5-2021

The effect of the apolipoprotein A1 (APOA1): the stability and folding in potassium chloride environment

Alexandra Paladian
University of Tennessee at Chattanooga, hgb143@mocs.utc.edu

The Effect of the Apolipoprotein A1 (APOA1):

The Stability and Folding in Potassium Chloride Environment

Alexandra Paladian

April 2021

Departmental Honors Thesis

University of Tennessee at Chattanooga

Department of Biology

Dr. Luis E. Sanchez Diaz

Project Director

Dr. Margaret Kovach

Departmental Examiner

Dr. Fernado Alda

Departmental Examiner

Abstract

Healthy levels of potassium chloride (KCl) can significantly affect the workings of the cholesterol level of the human body and how they pertain to an individual person. The search for a better salt additive for the human diet can provide a better option for people who experience high cholesterol levels and heart disease. The study focuses on the experimental design of the Molecular Dynamic (MD) simulation of the Apolipoprotein A1 (APOA1) in the potassium ion solution environment to determine the stability and folding of the protein. The study also compares its data to the previous experimental design of chloride ions to determine the stability and folding of APOA1. The simulations revealed that the potassium ion solution environment makes APOA1 stable, but it does not fold, unlike the chloride ion solution system that makes the protein stable and fold. The hypothesis of the research was proven wrong, due to the initial thought that the APOA1 would fold in the experimental conditions of potassium ions solution. However, there might be more to the behavior of the APOA1 if the potassium and chloride ions were together in the solution for the future experimental designs to help understand how the proper ions mix can still stimulate protein folding along with the stability to produce healthy cholesterol levels in the human body.

Table of Contents

Abstract……...……....2
Chapter 1: Introduction……4
1.1: Background……...5
1.2: Selection of the Apolipoprotein A1 (APOA1)……6
1.3: Molecular Dynamic Simulation and GROMACS……..……..7
1.4: Protein Folding……...…..8
1.5: Previous Work……12
1.6: Objectives…….……..13
Chapter 2: Materials and Methods……15
2.1: Apolipoprotein A1 Structure……..……16
2.2: GROMACS…….……...……17
2.3: Solvents……...……...…19
2.4: Simulation Details……...…….……..20
2.5: Data Analysis Techniques…….…….23
Chapter 3: Results……...……28
3.1: Ionic Solvent Concentrations……...……..29
3.2: Simulation Data of K Environment……....29
Chapter 4: Discussion and Conclusion……..36
4.1: Discussion and Conclusion……...……..…37
4.2: Future Work……...…42
References……...…..……44
Appendix 1……...……47

Chapter 1:

Introduction

1.1 Background:

Healthy levels of potassium chloride (KCl) can significantly affect the workings of the cholesterol level of the human body and how they pertain to an individual person. The search for a better salt additive for the human diet can provide a better option for people who experience high cholesterol levels and heart disease. Moreover, the type and the amount of salt can dictate the levels of the blood cholesterol and how the neutral salt in the physiological pH interacts with the protein that is responsible for the control of the cholesterol levels (Metcalf et al., 1990).

Potassium chloride is a suitable candidate for the solution because potassium (K^+) maintains intracellular tonicity, the requirement for nerve conduction, cardiac, skeletal, and smooth muscle contraction, production of energy, the synthesis of nucleic acids, and maintenance of blood pressure and cholesterol ("Potassium chloride," 2020). The substance is a metal halide that has antihypertensive effects when taken as a nutritional supplement. One of its properties lies with the potassium component. The potassium portion of the salt is intracellular cation with a concentration of 140-150 mmol/l in the human cells and 3.5-5 mmol/l in the extracellular fluid of the cells (Zacchia et al., 2016).

Paired with Apolipoprotein A1 (APOA1), the molecule that transports cholesterol through the bloodstream ("APOA1 gene", 2020), there is a correlation for testing how this salt and protein interact with each other. The function of the protein in the human system pertains its composition of a high-density lipoprotein that is responsible for the lipid metabolism in the blood stream. Moreover, the protein is a key feature in the fatty acid efflux that transports cholesterol from the human tissues to the liver to be filtered out of the body (Balderas Altamirano et al., 2015). The

aim is to discover the effects of the potassium chloride solution on the stability and folding of APOA1 protein, the relationship between this interaction, and how it corresponds to the human system concerning the levels of cholesterol in the bloodstream.

1.2 Selection of the Apolipoprotein A1 (APOA1):

The properties of the Apolipoprotein A1 (APOA1) make it a desirable candidate for the continuation of the work previously done on the conformation and stability of the protein in the ionic solutions (Smith, 2019). Due to the property of the protein being a high-density lipoprotein, the subject might yield potential results towards the levels of cholesterol and how it affects the good cholesterol in the individuals that potentially use potassium chloride over their dietary limit. The overuse of the substance can also cause various cardiovascular health issues, including heart disease (Ansell et al., 2015).

Previous research on this protein involved computerized simulations of the organic molecule in ionic solutions and their effect shown on the concentration (Ansell et al., 2005). Folding of the proteins dramatically depends on their environment, precisely on the solvent factor. The hydrophobic effect of the water drives protein folding and is a suitable solvent for unfolded states, which might be beneficial for the supplement of misfolding (Shahzad, 2018).

Moreover, the factor of acidity in hypochlorous acid can be potent to the DNA in amino acids of the protein and the influence of its folding process, thus determining the potential of apoptosis cell death through the human coronary artery endothelial cells (Shahzad, 2018).

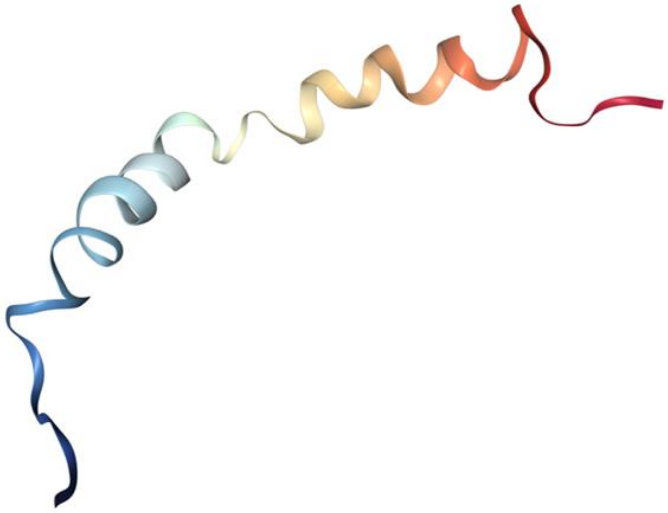

Figure 1: The native structure of APOA1 from Protein Data Bank (Smith, 2019). The task is to determine how the protein will change its shape and folding confirmation depending on the concentration of the experimental environment system by using molecular dynamics in GROMACS.

1.3 Molecular Dynamic Simulation and GROMACS:

Molecular Dynamic (MD) simulations and thermodynamic analysis show the protein folding or unfolding in the specified concentration, which is essential in the proper function of the molecule and how it can affect the human body (Abraham et al., 2015). The multi-step process of protein folding provides a comprehensive picture of the relationship and the interaction of the molecule with its environment and how it can potentially affect the health of a human body. The MD is often used in biophysical research, as it is a preferable way to gain data on the subject in a quicker manner (Smith, 2019).

Ultimately, computer simulations are the best way to test the folding properties of the protein, analyze data, and compare it with various outside factors. The data will be further applied to the inner workings of the proteins concerning the levels of cholesterol (Smith, 2019). The

technological limitations still exist in this scientific area concerning the accuracy of the data presented, but it can be monitored through the software coding system and simulations using GROMACS (Groningen Machine for Chemical Simulations) (Abraham et al., 2015; Smith, 2019). Also, the frequency of the simulations run is used as the way to solidify the results of the previous data computations, and the data can be compared to the findings in similar research.

The Protein Data Bank allows the download of the APOA1 in its native conformation from NMR spectroscopy before a simulation can begin. The properties of the protein are known and understood by the program, such as mass, charge, and amino acids (Smith, 2019). The simulation can begin whenever the desired properties of volume, molarity, heat, pressure, and potential are established in the system. A detailed description of the command lines is described in **Appendix 1**.

The steps needed to accomplish the simulation require processing time and computers that would run the data quickly. The computing power of a 24-core processor of the laboratory computer for the simulation run alleviates the long run times of the simulation, which allows to derive the data sooner and to confirm the experiment before starting the next stages of the research (Smith, 2019).

1.4 Protein Folding:

The APOA1 folding greatly depends on its stable environment; thus, the function of the protein will not be achieved if the apolipoprotein is not correctly established in the system. The solvent for the protein used is water, which is argued, by Clark et al. (2020), that the unfolded states of

proteins are expanded in water, and that suggests that water is a good solvent. Also, the folding and unfolding depend on the hydrophobic effect that promotes the action, thus makes water the neutral, good solvent that prevents misfolding and non-specific interactions (Clark et al., 2020).

The folding and unfolding of the protein include a complex process that requires high-resolution characterization by molecular dynamics simulations that are used to analyze the conformational changes in the protein and how the solvent can be responsible for the folding and unfolding of the protein (Ferina & Daggett, 2019). APOA1 protein is a complex structure because it consists of positive, negative, and neutral sequences in its structure, along with residues of hydrophilic and hydrophobic properties (Balderas Altamirano, 2015). The meaning of these diverse enigmatic properties can be beneficial in the protein folding in different charged environments.

A stable protein structure depends on the types of ions that will keep it in homeostasis and in a preferable conformation (Ibragimova & Wade, 1998). Some kinetics of breaking a salt-bridge of Arg-Glu can be observed as a potential barrier to protein folding and unfolding that can be analyzed by using molecular dynamics simulations (Gruia et al., 2004). Moreover, the gene that is responsible for the folding can be influenced by the environmental system and cause the folding and unfolding of the protein.

The gene that is responsible for the APOA1 folding, the APOA1 gene, contains the information for the creation of this protein (APOA1 gene, 2020). The Apolipoprotein A1 is a part of the high-density lipoprotein (HDL). HDL is responsible for the transfer of cholesterol and phospholipids from the human tissues to the liver (APOA1 gene, 2020).

The transport occurs through the bloodstream, and the liver is the point where the cholesterol and phospholipids either get removed from the body through waste or get moved to a different location (APOA1 gene, 2020). This process is essential because HDL is known as the good cholesterol of the body and its high levels reduce the potential occurrence of diseases such as heart and cardiovascular disease (APOA1 gene, 2020).

APOA1 protein interacts with the cell membranes, which moves the cholesterol and phospholipids from the inner to the outer part of the cell. Cholesterol and phospholipids interact with APOA1 protein to create HDL. The protein also converts the cholesterol to the type of substance that can be linked into HDL and moved through the bloodstream, which is the process of cholesterol esterification (APOA1 gene, 2020).

The proper functioning protein will fold in the right environment to produce these steps towards increasing HDL cholesterol into the cells and human body. The importance of having healthy cholesterol levels and conditioned cardiovascular health depends on the normal function of the protein. However, familial HDL deficiency and disorders are common when there is a mutation or unfavorable environment that can cause the APOA1 protein to not properly fold and function (APOA1 gene, 2020).

The mutation in the APOA1 gene is responsible for familial HDL deficiency, which is a condition that can be inherited by having low levels of HDL due to the gene mutation and an improper fold of the protein (APOA1 gene, 2020). Early cases of cardiovascular disease are

possible due to the APOA1 protein not able to fully stimulate HDL cholesterol and phospholipids into the cells and human body. Moreover, there is a case when the protein will not fold to provide cholesterol esterification, which will not let cholesterol be linked into HDL particles, and there will be less or no traffic of HDL through the bloodstream (APOA1 gene, 2020).

Other mutations in the APOA1 gene correlate with the occurrence of familial visceral amyloidosis, which is the abnormal development of proteins inside the human body and especially in the internal organs (APOA1 gene, 2020). Also, the study by Doss et al. (2012) argues that the substitutions in the amino acid, specifically in Fibroblast Growth Factor Receptor 1 (FGFR1) in some proteins, can have impactful genetic disorders due to the failure of the protein to be able to compute a fold based on the amino acid change. The way to detect the outcome of the protein in a timely manner, the molecular dynamics simulations can be used to understand if either a mutation or environmental agent caused its new behavior.

As the mutation can affect the folding of the APOA1 protein, these faulty proteins can stack up affect organs through their amyloid deposits and disturb the liver, kidneys, and/or heart. While the condition can present either mild or no symptoms, some of the affected people can develop hepatomegaly (enlarged liver), chronic kidney disease, and/or cardiomyopathy (heart disease) (APOA1 gene, 2020). Tang et al. (2019) also mention the effects on the cellular level by describing the effect that the unfolded protein can have on the endothelial cells in the human coronary artery and their potential apoptosis due to the incorrect layout of the protein, as mentioned in section 1.2.

1.5 Previous Work:

The study conducted by Balderas Altamirano et al. (2015) used the dynamics simulations to research the folding of the APOA1 and how it affects cholesterol. To figure out the data presented, they calculated the radius of gyration (Rg) and root mean square deviation (RMSD). Their report includes that there are two configurations that the protein takes on in the environment, which are the extended structure and the looped structure. The research argues for the looped structure to be the key factor of the protein function and the associated with the cholesterol and based on the hydrophilic and hydrophobic factors of the protein face and its interaction with the environment.

The research with APOA1 protein has been studied before in various simulated environments. The molecular dynamics (MD) expanded the work in protein behavior by simulating the accurate setup that has not been available in previous experiments before the computerization (Abraham et al., 2015). By choosing molecular configuration, analyzing atomic interactions, running simulations, and deriving the data from the MD run, the software provides customization in the experimental methods. The software applications that can be used include GROMACS, AMBER, NAMD, CHARMM, LAMMPS, and Desmond (Abraham et al., 2015).

One of the previous works mentions using CHARMM27 as a popular force field for biomolecular simulation package in GROMACS. CHARMM was compared to traditional TIP3P that concerns the effects of both correction maps and the effect of Lennard-Jones interactions on hydrogens (Bjelkmar et al., 2010).

The protein-specific study focused on the molecular dynamics simulation of APOA1 in ionic liquids, specifically the liquids in sodium, chloride, and 1-ethyl-3-methylimidazolium (EMIM), to analyze how the protein behaves in those solutions (Smith, 2019). Following the style of the study, the experiment will focus on the molecular dynamics in simulation of APOA1 in ionic liquids of potassium and how the protein will behave in this custom ionic environment. The study of using the chloride ions by Smith (2019) is used to compare the environment with the potassium ions in this experiment to determine the behavior of APOA1.

1.6 Objectives:

The research and simulation experiment will serve to further explain and compare previous studies on the effect of different ionic solutions on the conformation and stability of the Apolipoprotein A1 (APOA1). Through Molecular Dynamic Simulation and GROMACS, there will be new data regarding the interaction of the protein in the environment with potassium ions and how these findings can correspond to the human system concerning the levels of cholesterol.

The results will be compared to the study done by Smith (2019) to understand the similarity and difference in the behavior of APOA1, specifically in the part of the ionic compound of chloride, which was previously studied. The potassium factor is a new component in this experiment, and it will be heavily considered for review during the results and discussion sections. The goal is to further understand the interaction of APOA1 in the potassium solution and refer to the studies to seek an explanation for the potential behavior of the protein in the KCl solution. Based on the

information, the hypothesis of the research is that the protein might have an active interaction in the potassium ion solution on its stability and folding in the simulation system.

Chapter 2:

Materials & Methods

2.1 Apolipoprotein A1 Structure:

The structure of APOA1 is shown below (Figure 2). The fragment of the apolipoprotein consists of bends represented by a light green line, turns represented by a dark green curved line, 3/10-helix represented by a pink wavy line, and alpha-helix represented by red wavy line (Smith, 2019).

The human APOA1 has the conformation of the helix-hinge-helix structural motif. The composition of the amino acids is twenty-two hydrophobic and seven hydrophilic configurations (Smith, 2019). The charges of the amino acids include ten positive and six negative charged entities. There are two alpha helixes at the amino acid positions of 14-21 and 27-35 (Smith, 2019). Those regions are limited in movement due to being stabilization of hydrogen bonds in the secondary structure of the protein, and other sections can be prone to more structural influence (Smith, 2019).

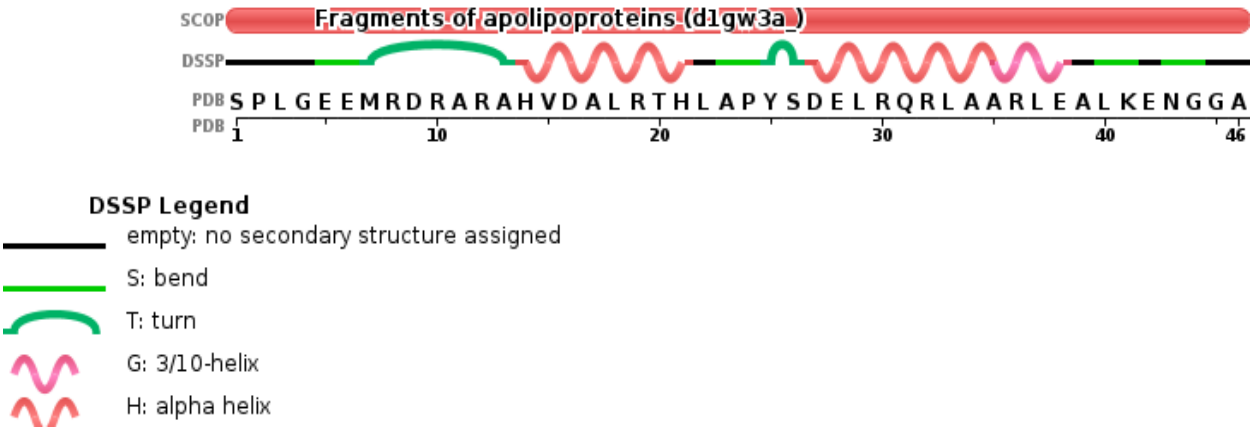

Figure 2: Amino acid sequence and secondary structure of APOA1 (Smith, 2019).

2.2 GROMACS:

The objective of the GROMACS, the MD package software, is to simulate the environment and test the outcome of the APOA1 behavior in the high halite concentration. There were preparatory steps taken before the start of the simulations. The setup included loading data, solvation, the addition of water, and the appropriate ion structure. Then, the protein needs to be programmed in, and both the solvent and the protein will be equilibrated before the production simulation and the outcome of the trajectory (Abraham et al., 2015). The trajectory, acting as a binary file, records the atomic coordinates at multiple time steps and shows the dynamic motion of the molecule, which will demonstrate how the protein interacted with the halite and if it still has a proper folding for its normal function (Abraham et al., 2015).

Along with the setup, GROMACS is important in the calculation of every particle in the system that is in motion. Simulation analysis allows for the calculation of these equations in a more timely and efficient manner than manually. The function of time, r(t), is the trajectories given for the simulation. Also, bonding, and non-bonding interactions in the protein can be calculated through these equations. Water implementation and ionic interactions in the system are other factors to consider that would require making changes in these equations (Smith, 2019).

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

Figure 3: The Newtonian equations of motion for N atoms (Smith, 2019).

To achieve the proper environment in the simulation, there are specific commands need to be followed regarding the selection of the force field. The establishment of potential energies between molecules and any application of the restraints for the system drives the main functions of the force field type selection and implementation in the environment (Smith, 2019). OPLS-AA/L all-atom force field was selected in GROMACS to continue the simulation (Lemkul, 2018).

Equation 2.2.3:

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

Figure 4: The calculation of the energy in the interaction for all atoms in the system. The spring potential represents the first three sections of the calculation, and the last section represents Leonard Jones force calculation from Pauli repulsion and Van Der Waal interaction (Smith, 2019).

Moreover, a detailed setup is required for the simulation to run successfully, and it uses the PDB (Protein Data Bank) input to create three files for the MD simulation. A topology of the APOA1 protein will be prepared, which includes the information needed to describe the molecule for the simulation: atomic mass, bond length, angles, and charges. Also, a force field and water model must be present for the topology calculation. It is recommended to use the OPLS/AA force field and SPC/E water model for the computer run (Bray, 2020). The PDB protein structure will be converted into a GRO file, its structure centered in a simulation box.

The additional steps in the programming will include solvation using GROMACS and adding ions tool. The coding will ensure that the water molecules are added to the structure, and

topology files fill the unit cell with the addition of potassium ions. Moreover, the structure must be relaxed by running an energy minimization (E.M.) algorithm to remove any steric clashes. NVT equilibration will be performed to stabilize the environment around APOA1 protein and to hold it in place in the system. Only after those measures are considered, the position restraints can be released, and a production MD simulation will be run to attain data for analysis.

2.3 Solvents:

The selection of solvent in the environment system of the APOA1 in GROMACS depended on the type of simulation needed to observe the interaction between the protein and potassium chloride environment. Water was selected as a good solvent due to the high percentage of water found in the human body and its bloodstream, which is a way to mimic the environment in the simulation (Smith, 2019). Also, this medium is shown that it reduces misfolding and non-specific interaction of the protein (Clark et al., 2020).

Since the protein folding thrives based on the hydrophobic effect, adding water as the solvent should provide the motivation for the APOA1 to behave naturally when in contact with water. The hydrophobic surface area decreases on the outside of the protein when in water because the apolar side chains in the soluble protein bury themselves in the interior portion of the structure (Clark et al., 2020).

The burial of the hydrophobic area of the protein increases the solvent (water) entropy because the water molecules do not need to adapt to the hydrophobic outer surface of the protein because those parts are folded inside the protein (Clark et al., 2020). This allows for the best environment

to experiment with to figure out how the addition of potassium chloride ions can influence the solvent-protein relationship and the behavior of the folding and unfolding of the protein.

2.4 Simulation Details:

To create the environment for testing the APOA1 protein in the potassium solution, a series of simulations were conducted to provide findings to the research. The GROMACS Tutorial, *Lysozyme in Water,* was used as a guide to code in the commands and run the simulation data with a minor modification in the addition of the ions stage (Lemkul, 2018). The following will discuss the steps taken to achieve the proper K^+ environment for the APOA1 protein. The detailed codes and force field selections of the program needed to achieve the desired simulation environment can be found in **Appendix 1**.

Before the Ubuntu software can be used to start the simulation with the *Lysozyme in Water* tutorial, the APOA1 was downloaded from the Protein Data Bank on the RCSB website. The structure can be assessed visually through the VMD, Chimera, or PyMOL program. This determines whether the molecule needs to be stripped out of the crystal waters (Lemkul, 2018). In the case of the APOA1, the entity was already clean of crystal water and ready for the preparation of its topology.

The preparation of the topology for the molecule also requires checking the .pdb file for any components, such as atoms or whole residues, that might be missing, which can interfere with the dynamics process (Lemkul, 2018). Once the PDB file contains only protein atoms, then the next step in the tutorial will include preparing the pdb2gmx module. The purpose of this step is

to establish the molecular topology (topol.top file). The entity will contain the necessary information for the definition of molecule in the simulation (Lemkul, 2018).

When the protein structure is prepared by the pdb2gmx file, then the software program prompts several choices of force fields to select. The selection of the force field is a critical choice that can determine the direction of the simulation being conducted (Lemkul, 2018). After completion of these steps, there are three new files added to the structure file of the protein via GROMACS-formation with the force field that has all the atoms defined inside (Lemkul, 2018).

The next step includes the examination of the topology of the protein. The interpretation of the information will include the atom number, atom type and name, amino acid residue number and name, the charge and the charge group number, and mass (Lemkul, 2018). Following the examination, the GROMACS needed to define the box dimensions of the APOA1 and fill the box with water using the solvate module (Lemkul, 2018). This means that the topology of the protein does not change if any other non-water solvent is added to the environment.

The addition of the ions within GROMACS followed the previous step, which is usually called genion (Lemkul, 2018). The stimulation ran the environment on the potassium ions environment and will compare the chloride environment that has been derived from the previous research in the discussion and conclusion chapter.

The protein structure was then relaxed through the next process of energy minimization. The step is similar to the addition of the ions, but this ensured that there is no steric clashes or improper

geometry in the subject of the simulation (Lemkul, 2018). Both steps were used to assemble the structure, topology, and simulation parameters of the APOA1, but energy minimization ran this process through a GROMACS MD engine called mdrun. The following steps used the mdrun command to gather data graphs for the analysis technique.

2.5 Data Analysis Techniques:

Once the GROMACS reaches the Energy Minimization step, then the process of mdrun is used to analyze the genion and to produce a data graph of potential from the energy minimization. The data graph is presented as the Energy Minimization Step vs. Potential Energy in kJ mol^-1 (Lemkul, 2018).

Potential

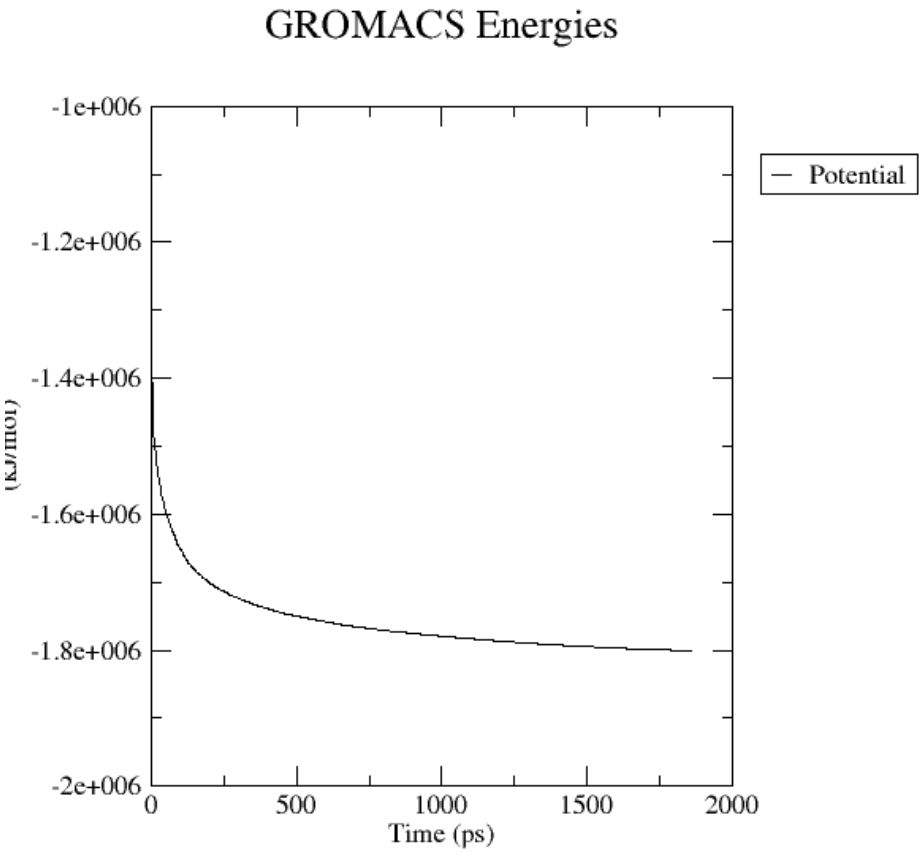

Figure 5.1: The data graph is presented as the Energy Minimization Step "Time in ps vs. Potential Energy in kJ mol^-1." The total time run is 2000 ps.

The following step ensured the equilibration of the solvent and ions around the protein, which is called the process of NVT Equilibration (Lemkul, 2018). This will establish the proper temperature, based on kinetic energies, for the protein to be used in the simulation for the interaction with the solvent. The mdrun will establish a file for another manifestation of data. The data graph is presented as the Time in ps vs. Temperature in Kelvin (Lemkul, 2018).

Temperature

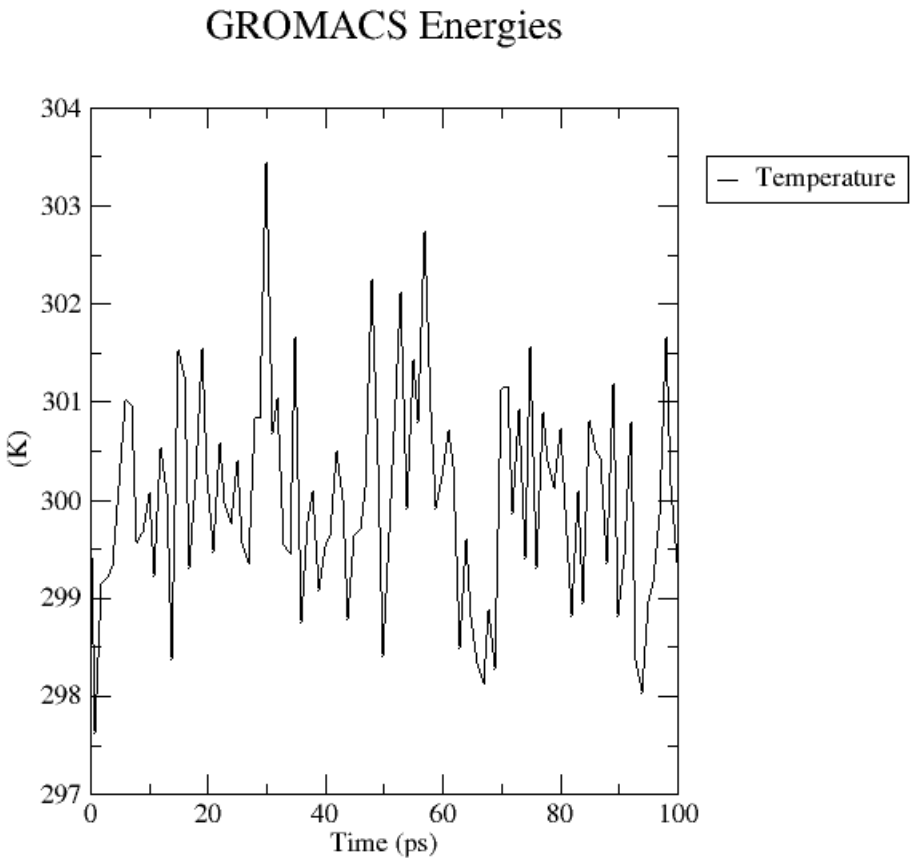

Figure 5.2: The data graph is presented as the Time in ps vs. Temperature in Kelvin. The total run time is 100 ps.

Pressure in the system needed to be stabilized next for the proper density of the protein, which is also called the “isothermal-isobaric” ensemble. The process is similar to the establishment of the right temperature in the system. This continuation from the NVT Equilibration ends with the other mdrun that will produce data for pressure (Lemkul, 2018). The stability of the protein also depends on the type of pressure exerted in the system, which makes the calibration of pressure important (Hatch et al., 2014). The data graphs are presented as Time in ps vs. Pressure in bars.

Pressure

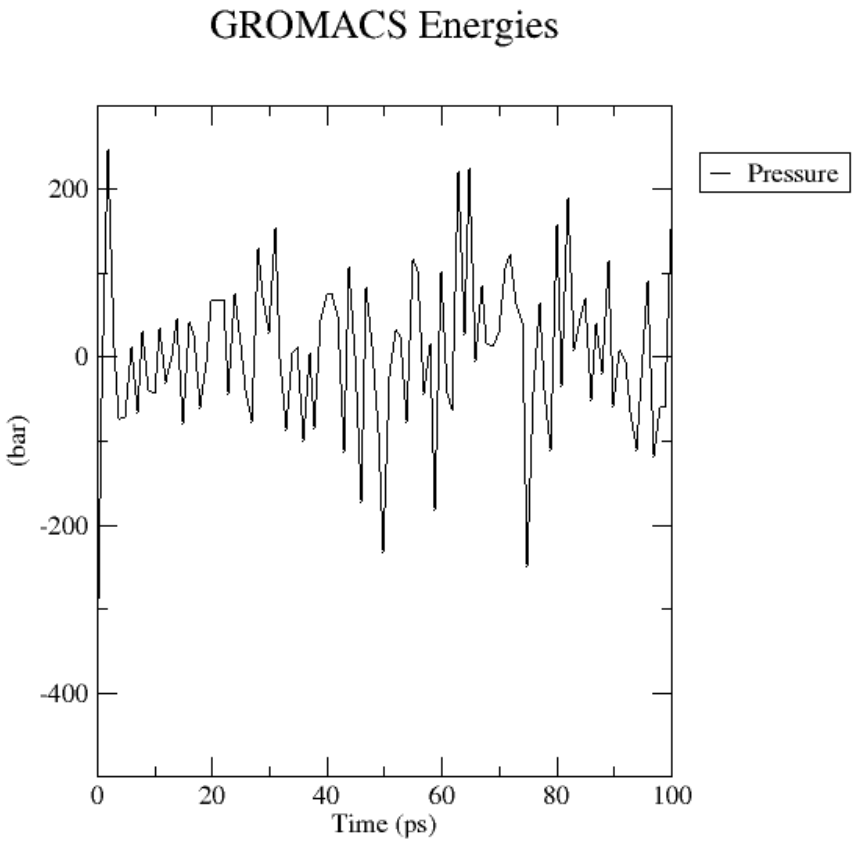

Figure 5.3: The data graphs are presented as Time in ps vs. Pressure in bars. The total time run is 100 ps.

The completion of the two equilibration stages ensured that the system is at the desired temperature and pressure for the next stage that will release the position restraints for the MD run production and data collection (Lemkul, 2018). Like the previous steps, the Production MD creates a specific file for data retrieval and analysis.

Along with the previous data graphs, the radius of the gyration graph is presented to analyze the compactness of the protein, which will explain if the APOA1 stayed relatively folded or unfolded based on the steady value of gyration (Rg). The data graph is presented as Time in ns vs. Rg in nm. in the results chapter (Lemkul, 2018). The calculation first needs the value for the center of the mass equation shown below, which is for all the atoms within the backbone of the protein. The purpose of this is to know that the radius has a zero value to the center mass of the protein backbone. This allows the m value to be a mass of a single atom and for gyration to be calculated in real space (Smith, 2019).

Equation 2.5.1:

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

Figure 6.1: The center of mass equation (R(c)) (Smith, 2019).

Equation 2.5.2:

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

Figure 6.2: The radius of gyration equation for the average distance of the center of mass in the backbone of the protein, which includes each atom (Smith, 2019).

The RMSD, Root Mean Squared Distribution, a graph is derived to analyze if the protein structure was stable in the given environment. The data graph is presented as Time in ns vs. RMSD in nm. in the results chapter (Lemkul, 2018). The process is similar to the calculation of gyration because it also can be put into the GROMACS for the computerized data output. Both calculations and graphs are needed to get a complete picture of the effect of the protein in the simulated environment (Smith, 2019).

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 $t_2 = t_1 + \Delta t$

Figure 7: The RMSD equation for the analysis of the topology from the simulations. Just as gyration, it is used to study the backbone of the protein but also shows how the protein changes from its original positioning (Smith, 2019).

Chapter 3:

Results

3.1 Ionic Solvent Concentrations:

As the water and ions were inserted into the environment around the protein, following the creation of the box around the protein, the topology file of the APOA1 is used to determine the ion concentration in the system environment. Due to the slight variation of the amount of water molecules in the simulation, the ion concentration number is different based on the amount and type of ions present.

Ions of K^+	Water Molecules
4	33,339
8	33,323
16	33,307

Figure 8: The amount of water molecules that are present in solution with the ion count based on the topol.top file derived from the MD simulation of K^+ ions. The gathered data represents an inverse relationship between the ions of potassium and the water molecule count.

3.2 Simulation Data of K Environment:

The RMSD graph is derived to analyze if the protein structure was stable in the given environment. Also, the radius of the gyration graph is presented to analyze the compactness of the protein, which will explain if the APOA1 stayed relatively folded or unfolded based on the steady value of gyration (Rg). Each calculation will be based on the environment of 4, 8, or 16 ions of potassium molecules. One type of the graphs is not enough to figure out the state of the APOA1, but both of the graphs can confirm with each other the folding and the stability of the

protein. The following graphs are going to be compared in the discussion and conclusion chapter to the study done by Smith in 2019.

Gyration (R_g) of APOA1 in Potassium 4 Ions System

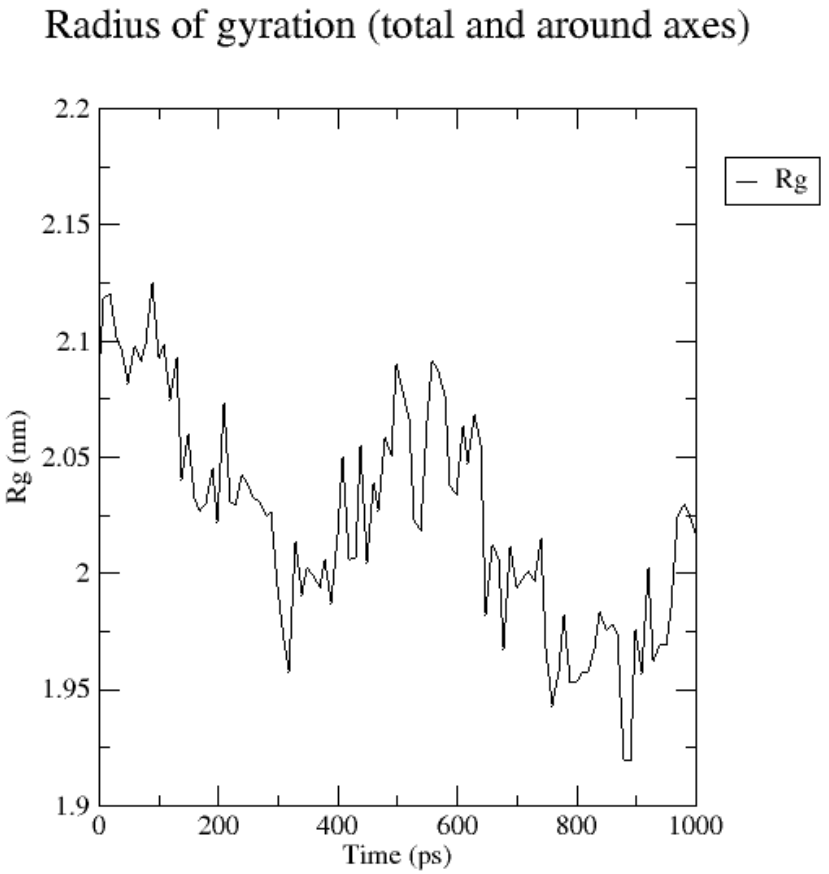

Figure 9.1: R_g (in nm) of APOA1 in potassium 4 ions environment simulated over 1000 ps. The protein is shown not to be compact in the graph because the values are not constant and above the values that are needed for the protein to be folded. This means that the protein is not folding in the potassium 4 ions system. The values of R_g below 1 nm mean that the protein is folded, unlike the values that are above 1 nm mean that the protein is unfolded.

RMSD of APOA1 in Potassium 4 Ions System

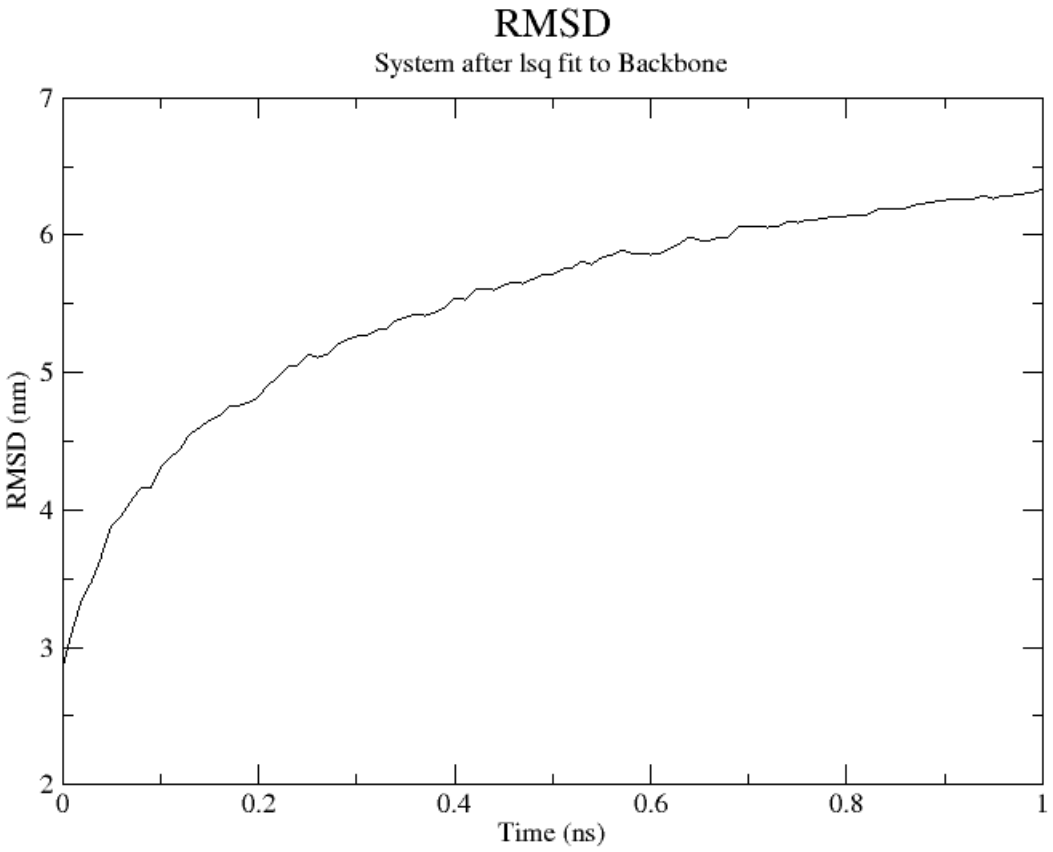

Figure 9.2: RMSD (in nm) of APOA1 in potassium 4 ions environment simulated over 1 ns. The graph shows the stability of the protein based on the constant exponential value, but comparing with the gyration graph, it shows that the protein is not folding. The values of RMSD below 1 nm mean that the protein is unfolded, unlike the values that are above 1 nm mean that the protein is folded. The values confirm the stability of the protein, but in conflict with the radius of gyration results.

Gyration (R_g) of APOA1 in Potassium 8 Ions System

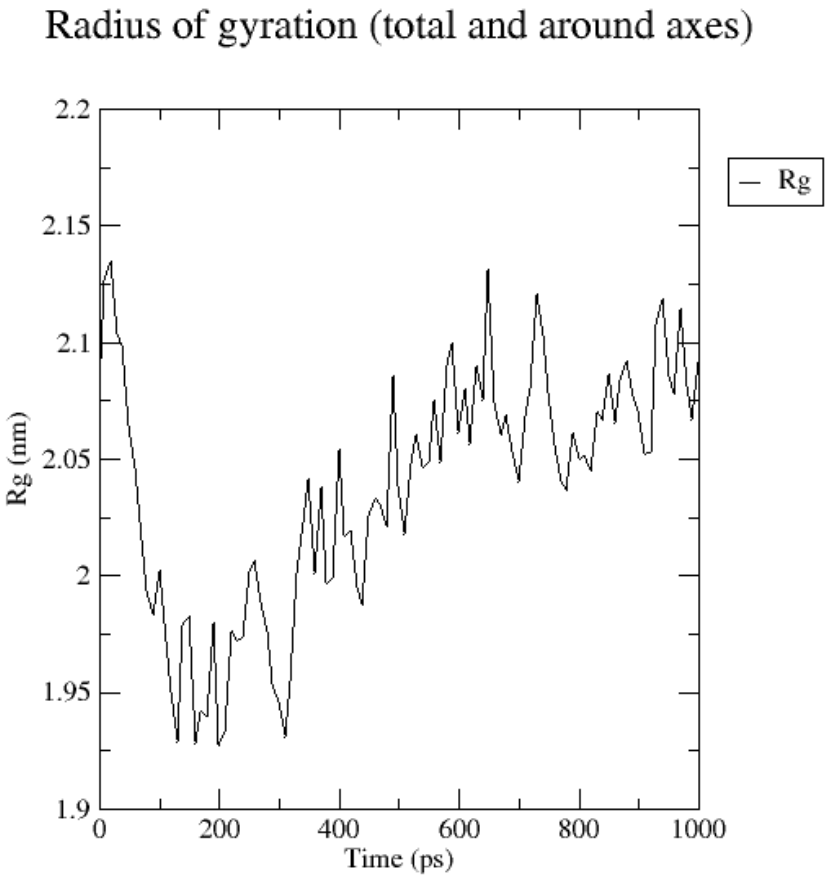

Figure 10.1: R_g (in nm) of APOA1 in potassium 8 ions environment simulated over 1000 ps. The protein is shown not to be compact in the graph because the values are not constant and above the values that are needed for the protein to be folded. This means that the protein is not folding in the potassium 8 ions system. The values of R_g below 1 nm mean that the protein is folded, unlike the values that are above 1 nm mean that the protein is unfolded.

RMSD of APOA1 in Potassium 8 Ions System

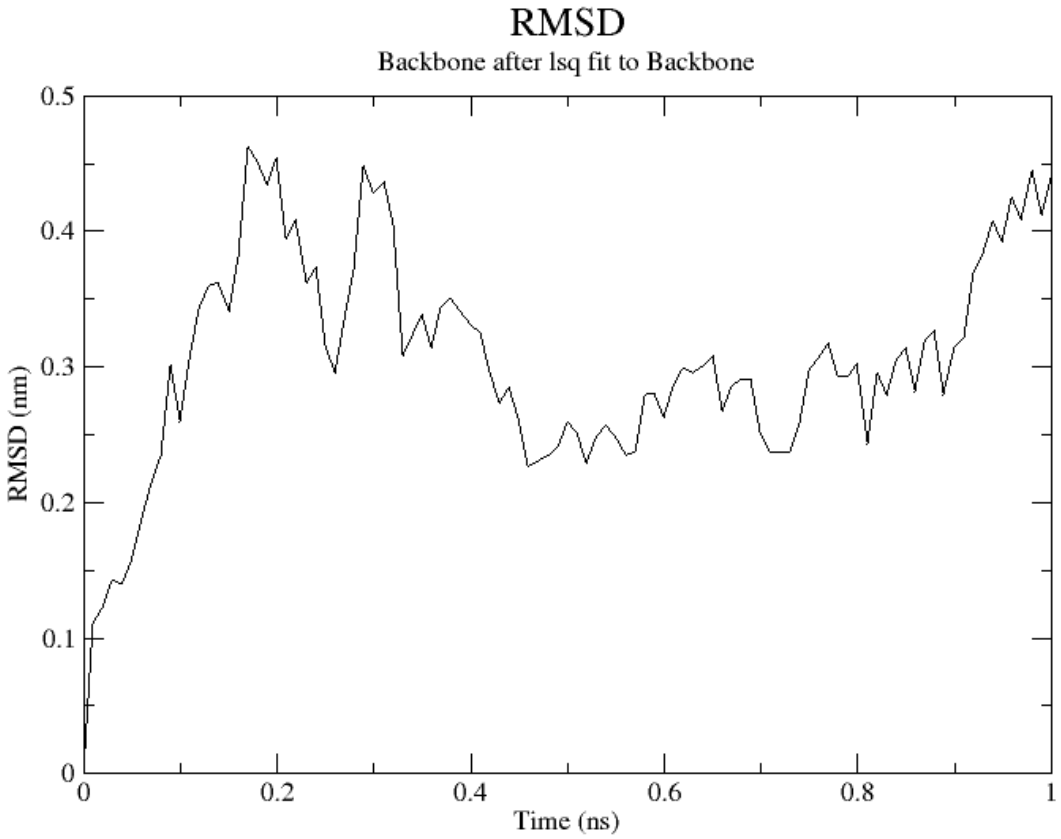

Figure 10.2: RMSD (in nm) of APOA1 in potassium 8 ions environment simulated over 1 ns. The graph shows the stability of the protein based on the constant exponential value, but comparing with the gyration graph, it shows that the protein is not folding. The values of RMSD below 1 nm mean that the protein is unfolded, unlike the values that are above 1 nm mean that the protein is folded. The protein is unfolded in the graph results.

Gyration (R_g) of APOA1 in Potassium 16 Ions System

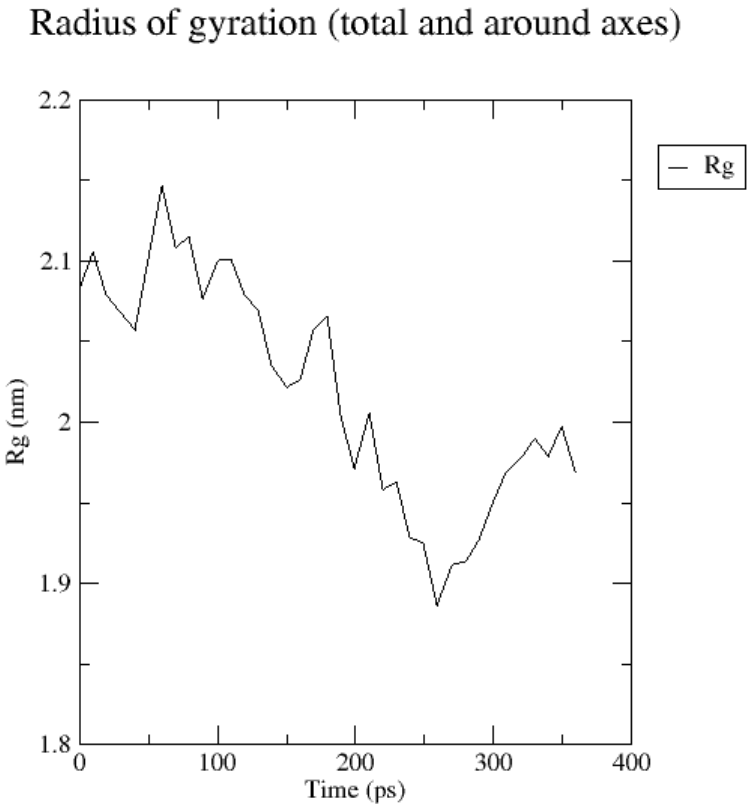

Figure 11.1: R_g (in nm) of APOA1 in potassium 16 ions environment simulated over 400 ps. The protein is shown not to be compact in the graph because the values are not constant and above the values that are needed for the protein to be folded. This means that the protein is not folding in the potassium 16 ions system. The values of R_g below 1 nm mean that the protein is folded, unlike the values that are above 1 nm mean that the protein is unfolded.

RMSD of APOA1 in Potassium 16 Ions System

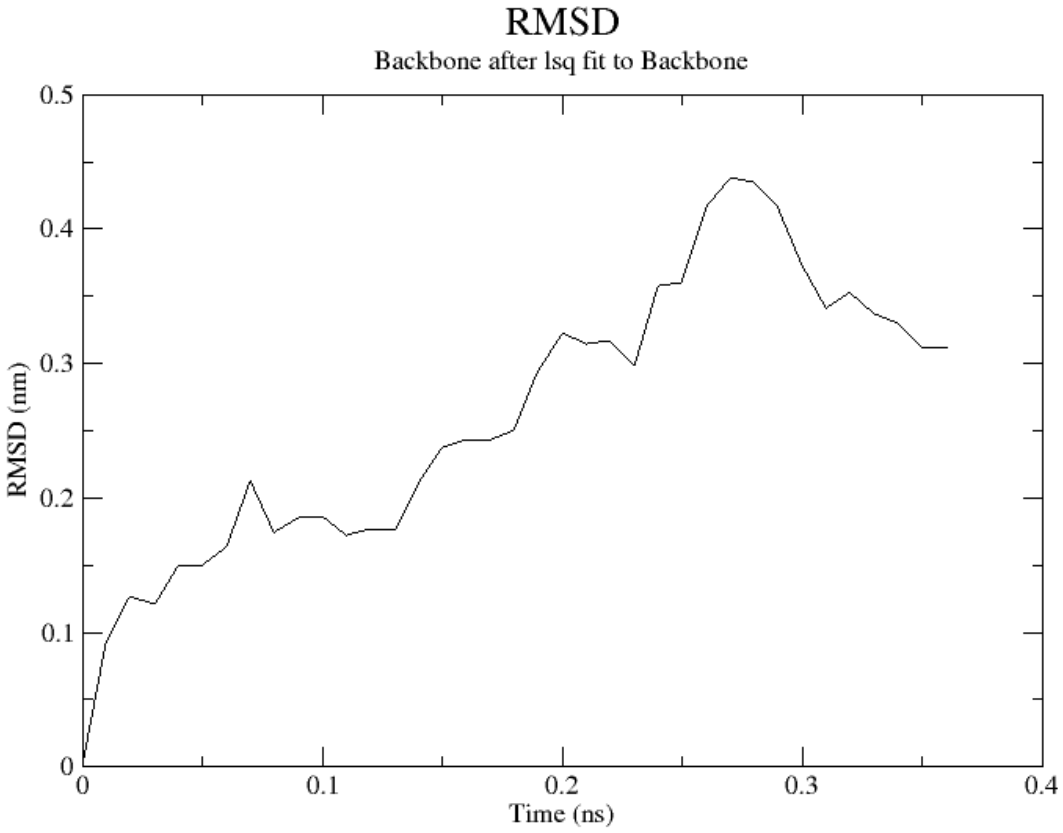

Figure 11.2: RMSD (in nm) of APOA1 in potassium 16 ions environment simulated over 0.4 ns. The graph shows the stability of the protein based on the upward trend value, but comparing with the gyration graph, it shows that the protein is not folding. The values of RMSD below 1 nm mean that the protein is unfolded, unlike the values that are above 1 nm mean that the protein is folded. The protein is unfolded in the graph results.

Chapter 4:

Discussion and Conclusion

4.1 Discussion and Conclusion:

The goal of this work is to understand the behavior of APOA1 in the potassium environment based on the increments of the number of ions that would be present in the system. The following findings are going to be compared to the study done by Smith (2019) to figure out how potassium ions can fit with the chloride ion and the ultimate behavior of APOA1 in potassium chloride solution. Due to time constraints, both data sets are discussed with the intention of hypothetical conclusion about the behavior of the protein when both types of ions would be present.

The methodology of the simulations also suggested doing the simulation of water as a control environment for the APOA1. The study by Smith (2019) provides the gyration and RMSD graphs of the APOA1 in the water solution. The graphs and the image on the bottom show that there is not much change in the structure of the protein, it is stable, and it is not folding.

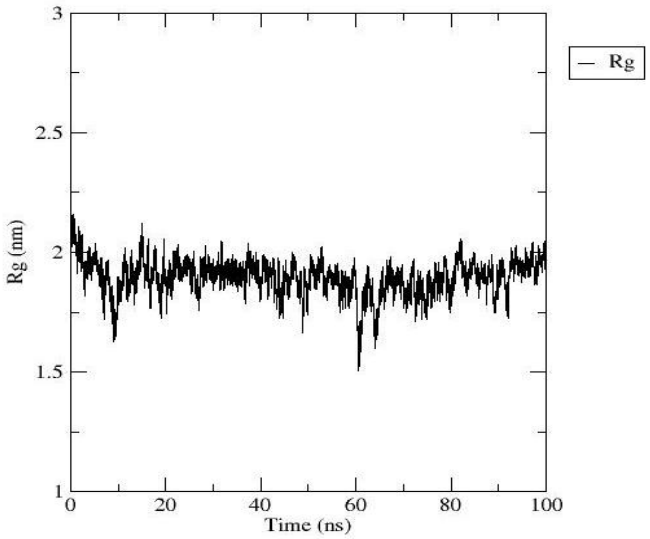

Figure 12.1: R_g (in nm) of APOA1 in pure water solution simulated over 100 ns (Smith, 2019). The protein is stable, but it is not folded because the radius of gyration value is over 1 nm.

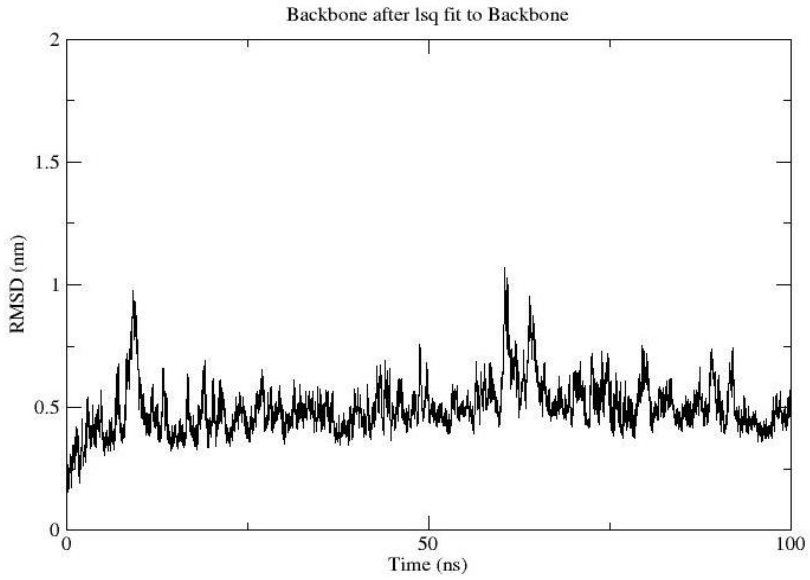

Figure 12.2: RMSD (in nm) of APOA1 in pure water solution simulated over 100 ns (Smith, 2019). The following represents a control for an environment without any addition of the ions in the system. The protein is stable, but not folded because the value is below 1 nm. Any short periods of spikes into above 1 nm range can be described as anomalies.

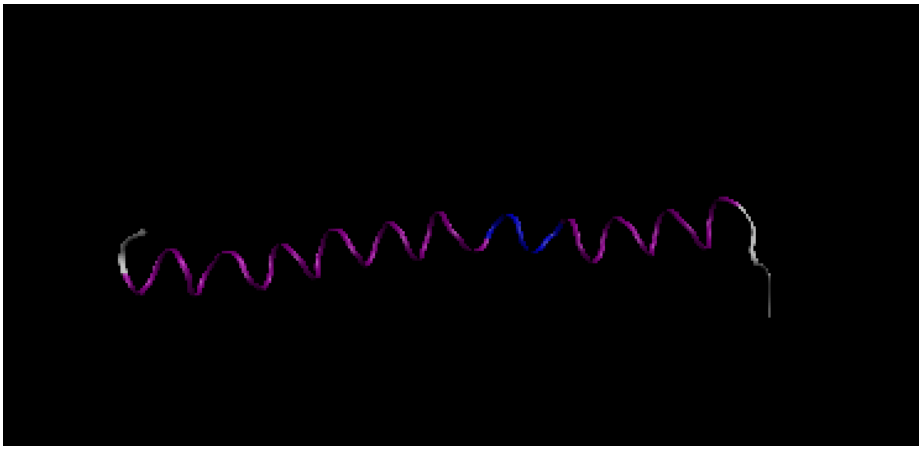

Figure 12.3: The digital representation of APOA1 in water (Smith, 2019). There is no folding present on the image, but the protein is in a stable configuration. This set up of the protein is preferred before the introduction of any ions in the system.

The chloride ion is now going to be compared to the potassium chloride. Both gyration and

RMSD will be discussed to understand their potential similarities and differences between those

two ions. The study by Smith (2019) mentions that at the lower concentrations, the drop of gyration would be evidence by the first 40 nanoseconds drop in gyration in the lower concentrations. Figures 9.1, 10.1, and 11.1 also show the drop, but only figure 10.1 has a steeper drop in the first 40 nanoseconds. Moreover, the chloride ion gyration graphs have a more close-knit trend than the gyration graphs in the potassium ion. The outcome is that the potassium ion environment provides the open configuration due to the up and down jump of gyration data, and this configuration can show that the protein is not compact and not folding. The chloride ion study is showing the folding process by the more stable data than potassium ions. Moreover, the data by Smith (2019) shows an eventually decline of the radius of gyration to the values below 1 nm, which means that the protein started to fold in the chloride ion environment. The radius of gyration in the potassium ion environment does show the decline also, but it does not lower to the value below 1 nm, and that will leave the protein in an unfolded confirmation.

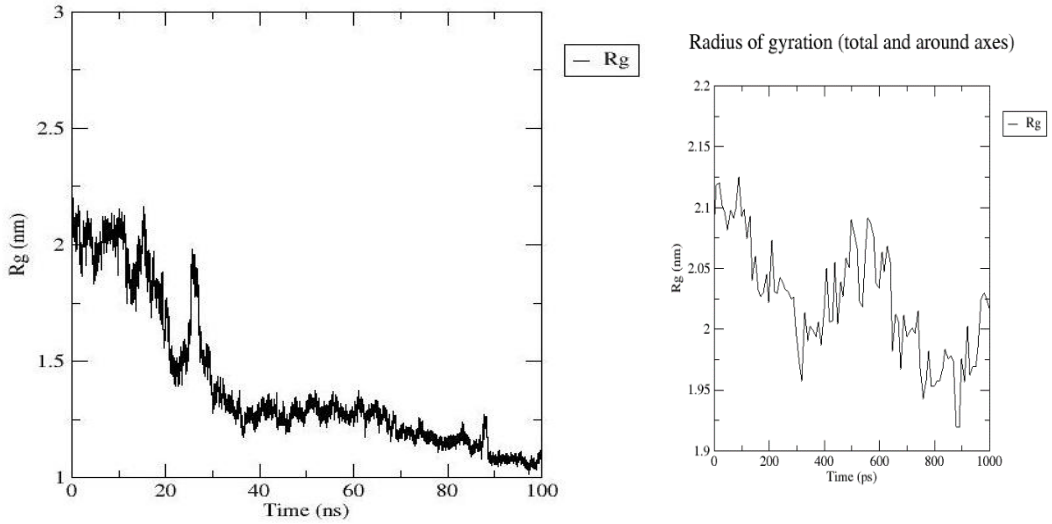

Figure 13.1: Comparison of gyration graph by Smith, 2019 (on the left) of APOA1 in 0.024% chloride anion ionic liquid simulated over 100 ns and personal gyration graph of APOA1 in 4 ions of potassium simulated over 1000 ps. The graphs were picked based on the lowest ion concentration, meaning that the higher ions comparison would follow the trend set up by the individual studies. The data by Smith (2019) shows an eventually decline of the radius of gyration to the values below 1 nm, which means that the protein started to fold in the chloride ion environment. The radius of gyration in the potassium ion environment does show the decline also, but it does not lower to the value below 1 nm, and that will leave the protein in an unfolded confirmation.

The RMSD data will verify the outcomes from the gyration graphs to determine the correct course for the behavior of the proteins in the different ionic environments. Same concentrations will be used to keep the level comparison and assume that the trends in the studies follow the inverse relationship between the ion compound and the water molecules. The study by Smith (2019) shows the data set that inclines to the values above 1 nm, which means that the protein is folding in the chloride ion environment. The data set of the potassium ion environment starts with the values above 1 nm, which does confirm the folding of the protein; however, the data results clash with the results of the radius of gyration of the potassium 4 ions system, and that would require further examination to determine if there was any simulation error.

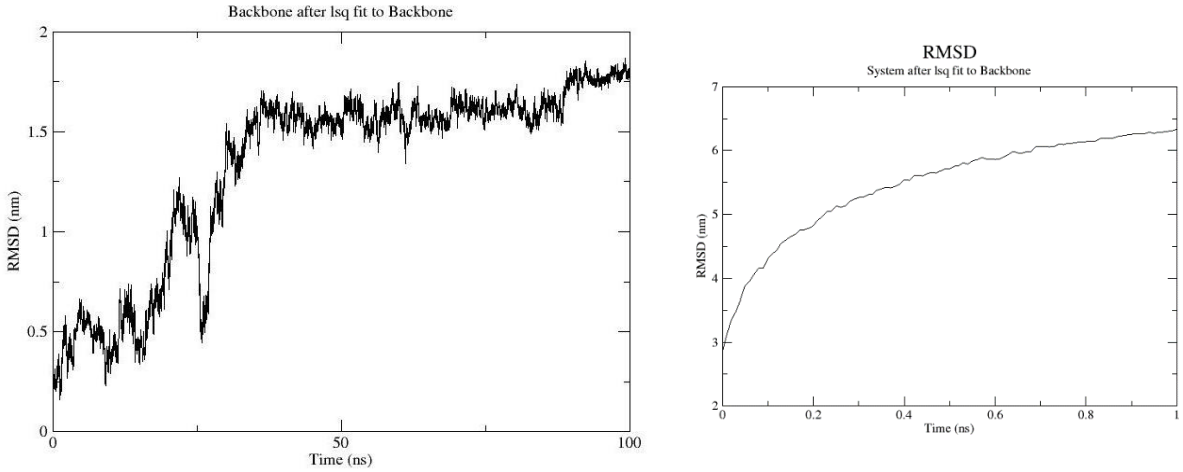

Figure 13.2: Comparison of RMSD graph by Smith, 2019 (on the left) of APOA1 in 0.024% chloride anion ionic liquid simulated over 100 ns and personal gyration graph of APOA1 in 4 ions of potassium. The graphs were picked based on the lowest ion concentration, meaning that the higher ions comparison would follow the trend set up by the individual studies. The study by Smith (2019) shows the data set that inclines to the values above 1 nm, which means that the protein is folding in the chloride ion environment. The data set of the potassium ion environment starts with the values above 1 nm, which does confirm the folding of the protein; however, the data results clash with the results of the radius of gyration of the potassium 4 ions system, and that would require further examination to determine if there was any simulation error.

Even though there are similarities between the RMSDs of this study and study by Smith (2019), looking at both gyration and RMSD of the potassium environment shows that the open configuration of the protein causes it not to fold and essentially not to work. RMSD confirms an open, linear protein on the potassium side and folded on the chloride side. The difference between the radius of gyration and RMSD results of the potassium 4 ions environment would require additional simulation runs to determine if the difference in the data results is caused by the simulation error or if the protein needed more simulation time to fold.

The matter of the charges might be the factor that distinguishes the two structures. Potassium ion is a positive molecule, and chloride is a negative molecule. This alone can speculate that positive molecules of potassium do not cause folding, but the negative molecules of chloride do cause the protein to fold. Previous studies do mention the strength of the hydrophilic and hydrophobic faces of the protein but do also mention that the ionic strength is another factor for protein folding. The previous study also worked with a sodium ion, which is also a positive ion molecule and that could be another factor to compare between the sodium and potassium ions.

The electron charges of chloride and potassium are opposite, so there is a reason to believe that APOA1 would interact with the chloride ion environment to stabilize and fold for the continuation of the cholesterol in the human body. The potassium ion environment does have the stability factor, but the protein did not fold in the simulation. The hypothesis of the research was proven wrong, due to the initial thought that the APOA1 would fold in the experimental conditions of potassium ions solution. However, there might be more to the behavior of the APOA1 if the potassium and chloride were together in the solution for the future experimental designs to help understand how the proper ions mix can still stimulate protein folding along with the stability of the protein to function and produce HDL cholesterol in the human body.

4.2 Future Work:

The next avenue to explore concerning the effect of the APOA1 in the given system environment would be to run the potassium ions simulation a few more times to verify that the data was accurate and there are no hidden anomalies. Then, the potassium and chloride ions simulations could be run side by side to analyze the graphs produced. The next challenge would be to figure

out the molarity of the salt ions in the biological system to simulate the natural microenvironment based on the data achieved through molecular dynamics. Also, since the charges might be a factor in the study about the folding of the protein, the suggested experiment might be needed to compare the protein behavior between sodium and potassium ions due to the positive charge that both molecules exhibit.

By working on the previous steps, there can be more system runs that could span over more than 1 ns or 1000 ps. The interaction between the negative charge of the chloride and the positive charge of the potassium should present a new behavior of the protein concerning the new environment and if it will be able to fold to be viable for the HDL cholesterol production in the human body. Also, other force fields can be run to compare and contrast the ability of the protein to behave based on the manipulation from the GROMACS and not only based on the ions. Moreover, the longer and multiple run times can provide a deeper understanding of the APOA1 stability and folding in the potassium chloride environment and if this salt type is suitable for the human body and its levels of healthy cholesterol.

References

Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 1-2, 19-25. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352711015000059

Ansell, B. J., Watson, K. E., Fogelman, A. M., Navab, M. & Fonarow, G. C. (2005). High-Density Lipoprotein Function: Recent Advances. *Journal of the American College of Cardiology*. 46 (10), 1792-1798. Retrieved from https://www.sciencedirect.com/science/article/pii/S0735109705019315?via=ihub

"APOA1 gene." (2020). Medline Plus. *U. S. National Library of Medicine*. Retrieved from https://medlineplus.gov/genetics/gene/apoa1/#conditions

Balderas Altamirano, M. A., Gama Goicochea, A., & Pérez, E. (2015). Folding of the Apolipoprotein A1 Driven by the Salt Concentration as a Possible Mechanism to Improve Cholesterol Trapping. *Selected Topics of Computational and Experimental Fluid Mechanics*, 281-291.

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. *J. Chem. Theory Comput.*, 6 (2), 459-466. Retrieved from https://pubs.acs.org/doi/abs/10.1021/ct900549r

Clark, P. L., Plaxco, K. W., & Sosnick, T. R. (2020). Water as a good solvent for unfolded proteins: Folding and collapse are fundamentally different. *Journal of Molecular Biology*. 432 (9), 2882-2889. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022283620300954

Doss, C. G. P., Rajith, B., Garwasis, N., Mathew, P. R., Raju, A. S., Apoorva, K., William, D., Sadhana, N. R., Himani, T., & Dike, I. P. (2012). Screening of mutations affecting protein stability and dynamics of FGFR1-A simulation analysis. *Applied & Translational Genomics*, 1, 37-43. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5121281/

Ferina, J., & Daggett, V. (2019). Visualizing Protein Folding and Unfolding. *Journal of Molecular Biology*, 431 (8), 1540-1564. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022283619301044

Gruia, A. D., Fischer, S., & Smith, J. C. (2004). Kinetics of breaking a salt-bridge critical in protein unfolding. *Chemical Physics Letters*, 385 (5-6), 337-340. Retrieved from https://www.sciencedirect.com/science/article/pii/S0009261403021717

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*, 118 (28), 7761-7769. Retrieved from https://pubs.acs.org/doi/10.1021/jp410651u

Ibragimova, G. T., & Wade, R. C. (2009). Importance of Explicit Salt Ions for Protein Stability in Molecular Dynamics Simulation. *Biophysical Journal*, 74 (6), 2906-2911. Retrieved from https://www.sciencedirect.com/science/article/pii/S0006349598779974

Lemkul, J. A. (2018). From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package, v1.0. *Living J. Comp. Mol. Sci*. In Press.

Metcalf, R. C., Peck, D. V., & Arent, L. J. (1990). Effect of potassium chloride additions on pH measurements of dilute sulphuric acid standards. *Analyst*, 7. Retrieved from https://doi.org/10.1039/AN9901500899

"Potassium chloride." *U.S. National Library of Medicine: National Center for Biotechnology Information*. (2020). Retrieved from https://pubchem.ncbi.nlm.nih.gov/compound/Potassium-chloride

Shahzad, M. A. (2018). Free energy calculations of protein-water complexes with Gromacs. *Cold Spring Harbor Laboratory: bioRxiv*. Retrieved from https://www.biorxiv.org/content/10.1101/305037v1

Smith, B. (2019). *Molecular dynamics simulation of apolipoprotein A1 in ionic liquids.* The University of Tennessee at Chattanooga. In Print.

Tang, V., Fu, S., Rayner, B. S., & Hawkins, C. L. (2019). 8-Chloroadenosine induces apoptosis in human coronary artery endothelial cells through the activation of the unfolded protein response. *Redox Biology*, 26, 1-11. Retrieved from https://www.sciencedirect.com/science/article/pii/S2213231719304306

Zacchia, M., Abategiovanni, M. L., Stratigis, S., & Capasso, G. (2016). Potassium: From Physiology to Clinical Implications. *Kidney Diseases*, 2 (2), 72-79. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4947686/

Appendix 1:

The instructed commands for potassium and chloride simulations in GROMACS. The coded instructions follow the methodology of a simulation of lysozyme in water (Lemkul, 2018).

Generate Topology

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

Choose force field 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 with (#) number of ions

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