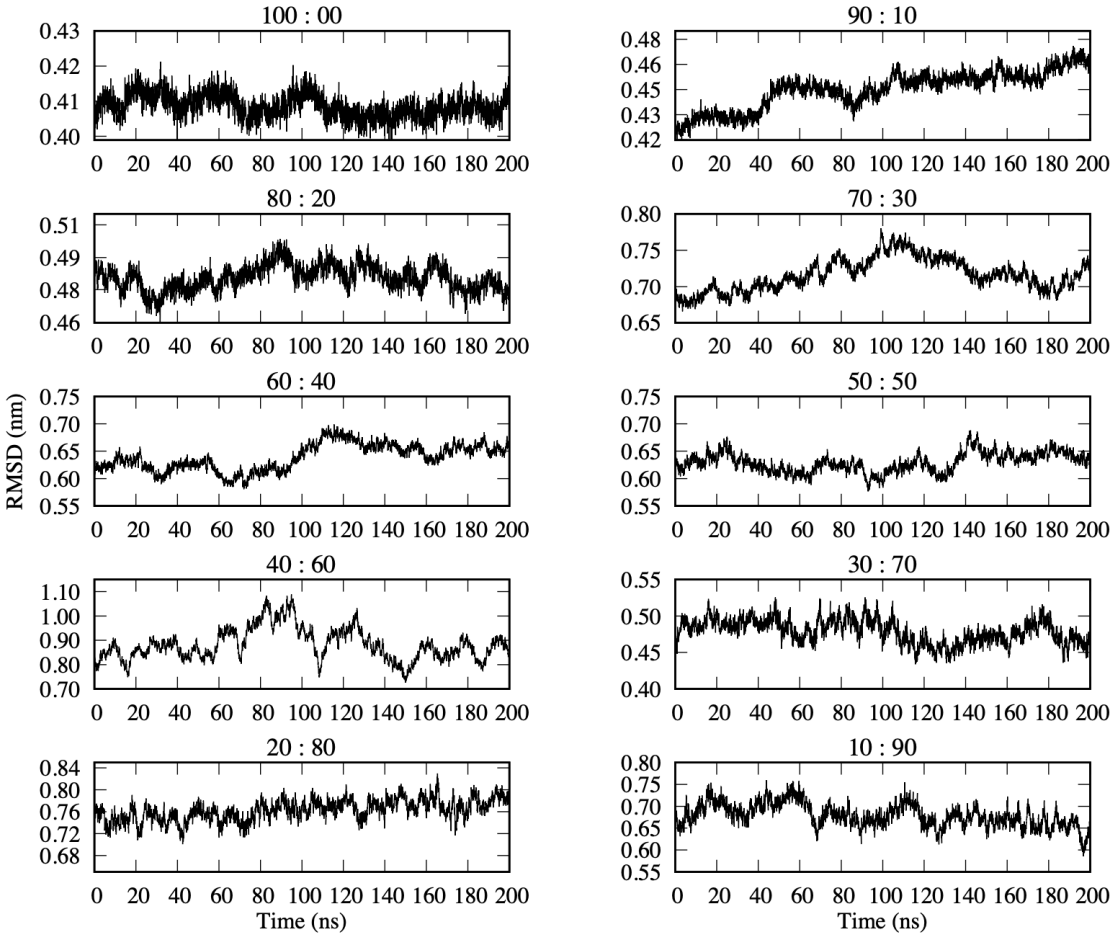


Figure 5.4: Conformational change in RVFV protein domain of the MD simulation, measured as the root-mean-square deviation (RMSD) along the last 200 ns simulation

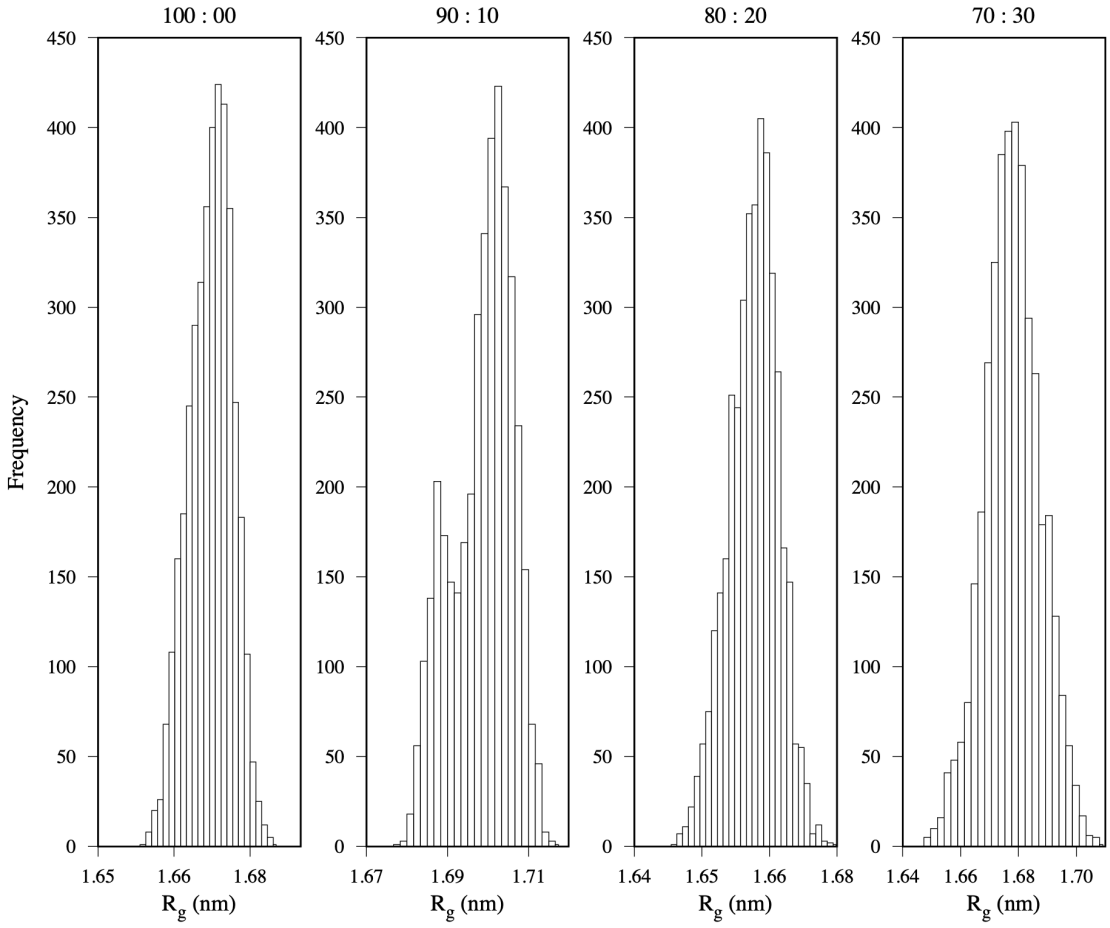


Figure 5.5: Comparison of the peptide R_g distributions as a function of glycerol concentration at 298.15 K and the corresponding equilibrated densities.

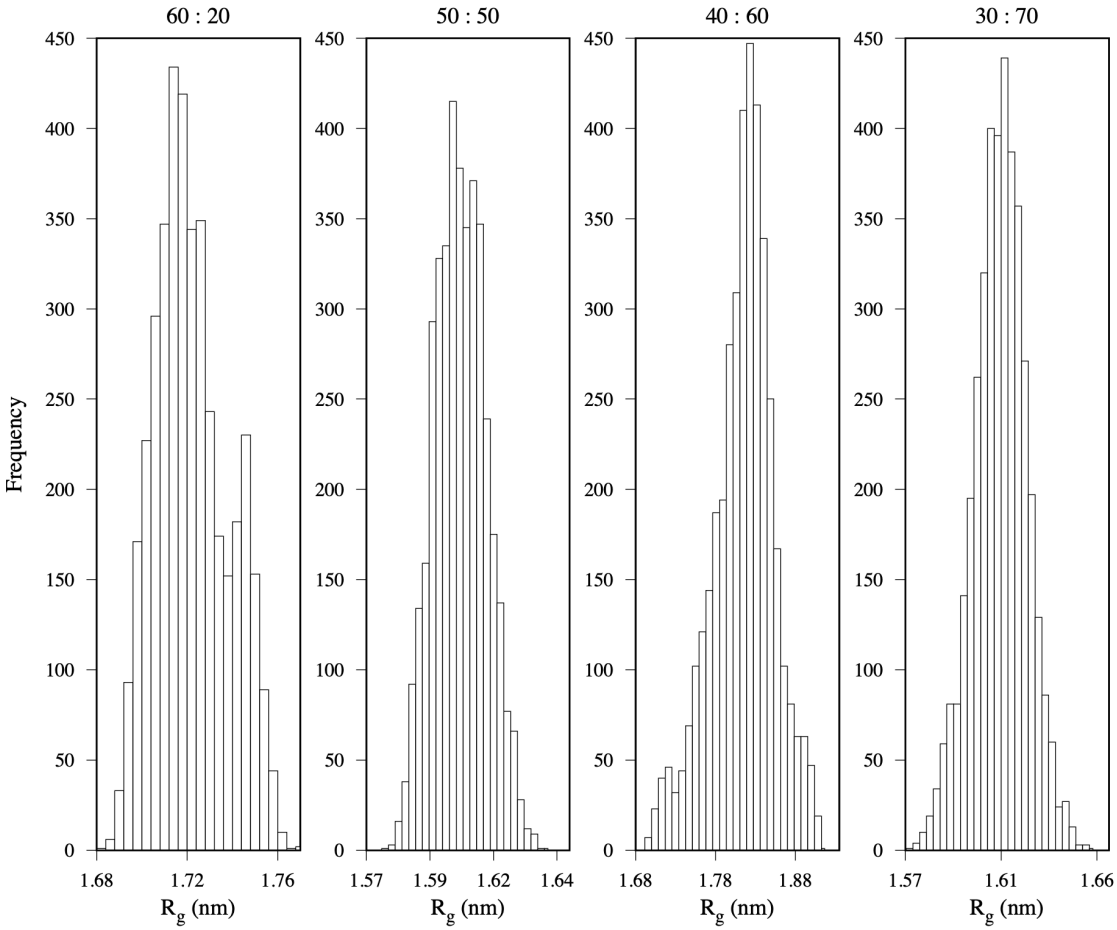


Figure 5.6: Comparison of the peptide R_g distributions as a function of glycerol concentration at 298.15 K and the corresponding equilibrated densities.

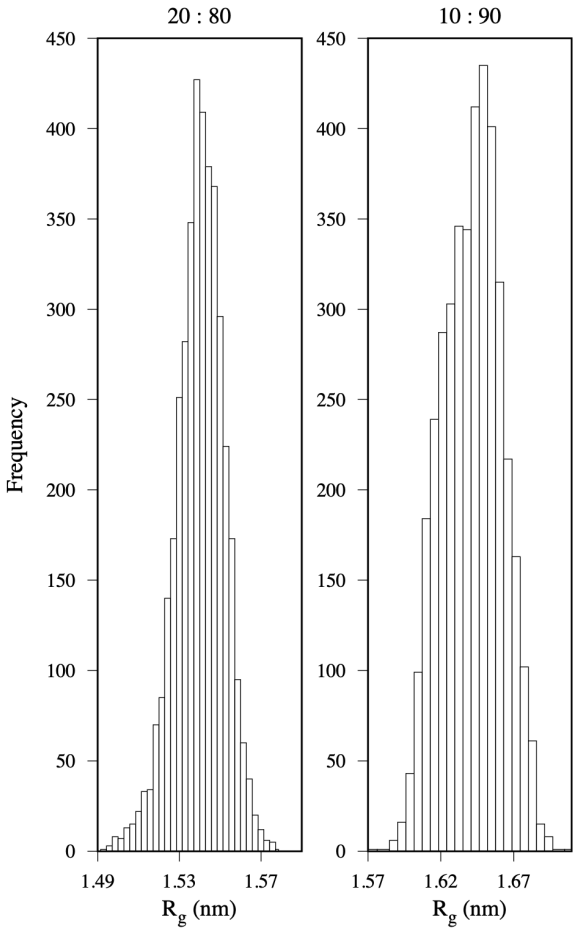


Figure 5.7: Comparison of the peptide R_g distributions as a function of glycerol concentration at 298.15 K and the corresponding equilibrated densities.

5.3.4 Secondary Structure Analysis

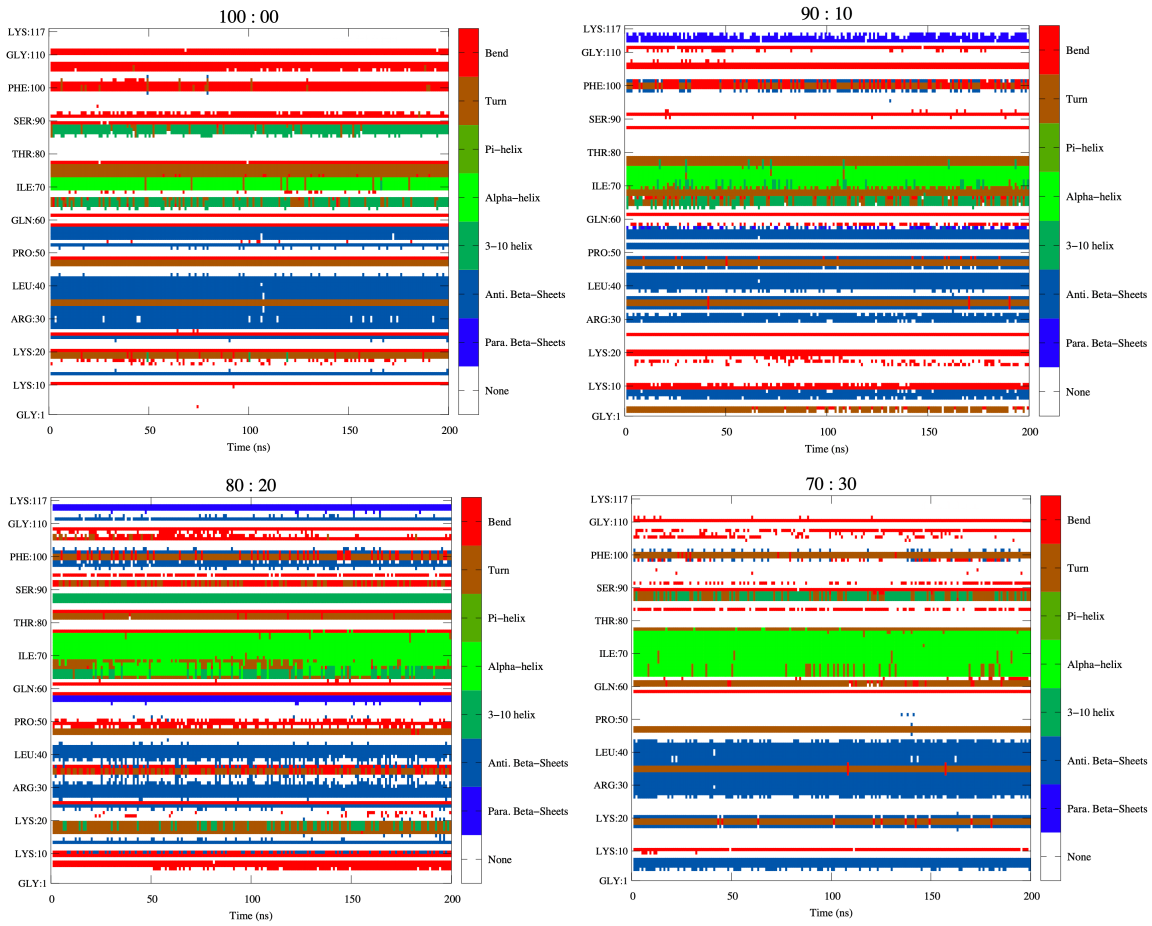
Having established that the degree of conformational change in the RVFV protein domain is modest in the solvent environment, and comparable to that in simulations of other proteins [112], it is informative to examine the extent of conformational drift for the individual helices and sheets. To investigate this, secondary structure analysis was carried out on the peptide in the solvents. α-helices, β-sheets, and turns are the common secondary structures in proteins with the common element of most of these structures being the presence of characteristic hydrogen bonds. Because their backbone ϕ and ψ angles repeat, helices

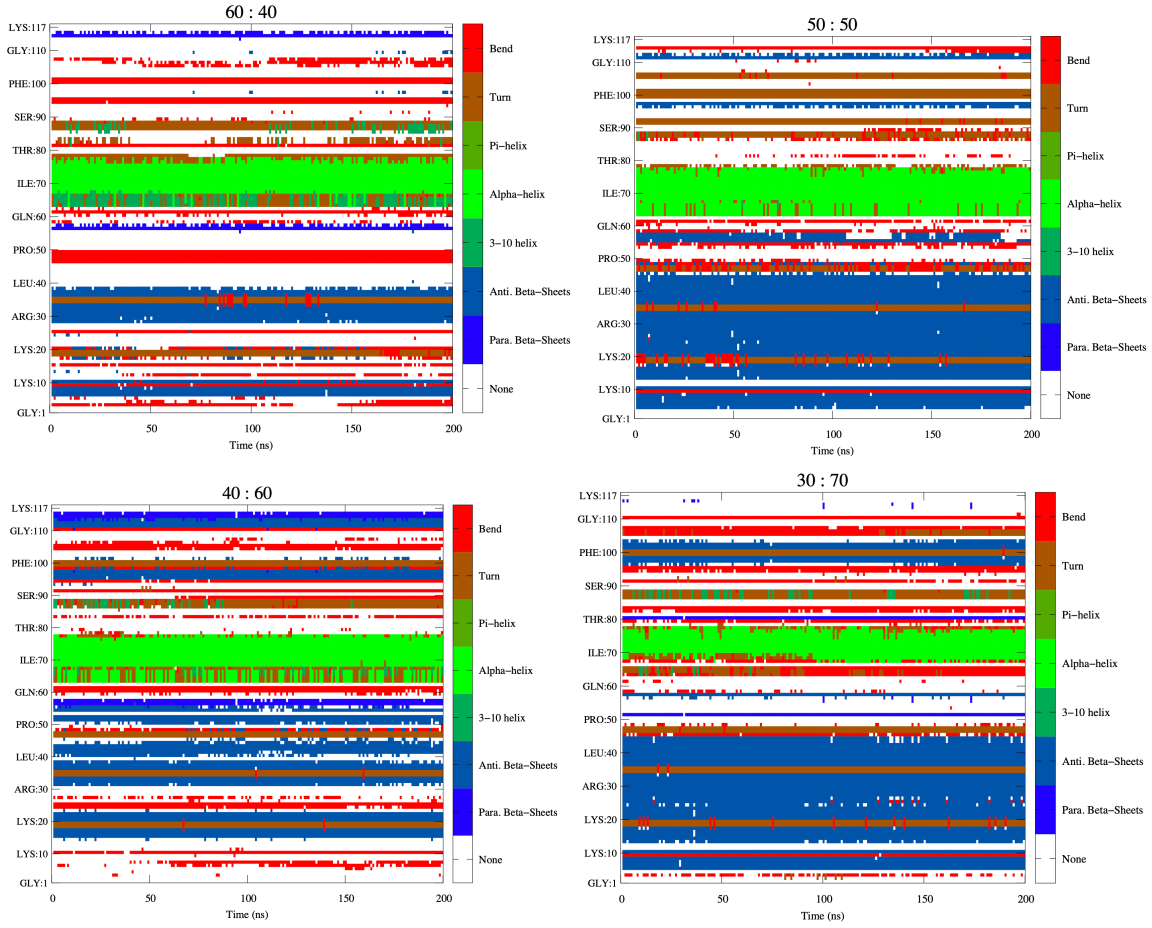
are classified as repetitive secondary structure. Alternatively, if the backbone dihedral angle pairs are the same for each residue, the resulting conformation will assume a helical conformation about some axis in space [88].

Figure 5.8 shows the results from the secondary structure analysis of the peptide. We find that the peptide structure does not deteriorate much from the initial structure with time. Comparison of pure glycerol with glycerol/water solutions however shows some amount of difference. Compared to pure glycerol, the largely α-helical conformation of the peptide is kept throughout the last 200 ns of simulations. However, some local deviations from α-helicity were observed in the C- and N-termini of the peptide.

The β-sheets are another major structural element in globular proteins [3, 89]. They are found in two forms designated as Parallel or Antiparallel based on the relative directions of two interacting beta strands. The basic unit of a beta sheet is a β strand with approximate backbone dihedral angles $\phi = -120$ and $\psi = +120$, producing a translation of 3.2 to 3.4 Å/residue for residues in antiparallel and parallel strands, respectively. Due to the more optimal orientation of the interstrand hydrogen bonds, antiparallel β-sheets are thought to be intrinsically more stable than parallel sheets. Hydrogen bonds in a parallel β sheet are not perpendicular to the individual strands resulting in components parallel to the strand [90].

The analysis of secondary structure elements presented in Figure 5.8 shows that the helices in this region of the L protein are relatively stable in the different glycerol concentrations with increasing stability as glycerol concentration decreases. This implies that the observed conformational changes or large RMSD values observed in Figure 5.4 are not generally caused by unfolding of the structure. Therefore, it can be concluded that in this cap-binding domain of RVFV L protein, relatively large motions occur as the glycerol concentration decreases or when the solvent density decreases and the observed flexibility is inherent to the structure. Conformational change appears to originate from an opening of the helix-loop regions. The secondary structure analysis also shows some preservation of the β-sheets along the simulation in all solvents with some few disappearance in some residues.


100 : 00
90 : 10
80 : 20
70 : 30
LYS:117
GLY:110
PHE:100
SER:90
THR:80
ILE:70
GLN:60
PRO:50
LEU:40
ARG:30
LYS:20
LYS:10
GLY:1
0
50
100
150
200
Time (ns)
Bend
Turn
Pi−helix
Alpha−helix
3−10 helix
Anti. Beta−Sheets
Para. Beta−Sheets
None


60 : 40
50 : 50
40 : 60
30 : 70
LYS:117
GLY:110
PHE:100
SER:90
THR:80
ILE:70
GLN:60
PRO:50
LEU:40
ARG:30
LYS:20
LYS:10
GLY:1
0
50
100
150
200
Time (ns)
Bend
Turn
Pi−helix
Alpha−helix
3−10 helix
Anti. Beta−Sheets
Para. Beta−Sheets
None

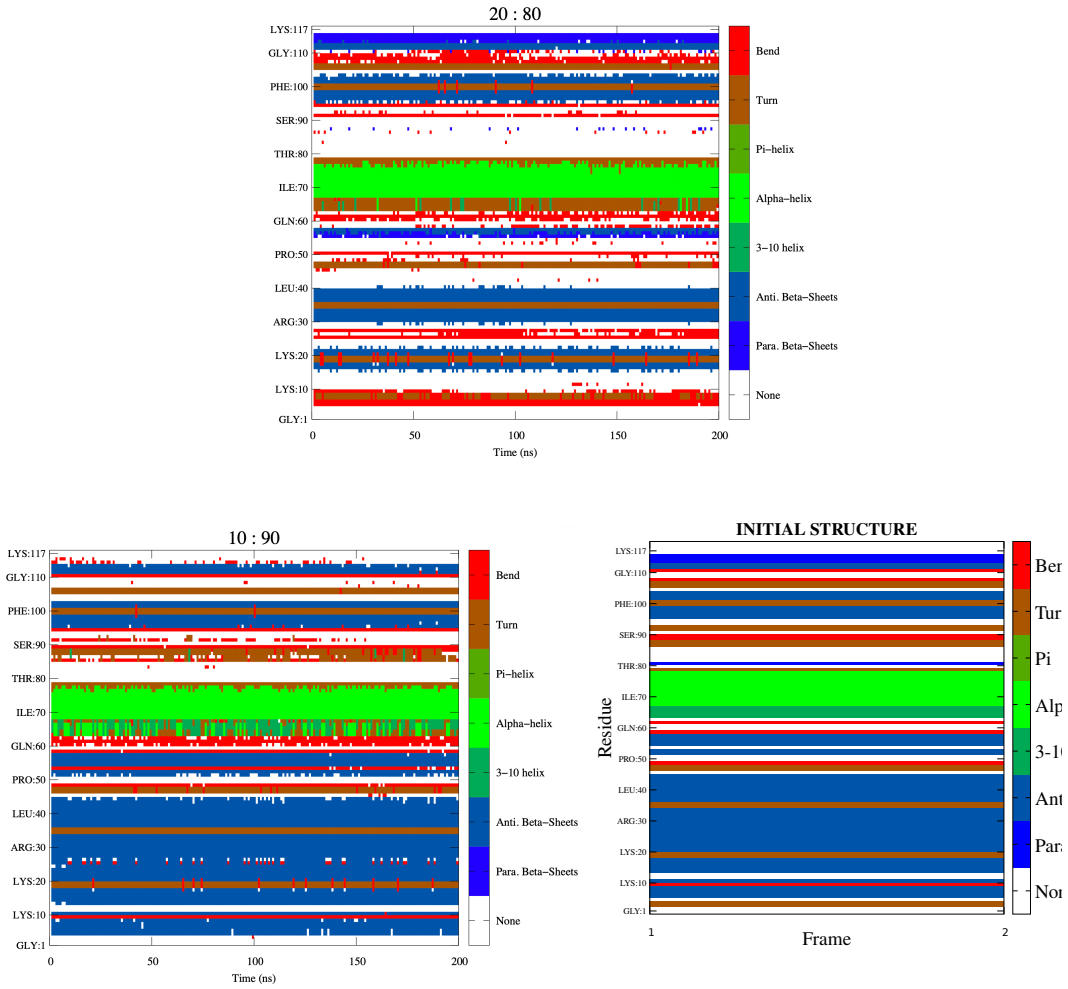


Figure 5.8: Secondary Structure Analysis (according to the Kabsch and Sander procedure [3]) of the RVFV L protein domain at 298.15 K in the pure glycerol, and aqueous glycerol solutions.

5.3.5 Cluster Analysis of the MD trajectory of the RVFV L protein peptide

Clustering is a general Machine Learning technique that can be applied to any collection of data elements or points where a function measuring distance between pairs of points is available [113, 114]. A clustering algorithm partitions the data points into a disjoint collection of sets called clusters. The points in one cluster are ideally closer, or more similar, to each other than to points from other clusters. The use of clustering algorithms to group

similar conformations visited during a MD simulation is not a new concept [115, 116]. A wide variety of cluster algorithms have been applied in many studies to MD trajectories in order to search for similarities among structures by grouping similar conformations. A subset of publications developing and applying machine learning algorithms to analyze MD trajectories covers some of the earliest MD simulations to very recent studies.[115, 117–121]. When clustering the molecular configurations from a MD trajectory, ideally each clustering algorithm should group similar molecular configurations into distinct sets or groups. This gives a refined view of how a given molecule is sampling conformational space and allows direct characterization of the separate conformational substates visited by the MD simulation [122]. It should be noted here that large-scale conformational change during the MD can lead to high variance for the calculation of time-independent properties such as estimation of free energetics [123]. By clustering the trajectory into distinct sub-state populations. We can minimize this variance and provide more useful information about the ensemble of conformations sampled by MD.

This work applied the well-known pairwise distance metric clustering algorithm, Agglomerative Hierarchical Clustering, to the MD trajectories. The bottom-up hierarchical clustering approach is employed to cluster the trajectories of the RVFV L protein domain solvated in the glycerol and its aqueous solutions. The clustering on the backbone atoms of the peptide using average-linkage with stopping when either 5 clusters are reached or the minimum euclidean distance,

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2},$$

between clusters $\mathbf{p}$ and $\mathbf{q}$ is 0.3 nm was used. A visualization of the results obtained from the clustering analysis is presented as a radius of gyration versus hydrodynamic radius correlation plot shown in Figure 5.9. The individual clusters are represented by colors. In all the solvent concentrations, 5 clusters were obtained based on the stopping criteria used in the clustering.

A comparison of cluster sizes, average distances of each conformation in the cluster to its centroid, and average distances between clusters within each solvent is presented in Table 5.5. It was found that the average distance of each conformation in the cluster to its centroid spans a large range. The range is from 0.103 ± 0.004 nm to 0.393 ± 0.023 nm with the larger distances observed in the water-dominant solvents. This result further explains the large RMSD and R_g values we noticed earlier in the water dominant solvents. The result also reflects the relevance of understanding the transport mechanism of the RVFV peptide in aqueous glycerol environments.

Table 5.5: Comparison of the cluster sizes of the RVFV L protein peptide in each of the {glycerol (1) + water (2)} solvents: Average distance between clusters within solvents, d_s (nm) and Average distance to centroid, d_c (nm)

$x_1 : x_2$	d_s (nm)	d_c (nm)	Cluster Sizes
100 : 00	0.083 ± 0.002	0.103 ± 0.004	1374, 975, 631, 533, 487
90 : 10	0.096 ± 0.004	0.143 ± 0.008	1071, 960, 728, 660, 581
80 : 20	0.110 ± 0.002	0.157 ± 0.010	1089, 813, 803, 656, 639
70 : 30	0.146 ± 0.066	0.214 ± 0.008	1112, 863, 834, 597, 594
60 : 40	0.018 ± 0.012	0.283 ± 0.011	1237, 867, 755, 689, 452
50 : 50	0.171 ± 0.009	0.262 ± 0.010	1176, 1061, 1037, 431, 295
40 : 60	0.241 ± 0.020	0.393 ± 0.023	1026, 994, 785, 672, 523
30 : 70	0.184 ± 0.006	0.262 ± 0.016	1578, 1140, 593, 435, 254
20 : 80	0.192 ± 0.009	0.264 ± 0.012	1266, 903, 858, 710, 263
10 : 90	0.200 ± 0.026	0.281 ± 0.020	2008, 1149, 527, 256, 60

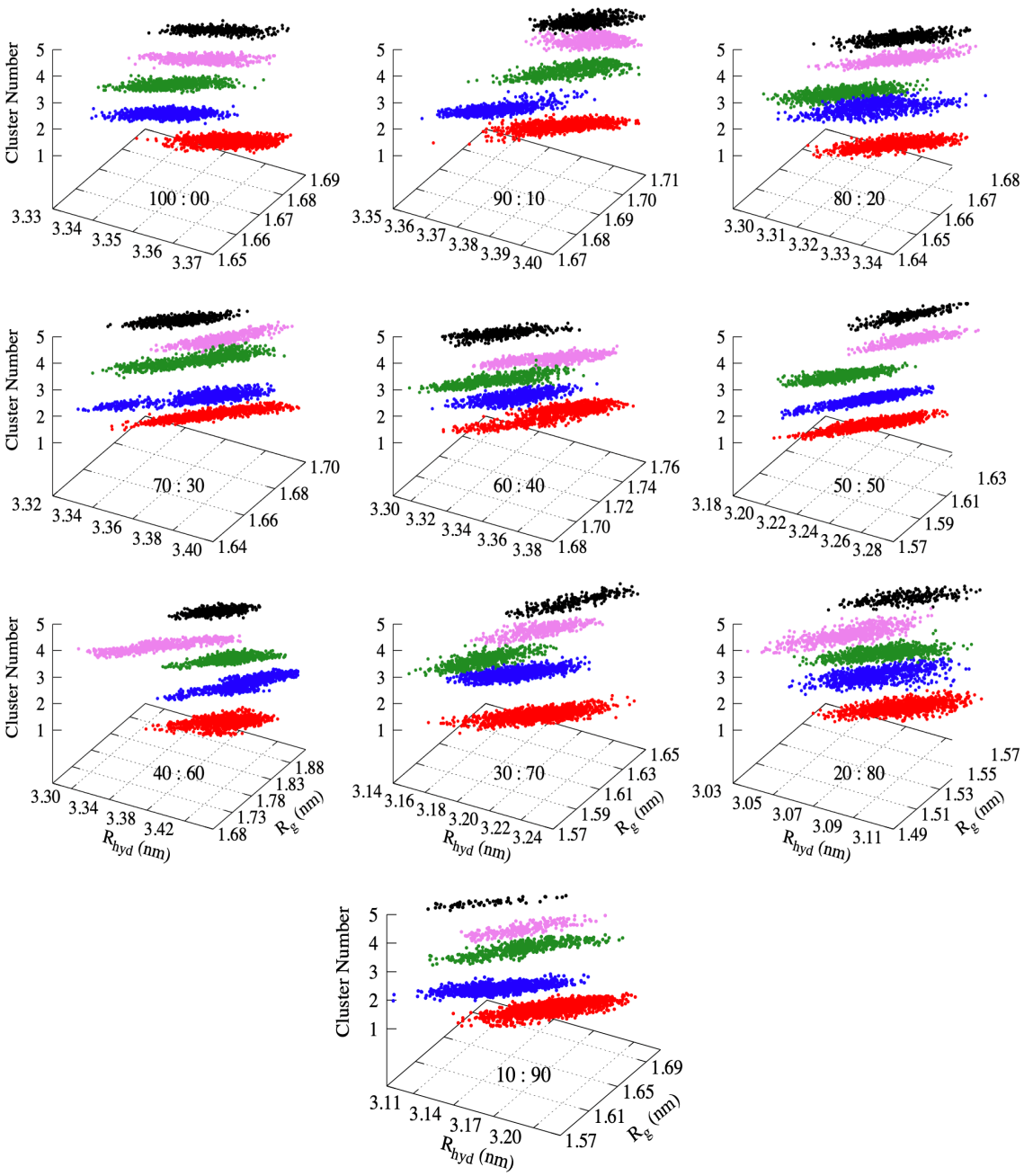


Figure 5.9: Cluster distribution along the MD trajectory of the RVFV domain from the hierarchical agglomerative clustering. Radius of gyration vs Hydrodynamic radius of peptide over the trajectory which is colored based on their cluster memberships along the 200 ns MD runs at 298.15 K

5.4 Conclusions

In this work, we have presented a computational investigation of the structural dynamics and energetics of the Rift Valley Fever Virus L protein domain in glycerol solutions with the goal of understanding and explaining the sensitivity of the peptide to viscous liquids such as glycerol and its mixture with water at both high and low concentrations. We find that solvents have a significant effect on the conformation of the peptide, hence we conclude that the structural conduct and preference of this protein domain is highly sensitive to its accommodating environment.

These effects play an important role in protein folding in the presence of glycerol. We predict that the peptide structure is maintained only when the modeling strategy considers the solvent with less glycerol concentration. We also found that the solvent-peptide interaction becomes more cohesive with decreasing glycerol concentrations. The density and radial distribution function of glycerol solutions calculated with the modified generalized amber force field by including restrained electrostatic potential atomic charges for the glycerol molecules shows a very good agreement with the experimental results at 298.15 K and other simulations. From this, we can conclude that the protein structure studied here undergo relatively little conformational changes in higher glycerol concentrated solvents as compared to its water-dominant equivalent.

Based on the observed energetics, the more stable structures were those obtained from the simulations with higher water concentrations. The RVFV L protein domain structures produced from the water dominant solvent models turn to have the lowest interaction energy showing a strong attraction between the peptide and solvent molecules. This shows that the solvent-peptide become more cohesive with decreasing glycerol concentrations. We predict that the peptide helices and sheets are maintained only when the modeling strategy considers the solvent with less glycerol concentration.

Chapter 6: Concluding Remarks and Future Work

6.1 Concluding Remarks

This last chapter summarizes the findings presented in this dissertation. The research successfully studied the behavior of different lengths of polyacrylamide and also a protein domain from the Rift Valley fever virus L protein in different solvents. The solvents considered are pure glycerol, water and aqueous glycerol solutions in different concentrations by molecular weight. This was done by investigating the energetics, structural and thermodynamic properties of polymers and the biomolecule in the solvents using molecular dynamic simulations. This work was broadly divided into three parts.

The first part presented a dynamical modeling of the structure and energetics of PAM oligomers in glycerol solvents and in water with the goal of elucidating the sensitivity of the PAM structure to viscous non-Newtonian liquids such as glycerol and its mixture with water at high concentration. A modification of the atomic charges of the GAFF was used and also introduced a set of RESP atomic charges that effectively permit a strong temporal localization of the polymer chain in these glycerol solvents. We have also modified the established GAFF atomic charges of the glycerol molecules by the RESP values, obtaining a very good representation of the glycerol liquid at ambient temperature. It is the combination of these two modeling strategies that enable the PAM oligomers to acquire locked-in, swollen, and elongated structures in the glycerol solvent while they are more flexible and prone to remain as less elongated random coils in the glycerol-water solvent. In water, the oligomers are very flexible changing frequently their random coil structure without fully collapsing into a compact globuler. Both the glycerol-water and the pure water solvents behave as θ solvents for the PAM-10, PAM-20, and PAM-30 oligomers considered in this work. In

contrast, the BCC atomic charges case fails to provide a clear and distinct behavior of the solvated oligomers in the different solvents.

The second part again used MD to investigate the structural and dynamic behavior of RVFV L protein C-terminal domain in both explicit and implicit solvent. Here five water models, TIP3P, TIP4P, SPC/E, SPC/Fw and OPC are chosen as solutions for the explicit solvent simulations. It was found that, in the SPC/E water model and the implicit solvent, the protein domain likes to remain compact with the α-helices and β-sheets somewhat in place along the simulation overtime. Based on energetics, the more stable structures were those obtained from the implicit solvent simulation with the next being structures from the SPC/E water model simulation. The RVFV L protein domain structures produced from the OPC water models turn to have the lowest interaction energy showing a strong attraction between the peptide and water molecules. The interaction energy also shows how the models obtained from the TIP3P water model simulation demonstrate the most repulsiveness as compared to the other systems. The structural characterization of the atomic trajectories enabled a better understanding of the structural and dynamical behavior of the RVFV domain under the conditions in the water models studied along time. The findings from this part of the research informed the choice of SPC/E as the water component of the aqueous glycerol solutions used in the last part of the dissertation.

Finally, the last part of the dissertation presented a computational investigation of the structural dynamics and energetics of the RVFV L protein domain in glycerol solutions with the goal of understanding and explaining the sensitivity of the peptide to viscous liquids such as glycerol and its mixture with water at both high and low concentrations. It was found that solvents have a significant effect on the conformation of the peptide in solvation and hence we conclude that the structural conduct and preference of a protein are highly sensitive to its accommodating environment. These effects play an important role in protein folding in the presence of glycerol. From the results of this part of the research, it is reasonable to predict that the peptide structure is maintained only when the modeling strategy considers the solvent with less glycerol concentration. It was also found that the

solvent-peptide interaction becomes more cohesive with decreasing glycerol concentrations. The densities and radial distribution function of glycerol and its solutions calculated with a modified generalized amber force field by including restrained electrostatic potential atomic charges for the glycerol molecules shows a very good agreement with the experimental results at 298.15 K and other simulations. Based on the observed energetics, the more stable structures were those obtained from the simulation with higher water concentration in the solvent. The RVFV L protein domain structures produced from the water dominant solvent models turn to have the lowest interaction energy showing a strong attraction between the peptide and solvent molecules. This shows that the solvent-peptide become more cohesive with decreasing glycerol concentrations. These findings suggest that the peptide helices and sheets are maintained only when the modeling strategy considers the solvent with less glycerol concentration.

6.2 Future Work

In the immediate future, I would like to further explore and analyze other polymers and biomolecules (including peptides from the RVFV L protein) and incorporate machine learning with molecular dynamics for accurate, faster, and on-the-fly analysis of the MD simulations trajectories. Machine learning is transforming almost every area of computational science. The complex and time-consuming calculations in molecular dynamics simulations are particularly suitable for a machine learning revolution and have already been profoundly impacted by the application of existing machine learning methods. I hope to review recent machine learning methods such as random forest, neural networks, support vector regression, Kernel regression, and $k-$nearest neighbors for the prediction of coarse-grained molecular dynamics, quantum-mechanical energies and forces, the extraction of free energy surfaces and kinetics and also generate other network approaches to sample molecular equilibrium structures and compute structural and thermodynamic properties.

Appendix A: *mol2* file showing BCC Atomic Charges for the Glycerol Molecule

```
@<TRIPOS>MOLECULE
GOL
14    13     1     0     0
SMALL
bcc


@<TRIPOS>ATOM
 1 C1          -0.0140    0.0010   -0.2600 c3        1 GOL      0.100100
 2 O1          -0.0680    1.2950    0.2860 oh        1 GOL     -0.612800
 3 C2          -1.2800   -0.7030    0.1940 c3        1 GOL      0.115900
 4 O2          -2.4170   -0.0190   -0.2380 oh        1 GOL     -0.603800
 5 C3           1.2530   -0.7020    0.1870 c3        1 GOL      0.115900
 6 O3           2.3290    0.0910   -0.2520 oh        1 GOL     -0.603800
 7 H1          -0.0140    0.0590   -1.3460 h1        1 GOL      0.061700
 8 H2          -1.2700   -0.7940    1.2780 h1        1 GOL      0.043450
 9 H3          -1.3300   -1.6990   -0.2290 h1        1 GOL      0.043450
10 H4           1.3050   -1.6990   -0.2430 h1        1 GOL      0.043450
11 H5           1.2590   -0.7910    1.2710 h1        1 GOL      0.043450
12 H6           0.7310    1.7510    0.0550 ho        1 GOL      0.425000
13 H7          -2.3250    0.8900    0.0220 ho        1 GOL      0.414000
14 H8           3.1420   -0.2320    0.1080 ho        1 GOL      0.414000
@<TRIPOS>BOND
1    1    2 1
2    1    3 1
3    1    5 1
```

```
4    1    7 1
5    2   12 1
6    3    4 1
7    3    8 1
8    3    9 1
9    4   13 1
10    5    6 1
11    5   10 1
12    5   11 1
13    6   14 1
@<TRIPOS>SUBSTRUCTURE
1 GOL         1 TEMP              0 ****  ****    0 ROOT
```

Appendix B: *mol2* file showing RESP Atomic Charges for the Glycerol Molecule

```
@<TRIPOS>MOLECULE
GOL
14    13     1     0     0
SMALL
resp

@<TRIPOS>ATOM
1 C1          -0.0140    0.0010   -0.2600 c3        1 GOL      0.143911
2 O1          -0.0680    1.2950    0.2860 oh        1 GOL     -0.671611
3 C2          -1.2800   -0.7030    0.1940 c3        1 GOL      0.205929
4 O2          -2.4170   -0.0190   -0.2380 oh        1 GOL     -0.679681
5 C3           1.2530   -0.7020    0.1870 c3        1 GOL      0.205929
6 O3           2.3290    0.0910   -0.2520 oh        1 GOL     -0.679681
7 H1          -0.0140    0.0590   -1.3460 h1        1 GOL      0.071700
8 H2          -1.2700   -0.7940    1.2780 h1        1 GOL      0.018920
9 H3          -1.3300   -1.6990   -0.2290 h1        1 GOL      0.018920
10 H4          1.3050   -1.6990   -0.2430 h1        1 GOL      0.018920
11 H5          1.2590   -0.7910    1.2710 h1        1 GOL      0.018920
12 H6          0.7310    1.7510    0.0550 ho        1 GOL      0.444191
13 H7         -2.3250    0.8900    0.0220 ho        1 GOL      0.441816
14 H8          3.1420   -0.2320    0.1080 ho        1 GOL      0.441816
@<TRIPOS>BOND
1    1    2 1
2    1    3 1
3    1    5 1
```

```
4    1    7 1
5    2   12 1
6    3    4 1
7    3    8 1
8    3    9 1
9    4   13 1
10    5    6 1
11    5   10 1
12    5   11 1
13    6   14 1
@<TRIPOS>SUBSTRUCTURE
1 GOL         1 TEMP              0 ****  ****    0 ROOT
```

Bibliography

Bibliography

[1] M. Wolfe and J. Jonas, "Reorientational motions in compressed viscous fluids: Selectively deuterated glycerol," *The Journal of Chemical Physics*, vol. 71, no. 8, pp. 3252–3262, 1979.

[2] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera—a visualization system for exploratory research and analysis," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[3] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers: Original Research on Biomolecules*, vol. 22, no. 12, pp. 2577–2637, 1983.

[4] M. P. Taylor, Y. Ye, and S. R. Adhikari, "Conformation of a flexible polymer in explicit solvent: Accurate solvation potentials for lennard-jones chains," *The Journal of chemical physics*, vol. 143, no. 20, p. 204901, 2015.

[5] R. Affleck, C. A. Haynes, and D. S. Clark, "Solvent dielectric effects on protein dynamics," *Proceedings of the National Academy of Sciences*, vol. 89, no. 11, pp. 5167–5170, 1992.

[6] C. Finch, "Computer simulation of polymers edited by rj roe prentice-hall inc., englefield cliffs, new jersey, 1991. pp. xii+ 404, price e82. 40. isbn 0-1 3 1-161480-0," *Polymer International*, vol. 30, no. 1, pp. 140–140, 1993.

[7] T. Weidner, N. T. Samuel, K. McCrea, L. J. Gamble, R. S. Ward, and D. G. Castner, "Assembly and structure of α-helical peptide films on hydrophobic fluorocarbon surfaces," *Biointerphases*, vol. 5, no. 1, pp. 9–16, 2010.

[8] J. M. Haile, *Molecular dynamics simulation: elementary methods*. John Wiley & Sons, Inc., 1992.

[9] G. J. Morris, M. Goodrich, E. Acton, and F. Fonseca, "The high viscosity encountered during freezing in glycerol solutions: effects on cryopreservation," *Cryobiology*, vol. 52, no. 3, pp. 323–334, 2006.

[10] Z. Hubalek, "Protectants used in the cryopreservation of microorganisms," *Cryobiology*, vol. 46, no. 3, pp. 205–229, 2003.

[11] W. Ou, Y. Park, F. Meng, and H. Zhou, "Effects of glycerol in the refolding and unfolding of creatine kinase," *Tsinghua Science and Technology*, vol. 7, no. 4, pp. 352–362, 2002.

[12] M. Bishop, M. Kalos, and H. Frisch, "Molecular dynamics of polymeric systems," *The Journal of Chemical Physics*, vol. 70, no. 3, pp. 1299–1304, 1979.

[13] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[14] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb," *Journal of chemical theory and computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

[15] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *Journal of the American Chemical Society*, vol. 112, no. 16, pp. 6127–6129, 1990.

[16] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson *et al.*, "Amber 2018; university of california: San francisco, ca, usa, 2018.© 2019 by the authors," *Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/)*, 2018.

[17] D. R. Roe and T. E. Cheatham III, "Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data," *Journal of chemical theory and computation*, vol. 9, no. 7, pp. 3084–3095, 2013.

[18] A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, "Routine microsecond molecular dynamics simulations with amber on gpus. 1. generalized born," *Journal of chemical theory and computation*, vol. 8, no. 5, pp. 1542–1555, 2012.

[19] R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, and R. C. Walker, "Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald," *Journal of chemical theory and computation*, vol. 9, no. 9, pp. 3878–3888, 2013.

[20] S. Le Grand, A. W. Götz, and R. C. Walker, "Spfp: Speed without compromise—a mixed precision model for gpu accelerated molecular dynamics simulations," *Computer Physics Communications*, vol. 184, no. 2, pp. 374–380, 2013.

[21] M. Chanda and S. K. Roy, *Industrial polymers, specialty polymers, and their applications*. CRC press, 2008, vol. 74.

[22] T.-H. Yang, "Recent applications of polyacrylamide as biomaterials," *Recent Patents on Materials Science*, vol. 1, no. 1, pp. 29–40, 2008.

[23] A. K. Denisin and B. L. Pruitt, "Tuning the range of polyacrylamide gel stiffness for mechanobiology applications," *ACS applied materials & interfaces*, vol. 8, no. 34, pp. 21 893–21 902, 2016.

[24] J. E. Woodrow, J. N. Seiber, and G. C. Miller, "Acrylamide release resulting from sunlight irradiation of aqueous polyacrylamide/iron mixtures," *Journal of agricultural and food chemistry*, vol. 56, no. 8, pp. 2773–2779, 2008.

[25] B. Qin and P. E. Arratia, "Characterizing elastic turbulence in channel flows at low reynolds number," *Physical Review Fluids*, vol. 2, no. 8, p. 083302, 2017.

[26] D. Krapf and R. Metzler, "Strange interfacial molecular dynamics," *Physics today*, vol. 72, no. 9, pp. 48–55, 2019.

[27] H. Wang, H. Zhang, C. Liu, and S. Yuan, "Coarse-grained molecular dynamics simulation of self-assembly of polyacrylamide and sodium dodecylsulfate in aqueous solution," *Journal of colloid and interface science*, vol. 386, no. 1, pp. 205–211, 2012.

[28] G. Wu, Q. Zhu, C. Yuan, H. Wang, C. Li, S. Sun, and S. Hu, "Molecular dynamics simulation of the influence of polyacrylamide on the stability of sodium dodecyl sulfate foam," *Chemical Engineering Science*, vol. 166, pp. 313–319, 2017.

[29] T. E. de Oliveira, D. Mukherji, K. Kremer, and P. A. Netz, "Effects of stereochemistry and copolymerization on the lcst of pnipam," *The Journal of chemical physics*, vol. 146, no. 3, p. 034904, 2017.

[30] A. Asadujjaman, T. E. de Oliveira, D. Mukherji, and A. Bertin, "Polyacrylamide "revisited": Ucst-type reversible thermoresponsive properties in aqueous alcoholic solutions," *Soft matter*, vol. 14, no. 8, pp. 1336–1343, 2018.

[31] M. P. Allen and D. J. Tildesley, *Computer simulation in chemical physics*. Springer Science & Business Media, 2012, vol. 397.

[32] J. Andrews and E. Blaisten-Barojas, "Exploring with molecular dynamics the structural fate of plga oligomers in various solvents," *The Journal of Physical Chemistry B*, vol. 123, no. 48, pp. 10 233–10 244, 2019.

[33] A. V. Egorov, A. P. Lyubartsev, and A. Laaksonen, "Molecular dynamics simulation study of glycerol–water liquid mixtures," *The Journal of Physical Chemistry B*, vol. 115, no. 49, pp. 14 572–14 581, 2011.

[34] R. Chelli, P. Procacci, G. Cardini, R. G. Della Valle, and S. Califano, "Glycerol condensed phases part i. a molecular dynamics study," *Physical Chemistry Chemical Physics*, vol. 1, no. 5, pp. 871–877, 1999.

[35] J. Blieck, F. Affouard, P. Bordat, A. Lerbret, and M. Descamps, "Molecular dynamics simulations of glycerol glass-forming liquid," *Chemical Physics*, vol. 317, no. 2-3, pp. 253–257, 2005.

[36] S. Reiling, M. Schlenkrich, and J. Brickmann, "Force field parameters for carbohydrates," *Journal of Computational Chemistry*, vol. 17, no. 4, pp. 450–468, 1996.

[37] E. R. Hatcher, O. Guvench, and A. D. MacKerell Jr, "Charmm additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol," *Journal of chemical theory and computation*, vol. 5, no. 5, pp. 1315–1327, 2009.

[38] G. I. Egorov, D. M. Makarov, and A. M. Kolker, "Volume properties of liquid mixture of water+ glycerol over the temperature range from 278.15 to 348.15 k at atmospheric pressure," *Thermochimica acta*, vol. 570, pp. 16–26, 2013.

[39] D. A. Jahn, F. O. Akinkunmi, and N. Giovambattista, "Effects of temperature on the properties of glycerol: a computer simulation study of five different force fields," *The Journal of Physical Chemistry B*, vol. 118, no. 38, pp. 11 284–11 294, 2014.

[40] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11 225–11 236, 1996.

[41] R. A. Handler, E. Blaisten-Barojas, P. M. Ligrani, P. Dong, and M. Paige, "Vortex generation in a finitely extensible nonlinear elastic peterlin fluid initially at rest," *Engineering Reports*, vol. 2, no. 3, p. e12135, 2020.

[42] R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of polymeric liquids, volume 2: Kinetic theory*. Wiley, 1987.

[43] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chemical reviews*, vol. 105, no. 8, pp. 2999–3094, 2005.

[44] U. C. Singh and P. A. Kollman, "An approach to computing electrostatic charges for molecules," *Journal of computational chemistry*, vol. 5, no. 2, pp. 129–145, 1984.

[45] B. H. Besler, K. M. Merz Jr, and P. A. Kollman, "Atomic charges derived from semiempirical methods," *Journal of Computational Chemistry*, vol. 11, no. 4, pp. 431–439, 1990.

[46] G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, H. Nakatsuji *et al.*, "Gaussian," *Inc., Wallingford CT*, 2009.

[47] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method," *Journal of computational chemistry*, vol. 21, no. 2, pp. 132–146, 2000.

[48] A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation," *Journal of computational chemistry*, vol. 23, no. 16, pp. 1623–1641, 2002.

[49] H. Berendsen, J. Grigera, and T. Straatsma, "The missing term in effective pair potentials," *Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.

[50] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[51] I.-C. Yeh and G. Hummer, "System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions," *The Journal of Physical Chemistry B*, vol. 108, no. 40, pp. 15 873–15 879, 2004.

[52] K. Schröter and E. Donth, "Viscosity and shear response at the dynamic glass transition of glycerol," *The Journal of Chemical Physics*, vol. 113, no. 20, pp. 9101–9108, 2000.

[53] M. L. Sheely, "Glycerol viscosity tables," *Industrial & Engineering Chemistry*, vol. 24, no. 9, pp. 1060–1064, 1932.

[54] J. R. Rumble, D. R. Lide, and T. J. Bruno, *CRC handbook of chemistry and physics*. CRC press, 2019.

[55] A. Y. Grosberg and A. Khokhlov, *Statistical Physics of Macromolecules (AIP, Woodbury, NY)*. Woodbury, NY : AIP Press, 1994.

[56] G. R. Strobl and G. R. Strobl, *The physics of polymers*. Springer, 1997, vol. 2.

[57] P. Mark and L. Nilsson, "Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k," *The Journal of Physical Chemistry A*, vol. 105, no. 43, pp. 9954–9960, 2001.

[58] J. Towey, A. Soper, and L. Dougan, "The structure of glycerol in the liquid state: a neutron diffraction study," *Physical Chemistry Chemical Physics*, vol. 13, no. 20, pp. 9397–9406, 2011.

[59] K. Amann-Winkel, M.-C. Bellissent-Funel, L. E. Bove, T. Loerting, A. Nilsson, A. Paciaroni, D. Schlesinger, and L. Skinner, "X-ray and neutron scattering of water," *Chemical reviews*, vol. 116, no. 13, pp. 7570–7589, 2016.

[60] G. D'Errico, O. Ortona, F. Capuano, and V. Vitagliano, "Diffusion coefficients for the binary system glycerol+ water at 25 c. a velocity correlation study," *Journal of Chemical & Engineering Data*, vol. 49, no. 6, pp. 1665–1670, 2004.

[61] D. Tomlinson, "Temperature dependent self-diffusion coefficient measurements of glycerol by the pulsed nmr technique," *Molecular Physics*, vol. 25, no. 3, pp. 735–738, 1973.

[62] J. Stepišnik, C. Mattea, S. Stapf, and A. Mohorič, "Molecular velocity autocorrelations in glycerol/water mixtures studied by nmr mgse method," *Physica A: Statistical Mechanics and its Applications*, p. 124171, 2020.

[63] M. Holz, S. R. Heil, and A. Sacco, "Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1h nmr pfg measurements," *Physical Chemistry Chemical Physics*, vol. 2, no. 20, pp. 4740–4742, 2000.

[64] D. Sornette, *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media, 2006.

[65] C. Tanford *et al.*, *Physical chemistry of macromolecules*. Wiley, 1961.

[66] B. H. Bird and S. T. Nichol, "Breaking the chain: Rift valley fever virus control via livestock vaccination," *Current opinion in virology*, vol. 2, no. 3, pp. 315–323, 2012.

[67] M. Bouloy and F. Weber, "Molecular biology of rift valley fever virus," *The open virology journal*, vol. 4, p. 8, 2010.

[68] G. K. Gogovi, F. Almsned, N. Bracci, K. Kehn-Hall, A. Shehu, and E. Blaisten-Barojas, "Modeling the tertiary structure of the rift valley fever virus l protein," *Molecules*, vol. 24, no. 9, p. 1768, 2019.

[69] N. Gogrefe, S. Reindl, S. Günther, and M. Rosenthal, "Structure of a functional cap-binding domain in rift valley fever virus l protein," *PLoS pathogens*, vol. 15, no. 5, p. e1007829, 2019.

[70] M. Guo and Y. Mei, "Equilibrium and folding simulations of ns4b h2 in pure water and water/2, 2, 2-trifluoroethanol mixed solvent: examination of solvation models," *Journal of molecular modeling*, vol. 19, no. 9, pp. 3931–3939, 2013.

[71] A. Hazel, C. Chipot, and J. C. Gumbart, "Thermodynamics of deca-alanine folding in water," *Journal of chemical theory and computation*, vol. 10, no. 7, pp. 2836–2844, 2014.

[72] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, "Free energy calculation from steered molecular dynamics simulations using jarzynski's equality," *The Journal of chemical physics*, vol. 119, no. 6, pp. 3559–3566, 2003.

[73] R. B. Nellas, Q. R. Johnson, and T. Shen, "Solvent-induced α-to 310-helix transition of an amphiphilic peptide," *Biochemistry*, vol. 52, no. 40, pp. 7137–7144, 2013.

[74] M. Perutz, "Electrostatic effects in proteins," *Science*, vol. 201, no. 4362, pp. 1187–1191, 1978.

[75] B. Garcia-Moreno, "Adaptations of proteins to cellular and subcellular ph," *Journal of biology*, vol. 8, no. 11, p. 98, 2009.

[76] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[77] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building water models: a different approach," *The journal of physical chemistry letters*, vol. 5, no. 21, pp. 3863–3871, 2014.

[78] Y. Wu, H. L. Tepper, and G. A. Voth, "Flexible simple point-charge water model with improved liquid-state properties," *The Journal of chemical physics*, vol. 124, no. 2, p. 024503, 2006.

[79] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of chemical physics*, vol. 126, no. 1, p. 014101, 2007.

[80] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," *The Journal of chemical physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[81] V. Tsui and D. A. Case, "Theory and applications of the generalized born solvation model in macromolecular simulations," *Biopolymers: Original Research on Biomolecules*, vol. 56, no. 4, pp. 275–291, 2000.

[82] S. Y. Mashayak and D. E. Tanner, "Comparing solvent models for molecular dynamics of protein," *University of Illinois at Urbana-Champaign, Champaign, IL.*, 2011.

[83] J. Weiser, P. S. Shenkin, and W. C. Still, "Approximate atomic surfaces from linear combinations of pairwise overlaps (lcpo)," *Journal of Computational Chemistry*, vol. 20, no. 2, pp. 217–230, 1999.

[84] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *The Journal of Chemical Physics*, vol. 112, no. 20, pp. 8910–8922, 2000.

[85] J. L. Abascal and C. Vega, "A general purpose model for the condensed phases of water: Tip4p/2005," *The Journal of chemical physics*, vol. 123, no. 23, p. 234505, 2005.

[86] R. Mills, "Self-diffusion in normal and heavy water in the range 1-45. deg." *The Journal of Physical Chemistry*, vol. 77, no. 5, pp. 685–688, 1973.

[87] W. S. Price, H. Ide, and Y. Arata, "Self-diffusion of supercooled water to 238 k using pgse nmr diffusion measurements," *The Journal of Physical Chemistry A*, vol. 103, no. 4, pp. 448–450, 1999.

[88] W. L. Bragg, J. C. Kendrew, and M. F. Perutz, "Polypeptide chain configurations in crystalline proteins," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 203, no. 1074, pp. 321–357, 1950.

[89] T. E. Creighton, *Proteins: structures and molecular properties*. Macmillan, 1993.

[90] E. Baker and R. Hubbard, "Hydrogen bonding in globular proteins," *Progress in biophysics and molecular biology*, vol. 44, no. 2, pp. 97–179, 1984.

[91] W. Burchard, M. Schmidt, and W. Stockmayer, "Information on polydispersity and branching from combined quasi-elastic and intergrated scattering," *Macromolecules*, vol. 13, no. 5, pp. 1265–1272, 1980.

[92] D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith, "Hydrodynamic radii of native and denatured proteins measured by pulse field gradient nmr techniques," *Biochemistry*, vol. 38, no. 50, pp. 16 424–16 431, 1999.

[93] D. Lucena, M. Mauri, F. Schmidt, B. Eckhardt, and P. L. Graumann, "Microdomain formation is a general property of bacterial membrane proteins and induces heterogeneity of diffusion patterns," *BMC biology*, vol. 16, no. 1, pp. 1–17, 2018.

[94] K. Gekko and S. N. Timasheff, "Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures," *Biochemistry*, vol. 20, no. 16, pp. 4667–4676, 1981.

[95] R. Sinibaldi, M. G. Ortore, F. Spinozzi, F. Carsughi, H. Frielinghaus, S. Cinelli, G. Onori, and P. Mariani, "Preferential hydration of lysozyme in water/glycerol mixtures: a small-angle neutron scattering study," *The Journal of Chemical Physics*, vol. 126, no. 23, p. 06B609, 2007.

[96] H. Betting, M. Häckel, H.-J. Hinz, and M. Stockhausen, "Spectroscopic evidence for the preferential hydration of rnase a in glycerol–water mixtures: dielectric relaxation studies," *Physical Chemistry Chemical Physics*, vol. 3, no. 9, pp. 1688–1692, 2001.

[97] E. J. Loveridge, L.-H. Tey, and R. K. Allemann, "Solvent effects on catalysis by escherichia coli dihydrofolate reductase," *Journal of the American Chemical Society*, vol. 132, no. 3, pp. 1137–1143, 2010.

[98] D. R. Canchi and A. E. García, "Cosolvent effects on protein stability," *Annual review of physical chemistry*, vol. 64, pp. 273–293, 2013.

[99] N. Chéron, M. Naepels, E. Pluharova, and D. Laage, "Protein preferential solvation in water: glycerol mixtures," *The Journal of Physical Chemistry B*, vol. 124, no. 8, pp. 1424–1437, 2020.

[100] V. Vagenende and B. L. Trout, "Quantitative characterization of local protein solvation to predict solvent effects on protein structure," *Biophysical journal*, vol. 103, no. 6, pp. 1354–1362, 2012.

[101] R. V. Rariy and A. M. Klibanov, "Correct protein folding in glycerol," *Proceedings of the National Academy of Sciences*, vol. 94, no. 25, pp. 13 520–13 523, 1997.

[102] M. Farnum and C. Zukoski, "Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor," *Biophysical journal*, vol. 76, no. 5, pp. 2716–2726, 1999.

[103] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model," *The Journal of Physical Chemistry*, vol. 97, no. 40, pp. 10 269–10 280, 1993.

[104] F. O. Akinkunmi, D. A. Jahn, and N. Giovambattista, "Effects of temperature on the thermodynamic and dynamical properties of glycerol–water mixtures: A computer simulation study of three different force fields," *The Journal of Physical Chemistry B*, vol. 119, no. 20, pp. 6250–6261, 2015.

[105] F. Mehrnejad, H. Naderi-Manesh, and B. Ranjbar, "The structural properties of magainin in water, tfe/water, and aqueous urea solutions: molecular dynamics simulations," *Proteins: Structure, Function, and Bioinformatics*, vol. 67, no. 4, pp. 931–940, 2007.

[106] S. K. Kandasamy and R. G. Larson, "Binding and insertion of α-helical anti-microbial peptides in popc bilayers studied by molecular dynamics simulations," *Chemistry and physics of lipids*, vol. 132, no. 1, pp. 113–132, 2004.

[107] F. Mehrnejad, M. M. Ghahremanpour, M. Khadem-Maaref, and F. Doustdar, "Effects of osmolytes on the helical conformation of model peptide: molecular dynamics simulation," *The Journal of Chemical Physics*, vol. 134, no. 3, p. 01B620, 2011.

[108] W.-Y. Choy, F. A. Mulder, K. A. Crowhurst, D. Muhandiram, I. S. Millett, S. Doniach, J. D. Forman-Kay, and L. E. Kay, "Distribution of molecular size within an unfolded state ensemble using small-angle x-ray scattering and pulse field gradient nmr techniques," *Journal of molecular biology*, vol. 316, no. 1, pp. 101–112, 2002.

[109] Y. Oono and M. Kohmoto, "Renormalization group theory of transport properties of polymer solutions. i. dilute solutions," *The Journal of Chemical Physics*, vol. 78, no. 1, pp. 520–528, 1983.

[110] K. Lindorff-Larsen, S. Kristjansdottir, K. Teilum, W. Fieber, C. M. Dobson, F. M. Poulsen, and M. Vendruscolo, "Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein," *Journal of the American Chemical Society*, vol. 126, no. 10, pp. 3291–3299, 2004.

[111] M. Nygaard, B. B. Kragelund, E. Papaleo, and K. Lindorff-Larsen, "An efficient method for estimating the hydrodynamic radius of disordered protein conformations," *Biophysical journal*, vol. 113, no. 3, pp. 550–557, 2017.

[112] W. L. Ash, M. R. Zlomislic, E. O. Oloo, and D. P. Tieleman, "Computer simulations of membrane proteins," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1666, no. 1-2, pp. 158–189, 2004.

[113] R. M. Cormack, "A review of classification," *Journal of the Royal Statistical Society: Series A (General)*, vol. 134, no. 3, pp. 321–353, 1971.

[114] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[115] M. E. Karpen, D. J. Tobias, and C. L. Brooks III, "Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of ypgdv," *Biochemistry*, vol. 32, no. 2, pp. 412–420, 1993.

[116] P. S. Shenkin and D. Q. McDonald, "Cluster analysis of molecular conformations," *Journal of computational chemistry*, vol. 15, no. 8, pp. 899–916, 1994.

[117] A. E. Torda and W. F. van Gunsteren, "Algorithms for clustering molecular dynamics configurations," *Journal of computational chemistry*, vol. 15, no. 12, pp. 1331–1340, 1994.

[118] J. Gabarro-Arpa and R. Revilla, "Clustering of a molecular dynamics trajectory with a hamming distance," *Computers & Chemistry*, vol. 24, no. 6, pp. 693–698, 2000.

[119] M. C. Lee, J. Deng, J. M. Briggs, and Y. Duan, "Large-scale conformational dynamics of the hiv-1 integrase core domain and its catalytic loop mutants," *Biophysical journal*, vol. 88, no. 5, pp. 3133–3146, 2005.

[120] S. K. Singh, K. K. Bejagam, Y. An, and S. A. Deshmukh, "Machine-learning based stacked ensemble model for accurate analysis of molecular dynamics simulations," *The Journal of Physical Chemistry A*, vol. 123, no. 24, pp. 5190–5198, 2019.

[121] T. W. Whitfield, D. A. Ragland, K. B. Zeldovich, and C. A. Schiffer, "Characterizing protein–ligand binding using atomistic simulation and machine learning: application to drug resistance in hiv-1 protease," *Journal of Chemical Theory and Computation*, vol. 16, no. 2, pp. 1284–1299, 2019.

[122] M. Poncin, B. Hartmann, and R. Lavery, "Conformational sub-states in b-dna," *Journal of molecular biology*, vol. 226, no. 3, pp. 775–794, 1992.

[123] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang *et al.*, "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models," *Accounts of chemical research*, vol. 33, no. 12, pp. 889–897, 2000.

Curriculum Vitae

Gideon K. Gogovi graduated with a B.Sc. degree in Mathematics in 2010 from the University of Cape Coast, Cape Coast, Ghana. This was followed by an MPhil. degree in Applied Mathematics in 2013 from Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. From there, he moved to West Virginia University in Morgantown WV, where he obtained his second Maters degree in Mathematics in 2016. He continued at George Mason University in spring 2017 for PhD studies in Computational Sciences & Informatics and completed in fall 2020.

List of Publications

1. **Gogovi, G.K**., Silayi, S., and Shehu, A., 2020. Computating the Structural Dynamics of RVFV L protein domain in Aqueous Glycerol Solutions. *In preparation*

2. **Gogovi, G.K**., 2020. Structural Exploration of RVFV L protein Domain in Implicit and Explicit Solvents by Molecular Dynamics. *Accepted*

3. Chellathurai, M., **Gogovi, G.K.** and Papaconstantopoulos, D.A., 2020. Electronic Structure and Tight-Binding Molecular Dynamics Simulations for Calcium and Strontium. *Materialia*, p.100915.

4. Hopkins, S.D., **Gogovi, G.K.**, Weisel, E., Handler, R.A. and Blaisten-Barojas, E., 2020. Polyacrylamide in glycerol solutions from an atomistic perspective of the energetics, structure, and dynamics. *AIP Advances*, 10(8), p.085011.

5. **Gogovi, G.K.**, Almsned, F., Bracci, N., Kehn-Hall, K., Shehu, A. and Blaisten-Barojas, E., 2019. Modeling the Tertiary Structure of the Rift Valley Fever Virus L Protein. *Molecules,* 24(9), p.1768.

6. Almsned, F., **Gogovi, G.K.**, Bracci, N., Kehn-Hall, K., Blaisten-Barojas, E. and Shehu, A., 2018, August. Modeling the Tertiary Structure of a Multi-domain Protein: Structure Prediction of Multi-domain Proteins. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 615-620). ACM.