5-2020

Molecular dynamics simulations of apolipoprotein A1 in ionic liquids

Benjamin Smith
University of Tennessee at Chattanooga, jln766@mocs.utc.edu

Molecular dynamics simulations of apolipoprotein A1 in ionic liquids

Benjamin Smith

Departmental Honors Thesis

University of Tennessee at Chattanooga

Department of Chemistry and Physics

Dr. Luis Sanchez-Diaz

Professor of Physics

Project Director

Dr. Jisook Kim

Professor of Chemistry

Departmental Examiner

Abstract

High density lipoprotein function is essential for healthy processing of fats in the blood. Function of these proteins is determined by their conformational structure of chains of amino acids. Certain ionic liquids have been shown to interact with individual amino acids to affect the tertiary shape of the protein. These interactions can disrupt the native hydrogen bonds between amino acids which can either promote folding or denaturing of the protein. Our work focused on apoA-1, a high-density lipoprotein that binds to cholesterol for efflux from the body. We studied this protein in different ion-based liquids with sodium, chloride, and 1-ethly-3-methylimidazolium (EMIM). The different solutions will be studied at different concentrations in water. This work will be done using molecular dynamic simulations of apoA-1 immersed in solution. Our simulations revealed which ions promote folding and which would be potential candidates for making solutions that stabilize apoA-1. The systems with sodium cations and others with chloride anions in solution reveal the effects of charge on the shape of the protein. EMIM cation with chloride counter anions in solutions show that as the solvent concentration increases, the protein folds unstably but maintains its native secondary structures.

Table of Contents

Chapter 1:

Introduction

1.1 Background:

Molecular dynamics (MD) is a simulation method that has proven key to the understanding of protein folding and stability. [1] This computational approach answers many biophysical questions that are difficult to discover otherwise. MD can create massive amounts of topological data which can be used to analyze the relationship of a sample and solvent to better understand the molecular interactions. [2] This information can be used to determine how to best stabilize a protein or learn what properties of solvents promote conformational changes. [3] There is quantifiable data produced from each atom in the system and we can analyze this to determine crucial information on protein functions. In this way, MD allow for the study of macromolecules while still gaining information on individual interactions, and they work alongside other biochemical techniques such as crystallography, neutron scattering, and X-Ray scattering to better understand proteins. [4]

1.2 Apolipoprotein A1:

Apolipoprotein A1 (apoA-1) is our protein of interest because of its important role in removing cholesterol from the bloodstream. [5] ApoA-1 is a major structural protein for high-density lipoproteins (HDL) which are what are now referred to as the “good” cholesterol. [6] HDL exists in blood plasma and helps to remove lipids from the body. They bind to fats released by other cells and transport them to the liver for release. Because of their binding affinity to other proteins and lipids, apoA-1 is classified as a monomeric protein. HDL structure has been characterized by having multiple apoA-1 fragments which are folded and largely consist of alpha helixes surrounding inner lipids. [7] The proper function of HDL is important to many different health issues. HDL can prevent heart diseases and plaque build-up and has been shown to act as an anti-inflammatory as well. [8] Doctors are now testing the concentration of this cholesterol in

patients to determine healthiness and make dietary recommendations. The presence of these in plasma helps to determine your relative risk for certain diseases. A variety of HDL, as well as apoA-1 specifically, have been shown to be much more effective at preventing these diseases at higher concentrations in the blood. Not only does this mean that the production of apoA-1 is essential, but also the structure of this protein is important in determining its function. This structure determines the binding ability to lipids and effectiveness in forming HDL. Tertiary structure and conformational results from MD can help to provide insight into the specific interactions between the protein and solvent. We can change the properties of the solvents to isolate properties that promote folding or denaturing. This also allows us to identify potential regions for these interactions and understand more about the structure of apoA-1. A key component to the secondary structure of apoA-1 is its amphipathic α helixes which make up 43% of the chain. [9] These structures are common in proteins involved in lipid binding and should help make determinations for less prominent HDL proteins. [10] We use different simulations to see how our protein changes shape in different solutions to make conclusions on specific methods of folding and stabilization.

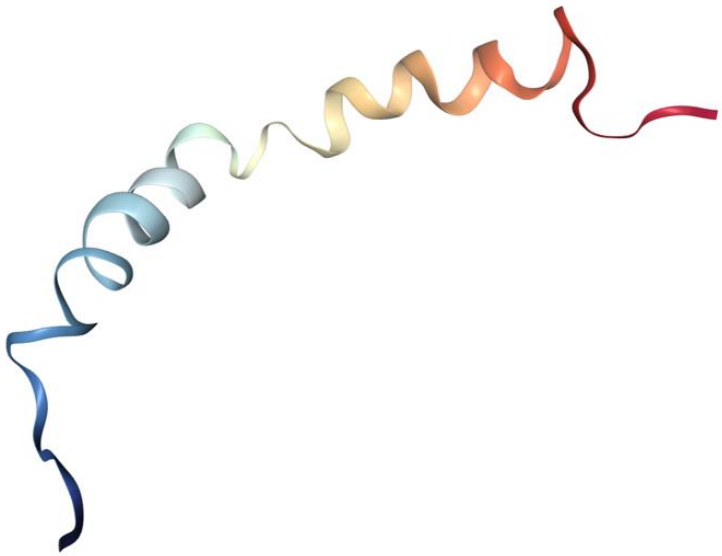

Figure 1.2.1: The native structure of apoA-1 from Protein Data Bank

1.3 Molecular dynamic simulations:

MD are now commonplace in biophysical research; they allow researchers to speed up research and find data much quicker than previously physically possible. [11] Molecular dynamic simulations are a subclass of general chemical simulations and implement many of the same principles. However, the MD software that we use is more geared towards biochemical interactions and biological questions. Advancements in computational biology allow us to download our apoA-1 directly from the Protein Data Bank and view it on our lab computer in seconds. ApoA-1, labeled as 1GW3 by the Protein Data Bank, is available with its native structure from NMR spectroscopy. This protein file contains all the amino acids and its native structure. The properties of amino acids and molecules such as mass and charge are known and interpreted by the simulation program. We then put this protein in a "box" containing a certain amount of solvent. The exact heat, molarity, pressure and volume is all set and controlled by researchers and the advantage to this technique is the control that researchers have on the system parameters. Different parameters can be held constant without the physical difficulties of non-computational studies. The simulation software is able to generate data that can then be analyzed by researchers to understand physical properties of the interaction of atoms of the protein with the solvent. This process is detailed in the simulation details section and the exact command lines are given in Appendix 1 and 2. Programs are able to calculate the potential energies of atoms and show their progression over time. These interactions are mapped out as trajectories for each atom and are taken at each incremental time step. Because of this, simulations often take a lot of computing power to run quickly. The demand for processing power makes MD a popular application of parallel high-performance computing. Oftentimes researchers will choose to run

their simulations at supercomputers such as Summit at Oak Ridge National Laboratory. The laboratory computer here allows us to run on 24 core processors to greatly reduce run times. MD are designed to help guide early stage research to prevent wasted time and also, they serve to confirm physical experiments. The accepted target time length for MD protein folding is 100 ns which is broken down by very small time increments. [12] This allows for the dynamics of the protein to be studied and for the randomness of the solvent to not skew data.

1.4 Protein Folding:

Proteins are made of combinations of the twenty essential amino acids present in humans that are joined by peptide bonds. The properties of each amino acid group are determined by the structural differences among them. All amino acids contain a carbon atom attached to amine group and carboxyl group and differ based on the -R group also attached to the central carbon. The interactions of these -R groups with the surrounding environment created interaction forces that caused structural changes to the protein. Electrostatic and ionic interactions, disulfide bonds, van der Waals forces and hydrogen-bonding must all be considered, and all play a role in determining tertiary structure of proteins. [13] The tertiary structure is the three dimensional shape of a protein and is a result the binding and repulsive forces mentioned. The structure of these proteins can be denatured by salts, pH levels, and some organic solvents that disrupt interactions. The main cause for destabilization is a loss of conformational entropy. There are other formations such as salt bridges that work to stabilize the protein. [14] Folding can be studied by calculations of the Gibbs Free Energy as a function of time in a simulation. Also, denaturation is found to be proportionate to the accessible surface area. Due to the nature of protein study, these calculations are all done as averages for many data points. Some of these simulations are done with known models of proteins. However, our simulation work is an *ab*

initio method as the structure is calculated through the simulation and not numerically compared with a known model for denaturation.

1.5 Previous Studies

ApoA-1 folding is not new topic of question for researchers. Balderas has shown an oscillating pattern in radius of gyration as NaCl concentration is increased in simulation. [15] This pattern indicates an open and closed position of the protein which is affected by the solvent concentration. Balderas suggests the formation of solvation spheres affecting the native electrostatic interactions which causes the shape to denature. This work is what inspires the study of the sodium cation and chloride anion individually. ApoA-1 has also been previously simulated in solution at high solvent concentrations (up to 75% ions in solution) and over 50 nanoseconds with 1-ethly-3-methylimidazolium (EMIM) cations with chloride counter anions. [10] Their work revealed that for their parameters, the EMIM-Cl solution and similar organic cations with short chains promote the unfolding and denaturing of apoA-1. This work also served as a general guide to performing our simulations in GROMACS.

1.6 Objectives:

Our work will serve to further previous studies on the effect of ionic solutions on the conformation of apoA-1. We implement specific ions in solution to determine how each effects the protein over time. Na^+, Cl^-, and EMIM-Cl based ionic solutions will be run through simulations over the course of 100 ns. From this work, we hope to identify certain factors of the solvent that promotes a folded state and how the protein can be stabilized. This research will also create a complete detailed procedure for how to implement ions in simulations of apoA-1 using GROMACS. We also will provide data to better identify the specific areas of the protein which are active in the formation of secondary and tertiary structures.

Chapter 2:

Materials & Methods

2.1 Apolipoprotein A1 Structure:

The sequence of amino acids and secondary structures for regions of the protein are shown below. [8] The initial conformation is from the helix-hinge-helix structural motif in human apoA-1. Twenty-two of the amino acid residues are hydrophobic while seven are hydrophilic. Ten have a positive charge while six are charged negatively. [10] Two α helices can be seen from amino acid 14 to 21 and 27 to 35. This secondary structure limits movement in this region as the strand is stabilized by hydrogen bonds. α helixes have amine groups binding to the carboxyl groups of amino acids ahead in the sequence. Notice that some sections do not begin with a specific secondary structure and may be more inclined to structural changes.

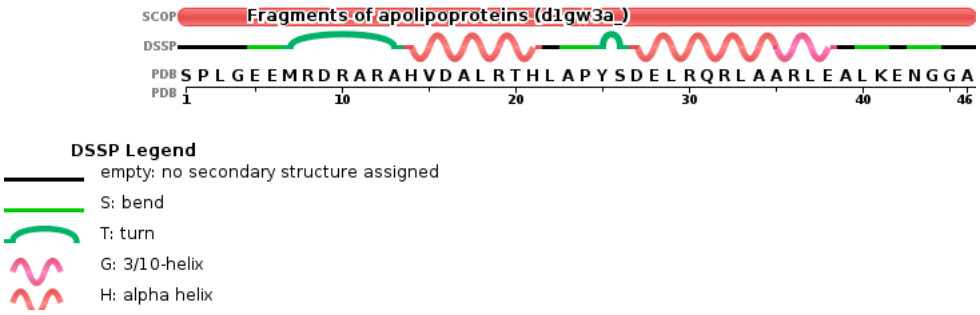

Figure 2.1.1: Amino acid sequence and secondary structure of apoA-1

2.2 GROMACS:

The Groningen Machine for Chemical Simulations, or GROMACS, is a free MD package for the study of molecular dynamics. It calculates the equations of motion for each particle in the system. [16] In doing so, the package solves the Newtonian equations of motion for N atoms and that force is found to be the inverse derivative of the energy. [17]

Equation 2.2.1:
$F = ma = m\left(\frac{dv}{dt}\right) = m\left(\frac{d^2r}{dt^2}\right)$
Equation 2.2.2:
$F = -\nabla E$

Solving these equations for all the molecules in the system is virtually impossible to do in a reasonable amount of time without simulation data analysis. Our trajectories are given by positions as a function of time, r(t). The program can calculate bonding and non-bonding interactions and is specifically designed for use with proteins. There is also a cut off radius long range for Leonard-Jones interactions to reduce computing time and limit interaction calculations. GROMACS is controlled via command lines in the terminal with input and output files. This package is well-developed and is much more robust than other simulation methods. For this reason, GROMACS is a popular package in for chemical simulations. Multiple cores can be implemented to run simulations simultaneously and greatly reduce computational time. Periodic boundary conditions can also be implemented to prevent hinderance from the size of our "box." GROMACS also allows for the change in the equations used for water implementation and ionic interactions in the system.

Force Field:

Simulations have to have some sort of specific guidelines for simulation interactions; these are given by the force field. The main functions of the force fields are to establish potential energies between molecules and to apply restraints to the system. This will be how the simulation will calculate the energy landscape of interactions. We implemented the AMBER94 force field in GROMACS, similarly to previous protein studies. [15] This force field has been

found to work well with helix structures in proteins well. [18] AMBER uses a simple point charge (SPC) water model to calculate the interactions between the protein and ions with the water molecules in the system. AMBER also contains the properties of atoms and many macromolecules, because many are already implemented, errors are less frequent in energy potentials. This is a rigid water model and is the default system in GROMACS as well which makes it easier to add water this way. The system is able to find potential energy as a function of position for each atom interaction.

Equation 2.2.3:

$$V(r) = \sum_{1,2\ pairs}\frac{1}{2}K_b(b - bb_0)^2 + \sum_{angles}\frac{1}{2}K_\Theta(\Theta - \Theta_0)^2 + \sum_{dihedrals}K_\phi(1 + \cos(n\phi - \delta)) + \sum_{i,j}\left\{4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon D r_{ij}}\right\}$$

The above equation shows how the energy is calculated for interactions in the system for all atoms. The first three terms are variations of spring potential with a restoring force K and represent attractive forces. [13] [18] The last term in the equation represents how the Leonard Jones force is calculated from Pauli repulsion and Van Der Waal interactions. The non-bonded interactions require the most computing time which is why a cutoff radius is needed. The position of atoms based on these equations can be done using a Taylor Series of $(r(t + \Delta t))$ where the known position at a velocity v(t) can be multiplied by the time step to create a new position in the box. The size of the time step determines the accuracy of the simulation as a smaller time step will give better results. Most time steps are less than 1 fs. AMBER has many common molecules already implemented into its supporting files so typically there is no need to adjust the files. However, in our work with lipids we have to add the lipids so

that the system can understand our inputs. Our lipids parameters are implemented in the same way as previous research to provide consistency. [10] AMBER also has the known structures and interactions of amino acids so that when our protein is download from PDB each section has quantifiable information for the simulation.

2.3 Solvents:

Ionic liquids (IL) are used in MD to attempt to mimic environments in the blood and stabilize the protein. [19] Many solvents have been used previously, with mixed effects on the stability based on charge and molarity. [20] Our initial study works with sodium cations in an aqueous solution which are compared chloride based ionic solutions. This work is intendent to provide some information regarding the effects of charge on the stability and folding of apoA-1. The other simulation will focus on the lipid cation, 1-ethly-3-methylimidazolium (EMIM). This all use chloride as the corresponding anion in the IL. [10] EMIM based IL have been studied previously in comparison with similar organic molecules. The only difference between these solutions are the cations which have different length of amphipathic chains and polarity.

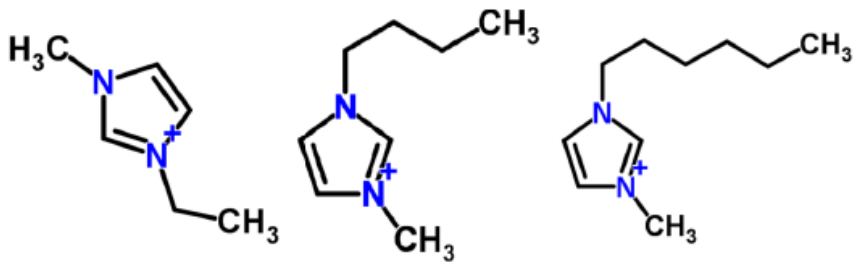

Figure 2.3.1: Molecular structures EMIM (left), 1-butyl-3-methylimidazolium (BMIM) (center), and 1-hexyl-3-methylimidazolium (HMIM) (right) shown. [10]

This work studied variance in the chain length for these molecules to provide insight into how the protein reacts differently with hydrophobic molecules (HMIM) compared more

hydrophilic molecules (EMIM). Our work will be to recreate their simulation of EMIM, by doing so, we will establish a more detailed process setting up simulations of apoA-1 with outside molecules than is currently available. We will also increase the simulation run time up to 100 ns to give a more complete data set of atom trajectories and protein structure. This will double the amount of time EMIM based IL have been shown with apoA-1. Our work will provide a better picture of the change in protein structure over time. We also see how the different charges associated with each solvent affect stabilization and folding differently. This will also allow for us to designate certain regions of apoA-1 where significant changes occur in structure to see which sections of the protein are interacting with the solvent. [21] Interactions with solvents can disrupt native hydrogen binding and change secondary and tertiary structure. Depending on the properties of these amino acids we can develop a better understanding in optimization of IL to use with apoA-1. The IL studied will be compared with simulations of a pure water solution.

2.4 Simulation Details:

The first technique used to study the sodium and chloride based ionic solutions was done relatively easily compared to EMIM because all the molecules were already implemented into GROMACS. This process followed a previous simulation of a lysosome in ionic liquids. [22] The first step is to create the system. Making the system requires apoA-1 to be downloaded as a pdb file and put into a box with periodic boundary conditions on each side. This file type is designed by the protein data bank to describe the three-dimensional structural information of proteins as well as specific amino acid sequence. The box is then solvated using the SPC water method and the genion function in GROMACS is used to add select quantities of each solvent. Note that the specific amount of water molecules will change depending on the amount of solvent inserted. Once we have created our box, we can use the pdb2gmx command to create a

topology file to the new system. This topology file will contain the coordinates for all water, protein, and solvent molecules at all times through our simulation and will provide raw data to be analyzed later. The topology file is updated through each following step. Our system then undergoes energy minimization to ensure a consistent starting energy for each simulation and reduce steric clashes that occur otherwise. This minimization is done at 100 atms the temperature is dropped from 1000 K to 300 K in 200 K steps with 2 ns spent at each step. Now, it must be noted that the number of calculations performed during each step is determined by the .mdp file used for each step. This has a significant impact on the amount of raw data collected. Following this, you can view the change in potential energy over time to see the energy as an exponential decay with an asymptote at the minimum energy produced. Next, we performed NVT equilibration where volume was held constant with simulated annealing temperatures rising up to 2000 K back to 300 K with 200 K steps and 2 ns spent at each step. We can then see the temperature over time for this as well and ensure it ends on 300 K. Next, we use NPT equilibration at 300 K for 6 ns with a constant pressure. After this, we can view the total energy for the system which should be relatively flat now and should resemble the graph when this process is repeated. Now, due to the randomness of molecules in the system, the graphs will not be identical, but they should be at similar energy levels. The final command actually runs the production of data for our simulation. This command is also controlled by another mdp file and has specific step number and time. To reach our optimal total simulation time of 100 ns, it required 500 million steps with 0.002 fs steps. This process was by far the most time-consuming computing part and took nearly a month on 24 cores with constant power. This process was repeated for specific quantities of sodium and chloride ions. At the completion of the data production, our sodium and chloride ions have data that can be analyzed numerically as well as

VMD representations to view the final structure. The original apoA-1 box system file can also be put into VMD and then provided with the topological data and the entire process can be shown as a movie over the 100 ns. The details for this process are shown in Appendix 1.

The simulations of EMIM is more difficult to prepare because these molecules are not already implemented into GROMACS or our force field. So, to fix this, we followed previous work [10] to update the topology files within the program so that the system would recognize what we were trying to insert into the box and also how it treats the protein and solvent interactions with these molecules. We then attempted to follow the previous technique as closely as possible but had to change certain steps due to errors in the simulation. Many of these errors occur due to the size of these molecules which can create steric clash issues. Once we have downloaded our protein, the first command is inserting the molecules into a box with our apoA-1 with the same size and periodic boundary conditions. Again, the genion function of GROMACS inserts the ions into solution. It is worth noting that this was prior to solvation, differing from our process previously used. We then had to create a buffer between our solution and the outer edge of the box to prevent an error that kept occurring in solvation which follows. Solvation also used the simple point change form of water to solvate our box and leave us with all parts necessary to begin simulation. We used pdb2gmx to create an initial topology file from this system which we will continuously update in each new command. From here, we followed the previous process. We minimized this system as done previously. Then, we did NVT equilibration followed by NPT equilibration for each different concentration to create a consistent starting system for each simulation. We verified energy and temperature visually prior to MD production for each system. Now, the files were subjected to MD production with 500 million steps to produce 100 ns of real simulation time to compare with previous studies of these molecules at shorter time

periods as well as our simple ionic solvents we have already simulated. This simulation took just as much time on our laboratory computer as the simple ionic solutions. The details for this process are given in Appendix 2.

2.5 Data Analysis Techniques:

Once MD production has completed, the simulation itself is over and we can now figure out what actually happened with our protein in solution. The first, and easiest, way to do this is visually. Each system can be pulled up in Visual Molecular Dynamics which will show us the final confirmation of our system and allows us to see the protein in box. This file can then be loaded with the topology file which will allow us to see the protein over the course of the 100 ns of simulation. Now, this process does not provide any data to us, it serves as a tool to check the impact of the analyzed data we will find. [23] To analyze our data, we calculated the Radius of Gyration (R_g) as well as the Root Mean Squared Distribution (RMSD).

Radius of Gyration:

R_g describes the spread of the backbone of the protein from the center of mass (R(c)). To do this, it must first calculate the center of mass shown below.

Equation 2.5.1:

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

This calculation is of all atoms found within the protein backbone. This equation essentially finds the point at which the radius is exactly zero compared to the center of mass of the protein backbone. In this equation m represents the mass of a single atom. This step allows for the calculation of the R_g in real space. [24]

Equation 2.5.2:

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

Here, M represents the total mass of the atoms in the backbone and the actual R_g is defined by the square root of this value. The radius of gyration will provide us with an average distance from the center of mass for each atom in the protein and will serve to explain how compact or spread out the protein is over time. A lower R_g should indicate that the protein is more folded and compact which allows it to be closest to its functional form. A high R_g indicates that the protein is more spread out and is in more of a linear shape rather which would not be an effective shape for lipid binding and would indicate a poor solution for future protein study. Radius of gyration can also be used to determine the general stability of the protein in solution. When R_g values are constant then the protein is in a stable confirmation. At higher values, this does not tell us if the ionic solution interactions are the cause, but at lower values, a relatively constant value through simulation further indicates a solution which stabilizes the protein well. The radius of gyration for our ionic systems are directly compared with our control system in just water to see if these changes are because of the water or ions in solution. We did not normalize our data directly but used the control as a comparison on our graphs of R_g. This method is fairly common for protein study and is already implemented into GROMACS. [25] The program allows us to command the program to create a xvg file which shows our average radius in nanometers as a function of time.

Root Mean Squared Distribution:

RMSD is the other method used to analyze our topology from simulations. Similar to R_g, this calculation is built into GROMACS and can be performed with a simple command line. RMSD is the least squared fit of the structure compared to the initial confirmation. [10]

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 $t_2 = t_1 + \Delta t$

This equation will find the average change in position in nm of all atoms in real space over time. Again, we focus this calculation on the backbone of the protein, but it could be used to see how any group of atoms is specifically affected in the simulation. This data will provide a general picture of how much the protein changes from its initial confirmation. Thus, high RMSD values will indicate significant change at certain times in the simulation. Low values will indicate points where the protein is similar to the initial conformation. RMSD values also can tell us when the protein is stable. Long time intervals with consistent RMSD levels indicates times when there is little change in the total structure of the protein while unstable spikes in the plot over time indicate instability. However, the areas of flatness must be compared visually as well as with R_g to determine if the stability was because it was folded or because it remained in the initial conformation. None of these graphs will tell us the entire picture of the simulation, but together they are used alongside VMD visualizations to give us a better picture of the interactions with our protein.

Chapter 3:

Results

3.1 Ionic Solvent Concentrations:

Following the creation of the box and insertion of water and ions, we can use the topology file to determine concentrations. Although we had a known value for the ions inserted, the amount of water was slightly varied based on the number of ions in the system. The concentrations for each system are shown below. It is worth noting that the system added more water molecules for the simpler ions due to spatial issues in the box.

Ion	Water Molecules	Ion Concentration (%)
Cl^-	33315	0.0240
Na^+	33315	0.0240
Cl^-	33291	0.0961
Na^+	33291	0.0961

Figure 3.1.1: Chloride and Sodium residue concentrations as a percentage of the water molecules present in solution.

Cation	Water Molecules	Ion Concentration (%)
EMIM	15888	0.629
EMIM	15122	1.32
EMIM	14334	2.09

Figure 3.1.2: EMIM concentrations residue concentrations as a percentage of the water molecules present in solution.

3.2 Simulation Data:

Radius of gyration was calculated from the topography file for the entire protein structure. Root Mean Squared distribution was taken from the backbone of the protein which excludes side chains of the amino acid sequence. This is the same for all simulations performed.

Water

The simulation of water served as a control to see how the protein acts without any ions in the system. Our graphs reveal relatively small changes to the initial tertiary structure. There do not appear to be any new secondary structures present, indicating that the protein is stable, and is not being forced to fold. This is confirmed visually as we can the two disconnected α helixes which are consistent with the original structure. Visually, it is also apparent that the protein remains in a relatively linear tertiary shape and the ends of the fragment do not move together.

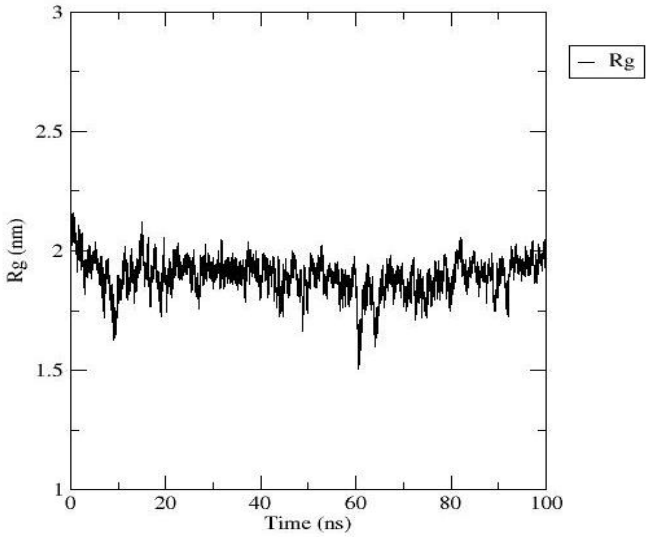

Figure 3.2.1: R_g (in nm) of apoA-1 in pure water solution simulated over 100 ns

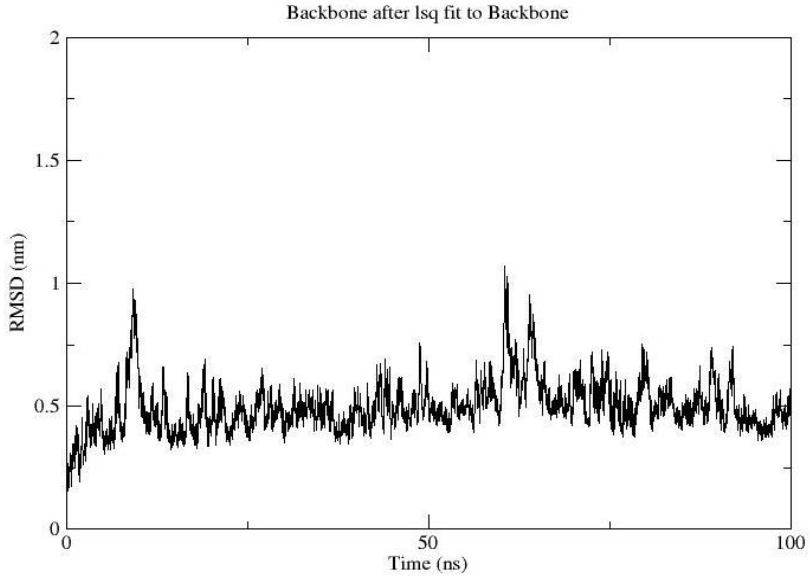

Figure 3.2.2: RMSD (in nm) of apoA-1 in pure water solution simulated over 100 ns

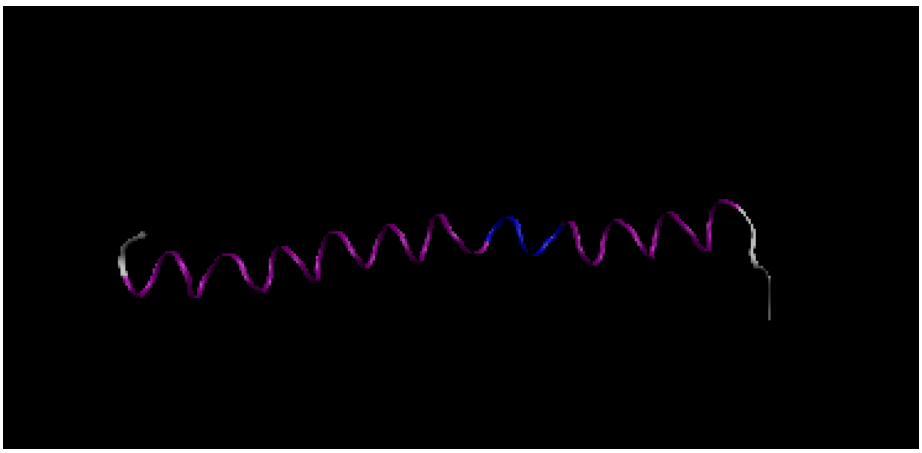

Figure 3.2.3: Final conformation of apoA-1 in water

Sodium Cation

The data collected from simulations of ionic liquids with the sodium cation embedded are shown below. At the lowest concentration of sodium ions, the R_g and RMSD fluctuate at the beginning and slightly near the end of the simulation. The changes of the backbone shown in Figure 3.2.4 seem to follow the pattern of oscillation which has been shown previously. [15] At the higher concentration, this pattern is not seen as clearly. For most of this simulation, the graphs resemble that of apoA-1 in water. However, just before the end there is a drop of R_g and increase of RMSD indicating that the protein is folding here. This is likely why our image of the final conformation shows definite folding as the ends of the fragment move together. The visualization also reveals that the α helixes have been connected across the central amino acids.

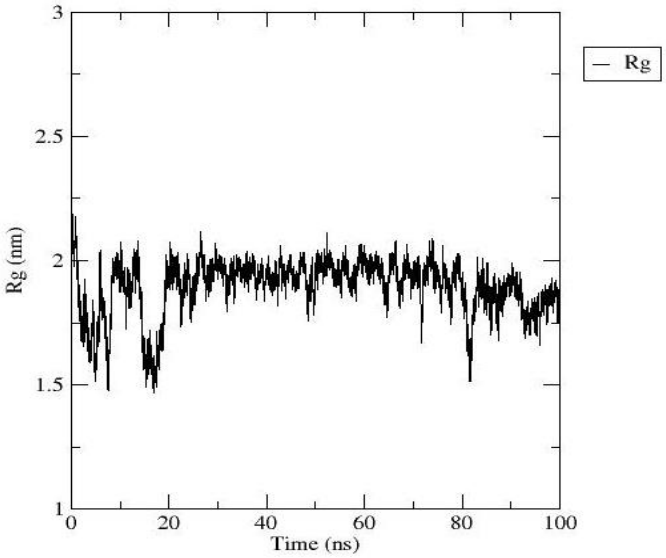

Figure 3.2.4: R_g (in nm) of apoA-1 in a 0.024% sodium cation ionic liquid simulated over 100 ns

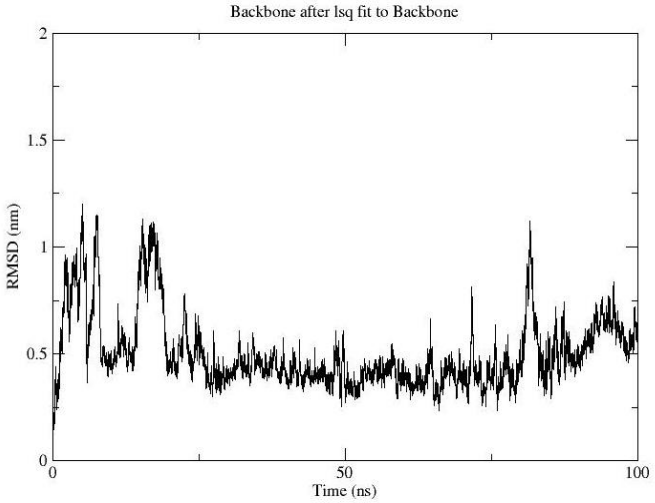

Figure 3.2.5: RMSD (in nm) of apoA-1 in a 0.024% sodium cation ionic liquid simulated over 100 ns.

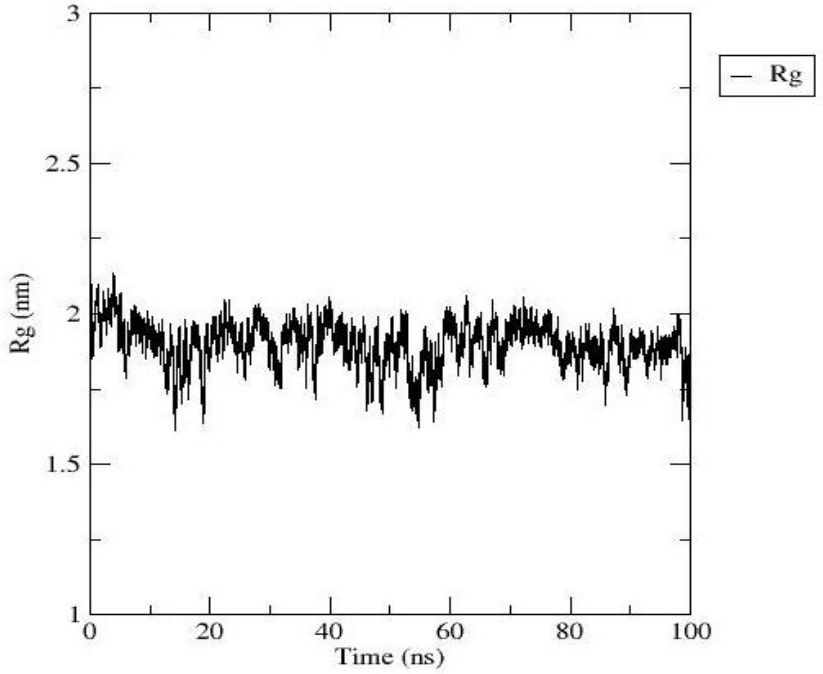

Figure 3.2.6: R_g (in nm) of apoA-1 in a 0.096% sodium cation ionic liquid simulated over 100 ns

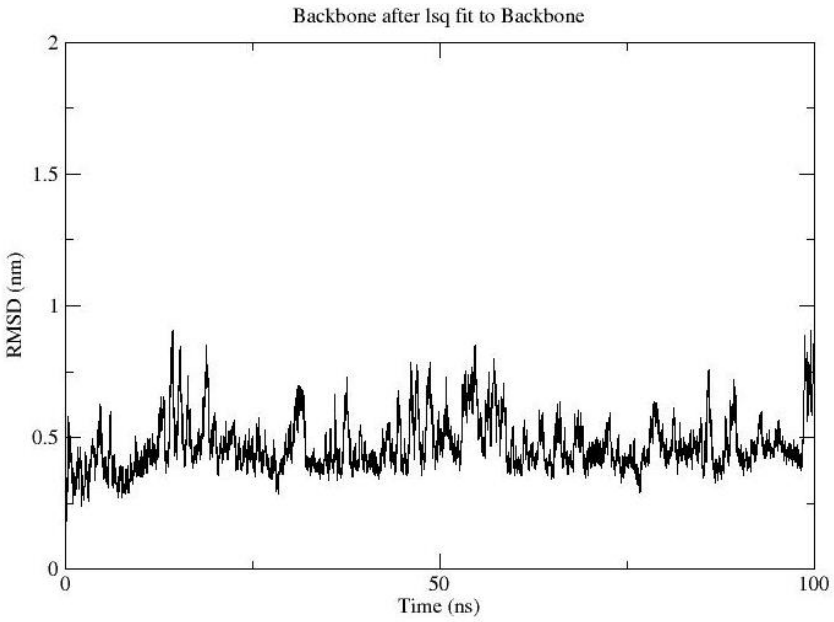

Figure 3.2.7: RMSD (in nm) of apoA-1 in a 0.096% sodium cation ionic liquid simulated over 100 ns

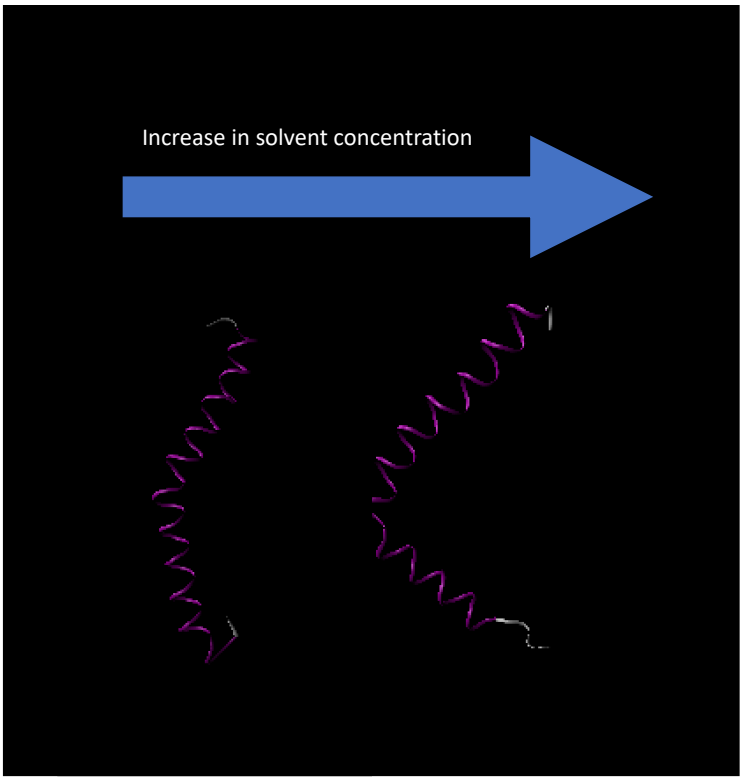

Figure 3.2.8: Final conformation of apoA-1 in sodium cation ionic liquids. The 0.024% solution is shown on the left and the 0.096% solution is shown on the right.

Chloride Anion

The simulations in which chloride anions were implemented to our solution show significant changes to the structure of apoA-1. At the lower concentration, the radius of gyration drops significantly in the first 40 nanoseconds. The shape then stabilizes which is seen in the flat section of the RMSD graph of the backbone. For the simulation at 0.096% ions, the graphs follow similar trends as that for 0.024% for the first 40 ns. After that, the R_g increases and the RMSD returns closer to its original value. Both of these simulations cause the denaturation of the native secondary structures, especially for the lower concentration. Visually, the compact folded shape for 0.024% agrees with our graphs and the general final shape for 0.096% is similar to the original shape.

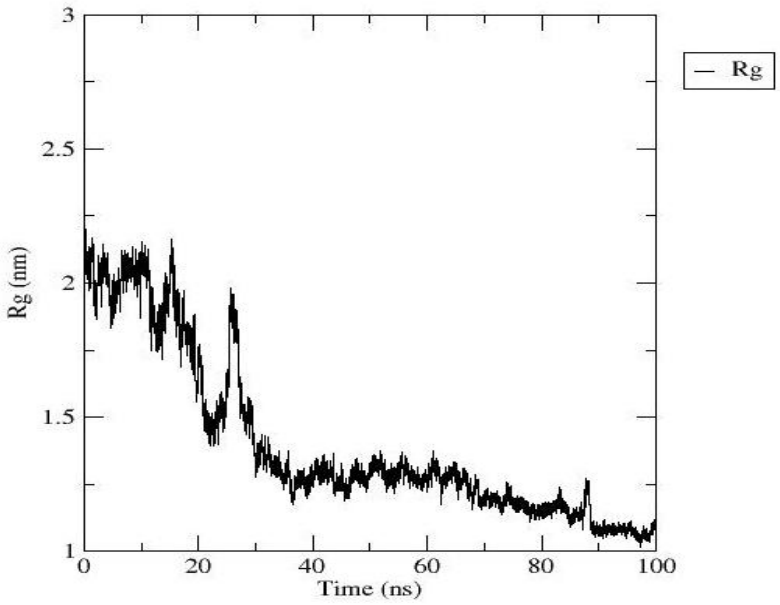

Figure 3.2.9: R_g (in nm) of apoA-1 in 0.024% chloride anion ionic liquid simulated over 100 ns

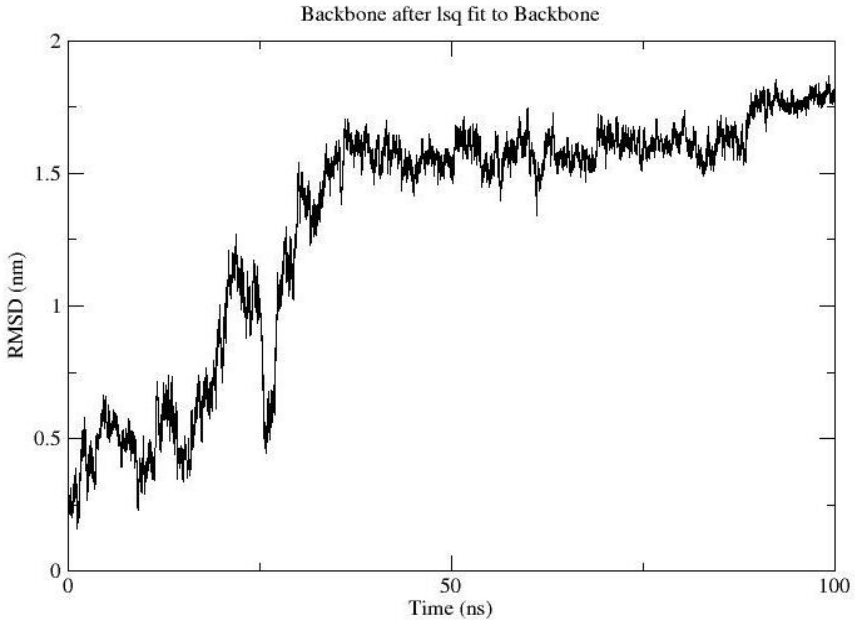

Figure 3.2.10: RMSD (in nm) of apoA-1 in 0.024% chloride anion ionic liquid simulated over 100 ns

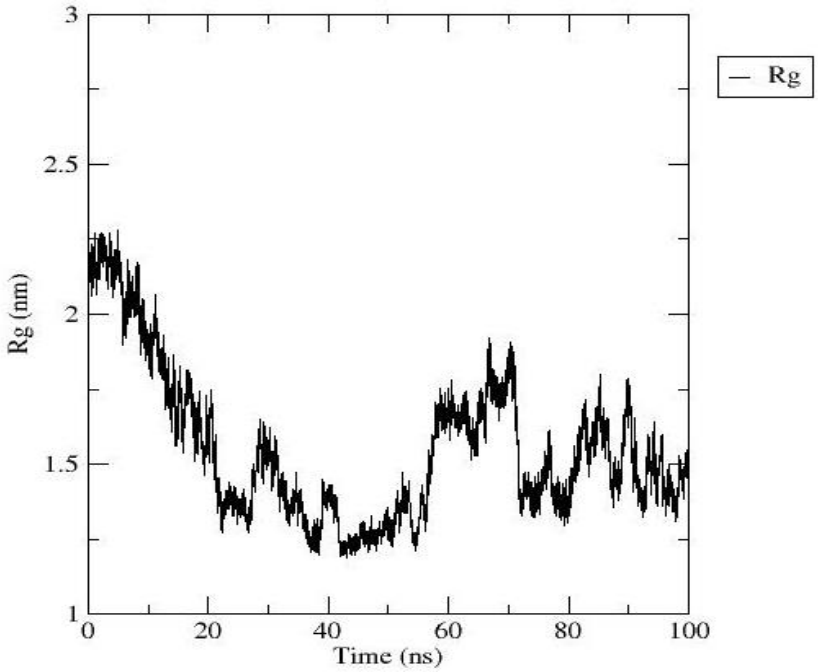

Figure 3.2.11: R_g (in nm) of apoA-1 in 0.096% chloride anion ionic liquid simulated over 100 ns

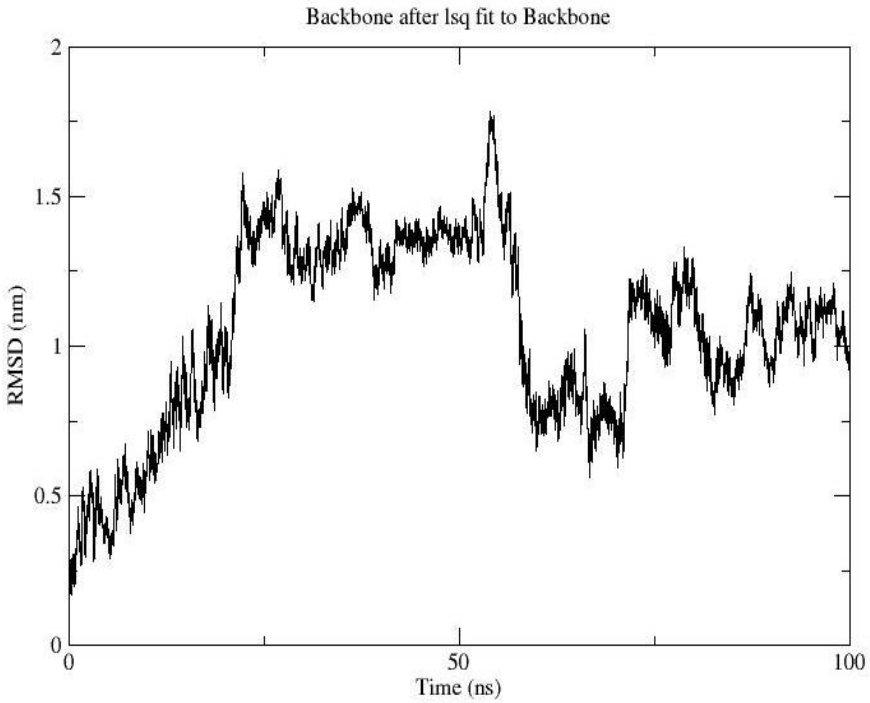

Figure 3.2.12: RMSD (in nm) of apoA-1 in 0.096% chloride anion ionic liquid simulated over 100 ns

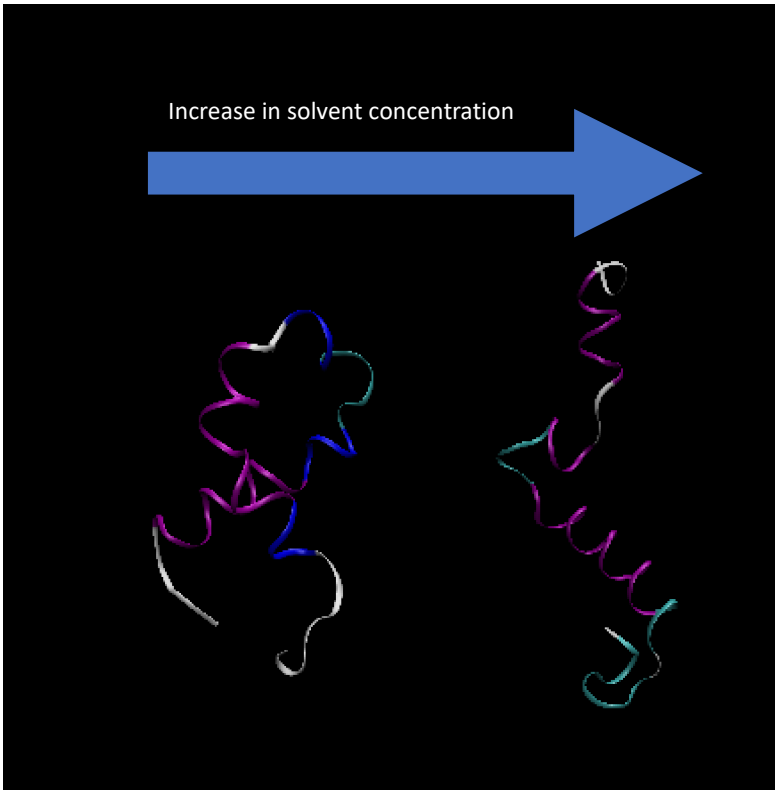

Figure 3.3.13: Final conformation of apoA-1 in chloride anion ionic liquids. The 0.024% solution is shown on the left and the 0.096% solution is shown on the right.

EMIM

In simulations in which EMIM-Cl served as the ion, there were very minimal changes in structure at 0.63% and 1.32% ions. The R_g and RMSD do not differ significantly from the water simulation and the lack of change is also confirmed visually. When we increased the concentration to 2.01% there were more obvious changes. In the final 40 ns, the R_g decreases and the protein shifts to a more folded state. This state is obviously unstable however as the graph fluctuates during this time and at 100 ns the shape even shifts back towards its original shape. Meaning that the visualization on the right, even though it is the most folded of the three, does not show the maximum folded structure. All of these simulations cause the α helixes to join which makes the protein less prone for folding in the center.

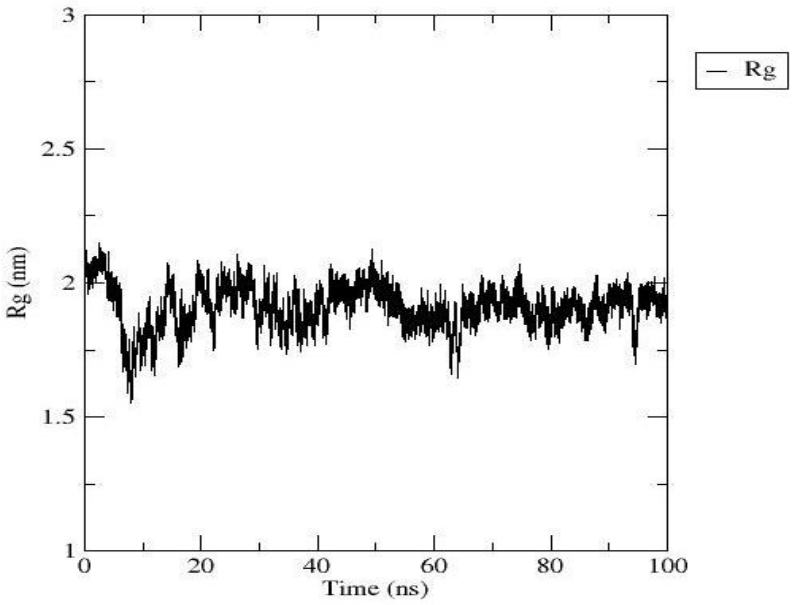

Figure 3.2.14: R_g (in nm) of apoA-1 in a 0.63% EMIM-Cl ionic liquid simulated over 100 ns

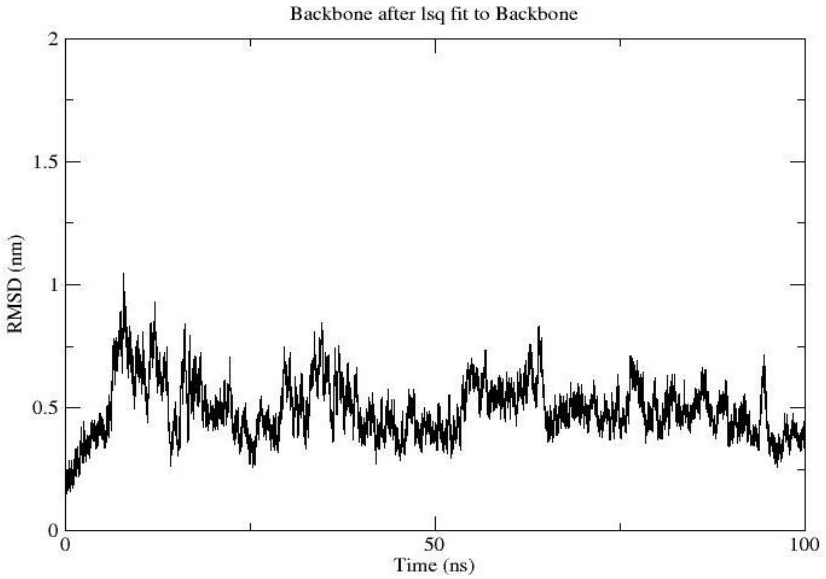

Figure 3.2.15: RMSD (in nm) of apoA-1 in a 0.63% EMIM-Cl ionic liquid simulated over 100 ns

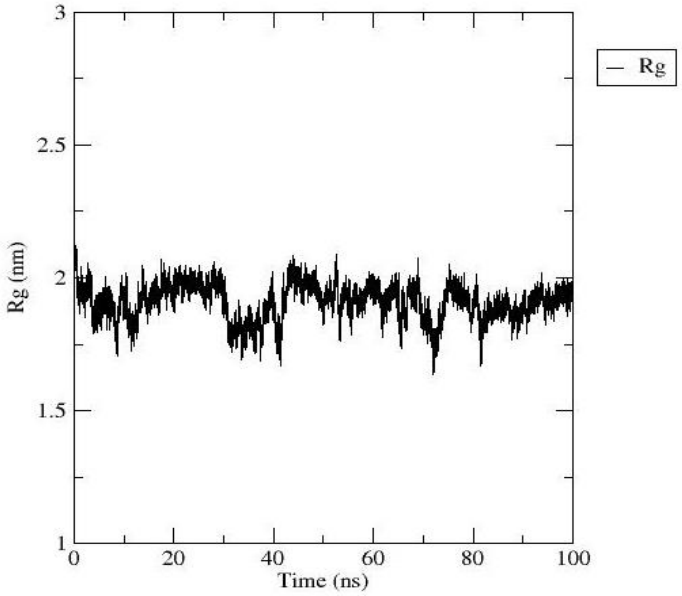

Figure 3.2.16: R_g (in nm) of apoA-1 in a 1.32% EMIM-Cl ionic liquid simulated over 100 ns

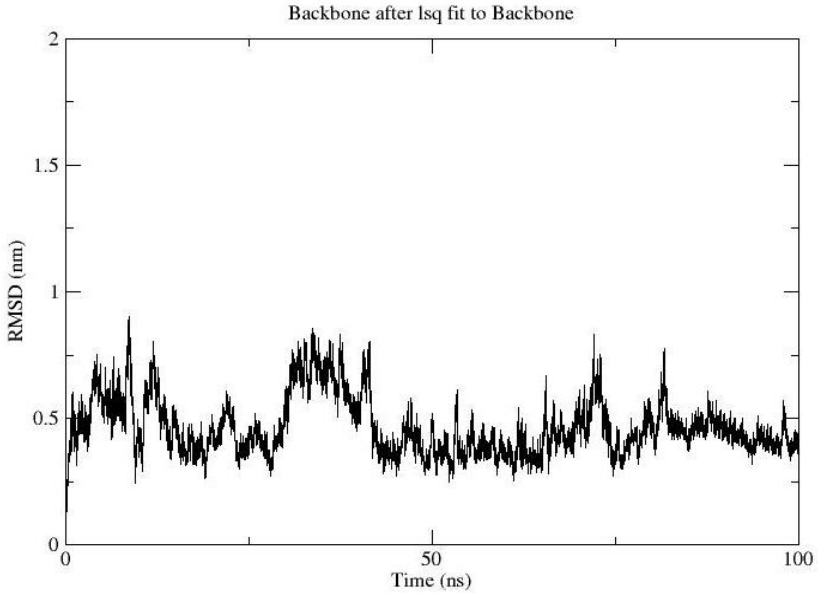

Figure 3.2.17: RMSD (in nm) of apoA-1 in a 1.32% EMIM-Cl ionic liquid simulated over 100 ns

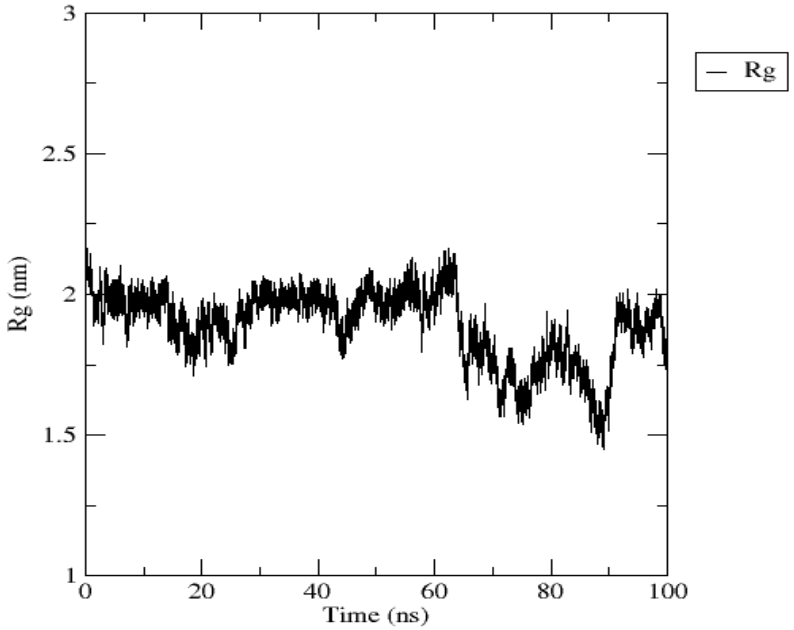

Figure 3.2.18: R_g (in nm) of apoA-1 in a 2.01% EMIM-Cl ionic liquid simulated over 100 ns

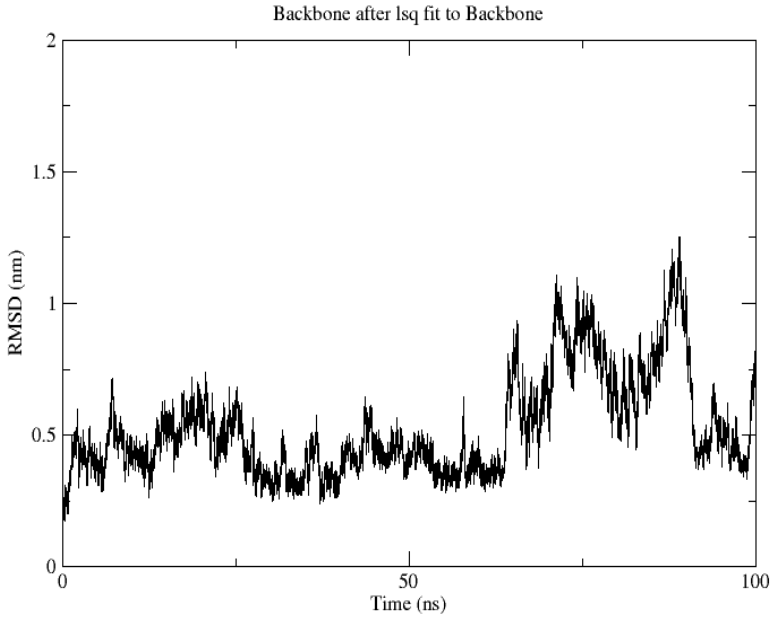

Figure 3.2.19: RMSD (in nm) of apoA-1 in a 2.01% EMIM-Cl ionic liquid simulated over 100 ns

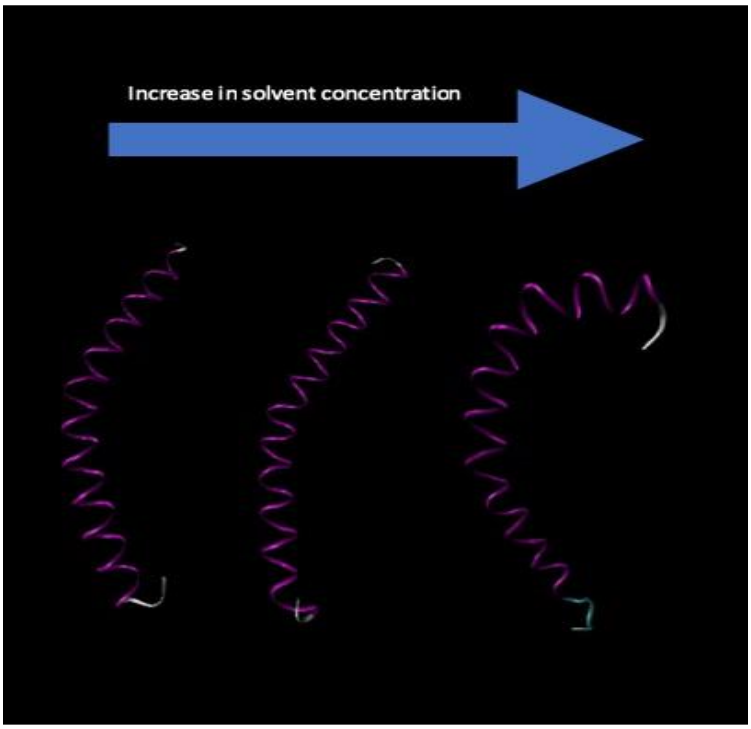

Figure 3.3.20: Final conformation of apoA-1 in EMIM-Cl ionic liquids. The 0.63% solution is shown on the left, the 1.32% solution in the center, and the 2.01% solution is shown on the right.

EMIM-Cl average over entire simulation

To further compare the data produced from the EMIM-Cl simulations, we analyzed the average values for both R_g and RMSD for each concentration. These charts will be able to clearly show how concentration changes these values. Our R_g shows what we were able to infer from the individual graphs, the highest concentration changes significantly more than the other two. The RMSD is actually fairly similar for each of the three concentrations. This indicates that none of the solutions are causing significant changes to the backbone of apoA-1.

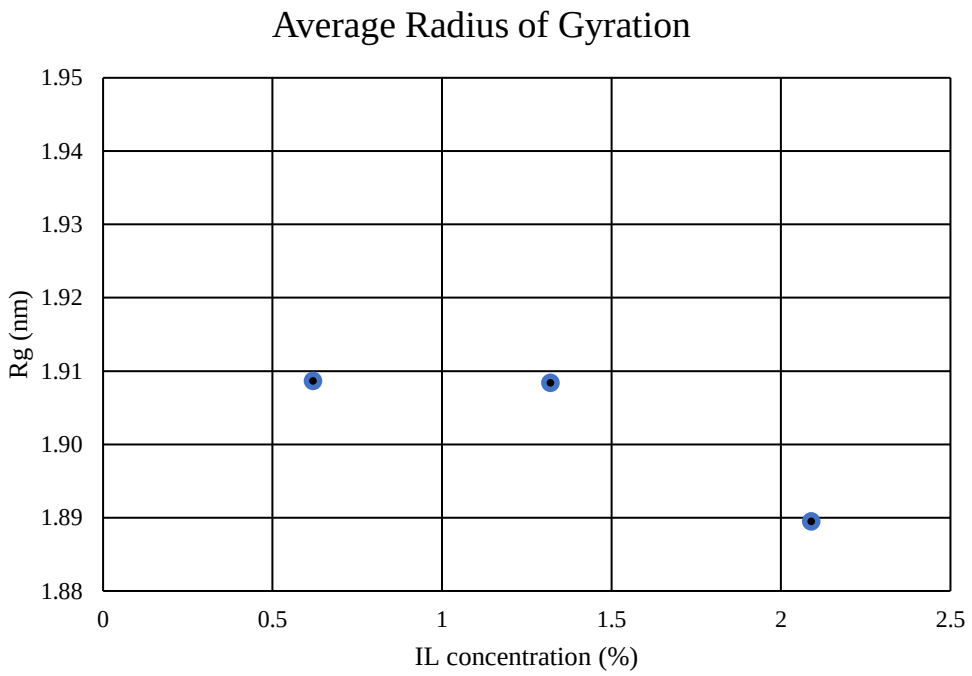

Figure 3.2.21: Average R_g (in nm) of apoA-1 in EMIM-Cl ionic liquids from 100 ns simulations at different ion concentrations.

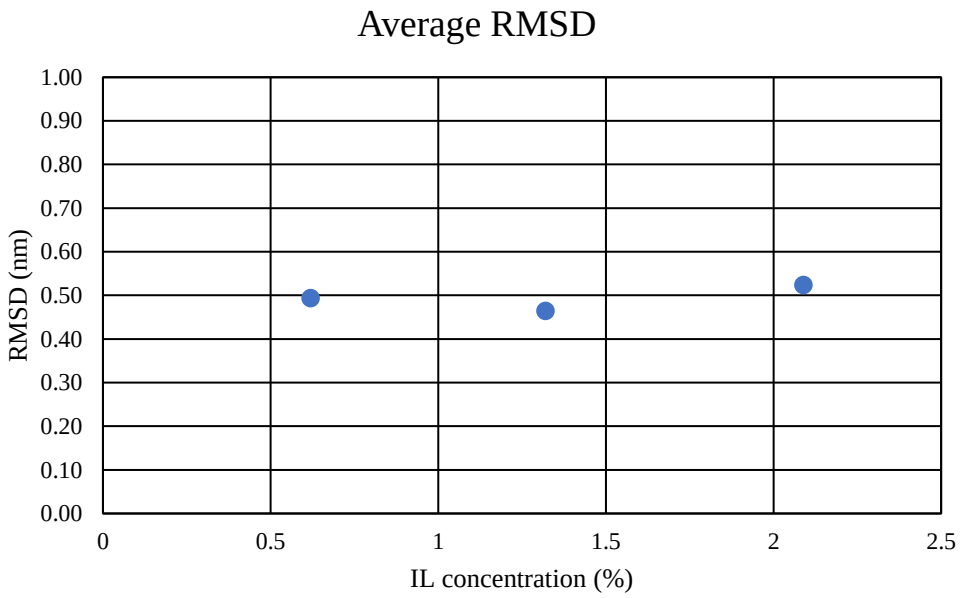

Figure 3.2.22: Average RMSD (in nm) of apoA-1 in EMIM-Cl ionic liquids from 100 ns simulations at different ion concentrations.

Chapter 4:

Discussion

4.1 Conclusions:

The main objectives of this work are to learn more about the folding of apoA-1 in ionic liquids and establish simulation method for implementing new molecules into the force field. The process for adding ions already in AMBER had already been well established and those simulations required following the known process. Our goal was to see how to change this process for new molecules. The methodology for simulations with EMIM demonstrates that our process is repeatable for similar solvents at similar concentrations. Our topology files as well as the VMD visualizations reveal that the properties of the ion were taken into account in trajectory calculations. However, we had difficulties with higher concentrations of EMIM in solution. These difficulties were caused by the magnitude of forces which created errors within GROMACS when trying to calculate trajectories. Frustratingly, these errors are not obvious and can occur at any point in simulation. This makes the refinement process for this process very time consuming. We believe these issues can be corrected by decreasing the size of the time step in MD production, but that will increase computation time. Currently, a system with 2.96% EMIM is running with a 0.001 fs step size for twice as many steps to produce the same amount of time.

The simulations where sodium served as the cation showed relatively small changes to the tertiary structure of apoA-1. The graphs for radius of gyration have very similar average values to the water simulation especially at 0.024% ions in solution. Visually, we can see that there is a slight turn in the chain in the region where the α helixes had been separated originally. Also, the α helix is now extended through the majority of the chain. Thus, we have more hydrogen bonding between amino acids in this region. So, for at least this area, the ionic liquid

helped to stabilize the structure. Aside from this region, our data does not reveal significant interactions with the solvent as the tertiary structure does not change much.

The chloride ionic solutions, which are negatively charged, show significant structural changes to the protein. Many different sections of the chain are folded and brought closer to other sections of the chain. This is true more so at the lower ion concentration. We can see this clearly in the radius of gyration graph where the shape becomes compact and stays in a folded state similar to the final conformation. For the 0.024% Cl solution, one of the α helixes are destroyed, and at 0.096% the middle of one is denatured while the sides retain their secondary shape. The higher concentration does reveal a turn between the two original helixes however, one was likely present where the R_g was lowest. While it is good to see that apoA-1 folds when in the presence of this cation, it also reveals that the protein becomes more unstable. That instability is more prevalent in the higher ion concentration.

Remember, previous work showed that increasing NaCl concentration in solution created an oscillating pattern of folded and unfolded states. Based on our work, it seems as though the chloride anion is largely responsible for causing the denaturing and destabilization of apoA-1. The chloride molecules are also the cause of folding in our simulations. Our work showed the folded and unfolded state over time in the simulation. We did not have concentrations nearly as high as Balderas's work with NaCl and did not see a concentration that specifically went to a folded or unfolded. That is likely due to the relatively low molarity of our solution. I would anticipate that trend to be similar for chloride solutions if we had similar concentrations to previous work. It is also important to understand that these solutions may not be physically realistic but rather serve to establish the relationship between charge of the solution and its effect

on the protein. The data from sodium and chloride simulations serve to guide future solution preparations that may be used.

EMIM-Cl ionic liquids are found to show properties seen in each of our simpler ions. Firstly, we can see through the average radius of gyration graph that as concentration increases, apoA-1 folds and becomes more compact. This is confirmed visually as well. The results also show that the α helixes are preserves and combined similarly to sodium simulations. This makes sense as both solutions had a net positive charge which indicates that solution charge causes interactions with these amino acids. At 2.09% there is also a turn now shown in the tail of the chain, indicating that this solution may be the most effective at promoting folding and that it can cause folds in regions other than between helixes. While the radius of gyration shows how compact the protein becomes, the root mean squared distribution reveals that the structural changes are not as drastic as seen with chloride. As the backbone of the protein structure does not change as significantly. Meaning that the radius of gyration decreases largely due to amino acid side chains. Thus, we can see this is a less extreme folding process rather than the randomness seen previously, but at 2.01% the graphs reveal the instability in the folding process. The data does follow similar folding trends and shape of EMIM in previous investigations. [10] Previous work showed EMIM solutions causing apoA-1 to move to unstable shapes. It is likely that if we continue to increase the concentration of EMIM-Cl we may get more folded structures, but they will be very unstable. Because of the increased time of the simulation, we are able to see that the protein shape changes after 60 nanoseconds. The EMIM-Cl solution is one which the charges were represented consistent with chemical constraints.

4.2 Future Work:

In continuing this research our next molecules of interest will be 1-butyl-3-methylimidazolium (BMIM) and 1-hexyl-3-methylimidazolium (HMIM). These two molecules are the other two similar to EMIM and were studied alongside it previously. [10] However, like EMIM, they have never been simulated over an entire 100 ns, nor have they been looked at for low concentrations. We hope to replicate the simulation we performed for EMIM for these other two molecules, by doing so we will learn more about the effect of hydrophobicity of ions in solution on apoA-1. This work will also confirm the method of implementing ions that are not already known by AMBER and simulate over longer time than done previously. Also, we want to continue creating liquids with higher EMIM concentrations to see if this solution continues its trend of increased folding or if it reverts to denaturing secondary structures similarly to chloride.

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. Insights from Molecular Dynamics Simulation Studies. *The Journal of Physical Chemistry B, 112*(9), 2566-2572. doi:10.1021/jp0766050

[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] Ibragimova, G. T., & Wade, R. C. (1998). Importance of Explicit Salt Ions for Protein Stability in Molecular Dynamics Simulation. *Biophysical Journal, 74*(6), 2906-2911. doi: https://doi.org/10.1016/S0006-3495(98)77997-4

[5] 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.

[6] LDL & HDL: Good & Bad Cholesterol. (n.d.). Retrieved August 1, 2019, from https://www.cdc.gov/cholesterol/ldl_hdl.htm

[7] 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

[8] APOA1 apolipoprotein A1 [Homo sapiens (human)] - Gene - NCBI. (n.d.). Retrieved August 1, 2019, from https://www.ncbi.nlm.nih.gov/gene/335

[9] Wang, & R.j. (n.d.). The helix-hinge-helix structural motif in human apolipoprotein A-I determined by NMR spectroscopy. Retrieved August 1, 2019, from https://www.rcsb.org/pdb/explore/remediatedSequence.do?structureId=1GW3

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

[11] 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

[12] Hatch, H. W., Stillinger, F. H., & Debenedetti, P. G. (2014). Computational Study of the Stability of the Miniprotein Trp-Cage, the GB1 β-Hairpin, and the AK16 Peptide, under Negative Pressure. *The Journal of Physical Chemistry B, 118*(28), 7761-7769. doi:10.1021/jp410651u

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

[14] M. Compiani and E. Capriotti, “Computational and Theoretical Methods for Protein Folding,” *Biochemistry*, vol. 52, no. 48, pp. 8601–8624, 2013.

[15] Shahzad, M. A. (2018). Free energy calculations of protein-water complexes with Gromacs. *bioRxiv*.

[16] R. H. Stote, “Introduction to Molecular Dynamics Simulations.” [Online]. Available: http://www.ccp4.ac.uk/maxinf/mm4mx/PDFs/MM4MX.MD.Stote.pdf. [Accessed: 16-Sep-2019].

[17] 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.

[18] Sorin & Pande (2005), AMBER FORCE FIELD PORTS FOR THE GROMACS MOLECULAR DYNAMICS SUITE, Biophysical Journal, 88, 2472-2493.

[19] 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

[20] 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

[21] Winger, M., & van Gunsteren Wilfred, F. (2008). Use of Molecular-Dynamics Simulation for Optimizing Protein Stability: Consensus-Designed Ankyrin Repeat Proteins. *Helvetica Chimica Acta*, *91*(9), 1605-1613. doi:10.1002/hlca.200890175

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

[23] 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

[24] 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.

[25] 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

Appendix 1:

Specific commands for sodium and chloride simulations in GROMACS. Work follows the methodology of a simulation of a lysozyme in water. [22]

Generate topology

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

Choose force field 2

```
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 sodium with (#) number of ions

```
gmx genion -s ions.tpr -o apoa_solv_ions.gro -p topol.top
-pname NA -np (#)
```

For chloride with (#) number of ions

```
gmx genion -s ions.tpr -o apoa_solv_ions.gro -p topol.top
-nname CL -nn (#)
```

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
```

Appendix 2:

Specific commands for simulations of apoA-1 in EMIM-Cl ionic liquids

Add ions to protein

```
gmx insert-molecules -f apoa.gro -ci EMI3.gro -o
apoaemi3.gro -nmol (#) -box 8.0
```

Create a “box” containing our system

```
gmx editconf -f apoaemi3.gro -o apoaedit.gro -c -d 1.0 -
bt cubic
```

Add water

```
gmx solvate -cp apoaedit.gro -cs spc216.gro -o
apoasolv.gro -box 8.0 8.0 8.0
```

Create topology

```
gmx pdb2gmx -f apoasolv.gro -o apoaprocessed.gro -
waterspce
```

Choose force field 2

```
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)
```

Energy minimization (mimim.mdp step size had to be decreased for higher concentrations)

```
gmx grompp -f minim.mdp -c apoaprocessed.gro -p topol.top
-o em.tpr -maxwarn 1
```

```
gmx mdrun -v -deffnm em
```

NVT equilibration

```
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o
nvt.tpr -maxwarn 1
```

```
gmx mdrun -v -deffnm nvt
```

NPT equilibration

```
gmx grompp -f npt.mdp -c em.gro -r em.gro -p topol.top -o
nvt.tpr -maxwarn 1
```

```
gmx mdrun -v -deffnm npt
```

MD production

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -
o md_0_1.tpr -maxwarn 1
```

```
gmx mdrun -v -deffnm md_0_1
```

Data analysis was done identically to the process shown in Appendix 1.