2012

Computational Modeling Of Mucin 1 (Muc1) Peptide And Anti-Muc1 Aptamer Binding

Kristen Lorraine Rhinehardt
North Carolina Agricultural and Technical State University

COMPUTATIONAL MODELING OF MUCIN 1 (MUC1) PEPTIDE AND ANTI-MUC1 APTAMER BINDING

Kristen Lorraine Rhinehardt

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

Department: Nanoengineering

Major: Nanoengineering

Major Professor: Dr. Ram Mohan

Greensboro, North Carolina

2012

School of Graduate Studies

North Carolina Agricultural and Technical State University

This is to certify that the Master's Thesis of

Kristen Lorraine Rhinehardt

has met the thesis requirements of

North Carolina Agricultural and Technical State University

Greensboro, North Carolina

2012
Approved by:

_____	_____
Dr. Ram Mohan Major Advisor	Dr. Goundla Srinivas Committee Member
_____	_____
Dr. Marinella Sandros Committee Member	Dr. Ajit Kelkar Committee Member/Department Chair

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

Biographical Sketch

Kristen Rhinehardt was born on August 3, 1988 in High Point, North Carolina. After graduating valedictorian from T. Wingate Andrews High School she went on to receive a Bachelors of Science in Biological Engineering from Cornell University. She is a candidate for the Masters in Science in Nanoengineering.

Acknowledgements

I would like to thank

Dr. Ram Mohan

Dr. Marinella Sandros

Dr. Srinivas Goundla

Dr. Ajit Kelkar

Kenneth Flurchick

Rev. April Rhinehardt

For their help, guidance, time and continued support. Without their help and belief in me this thesis would not have been realized.

Table of Contents

List of Figures

List of Tables

Abstract

Biomarkers are molecules that indicate changes in a physiological state and are detected by biosensors. Aptamer based biosensors are highly efficient, with high specificity and reusability. An aptamer library for a 25-mer aptamer contains 10^{15} possible sequences. Experimentally, the procedure known as Systematic Evolution of Ligands by Exponential Enrichment (SELEX) is used. Selecting an aptamer from such a huge library is highly involved and time consuming. A single round of SELEX uses a few hundred aptamers and can take few hours to weeks. Use of computational modeling may simplify this aptamer selection process. Prior to computational modeling of the aptamer selection process, aptamer binding must be simulated and understood as the selection depends on the ability of an aptamer to bind to a target molecule. In the present study, we used Molecular Dynamics modeling to simulate and subsequently visualize the well-established aptamer binding combination of mucin 1(MUC1) peptide and Anti-MUC1 aptamer. During the simulation it was seen that the peptide associated twice with the aptamer. In particular, the peptide associated with the 12^{th} tyrosine residue of the aptamer loop after 25ns before dissociating and binding with the 3' and 5' ends of the aptamer. Post simulation analysis of the Radius of Gyration, atomic distance to the wet lab surface plasma resonance imaging (SPRi) results corroborated with the observations of the simulation results. Current foundational study shows that computational molecular dynamics simulations can provide molecular level insight for aptamer-peptide binding process, which is difficult to probe directly in wet lab experiments.

CHAPTER 1

Introduction

Biomarkers are molecules that correspond with bio-chemical changes in the body. Biomarkers undergo changes in concentration, physiology and morphology as they track disease progression and drug effectiveness in the body (1). These molecules have opened a window of opportunity in the diagnosis of many diseases, like traumatic brain injuries. Traumatic Brain Injury (TBI) is classified as an acute injury to head which results in physiological dysfunction of the brain (2). This injury can vary from simple falls to athletic contests or battle field injuries. TBI is can be classified as mild, moderate or severe (3). The nature of the injury in the case of mild or moderate TBI makes it difficult to detect by conventional methods such as neural imaging, behavioral observation or standardized testing. When undiagnosed, mild/moderate TBIs could cause a patient to encounter comas, permanent behavioral changes or even death. In the absence of a physiological manner to readily diagnose such TBIs, research has turned to the molecular level for answers. Changes in a selected group of proteins have been shown to coincide with those cases of known and medically diagnosed TBI (4). The proteins S100β, Serum Amyloid A (SAA), C-reactive Protein (CRP) and Retinol Binding Protein (RBP4) have been identified as biomarkers for this disease (5; 6). CRP and SAA are proteins that respond to inflammation while RBP4 is a vitamin A transporter that is reduced in TBI cases (6). S100β is a protein which found in the brain and the body but is elevated in the brain in cases of TBI (5). Testing this protein elevation in addition to those of the SAA, CRP and RBP4 can be used in confirming the presence of TBI and its severity. These biomarkers can be used as a detector with high specificity and repeatability in a biosensor device. Currently, glucose meters and pregnancy tests have become the poster devices for biosensors because of their ease and

immediacy of use. Though TBI has an identified biomarker, it has not been thoroughly studied for its implications and full potential use in diagnostic devices.

Having a biomarker that has been thoroughly researched is thus essential in executing the present foundational study. For example, the well-studied transmembrane protein mucin 1 is a biomarker for breast cancer. In breast cancer patients, mucin 1 isoform is overexpressed in the blood stream. Discovering a biomarker alone is not sufficient to create a diagnostic device. Selecting a bio-receptor for the biomarker is essential as it will specifically recognize the biomarker among millions of other molecules. The extensive study of this biomarker has revealed antibodies and aptamers that are bio-receptors for this protein and can be incorporated in a diagnostic device.

Computational modeling built upon the associated molecular systems, and understanding of the associated energies and interactions can potentially be used to select for these biomarkers or bio-receptor molecules, understand their behavior and their bindings. Applying computational modeling to improve our understanding of the potential target molecules could greatly facilitate the selection of target molecules for any biomarker; could aid in making an efficient diagnostic tool for a given disease. The computational modeling could thus facilitate the identification of the potential high probability target molecule candidates for the focused biomarkers and understand their behavior and bindings for the enhancement of biosensor devices.

1.1 Biosensors

Investigating the genetic anomalies between protein production and disease can identify biomarkers for diseases. Many diseases like TBI cause changes in protein production and morphology as a result of the disease. These proteins can thus be used as biomarkers for the disease where the concentration of protein coincides with the severity disease. Along with the

presence of protein is the presence of antibody. When a foreign invader or imbalanced concentration of host molecules enters the blood stream, a specific cocktail of antibodies are released in the body. Each antibody protein combination can act as a lock and key to create a biological complex. A biosensor as a receptor-transducer device that provides a quantitative information using a bio-recognition element and a transducer (7). The transducer is based on electrochemical, mass, optical or thermal principles while the bio-recognition element or bio-receptor acts on some biochemical mechanism (7; 8). The bio-recognition element of the biosensor works off the formation of the biological complex. The antibody can act as the bio-recognition element and is bound to the biosensor surface. When a biological sample is loaded into the sensor, the bio-recognition element/bio-receptor recognizes the target or key in the sample and binds to it. The transducer registers the change which is quantified and displayed for the user (see Figure 1).

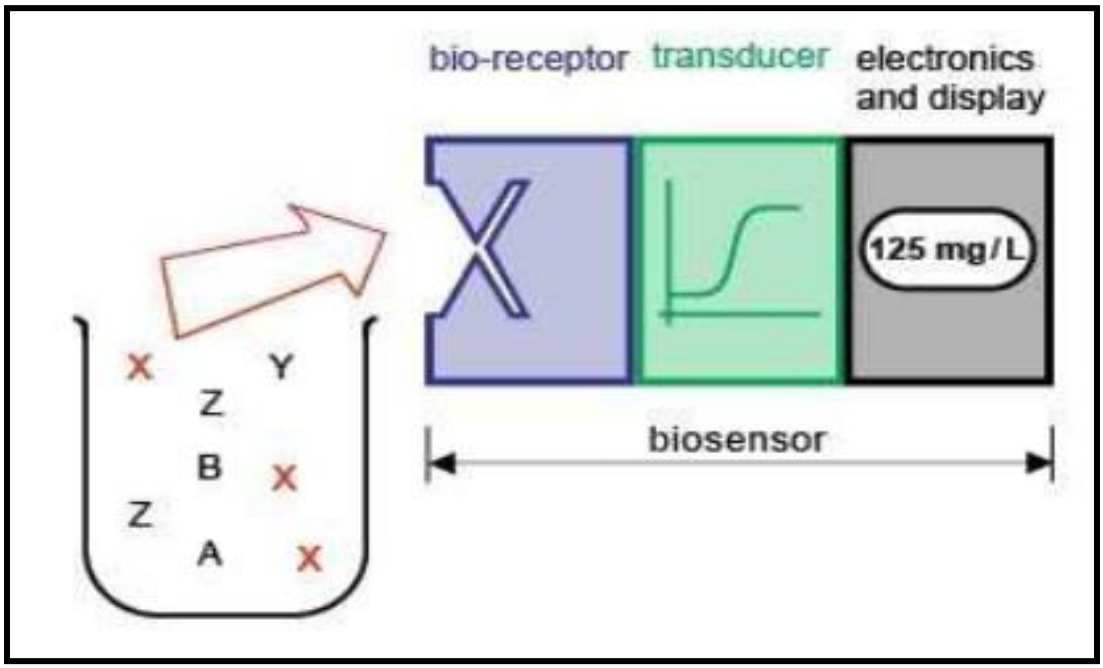

Figure 1: Biosensor Flow chart (7)

Difficulty in sensor development occurs with picking a proper bio-recognition element (7). When such proteins or genes have been linked to specific conditions they can be used as biomarkers for sensors. The most common method of making biosensors is using antibodies as the bio-recognition elements (7). Antibodies are freely floating in the blood stream and when the proper antigen is recognized it binds to it. Antibodies can take two forms of defense. Some antibodies bind to hinder biological processes while others recruit macrophages and other molecules to destroy the target. Antibodies are specific in their binding. By targeting the antibody we can create a panel of what is present in the blood and at what concentrations. Though antibody use is common there are some disadvantages of using them as bio-receptors. Antibodies are large molecules that are not readily synthesized and can be chemically unstable (9; 10). Instability can cause errors and inaccuracies in readings of the biosensor. The size of the antibody limits the number that can be placed on the surface of the biosensor. This creates a low density sensor device which may not function with high accuracy. Not only are antibodies large, they are only good for a single use in a biosensor (10). This increases the cost of the biosensor for the manufacturer and the consumer. If the biosensor is not cost effective and efficient then there is little reason for it to come to market. With these challenges better bio-recognition elements have been researched. One such element choice is aptamers (11).

1.2 Aptamers

Another possible type of bio-recognition elements are aptamers. Aptamers are oligonucleotides sequences made of single stranded DNA or RNA (12). Aptamers are advantageous as a bioreceptor as they are small, chemically stable and have a high binding affinity (9). The binding affinity of aptamers rival or better than that of antibodies (12). This high binding affinity is not only due to their ability to bind to a structure but also to fold in

correspondence to that binding. Moreover, unlike antibodies, DNA aptamers are reusable (7; 9). RNA aptamers are susceptible to ribonuclease degradation which restricts them for multiple use (7). Because of small size, it is possible to affix large number of aptamers in a single location creating a high density receptor area. Aptamers can also be easily functionalized and immobilized to surfaces to create highly ordered receptor layers (7).

A compilation of these oligonucleotides has been made into aptamer libraries. These aptamers can be made from DNA or RNA and stored in an aptamer library. A standard 25-mer library compilation currently stands at 10^{15} available aptamers (13). In solution, these aptamers are quite flexible and adopts a 3D conformation that complements the target molecule (12). To select a single aptamer for use in a biosensor device, one must filter through this entire library.

In 1990, a reasonable experimental solution has been provided by the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) process, in which developed libraries undergo incubation with the desired target molecule (13). Those aptamers that do not bind to the target are removed while bound aptamers are separated from the target and amplified using polymerase chain reactions (PCR). In this process primers are added to the aptamers and they are replicated making many double stranded copies of these aptamers. These double strands are then separated, transcribed and purified into single stranded DNA (ssDNA) (13). This pool of aptamers goes through several more rounds of SELEX until the pool is reduced to a handful of sequences (see Figure 2). As aptamers have not yet been determined for TBI this process must be done for selection. Target features, concentration, design of the initial library, experimental environment, and specificity of the binding are all determinants for the number of SELEX rounds that need to be done (13). The resulting SELEX aptamers should result in a select group that has the highest binding affinity for the target molecule. However these aptamers must be

sequenced for this process to be successful. Due to the massive size of the aptamer library, SELEX must be done in small batches and there are risks of damaging the aptamers during the process. During the PCR process of SELEX, the aptamers are amplified with the addition of primers and extension regions. These later have to be removed which further exacerbates the process. If these were not removed the aptamer folding would be changed.

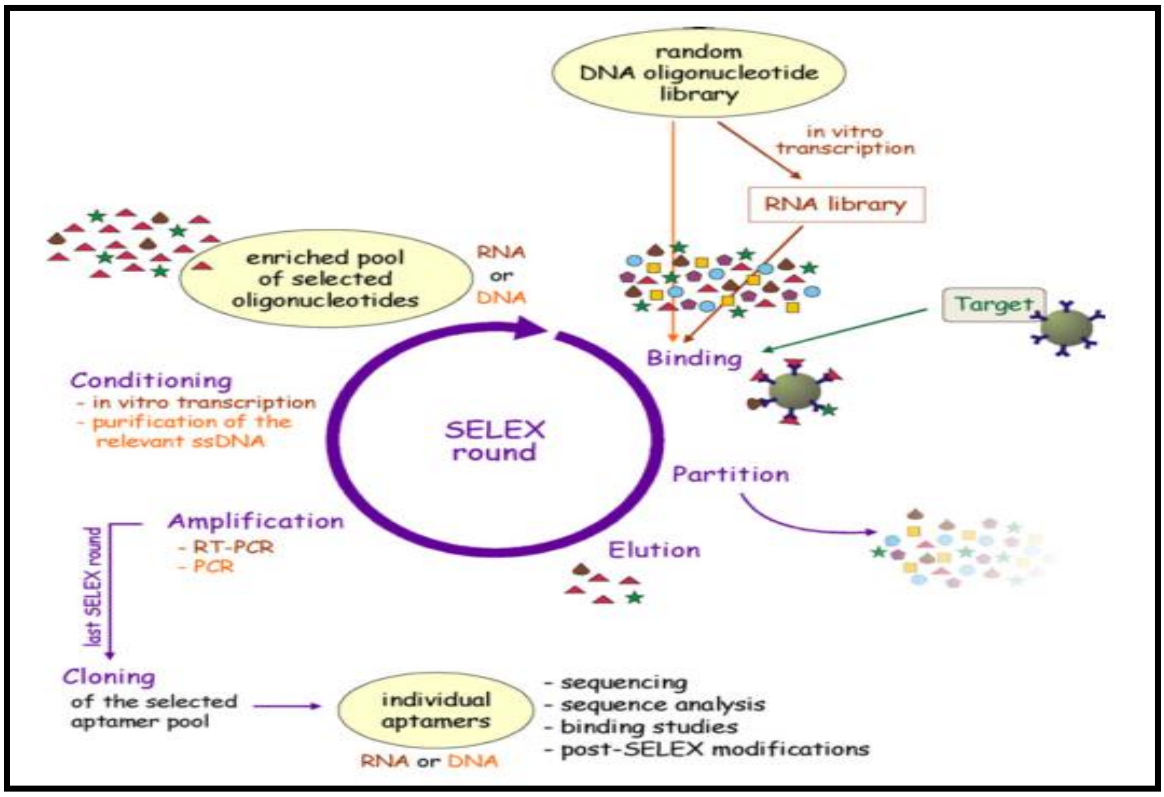

Figure 2: SELEX process (13)

SELEX is the most common way of selecting aptamers, however many variations and alternatives have been developed in the last 20 years (14-16). Non-SELEX capillary electrophoresis has reduced the time needed for aptamer selection and has removed the PCR process (17). Though this process can select a group of aptamers over time, none of these methods have shown how they bind or the preferential binding sites. Additionally, it is not

possible to analyze visually the natural binding process. Computational modeling may provide insights into aptamer-peptide binding from a molecular view point.

1.3 Computational Modeling

As devices are moving towards the nanoscale and below, many questions of how things work need to be addressed. Biomolecules that form our existence can now be tested. Knowing their properties and behavior is a necessity to see if as they are viable for use in current and future technology. These tests are currently being performed in many wet lab experiments. Through these tests it is possible to quantify the end results but in most cases we cannot see what is happening during the process. The biomolecular motion can be visualized and analyzed using computational modeling. Simulations are modeled to include the effect of the associated processes involved and can provide an emulation of the real behavior based on the parameters and account for the associated conditions of the physical problem. In the case of molecular structures involving protein – aptamer interactions, simulations can provide visual conformation of the step by step progression of an experiment which is not always possible in the case of wet lab experiments for the case of aptamer bindings.

One computational modeling technique applicable for the analysis of biomolecular motion and interactions is based on Molecular Dynamics (MD) modeling (18; 19). This modeling methodology has been recently applied to determine the chemical, physical and mechanical properties of materials. Other techniques such as finite element methods have been regularly employed for the analysis of various physical phenomena involving solids, fluids and the associated mass, momentum and energy transfer over the years, and are now commonly used in many engineering applications. Having these computational solutions provides infrastructure and understanding on which companies and laboratories can build and solve many of today's

complex inter- and multi- disciplinary problems. Computational modeling, for example, provides the means to see how a material or structure is formed, deformed and affected based on the associated influencing parameters that can be changed as needed.

Based on the time and length scales involved, computational modeling and analysis approaches can be broadly divided into three main categories; Quantum Mechanics, Molecular Dynamics, and Mesoscale Dynamics (20). Mesoscale Dynamics focuses on systems that involve billions of atoms and are generally based on larger geometrical sizes that are represented by appropriate physical laws (20). Solid, liquid and gas systems that have complexities at length scale much above the atomic level can be modeled in this system (20). Large scale interface dynamics of such complexes can be simulated on this level. The algorithms in this model are generally based on Newtonian Physics. This size dependence has the limitation in length scale. Modeling small atomic scale interactions on the mesoscale would cause inaccuracy since singular or small groups of atoms do not follow Newtonian Physics.

Molecular Dynamics (MD) modeling is suitable for small systems where individual atoms and or small clusters of atoms are involved and the phenomena influenced by the motion of individual atoms (20). These models can routinely explore a time scale of picoseconds (10^{-12}s) to hundreds of nanoseconds. The governing equations in Molecular Dynamics follow classical Newtonian Physics. This method is derived from Newton's equation of motion based on the selected force fields that defines the associated forces in the computational models. This method is popularly used as means to model biological structures and interactions based on the associated molecular configurations. Not only does Molecular Dynamics fit the size of most individual biological reactions at the molecular level, but it cuts down on laboratory time and cost.

Quantum Mechanics (QM) Methods are highly suitable when the electronic structure and properties of the system need to be included in the simulation study. Generally, chemical bond formation/breakage involve electron interactions between atoms (20). Such bond formation and breakage are accurately modeled in this system. This method is the most accurate of the three models in estimating the molecular properties of a system but also is computationally expensive and is well suited for extremely small systems. The high accuracy of this system is due to its ability to account for electron interactions through appropriate quantum physics equation approximations. However, this does not come without a significant cost in computational time. The intricate equations of this model create the need for many unknown degrees of freedoms in the computational model. In addition many real systems operating at the quantum level occur in very small time periods and require extremely small time steps to firmly fit the model accurately. This model can also be combined with molecular dynamics which combines the advantages of both individual models (20).

These three modeling approaches in hand allow for modeling of almost any system if one picks the appropriate parameters. Before a simulation system is constructed one must consider the parameters for the specific application. First the size and nature of the system will determine the modeling approach to be used. In addition, the computational time for the modeling analysis is also influenced by the number of users working on the problem and how much data will be accumulated and assessed even with multiprocessors. That paired with the size of the expected final outputs and storage space must all be considered before one can determine the computational power needed to run the simulation. Additional system users, intermediate steps, size and storage can cause an increase in the computational power and time required for the analysis. Considerations also need to be made if the simulation results need to be visualized and

if customized algorithms and tools will be needed. This added to the needed computational power calls upon the aide of visualization programs that need to be compatible with the computational analysis solver used in the modeling process. For modeling to be successful and useful it not only has to provide a valid solution but also do so in an effective manner. By answering the "what if's" and "how does" questions, computational modeling is providing a solid foundation for understanding today's complex problems.

The power of today's computational modeling can be an avenue to test, analyze and visualize the aptamer binding that forms the basis of the aptamer selection process, and provides the foundation and methodology in the present work. The size of an aptamer library depends on the length of the variable region and can be approximated as

$$library\ size = 4^n$$

where n is the length of the variable region in the aptamer (7). Going through each sequence of such a library is near impossible using the current wet lab procedures. Using aptamers offers high selectivity and specificity within the biosensor and by reducing the multi-trillion aptamer pool through prior/concurrent computational modeling and analysis to a smaller consortium cuts down on the time and resources needed to find the optimum binding aptamers for the biomarker of interest. All aptamers must have primers added during the amplification process of SELEX. This addition could cause a change in the binding characteristics or location which won't be necessary in a computational model. In addition RNA aptamers must be pretreated before they can be used making them more difficult to work with. Computational modeling avoids the need of such preprocessing and enables one to analyze and understand the binding process. Open regions of the aptamer's 3D structure provide the binding sites for peptides. There can be several of these sites but it is unclear which site is used and whether the site changes under

varying conditions. Using aptamers in a biosensor has the additional challenge of identifying the optimum orientation that favors binding. Within a device an aptamer is bound to surface. Surface chemistry and target orientation influence the effectiveness of biosensors. Modeling how binding will occur will aid in the development by showing where and how this binding will occur. When a protein is cleaved in the body there is a corresponding antibody that is released to find and bind the target protein. Modeling the antibody could also attest to the presence of the target protein. All these can be analyzed and understood through computational models.

In the present thesis, using a bottom up approach we will first validate the ability of computational modeling to simulate aptamer binding. This is important as aptamer selection and bio-recognition depends on aptamer binding. Before we can simulate aptamer selection and bio-recognition we must ensure that we can indeed simulate the natural progression of binding. To further validate the accuracy of this modeling and visualization we must use a well-known combination of protein-aptamer binding. Though the motivation of this thesis is towards application of the modeling hypothesis for protein – aptamers systems of interest to TBI and mTBI, knowledge of potential aptamers and structures for this application are still limited and we cannot use it as a means to test the ability of the computational modeling approach envisioned in this process. In the case of TBI known biomarkers are available, but an aptamer for those respective proteins has not yet been identified. To prove that the proposed hypothesis of computational modeling of protein-aptamer interactions is indeed accurate and viable, we need an aptamer and protein combination that has been thoroughly tested in the present study. One such combination exists with the protein mucin 1.

1.4 Mucin 1

Breast Cancer is the most common cancer amongst women (21). There has been a lot of research at the genetic and molecular level for this type of cancer (21). One gene in particular that is unregulated in the breast cancer cells is the MUC1 gene (21). This gene codes for the transmembrane protein MUC1. Transmembrane mucin proteins are thought to be sensors for environmental changes and as a second line of defense (22). MUC1 has many alternative spliced forms: MUC1/SEC, MUC1/Y, and MUC1/X/Y (22). MUC1/SEC is a secreted form of MUC1 and lacks hydrophobicity while MUC1/Y, MUC1/X/Z lack the tandem repeat regions (21; 22). In breast cancer there is an up regulation of the MUC1/Y isoform (22). This correlation between breast cancer and MUC1 has been found as possible biomarker for breast cancer. For carcinomas there is an overexpression of the antigenically distinct isoform of MUC1. This version of the MUC1 has O-glycans which are shorter and expose the core of the protein that contains a variety of peptide epitopes (23-25). The antibody, SM3, that has a high affinity for this version of MUC1 (22; 26; 27). This antibody is crystallized with a peptide antigen that is exposed in the carcinogenic isoform of MUC1 creating a protein-peptide complex (23). The antibody is made of 2 immunoglobulin molecules that bind to a peptide sequence APDTRPAP by hydrogen bonds and Vander Waals interactions (23; 28). The peptide can vary from the APDTRPAP sequence but the sequences are similar to this epitope in the membrane portion of the MUC1 protein (29). It is this exposed epitope region in cancerous MUC1 that is targeted by the SM3 antibody. In cases of cancer this isoform is free floating in the blood and can be targeted by the SM3 antibody. There is an inverse relationship between antibody concentration and disease severity (22; 30). Those with a decreased severity of breast cancer will have more

free antibodies than bound antibodies. Higher concentration of free antibody have been shown to inhibit distant metastases (22; 26).

Purchasing the entire MUC1 protein for the wet lab experiments can be expensive. Instead, testing is done with the peptide epitopes as they are the target within the protein. This is a less expensive way to test the binding of the aptamer. However this binding may be different with the antibody present. MUC1 has multiple binding epitopes. However, how this binding will naturally occur and how this will aid recognition in a sensor raises interesting questions. Current research has shown that MUC1 peptide and anti-MUC1 aptamer do indeed bind. However, the orientation of binding between the unbound MUC1 peptide and aptamer is unknown. There are multiple binding sites on an antibody but where they bind also depends on what else is present. Antibodies that target the peptide sequence of MUC1 could already be bound which could affect how and where the aptamer will bind. By computationally modeling aptamer-peptide binding along with aptamer-peptide-antibody binding, we get a better understanding on how this binding occurs and how we can change the surface chemistry in biosensors to optimize target and bio-receptor recognition and binding. Such questions will be answered in this thesis with the aid of computational modeling methods.

CHAPTER 2

Computational Modeling Methodology

2.1 Molecular Dynamics Force Field

In this research, we use Molecular Dynamics (MD) simulation that is most suitable for studying proteins and aptamers from a molecular level. We excluded quantum mechanical calculations as there are no chemical reactions or electron level interactions involved in this study but rather the behavior of the atoms within the macromolecules and how binding occurs under given conditions. Using the Newtonian equations of motion, Molecular Dynamics simulations can predict the movement of the atomic behavior of molecules as closely as possible to that exist under lab conditions (31). For accuracy of such simulations we must consider the force field that is used for calculating the energy changes in the system.

There is a need for computational techniques as we investigate more into the components that make up specific pieces of matter. The how, why and what questions can be answered with aid of computational simulations and models. With each passing year there is an increase in computational power that allow for more intricate models which calls on more detailed algorithms and components that further improve the fidelity and modeling size of such simulations. This increase in computing power also brings into account the specificity that is available to increase the accuracy of computational models. For MD simulations, force fields are an important part completing this task and influence the accuracy of these simulations. The form of the force field is determined by the potential energy of a molecule in a particular geometry (32). To generate this force field there is a need to account the following energy interaction parameters that exist in a molecular system. This energy is the sum of the bonded interactions

due to bond stretching, angle bending, torsion and inversion, and the non-bonded interactions due to van der Waals and electrostatic forces.

Various force-fields such as AMBER, CHARMM, and GROMOS have been successful in simulating various physical, biological and material systems. Most of these force-fields are available for free usage. For example, the AMBER99sb force field, a modified version of AMBER force field, was developed for biological applications, and hence used in the present study (33). This force field has been improved over the years with the increasing ability to analyze and run simulations of the molecular dynamics over hundreds of nanoseconds. This force field is often used with proteins and DNA structures.

The force field defines how the potential energy will be calculated. This energy summation is written as

$$E = E_{bonded} + E_{nonbonded} + E_{other}$$

Bonded energies include bond stretching and bond angle and torsional energy (34). Non bonded interactions include Van der Waals interactions and electrostatics forces. Force field specific terms are contributed in the E_{other}term.

The E_{bonded} term is the summation of the energy due to 3 parameters. The first parameter in bonded interactions is the bond stretching (E_S). This 2 atom interaction can be described as

$$E_s = \sum_{bonds} K_r \, (r - r_{eq})^2$$

This equation is modeled from the quadratic equation for Hooke's Law of the potential energy of a spring where the force constant (K_r), the bond radii (r) and the equilibrated radii is (r_{eq}) are the parameters (33; 34). The force constant is based per atom on the Badger's rule which is

determined by an assumption that at equilibrium each atomic bond is under the influence of short range Pauli repulsive forces and attractive ion forces (33) and can be written as

$$K_r = K_{ij}^{ref}(\frac{r_{ij}^{ref}}{r_{ij}})^6$$

The natural bond length or bond radius is the summation of the atom type specific single bond radii

$$r_{IJ} = r_J + r_I + r_{BO} + r_{EN}$$

where r_I and r_J are the single bond radii determined from x-ray crystallography or NMR calculations and the bond order correction and the electronegativity correction are r_{BO} and r_{EN} (35). The second parameter is the angular distortion due to bending which is a 3 atom interaction. The cosine Fourier expansion was selected to define the bond bending over other means and can be simplified for both linear and nonlinear coordination environments. For the linear and planer cases the equation can be simplified to a two term Fourier expansion (35)

$$E_\theta = \frac{k_{IJK}}{n^2}[1 - cosn\theta]$$

The nonlinear case restricts the angle bending due to repulsion forces to be between 0^o and 180^o. The force constants were determined under Badger law. This range introduces a third term to be Fourier expansion given by

$$E_\theta = \sum_{angles} K_\theta\,(\theta - \theta_{eq})^2$$

where the force constant describes the stiffness and equilibrium geometry and of the angle of equilibrium is θ_{eq} (33).

Torsion or dihedral angle displacements are the last bonded interactions in the potential energy summation. Torsion is described by a Fourier expansion where the rotational barriers vary with periodic trends of the structure

$$E_{\varphi} = \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)]$$

Where γ is the phase angle, V_n is the rotational barrier and n is the multiplicity term (33). The structure not only changes the rotational barriers but also the angle and *n* value for each.

Non-bonded interactions are also accounted for with two specific terms. Van der Waals interactions are incorporated into the potential energy. These interactions are for atom interactions and are treated with Lennard-Jones 6-12 expression (33)

$$E_{vdw} = \epsilon_{IJ}[\left(\frac{r_{IJ}}{r}\right)^{12} - \left(\frac{r_{IJ}}{r}\right)^{6}]$$

Electrostatic force is another non-bonded term included in the potential energy calculation and is written as

$$E_{el} = (\frac{Q_I Q_J}{\epsilon R_{IJ}})$$

This term takes into account the partial charges along a molecule (Q_I and Q_J) in electron units as well as the dielectric constant (ϵ) and the distance (R_{IJ}) (33).

Force fields such as AMBER, CHARMM, and GROMOS incorporate and extrapolate based on the above general forms to account for more than nucleic acid and protein molecules to include DNA and other single atom calculations. The AMBER99sb force field has the added improvement based on quantum-chemical representations of specific tetra peptides (36; 37). This force field is more refined to improve the representation of folded protein systems (36). Accordingly, we use AMBER99sb force-field in the present research work.

2.2 MD Simulation Methodology

The molecular system configurations for all the simulation studies are derived from a topology file. These files are usually in the formats of .pdb, .mol2 and .xml and define the individual atoms, types, bonds that form the molecular structure. These molecular structures are

generated from NMR or X-Ray crystallographic solutions of the associated material systems. These files are used as a starting configuration for the simulations in molecular dynamics. MD simulation analysis based on the Newtonian equations of motion can be completed using several MD simulation codes developed over the years. The choice of MD simulation analysis package depends upon the material system studied and the availability of associated force field for the material system of interest. In general, the starting topology files must then be converted into a readable format for the respective MD simulation analysis package. MD computational modeling simulations in the present study employed MD analyzer GROMACS.

From the starting molecular conformations, the enclosed work space must be defined. We define sufficient work space by generating a box to enclose the entire molecule. The simulation is designed to mimic the experimental system as closely as possible. The thermodynamic and physical parameters that exist in the physical problem must be incorporated in the computational analysis. These parameters include the conditions and environment in which the wet lab experiments are done such as water, buffers, solvents and additional biomolecules. In the case of solvents individual ion concentrations are added to the water environment. Once the environment has been defined, the system is subsequently minimized, equilibrated and simulated.

After the initial construction, we need to make sure that the system is structurally and energetically relaxed. This process is done during energy minimization in MD simulations. For every conformation there is a corresponding stable, minimal potential energy state. The selected force field defines how this potential energy is calculated. As the system relaxes, its potential energy decreases until it reaches a stable conformation. The most stable conformation will also be at the lowest potential energy state. With each conformation of the system a potential energy

is calculated. When the first derivative of the force field is equal to zero corresponding to a particular molecular configuration, there is a minimum and when the second derivative of the potential is at a maximum, a stable molecular confirmation exists. A system can have several local minimum conformations but only one absolute minimum conformation. During minimization, the corresponding energy is calculated using these mathematical energy minimization methods. The molecular dynamics parameter (mdp) files are used as instruction files that contains the user's choice for this minimization method when GROMACS is used as the MD analyzer. After minimization the molecular system is in its most stable conformation.

Using this new conformation the molecular system is equilibrated. The minimized structure has an energetically stable system in terms of geometry and solvent orientation. However these are still separate systems in a single space. We must equilibrate the solvent and ions around the protein to have a harmonious system. If we were to attempt unrestrained dynamics at this point, the system may collapse. The reason is that the solvent is mostly optimized within itself, and not necessarily with the molecule. It needs to be brought to the specified thermodynamic parameters given in the .mdp files so that the system has the proper orientation about the molecule. The system is brought to these parameters slowly using various ensembles. There are three types of ensembles in MD modeling. The microcanonical ensemble is also called the NVE ensemble (38). In this case the number of atoms, volume and energy are held constant in the simulation. Controlling the constant energy in this system is done by convection, conduction and radiation. The canonical ensemble is also known as the NVT ensemble (38). When a system is subjected to this ensemble the number of atoms, volume and temperature are held constant. Changes in temperature can be expressed in terms of energy so

that the temperature remains constant. Temperature is related to the kinetic energy as shown in the equations below.

$$E(p,r) = K(p_s) + U(r_s) = constant$$

$$K = \frac{1}{2m}\sum_{i}^{N} p_i^2 = \frac{3}{2}Nk_bT$$

$$p = momentum = mv$$

$$\frac{m}{3Nk_b}\sum_{i}^{N} v_i^2 = T$$

To control the temperature computationally one must choose a thermostat. These thermostats vary in how they define the changes in temperature in a constant temperature system. A velocity scaling thermostat adjusts the velocity in the system to control the temperature. The Berendsen thermostat creates a ghost heat bath at the desired temperature that acts as a source for the removal and addition of heat (39). The velocity is scaled so that the change in temperature over time is proportional to the difference between the kinetic temperature and the heat bath over a coupling parameter that corresponds to the coupling strength of the system. The Andersen thermostat also uses a ghost heat bath that emits thermal particles. The coupling strength is proportional to the frequency of collisions between particles in the heat bath and those of the system (38). The Langevin method introduces a friction and random term to the Newtonian Equation of motion to simulation the interactions of atoms with the solvent (40). The friction removes energy from the system while the random term accounts for the addition of energy due to collision of particles as they move within a solvent. The last thermostat is the Nose-Hoover method which introduces additional degrees of freedom as a way to calculate the changes in temperature (41). The final ensemble used in computational modeling is the isobaric-

isothermal/NPT ensemble (38). By holding the number of atoms, pressure and temperature constant the molecular system is expressed as a function of volume.

$$\frac{PV}{T} = nR = constant$$

$$\Delta T = 0 \; and \; \Delta P = 0$$

$$\frac{\Delta P}{\Delta T} = constant = \frac{nR}{\Delta V}$$

This change in volume with respect to pressure and temperature can be expressed through the compressibility of the system.

$$\kappa = -\frac{1}{V}(\frac{\partial V}{\partial P})_T$$

Changes in pressure will be expressed as a function of the position of the atoms and the gradient of the potential energy. Barostats that control how the pressure is calculated are applied in the molecular systems employed in the present study. The Berendsen and Andersen methods for NVT are adapted for pressure and can be applied in this NPT system with additional consideration of the volume changing with the potential energy (42). This change in volume can also effect the density calculation of the system.

After equilibration we can begin to dynamic analysis (production) phase of the MD simulations. During the equilibration steps the atoms bond are held in their fixed positions. During the production stage of the MD simulations these atoms are released from their constraints to move in their natural progression in the given system. These MD production simulations can take hours or weeks depending on time duration of the dynamic analysis and the molecular system sizes. In other words, the dynamic time duration of the analysis depends on the size of the system, how long simulations need to be carried out, how often data is collected and how many processors are at work. Typical desktop computer has single central processing

unit (CPU). The calculation speed and computing power is limited to that single computer and its characteristics. The speed of the calculations in single CPU systems depends on the complexity and size of the molecular system. MD analysis of large molecular systems is computationally expensive and requires the use of multi-processor computing systems. A multiprocessor computing system has a collection of connected processors that allows the user to use as many CPU as available. This allows that the use of multiple CPUs and allows the simulation to be distributed and completed more quickly employing suitable multi-processor versions of MD analysis codes such as GROMACS. Multiprocessors provide the advantage of increased computational power that is required for large scale analysis. Increased computational power gives the opportunity to run longer time duration simulations as well as introduce more sophisticated force fields. These advantages allow for better models and results that are more accurate for the real world problems.

Post production analysis will assess the results from the dynamic simulation study. Both quantitative parameters and visual analysis of the dynamic atom structure changes are studied. Fortunately, there is a variety of visualization software's available that are compatible with the output files of commonly used computational modeling packages. Viewing these output files not only provides a visualization of the molecular process in the simulation but also shows the dynamics of each step of the MD analysis process. Recent versions of simulation packages such as GROMACS have an added advantage that include post processing analysis codes built into the package. These quantitative analysis codes and analysis parameters studied vary based on what is being observed in the real problem. These quantitative analysis results act as a verification method for the simulation and are to be studied and interpreted by the modeler to draw upon the required conclusions. This MD simulation method was used in the present work

to simulate and analyze MUC1 peptide and anti-MUC1 aptamer binding process both in presence and absence of antibody.

CHAPTER 3

Anti-MUC1 Aptamer, MUC1 Peptide and SM3 Antibody Analysis

GROMACS, an open source MD analysis software, is one of the fastest molecular dynamics package to date and is commonly used with biomolecules (43). In this work, we use GROMACS to simulate the dynamics of the aptamer-peptide binding. Currently in the Joint School of Nanoscience and Nanoengineering (JSNN) genomics lab, aptamers are being tested in a .15M solution of sodium chloride (NaCl). All simulations were done with a .15M NaCl solvent at room temperature (300K) and absolute pressure (1bar). Before binding could be tested, simulations using the individual aptamer, peptide modified aptamer and antibody complex were performed to establish and understand their individual behaviors in this solvent environment. Each of these configurations for the molecular systems studied was obtained from the RCSB Protein Data Bank as .pdb files. We used the NMR configuration of the Anti-MUC1 aptamer s2.2 for this testing. The SM3 antibody complex is an X-ray crystallographic structure of the SM3 antibody along with the MUC1 peptide sequence APDTRPAP. This peptide sequence was isolated from the antibody complex using the visualization and analysis code Pymol (44). The modified aptamer was built in Pymol by removing 2 nucleotides from the 5' end of the Anti-MUC1 aptamer. Following the individual runs we tested a series of combinations to see if the following aptamer bindings can be simulated, namely: aptamer-peptide, aptamer-antibody complex and modified aptamer-peptide. These combined molecular systems were obtained from the original .pdb files and combined virtually within Pymol.

Within GROMACS each molecule configuration was added and converted into a .gro file from the .pdb starting configuration. This conversion was not possible with the antibody complex due to the missing atoms in the original .pdb configuration files. This was overcome

through the use of molecular analysis visualization software Maestro that permitted the addition of missing bonds and atoms to create a new configuration. This new configuration was then converted in GROMACS by ignoring the hydrogen atoms and subsequently treated as all other MD simulation analysis runs. All simulation employed cubic cells with a minimum distance of 1nm from the box edge (see Table 1).

Table 1

Atom Count and Box Size for Each Simulation

Configuration	Number of Starting Atoms	Box size (nm^3)	Number of Solvent Atoms	Total Number of Atoms
Aptamer	728	389.1	38317	39045
Peptide	115	144.6	14402	14517
Antibody Complex	6269	1500.9	44854	51123
Modified Aptamer	663	200.8	19571	20234
Aptamer-Peptide	853	462.8	45570	46423
Aptamer-Antibody Complex	7009	2834.3	277509	284518
Modified Aptamer-Peptide	778	445.8	44106	44884

The cell box size was increased as needed for the accommodation of the molecules. In the next step of the MD simulation analysis, water was added to the system. In the solvated system, water atoms were replaced with sodium and chloride atoms until the solvent concentration of NaCl rose to .15M. Each configuration contained a different number of atoms depending on the box size and the number of atoms in the molecule (see Table 1).

The protein structure was minimized for a minimum of 100 ps. The potential energy was analyzed to make sure the system reached its most stable configuration. This configuration was then constrained and equilibrated using NVT and NPT ensembles. The NVT equilibration was done with all bonds constrained and temperature coupled by a Velocity rescale thermostat which is a modified Berendsen thermostat specific to GROMACS (45). This was done for each configuration with the exception of the antibody complex. The antibody complex structure required two sets of NVT equilibrium runs. The first was with unconstrained bonds followed by the second set of NVT with constrained bonds for a longer time. Using the resulting data from the NVT, the system was equilibrated under NPT thermo dynamical ensemble. Equilibration was done with the same velocity-rescale temperature coupling in addition to the Parrinello-Rahmen pressure coupling. The Parrinello-Rahmen method is an extension of the Anderson thermostat. As volume is no longer constrained the Parrinello-Rahmen barostat allows the simulation cell box to change its shape accordingly (46). Both NVT and NPT equilibrations were done for 1ns each. The temperature and pressure dynamic variations were verified to ensure the system was indeed equilibrated.

The fully equilibrated system is used as the starting configuration for the MD production simulation. In the production simulations atoms are unconstrained and are free to move in their most energetically favorable positions in a dynamic process. The system was initially tested for

1ns at standard temperature and pressure. After examining these results, we extended the MD dynamic analysis runs for the combined molecular systems defined earlier in Table 1, as binding takes longer time. All the long term MD analysis simulations were completed using the multiprocessor cluster (HERMES) at North Carolina A&T State University. The number of processors used depended on the size of the molecular system studied. The aptamer-peptide combinations and aptamer- antibody complex were simulated for a total length of 110ns and 15ns respectively during the production run.

Post processing analysis tools were applied to visualize and quantify the configurations of the system using Visual Molecular Dynamics (VMD) software and GROMACS (47). We have used several quantities to analyze aptamer-peptide binding. In the following, we describe each of the quantities in detail. Atomic Distance was also considered during these post-processing analyses. Spatial confinement and periodic boundary conditions can make visualized data appear unbound as the molecules approach the edges of the box. To ensure that the distance between the two molecules are remaining constant throughout the simulation further reaffirms that the molecules are still in close proximity. To assure that these are indeed bound we must look at the formation of bonds. We can do this by calculating the amount of hydrogen bonds formed between the 2 neighboring molecules. Binding occurs when there is an attractive force between 2 molecules due to their close proximity. Hydrogen Bonds are formed by a favorable chemical interaction between atoms in a system. Though we are looking for binding not bonding we can still calculate the number of chemical bonds formed during the dynamics simulation. In GROMACS hydrogen bonds are defined as bonds shorter than .35nm and at an angle less than 30 degrees. The formation of hydrogen bonds indicates that favorable interaction exists between the two distinct molecules. Non bonded interactions such as Van der Waals forces are still

present in the system. Checking for the hydrogen bonds not only allows us to visualize binding but also to see that the system itself did not implode during processing. Another parameter that defines the deviation of atomic positions compared to the starting structure is root mean square deviation (RMSD) which is a least squares approximation of the distance between atoms (48). In the cases of individual atoms, if they are properly settling and are not highly flexible we expect the RMSD to reduce over time. This reduction should be due to the atoms folding on themselves into a more stable state. In the cases of binding there should also be a reduction in the RMSD over time as the binding restricts atomic movement compared to unbound state. If there was an increase it would indicate that the molecules are moving apart or the structure has to become more open to bind. Fluctuations in the RMSD show a highly flexible molecule in constant movement. These fluctuations can also occur when the system has not reached stable final state and indicates a need for dynamics simulations to be continued for a longer period of time. In order to measure the compactness of protein we used another quantity, namely Radius of Gyration (R_g) (49). As molecules become closer together the Radius of Gyration decreases and as they move further apart there is an increase in the Radius of Gyration. When the molecule is stable the Radius of Gyration stabilizes at a finite value with little fluctuation indicating the minimal atomistic movements within the stabilized structure.

CHAPTER 4

Results and Discussions

After completing MD production simulation, a complete analysis of the molecular systems defined in the previous chapter was conducted. The required computing time for the simulations of aptamer, peptide, antibody and combination of binding varied. The initial simulations of the aptamer and the peptide used a single processor and required 5 hour of single processor computing run time (see Table 2).

Table 2

Computing Time and Processors for MD Dynamic Simulations

Configuration	**Number of Atoms**	**Time**	**Number of Processors**	**Rate (ns/day)**	**Real Time (hours:min)**	**GFlops**
Aptamer	39045	1ns	1	4.905	4:54	12.121
Peptide	14517	1ns	1	5.336	4:30	4.879
Antibody Complex	151123	1ns	60	2.517	9:53	24.397
Modified Aptamer	20234	1ns	60	88.6	:16	115.122
Aptamer-Peptide	46423	100ns	24	20.192	118:54	59.113
Aptamer-Antibody Complex	284518	100ns	60+	----	----	----
Modified Aptamer-Peptide	44884	100ns	96	66.149	34:42	194.293

However with larger systems this time would be exacerbated which would make simulating inefficient. These larger systems were completed using multi-processors. Table 2 presents the computing time, number of processors, normalized MD dynamic time completed.
The analysis results for the individual components established first and binding systems are discussed next.

4.1 Anti-MUC1 Aptamer

Upon completion of each MD simulation the results were visualized to analyze and understand the molecular configuration. The configuration remained same through the addition of solvent and equilibration (see Figure 3A-D). Post simulation configuration of the aptamer appears to be more compact with the loop region flattened (see Figure 3E).

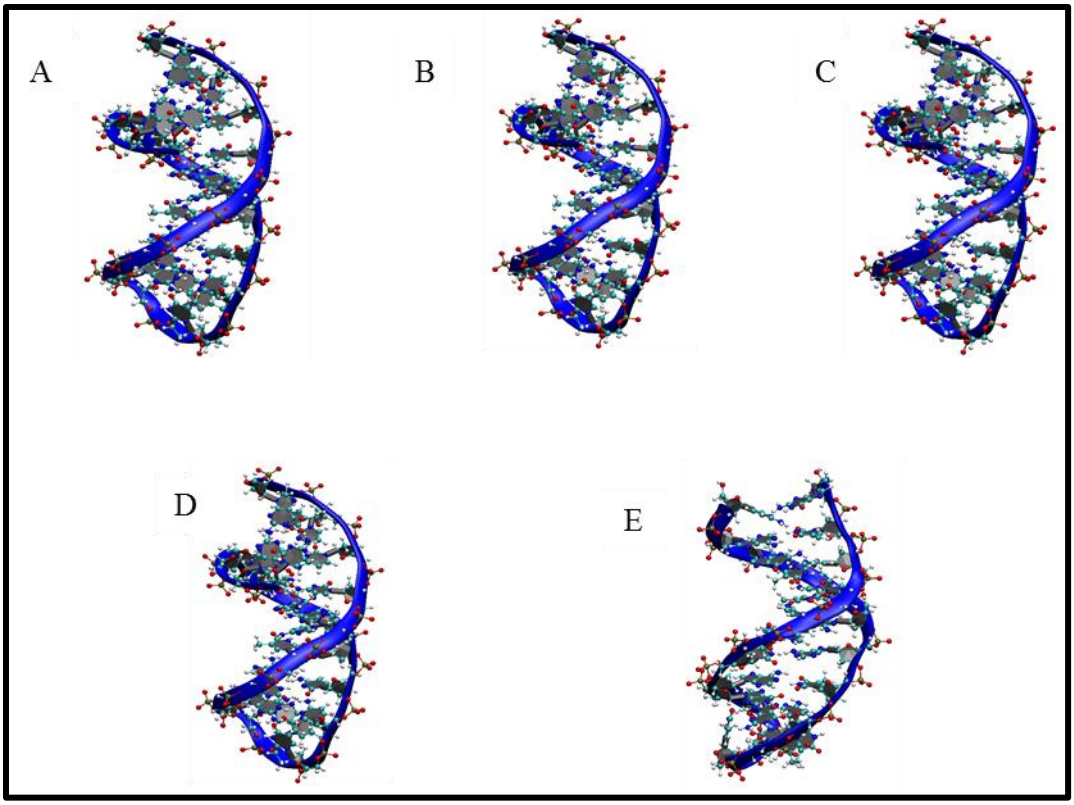

Figure 3: MD simulation images of the Anti-MUC1 Aptamer (water and ions are not shown for clarity) A) Solvent Structure B) Energy Minimized Structure C) NVT equilibrated Structure D) NPT equilibrated Structure and E) Post Simulation Structure

The potential energy for the starting aptamer system was relatively high, which gradually decreased in less than .4ns (see Figure 4). The convergence of the potential energy shows that the minimization has reached and resulted in a stable configuration.

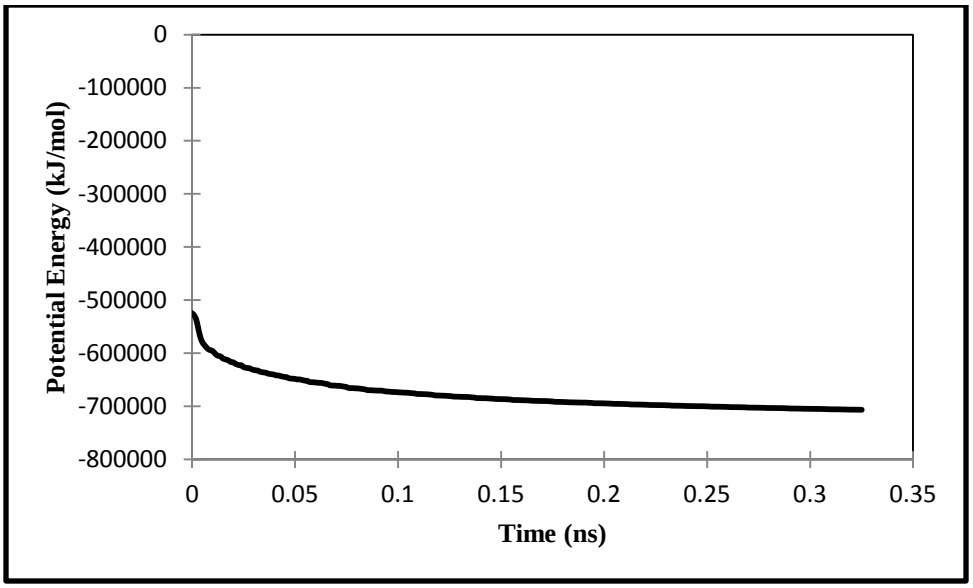

Figure 4: Anti-MUC1 Aptamer Potential Energy during Minimization

During equilibration the temperature fluctuated close to 300K and is within the accepted dynamic fluctuations (see Figure 5).

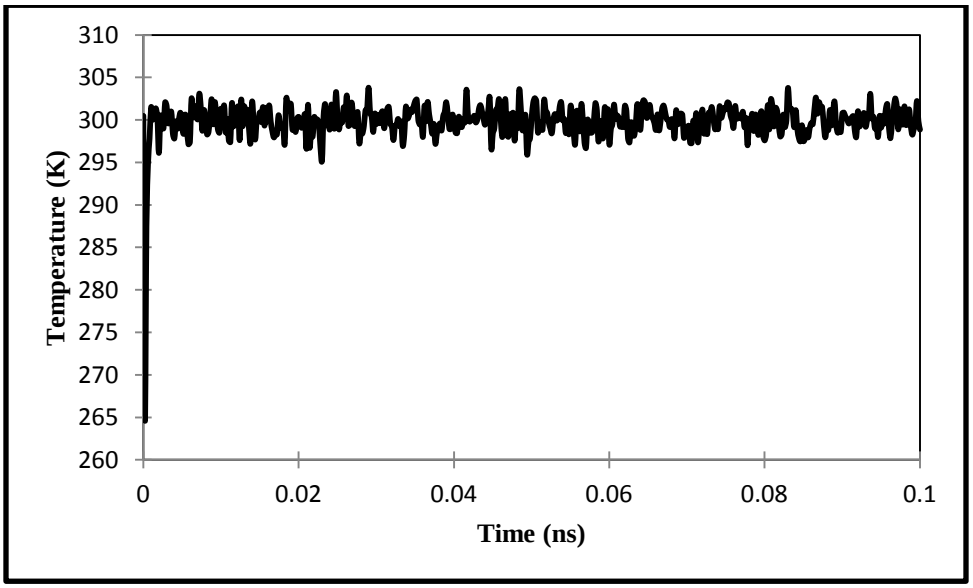

Figure 5: Anti-MUC1 Aptamer temperature during NVT equilibration

The pressure during the NPT equilibration was highly fluctuating (see Figure 6). The average pressure was near 1 bar but the individual points were between 500 and -500bar. The density

during this equilibration was also varied but was consistently between 1020 and 1025 kg/m^3 which is a variance less than 10 percent (see Figure 7).

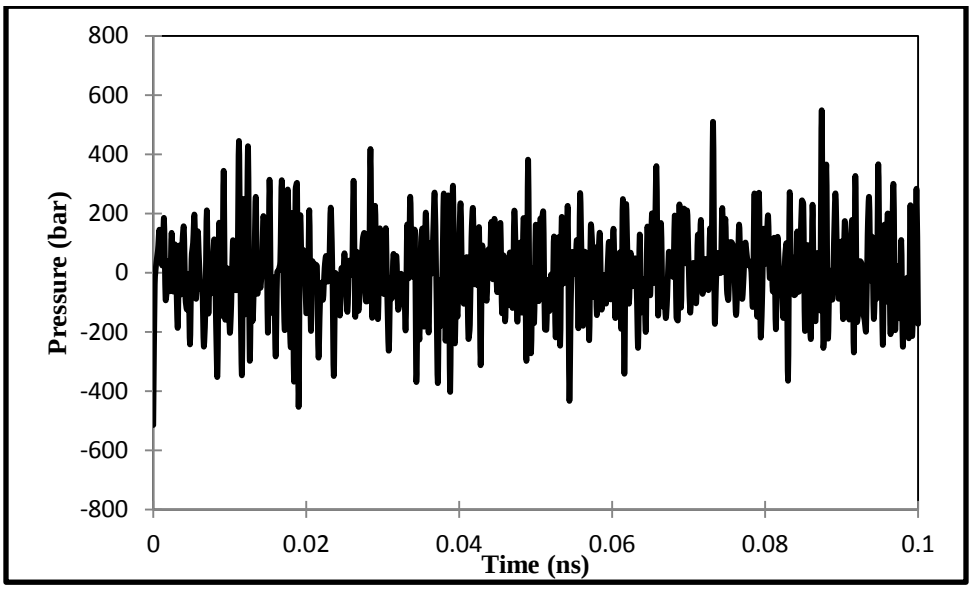

Figure 6: Anti-MUC1 Pressure during NPT equilibration

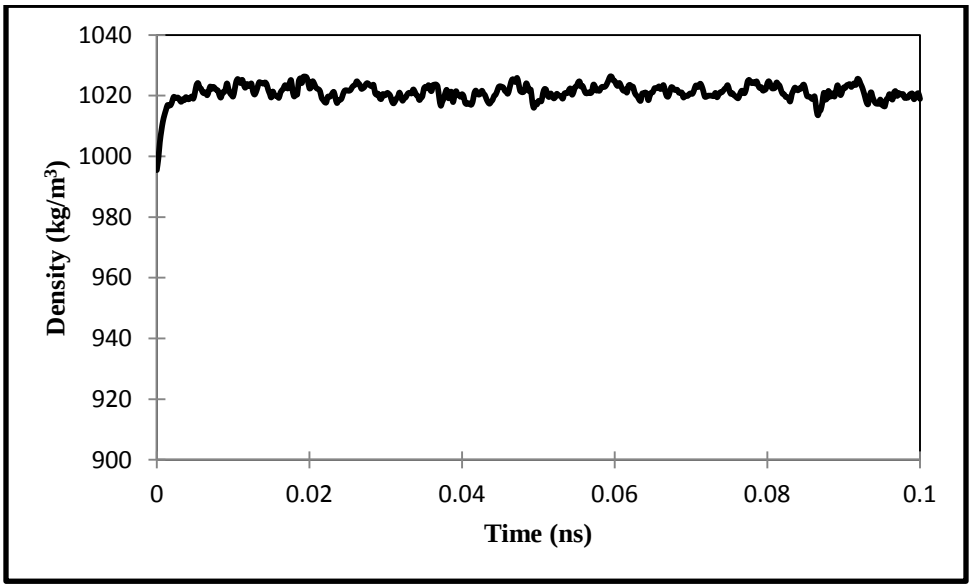

Figure 7: Anti-MUC1 Aptamer Density during the NPT Ensemble

The Root Mean Square Deviation (RMSD) of the aptamer increased over time before settling near .25nm (see Figure 8). However in the final stages of the simulation, the RMSD

shows an increase. This increase may be a fluctuation or the start of another change in the RMSD. Resolving this would require a larger time dynamic simulation analysis. The Radius of Gyration was consistently near 1.3nm with a slight dip at .7ns (see Figure 9). This shows that the atoms though more spread apart have still settled in a less compact space.

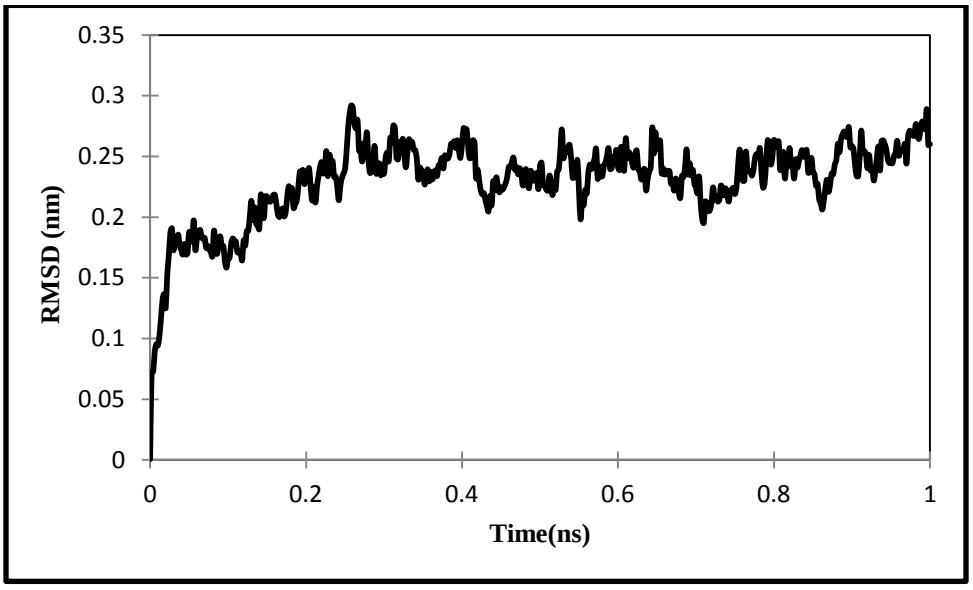

Figure 8: Anti-MUC1 Aptamer RMSD Variation during 1ns Simulation

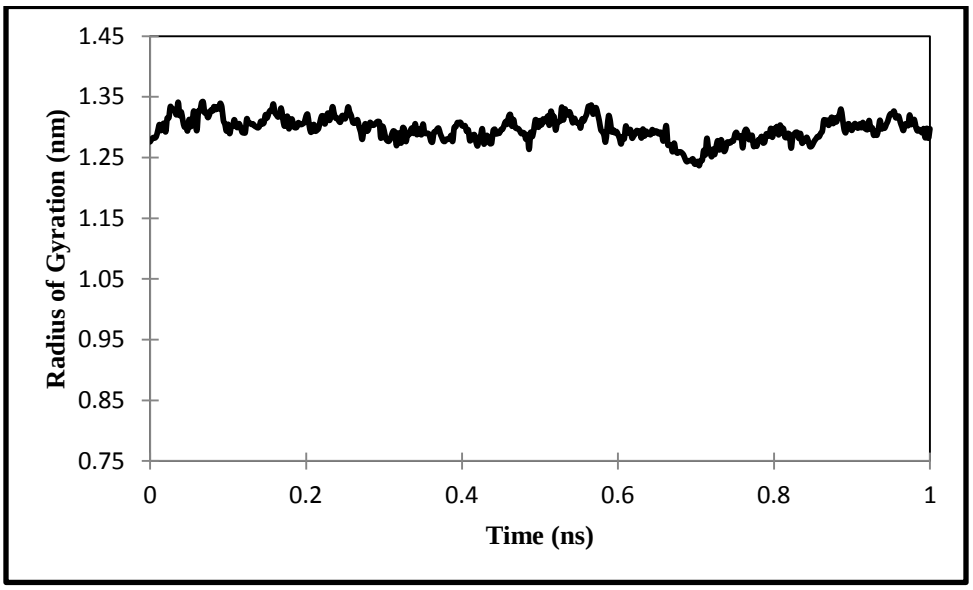

Figure 9: Anti-MUC1 Aptamer Radius of Gyration during 1ns Simulation

4.2 MUC1 Peptide

The MUC1 Peptide is very flexible. Though there was little change in the configuration during the first stages of simulation the peptide changed greatly during the dynamic simulation (see Figure 10). In the initial stages of simulation peptide starts expanding, but eventually becomes more compact (see Figure 10E-J). With its highly flexible backbone the peptide spreads and changes conformation often.

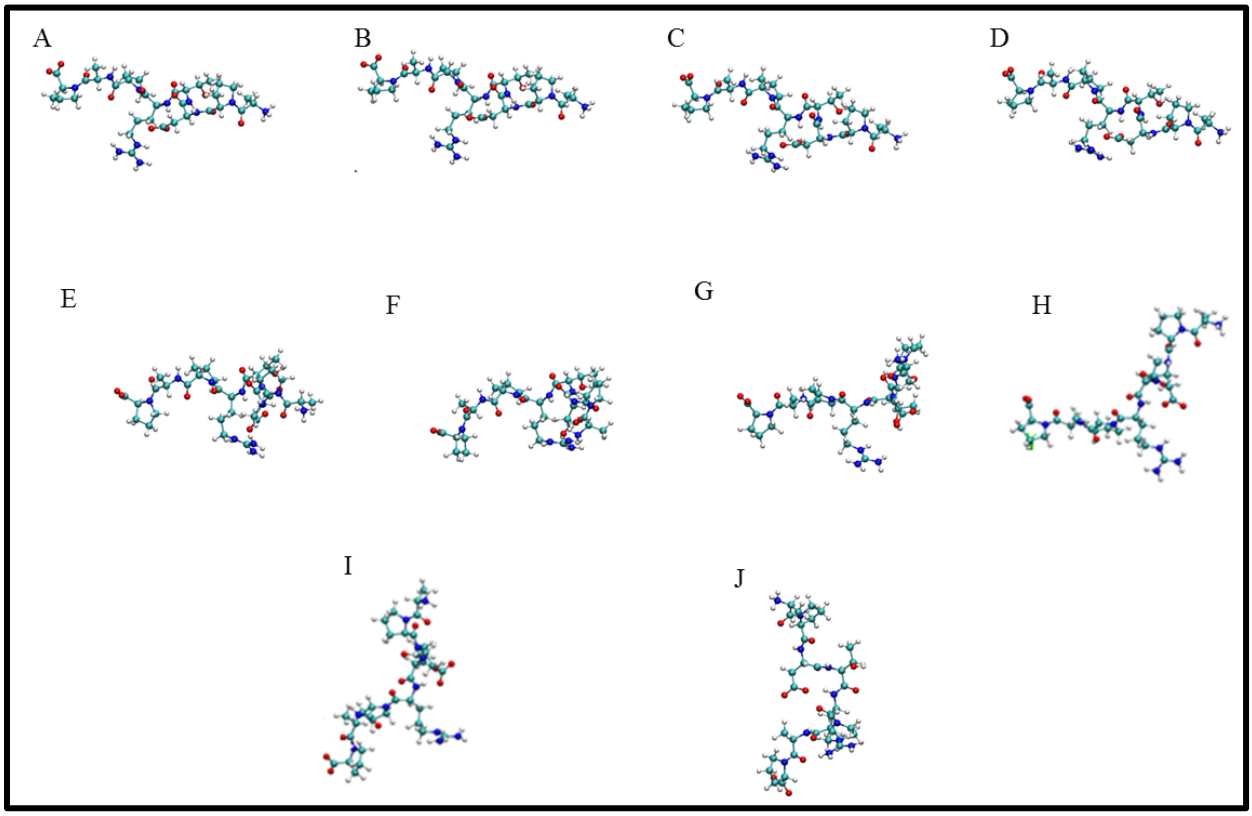

Figure 10: MUC1 Peptide A) Starting Configuration B) Minimized Configuration C) Post NVT Configuration D) Post NPT Configuration E) Configuration at F) 272ps F) 304ps G) 408ps H) 564ps I) 730ps and J) Configuration at 1ns

Minimization was successfully achieved without any errors or warnings. The potential energy reduced and stabilized around -250000 kJ/mol (see Figure 11).

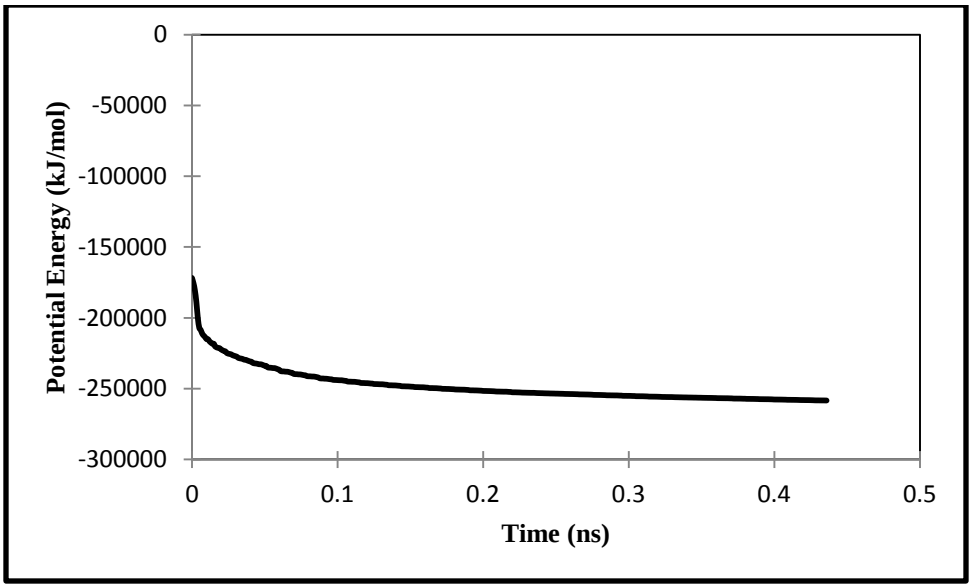

Figure 11: MUC1 Peptide Potential Energy during Minimization

During equilibration the temperature stayed consistently near 300K (see Figure 12). However the pressure post NPT was fluctuated greatly (see Figure 13). This could be due to the Parrinello-Rahmen barostat applied to the system as it releases pressure by allowing the volume to fluctuate.

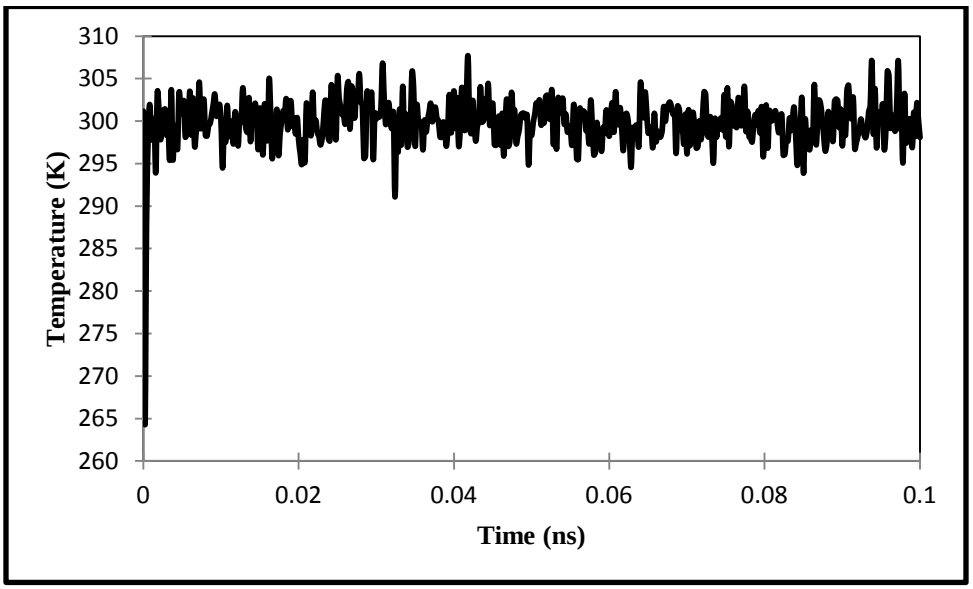

Figure 12: MUC1 Peptide Temperature during NVT Equilibration

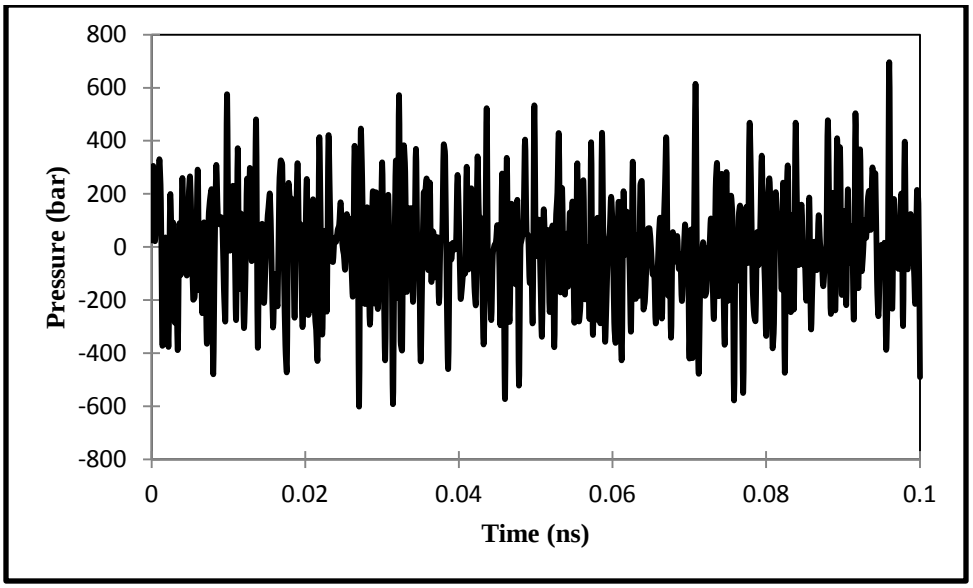

Figure 13: MUC1 Peptide Pressure during NPT Equilibration

The density during the NPT ensemble was consistent (see Figure 14). Density is based on the kg of the molecule occupying a unit volume. As an NPT simulation can fluctuate in volume the consistency of the density correlates to the changes in the pressure. This density resulted in a change less than 10 percent throughout the equilibration despite the variation in volume that is due to the Parrinello-Rahmen barostat.

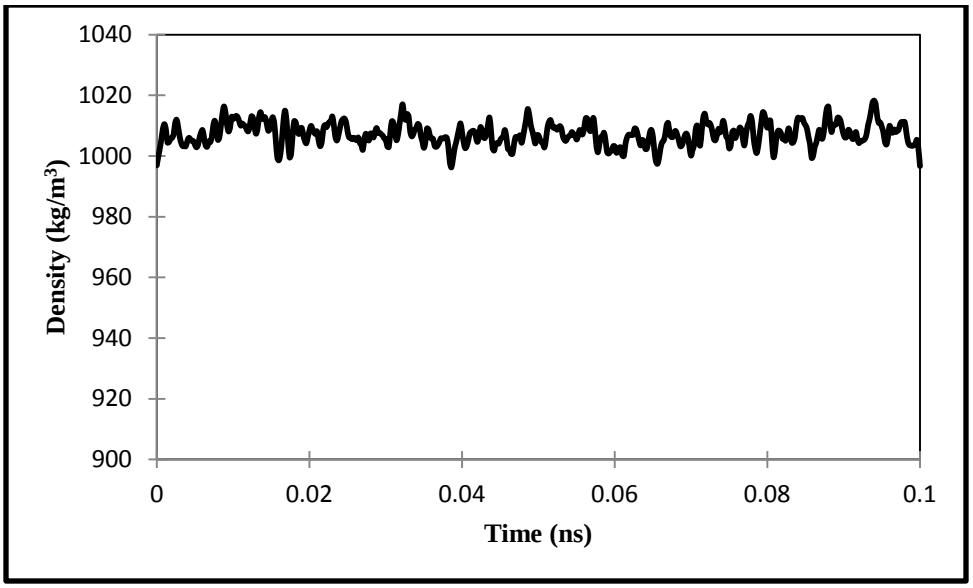

Figure 14: MUC1 Peptide Density during NPT Equilibration

The peptide has a highly flexible backbone which will change the distance between the atoms. The RMSD of the backbone showed a lot of variation throughout the dynamic simulation (see Figure 15).

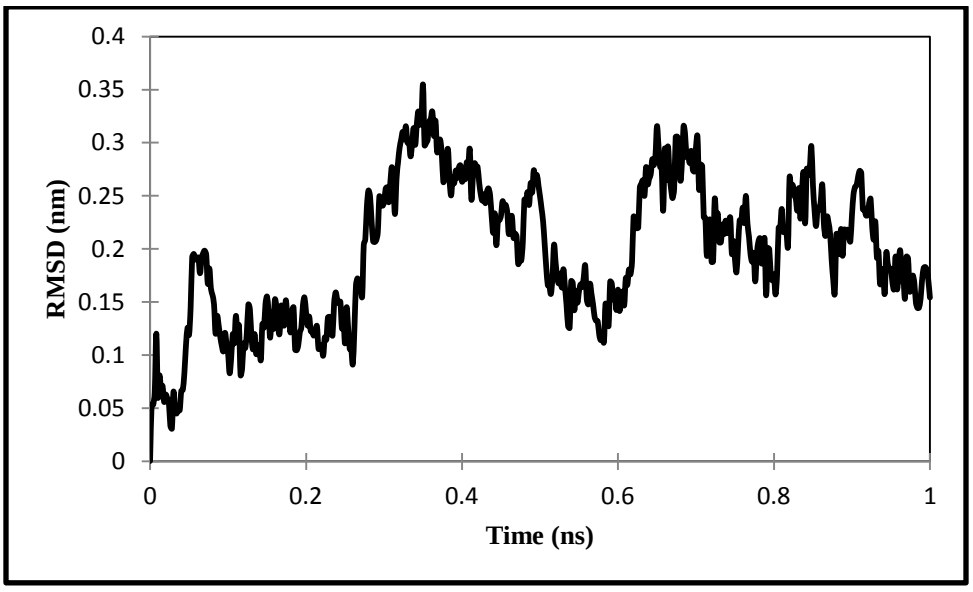

Figure 15: MUC1 Peptide RMSD during a 1ns Simulation

The Radius of Gyration was more consistent for the peptide (see Figure 16). The variations observed in the figure are due to the backbone movement. When the backbone began to fold over the peptide became more compact and corresponds to the slight decreases seen in the graph. The increases occurred because the folding was only temporary and the peptide began to unfold.

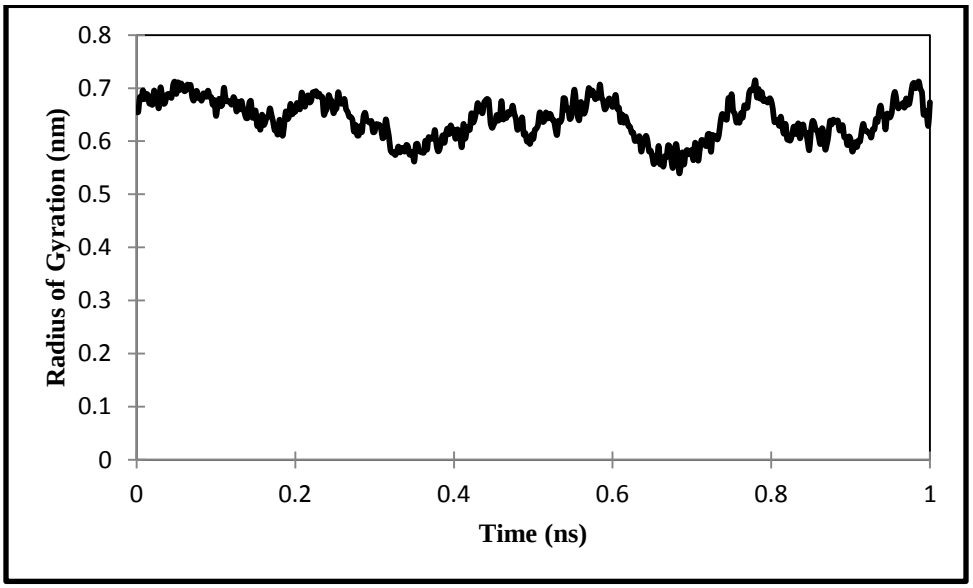

Figure 16: MUC1 Peptide Radius of Gyration during a 1ns Simulation

4.3 SM3 Antibody Complex

In solving for the antibody complex there were many simulation issues encountered. The protein pdb structure did not readily match the conformation spectrum given in the GROMACS force fields. Atoms appeared to be missing from the structure. Using VMD's psf converter application the missing atoms in the system were added which increased the number of atoms in the system to 6269. This increase in size made the dynamic simulation more time consuming with increased computing time required for the dynamic simulations (see Table 2). A simple cubic periodic cell structure was used as the defined unit cell in all the dynamic simulations. This structure introduced a lot of water molecules into the system. Another structure like a

rhombic dodecahedron would be computationally more favorable and are to be considered in future studies

Visual analysis of the configurations of this structure displayed some unexpected features. When ions were added to the system the peptide structure got detached (see Figure 17A). Within the structure we found several alpha helices and an extended section of beta sheet. When we create a solvent system, sometimes the atoms embed themselves into the protein space which could have interrupted the binding between the peptide and antibody. However during minimization the peptide structure though extended was again closely associated to the antibody and the number of helices has reduced (see Figure 17B). During both NVT and NPT equilibrations, the peptide structure was more closely associated to the antibody (see Figure 17C-D). The structure of the antibody changed slightly between the NVT and NPT equilibration as there was a movement in the location of the helical structures. During dynamic simulation the peptide appeared to be in association with the antibody but was extended out in the space (see Figure 17 E-G).

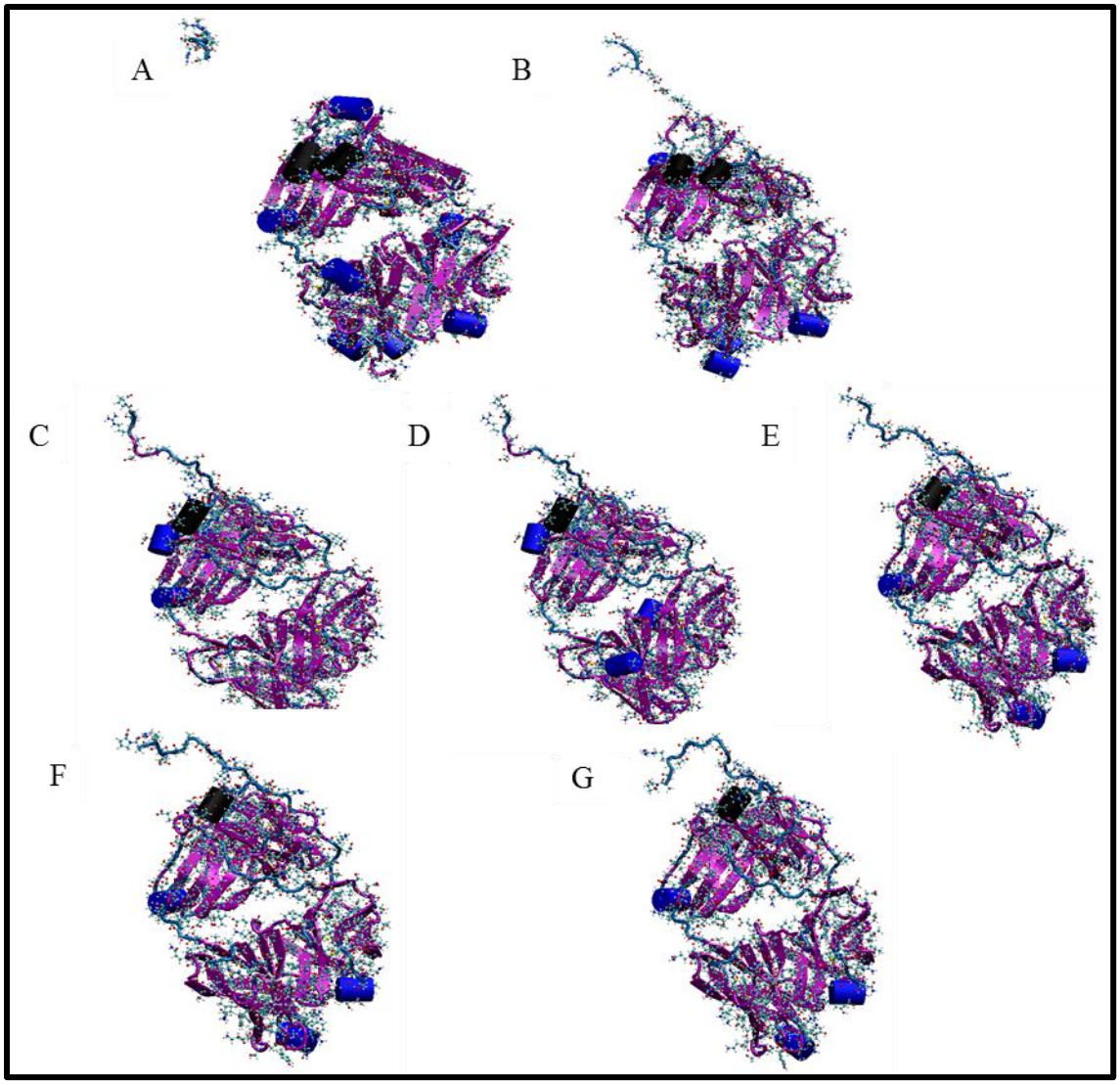

Figure 17: SM3 Antibody Complex A) Configuration after the addition of ions B) Minimized Structure C) NVT Configuration D) NPT Configuration E) Simulation Configuration at 382ps F) Simulation Configuration at 608ps G) Simulation Configuration at 1ns

The minimization process of the antibody complex took much longer than other structure minimizations. The potential energy did reach a minimization near -2700 MJ/mol which took 2.5ns (see Figure 18).

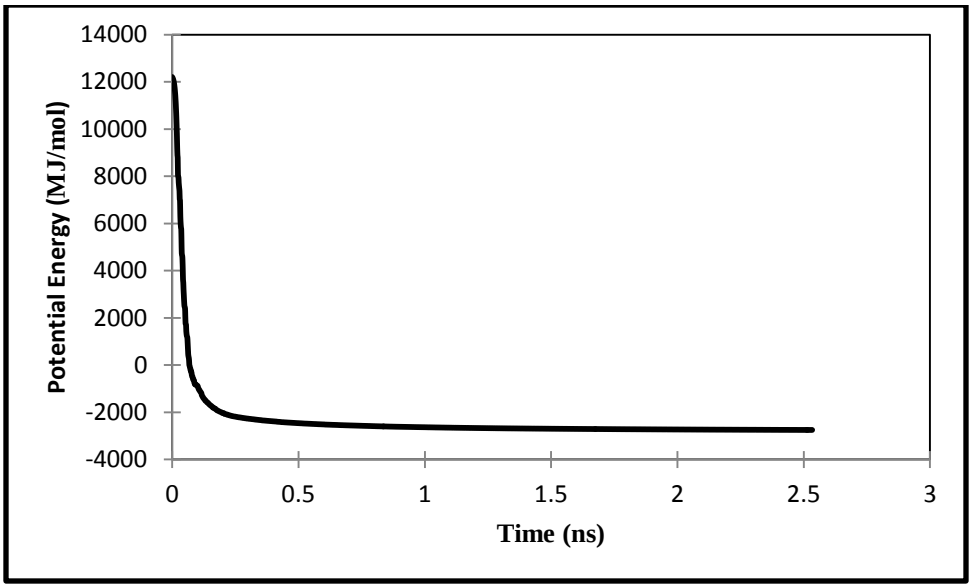

Figure 18: SM3 Antibody Complex Potential Energy during Minimization

Another issue during modeling analysis with this antibody complex was with the configuration of the input mdp files. The time steps for equilibration initially had to be very small because the atoms were not settling. The simulation would terminate within seconds and would only run a few thousand steps before termination. In each run of the NVT ensemble caused a warning of constraints after a time duration of 5ns. The temperature was higher than 300K specified (see Figure 19). This equilibration averaged at 304K.

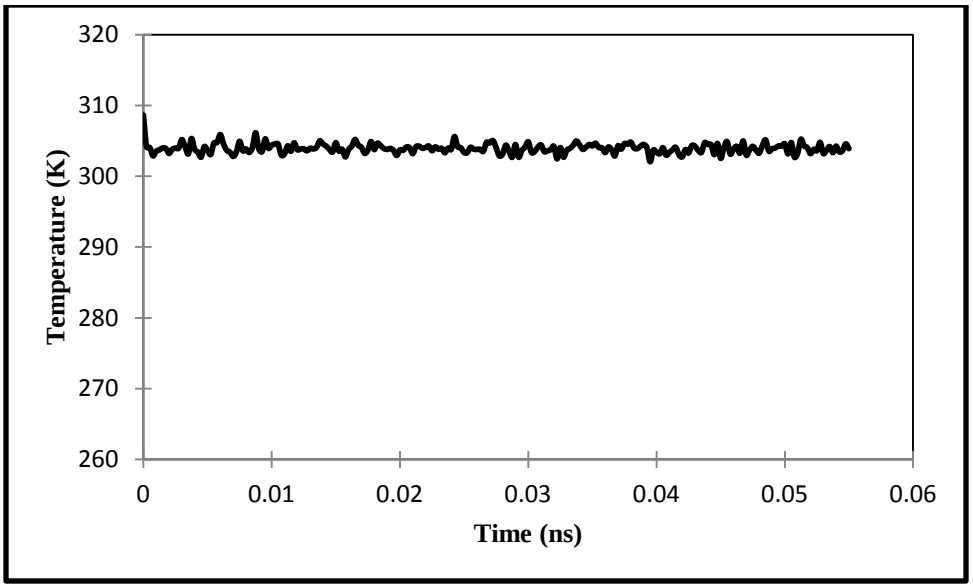

Figure 19: SM3 Antibody Complex Temperature during a 500ps NVT Equilibration

In the NPT runs there were warnings of the pressure coupling which could be attributed to the constraints in the system. This added to the sporadic behavior of the pressure over time (see Figure 20). Because we have constrained the system for the number of atoms, volume and temperature the only way for changes in the system are through energy or pressure, as clearly seen in significant pressure variations

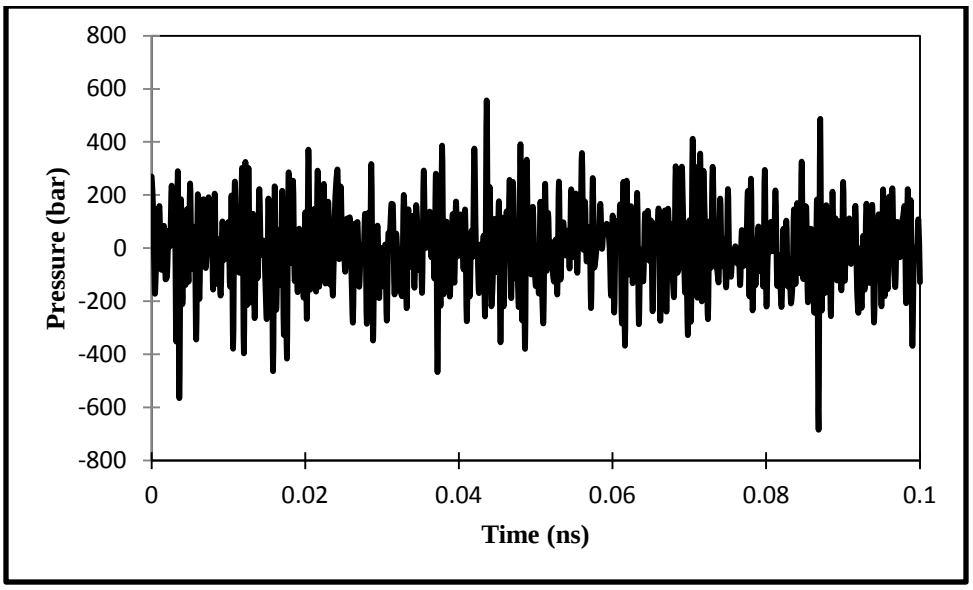

Figure 20: SM3 Antibody Complex Pressure during a 1ns NPT Equilibration

In checking the density of the molecule there was an increase over time which shows that the atoms in the molecule are occupying a smaller space (see Figure 21).

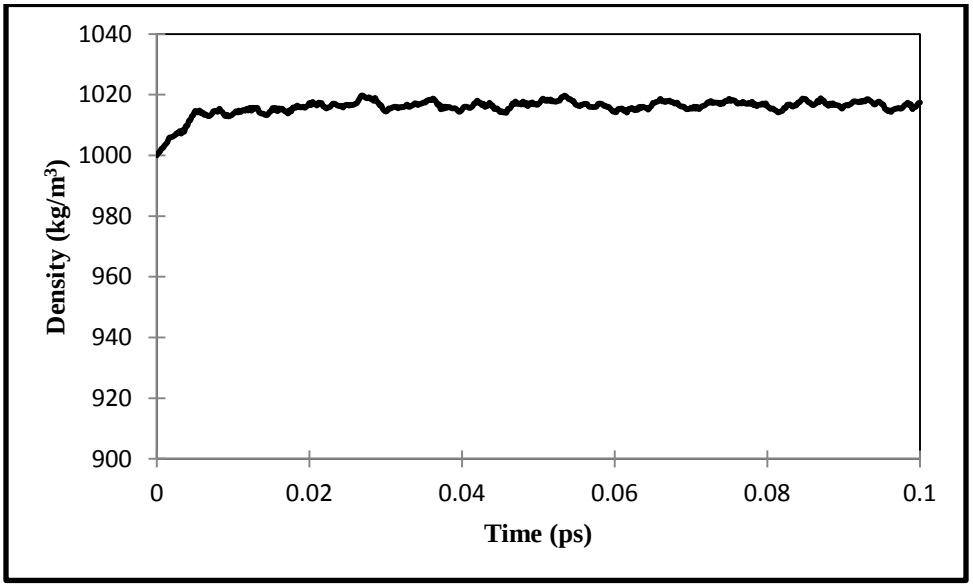

Figure 21: SM3 Antibody Complex Density during a 1ns NPT Equilibration

The RMSD for the Antibody Complex increased throughout the system (see Figure 22). This indicated that during the dynamics simulation, the distance between the atoms in the systems is moving further apart. Clearly, this complex system has not reached convergence yet. Longer time duration dynamic simulation is needed to see if this behavior would continue.

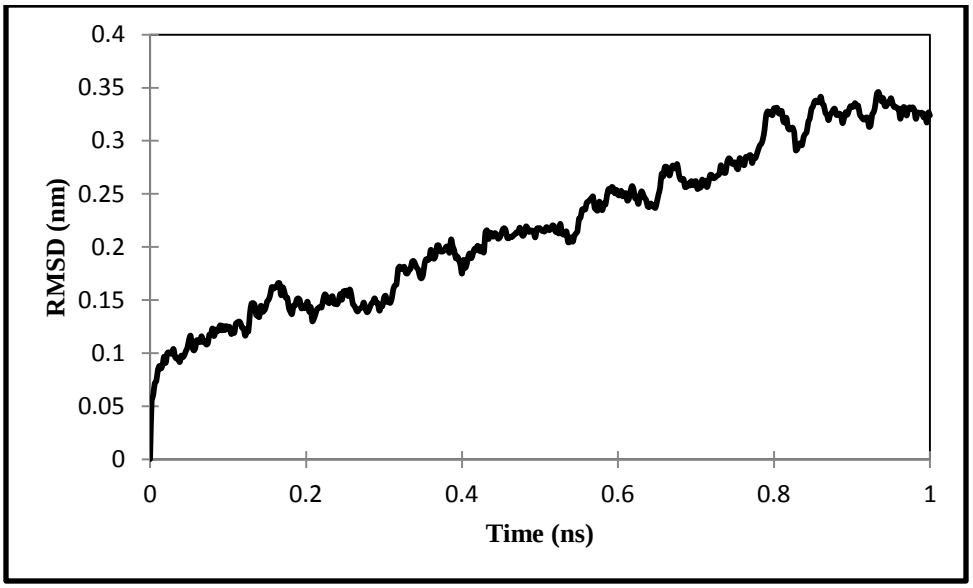

Figure 22: SM3 Antibody Complex Root Mean Square Deviation during a 1ns Simulation

The Radius of Gyration of the Antibody Complex decreased over the dynamic analysis time (see Figure 23). During the dynamic simulation time studied, the system became more compact. This negligible decrease could easily change as the results are still showing transient variations. Without further longer time duration simulation, it is unclear if this behavior will continue.

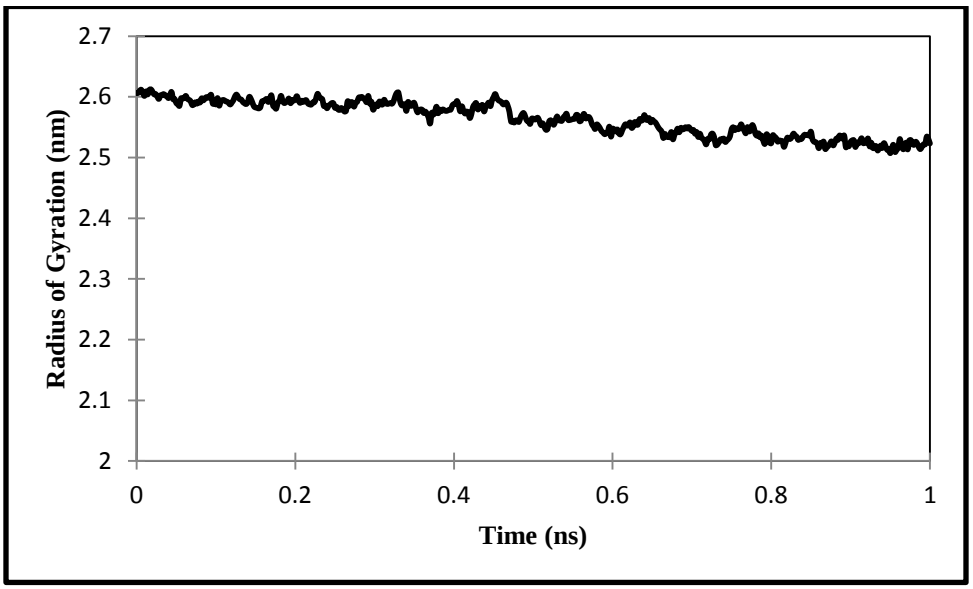

Figure 23: SM3 Antibody Complex Radius of Gyration during a 1ns Simulation

4.4 Modified Anti-MUC1 Aptamer

In experimental selection of aptamers using SELEX, the aptamers are amplified using PCR. The primers added in this reaction are then removed and the double stranded aptamers are separated. As the primers are removed, part of the aptamer may also get removed. This in turn may affect the binding of the aptamer. The effect of this variation is studied next. Within Pymol this was achieved by removing 2 nucleotides from the 5' end of the aptamer to generate a modified Anti-MUC1 aptamer. Visually the aptamer changed very little after the addition of ions and minimization (see Figure 24A-B). During equilibration the structure appeared slightly different as the helices were narrower in comparison to the minimized structure (see Figure 24C-D). During dynamic simulation there was fluctuation in the 3' and 5'ends of the aptamer. In the final frames the ends seem to bend more towards each other and the helical region became much wider (see Figure 24E).

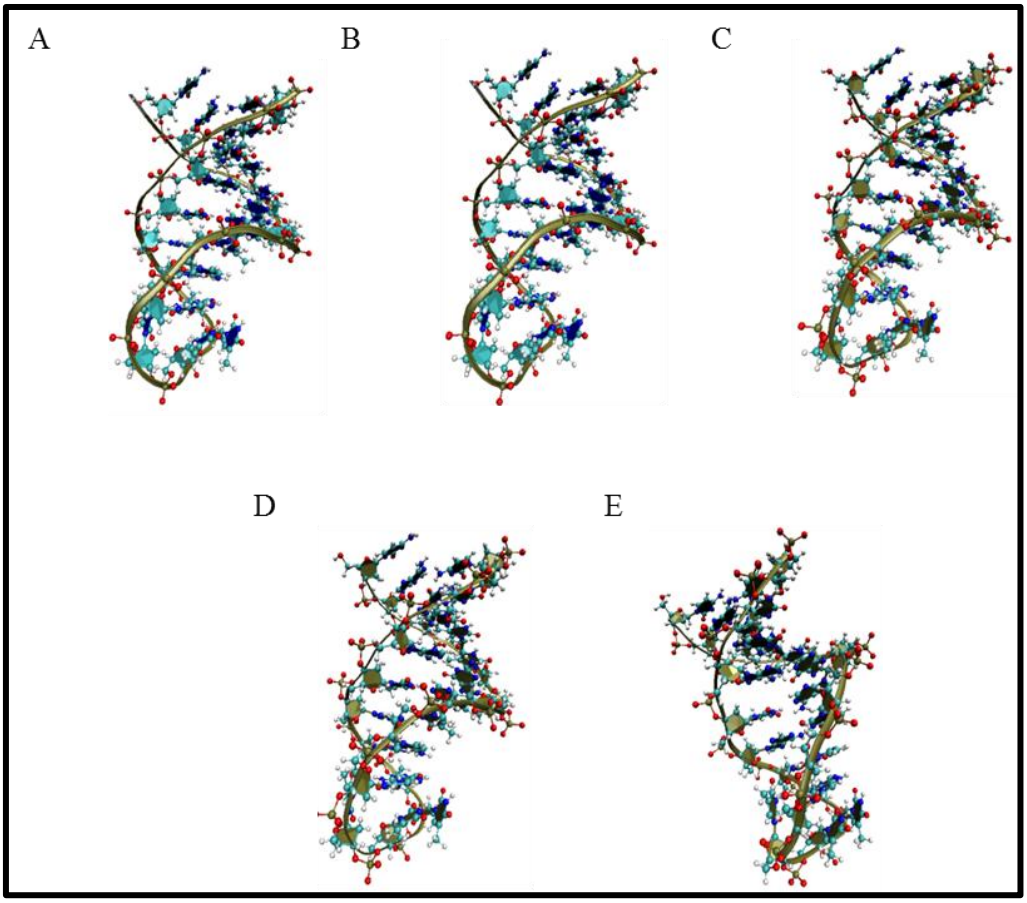

Figure 24: Modified Anti-MUC1 Aptamer Configuration A) After the Addition of Ions B) After Minimization C) After the NVT Equilibration D) After the NPT Equilibration E) After a 1ns Simulation

The potential energy of the modified aptamer decreased but did not show a significant asymptotic flattening region over time (see Figure 25). This could potentially be attributed to the parameters set in the minimization. The minimization stops when the difference between 2 configurations is less than the set value as convergence criteria A stringent convergent criteria would increase the computing time and cost but could lead to a better minimization.

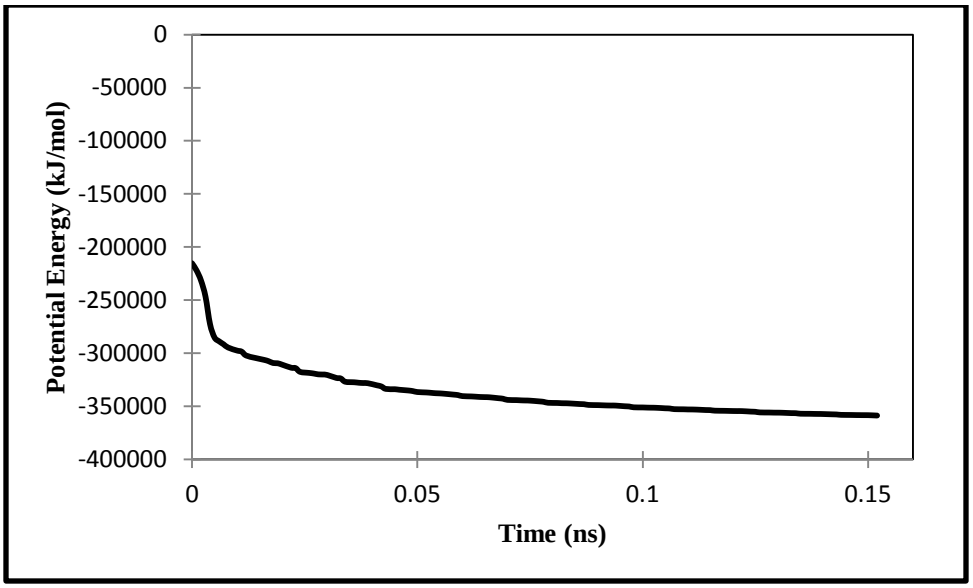

Figure 25: Modified Aptamer Potential Energy during a .15ns Minimization

The temperature during this simulation was very close to the 300K with normal dynamic variations. (see Figure 26).

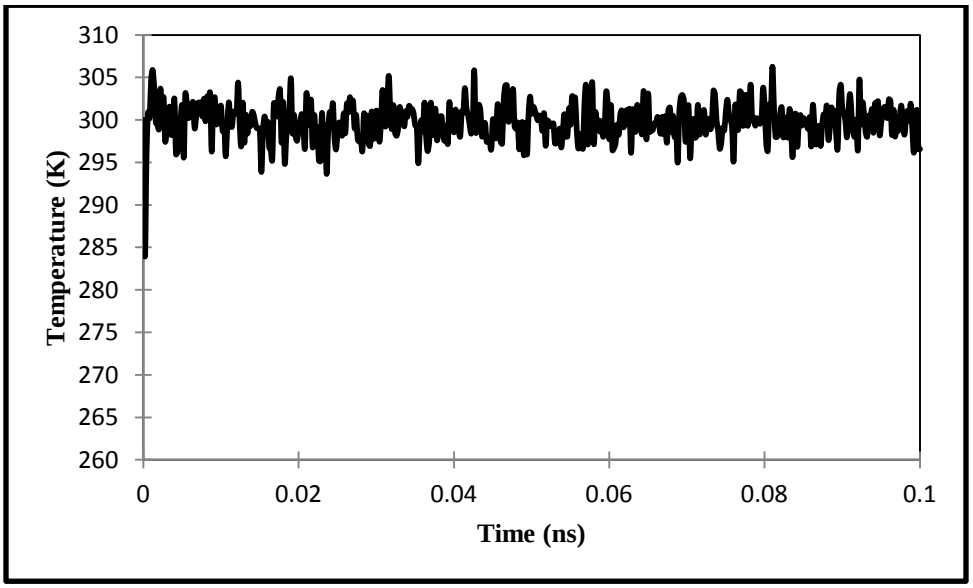

Figure 26: Modified Aptamer Temperature during a .1ns NVT Equilibration

Unlike the temperature, the pressure showed significant fluctuations throughout the .1ns NPT equilibration of the modified aptamer (see Figure 27).

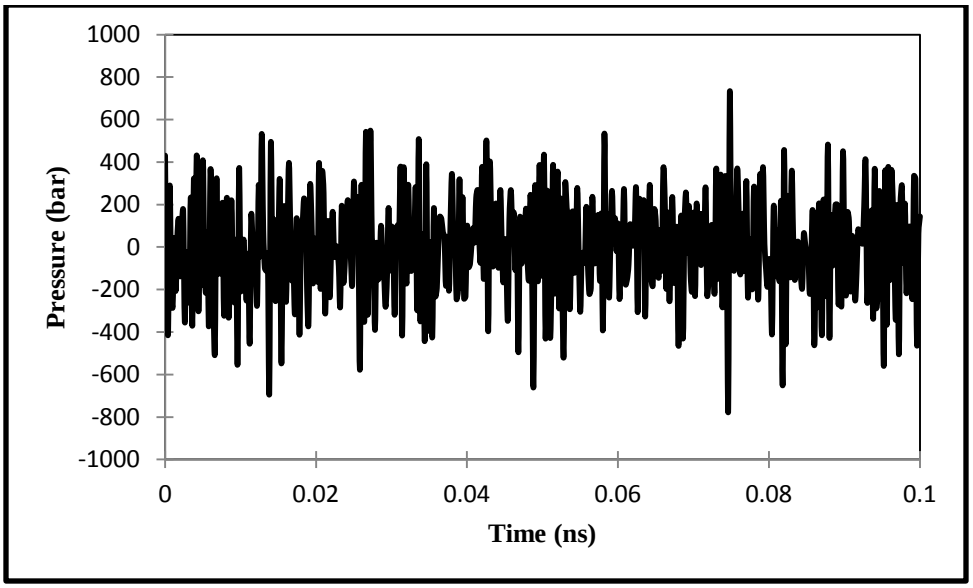

Figure 27: Modified Aptamer Pressure during the .1ns NPT Equilibration

The Anti-MUC1 Aptamer density gravitated around 1020 kg/m^3 during the NPT simulation. For the modified Aptamer the density fluctuates near 1035 kg/m^3 (see Figure 28).

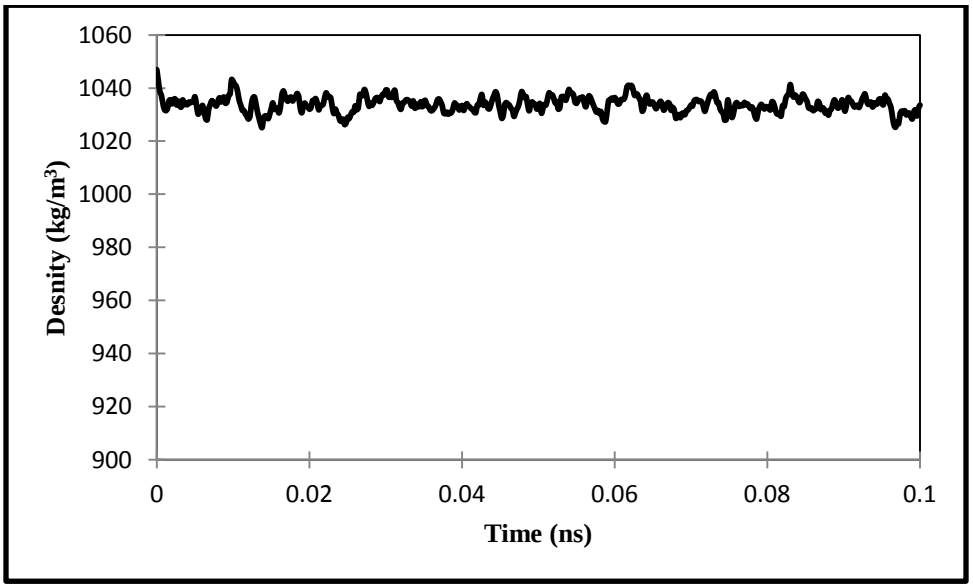

Figure 28: Modified Aptamer Density during the .1ns NPT Equilibration

The RMSD increased over the time of the simulation and continue to show fluctuations over several angstroms (see Figure 29). This increase shows that the distance between the atoms has

increased over the time. However the latter portion of the simulation showed a lot of fluctuation which could be due to the open 3' region of the aptamer.

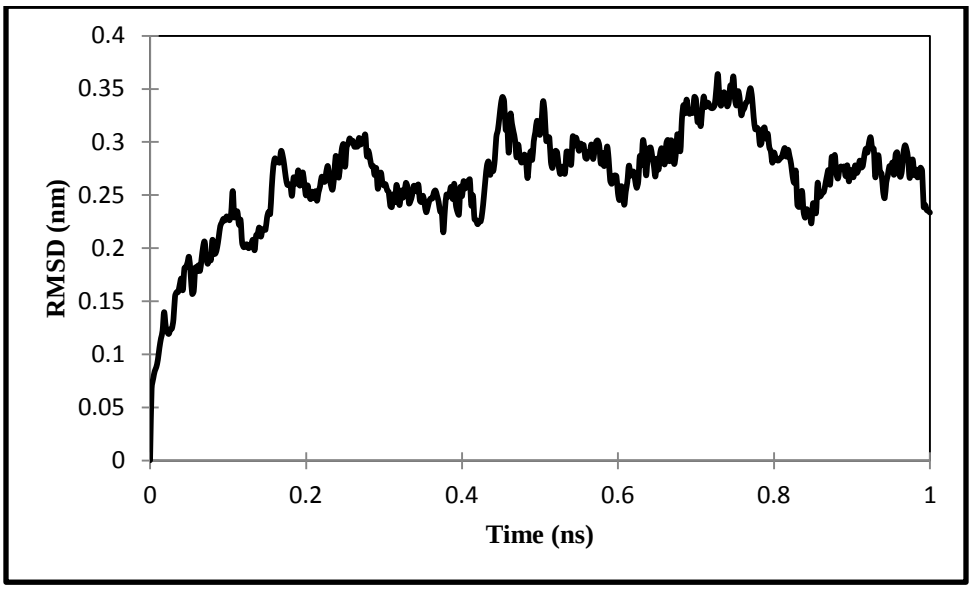

Figure 29: Modified Aptamer Root Mean Square Deviation during a 1ns Simulation

The Radius of Gyration also increased during the simulation in an irregular fashion before showing a slight decrease in the final nanoseconds of the dynamic analysis (see Figure 30). This increase indicates that the molecule has become less compact in the simulation. The decrease towards the end is inconclusive and to understand its behavior a longer time duration for the dynamic simulation is required.

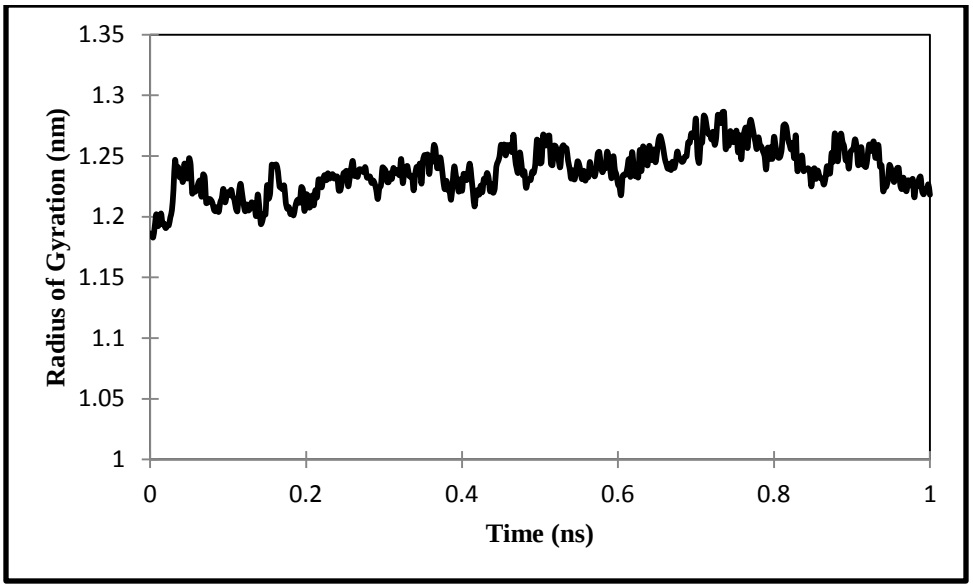

Figure 30: Modified Aptamer Radius of Gyration during a 1ns Simulation

4.5 Anti MUC1 Aptamer-MUC1 Peptide

Now that we have established the individual behaviors of the Anti-MUC1 Aptamer, MUC1 Peptide, SM3 Antibody Complex and Modified Aptamer, we investigate the binding between these molecules. The binding of Anti-MUC1 Aptamer and MUC1 peptide is considered first. The visual examination of the molecular configuration revealed little change in the configuration after solvation; minimization and equilibration (see Figures 31A-D). During the dynamic simulation the aptamer and peptide though initially in close proximity moved apart (see Figure 31E). After 25ns the aptamer associated briefly with the 12th Tyrosine residue in the aptamer loop (see Figure 31F). The peptide then disassociated before quickly associating with the 5'end of the aptamer at 44.6ns (see Figure 31G). The peptide is then taken in by the 3' end of the aptamer and is visually parallel to the aptamer (see Figure 31H). The peptide rotates so that it is orthogonal to the aptamer but remains bound to the open 5' and 3' ends of the aptamer throughout the simulation (see Figure 31I-J).

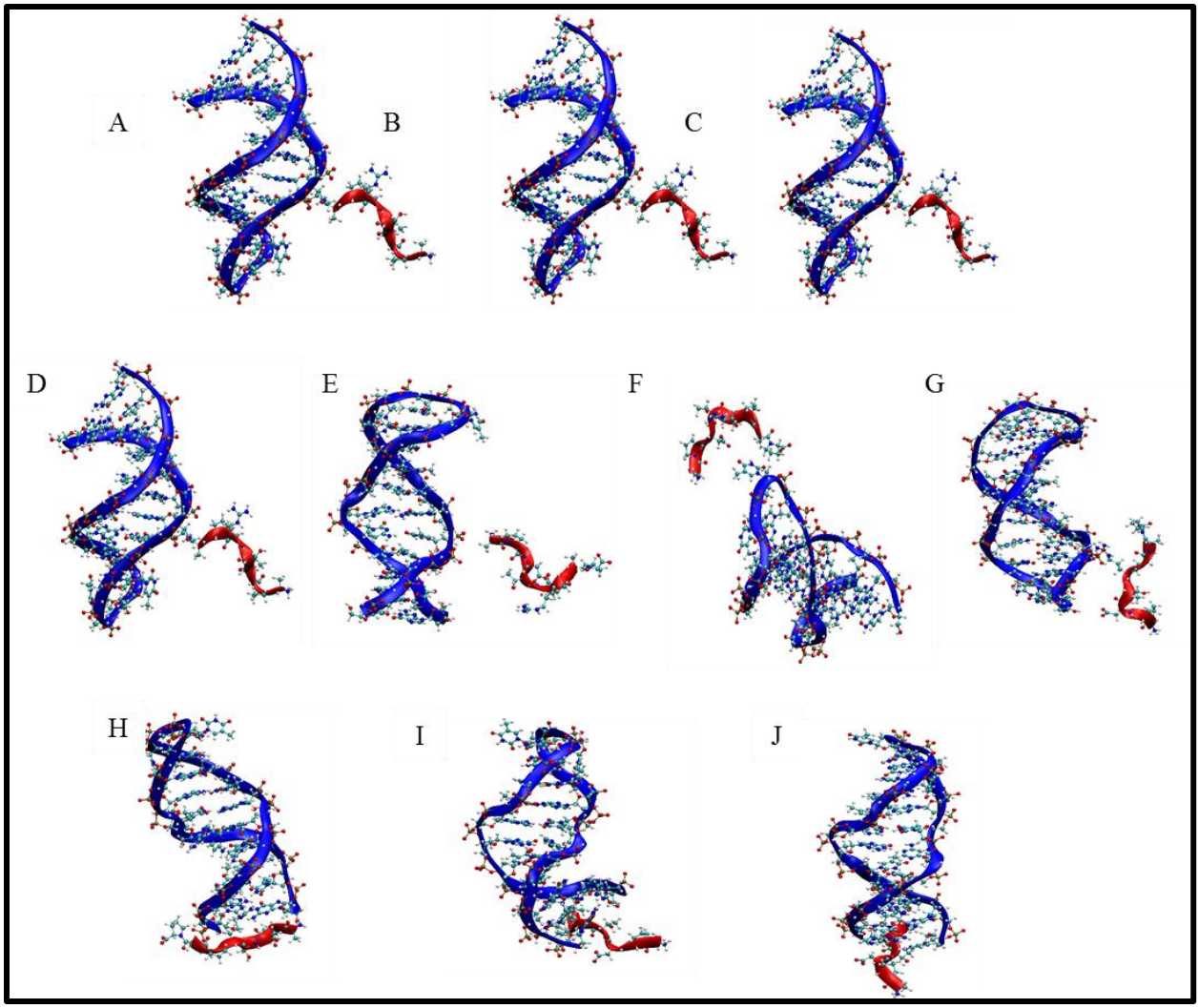

Figure 31: Anti-MUC1 Aptamer (blue) MUC1 Peptide (red) Visual Configurations A) After the addition of ions B) After Minimization C) After the NVT Equilibration D) After the NPT Equilibration E) at 10ns during Simulation F) at 23ns during the Simulation G) at 35ns during the Simulation H) 45ns during the Simulation I) at 76ns during the Simulation J) at the end of 110ns Simulation

The potential energy started at a high value and quickly descended to a plateau. The continuous plateau at a specified value indicated that the change is energy was minimal between configurations and a stable state was reached by the molecule (see Figure 32A). This high initial

potential could be due to the steric interactions involved in the initial state of simulations that used arbitrary starting configurations. Visually these changes were slight but small changes in the angle between the 5' and 3' ends and the loop region could have attributed to this high initial potential. In this simulation there were 6 initial energy values that were very high. Removing these values allowed a better visual on the behavior of the potential (see Figure 32B) and shows a quick asymptotic flat region in a short period.

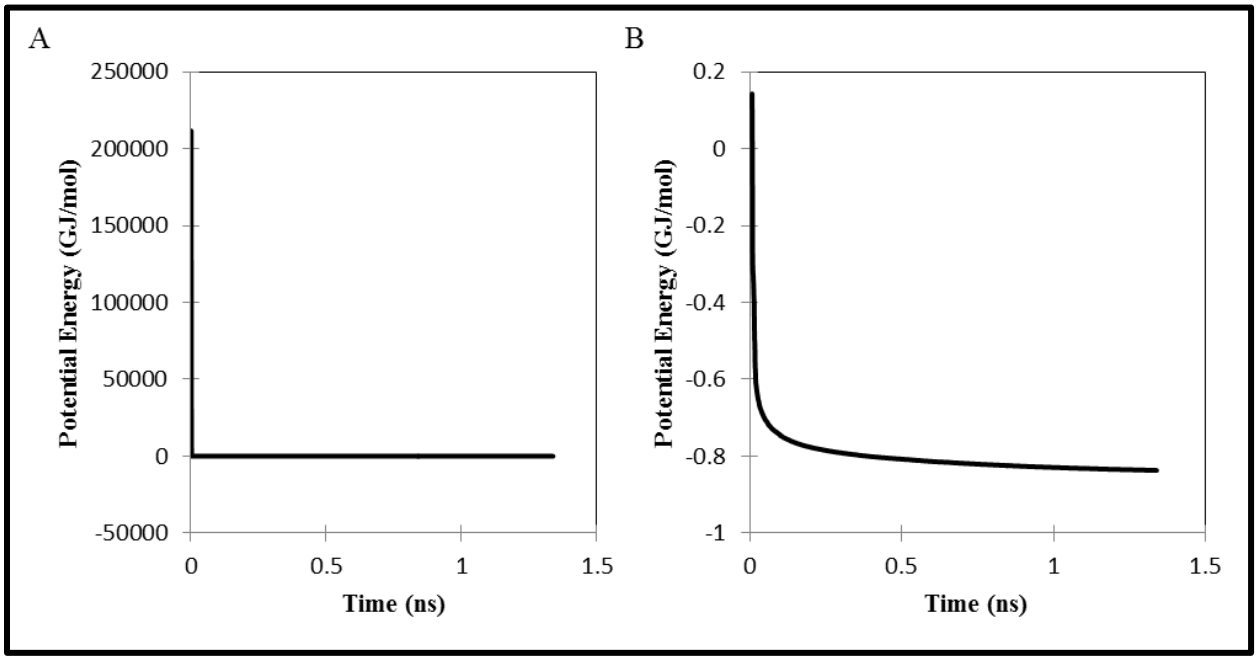

Figure 32: Anti-MUC1 Aptamer and MUC1 Peptide A) Potential Energy B) Potential Energy with the first five data points removed

The temperature during the NVT equilibration was consistently near the 300K with minor fluctuations around this value. The time averaged value as seen from figure (see Figure 33) is 300K.

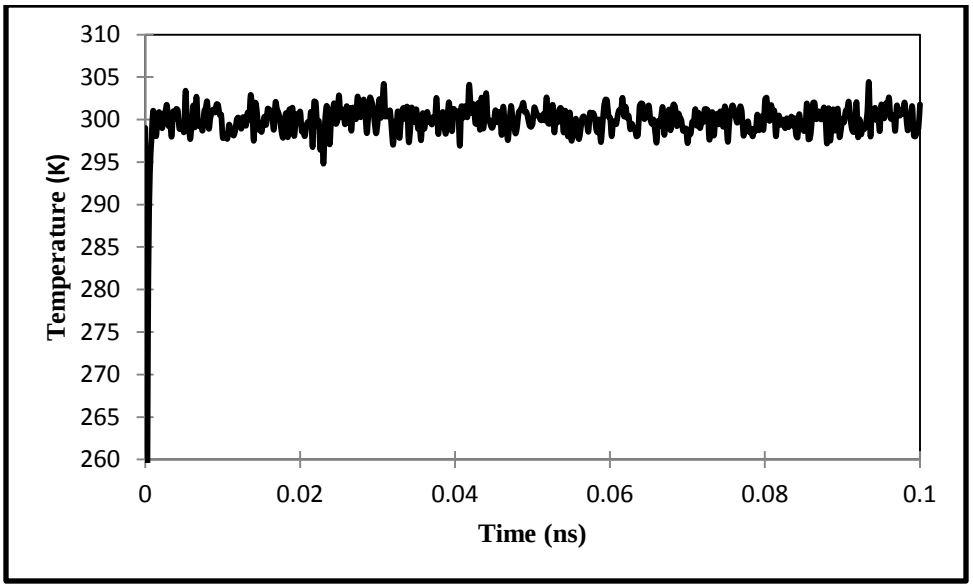

Figure 33: Anti-MUC1 Aptamer and MUC1 Peptide Temperature during NVT Equilibration

Unlike the temperature, the pressure during the NPT equilibration was very sporadic with significant variations though the average was 1bar (see Figure 34).

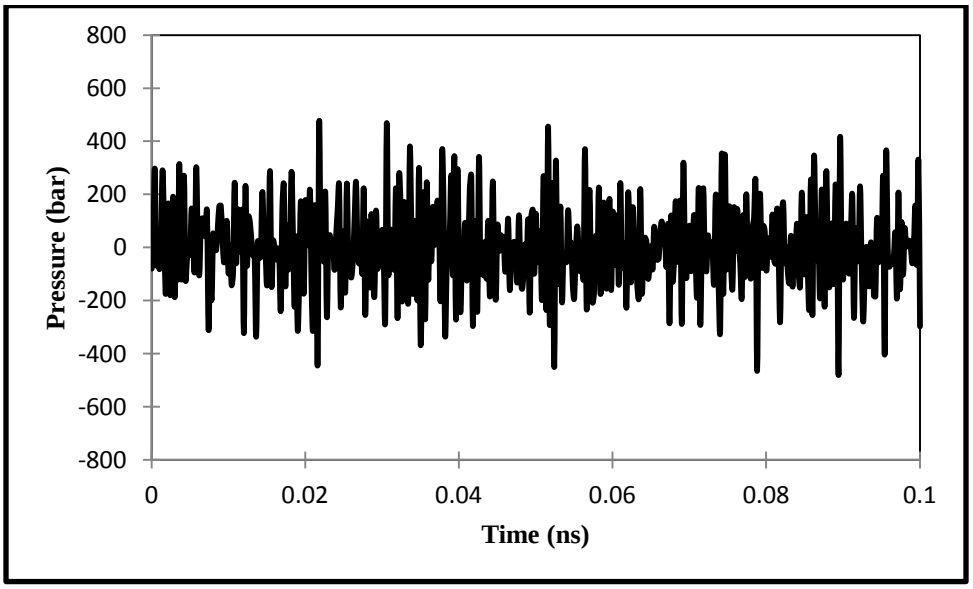

Figure 34: Anti-MUC1 Aptamer and MUC1 Peptide Pressure during the NPT Equilibration

Like the pressure the density had periodic spikes throughout the NPT equilibration (see Figure 35). This could be due to fluctuation in volume as during an NPT ensemble as this system is very transient.

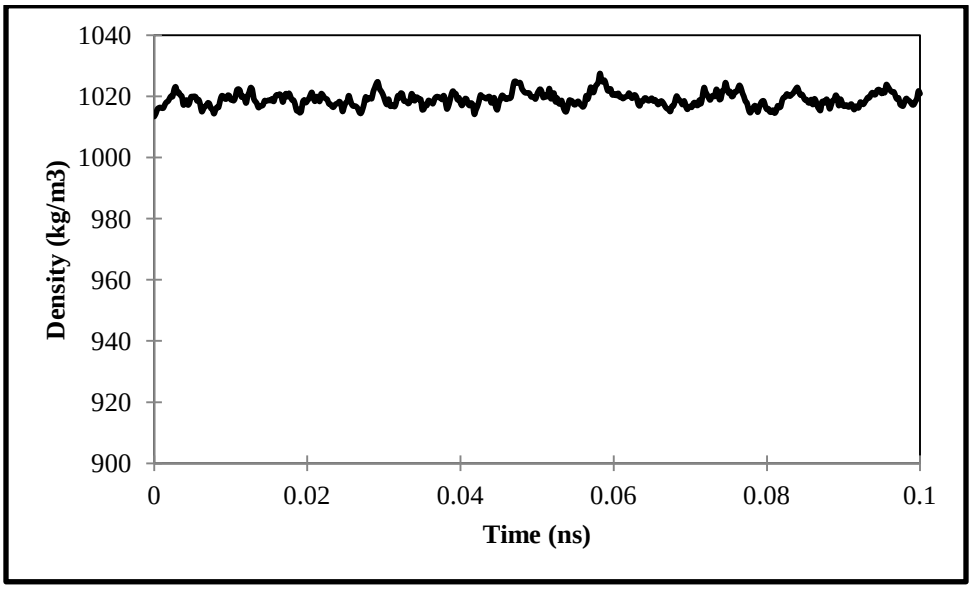

Figure 35: Anti-MUC1 Aptamer and MUC1 Peptide Density during the NPT Equilibration

A closer look at the visual configurations during the two associations between the aptamer and peptide reveals that there were selected atoms that interacted more than others (see Figure 36).

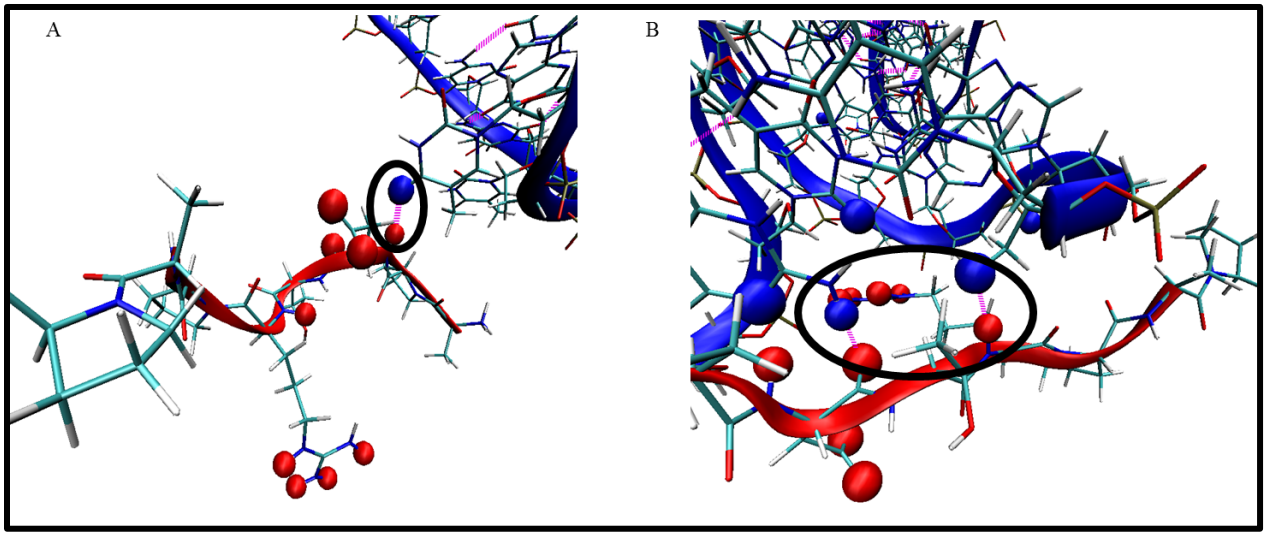

Figure 36: Anti-MUC1 Aptamer (depicted in blue) and MUC1 Peptide (depicted in red) A) during the first association at the 12th Tyrosine residue B) during the second association at the 5' and 3' ends

Figure 37 shows the Surface Plasmon Resonance Imaging (SPRi) results of an Anti-MUC1 aptamer covered surface interacting with MUC1 peptide. In this, there is a first small peak that indicates the surface effect of surface preparation followed by an increase in the reflectivity as the aptamer and peptide bind. The first association, at the top loop combined, with the binding at the 3’ and 5’ ends of the aptamer observed in the simulations clearly coincide with this increase in reflectivity.

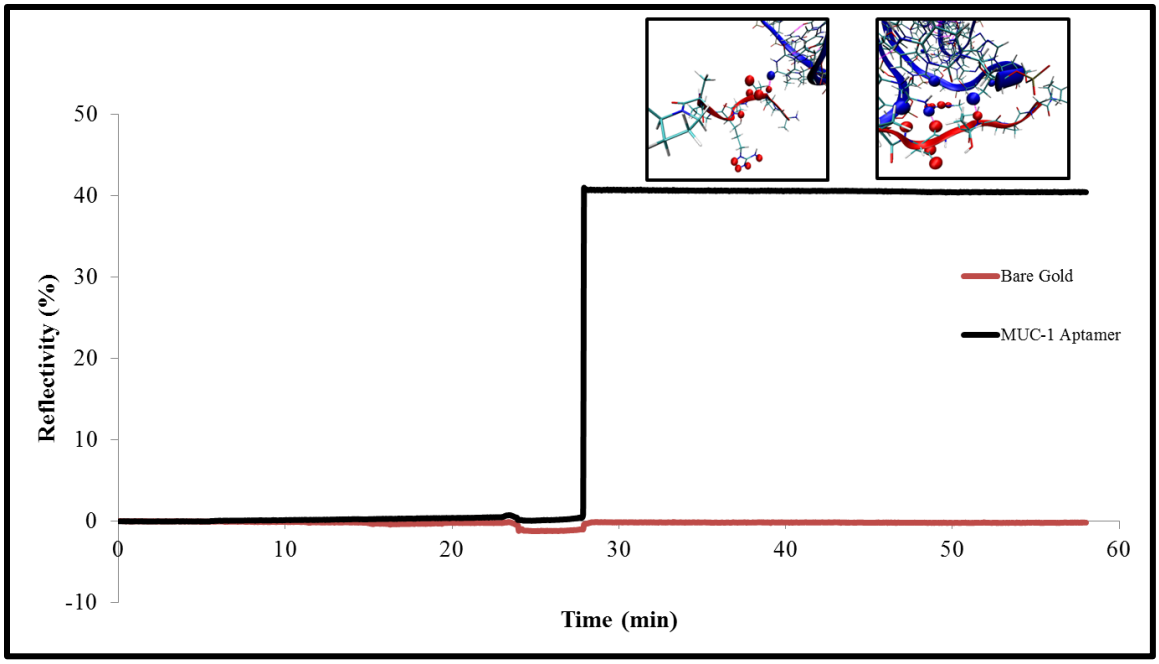

Figure 37: SPRi Analysis of a MUC1 Aptamer Surface and a Flowed Sample of MUC1 Peptide

It is possible to visually track the movement of the hydrogen bonds during the simulation. In carefully tracking these bonds, we observe that the open region in the helical loop forms a bond with the peptide. This bonding occurs for less than 25 ns before the peptide moves away. A second bonding occurs at 45ns. Two hydrogen bonds form as the peptide rotates and moves but often form and dissociate and subsequently reform. Though we can see and track these bonds it is possible that other non-bonded interactions keep the aptamer and peptide together as both molecules constantly move. Binding can occur without chemical bond formation. Though we are looking for binding in the system we still wanted to know if any hydrogen bonds were formed. The number of bonds increased near 25ns and 40ns which is where we observed visually the start of binding (see Figure 38). Figure 38 shows the number of hydrogen bonds formed during the transient dynamics of the aptamer-peptide system. There is a decrease in the number of hydrogen bonds in the latter half of the simulation. Visually the peptide, though still

bound, rotates between the 3’ and 5’ ends of the aptamer. This rotation could result in the breaking of these chemical bonds as the peptide takes this new conformation.

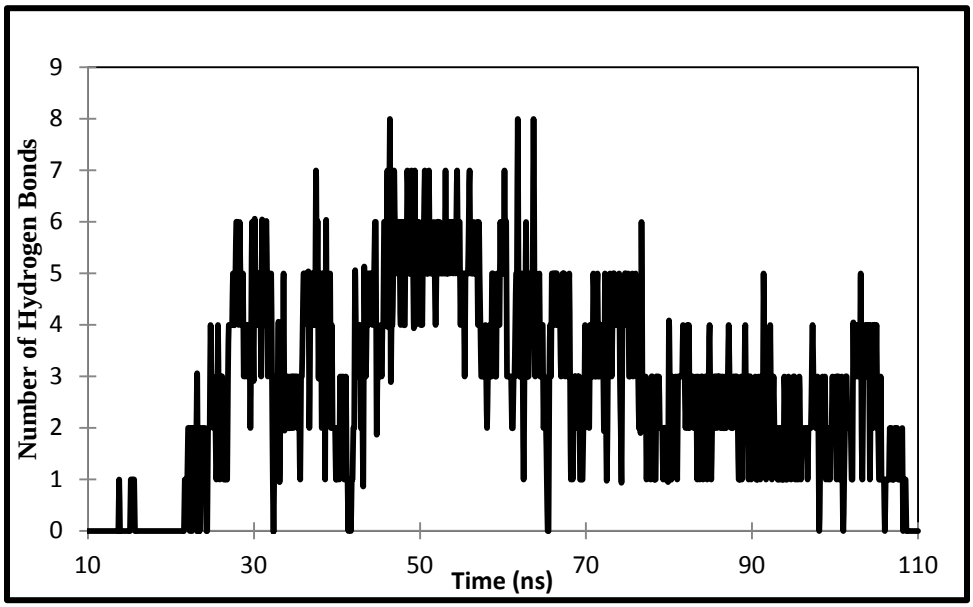

Figure 38: Number of Hydrogen Bonds Formed in the Anti-MUC1 Aptamer and MUC1 Peptide Simulation

To see if binding did occur we check the distance between the peptide and the aptamer. The distance between the aptamer and peptide decreased near 25ns and stayed consistently near 2ns (see Figure 39).

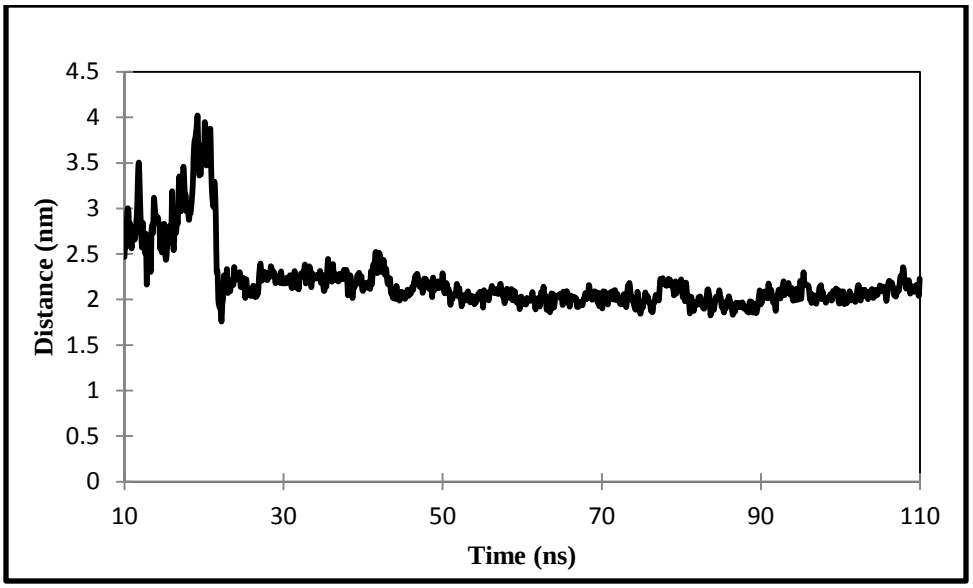

Figure 39: Distance between the Anti-MUC1 Aptamer and MUC1 Peptide Atoms

Though the aptamer and peptide appear to be bonded together, a closer examination was needed to confirm the same. A successful binding causes change in energy, RMSD, and bonds. The RMSD showed that the distance between the aptamer and peptide molecules has increased and stabilized after 25ns (see Figure 40). From the visual results 9 peptide and 6 aptamer atoms were identified that continuously participated in binding throughout the simulation. The RMSD of these selected atoms was calculated. The distance between the atoms also increased. However theses atoms are interacting at a greater distance than the overall system initially (see Figure 41).

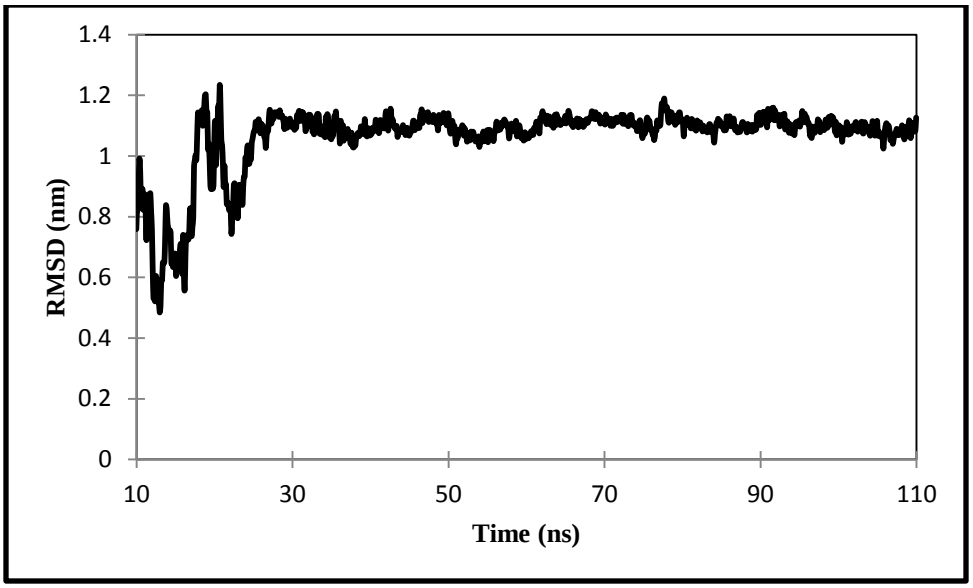

Figure 40: Anti-MUC1 Aptamer and MUC1 Peptide RMSD during 110ns Simulation

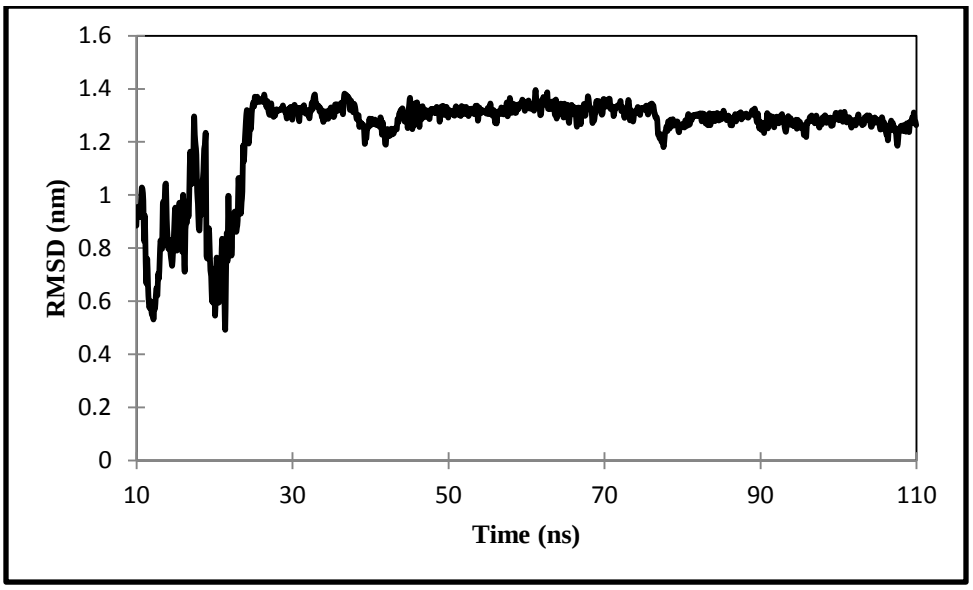

Figure 41: Selected Atoms of the Anti-MUC1 Aptamer and MUC1 Peptide RMSD during the 110ns Simulation

The Radius of Gyration was calculated for the aptamer and peptide and the previously selected 15 atoms. The Radius of Gyration decreased during the simulation for the aptamer and peptide as well as the selected 15 atoms (see Figure 42 and 43). The aptamer and peptide became very

compact and overall the atoms were 1.4nm apart. However the selected atoms were almost .5nm apart showing how close they were. Since binding occurs with atoms in close proximity, this decrease indicates that there is binding between the peptide and aptamer.

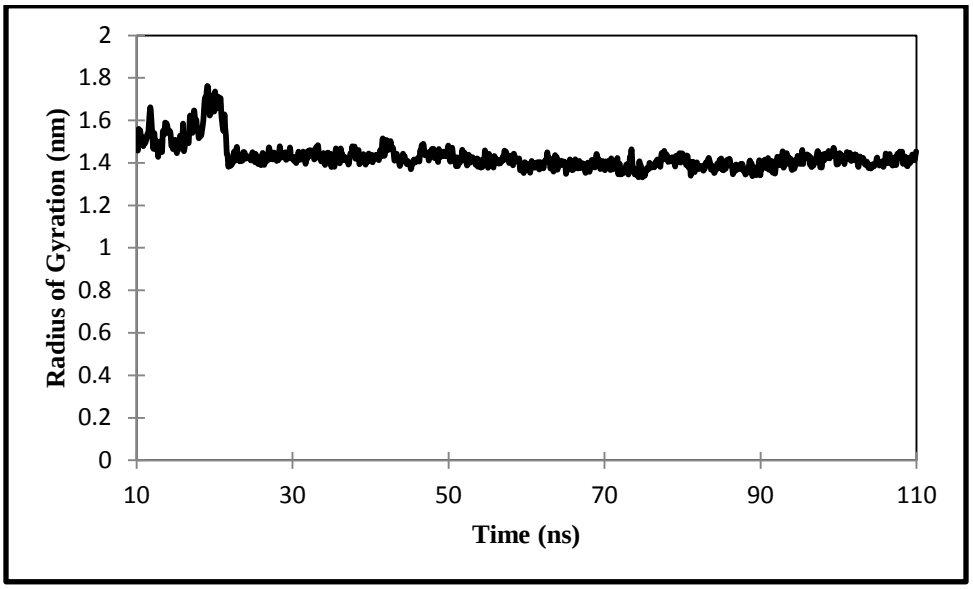

Figure 42: Anti-MUC1 Aptamer and MUC1 Peptide Radius of Gyration during the 110ns Simulation

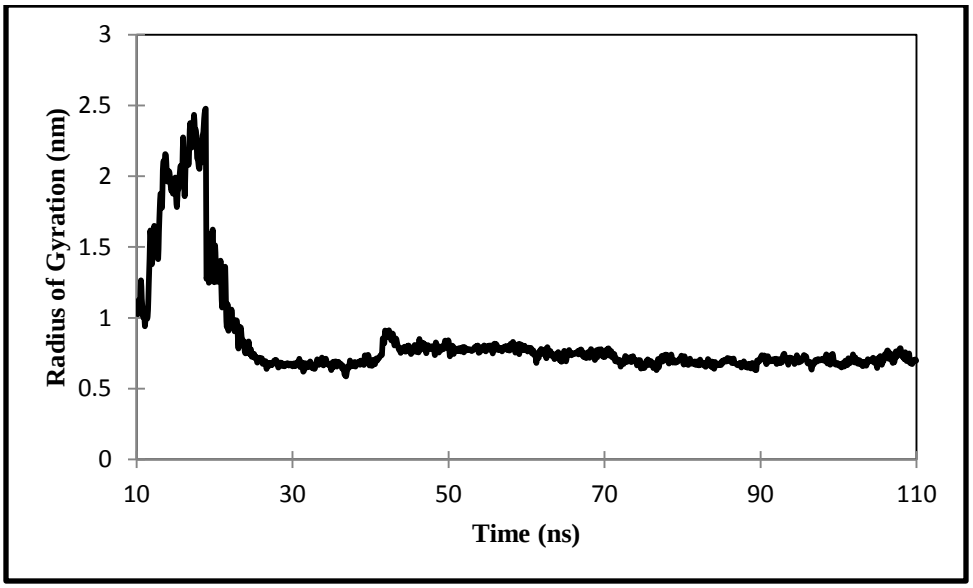

Figure 43: Selected Atoms of the Anti-MUC1 Aptamer and MUC1 Peptide Radius of Gyration during the 110ns Simulation

4.6 Anti-MUC1 Aptamer and SM3 Antibody Complex

When a blood sample is tested for the MUC1 there is a presence of the SM3 antibody. This Antibody can interact with the peptide sequences which could affect the binding of the anti-MUC1 aptamer. Results from Anti-MUC1 aptamer and SM3 antibody complex are presented and discussed next. During the early stages of simulation the peptide detaches from the antibody complex (see Figure 44A-C). During the NPT equilibration the helix becomes more open (see Figure 44D). The peptide is then released from the antibody complex before being taken up again by the antibody complex (see Figure 44F-G). The backbone of the antibody complex is more spread out throughout the simulation. Soon after the aptamer and peptide become closer together (see Figure 44H). The aptamer towards the end of simulation moves away from the antibody complex (see Figure 44I).

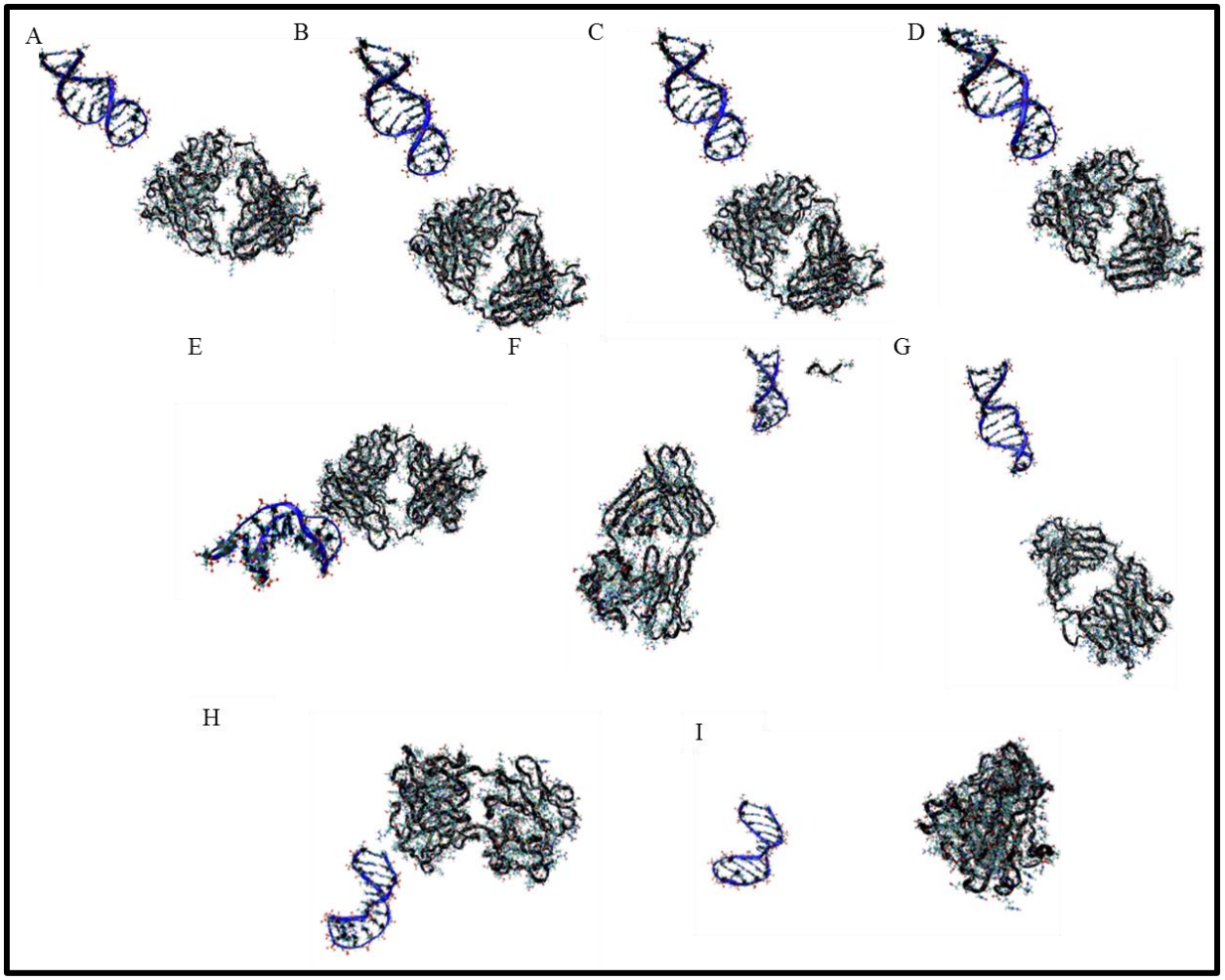

Figure 44: Anti MUC1 Aptamer and SM3 Antibody Complex Visual Configurations A) After the Addition of Ions B) After Minimization C) After a .1ns NVT Equilibration D) After a .1ns NPT Equilibration E) at 1ns of Simulation F) at 3ns of Simulation G) at 8ns of Simulation H) at 10ns of Simulation I) At 15ns of Simulation

Unlike the simulation of the antibody complex alone, this system minimized, equilibrated without the additional need to remove the bond constraints. With the increased size of this system, minimization took much longer and was performed in the multiprocessor computing environment (see Figure 45).

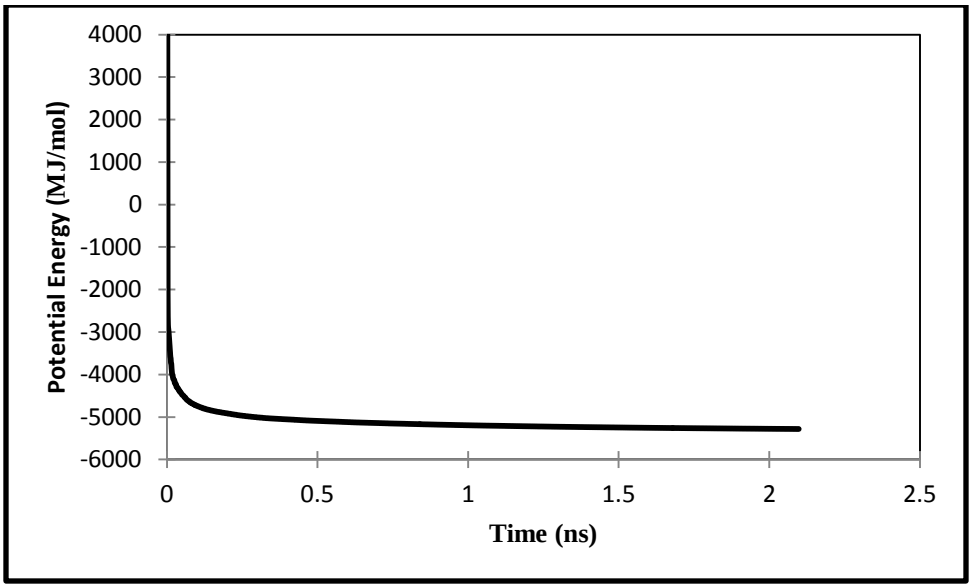

Figure 45: Potential Energy of the Anti-MUC1 Aptamer and Antibody Complex during Minimization

The temperature during this simulation stayed very close to 300K with minor fluctuations (see Figure 46).

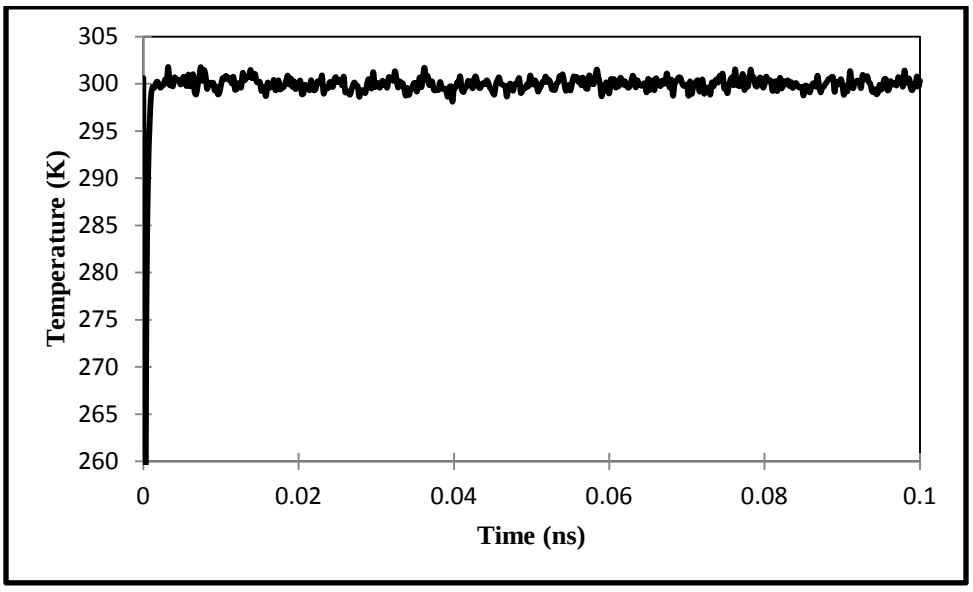

Figure 46: Temperature of the Anti-MUC1 Aptamer and Antibody Complex during a .1ns NVT Equilibration

The pressure was still varied despite the average being near 1 bar (see Figure 47). However when compared to other simulation these fluctuations were between 200 and -200 bar which is a much smaller range.

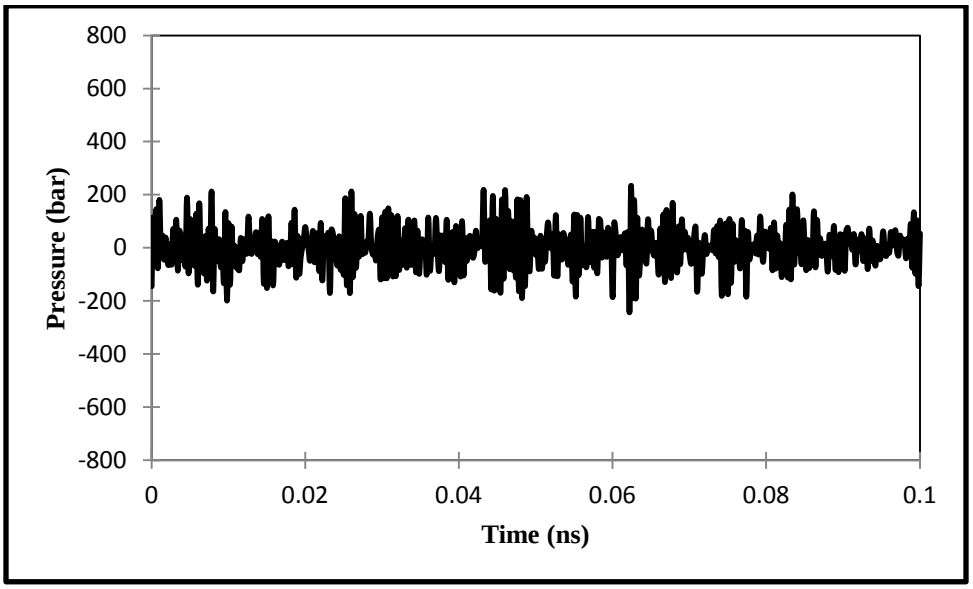

Figure 47: Pressure of the Anti-MUC1 Aptamer and Antibody Complex during a .1ns NPT Equilibration

The density varied between 1012 kg/m^3 and 1015 kg/m^3 but remained constant (see Figure 48). This is lower than other configurations which had densities over 1020 kg/m^3.

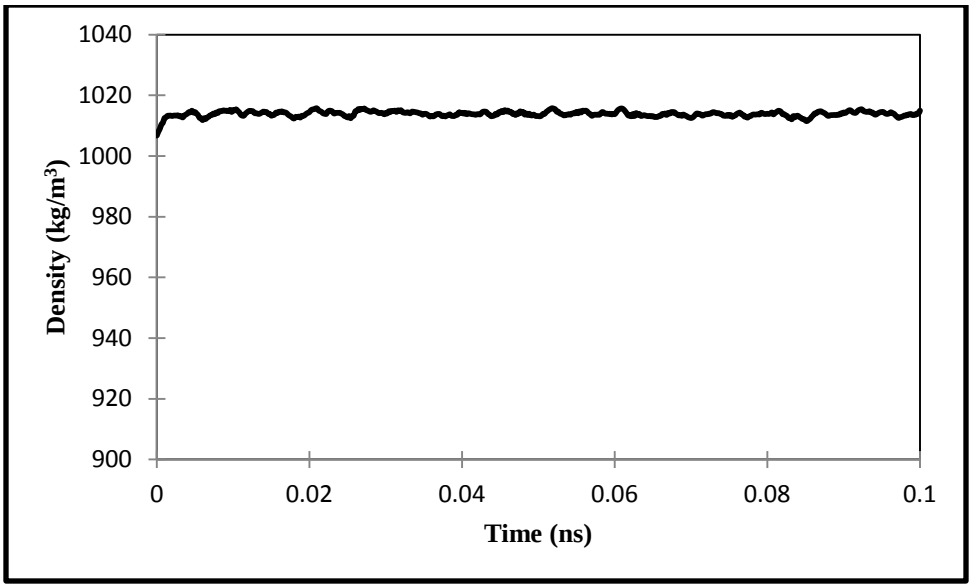

Figure 48: Density of the Anti-MUC1 Aptamer and Antibody Complex during a .1ns NPT Equilibration

The RMSD was unsettled for this simulation system (see Figure 49). Within the 15ns simulation time the distance between the atoms increased before promptly decreasing and increasing again. A longer dynamic simulation run is required to analyze the behavior further and will be continued in future efforts.

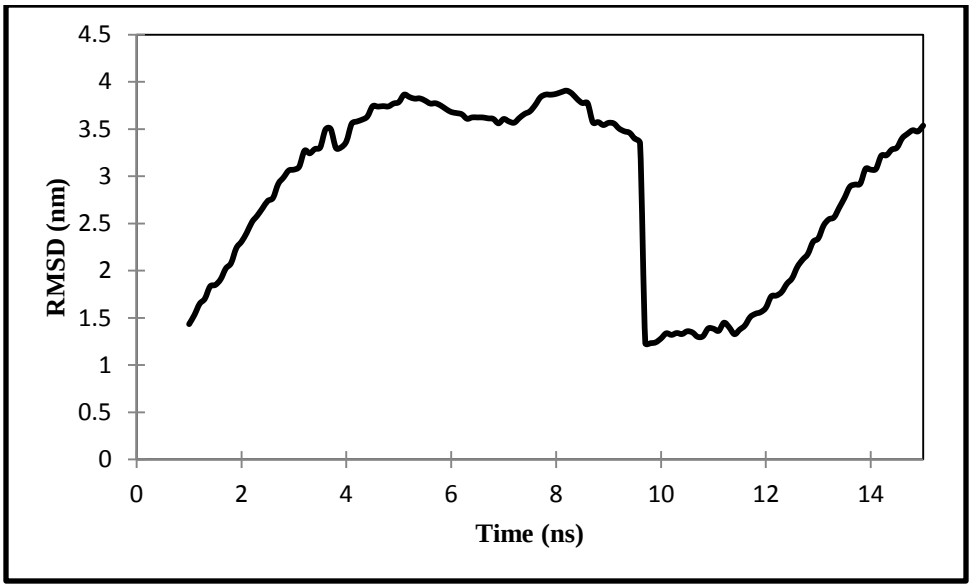

Figure 49: Root Mean Square Deviation of the Anti-MUC1 Aptamer and Antibody Complex during a 15ns Simulation

The Radius of Gyration for this system showed a slight increase before a decrease near 10ns (see Figure 50). This sharp decline was countered by a steep increase which indicates some relaxation in the compactness of the system.

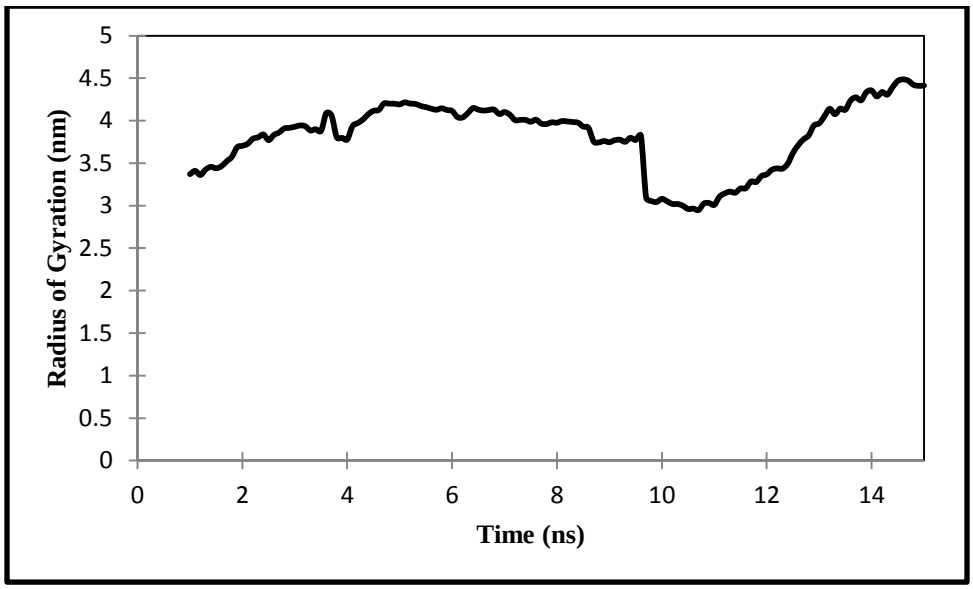

Figure 50: Radius of Gyration of the Anti-MUC1 Aptamer and Antibody Complex during a 15ns Simulation

Since our interest lies in binding, we looked at the distance between the atoms in the aptamer and antibody complex. These atoms were several nanometers apart with an increase in the distance towards the latter end of the simulation (see Figure 51). In previous simulations, binding occurred when the atoms were less than 1.5nm apart. The atoms in this simulation greatly exceed that value.

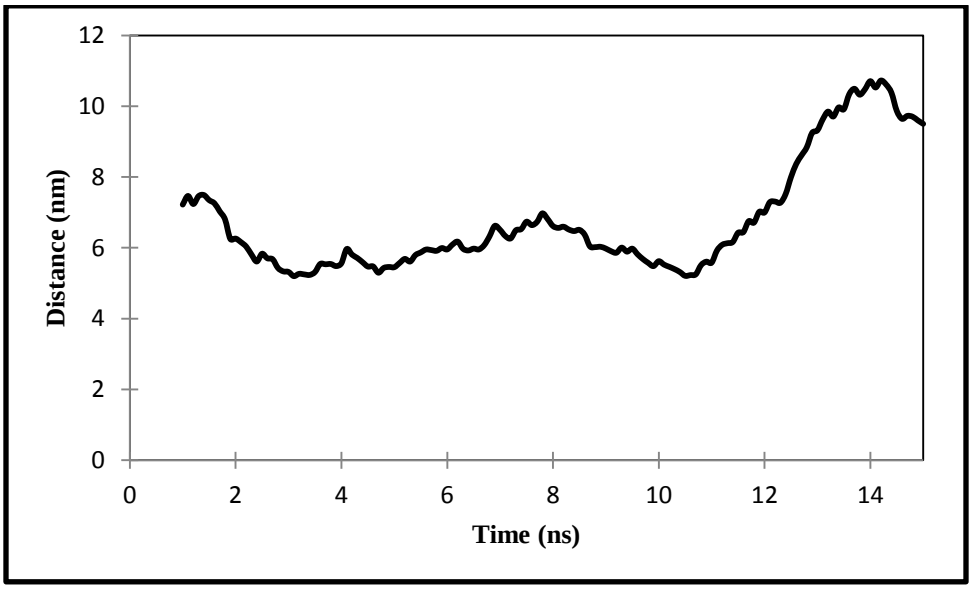

Figure 51: Distance between the Anti-MUC1 Aptamer and Antibody Complex Atoms during a 15ns Simulation

4.7 Modified Anti-MUC1 Aptamer and MUC1 Peptide

As we have established the behavior of the modified aptamer alone, we introduced the MUC1 peptide into the system with the modified aptamer. Visually the aptamer remained in a stable configuration through solvation; minimization and equilibration with the peptide in close proximity to helixes (see Figure 52A-D). At the start of minimization there is rotation of the 12th

tyrosine residue in the loop and the open 5’ and 3’ move toward each other (see Figure 51E). The aptamer and peptide separate before binding is attempted near the 12th tyrosine in the loop structure (see Figure 52F-G). The peptide attempts to bind in the helical region of the aptamer before moving into close proximity around the aptamer (see Figure 52H-I) Though the aptamer and peptide are close together for the second half of the simulation, it is unclear if any sort of binding occurred.

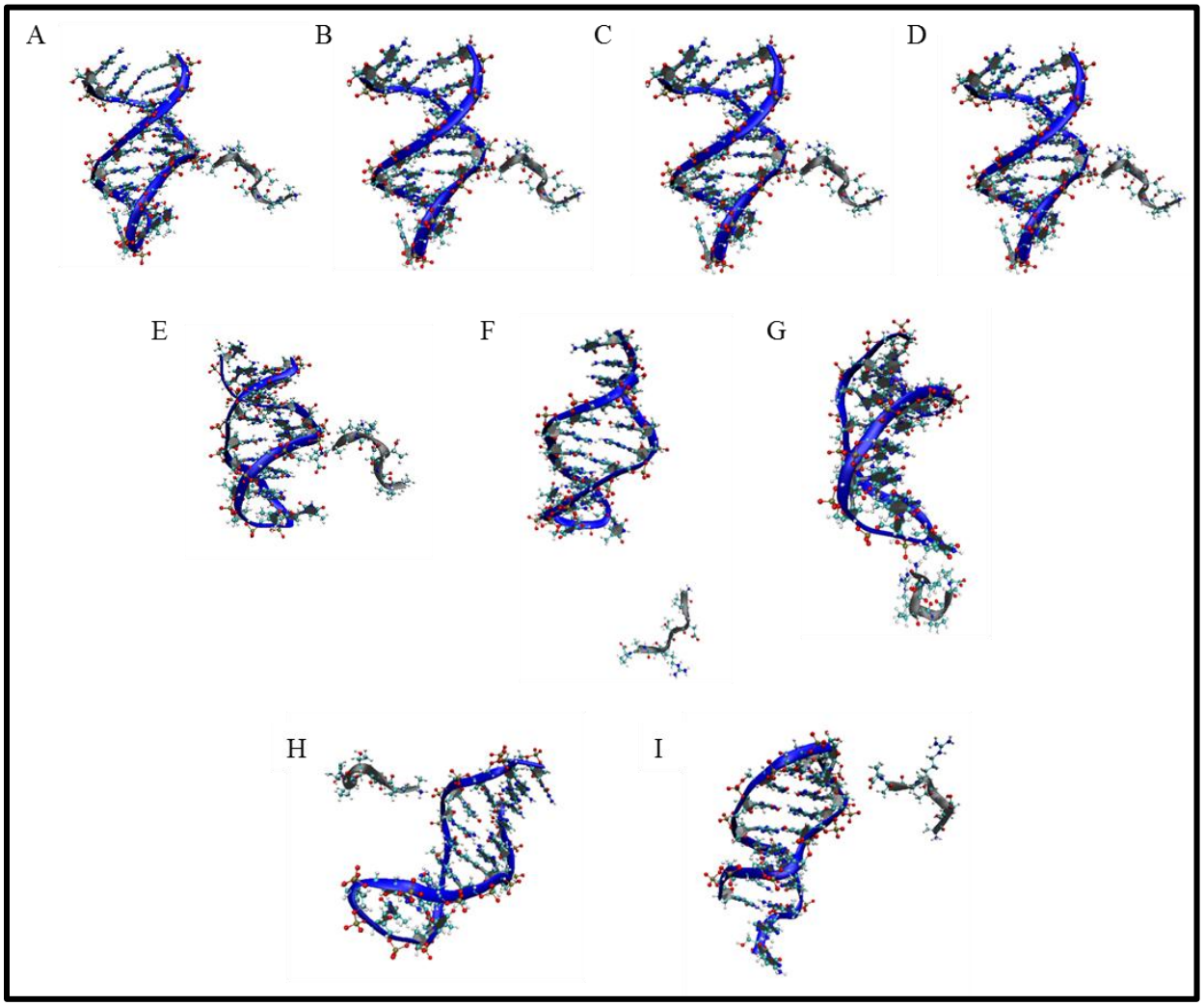

Figure 52: Modified Anti MUC1 Aptamer and MUC1 Peptide Visual Configurations A) After the Addition of Ions B) After Minimization C) After .1ns NVT Equilibration D) After .1ns NPT

Equilibration E) at 1ns of Simulation F) at 11ns of Simulation G) at 34ns of Simulation H) at 56ns of Simulation I) At 110ns of Simulation

Minimization occurred in just over 1ns (see Figure 53A). The arbitrary starting configuration yielded a very high potential energy. To better analyze the data the first 5 data points were removed from the potential and the graph showed the quick decrease followed by an asymptotic minimal leveling of the potential (see Figure 53B).

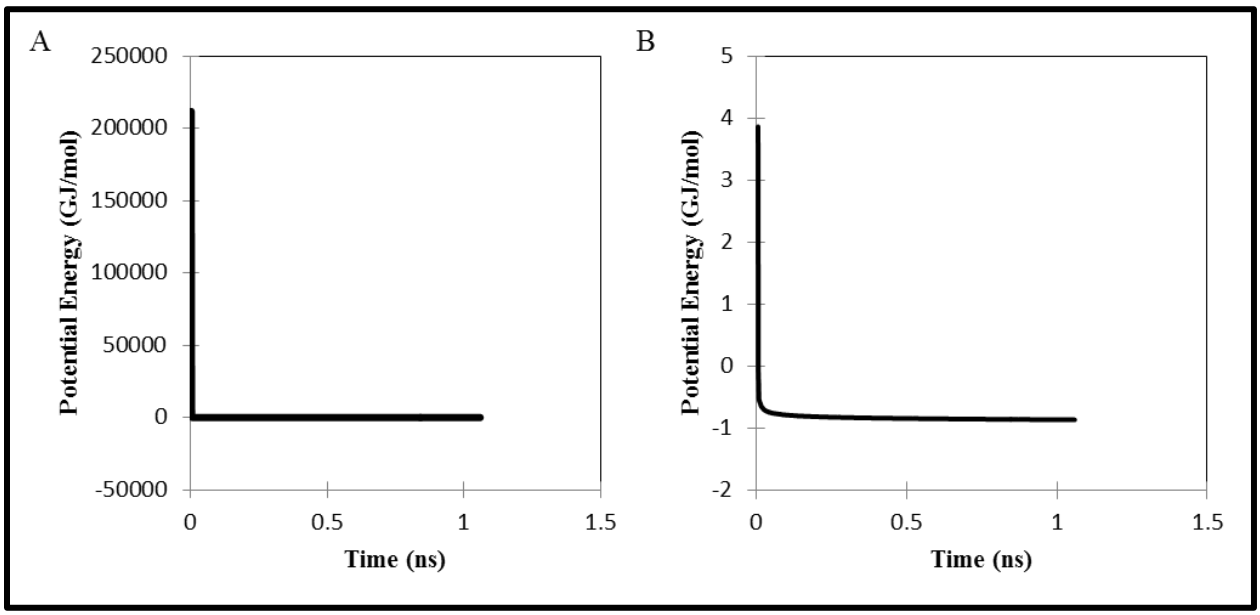

Figure 53: Modified Anti-MUC1 Aptamer and MUC1 Peptide A) Potential Energy during Minimization B) Potential Energy during Minimization with the First Five Data Points Removed

The equilibrated temperature was consistently near 300K (see Figure 54). The pressure like those before was very sporadic thought the average was 1 bar (see Figure 55).

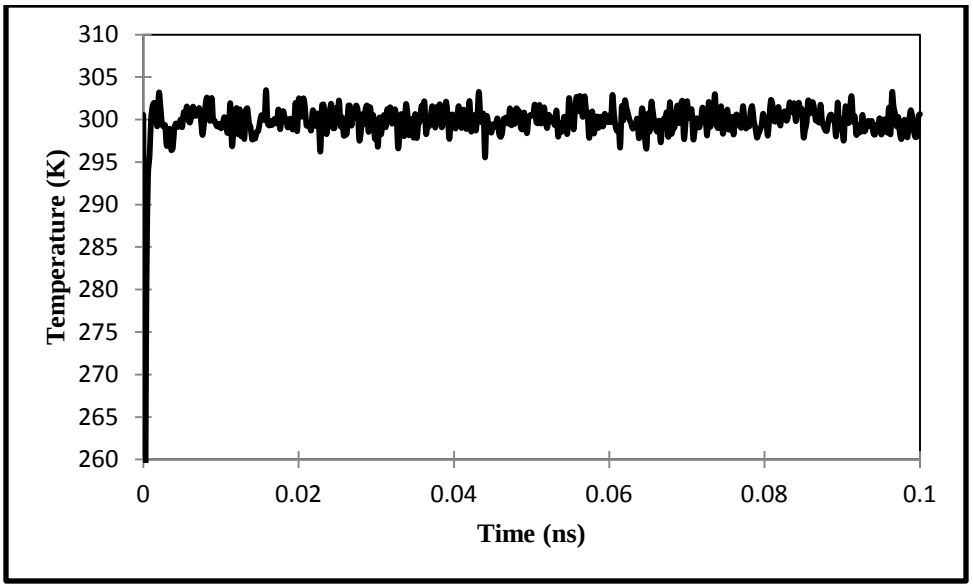

Figure 54: Temperature of the Modified Anti-MUC1 Aptamer and MUC1 Peptide during a .1ns NVT Equilibration

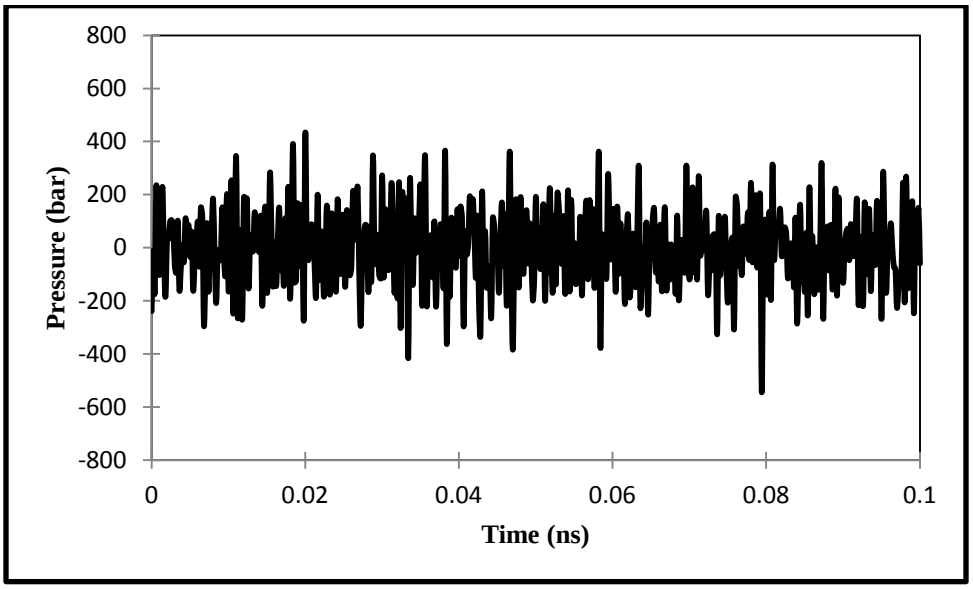

Figure 55: Pressure of the Modified Anti-MUC1 Aptamer and MUC1 Peptide during a .1ns NPT Equilibration

The density of this system varied throughout but its values were slightly higher than those of the previous simulations (see Figure 56).

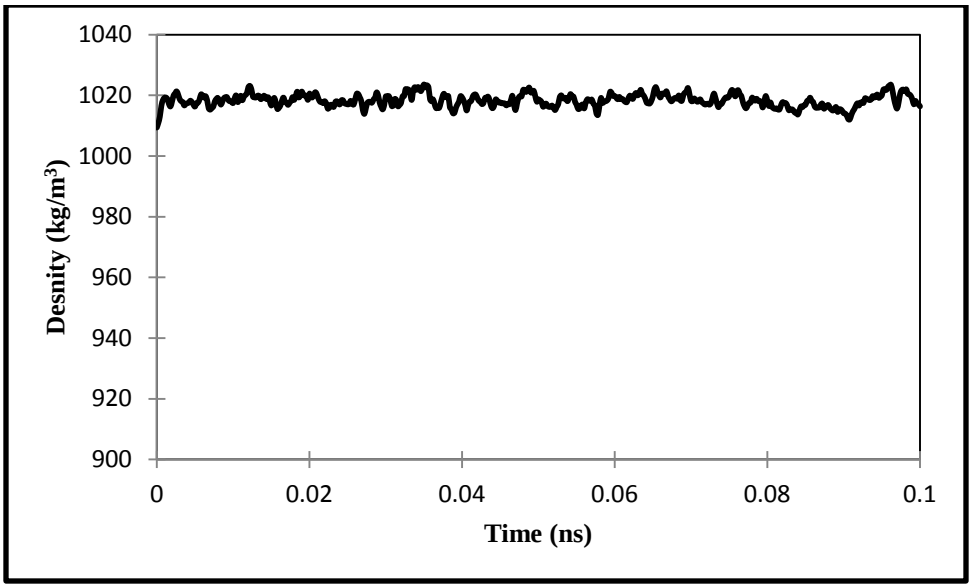

Figure 56: Density of the Modified Anti-MUC1 Aptamer and MUC1 Peptide during a .1ns NPT Equilibration

The RMSD for this simulation had significant variations during the dynamic simulation time period studied (see Figure 57). Towards the middle of the simulation time period, there was a decrease for a short time in the distance between the atoms in the system before an increase was noticed.

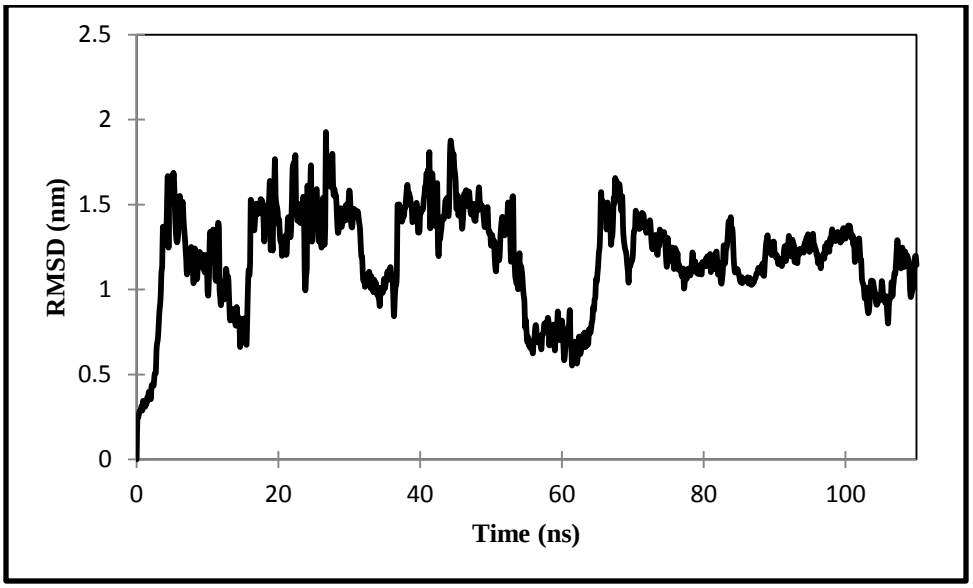

Figure 57: Root Mean Square Deviation of the Modified Anti-MUC1 Aptamer and MUC1 Peptide during an 110ns Simulation

The Radius of Gyration showed noticeable variations at the beginning of this simulation during the dynamic time period studied (see Figure 58). In the latter stages of the simulation the atoms became more compact which could potentially mean some sort of binding was attempted but not sustained.

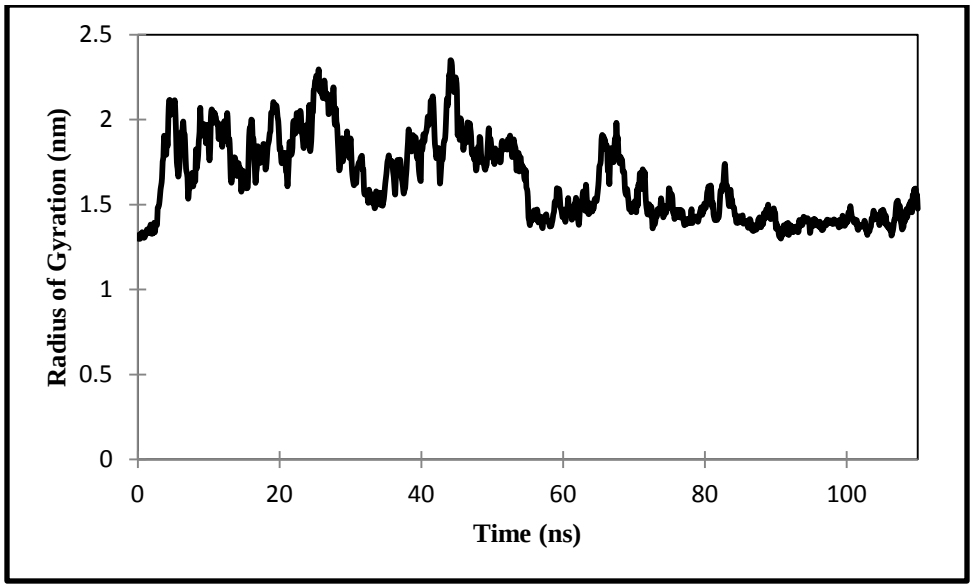

Figure 58: Radius of Gyration of the Modified Anti-MUC1 Aptamer and MUC1 Peptide during an 110ns Simulation

As with the previous aptamer and peptide combination the formation of the hydrogen bonds were explored. Visually the aptamer and peptide were distant in the first half of the current simulation time period studied but in latter points of the simulation the molecules seem to be in close proximity. Very few hydrogen bonds formed in this system and those that were formed were primarily in the latter stage of the simulation time period (see Figure 59).

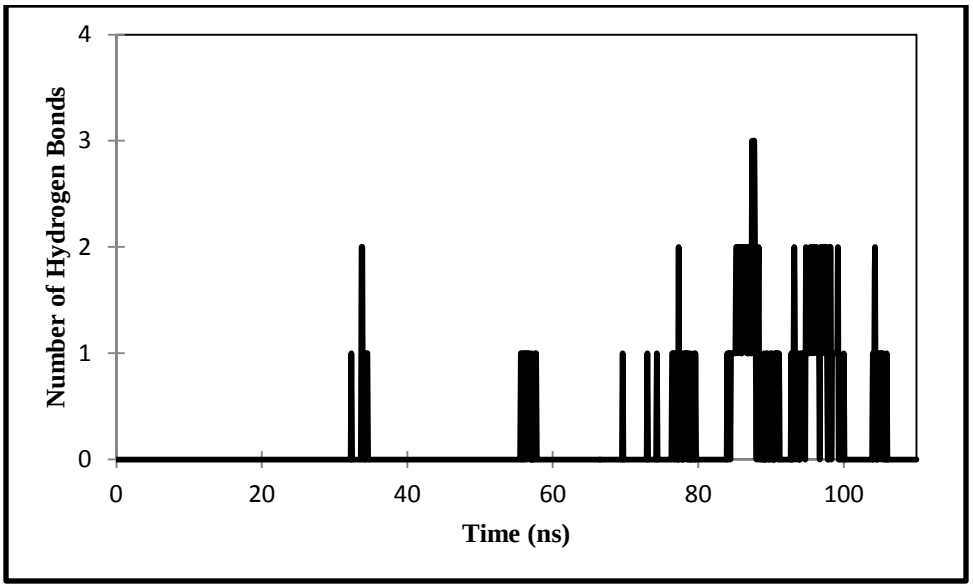

Figure 59: Formation of Hydrogen Bonds between the Modified Anti-MUC1 Aptamer and MUC1 Peptide during an 110ns Simulation

As binding is our goal in this simulation we looked at the distance between the modified aptamer and MUC1 peptide. The distance between the atoms was high and decreased over time but not below 1.7nm as was seen in the Anti-MUC1 Aptamer and MUC1 Peptide combination (see Figure 60).

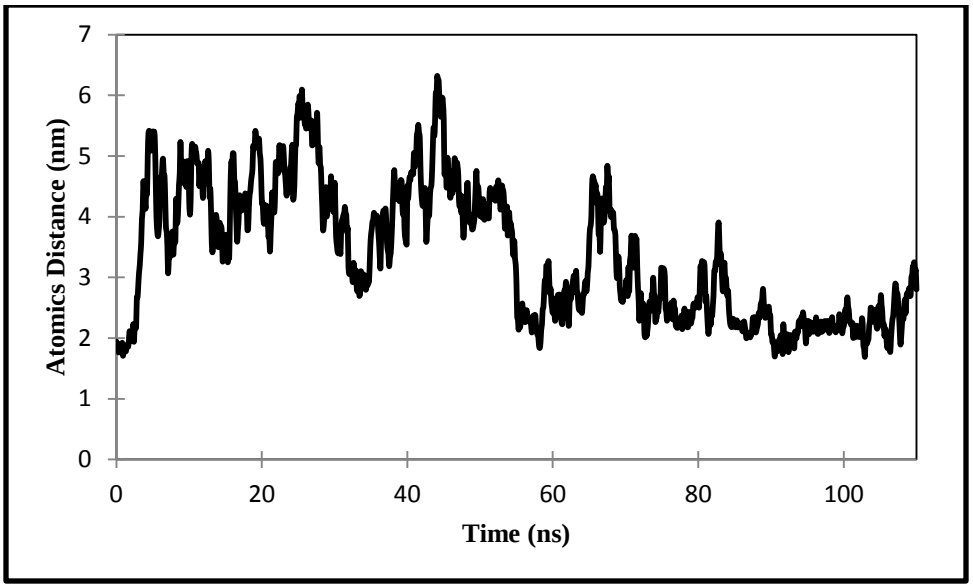

Figure 60: Distance between the Modified Anti-MUC1 Aptamer and MUC1 Peptide Atoms during an 110ns Simulation

Though bonds did form and the atoms did move closer together during this dynamic simulation study, binding appeared to be poor if not nonexistent. For a molecule to be truly bound it would need to sustain that binding over a long period of time. We find that there was no definitive time in which the peptide and aptamer experienced binding interactions.

CHAPTER 5

Concluding Remarks

Computational modeling and simulations based on molecular dynamics modeling could provide effective means to understand the biomarker aptamer bindings that are present as detection mechanisms in biosensors. Using the GROMACS Molecular Dynamics Package the present work simulated the individual behaviors of Anit-MUC1 aptamer, MUC1 peptide, a modified Anti-MUC1 aptamer and the SM3 antibody complex. After establishing these individual behaviors, extensive molecular dynamics simulations of aptamer-peptide binding were conducted. The dynamical simulation results were visually and quantitatively analyzed for the conformational changes and the overall behavior of an aptamer and peptide system was observed from a molecular view point, which is not always possible in wet lab experiments. The analysis and visualization of the natural progression of aptamer and peptide binding can aid in biosensor developments that are based on biomarker – aptamer detection mechanism. Extensive MD simulations performed in the present work has furthered the understanding of Anti-MUC1 aptamer and MUC1 peptide binding. In addition to aptamer peptide binding we simulated aptamer-peptide and antibody binding. Since the MUC1 peptide is likely to be bond to an antibody present in a blood sample we explored Anti-MUC1 aptamer and SM3 antibody complex binding. Though a long time duration dynamic analysis is needed, we were able to observe the behavior of the aptamer, peptide and antibody. The present MD simulations, analysis and discussions clearly show that aptamer and peptide binding could be simulated and analyzed using computational modeling. Preliminary comparison of the dynamic simulation results showed good correlation with the association and disassociation observed in the limited wet lab experiments. The demonstrated ability presented and discussed in the present thesis adds

to the potential to enhance selection of unknown aptamer combinations as well as act as a foundational study for using computational modeling and simulations for aptamer selection in new sensor developments for TBI and mild TBI, where the biomarker is known and identified but the suitable binding aptamer is still under investigation.

5.1 Future Directions

Building upon the present foundational study discussed in this thesis, several additional molecular dynamics based models can be expanded to various environments, conditions and molecular systems of interest. Though we started from the pdb 3D conformation of the aptamer and explored specifically isothermal room temperature and atmospheric pressure conditions, we must also consider wet lab experimental conditions that actually exist. In many cases during the experimental process, heating and cooling cycles are implemented with the aptamer in an attempt to break the bonds of the tertiary conformation and allow for more open regions of binding. However this may not be necessary if the aptamers and targets bind in the same fashion without heating. By incorporating the heating and cooling thermodynamics conditions in the computational models, future investigations can show if aptamers bind more efficiently with heating and cooling and where they bind in these processes. The actual process of binding is not possible to study via the experimental wet lab process. Determining the association and dissociation constants for theses simulations along with the foundational simulations are a necessity in determining the efficiency of binding (50-52).

In this present foundational study aptamer, peptide and antibody interactions were analyzed in the absence of actual MUC1 protein. This protein in its excreted form has multiple binding sites. How and where an aptamer will bind may be different when you have the entire protein structure present as opposed to the peptide region alone. With the varying epitopes

within this protein knowing how the antibody and aptamer bind to the protein would be helpful in determining the surface chemistry and allotted space need in the biosensor.

The success of this research work also raises interesting questions. We have modeled how an aptamer will behave naturally under simulated laboratory conditions that were employed in the present MD simulation analysis. However, we want to apply this to biosensors in which the aptamer will be fixed and bound to a surface. The computational MD models also need to account for this condition to understand if the same binding pattern would occur. There are also several theories on how the length of an aptamer affects the ability of the aptamer to bind more efficiently. Computer simulations could be employed to answer the where, how and if questions of size dependent aptamer binding. This could aid in the development of aptamers as researches design a library for testing. If we can find the binding energy dissociation constants and association constants of proteins and their ligands, it may provide a better estimate of the ideal system. By having the ability to test on the individual amino acid level, we can begin designing peptide and amino acid structures that are specific to a particular active site. This ability to build actual structures and test them will give not only additional options to aptamer libraries but eliminate the trial and error method of experimental selection.

An additional future direction would be scaling up this individual process. It is unknown how these aptamers and targets will behave in close proximity. When a solution of aptamer is bound to a surface for a biosensor we want to make sure they are not interacting with each other and what conformation reassured that each aptamer is acting individually and what happens when multiple aptamers are in close proximity meet a single target. From the design perspective there is also a need to have an optimum distance between the aptamers especially if their target is

large. We do not know if double binding will occur or if the aptamer will not bind at all and only the neighbor will be bound.

A lot of DNA structures are hard to crystallize in x-ray crystallography and NMR due size and charge constraints. Our ability to visualize real-time the natural movements of molecules allows us also to determine unknown locations and conformations. Having the ability to test these 2D structures and letting them naturally come into their ideal 3D structure gives some outlook determining the 3D conformation of small scale structures. Many proteins and large molecules have multiple sites for binding, thus being able to model the most likely scenario of binding to occur for different situations. These have many important implications in drug design, academic discovery as well as biosensor development, to name a few.

References

1. Jain KK. 2010. *The handbook of biomarkers*. New York: Springer
2. Pineda JA, Wang KKW, Hayes RL. 2004. Biomarkers of Proteolytic Damage Following Traumatic Brain Injury. *Brain Pathology* 14:202-9
3. Bruns JJ, Jagoda AS. 2009. Mild traumatic brain injury. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 76:129-37
4. Wang KKW, Ottens AK, Liu MC, Lewis SB, Meegan C, et al. 2005. Proteomic identification of biomarkers of traumatic brain injury. *Expert Review of Proteomics* 2:603-14
5. Pelinka LE, Kroepfl A, Leixnering M, Buchinger W, Raabe A, Redl H. 2004. GFAP versus S100B in serum after traumatic brain injury: relationship to brain damage and outcome. *J Neurotrauma* 21:1553-61
6. Hergenroeder G, Redell JB, Moore AN, Dubinsky WP, Funk RT, et al. 2008. Identification of serum biomarkers in brain-injured adults: potential for predicting elevated intracranial pressure. *J Neurotrauma* 25:79-93
7. Strehlitz B, Nikolaus N, Stoltenburg R. 2008. Protein Detection with Aptamer Biosensors. *Sensors* 8:4296-307
8. Erickson D, Mandal S, Yang A, Cordovez B. 2008. Nanobiosensors: optofluidic, electrical and mechanical approaches to biomolecular detection at the nanoscale. *Microfluidics and Nanofluidics* 4:33-52
9. Song S, Wang L, Li J, Fan C, Zhao J. 2008. Aptamer-based biosensors. *TrAC Trends in Analytical Chemistry* 27:108-17
10. Wang J. 2000. From DNA biosensors to gene chips. *Nucleic Acid Research* 28:3011-6

11. McCauley TG, Hamaguchi N, Stanton M. 2003. Aptamer-based biosensor arrays for detection and quantification of biological macromolecules. *Analytical Biochemistry* 319:244-50

12. Clark SL, Remcho VT. 2002. Aptamers as analytical reagents. *ELECTROPHORESIS* 23:1335-40

13. Stoltenburg R, Reinemann C, Strehlitz B. 2007. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular engineering* 24:381-403

14. Ye M, Hu J, Peng M, Liu J, Liu H, et al. 2012. Generating Aptamers by Cell-SELEX for Applications in Molecular Medicine. *International journal of molecular sciences* 13:3341-53

15. Maxim V. Berezovski ML, Michael U. Musheev, Tak W. Mak,, Krylov SN. 2008. Aptamer-Facilitated Biomarker Discovery (AptaBiD). *Journal of the American Chemical Society* 130:9137-43

16. Bishop GR, Ren J, Polander BC, Jeanfreau BD, Trent JO, Chaires JB. 2007. Energetic basis of molecular recognition in a DNA aptamer. *Biophysical Chemistry* 126:165-75

17. Maxim Berezovski MM, Andrei Drabovich, and Sergey N. Krylov. 2006. Non-SELEX Selection of Aptamers. *Journal of the American Chemical Society* 128:1410-1

18. Auffinger P, Westhof E. 1998. Simulations of the molecular dynamics of nucleic acids. *Current Opinion in Structural Biology* 8:227-36

19. Jayapal P, Mayer G, Heckel A, Wennmohs F. 2009. Structure–activity relationships of a caged thrombin binding DNA aptamer: Insight gained from molecular dynamics simulation studies. *Journal of Structural Biology* 166:241-50

20. Gomperts R, Renner E, Mehta M. 2005. Enabling Technologies for Innovative New Materials. *American Laboratory* 37:12-4

21. Baruch A, Hartmann M-l, Yoeli M, Adereth Y, Greenstein S, et al. 1999. The Breast Cancer-associated MUC1 Gene Generates Both a Receptor and Its Cognate Binding Protein. *Cancer Research* 59:1552-61

22. Gendler SJ. 2001. MUC1, The Renaissance Molecule. *Journal of Mammary Gland Biology and Neoplasia* 6:339-53

23. Dokurno P, Bates PA, Band HA, Stewart LMD, Lally JM, et al. 1998. Crystal structure at 1.95 å resolution of the breast tumour-specific antibody SM3 complexed with its peptide epitope reveals novel hypervariable loop recognition. *Journal of Molecular Biology* 284:713-28

24. Taylor-Papadimitriou J, Burchell JM, Plunkett T, Graham R, Correa I, et al. 2002. MUC1 and the Immunobiology of Cancer. *Journal of Mammary Gland Biology and Neoplasia* 7:209-21

25. Kirnarsky L, Prakash O, Vogen SM, Nomoto M, Hollingsworth MA, Sherman S. 2000. Structural Effects of O-Glycosylation on a 15-Residue Peptide from the Mucin (MUC1) Core Protein†. *Biochemistry* 39:12076-82

26. von Mensdorff-Pouilly S, Gourevitch MM, Kenemans P, Verstraeten AA, van Kamp GJ, et al. 1998. An Enzyme-Linked Immunosorbent Assay for the Measurement of Circulating Antibodies to Polymorphic Epithelial Mucin (MUC1). *Tumor Biology* 19:186-95

27. Pichinuk E, Benhar I, Jacobi O, Chalik M, Weiss L, et al. 2012. Antibody targeting of cell-bound MUC1 SEA domain kills tumor cells. *Cancer Research*

28. Murray A, Spencer DIR, Missailids S, Denton G, Price MR. 1998. Design of ligands for the purification of anti-MUC1 antibodies by peptide epitope affinity chromatography. *The Journal of Peptide Research* 52:375-83

29. Schuman J, Campbell AP, Koganty RR, Longenecker BM. 2003. Probing the conformational and dynamical effects of O-glycosylation within the immunodominant region of a MUC1 peptide tumor antigen. *The Journal of Peptide Research* 61:91-108

30. Hamanaka Y, Suehiro Y, Fukui M, Shikichi K, Imai K, Hinoda Y. 2003. Circulating anti-MUC1 IgG antibodies as a favorable prognostic factor for pancreatic cancer. *International Journal of Cancer* 103:97-100

31. Stavrakoudis A, Tsoulos I, Uray K, Hudecz F, Apostolopoulos V. 2011. Homology modeling and molecular dynamics simulations of MUC1-9/H-2K^b complex suggest novel binding interactions. *Journal of molecular modeling* 17:1817-29

32. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, et al. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* 117:5179-97

33. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004. Development and testing of a general amber force field. *Journal of Computational Chemistry* 25:1157-74

34. Guvench O, MacKerell AD. 2008. Comparison of Protein Force Fields for Molecular Dynamics Simulations Molecular Modeling of Proteins. ed. A Kukol, 443:63-88: Humana Press. Number of 63-88 pp.

35. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM. 1992. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* 114:10024-35

36. Yang W, Nymeyer H, Zhou H-X, Berg B, Brüschweiler R. 2008. Quantitative computer simulations of biomolecules: A snapshot. *Journal of Computational Chemistry* 29:668-72

37. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham Iii TE, et al. 2007. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical Journal* 92:3817-29

38. Hünenberger PH. 2005. Thermostat Algorithms for Molecular Dynamics Simulations Advanced Computer Simulation. ed. C Dr. Holm, K Prof. Dr. Kremer, 173:130-: Springer Berlin / Heidelberg. Number of 130- pp.

39. Berendsen HJC, Postma JPM, Gunsteren WFv, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81:3684-90

40. Bussi G, Parrinello M. 2008. Stochastic thermostats: comparison of local and global schemes. *Computer Physics Communications* 179:26-9

41. Martyna GJ, Klein ML, Tuckerman M. 1992. Nose--Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* 97:2635-43

42. Law RJ, Capener C, Baaden M, Bond PJ, Campbell J, et al. 2005. Membrane protein structure quality in molecular dynamics simulation. *Journal of Molecular Graphics and Modelling* 24:157-65

43. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* 26:1701-18

44. Delano WL. 2002. The PyMOL Molecular Graphics System.

45. Berendsen HJC, van der Spoel D, van Drunen R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* 91:43-56

46. Parrinello M, Rahman A. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52:7182-90

47. Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* 14:33-8

48. Carugo O. 2003. How root-mean-square distance (r.m.s.d.) values depend on the resolution of protein structures that are compared. *Journal of Applied Crystallography* 36:125-8

49. de Gennes P-G. 1979. *Scaling concepts in polymer physics*. Cornel University Press

50. Gouda H, Kuntz ID, Case DA, Kollman PA. 2003. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* 68:16-34

51. Gilson MK, Zhou H-X. 2007. Calculation of Protein-Ligand Binding Affinities*. *Annual Review of Biophysics and Biomolecular Structure* 36:21-42

52. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, et al. 2000. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research* 33:889-97