5-2021

Molecular dynamic simulation of the complex folding patterns of Apolipoprotein A1 in various concentrations of potassium chloride

Hannah Holmberg
University of Tennessee at Chattanooga, djc328@mocs.utc.edu

Molecular Dynamic Simulation of the Complex Folding Patterns of Apolipoprotein A1 in Various Concentrations of Potassium Chloride

Hannah Lindsey Holmberg

Departmental Honors Thesis
The University of Tennessee at Chattanooga
Chemistry and Physics Department

Examination Date: 04/15/21

Dr. Luis Sanchez-Diaz
Lecturer of Physics
Thesis Director

Dr. Joshua Hamblen
Associate Professor of Physics
Department Examiner

Abstract:

Apopliprotein or ApoA-1 is a complex lipoprotein that functions in the removal of cholesterol from the blood, removing cholesterol from the area around white blood cells and promoting the excretion of lipids through the lymphatic system. Previous research has found that ApoA-1 shows both folded and unfolded conformations depending on the concentration of NaCl in solution in the water around it. The protein was studied using molecular dynamics simulations. Once this state of equilibrium was reached, various structural properties of the protein were measured including the radius of gyration and the radial distribution function. The goal of the project was to determine if ApoA-1 had stable folding conformations in various concentrations of Potassium Chloride and how those concentrations compared with the concentration of Sodium Chloride around known stable conformations. Previous research has determined that in a range of concentrations from 0.5M to 2.0M ApoA-1 has both folded and unfolded conformations. We are now working to study Potassium Chloride in these concentrations to determine if the radius of gyration results will be the same as the ones found in studying the sodium chloride. This research will determine if the healthier salt potassium chloride, allows for the same folded and stable conformations of ApoA-1.

Table of Contents

Chapter 1:

Introduction

1.1 Background:

Molecular dynamics (MD) simulation is a method of computational research that has proven key to the study and visualization of protein folding and stability. This method answers many difficult biophysical questions that would be quite difficult to understand with other research methods due to the complexity of proteins and the rapid changes that occur in their structure from solvent molecule integrations and the thermal motion of the protein. Molecular Dynamics looks to overcome these problems by assuming that the majority of fine details of protein structure can be generalized into an average and chain folding can be observed and predicted based upon the variables that move the slowest but cause the greatest effect in a conformation. [1] Protein stability and activity is affected by several different factors including: hydrophobicity, polarity and hydrogen-bond capacity of ILs. These conditions can be used to determine the most favorable conditions for a stable tertiary structure of a particular protein. [2] These factors are important to the study of proteins including apoA-1 because of its multiple repeating but hydrophilic and hydrophobic residues. After each step in an MD simulation data analysis can be done to quantify the conformational and thermal changes in the protein- solvent system. [3]

1.2: Apolipoprotein A1:

Apolipoprotein A1 (ApoA-1) is a complex lipoprotein that is found in the human bloodstream that functions in the excretion of lipids through the liver. [4] ApoA-1 mainly functions in the removal of cholesterol from the arteries. ApoA-1 is a major protein in the much larger structures known as high density lipoproteins (HDLs), which are now known as “good” cholesterol. [5] ApoA-1 makes up approximately seventy percent of the structure of these HDLs. HDLs are known to contain many repeating sections of ApoA-1 that are folded into many alpha

helixes that further confine inner lipids. [6]. It is important in the prevention of many chronic diseases including atherosclerosis. The effects of ApoA- 1 are still not fully understood but it is believed that it might be a ligand for an HDL receptor in reverse cholesterol transport and cholesterol efflux. [7] This portion of the protein ApoA-1, 1GW3, contains a chain of 243 amino acid residues. These repeat in a sequence that allows the formation of hydrophilic and hydrophobic residues along the structure. These residues are believed to be used in conjunction with other factors like ionic strength to increase the bind affinity of ApoA-1 to lipids and allowing for a stronger hold to be kept on the lipid which helps remove the cholesterol from the blood. This protein also helps to remove cholesterol from the white blood cells in artery walls; it is essential for the protein to be properly folded for this process to be performed in its highest capacity. [1]

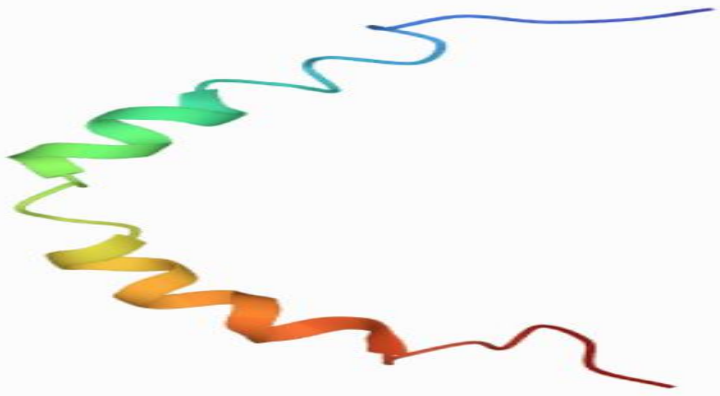

Figure 1.2.1: The native structure of apoA-1 from Protein Data Bank

1.3 Molecular dynamic simulations

Molecular dynamics simulations are more common in current research than they have ever been before. They allow research to be conducted much more efficiently and at a minimal cost in comparison to traditional research methods. [8] The MD software used in this research is known as GROMACS. It is used to research biophysical and biochemical questions in a computational capacity. Advancements in computational biophysics allowed for this research to be conducted using a portion of ApoA-1 that was downloaded from the Protein Data Bank. The file was then able to be viewed on the research lab computer using visual molecular dynamics (VMD) software. This software allows for the native conformation of the protein to be viewed while also allowing for files to be inputted to allow for the visualization of conformational changes. The protein can be rotated around all four of its different axes to allow for a three-hundred-and-sixty-degree view of the protein. The file from the Protein Data Bank has all of the amino acids in the native conformation of the protein. It also includes the amino acid properties including mass and polarity which is then interpreted by the GROMACS simulation software and used to allow for the movement of the amino acids in a natural environment to be recreated by the system. To begin the simulation process, the researcher puts the protein in a “box” of a specific size and containing a chosen number of solvent molecules. For the uses in this experiment the solvent was water. An advantage of this type of software is that the researcher has total control over the system. The researcher is free to set the heat, molarity, pressure and volume of the system and to then minimize or change these parameters at any time. The parameters are much easier to control in simulation as opposed to in noncomputational research. This control allows for parameters to be held constant over a long period of time or to be rapidly changed within an experiment.

1.4 Protein Folding

Early protein knowledge led to the understanding that proteins are macromolecules that are composed of peptide bond linked amino acids that have a charge. More recent studies and work across biochemistry and protein physics have allowed for a deeper understanding of the atomic level of protein structure, dynamics and folding. Furthermore, these studies have allowed for a more in-depth understanding of the dynamics behind protein folding. The sequence of amino acids in the polypeptide chain is known as the primary structure of a protein. This sequence then folds into a secondary structure that is composed of elements know as α- helix and β- sheets. The interactions between these elements then leads to the formation of the tertiary structure of the protein. A large component of the structure-function paradigm of structural biology is the relationship between a proteins tertiary structure and its function. The protein's structure allows for it to be more functional in its chosen role. A misfolded protein does not allow for proper functioning of the protein. In addition, the maintenance of tertiary structure of the protein is a result of the contribution of many physical interactions that work additively to hold the protein in its three-dimensional conformation. [9]

1.5 Previous Studies

Previous research has shown that along various concentrations of sodium chloride the protein will form either a folded more stable structure or it will form an unfolded less stable structure. [10] Through study with Dr. Luis Sanchez- Diaz I recreated this research using GROMACS molecular dynamics simulation software. I was able to run the simulation as

explained in the previous research and calculate the Radius of Gyration and Root Mean Squared Distribution of ApoA-1 over 50ns. [10] I was also able to visualize the final conformation state of the using VMD software. This software allowed me to view the final state of the protein to see whether it was folded or unfolded. Through this research we were able to see that at concentrations of 0.5 M and 1.8M the protein formed an unfolded pattern and had a high average value Radius of Gyration across the protein. In addition, we were able to see that when placed in a sodium chloride concentration of 1.7M and 2.0M the protein formed a folded conformation and had a low value for average Radius of Gyration across the protein.

This research furthers previous studies of ApoA-1 using molecular dynamics simulations in Groningen Machine for Chemical Simulations (GROMACS). First, we ran the protein through the same process as used in the previous research but, instead of using sodium chloride for the salt we used potassium chloride. This allowed us to verify if the potassium chloride will permit for the protein to fold into the same conformations at the same concentrations as it did in the previous research. We then ran other concentrations of the potassium chloride to find an ideal level for the most stable and favorable conformation of the ApoA-1 protein. Through this we established information on whether potassium chloride can be used in substitution for sodium chloride in people facing both high cholesterol and hypertension.

Previous research in both simulation and experimental forms has shown that ApoA-1 has several stable folded conformations. It has also proven that the 22- and 11- amino acid repeats in the flexible structure of Apolipoprotein contribute substantially to the functions of not only ApoA-1 but of HDL as a whole. These conformations allow for this protein to play a key role in reverse cholesterol pathways [7]. Through X-ray crystallographic analysis of a section of ApoA-1 that was crystalized in a solution of high salinity it was determined that under these conditions

the protein showed a helical structure that had any loops and overall had a more closed appearance similar to that of a U shape [3]. Computer simulations have become more widely used in recent years due to the ability to control the exact conditions of the space that the protein is modeled in. Additionally, the simulation allows for an exact solution to the mathematical equations needed to represent the protein's conformation and allows for the visualization of the protein's conformation changes over time. Research previously conducted by M.A. Balderas Altamirano, A. Gama Goicochea and E. Pérez, showed that ApoA-1 has both folded and unfolded conformations depending upon the concentration of the sodium chloride solution in which it is placed. The researchers used GROMACS computer simulation software to conduct their study. They studied how the complex ApoA-1 folded as the concentration of the salt was increased and how it contracted or expanded after the increase in concentration. It was found that the complex ApoA-1, which contains, positive, negative and neutral sequences along with both hydrophobic and hydrophilic residues, has many different conformations and overall sizes throughout the changes to its physiological environment. They used the section of ApoA-1 known as 1GW3 from the Protein Data Bank to do the simulations on. This section has 142-187 amino acid residues which allows for it to more easily computed than if the whole protein was used. [10]

1.6 Objectives:

Our work will serve to further the previous studies on the folding of ApoA-1 in different concentrations of ionic solution. We will focus on determining how the conformation of ApoA-1 changes with the change in concentration of Potassium Chloride in solution and how the simulation time effects the conclusiveness of the data collected. We place the protein in solution and determine how KCl affect the protein's folding over time. Solutions of concentration

0.5,0.6,1.0, 1.7 and 1.8 of KCl will be run through the simulations over times ranging from 1ns – 25ns. From this work, we hope to identify which concentrations of KCl allow for a stable folded or unfolded state of ApoA-1. We also hope to identify the most length of simulation time needed to collect the ideal amount of data to have conclusive results.

Chapter 2:

Materials & Methods

2.1 Apolipoprotein A1 Structure:

The sequence of amino acids and the regions of the secondary structures for the section of ApoA-1 known as 1GW3 are shown below. [11] The protein's beginning conformation is caused by the helix-hinge-helix structural motif in human ApoA-1. The 243 acid residues in ApoA-1 are organized into repeating units of 11-22 residues. These repeating sequences are proposed to form amphipathic helices, that is helices that have both hydrophobic and hydrophilic units or faces. Research has shown that the strong lipid binding of the C-terminus of ApoA-1 is caused by the large number of aromatic residues and hydrophobic residue pairs in the sequence. These allow for a strong anchoring of apolipoproteins to lipid [12].

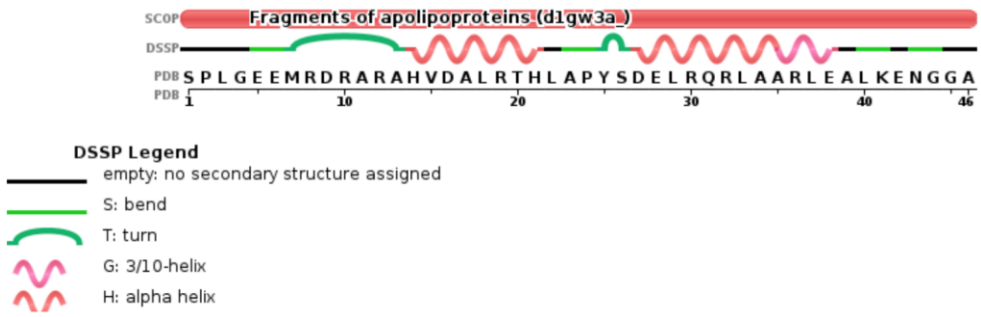

Figure 2.1.1: Amino acid sequence and secondary structure of apoA-1

2.2 GROMACS:

The Groningen Machine for Chemical Simulations, GROMACS, is a free computer simulation package for the study of molecular dynamics. Molecular dynamics simulations find the solution of the equations of motion, using numerical, step-by-step calculations [13]. In calculating this, the software solves the Newtonian equations of motion for the number of atoms, N, and finds that the force is the inverse derivative of the energy [10].

Equation 2.2.1:

$$F = ma = m\left(\frac{dv}{dt}\right) = m\left(\frac{d^2r}{dt^2}\right)$$

Equation 2.2.2:

$$F = -\nabla E$$

These equations would be virtually impossible to solve in a reasonable time setting without the assistance of simulation data analysis. The trajectories are given by positions as a function of time, r(t). Calculations of interactions are limited to help reduce computing time and to allow for a more generalized view of interactions. GROMACS is conducted by writing command lines into the terminal of the computer using input and output files. GROMACS is a popular package for protein simulations do to it being well developed and researched. Boundary conditions can be set to be periodic to allow for fluctuation in the size of our "box" to prevent hinderance. We can also change the equations for ionic interactions and water molecule addition in the system.

Force Field:

Simulations require a set of specific guidelines to direct the interactions inside of the simulation these are given by the force field chosen for the simulation. The force field mainly functions to apply restraints to the system and to set the potential energies between the molecules. This force field is how the simulation will calculate the energy of the interactions. We used the OPLS-AA/L all-atom force field. We choose an all atom force field to account for all of the columbic forces acting inside of the box. The OPLS-AA/L for proteins have been used in the simulation of many biological systems. This system allows for computations to be must faster. This force field has been found to perform well especially when used for quantities that are dependent upon nonbonded parameters. The nonbonded parameters in the system are optimized to reduce the liquid phase properties and fit the torsional parameters to the experimental data.

Local MP2 with a basis set that was larger than the previous version is used to evaluate single-point energies at optimized HF geometries. [14]

2.3 Solvents:

Ionic liquids (IL) are used as solvents in molecular dynamics simulations because they can get the closest comparison to the environments in the blood and can stabilize the protein. [15] It has been found that many solvents can be used in the study of proteins and that their effects can vary based upon the solvent and the concentration of anion or cation in it. [16] Our study works with a water solution with varying concentrations of Potassium Chloride. These were then compared to a previous study of the folding patterns of apoA-1 in varying concentrations of aqueous Sodium Chloride. This work aims to compare the folding of apoA-1 in Potassium Chloride to its folding in Sodium Chloride. It will also observe the stability of the protein at various concentrations of Potassium Chloride and how they compare to the stability of the protein in the same concentration of Sodium Chloride. The folding patterns and stability of apoA-1 in various molarities of Sodium Chloride has been previously studied however, we intend to compare these findings with the results from another group one salt, Potassium Chloride. This work studied the length of the protein backbone after the protein was simulated through a process inside a “box” containing various molarities of aqueous Potassium Chloride. Through this we will be able to see if the Potassium Chloride salt allows for the same folded and unfolded stable conformations of ApoA-1 as Sodium Chloride does.

2.4 Simulation Details:

To study the Potassium Chloride based ionic solutions we used GROMACS because all of the molecules were already implemented into it. This process followed a previous simulation study of lysozyme in a Sodium Chloride ionic liquid. [17] The first step in the simulation is to create the protein system. This begins by retrieving the section of ApoA-1 known as 1GWB from the protein data base as a pdb file. This file was then placed into a box with specific boundary conditions set to be periodic on each side. A pdb file is designed by the protein data bank to contain the structural information of the three-dimensional protein and the amino acid sequence of it. We then solvate the box using the SPCE water method and the genion function in GROMACS is used to add the chosen concentration of Potassium Chloride as directed by the command given. The system systematically replaces the solvent molecules with either Potassium or Chloride ions until the given concentration is reached. Therefore, the specific number of water molecules will vary based upon how many ions are added to reach the desired concentration. Once the box has been created the pdb2gmx command is used to create a topology file for the protein- solvent system that has now been created in the box. This file will contain the exact positions of all of the protein, solvent, and water molecules in our system for the duration of our simulation and will provide a set of raw data that can be later analyzed to observe the movement of these molecules as a function of time. The topology is rewritten in each step to show the movement of all of the molecules in the system. The system is then put through energy minimization where its energy is dropped to ensure the starting energy is consistent throughout each simulation. This reduces the number of steric clashes that could occur without this step. In energy minimization the temperature of the system is dropped in increments of 200K from 100K to 300K each step takes 2ns and is conducted at 100 atms. The .mdp file determines the number

of calculations that are conducted at each step and therefore dictates the overall amount of raw data that is collected. After this step a file is created that allows you to view the change in potential energy over the minimization time as an exponential decay that has an asymptote at the lowest produced energy. After this step, we conducted NVT equilibration in which volume was held constant with simulated temperatures dropping from 2000K to 300K in 200K steps that lasted 2ns each. After this step a file can be created that allows you to view the temperature over time and to allow you to confirm the temperature drops to 300K. Following this we use NPT equilibration at 300K for 6ns. In this step pressure is held constant. Following this step, a file can be created to view the total energy of our system over time, this energy graph should be relatively flat and should remain flat after the process is repeated. The graphs are not expected to be exactly identical due to the randomness of the molecules inside the system, but the energy levels of the graphs should be relatively the same. The final command of the simulation produces the data for our simulation. This command is controlled by another .mdp file and can be set to have a specific number of steps and time. For our simulation we wanted to run the simulation over a short time frame and over a long time because we wanted to find the optimal amount of time needed for conclusive data to be collected. For this reason, we ran some of the simulations for 25ns, some for 600ps, and some for 15ns with 0.002 fs steps. This process was very time consuming taking approximately a month for each simulation to be completed running on 24 cores 24 hours a day. This process was repeated for five molarities of Potassium Chloride, 0.5M, 0.6M, 1.0M, 1.7M and 1.8M. At the end of the data production each simulation has data that can be reviewed both mathematically using radius of gyration and radial mean distribution equations and visually using VMD software. VMD allows for the final structure of the system to be visualized. Additionally, the initial protein, ion file can be put into it and have the topology data

loaded in as well. This allows for a movie of the system throughout the simulation to be viewed. The details for the process are shown in Appendix 1.

2.5 Data Analysis Techniques

After MD production has been completed the simulation of the protein and IL system itself is complete. We then must begin to analyze the raw data that was collected throughout the simulation to see what changes occurred to the protein system throughout the simulation. The first way we will do this is by visualizing the proteins conformational changes using Visual Molecular Dynamics (VMD). VMD will allow us to view not only the final confirmation of the system, but also to view the systems changes throughout time. The protein-ion system file can be loaded with the topology file which will allow for us to visualize the protein over the full run time of the MD production. This process does not actually provide any data in itself, but it serves as a tool to check the results of our later data analysis [18]. To numerically analyze the data, we used the Radius of Gyration (R_g) and the Root Mean Squared Distribution (RMSD).

Radius of Gyration

Radius of Gyration (R_g) analyzes the spread of the molecules of the backbone of the protein around its center of mass (R©). This calculation begins by calculating the center of mass using the equation shown below.

Equation 2.5.1:

$$0 = \sum_{i=1}^{N} m_i(r_i - r_c)$$

This equation is looks at all of the molecules in the protein's backbone and finds the point at which their radius is zero as compared to the center of mass of the protein's backbone. [19]

Equation 2.5.2:

$$R_g = \sqrt{\sum_{i=1}^{N}\left(\frac{\mathrm{m}(r_i - r_c)^2}{\mathrm{M}}\right)}$$

In equation 2.5.2 “m” represents the total mass of the molecules in the backbone of the protein while the actual R_g is defined as the square root of this value. The radius of gyration explains how compact or widespread out the protein is over time by providing an average distance from the center of mass for each atom in the protein. A protein in its most functional form will have a low R_g value representing a more folded and compact folding pattern. A high R_g is representative of a less folded and more open conformation leading to the protein being in a less functional form. This almost linear protein does not allow for proper lipid binding and trapping making it not a functional concentration for the protein. Radius of Gyration can also be used as an indicator for the overall stability of a specific conformation of the protein in solution. R_g values that stay relatively constant are indicative of a stable conformation for the protein in solution. At lower values this constant R_g value indicates a solution which allows for a well stabilized protein. This however, cannot be used with higher values because it cannot be confirmed that the ionic interactions are the cause of the protein’s stability. The radius of gyration for our systems are directly compared to those of the systems study in the previous research using Sodium Chloride [20]. This method of study of R_g values is common place in protein study and is already a function in GROMACS. [21] GROMACS has a command to have the program create an xvg file that shows the average radius of the protein in nanometers as a function of time.

Root Mean Squared Distribution:

Root Mean Squared Distribution (RMSD) is the second method used to analyze our protein-ion system. Like R_g, RMSD is a calculation that is built into GROMACS to be used to analyze the topology from a simulation. It can be executed with a rather simple command line which produces another xvg file. RMSD is the least squared fit of the protein as compared with its initial conformation. [22]

Equation 2.5.3:

$$RMSD(t_1, t_2) = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i \parallel r_i(t_1) - r_i(t_2) \parallel^2}$$

Where $t_2 = t_1 + \Delta t$

Equation 2.5.3 finds the average change in position of all of the atoms in the system in nm over time. For our purposes, we will focus our calculation on the backbone of the protein however, it can be used for many other functions. This calculation will give us a general idea of how much change occurs in the protein over the time in the simulation. Therefore, low values of RMSD indicate little change in the protein's backbone over time while high values will indicate significant change over time. Consistent RMSD levels over long-time intervals indicates stretches of time where there is little to no change in the protein structure whereas, spikes in the plot over time indicate instability and changes in the conformation. The stretches of consistent RMSD must be compared to the visuals and R_g values at the time to discover whether the stability is a result of a stable folded conformation or a stable initial conformation that remains unchanged. These graphs must be used together in conjunction with the VMD visualizations to give the complete picture of the system interactions with our protein.

Chapter 3:

Results

3.1 Ionic Solvent Concentration:

After creating the box and inserting the ions and water into it, a topology file can be written to determine the number of solvents, potassium, and chloride molecules in the solution. Although we had a known molarity for each simulation the number of solvent molecules varied based upon how many ions were needed to reach the required molarity. The molarities for each system is shown below. It is important to note that the system added additional water molecules for some concentrations to fix spatial issues in the box.

Molarity	Solvent Molecules	K^+ Molecules	Cl^- Molecules
0.5	32581	371	371
0.6	32581	371	371
1.0	32705	309	309
1.8	31961	681	681

Figure 3.1.1: Number of Potassium, Chloride, and Water molecules for various concentrations.

3.2 Simulation Data:

Root Mean Squared Distribution was written as a function of the protein backbone and therefore excludes the side chains of the amino acid sequence. Radius of gyration was calculated from the topology file written for the entire protein. These parameters are the same for all simulations performed.

Water

This simulation of water was conducted in a previous study conducted by Benjamin Smith and Dr. Luis Sanchez- Diaz. The simulation of water serves as a control to see how the protein folds in a system that lacks any ion concentration. The data collected in their study showed small changes to the initial tertiary structure. They did not find any new secondary structures, which indicated that the protein was stable and was not being forced to fold. From a visual point of view, they found that the protein seemed to stay linear and did not show signs of the ends folding into initiate a folded configuration. [13]

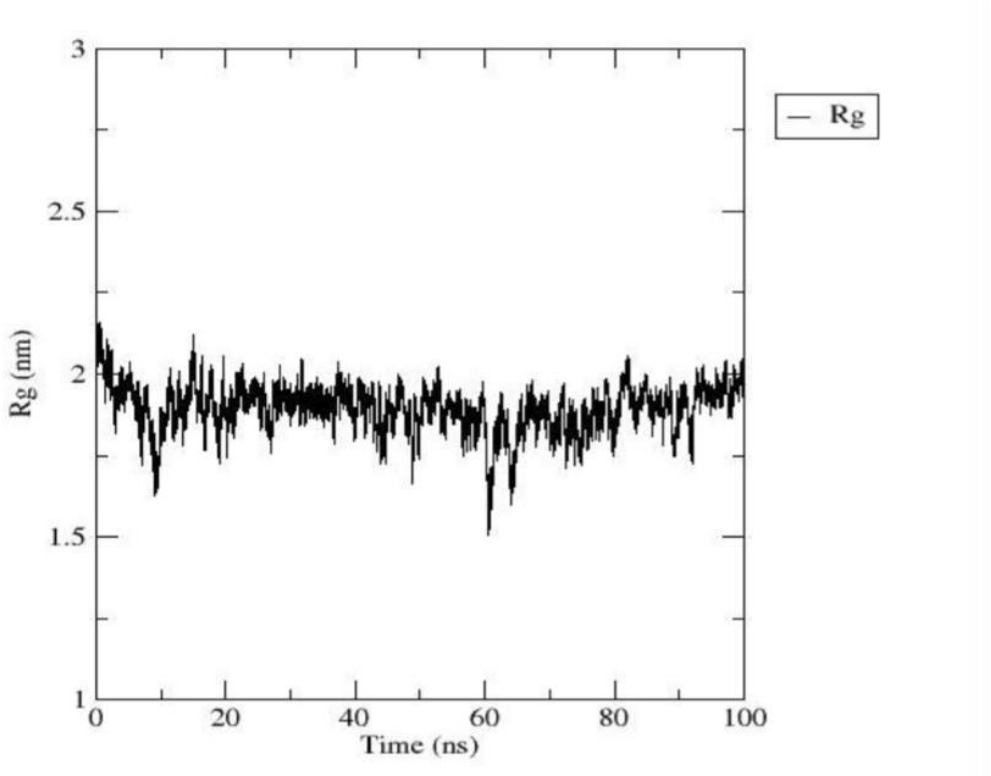

Figure 3.2.1: R_g (in nm) of ApoA-1 in water solution simulated over 100ns [13]

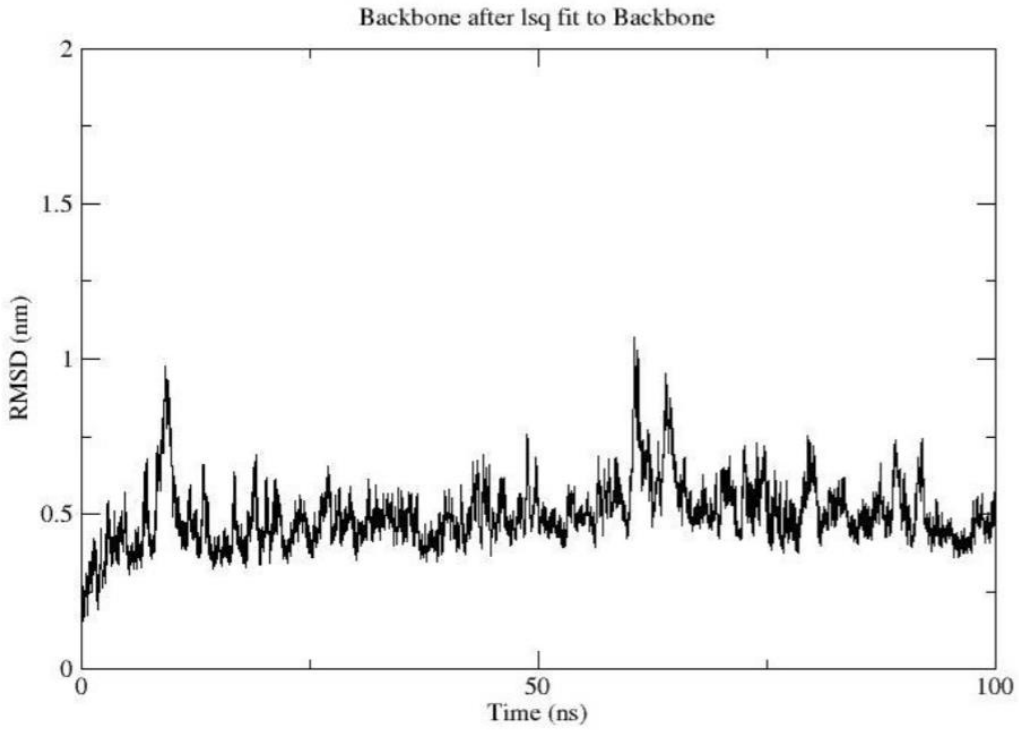

Figure 3.2.2: RMSD (in nm) of ApoA-1 in water solution simulated over 100ns [13]

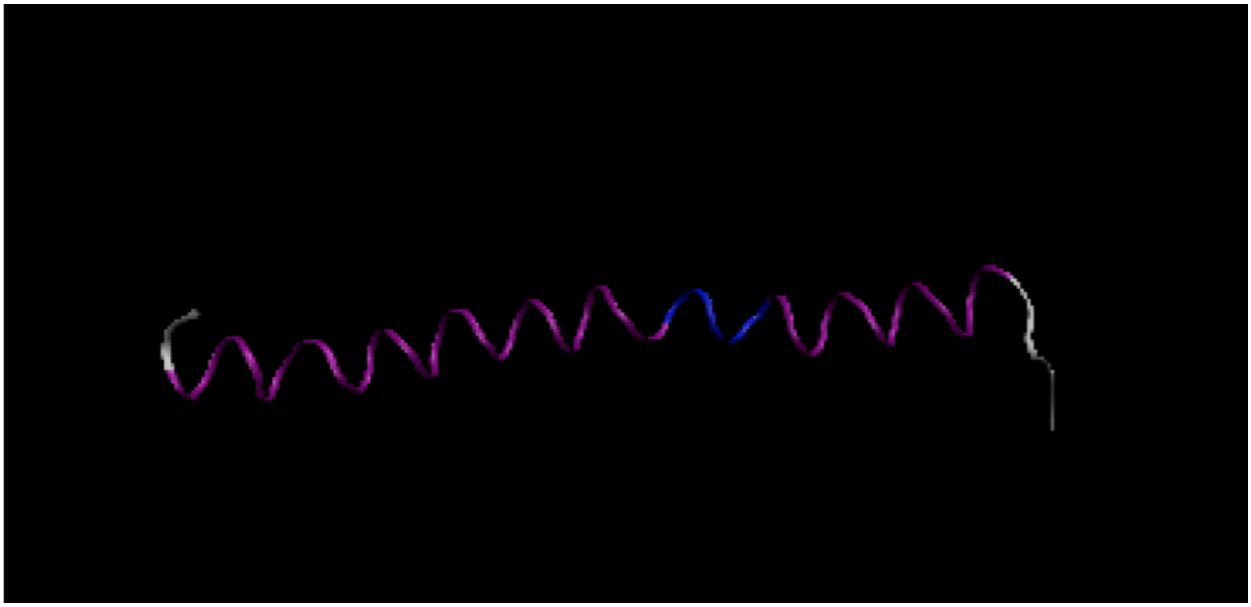

Figure 3.2.3: Final confirmation of ApoA-1 in water [13]

The data collected from the simulations conducted with Potassium Chloride are shown below. At the lowest concentration of Potassium Chloride, 0.5M both the RMSD and R_g show a relatively stable unfolded conformation of the protein when it is simulated for 600ps. However, when the run time was extended to 10000ps the data shows many more fluctuations in the conformation over time. From the longer simulation time we can see that in fact the final confirmation of ApoA-1 in a 0.5M concentration is folded this is shown visually in figure 3.2.6.

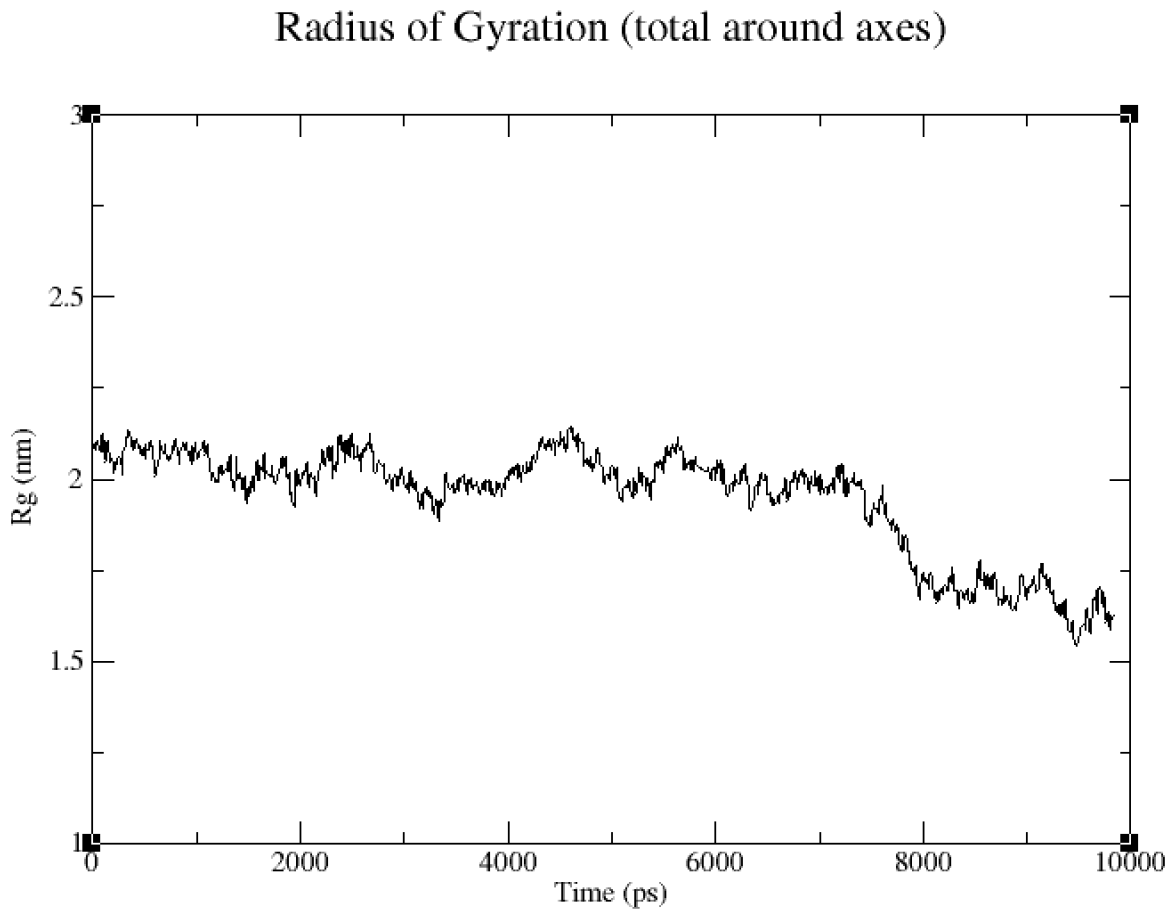

Figure 3.2.4: R_g (in nm) of ApoA-1 in a 0.5M potassium chloride ionic liquid simulated over 600ps

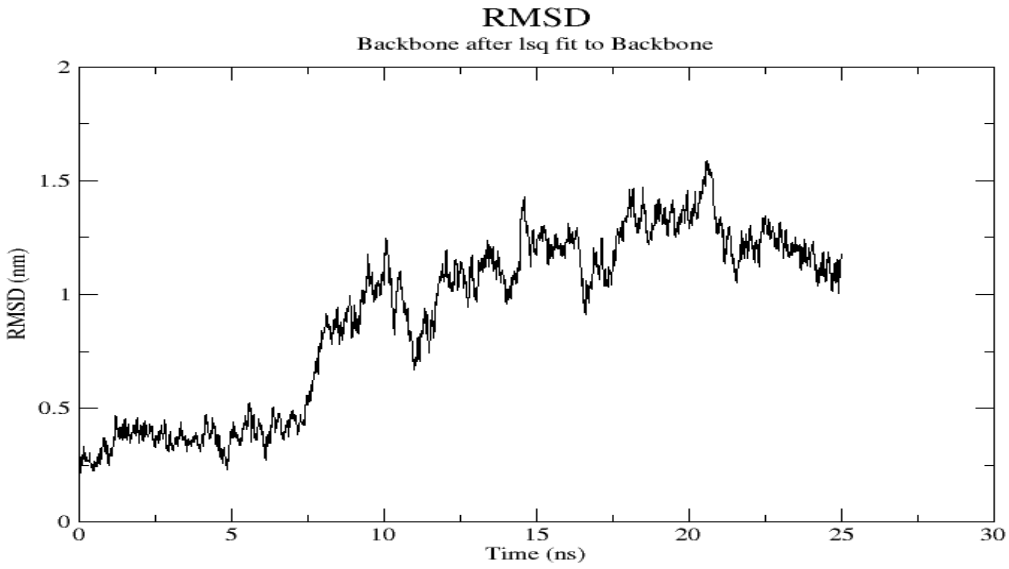

Figure 3.2.5: RMSD (in nm) 0f ApoA-1 in 0.5M potassium chloride ionic liquid simulated over 10000 ps (25 ns).

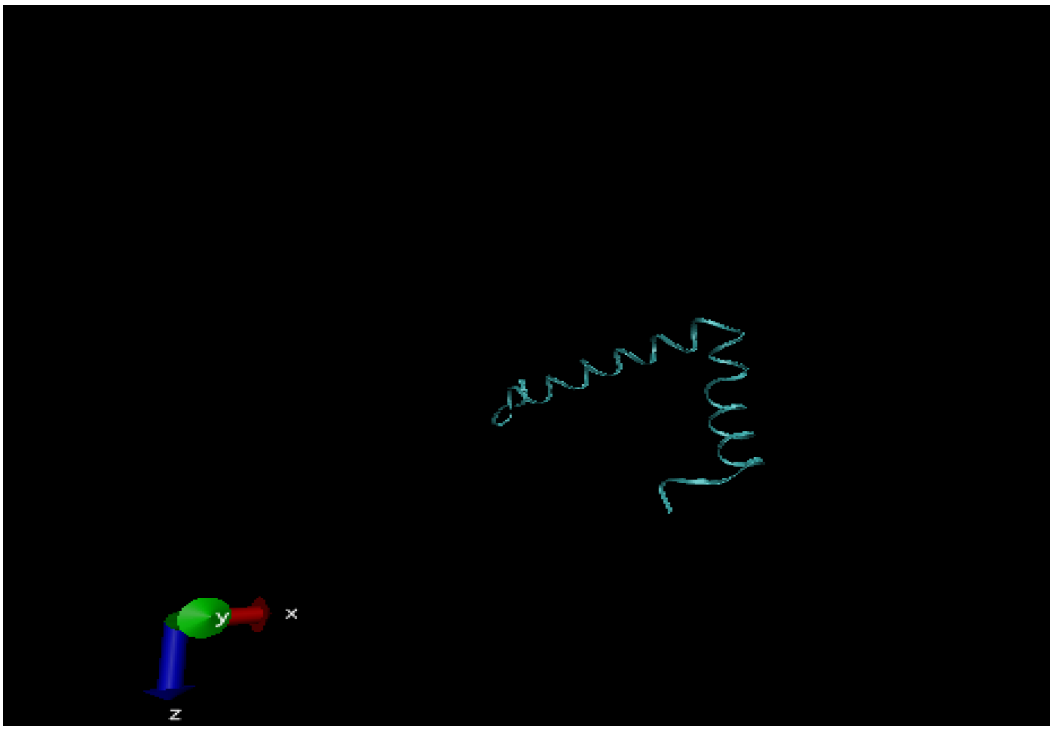

Figure 3.2.6: Final confirmation of ApoA-1 in 0.5M Potassium Chloride simulated over 10000 ps.

During the second lowest molarity, 0.6M the R_g shows that the protein begins rather linear with a R_g value of just over 2nm, but it can be seen that as the simulation continues toward the 15000 ps mark it stabilizes into a folded configuration with a final R_g in all directions between 1.3nm and 1nm. The RMSD data further shows this by starting with a low value around 0.20nm and then quickly rising and stabilizing around 1.25 nm. Anything above a value of 1nm in RMSD shows a folded conformation. This data is visualized in figure 3.2.9.

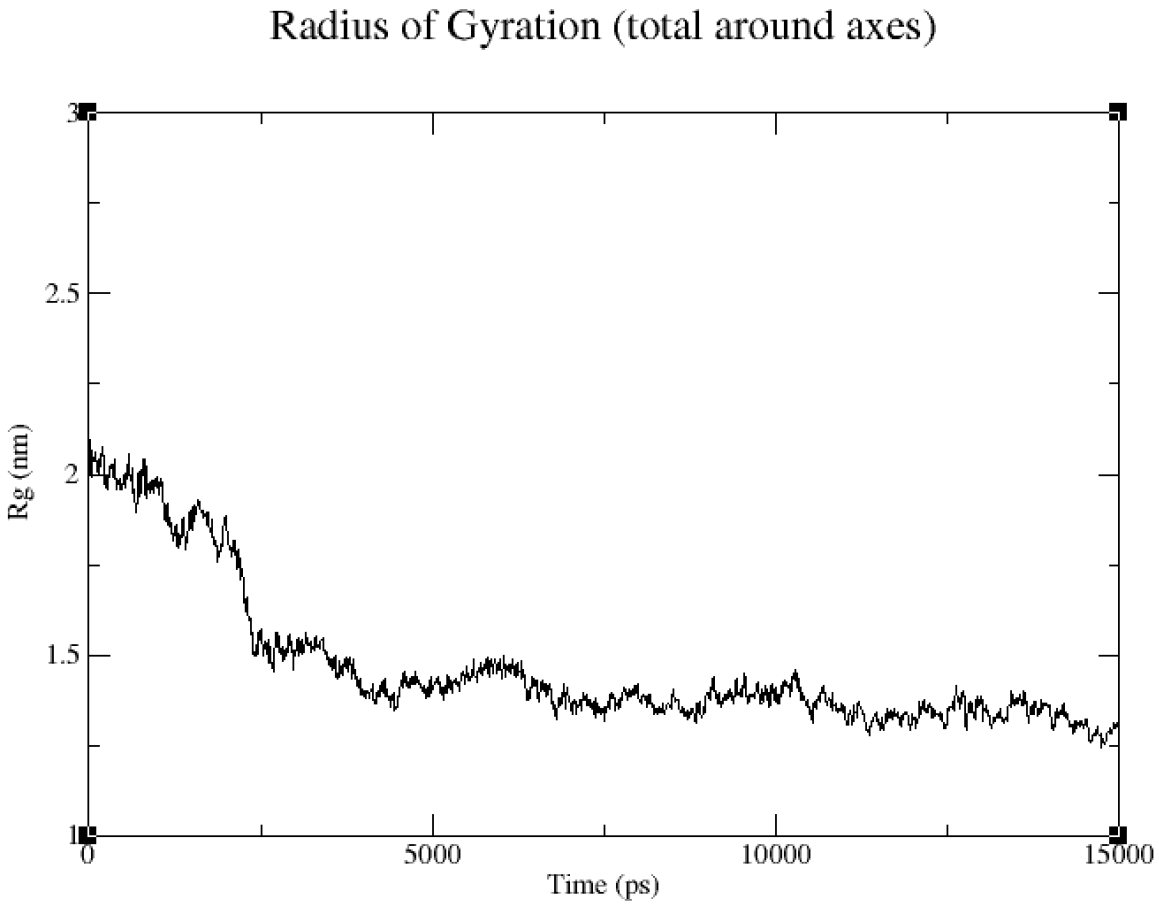

Figure 3.2.7: R_g (in nm) of ApoA-1 in 0.6M potassium chloride ionic liquid simulated over 15000 ps.

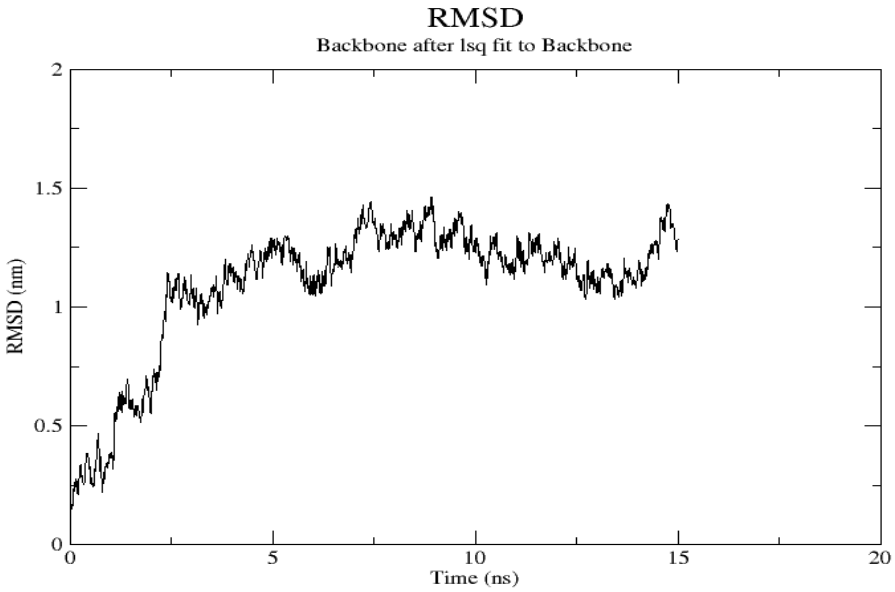

Figure 3.2.8: RMSD (in nm) 0f ApoA-1 in 0.6M potassium chloride ionic liquid simulated over 15000 ps (15 ns).

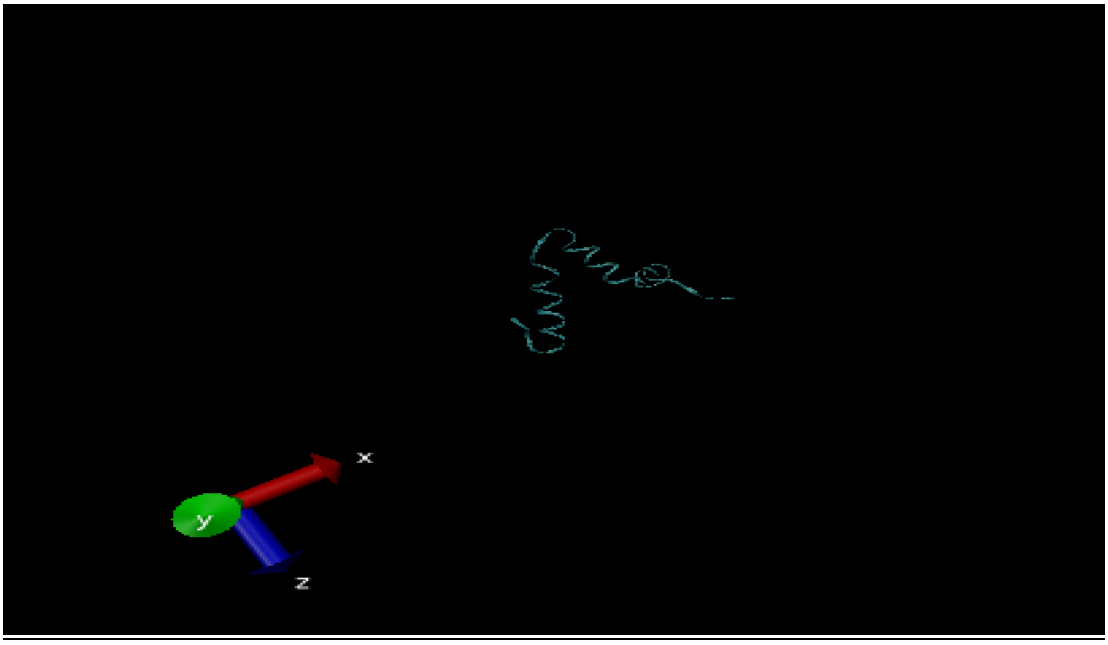

Figure 3.2.9: Final confirmation of ApoA-1 in 0.6M Potassium Chloride simulated over 15000 ps.

The third concentration studied was 1.0M. The R_g data showed a consistently unfolded conformation over time. The R_g values do not show much fluctuation at all throughout the time of the simulation. The RMSD data shows many spikes over time, but all at very low values all of which are under 0.35 nm showing a very linear protein conformation. The visualization of the protein in figure 3.2.10 further shows that at this concentration when the simulation is run for 600ns.

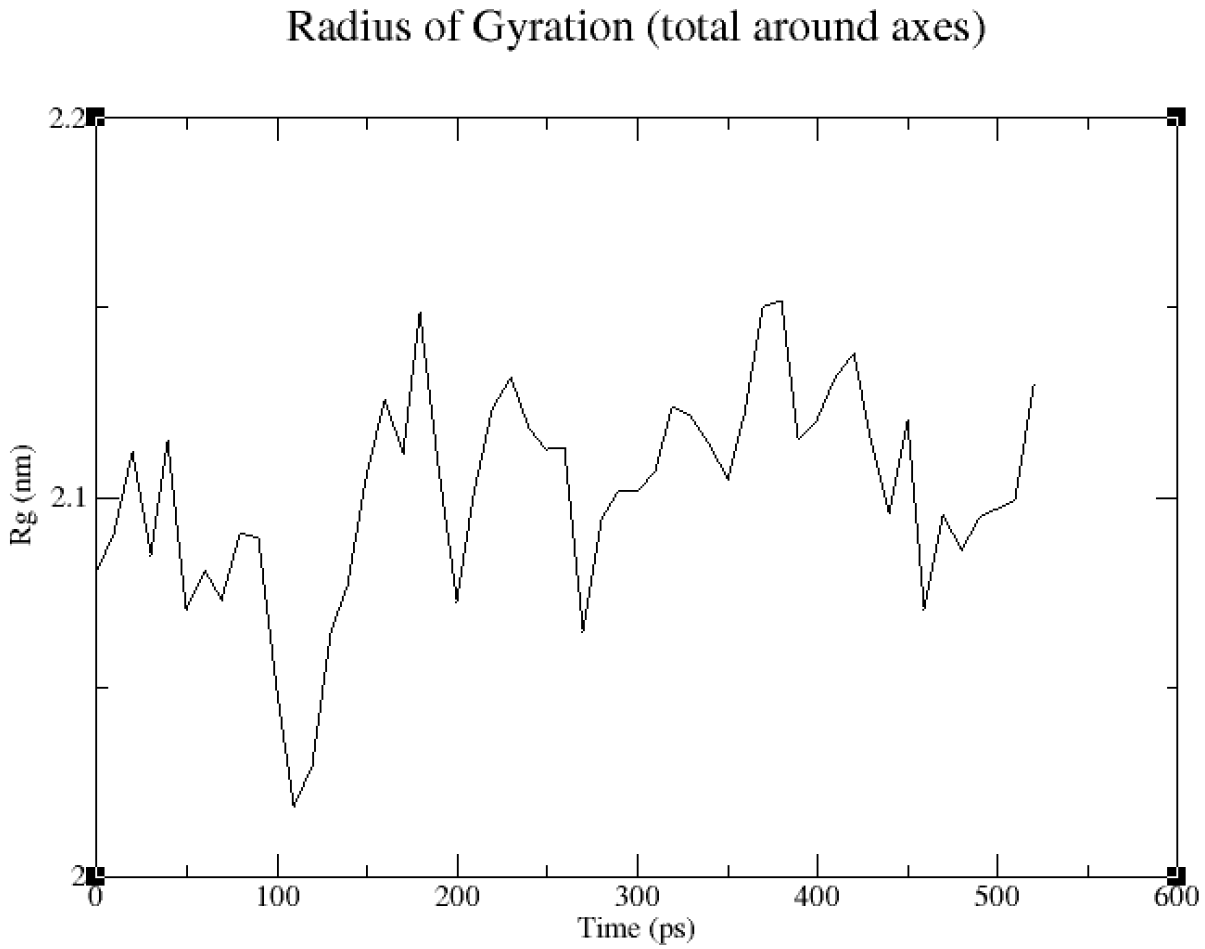

Figure 3.2.10: R_g (in nm) of ApoA-1 in 1.0M potassium chloride ionic liquid simulated over 600 ps

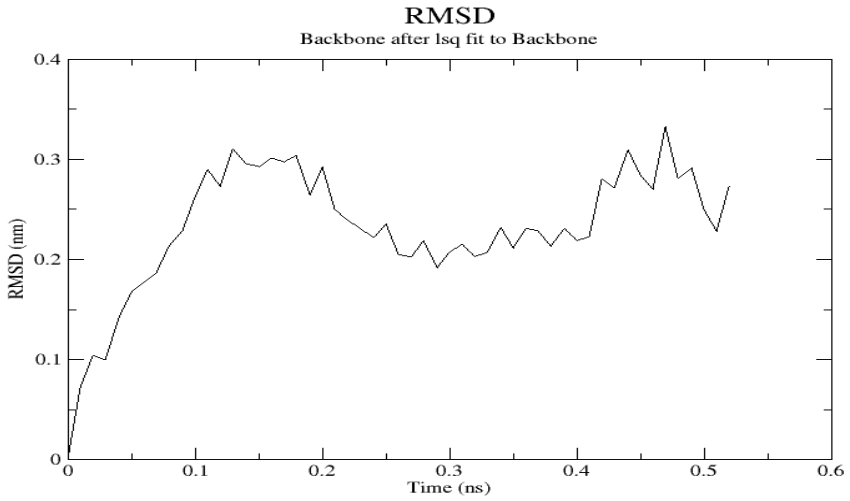

Figure 3.2.11: RMSD (in nm) 0f ApoA-1 in 1.0M potassium chloride ionic liquid simulated over 600 ps.

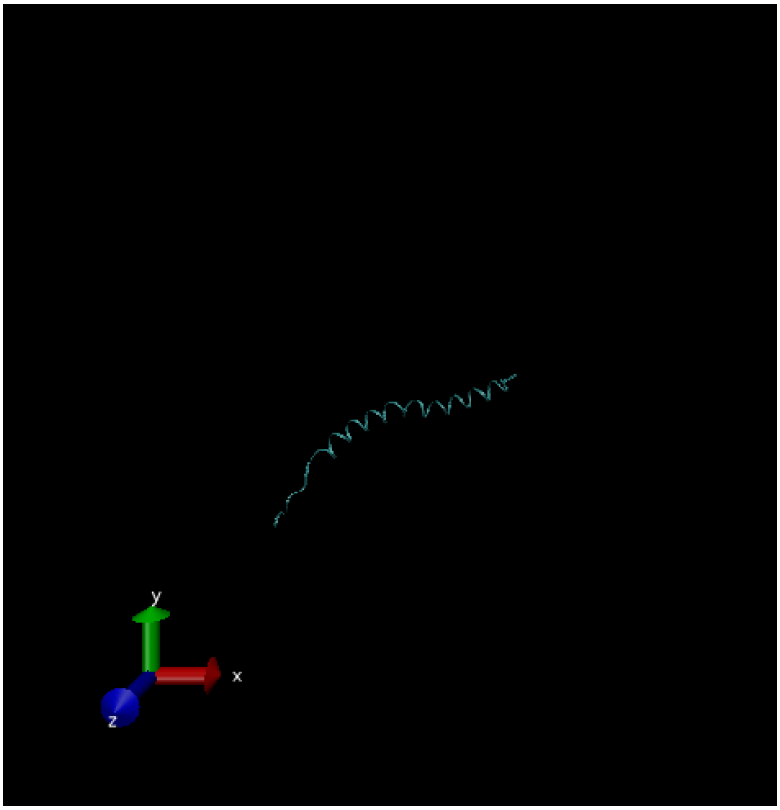

Figure 3.2.12: Final confirmation of ApoA-1 in 1.0M Potassium Chloride simulated over 600 ps.

On the upper end of the concentrations was 1.7M. The R_g value for this concentration shows a rather consistent unfolded conformation. It appears to show a slight bit of folding in on itself as the time approached 2000 ps. Figure 3.2.13 visualizes the final confirmation of ApoA-1 in a 1.7M concentration it shows a mostly linear protein with a slight bend in the middle of the protein.

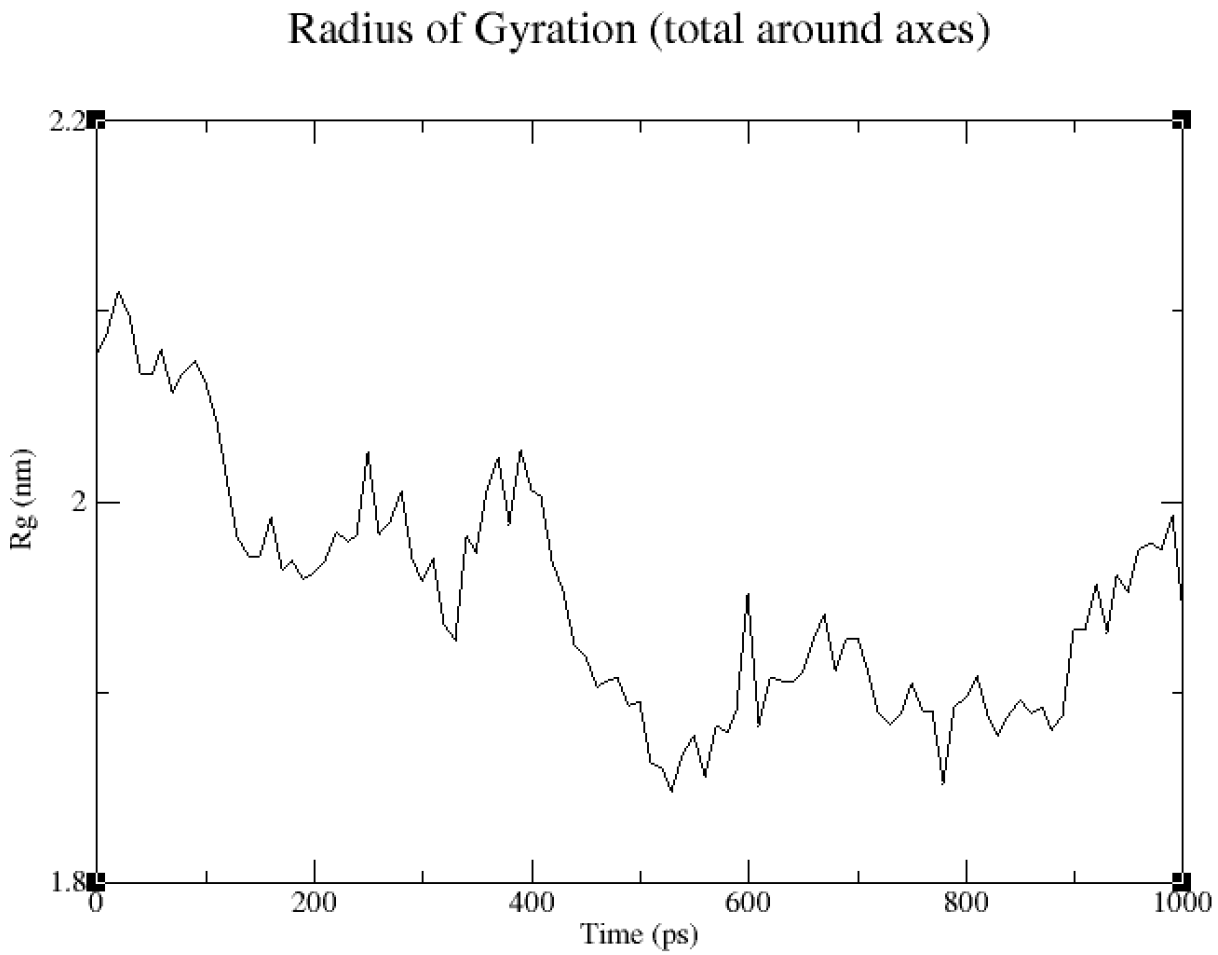

Figure 3.2.13: R_g (in nm) of ApoA-1 in 1.7M potassium chloride ionic liquid simulated over 2000 ps

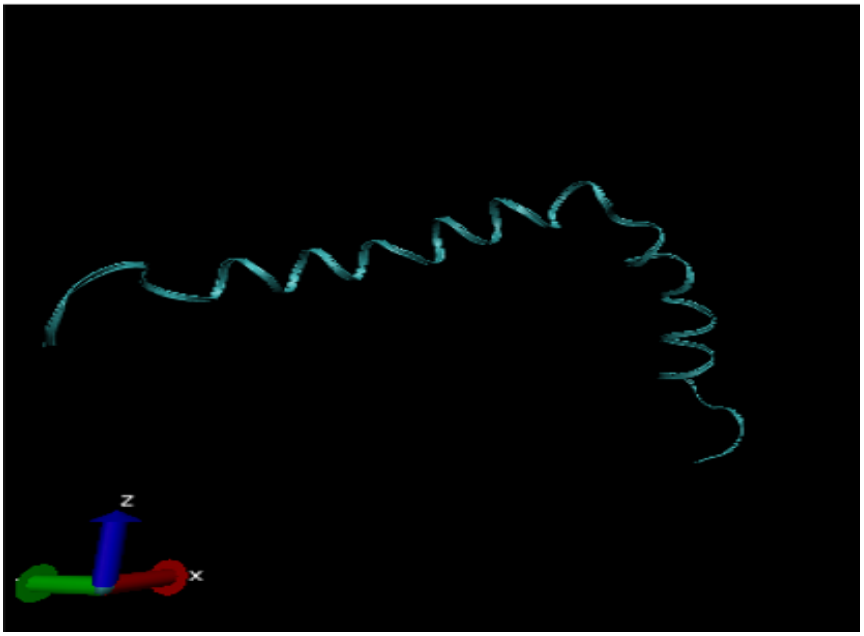

Figure 3.2.14: Final confirmation of ApoA-1 in 1.7M Potassium Chloride simulated over 2000 ps

The final concentration that we ran through the simulation was a 1.8M. We ran this simulation for 1 ns and found that this was not enough run time to collect enough data to create an accurate R_g and RMSD representation.

To further compare the data from the various concentration simulations, we analyzed the average R_g for three of the concentrations that showed the best results over time. This will more clearly and directly show how the change in concentration affected the R_g value of the protein. Our average R_g shows that over time the 1.0M system created the most unfolded protein conformation while the 0.5M also created a slightly more folded conformation. It finally showed that the 0.6M system created the most stabile folded conformation of apoA-1.

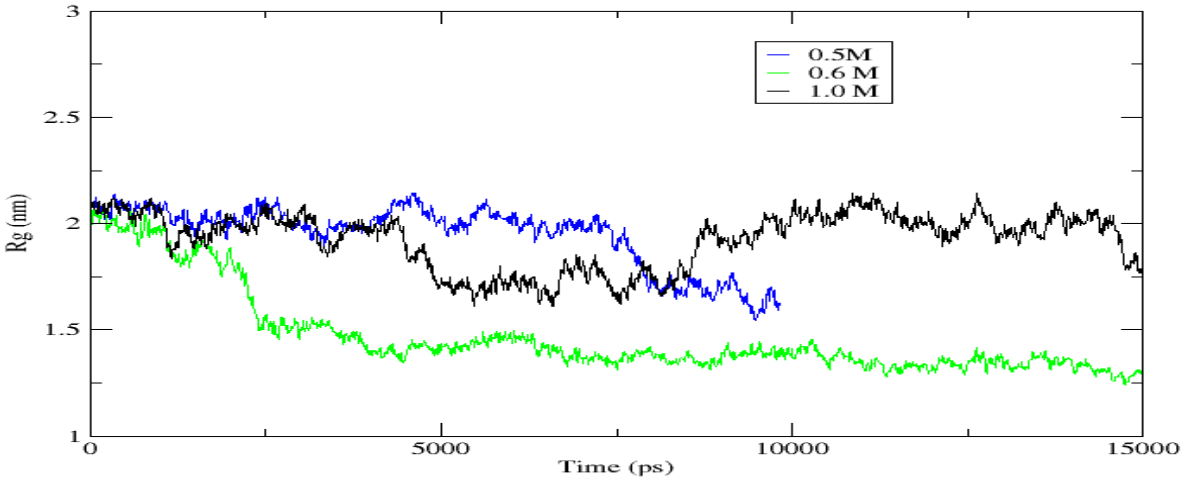

Figure 3.2.14: Comparison of average R_g for three concentrations of potassium chloride solution

Chapter 4:

Conclusions

4.1 Conclusions:

The main objectives of this work are to determine the effect that simulation time has on the overall conclusiveness of data collected in the topology of the simulation and to observe how the conformation of apoA-1 was affected by various concentrations of potassium chloride. The process for adding solvent and ions into the OPLS-AA force field has been well studied therefore those simulations required following the well-established procedure for the simulation. Our first goal was to see how a change in the ionic salt concentration would affect the folding conformations of ApoA-1. Additionally, we wanted to compare these findings to the previously documented data on the folding of ApoA-1 in sodium chloride. The methodology behind our process has been well documented by previous research and was repeated in our study, therefore, it is repeatable for similar ions at similar concentrations. The topology files and VMD visuals from our simulations showed that the ions effects were documented into the trajectory of the simulation and were used in the trajectory calculations. We did have some difficulties with lower concentrations. The system wanted to add solvent molecules to help with spatial stability. We believe this problem can be solved by using a different method for adding molecules by adding them by count as opposed to using the concentration command. Additionally, we found that a short run time, one under 25 ns, was not sufficient to provide the needed data to make a conclusive account of the protein's folding as a function of concentration and time. We believe that by increasing the run time at 100ns we can achieve more conclusive results about the folding patterns. This has been further proven by research conducted previously that showed more complex and complete data of the folding patterns of ApoA-1.

In the simulations that were ran for 25 ns we can see more definite folded or unfolded conformations. Previous research studied how ApoA-1 folded in sodium chloride. It is known that at some concentrations of sodium chloride ApoA-1 is unfolded and at others it is folded. Figure 4.1.2 shows the R_g values found in the previous study of ApoA-1 in sodium chloride. In these studies, they ran the simulation for 1000ps and followed the same simulation process as was followed in our study. Their study found that 0.5M and 1.8M sodium chloride solutions created a stable unfolded conformation for ApoA-1 while 0.6M, 1.0M and 1.7M created a stable folded conformation. In our study, we found that 0.5M & 0.6M solution created a stable folded conformation while 1.0M created a stable unfolded conformation. Although it did not have enough data to be conclusive the 1.8M potassium chloride solution appeared to create an unfolded conformation which can be seen in figure 4.1.3.

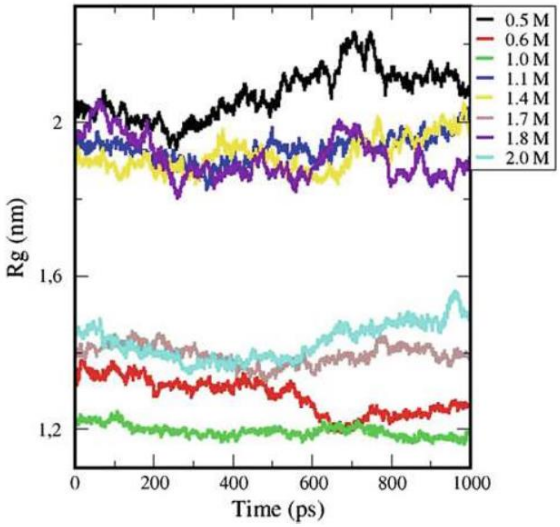

Figure 4.1.2: Time evolution R_g values of ApoA-1 in various concentrations of sodium chloride [10].

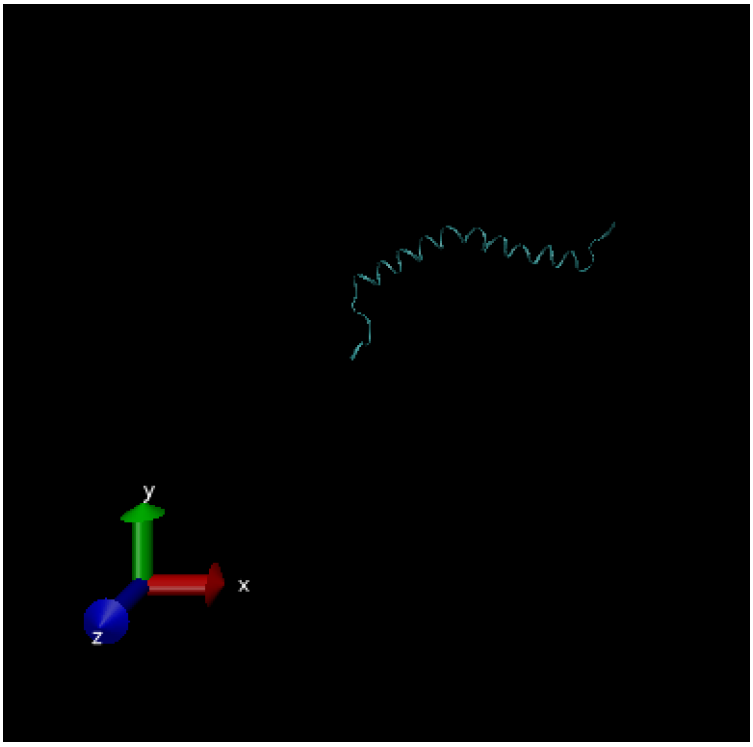

Figure 4.1.3: Final confirmation of ApoA-1 in 1.8M potassium chloride solution simulated over 1ns

Potassium Chloride is found to allow for and create a stable folded conformation of ApoA-1 at specific concentrations. It is also found to allow for and create stable unfolded conformations of ApoA-1 at other concentrations. This is confirmed by both the topology data and the visual representations of the final conformations of the protein. It was found that these concentrations did not fall into the same pattern as those of the previously conducted simulations with sodium chloride solutions. Although some at some concentrations ApoA-1 had similar conformations in both potassium chloride and sodium chloride there was not enough correlation to indicate that the concentration itself and not the ionic salt was responsible for the folding patterns. Additionally, it was found that a time of at least 25 ns was needed to collect enough data in the topology file for a conclusive picture to be created of the conformation of the protein

at the end of the simulation. Furthermore, a longer time is needed to allow for viewing and the protein's structural fluctuations over the time of the simulation.

4.2 Future Work:

In continuing our research, the next step will be to rerun the concentrations in this work for a longer time period, somewhere around at least 100ns. Additionally, we would like to run other concentrations of potassium chloride to see how the folding patterns match up with not only our results, but the results from the previous research involving sodium chloride. We hope to replicate the simulations performed with other group one ions paired with chlorine to make other ionic solvents. These studies will help to confirm that the type of ions used in the solution is important in the folding patterns of ApoA-1 and that concentration also plays a key role in the protein's conformational changes.

References

[1] Levitt, M., & Warshel, A. (1975). Computer simulation of protein folding. Nature, 253, 694. doi:10.1038/253694a0

[2] Micaêlo, N. M., & Soares, C. M. (2008). Protein Structure and Dynamics in Ionic Liquids.

[3] Patel, R., Kumari, M., & Khan, A. B. (2014). Recent Advances in the Applications of Ionic Liquids in Protein Stability and Activity: A Review. *Applied Biochemistry and Biotechnology*, *172*(8), 3701-3720. doi:10.1007/s12010-014-0813-6

[4] Ansell, B. J., Watson, K. E., Fogelman, A. M., Navab, M., & Fonarow, G. C. (2005). High- Density Lipoprotein Function. *Journal of the American College of Cardiology*,*46*(10), 1792.

[5] LDL & HDL: Good & Bad Cholesterol. (n.d.). Retrieved March 8, 2020, from https://www.cdc.gov/cholesterol/ldl_hdl.htm

[6] James C. Phillips, Willy Wriggers, Zhigang Li, Ana Jonas, and Klaus Schulten. Predicting the structure of apolipoprotein A-I in reconstituted high density lipoprotein disks. *Biophysical Journal*, 73:2337-2346, 1997

[7] Gorshkova, I. N., Liu, T., Kan, H. Y., Chroni, A., Zannis, V. I., & Atkinson, D. (2006). Structure and stability of apolipoprotein a-I in solution and in discoidal high-density lipoprotein probed by double charge ablation and deletion mutation. *Biochemistry*, *45*(4), 1242–1254. https://doi.org/10.1021/bi051669r

[8] Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B., & Lindahl, E. (2010). Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation, 6*(2), 459-466. doi:10.1021/ct900549r

[9] J. R. Claycomb and J. Q. P. Tran, *Introductory Biophysics: Perspectives on the living state*.Sudbury: Jones and Bartlett, 2011.

[10] Balderas Altamirano, M. A., Gama Goicochea, A., & Pérez, E. (2015, 2015//). *Folding of the Apolipoprotein A1 Driven by the Salt Concentration as a Possible Mechanism to Improve Cholesterol Trapping.* Paper presented at the Selected Topics of Computational and Experimental Fluid Mechanics, Cham.

[11] APOA1 apolipoprotein A1 [Homo sapiens (human)] - Gene - NCBI. (n.d.). Retrieved August 15, 2020, from https://www.ncbi.nlm.nih.gov/gene/335

[12] Wang, & R.j. (n.d.). The helix-hinge-helix structural motif in human apolipoprotein A-I determined by NMR spectroscopy. Retrieved February 1, 2020, from https://www.rcsb.org/pdb/explore/remediatedSequence.do?structureId=1GW3

[13] Smith, Benjamin, "Molecular dynamics simulations of apolipoprotein A1 in ionic liquids" (2020). Honors Theses

[14] R. H. Stote, “Introduction to Molecular Dynamics Simulations.” [Online]. Available: http://www.ccp4.ac.uk/maxinf/mm4mx/PDFs/MM4MX.MD.Stote.pdf. [Accessed: 20-Nov-2020].

[15] Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. *Journal of the American Chemical Society* **1996** *118* (45), 11225-11236. DOI: 10.1021/ja9621760

[16] Bye, J. W., & Falconer, R. J. (2013). Thermal stability of lysozyme as a function of ion concentration: A reappraisal of the relationship between the Hofmeister series and protein stability. *Protein Science: A Publication of the Protein Society, 22*(11), 1563-1570. doi:10.1002/pro.2355

[17] Bye, J. W., & Falconer, R. J. (2014). Three Stages of Lysozyme Thermal Stabilization by High and Medium Charge Density Anions. *The Journal of Physical Chemistry B, 118*(16), 4282-4286. doi:10.1021/jp412140v

[18] Lemkul, J. (2018). GROMACS Tutorial. Retrieved from http://www.mdtutorials.com/gmx/lysozyme/indexhtml

[19] Chevrot, G., Fileti Eudes, E., & Chaban Vitaly, V. (2015). Enhanced stability of the model mini-protein in amino acid ionic liquids and their aqueous solutions. *Journal of Computational Chemistry, 36*(27), 2044-2051. doi:10.1002/jcc.24042

[20] M. Yu. Lobanov, N. S. B., and O. V Galzitkaya. (2008). Radius of Gyration as an Indicator of Protein Structure Compactness. *Molecular Biology, 42*, 623-628.

[21] Ivankov, D. N., Bogatyreva, N. S., Lobanov, M. Y., & Galzitskaya, O. V. (2009). Coupling between Properties of the Protein Shape and the Rate of Protein Folding. *PLOS ONE, 4*(8), e6476. doi:10.1371/journal.pone.0006476

[22] P. E. Ramirez-Gonzalez, "On the stability of Apolipoprotein A1 in presence of ionic liquids," 2018.

Appendix 1:

Specific commands for potassium chloride simulations in GROMACS. Work follows the methodology of a simulation of a lysozyme in water. [15]

Generate topology

```
gmx pdb2gmx -f apoa.gro -o apoa_processed.gro -water spce
```

Choose forcefield 15

```
Select the Force Field:
From '/usr/local/gromacs/share/gromacs/top':
 1: AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
 2: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
 3: AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
 4: AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
 5: AMBER99SB protein, nucleic AMBER94 (Hornak et al., Proteins 65, 712-725, 2006)
 6: AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)
 7: AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
 8: CHARMM27 all-atom force field (CHARM22 plus CMAP for proteins)
 9: GROMOS96 43a1 force field
10: GROMOS96 43a2 force field (improved alkane dihedrals)
11: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
12: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
13: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
14: GROMOS96 54a7 force field (Eur. Biophys. J. (2011), 40,, 843-856, DOI: 10.1007/s00249-011-0700-9)
15: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
```

Create a “box” containing our system

```
gmx editconf -f apoa_processed.gro -o apoa_newbox.gro -c -d 1.0 -bt cubic
```

Add water

```
gmx solvate -cp apoa_newbox.gro -cs spc216.gro -o apoa_solv.gro -p topol.top
```

Add ions

```
gmx grompp -f ions.mdp -c apoa_solv.gro -p topol.top -o ions.tpr
```

For Potassium Chloride with specified concentration

```
gmx genion -s ions.tpr -o apoa_solv_ions.gro -p topol.top -pname K -nname CL
-conc
```

Energy minimization

```
gmx grompp -f minim.mdp -c apoa_solv_ions.gro -p topol.top -o em.tpr

gmx mdrun -v -deffnm em
```

NVT equilibration

```
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr

gmx mdrun -deffnm nvt
```

NPT equilibration

```
gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o
npt.tpr

gmx mdrun -deffnm npt
```

MD production

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr

gmx mdrun -deffnm md_0_1
```

Data Analysis

```
gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -pbc mol -center
```

RMSD

```
gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns
```

R_g

```
gmx gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -o gyrate.xvg
```