Computational Modeling of DNA Sequence Effects on the Nucleosome Core Particle

Levi CT Pierce

Submitted to the Department of Electrical Engineering and Computer Science and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirement for the degree of Master of Science

Thesis Committee

Terry Clark, Professor, EECS

Xue-wen Chen, Professor, EECS

David Andrews, Professor, EECS

The Thesis Committee for Levi CT Pierce certifies that this is the approved version of the following thesis:

Computational Modeling of DNA Sequence Effects on the Nucleosome Core Particle

Thesis Committee

Terry Clark, Professor, EECS

Xue-wen Chen, Professor, EECS

David Andrews, Professor, EECS

Computational Modeling of DNA Sequence Effects on the Nucleosome Core Particle

Abstract

The nucleosome particle is an essential biological macromolecule serving both a structural and gene regulatory roles in eukaryotic genomes. The nucleosome particle has a cylindrical shape, composed of 8 highly conserved histone proteins wrapped by a sequence of 147 base pairs of DNA. Considerable, experimental evidence has shown that different sequences of 147 base pairs have varying preferences for forming stable particles, yet atomic-level descriptions for the preferences are vague at best. Microscopic descriptions contribute to fundamental understanding of genomes process, facilitate rationale approaches to drug design for certain genetic diseases, and can contribute to genetic engineering. We have established a novel basis for computational modeling DNA interactions with the nucleosome core particle. Computational modeling can approach the complexity and vast number of DNA sequences of potential interest. Our method is the first to substitute DNA on the nucleosome core and explore the rotational degree of freedom, crucial to assessing a DNA helix in preferred, low-energy states. This work was carried out along with experimental work used as a reference for the computational studies. Specifically, we experimentally determined the relative binding affinities of 3 DNA sequences for forming nucleosomes. These experimental data provide a ranking of stability as a function of sequence on the free energy of binding. Accurate free-energies of binding for large biological systems are extremely difficult to compute. The computational modeling hinges on a high-resolution crystal structure (1.9 Å) of the nucleosome core particle. In our method, we substitute DNA molecules of interest on the crystal structure in order to study the structural dynamics using sophisticated computational models centered on molecular dynamics simulations. In this work, we performed three separate molecular dynamic simulations, one for each of the sequences, in order to explore the atomic-level basis for the differentials in binding, which were determined experimentally. Crucially, the rotational degrees of freedom are explored applying novel methods to a complex geometric and chemical problem. This method uses systematical sampling and dynamical modeling of the DNA rotational conformers. Through this work we have demonstrated the feasibility and methodology for atomic-level modeling of DNA with potential for high throughput. This thesis describes methods and results of this work.

Contents

1 Background

In this thesis, we contend that computational modeling of sequence effects on nucleosome structure, dynamics, and function are now within our reach. We demonstrate this with our development of a robust prototype system tested against experimental studies conducted along with the computations reported here.

First, we give an overview of the general nature of the project. This was a large molecular modeling project (~450,000 atoms per simulation system) of considerable complexity. Computationally, sustained access to 100 to 200 processors was required throughout the project. The nucleosome itself large by contemporary standards; the size increases considerable with the use of explicit solvent in the simulation. The particle is a tightly arranged molecular complex with 1.7 turns of DNA bent around a roughly cylindrical protein core. Formation of the nucleosome in vitro and in vivo is facilitated through the input of energy. Our work focused on the final configuration, that is, the DNA wrapped around the histone core; we did not approach the formation of the nucleosome in our work; this is a daunting problem with current computational models. However, it is now relatively commonplace to assess the stability of molecular complexes in their final configuration through a variety of computational means. *Stability* is often correlated with *binding affinity*; in this case the affinity of a particular sequence of DNA for forming a nucleosome particle. We developed protocols through experienced trial and error as is commonplace in simulations such as this, particularly in developing novel methods. The rotation of DNA with explicit atoms (including hydrogen atoms) on the nucleosome core has not been attempted, to our knowledge, prior to this work; the rotation is nevertheless an essential prerequisite to model novel DNA sequences. We emphasize that it is impractical to determine atomic-level structures of the many DNA-histone-core complexes of interest, and likely to continue to be so for the foreseeable future. Furthermore, important atomic-level details are difficult and impractical to obtain experimentally. It has been demonstrated experimentally that many DNA sequences are neutral in their binding effects. However, an enormous body of literature demonstrates dramatic effects of many DNA sequences on binding affinity. The nucleosome occurs approximately every 200 base pairs in bulk human genomic DNA (with the haploid sequence containing approximately 3 gigabases of DNA). This creates significant opportunities for sequence effects through nucleosome interactions on gene regulation, chromatin structure, and other genomic processes. It is important to keep in mind that genome function varies in large ways across the sequence from, for example, the centromeres central to nuclear division and the genes comprising only ~3% of the genome sequence. The landscape of concern for these studies is extremely rich and unexplored.

1.1 The Nucleosome Structure

The nucleosome particle is composed of two copies of each histone protein, H2A, H2B, H3, and H4 forming a cylindrical octamer core wrapped by approximately

147 base pairs of DNA (Figure 1). A super coil of approximately 1.7 turns is formed around the octamer core of proteins when the DNA is fully wrapped after an involved assembly process. The nucleosome particle is situated approximately every 200 bases throughout bulk genomic DNA. The nucleosome-DNA interaction is a first-order chromatin structural element, giving rise to higher order structures such as the 30-nm chromatin fiber; nearly 30-40 times compaction of linear DNA is achieved by the chromatin fiber (Figure 2).

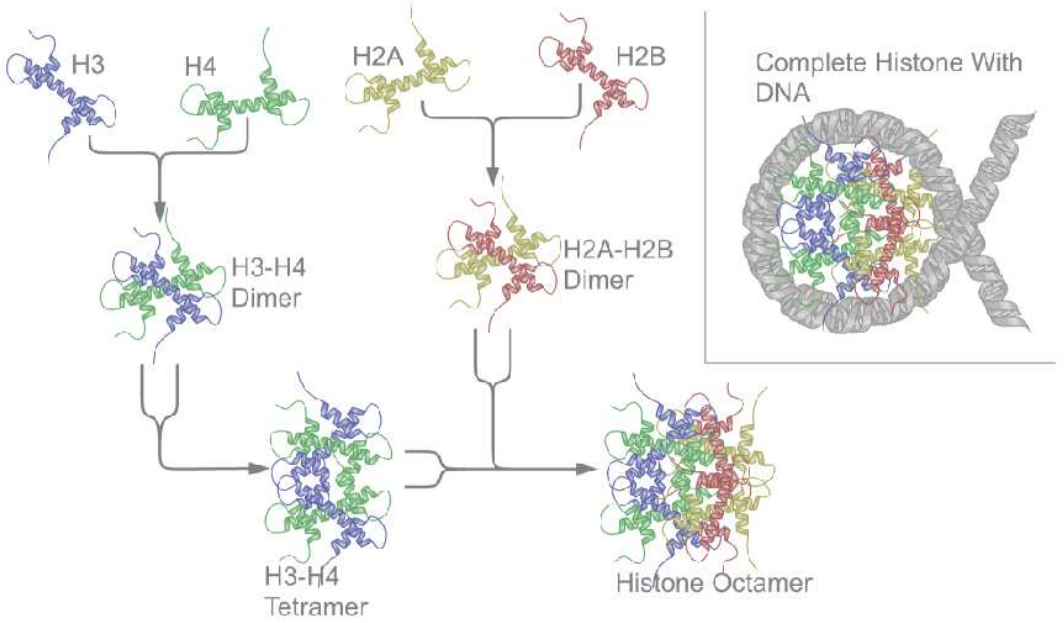


Figure 1. **Schematic representation of the formation of the histone octamer and the full nucleosome with DNA wrapped around. Not shown is the H2A-H2B tetramer which is formed in the same fashion as the H3-H4 tetramer.**

In the past decade, Richmond and colleagues have crystallized a series of nucleosome particles culminating in the 1.9 A structure of the *Xenopus laevis* (African clawed frog) recombinant histone core bound to a synthetic palindrome modeled after human alpha satellite DNA [8]. The core histone proteins are among the most highly conserved proteins throughout eukaryotic organisms. This fact permits reasonable extrapolation from the *Xenopus laevis* histone octamer, where, for example, in the 1KX5 crystal structure only 9 amino acids differ between the human histone counterparts. A detailed sequence analysis of the histone proteins found in *Xenopus laevis* and human histones has been compiled and is listed by us in Appendix 7.3. The crystallographers used palindromic DNA sequence was chosen for its favorable properties in forming nucleosome crystals [7-9].[1]

[1] Bishop notes this in his paper two references Harp, J. M. et. Al. and another paper Richmond…

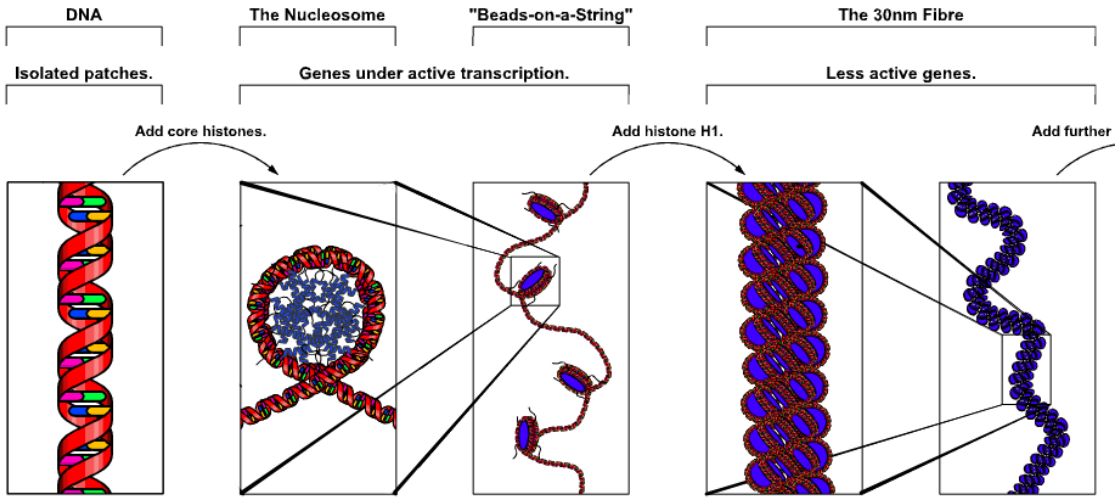


***Figure 2*. A Schematic representation showing the increasing levels of DNA compaction facilitated by the nucleosome.**

1.2 Biological Function of the Nucleosome

The long-studied relationship of nucleosomes with gene regulation has been firmly established [1], [2]. Highly transcribed genes such as housekeeping genes have been associated with nucleosome depleted regulatory elements [1]. Data mining applied to genome sequence has found enrichment in poly-dA/dT sequence upstream of housekeeping genes. In contrast, upstream regions of tissue-specific genes are relatively depleted in poly-dA/dT sequence [3]; poly-dA/dT sequence of sufficient length is thought to repel or destabilize nucleosomes [4-6]. Several classes of topological arrangements of nucleosomes relative to gene regulatory elements have been identified in recent reports involving the yeast genome. For example, transcriptional start sites were found concentrated at ~13 bp inside the border of the first upstream nucleosome, suggesting a role for the H3 histone N-terminal tail in transcription regulation; TATA boxes were found to have likely locations at positions on the nucleosome downstream border [1] and in regions unoccupied by nucleosomes [2]. The structural and regulatory aspects puts the nucleosome particle and its interaction with DNA in center stage for work in gene regulation, genetic engineering, drug design [7] and most biophysical events in the eukaryotic nucleus.

1.3 Sequence Preference

It has been shown that DNA sequence affects can significantly vary affinities for nucleosome binding. DNA bendability and histone contacts are considered two nucleosome stabilizing factors. Large free energies for bending DNA on the nucleosome during formation [10] can be reduced by more bendable DNA sequence [11][12]. The understanding of atomic level effects on DNA bendability is ongoing research, proving to be a difficult problem. Interestingly, the curvature of the DNA around the nucleosome are not uniform; a variety of approaches reach a common conclusion of more tightly super-coiled DNA around the pseudo dyad axis of about 10.5 bp per turn and about 10 bp

per turn to either side of the central region [13]. This asymmetry in curvature exemplifies the need for molecular modeling beyond simple sequence-based curvature models devoid of the nucleosome environment. For example, earlier work has attempted with little success to predict the capacity of DNA to form nucleosomes based on simple bendability models of free DNA. More recently, Johnathan Widom's group experimentally searched a library of 10^{12} randomly generated clones, discovering the high-affinity DNA sequence, clone 601 [14]. The sequence is listed in Appendix 7.4.

1.4 Medical Problems Associated with the Nucleosome

High resolution nucleosomes with custom DNA sequence provide a basis for modeling important situations in chromosome disorders. A disease that is directly related to improper compaction of the X chromosome due to abnormally low affinity for nucleosome formation is known as fragile X syndrome. The severity of the disease is directly proportional to the number of triplet repeats of the sequence CGG on the long arm of the X chromosome. The disease causes mental retardation, facial deformities, and macroorchidism in males. Increasing numbers of triplet repeats have been found experimentally to decrease the affinity for forming stable nucleosomes. Disrupting the first order level of nucleosome compaction leads to characteristic breaks in the X chromosome of fragile X patients (fig. 3). A sequence containing 49 CGG triplet repeats is one of the three sequences modeled in this study and is listed in Appendix 7.4.

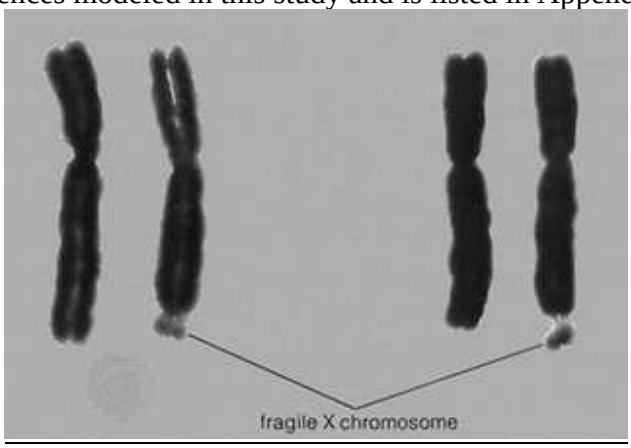


Figure 3. **Two examples of breaks in the X chromosome are shown due to mis-compaction of the chromatin fiber. A normal chromsoome is shown on the right in both cases.**

However, numerous other diseases are associated with triplet repeats. For example, the sole genetic defect known to be associated with myotonic dystrophy type 1 (DM1) is expansion of the unstable CTG/CAG repeat in the 3' untranslated region (UTR) of the myotonin kinase gene on human chromosome 19 [15]. DM1 is characterized by muscular dystrophy and myotonia in combination with features including heart defects and cognitive impairment [16]. The impact of the disease is significant, although estimates vary as widely as 1 in 500 to 1 in 20,000 individuals [15]. A multi-system disease, DM1 produces a pathological effect via loss-of-function mechanisms by shutting down the expression of DMPK, SIX5, and DMWD genes [17]. Wang and colleagues have demonstrated the strong nucleosome positioning effect by the expanded CTG/CAG

repeat [18]. Those studies showed nucleosome positioning strength increased with triplet length (ref). Free energy studies demonstrated that $(CTG/CAG)_n$ triplets are one of the strongest nucleosome-positioning signals known at n=75 [17]. However, the precise reason why the repeating triples generate such highly stable structure remains unclear.

In addition to genetic diseases such as DM1, the deregulation of the formation of compacted chromatin is implicated in many forms of cancer. Silencing of tumor suppressor genes through such aberrant chromatin structure is particularly critical in cancer progression. The inhibitors for histone deacetylase and DNA methyltransferase, which are key enzymes in the formation of compact chromatin, have been developed as therapeutic drugs to reactivate the repressed genes in genetic diseases and cancer (Cheng et al, 2004; Garcia-Manero and Issa, 2005; Marks and Jiang, 2005). However, because these drugs target all compacted chromatin regions including the regions in which heterochromatin is the norm, possible side effects are inevitably associated with these drugs. The methods developed and study conducted here characterizes the nucleosomal DNA-histone interactions at the atomic level. Unique interactions between DNA and histones that are found in this study could provide potential targets for small therapeutic chemicals.

2 Development of a Computational Model for the Nucleosome

With the advent of highly parallel computing environments large biological complexes can be studied at an atomic level. However, as with any model it must be tested and verified before conclusions can be made. A series of troubleshooting and refinement of the nucleosome model has been carried out producing a reliable robust testing environment.

2.1 Computational Methods for studying Macromolecular Systems

Molecular Dynamics (MD) is a well established method for investigating structural and dynamical properties of large macromolecular systems. MD has been successfully applied to systems of free DNA [24-26] and protein DNA complexes [27-30]. Within the past ten years, force fields capable of modeling B-DNA have been developed which are central to MD simulations. The MD studies presented in this work were carried out through the use of the highly parallelized package NAMD on the ITTC cluster [31]. The atomic-level interactions that MD generates have been well established for large macromolecular complexes such as DNA and protein [33][48]. The use of explicit solvent in the simulations is critical in order to understand the structural implications of DNA sequence effects in the structural and dynamical interactions. A recent study showed varying results from implicit solvent simulations of the nucleosome compared to previous explicit simulations [47]. Important interactions are lost when implicit solvent is used such as water bridges which have been implicated in stabilizing protein-protein complexes and DNA-protein complexes [42]. (The explicit waters increase the computational and data management costs by an order of magnitude.) The trajectories from an MD simulation are contiguous, sampled snapshots in time of every atom's position and velocity (a sampling frequency of 1 picosecond with 2 femtosecond time steps is a typical approach). These trajectories can be analyzed using numerous

techniques which elucidate time correlated properties of the system with thermodynamics, kinetics, and dynamical and structural features.

The rate at which the simulation can be propagated through time (the discretized time step) is determined by the highest frequency of vibration – that among bonded interactions involving the lightest in the system. Typically, the computed forces are applied in increments of 2 femtoseconds which move the system forward in time. Electrostatic forces are computed with long-range electrostatics based on a series expansion of the Coulomb potential, and simpler pairwise interactions for close atoms (those within about 15Å). The long-range electrostatic forces have been instrumental in accurately simulating DNA.

2.1 Previous Computational Studies of the Nucleosome

Thomas Bishop was one of the first to study the full nucleosome system with a molecular dynamics study using explicit solvent and the 1KX3 2.0Å crystal structure [20]. 1KX3 is one of the earlier crystal structures and does not include the tails of the histone proteins. More recently, Martin Zacharias' group used the 1KX5 1.9Å crystal structure and explored the structural effects of the nucleosome with full protein tails and with truncated protein tails [19]. The same study was performed and differences in our results are noted. Ramaswamy's group investigated the global motions of histone variants using an elastic network model and a low resolution structure of 2.8Å [44].

2.2 Initial Computational Studies of Substituted Sequences

X-ray crystallography and NMR are not suited to produce nucleosome structures at a pace with the vast DNA sequence combinations of potential interest to modeling research. Our computational approach presented here addresses this problem. The first models that were run with the 1KX5 system were two engineered sequences I1 and A7 which were hypothesized to show low binding and high binding do to the computed bendedness of the two sequences. The I1 sequence was constructed to have high bend by including sequence motifs that have been shown to contribute to nucleosome stability. The sequence motifs are characterized by dinucleotide biases that have been found experimentally at positions along the nucleosome where the minor grove faces inward. The dinucleotides which include AA/TT/TA and GC have been shown to produce sharp bends in free DNA. By positioning these locally bendable motifs in phase with the minor groves the sequence I1 was engineered to be highly bendable. Similarly, the motif $(AAACCCT)_n$ was chosen due to its locally low bendable properties. This was the basis for the engineered weak binding sequence, A7. The software package BEND was used to compute the relative bendability of these two sequences [39]. The DNA sequences were substituted onto the nucleosome core particle using a custom script to interface with the Tripos Sybyl Biopolymer module. The nucleotide substitutions along the 147 bases of the X-ray crystal structure use the orientation of the original DNA as a template for positioning the substituted nucleotides. Graphical inspection showed that the substitution retains the hydrogen bonding structure in the planes of the original sidechains, while also retaining the helical backbone orientation.

When first exploring the system the histone tails on H3 were clipped to reduce the total volume of the box used to simulate the system which would allow for faster throughput of the simulation. The length of the H3 tails was reduced from 135 amino acid residues to 101. The simulations with the clipped tails ran for 5ns before the ends of the DNA started to unwind from the protein core (fig 4).

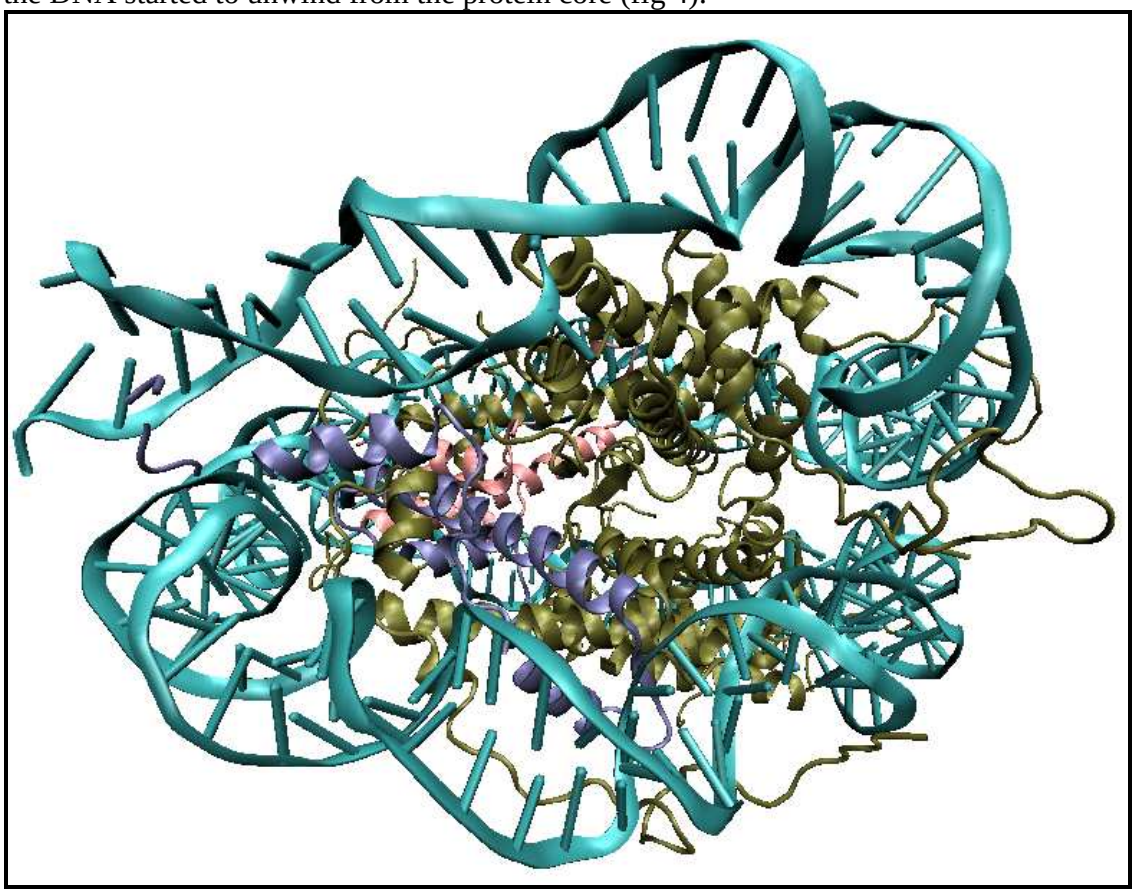


Figure 4. **DNA coming unbound from the core protein after 5ns. The H3 (A) in pink and H3` (E) in blue are highlighted to orient the reader and exhibit the truncated tails.**

The problem was first addressed by using a restraining force on selected phosphate atoms within the DNA backbone. If phosphates 147, 146, or 145 on chains I and J moved more than a defined distance from the protein core a restoring force was applied in the opposite direction. The restraints were included in the NAMD configuration file for each run in the simulation (Appendix 7.4). The simulation was started from the beginning and run for 25ns without the DNA unwinding. However, a harmonic wave could be seen propagating through the DNA structure due to the restraining forces. These simulations were terminated and could not be used for analysis.

A study of the original DNA, full length tails, and no constraints was conducted to provide a basis for comparison of any subsequent substituted systems. (The penalty was a larger, more computationally intensive system.) A second simulation was also run with the original DNA, truncated tails, and no constraints. The tails contain numerous positively charged amino acid residues including Arginine and Lysine which are attracted to the strong negatively charged DNA backbone.

An electrostatic potential map has been generated for the DNA (fig. 5) and protein (fig. 6) illustrating the negatively charged DNA and positively charged protein tails.

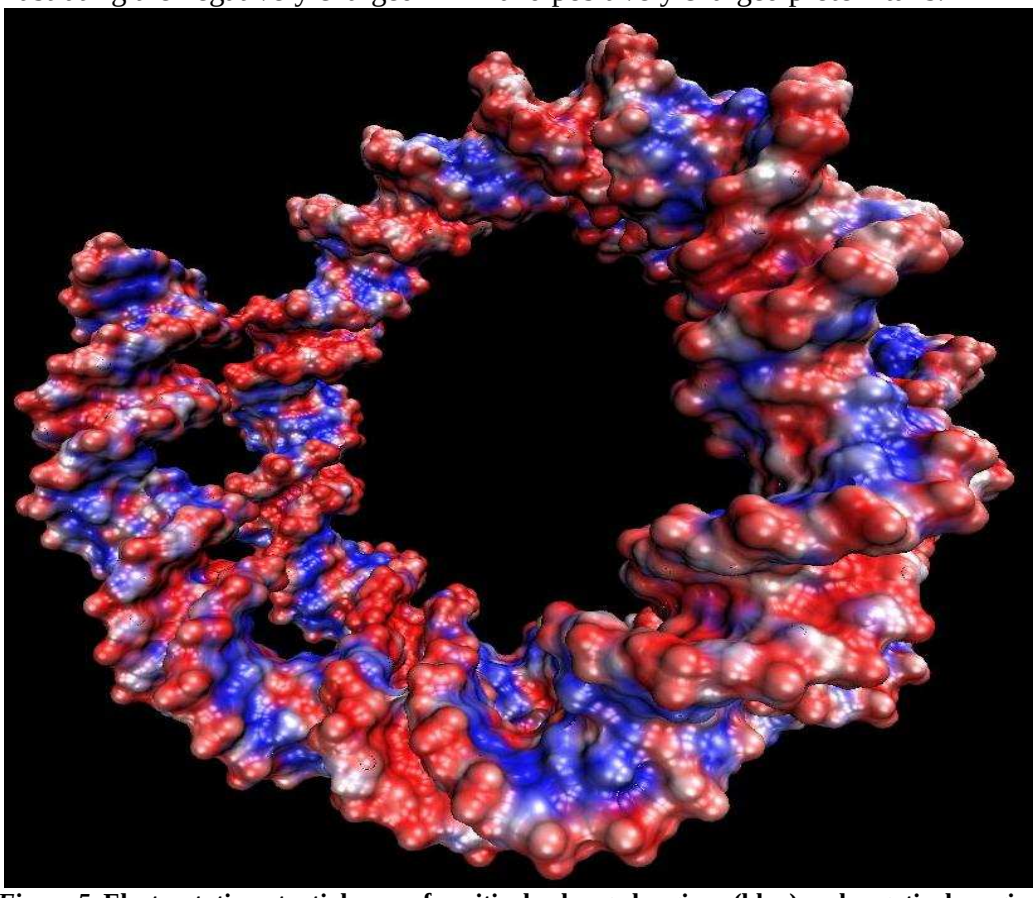


Figure 5. **Electrostatic potential map of positively charged regions (blue) and negatively regions (red) on the DNA. The negatively charged regions dominate the surface of the DNA.**

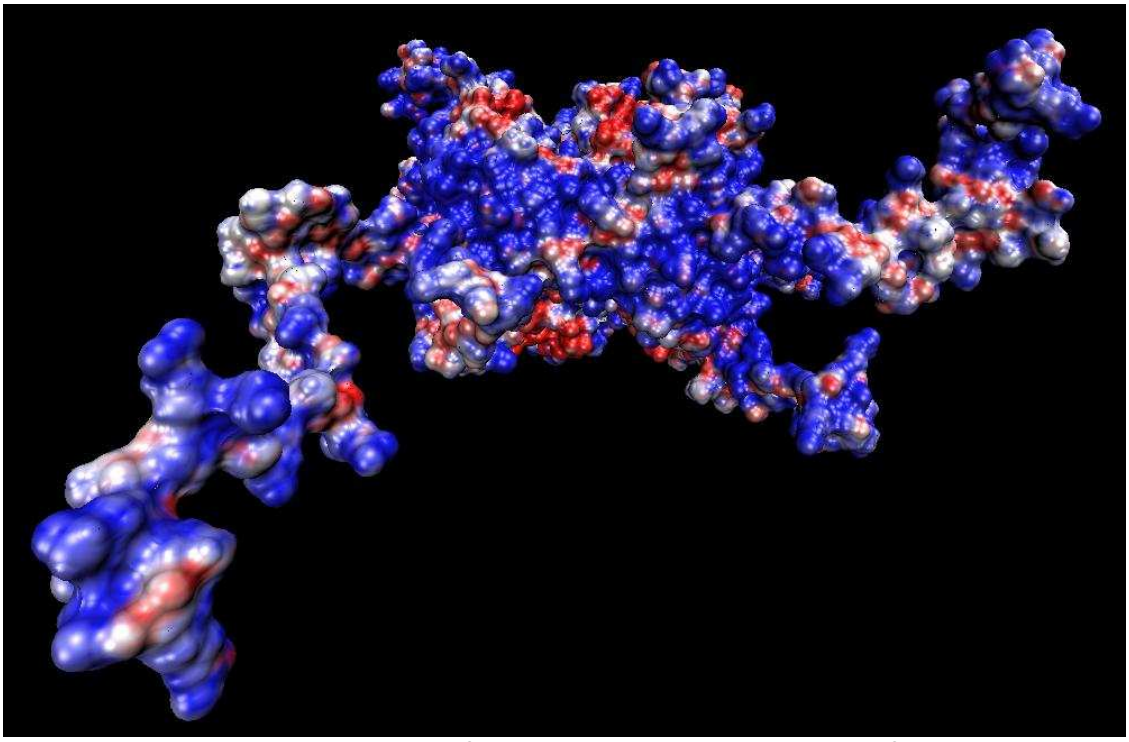


Figure 6. **Electrostatic potential map of the protein core. Dominating the surface landscape are positively charged regions (blue). Negatively charged regions (red) are found sparsely distributed on the surface. The H3 tails contain primarily positively charged regions.**

The simulation has run for an impressive 25ns without the DNA coming unbound from the protein core. Furthermore, with no restraints being applied the structure exhibits no harmonic waves propagated along the backbone. The tails can actually be seen cupping the ends of the DNA after several nanoseconds of simulation (fig. 7).

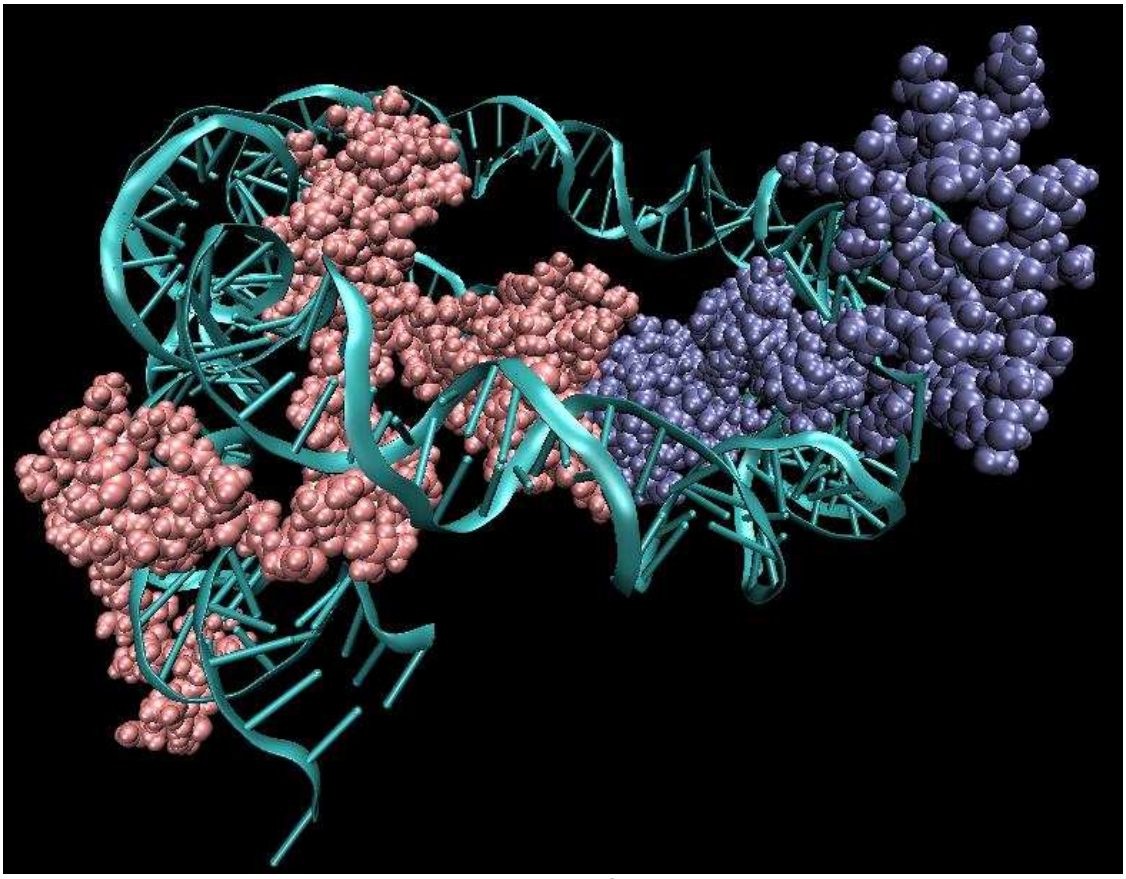


Figure 7. **Histone proteins H3 (A) colored pink and H3` (E) colored blue are seen binding to the DNA ends after only a few nanoseconds of simulation.**

As expected the truncated tail experiment came unbound from the core at 5ns into the simulation. The interaction of the tails with the ends of the DNA correlates with known transcriptional regulation mechanisms mediated by acetylating sites on the tails. The proteins in the 1KX5 crystal structure are in a recombinant state and are not acetylated therefore they should bind to the DNA preventing unwinding [23]. Another group came to the same conclusion running a truncated tail simulation alongside a full tail simulation. However, they did not carry their simulations out beyond 20ns and did not witness the dramatic unwinding event [19]. Previous computational studies of the nucleosome with truncated tails can not be reliably used for structural analysis.

3 Experimental Evidence

Once the full tail model had been proven to provide a reliable environment to model the 1KX5 system it was decided to explore sequences with experimentally determined affinities for nucleosome formation. By selecting sequences which can be experimentally ranked based on binding affinities for the nucleosome the computational model can then be applied to explore structural differences. For our experimental studies, we used: clone 601 (kindly provided by Jonathan Widom); CGG triplet repeat; 5S rRNA, and the Richmond and Davey human satellite palindrome DNA. The 5S rRNA sequence

is used as a reference point in the experiments and was not used in the computational models.

Relative free energy for the affinity of nucleosome formation has been obtained for clone 601 but no published data exists for the affinity of the palindrome sequence used in the 1KX5 crystal structure. Furthermore, in our reconstitution experiments, we used a novel approach of using trimmed sequences of 147 basepairs; nucleosome reconstitution experiments do not use sequences with lengths of 147 base pairs but instead add flanking DNA of up to 100 base pairs. In order to determine an accurate ranking of relative free energies of binding for sequences of 147 base pairs in length we visited Dr. Yuh-Hwa Wang's Biochemsitry laboratory at the Wake Forest Medical Center to carry out experimental free-energy of binding experiments with the same DNA we were modeling computationally. The experimental ranking provides a reference for correlating atomic-level structure and dynamics with nucleosome stability. **The hypothesis is that the sequence with a higher affinity for the nucleosome core protein will exhibit less fluctuation throughout the simulation, contain more stable contacts, and be appropriately curved.**

3.1 Competitive Nucleosome Reconstitution Theory

Shrader and Crothers pioneered the methodology using competitive nucleosome reconstitution to determine relative free energies [14]. The method measures the ability of a labeled DNA fragment to compete with bulk DNA for a limited amount of histone protein during nucleosome assembly. Free-energy differences can be obtained for the binding of a variety of sequences relative to binding of a standard reference sequence. Using the software package Image Quant[2] counts of nucleosome bound DNA and free DNA can be computed which defines the equilibrium constant K_{eq}.

$$\Delta G^{o}_{sample} = -RT\ln(K_{eq}), \text{ where } RT = 0.55 \text{ kcal/mol}^{-1} \text{ (for T=4°C)}$$

$$\Delta\Delta G^{o}_{sample} = \Delta G^{o}_{sample} - \Delta G^{o}_{reference}$$

This procedure has been applied to determine the relative free energies of several repeating sequences including the CGG repeats. Results from prior studies are listed at the right side of the free energy scale bar (Fig. 8), with the reference free energy of the vector DNA (pUC19 fragments) defined as 0 cal/mol [50][52][53]. Data which are at the left of the bar were adapted from Shrader and Crothers [12]. Importantly, denoted in red is the synthetic sequence clone 601, showing the strongest affinity for nucleosome formation.

[2] Image Quant was formerly made and sold by Molecular Dynamics. Amerhsam Biosciences is now supporting the software.

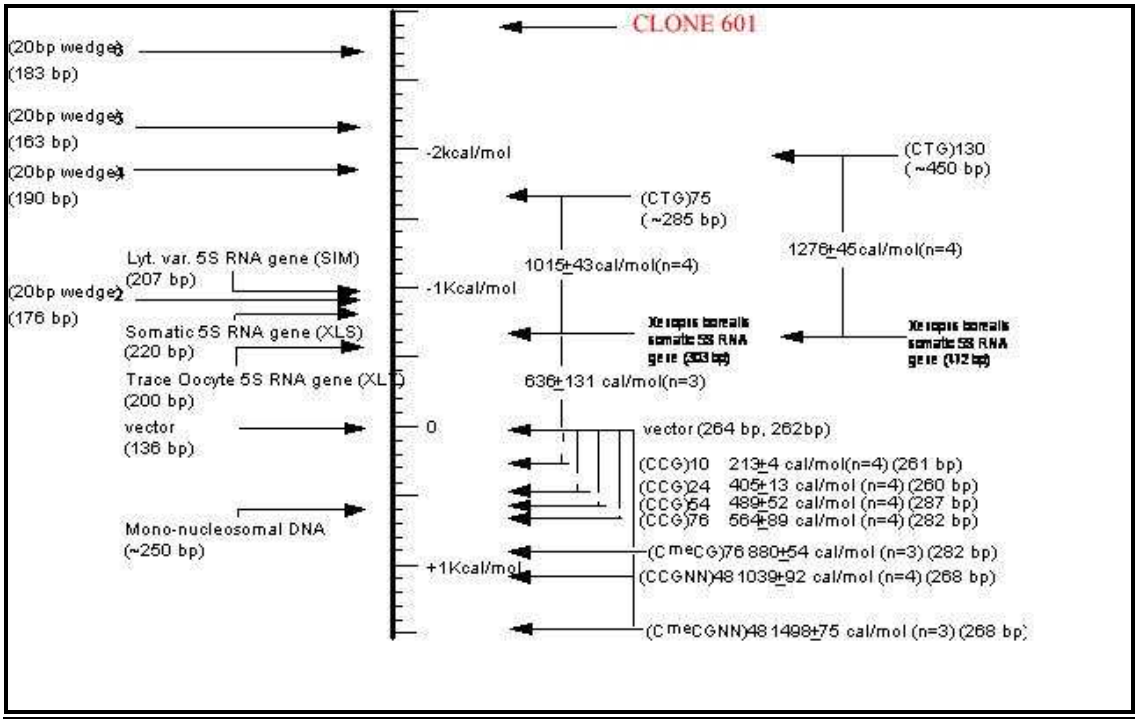


Figure 8. **Ranking of experimentally determined relative free energy binding of different DNA sequences to histone protein. Highlighted in red is clone 601 which has been found to be one of the strongest binders of histone protein.**

3.2 Selection of Sequences

The sequences chosen to be reconstituted with human histone proteins were Jonathan Widom's synthetic 601 clone, Richmond and Davey's synthetic palindrome sequence, Fragile X Syndrome $(CGG)_n$, and the 5S rRNA gene sequence (Accession number K01537). Jonathan Widom's sequence is the high affinity binder, the affinity of Richmond and Davey's was speculated to be low, the $(CGG)_n$ triplet repeat is a low binder, and the 5S rRNA gene sequence is what is used as the reference.

3.3 Preparation of Sequences

The experiments conducted at Wake Forest were carried out in 2 trips. The goal for the first trip was to prepare the chosen sequences for the nucleosome reconstitution experiments. The goal of the second trip was to conduct the reconstitution experiments and analyze the collected data.

3.3.1 Clone 601

The Widom sequence was received cloned into the HincII site of the plasmid pGEM3Z. The sequence of the 601 clone is as follows:

601nv Length: 282 April 1, 1999 07:27 Type: N Check: 1886[3] ..

```
  1  CGGGATCCTA ATGACCAAGG AAAGCATGAT TCTTCACACC GAGTTCATCC
 51  CTTATGTGAT GGACCCTATA CGCGGCCGCC CTGGAGAATC CCGGTGCCGA
101  GGCCGCTCAA TTGGTCGTAG ACAGCTCTAG CACCGCTTAA ACGCACGTAC
151  GCGCTGTCCC CCGCGTTTTA ACCGCCAAGG GGATTACTCC CTAGTCTCCA
201  GGCACGTGTC AGATATATAC ATCCTGTGCA TGTATTGAAC AGCGACCTTG
251  CCGGTGCCAG TCGGATAGTG TTCCGAGCTC CC
```

Three primers were selected to produce two sequences of lengths 190bp and 147bp (Appendix 7.1). The 147bp sequence is highlighted in red and was the sequence used in our computational model as well as the reconstitution experiments. The additional 43bp sequence encompasses the 190bp sequence and is shown in pink. Shown in figure 9 is the pGEM3Z vector map which the .

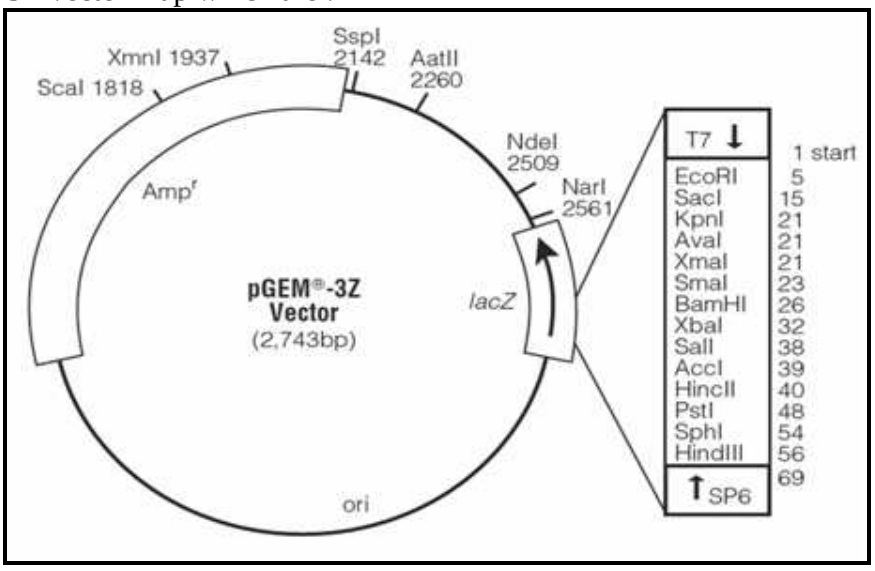


Figure 9. **pGEM-3Z Vector map displaying the restriction sites. The HincII site is where the clone 601 sequence was inserted.**

Transformation of the pGEM-3Z vector with the 601 clone insert into Super Competent E. Coli SURE cells was carried out using the following method. Chilling cells in the presence of divalent cations such as Ca^{2+} (in $CaCl_2$) causes the cell walls to become permeable to plasmid DNA. The cells are incubated on ice with the DNA and then briefly heat shocked (e.g. 42°C for 30–120 seconds), which causes the DNA to enter the cell. This method works well for circular plasmid DNAs. An excellent preparation of competent cells will give ~10^8 colonies per microgram of plasmid.

Transformed bacteria are then plated out onto two agar + ampicillin Petri dishes and grown overnight at 37°C ~8 hours. Six unique colonies are then scrapped (3 from each plate) with sterile toothpicks and grown in separate tubes overnight at 37°C in a gentle shaker. The growth medium contains 5ml LB broth and 5µl ampicillin (100µg/ml). A miniprep procedure is carried out to purify the plasmid DNA from the bacterial cells. The cells are lysed under alkaline conditions yielding both DNA and proteins. Addition of acetate-containing neutralizing buffer will cause the larger and less

[3] This clone originated form Lowery and Widom, 1998, New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-directed Nucleosome Positioning. JMB: 276, 19-42.

supercoiled chromosomal DNA and proteins to precipitate out leaving the small bacterial plasmids in solution.[4]

Once the plasmid DNA has been purified a test is performed to ensure the transformation process was successful. Restriction enzymes are used to cut out a region where the 601 clone was inserted. Using the pGEM-3Z map, EcoRI and HindIII restriction enzymes were selected. The expected fragment length after cutting these two sections in the circular plasmid is 333bp illustrated below:

(EcoRI)::::::**35bp**::::::(HincII insert site for 601 clone **282bp)**::::::**16bp**::::::(HindIII)

Figure 10 shows the results of the restriction enzyme experiments. Six bands can be distinctively seen at the expected 333bp mark. Three samples of 601-Lane 1, 601-Lane 2, and 601-Lane 3 were selected, frozen via a dry ice bath, and stored at -80ºC. The clones are then used for growing larger quantities of the 601 sequence in the future. The yield of the miniprep is approximately 25-50 μL and the yield for a midiprep is approximately 300μL.

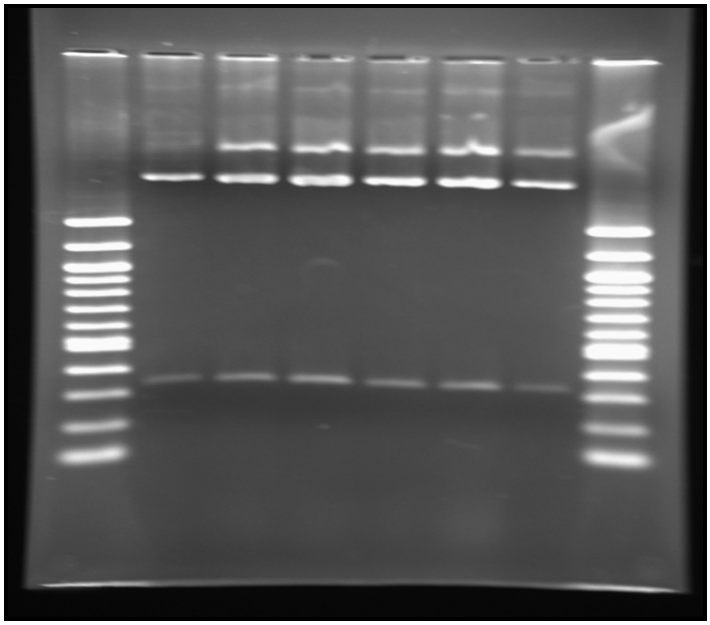


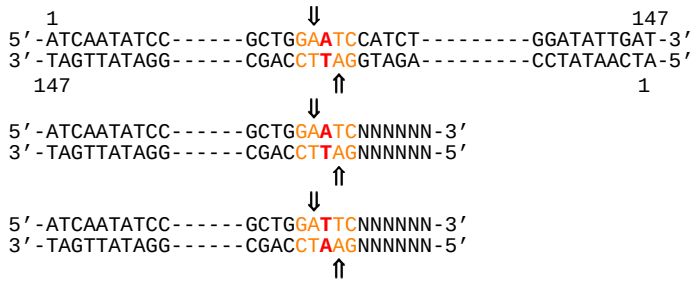
***Figure 10*. From left to right, lanes 1 and 8 were loaded with 100bp marker and the other lanes were loaded with purified plasmid DNA from the six colonies. The gel ran from top to bottom in a 1.5% agarose gel, 1XTE buffer, at 100V for 45 minutes.**

3.3.2 Palindrome 1KX5 Sequence

Starting from selected oligonucleotides the sequence of interest is generated and then inserted into the vector pGEM3zf(+). Transformation is performed using the SURE-2 cells. A miniprep is carried out to test the strategy. Midiprep is carried out for DNA sequencing of the fragments and reconstitution experiments.

[4] The Alkaline lysis method was developed by Birnboim and Doly in 1979.

The sequence is palindrome and cannot be synthesized by using simple primers. The strategy devised to generate the 147bp sequence involves creating two equal halves of the full length sequence and then ligating them together. Illustrated below are the full length palindrome sequence and the two half sequences that are generated. Extra sequence is added to the halves in order to allow proper binding of the restriction enzyme. The restriction enzyme HinfI cleaves the sequences at the location of the arrows.



Cutting the two half sequences yields the following “sticky” end sequences shown below which can be ligated together to form the desired 147bp sequence.

```
5′-ATCAATATCC------GCTGG
3′-TAGTTATAGG------CGACCTTA

5′-ATCAATATCC------GCTGG
3′-TAGTTATAGG------CGACCTAA
```

The primers to be synthesized are shown below:

```
P1 5′-ATCAATATCCACCTGCAGATACTACCAAAAGTGTATTTGGAAACTGCTC-3′
P2 5′-GGATCCGATTCCAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAA-3′
P3 5′-GGATCCGAATCCAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAA-3′
```

Primers **P1**+**P2** and **P1**+**P3** are annealed together pairing in the following orientation shown below.

```
P1 5′-ATCAATATC-----TATTTGGAAACTGCTC-3′
              P2 3′-AACCTTTGACGAGGTAGTTTT-----CTTAGAGCCTAGG-5′

P1 5′-ATCAATATC-----TATTTGGAAACTGCTC-3′
              P3 3′-AACCTTTGACGAGGTAGTTTT-----CTAAGAGCCTAGG-5′
```

After annealing the primers together the klenow fragment is used to fill in the nucleotides on the single strands. The two halves are referred to as 1KX5-(P)1/(P)2 and 1KX5-(P)1/(P)3, respectively. The characteristics of the primers are listed in appendix 7.1.

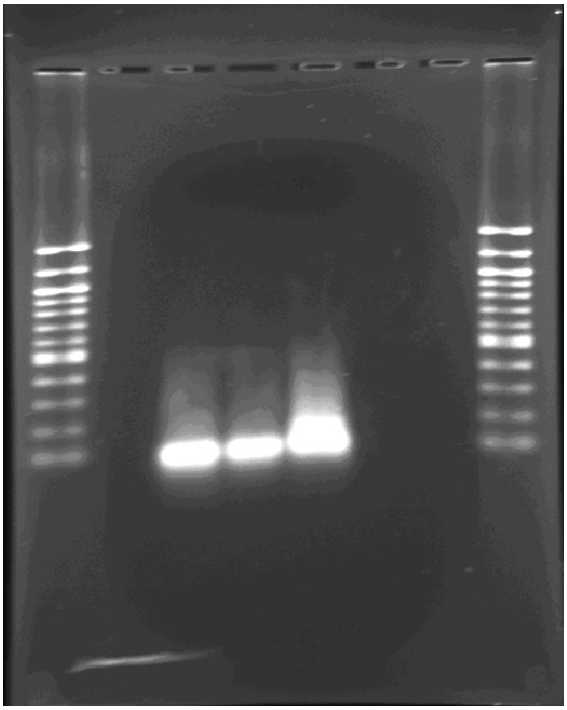


***Figure 11*. The annealing and ligation strategy using the half clones was tested by running the products on a 1.5% agarose gel, 1XTE buffer, at 100V for 45 minutes. Counting the lanes from left to right, 3 and 4 contain 1KX5-1/2 and 1KX5-1/3, respectively, after annealing and the Klenow fragment fill in reaction. The fragments in lanes 3 and 4 appear to be just to the right of the 100bp marker indicating they are the expected 79bp in length. The fragment in lane 5 is the product of HinfI cleavage of the two fragments in lanes 3 and 4 followed by ligation of the sticky ends. There are two bands, one faint at the expected 147bp and one stronger near the location of the 79bp fragments.**

The 1KX5-1/2 and 1KX5-1/3 fragments were inserted into plasmids. The site selected for insertion was the smaI site in the pGEM-3Zf(+) vector (Appendix 7.1). To ensure that this site did not exist within the 147bp palindromic sequence the program Stryder was used. The program analyzes the sequence for known restriction sites. The smaI site was not found in the sequence. Transformation and Growth of Colonies was carried out using the SURE2 Supercompetent Cells which are $lacZ^-$. The plasmid vector is $lacZ^+$ but an insert at the smaI site will disable the gene's function. Colonies which have the insertion in the vector will be white in color after incubation and those that do not have disrupted lacZ genes will be blue. Two plates were coated with the transformed cells and numerous white colonies were identified the next day. Six colonies were selected for miniprep and grown overnight at 37ºC in the shaker. The results of the miniprep were examined on two agarose gels which are included in Appendix 7.1. Two restriction enzymes were used to check for the 79bp inserts. EcoRI and HindIII were chosen which considering the total length of the fragments after cleavage should be 130bp. The 1KX5-1/2 samples showed a strong band at the expected 130bp length. The 1KX5-1/3 sample showed a strong band in lane 4 corresponding to the expected 130bp length. These were frozen down and saved for the reconstitution experiments.

3.3.3 CGG

FRAXA $(CGG)_n$ has been shown to be a low binder and the cause of Fragile X syndrome. The 484bp sequence containing the $(CGG)_n$ repeat, shown below, has been cloned into the vector pCR-Script Amp SK(+) (Appendix 7.1). The transformed vector insert is called pFXA53. Midiprep of pFXA53 was carried out using the pFXA53 bacterial vector clones. The concentration of the plasmid DNA was determined to be 330ng/µl from the midiprep.

484bp Insert

GGAACAGCGT TGATCACGTG ACGTGGTTTC AGTGTTTACA CCCGCAGCGG GCCGGGGGTT CGGCCTCAGT CAGGCGCTCA
GCTCCGTTTC GGTTTCACTT CCGGTGGAGG GCCGCCTCTG AGCGGGCGGC GGGCCGACGG CGAGCGCGGG CGGCGGCGGT
GACGGAGGCG CCGCTGCCAG GGGGCGTGCG GCAGCGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG
GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC
GGCGGCGGCG GCGGCGGCGG CTGGGCCTCG AGCGCCCGCA GCCCACCTCT CGGGGGCGGG CTCCCGGCGC TAGCAGGGCT
GAAGAGAAGA TGGAGGAGCT GGTGGTGGAA GTGCGGGGCT CCAATGGCGC TTTCTACAAG GTACTTGGCT CTAGGGCAGG
CCCC

BsahI cleaves the sequence at the position shown in red and XhoI cleaves the sequence in the region shown in magenta. Both of these restriction sites are also found in the vector presenting a resolution problem. Furthermore, fragments of length 147bp cannot be obtained due to no unique restriction site existing within the $(CGG)_n$ triplet repeat motif. Results of the restriction cleavage experiments are found in Appendix 7.1. The final preparation of purified DNA for the reconstitution experiments was conducted using polyacrylamide gel which could resolve 230bp fragments from 180bp fragments.

3.3.4 5S rRNA

This sequence has been cloned into the vector pSP64 and is referred to as pXP10. The 5S rRNA gene is highlighted below; in red is the first primer 5'-3' (17bp), in magenta is the second primer 5'-3' (20bp), and in orange is the complementary sequence of the third primer.

Xenopus Borealis 5S RNA gene

GGCCCCCCCC AGAAGGCAGC ACAAGGGGAG GAAAAGTCAG CCTTGTGCTC
GCCTACGGCC ATACCACCCT GAAAGTGCCC GATATCGTCT GATCTCGGAA
GCCAAGCAGG GTCGGGCCTG GTTAGTACTT GGATGGGAGA CCGCCTGGGA
ATACCAGGTG TCGTAGGCTT TTGCACTTTT GCCATTCTGA GTAACAGCAG
GGGGCAGTCT CCTCCATGCA TTTTTCTTTC CCCGAACAGC

The selected primers for the 5sRNA gene are listed to illustrate that this set of primers had a strong self dimmer 5S-2. Polymerase Chain Reaction experiments had to be modified to adjust for the dimerizing effects of the primer.

5S-1 17bp T_m: 55.5ºC GC:58.8% PD:No OD@260=0.068 OD@280=0.031 [DNA]=272ng/µl

5'-CCCAGAAGGCAGCACAA-3'

5S-2 20bp T_m: 63.5ºC GC:70% PD:**YES** OD@260=0.138 OD@280=0.088 [DNA]=552ng/µl

5'-CCTTGTGCTCGCCTACGGCC-3'

5S-3 20bp T_m: 51.0ºC GC:40% PD:No OD@260=0.263 OD@280=0.116 [DNA]=1052ng/µl

5'-GAATGGCAAAAGTGCAAAAG-3'

The prominent primer dimers formed within the 5S-2 sequence are displayed below.

```
5' CCTTGTGCTCGCCTACGGCC
                  ||||
3'                CCGGCATCCGCTCGTGTTCC

5' CCTTGTGCTCGCCTACGGCC
            |||   :::
3'          CCGGCATCCGCTCGTGTTCC
```

3.4 Competitive Nucleosome Reconstitution Experiments

The goal of the nucleosome reconstitution experiments was to determine the relative free energy of binding for the human histone core protein among the four sequences. Sequences of lengths longer than 147bp were not used with the exception of the CGG triplet reapeat. It was not possible given the short time in the lab to generate a triplet repeat of length 147bp therefore the previously prepared sequence of 180bp was used. The purified sequences are labeled with radioactive γ-P^{32} by using a kinase enzyme to add the radioactive phosphate to the 3' end.

Optical Density readings were taken for each purified sample to determine the concentration of DNA.

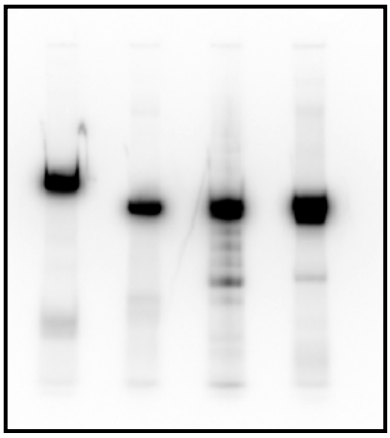
Sequence	OD@260nm	OD@280nm	Ng/uL
601	.016	.008	8.5
1KX5	.066	.039	35.2
(CGG)n	.022	.018	11.7
5S	.027	.021	14.4

The concentration of 601 was the lowest therefore more of that sample is used in the subsequent radio labeling experiment.

Kinase radio labeling is carried out by incubating the solution at 37ºC for 1 hour after adding the following to the samples:

601	1KX5	CGG	5S
43uL(365.5ng)DNA	10.4uL DNA	28.6uL DNA	25.4 uL DNA
5uL 10xBuffer	5uL 10xBuffer	5uL 10xBuffer	5uL 10xBuffer
1uL gamma-p32-ATP	1uL gamma-p32-ATP	1uL gamma-p32-ATP	1uL gamma-p32-ATP
1uL Kinase	1uL Kinase	1uL Kinase	1uL Kinase
0uL dH2O	32.6uL dH2O	14.4uL dH2O	17.6uL dH2O

After incubation 2µL (7.35ng/µL) of each sample was run on a 5% polyacrylamide gel @125V for 2 hours to verify they had been correctly labeled. The gel was transferred to filter paper and placed in the phosphor imaging plate overnight. The plate was then scanned the following morning and all samples showed excellent labeling. Reading from left to right, lane 1 is CGG, lane 2 is 5S, lane 3 is 601 and lane 4 is 1KX5. Note that lane 3 shows some contamination of other DNA fragments but does not affect the subsequent experiments. This could be due to 601 being purified from the agarose gel instead of a polyacrylamide gel. However, 5S which shows a well defined band was also purified from the same agarose gel. 1KX5 and CGG were purified from a polyacrylamide gel. Once the samples have been labeled the reconstitution experiments can be carried out.



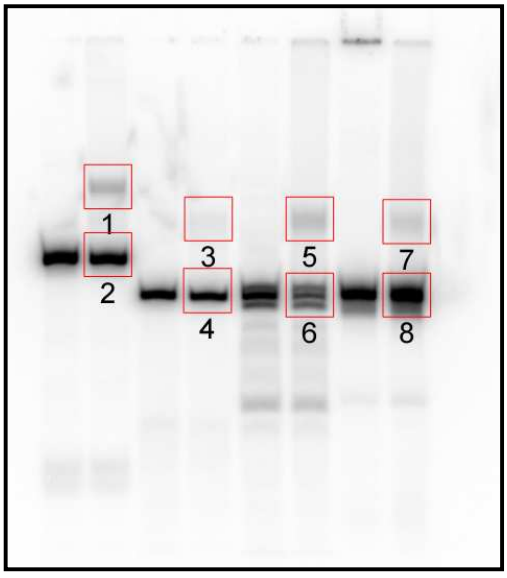
The first reconstitution experiment is carried out using histone protein denoted with a + and no histone protein denoted with a -, for each sample. The following table lists, in order, reagents added to carry out the reconstitution experiments.

	601+	**601-**	**1KX5+**	**1KX5-**	**5S+**	**5S-**	**CGG+**	**CGG-**
4M NaCl(uL)	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1
10% 1GEPAL(uL)	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
100XBSA(uL)	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
Calf Thymus DNA(10ug/uL)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Radio Labeled DNA 50ng total (uL)	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8
Buffer A(uL)	8.8	8.8	8.8	8.8	8.8	8.8	8.8	8.8
Histone Protein(uL)	5.0	0.0	5.0	0.0	5.0	0.0	5.0	0.0
2M NaCl(uL)	0.0	5.0	0.0	5.0	0.0	5.0	0.0	5.0
Total Volume(uL)	32.0	32.0	32.0	32.0	32.0	32.0	32.0	32.0

[NaCl](M)	Buffer B added (uL)	Total Volume (uL)
2.0	-	32.0
1.9	1.7	33.7
1.8	1.9	35.6
1.7	2.1	37.7
1.6	2.3	40.0
1.5	2.7	42.7
1.4	3.1	45.8
1.3	3.5	49.3
1.2	4.1	53.4
1.1	4.9	58.3
1.0	5.7	64.0
0.9	7.1	71.1
0.8	8.9	80.0
0.7	11.4	91.4
0.6	15.2	106.6
0.5	21.3	127.9
0.4	32.1	160.0
0.3	53.3	213.3
0.2	106.7	320
0.1	320	640

After adding the histone protein or 2M NaCl the solution sits for 5 minutes. The next series of steps in the experiment involve adding increasing amounts of Buffer B to each of the solutions which effectively lowers the concentration of NaCl. After each addition the solutions sat for 5 minutes before the next addition.

After completion of the reconstitution experiments samples can then be run on polyacrylamide gels to determine what percentage of DNA bound to the histone protein. Using 40uL of prepared sample and 8uL of dye the mixtures are loaded into a 5% polyacrylamide gel and run for 2 hours at 125V. The first gel that was run had to be thrown out because the concentration of TBE in the buffer solution and the concentration of TBE in the gel were different causing the samples to smear as they ran. Correcting for this error another gel was prepared using the correct buffer concentration which produced the following gel image. Lanes are counted from left to right; L1 CGG-, L2 CGG+ L3 5S-, L4 5S+, L5 601-, L6 601+, L7 1KX5-, L8 1KX5+. The odd numbered boxes highlight the retarded bands created by DNA bound to the nucleosome protein. The even numbered boxes are the free DNA samples which have run faster through the gel. Taking the ratio of intensity counts in the areas demarked by the boxes a general ranking can be established.



3.4 Analysis and Determination of Ranking

Since only 180bp constructs of CGG could be used in the reconstitution experiments analysis was only conducted on the 147bp sequences 1KX5, 5S rRNA, and the synthetic clone 601. Three separate reconstitution experiments were performed and several gels were run for each reconstitution yielding several sets of data. Using collected data from a total of eight trials the average relative free energy for 601 and 1KX5 was computed to be -0.63 kcal/mol and -0.25 kcal/mol with standard deviations of 0.29 and 0.08, respectively. The data supports the conclusion that the 601 sequence is the stronger binder compared to the palindrome 1KX5 sequence. Previous studies relative to

similar length sequences as the 180bp CGG triplet repeat have proved its affinity to be less than 5S rRNA. It is reasonable to place the affinity for the 147bp triplet repeat which is used in the computational model as the weakest binder in the set.

4 Simulations of Substituted Sequences

Three simulations were run using the same core histone proteins for comparison to experimentally determined free energy binding experiments. The three sequences were the palindrome sequence from 1KX5, the clone 601 sequence, and the triplet repeat CGG sequence. The clone 601 and CGG triplet repeat sequences were positioned in the same position as the palindrome sequence found in the crystal structure. The positioning of new substituted sequences on the crystal structure is considered in the next chapter. Here results of the three simulations are presented.

4.1 Preparation of Molecular Dynamic Simulations

The 1KX5 native crystal structure was solvated using GROMACS with all crystallographic histone tails intact. The recombinant proteins used to determine the crystal structure were without post-translational modification [21]; hydrogen atoms were added with the Tripos BioPolymer module and C-terminal and N-terminal ends were patched using CHARMM. The ions found in the crystal structure were included by (1) retaining the 4 Cl^{1-} ions and (2) substituting the 14 Mn^{2+} with 14 Mg^{2+}.[5] To neutralize the remaining charge, 120 Na^{1+} counter ions were added at random positions. The system was solvated in a 150 x 220 x 135 $Å^3$ rectilinear box. The final system contained 25064 solute atoms (15470 protein atoms, 9324 DNA atoms, and 138 counter ions) and 418653 water atoms (3130 of the water molecules were from the crystal structure). Thus, the total number of atoms in the 1KX5 simulation was 443855 atoms.

4.2 Root Mean Square Deviations and Fluctuations of DNA

Root mean square deviation is traditionally used to measure the scalar distance between atoms of the same type for two structures. In these calculations the center of mass of a group of atoms is used to compare the average spatial deviations between structures in time and the original structure (at time = 0ns). Drift and rotational aberrations of the entire simulation system are accounted for by first computing an optimal alignment of a particular structure at a point in time compared to the initial structure. The structure at a particular point in time is then moved to the optimally

[5] The Mn^{2+} ion parameters detected in the crystal structure were not available in the CHARMM force field. Therefore, the Mn^{2+} positions were substituted with Mg^{2+}. Mg^{2+} is the most prevalent divalent cation in physiological systems, and its volume is smaller than Mn^{2+}. Thus, modifying the force field was avoided, while retaining the divalent cations positioning consistent with the crystal structure. Although the coordination chemistry between the two ions differs, the classical model does not explicitly account for the molecular orbitals.

aligned position effectively removing the aberrations before computing the RMSD[6]. To calculate the RMSD of a pair of structures (say x and y), each structure must be represented as a 3N-length (assuming N atoms) vector of coordinates. The RMSD is the square root of the average of the squared distances between corresponding atoms of x and y. It is a measure of the average atomic displacement between the two conformations:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|^2}$$

In this RMSD analysis the position of each nucleic acid residue in a structure throughout the simulation is compared to the positions of the residues in the initial structure. Increased deviations from the initial structure exemplify more movement compared to the initial structure. A more stable structure will exhibit less deviation throughout a simulation. Experimentally, large RMS deviations of nucleosome crystal structures containing poly (dA/dT) sequences show distorted structures providing further evidence to our approach [35].

A *tcl* script has been used to unpack the trajectory data, determine the alignment matrix, and compute the RMSD for nucleic acid residues. A single base is defined as a nucleic acid residue enabling us to look at the average deviations of all 147 residues in each chain. (Chain I is defined to be the 5' to 3' orientation and the J chain is defined to be the 3' to 5' orientation of the sequences listed in Appendix 7.4.) A comparison of the average RMSD for the nucleic acid residues after 6.9ns and 23.7ns of simulation time is shown in figures 12 and 13, respectively. The large deviation of the 1KX5 palindrome sequence indicates instability in DNA bound to the protein core.

To investigate this further we looked at the contribution of individual nucleic acid residues and refer to the analysis as root mean square fluctuations. Here only 1KX5 and the 601 sequence are considered due to the lack of simulation time for the CGG sequence. Two representations are used to illustrate the data. The first two figures are surface plots corresponding to each nucleic acid residue along the X-axis for chain I, simulation time along the Y-axis, and RMSD along the Z-axis. These surface plots are then projected down into the XY plane and shown in figures 16 and 17. The overall lighter shade of blue in the 1KX5 surface plot indicates the overall higher average RMSD. The dark blue regions (low RMSD) exemplifying stability by the 601 sequence (fig 17) can be seen falling in tracts spanning residues {31-41}, {66-75}, {81-87}, {97-105}, and {121-125}. Both simulations show large deviations towards the ends of the DNA which is to be expected.

[6] This technique is often referred to as LRMSD for least root mean square deviation. A sample of the *tcl* script that performs this calculation is included in appendix 7.2.

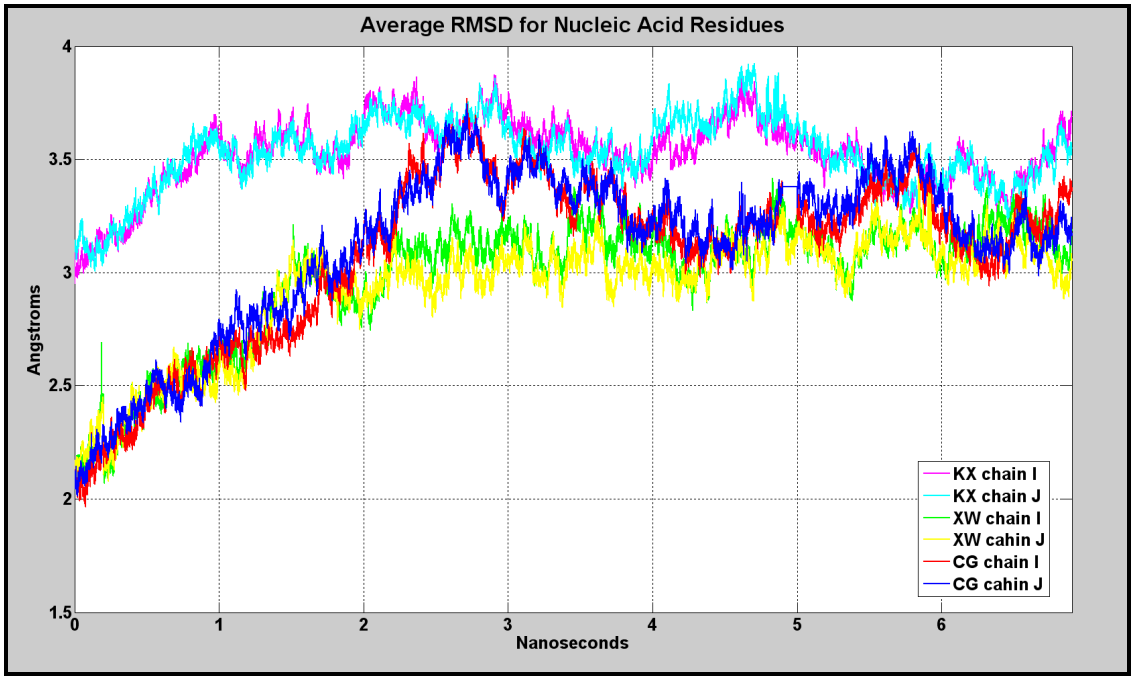


Figure 12. **This plot shows the average RMSD of each residue in the DNA sequence broken down into contributions of chain I and J for 6.9ns of simulation time. (In all figures, KX corresponds to the 1KX5 simulation DNA, XWF corresponds to clone 601 DNA, and CG corresponds to the CGG triplet repeat DNA.)**

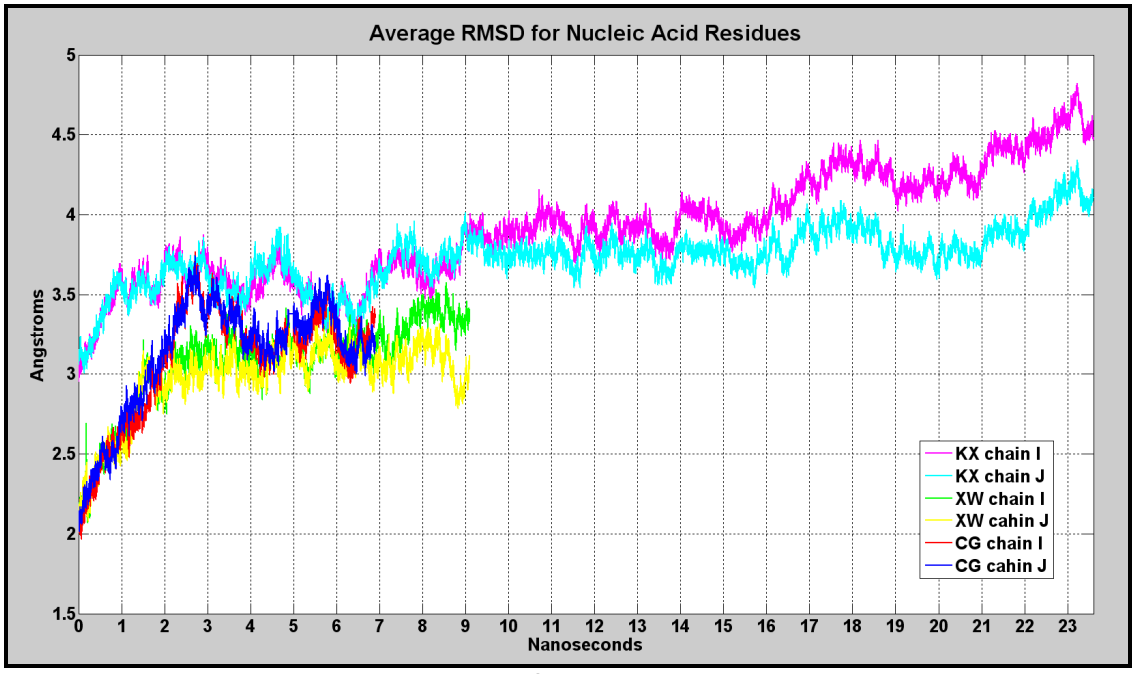


Figure 13. **This plot shows the average RMSD of each residue in the DNA sequence broken down into contributions of chain I and J for 6.9ns of simulation time.**

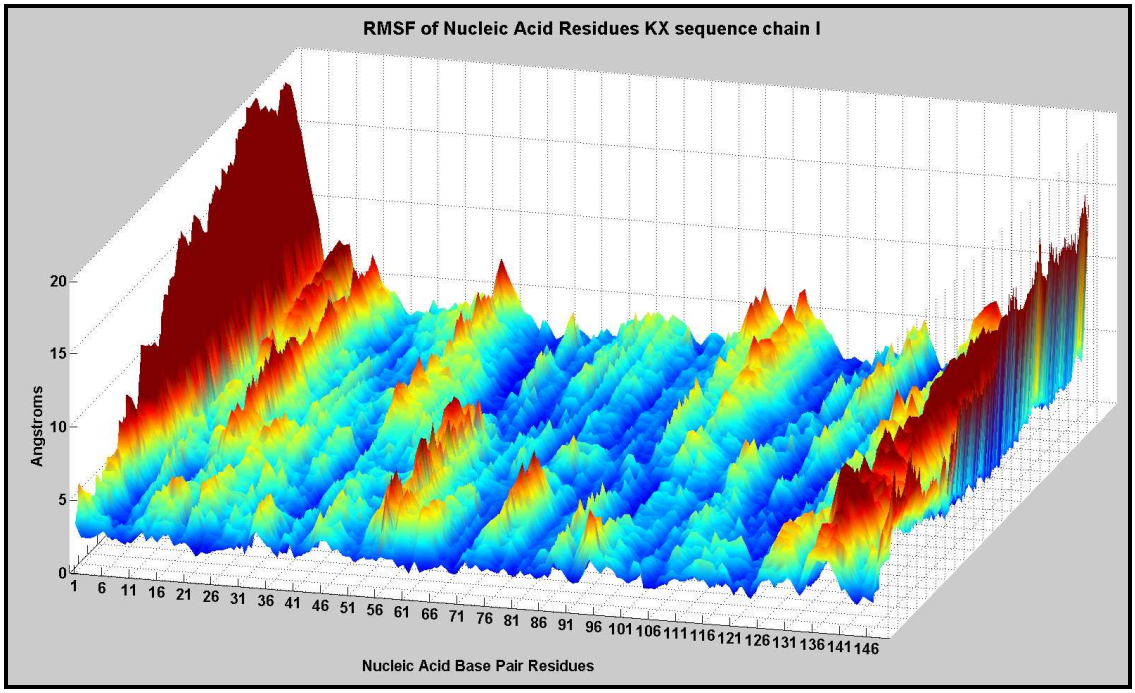


Figure 14. **Root Mean Square Fluctuations of the nucleic acid residues for 1KX5. Plotted along the Y-axis is simulation time out to 23.7ns.**

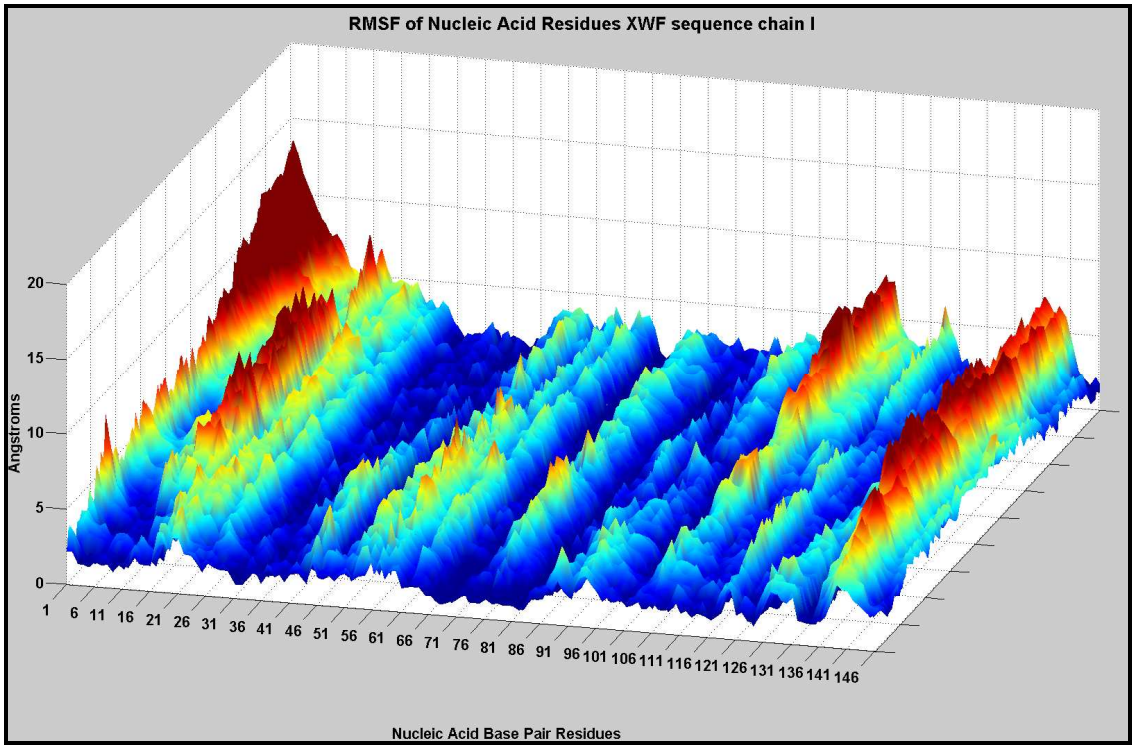


Figure 15. **Root Mean Square Fluctuations of the nucleic acid residues for 601. Plotted along the Y-axis is simulation time out to 9.7ns.**

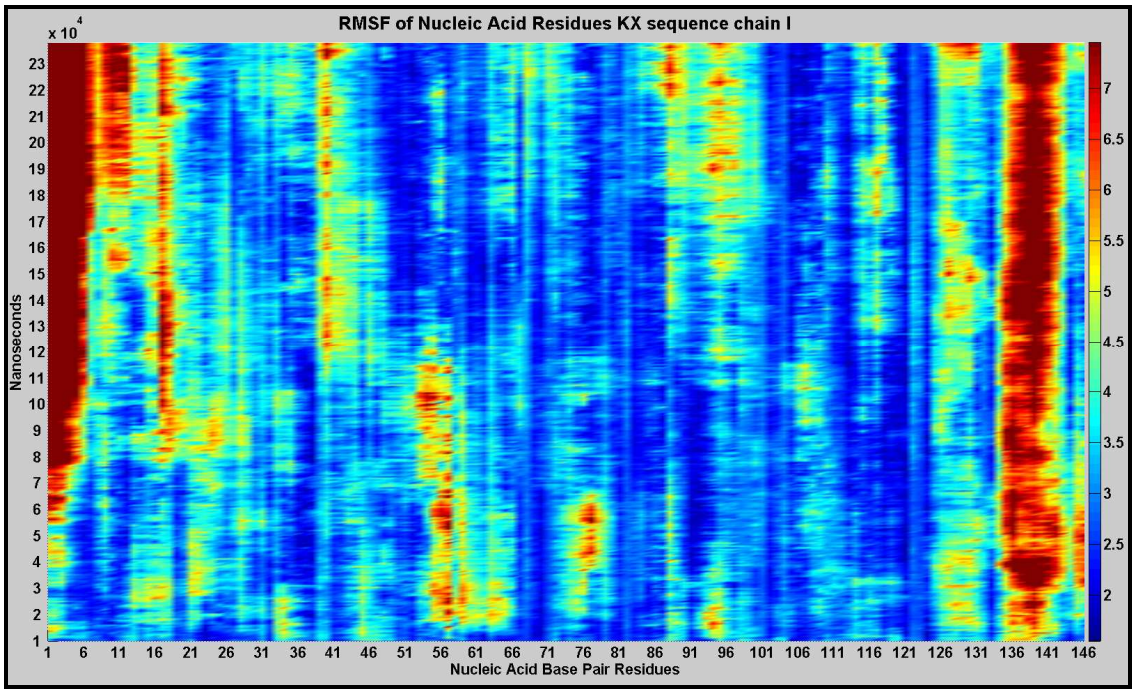


Figure 16. **Root Mean Square Fluctuations of the nucleic acid residues for 1KX5. Plotted along the Y-axis is simulation time out to 23.7ns.**

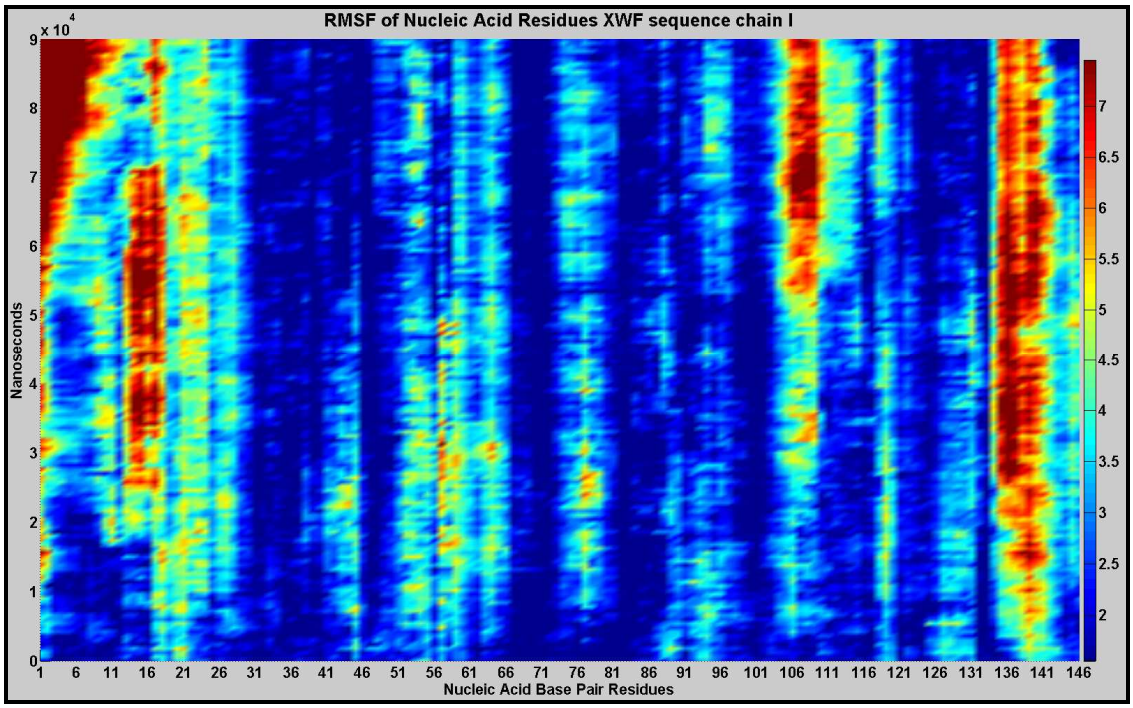


Figure 17. **Root Mean Square Fluctuations of the nucleic acid residues for 601. Plotted along the Y-axis is simulation time out to 9.7ns.**

4.3 Root Mean Square Deviations and Fluctuations of Histone Protein

Using the same approach used to calculate the RMSD for nucleic acids residues we computed the RMSD for protein residues throughout our simulations. Here only the proteins in 1KX5 and 601 are considered. The large deviations for the histone proteins H3 (A) and H3`(E) can be seen in the 1KX5 simulation (fig. 18). Similarly, 601 also shows large deviations in the H3 (A) histone protein and less in the H3`(E) protein at 9.7ns. These two proteins contain the long tails and start the simulations stretched out into the solvent where they have freedom to move about. A per residue RMSF analysis was performed on each residue in each of the two proteins characterizing the large deviations. XZ projections of the RMSF for the two proteins in each simulation are shown in (fig. 20 and fig. 21) for 1KX5 and (fig. 22 and fig. 23) for 601. Locations of the two proteins within the nucleosome can be seen in figure 7 where the colors used here correspond to those in figure 7. The low RMSF seen in residues {41-135} are characteristic of protein residues located away from the surface. A measure of solvent accessibility would reveal little surface exposure in these regions.

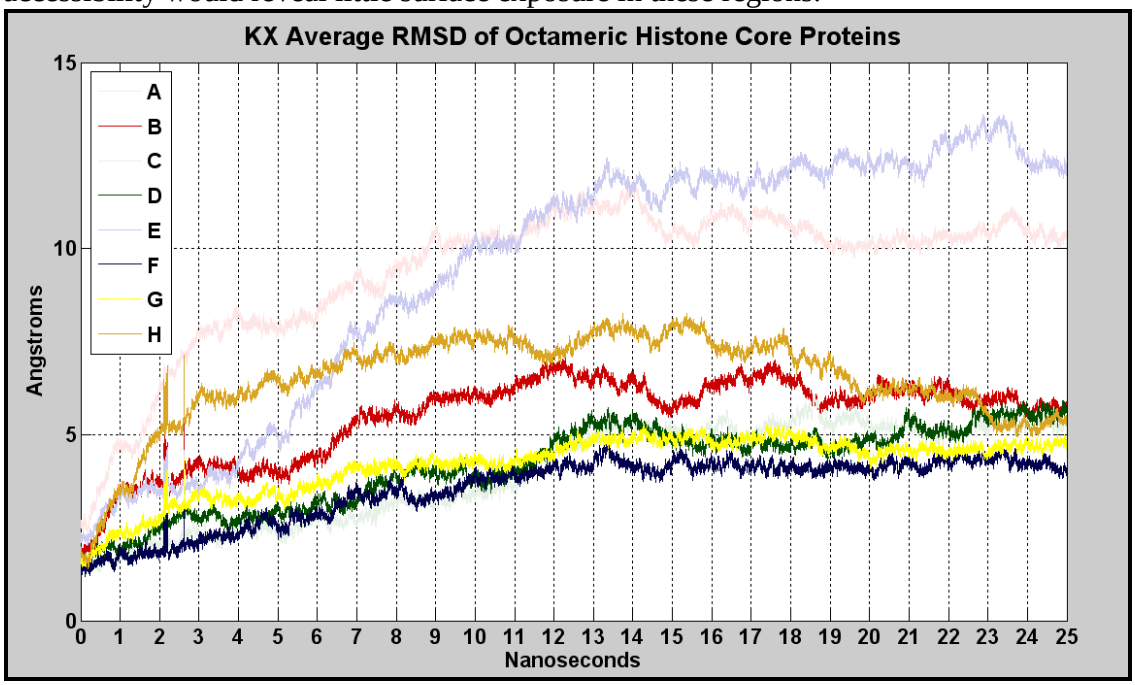


Figure 18. **Histone proteins plotted are H3 (A), H4 (B), H2A.1(C), H2B.2(D), H3`(E), H4`(B), H2A.1`(C), and H2B.2`(D). Plotted here are the RMSD for protein residues in 1KX5.**

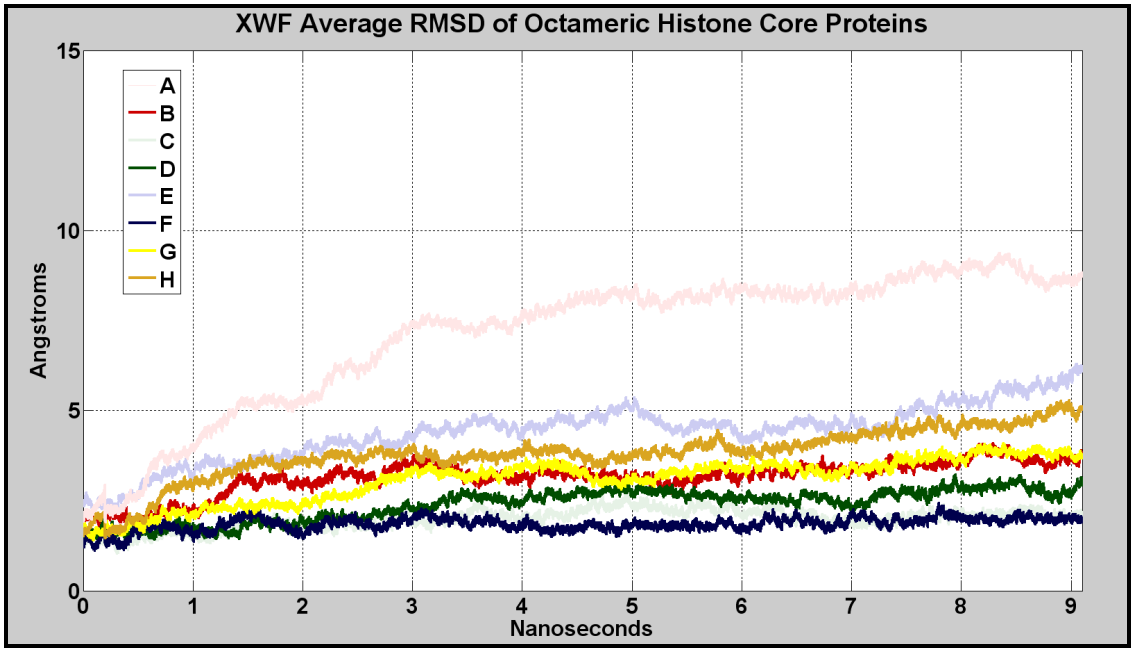


Figure 19. **Histone proteins plotted are H3 (A), H4 (B), H2A.1(C), H2B.2(D), H3`(E), H4`(B), H2A.1`(C), and H2B.2`(D). Plotted here are the RMSD for protein residues in 601.**

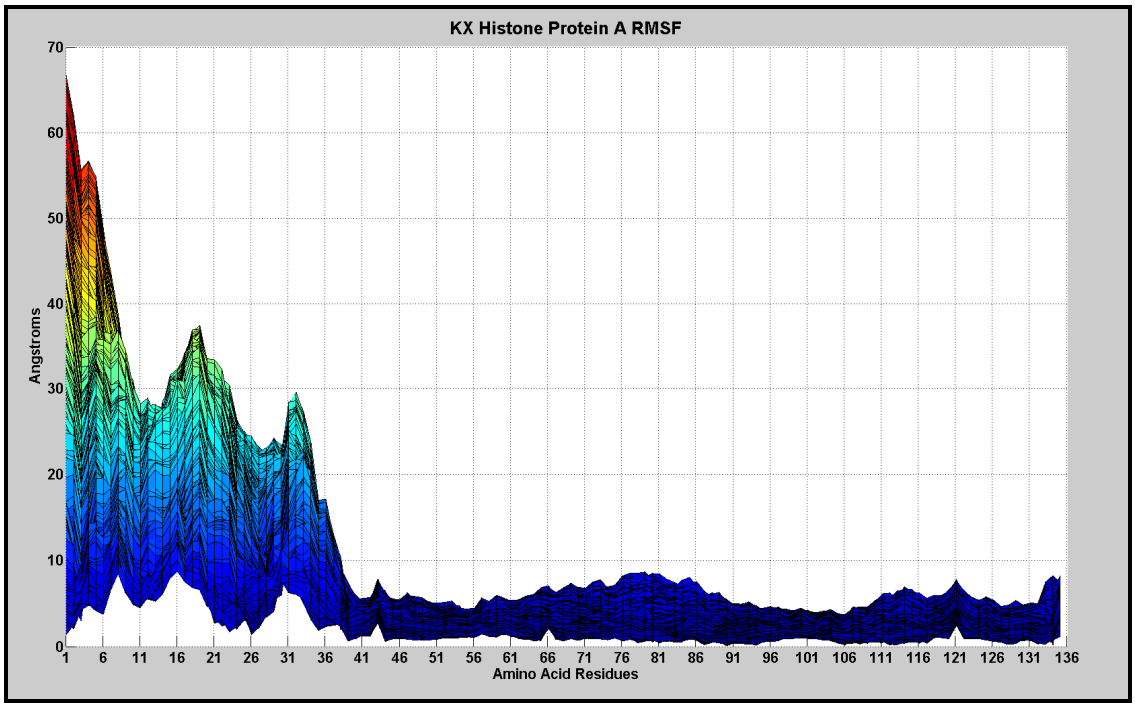


Figure 20. **XZ projection of histone protein H3(A) RMSF after 23.9ns for the 1KX5 system.**

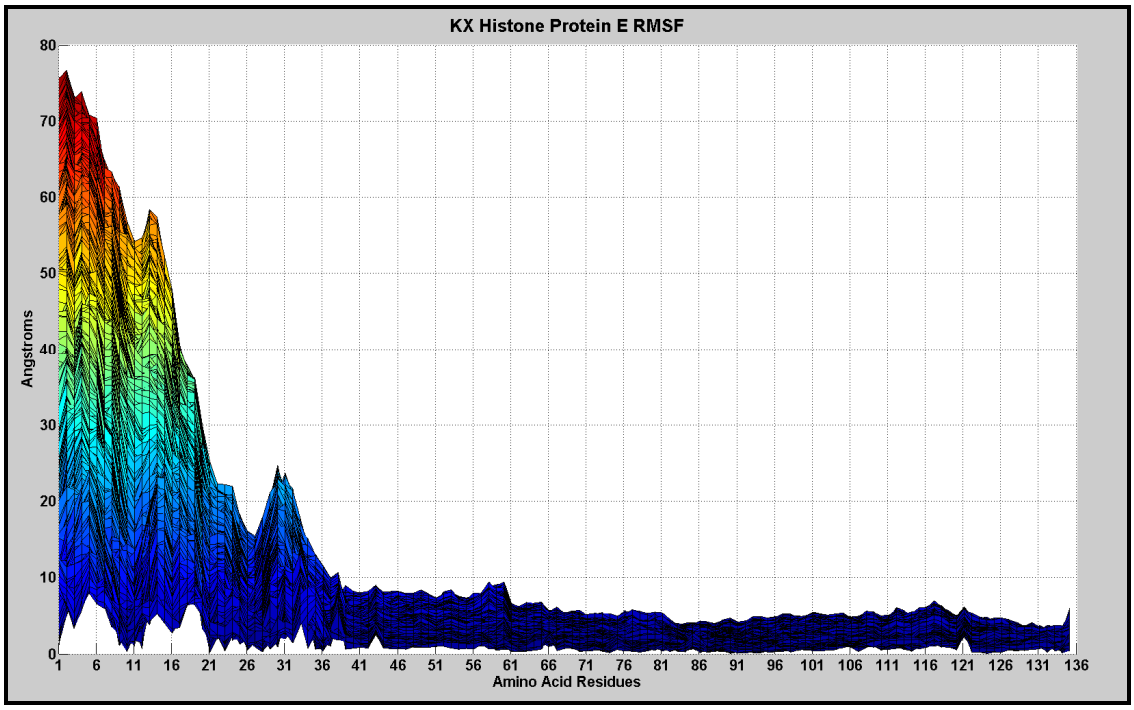


Figure 21. **XZ projection of histone protein H3`(E) RMSF after 23.9ns for the 1KX5 system.**

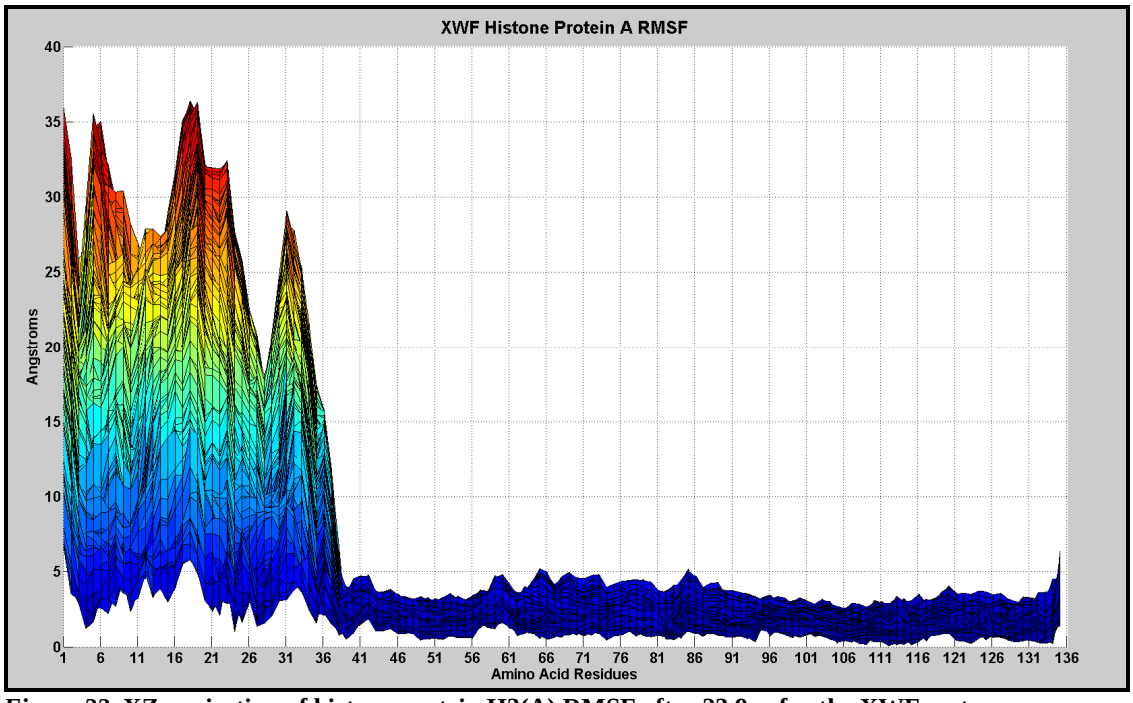


Figure 22. **XZ projection of histone protein H3(A) RMSF after 23.9ns for the XWF system.**

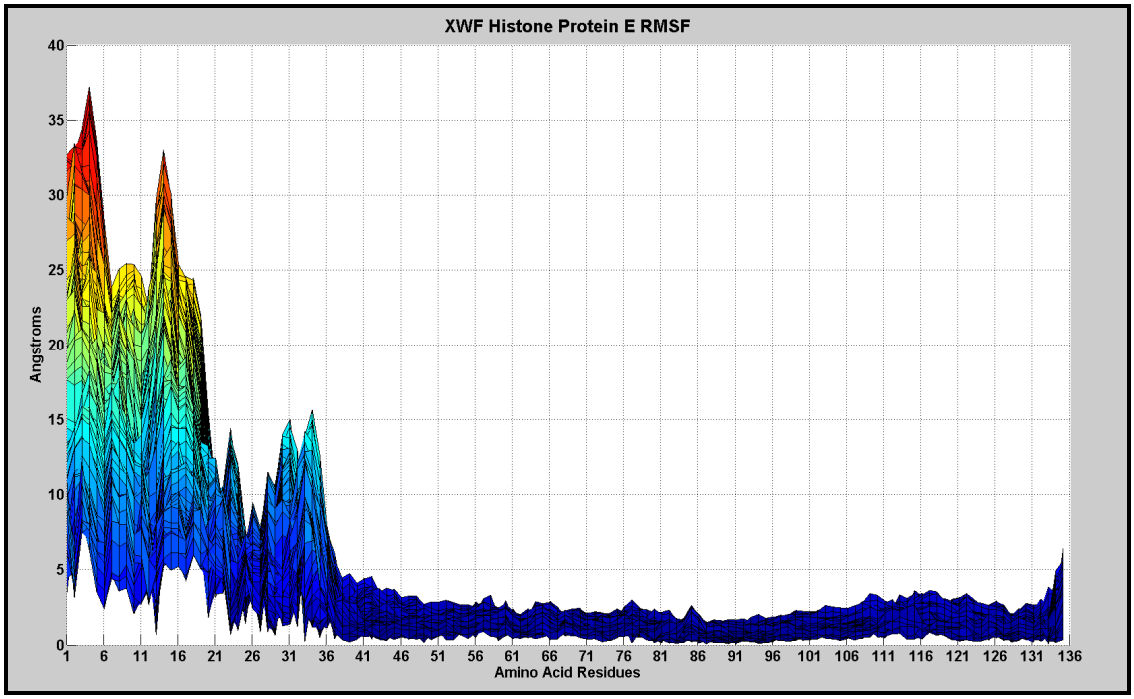


Figure 23. **XZ projection of histone protein H3`(E) RMSF after 23.9ns for the XWF system.**

4.4 Helical Parameters and Curvature

The helical parameters are a geometric representation of each individual basepair within the helix. The parameters are depicted in figure 24 and can be computed using both the Curves package and the X3DNA package. Here results of our two simulations are compared to several values computed by previous studies (Bishop and Zacharias). The parameters are first computed for each structure and then averaged over the entire course of the simulation. After this step the data has been reduced to 146 averages corresponding to the basepairs of the DNA sequence. Next an average is computed for all 146 base pairs providing a table of average roll, tilt, twist, shift, slide, and rise. Combining roll and tilt in the following manner provides an average bend per base pair:

$$Bend = \sqrt{(Roll)^2 + (Tilt)^2}$$

The results of the calculations are tabulated in figure 25. The necessary curvature can be created with a Bend of 360°/turn * 1.84turns/146bps ~ 4.5°/bps. Looking at the average bend after 10ns of simulation we see that the 601 sequence has slightly greater bend as compared to 1KX5.

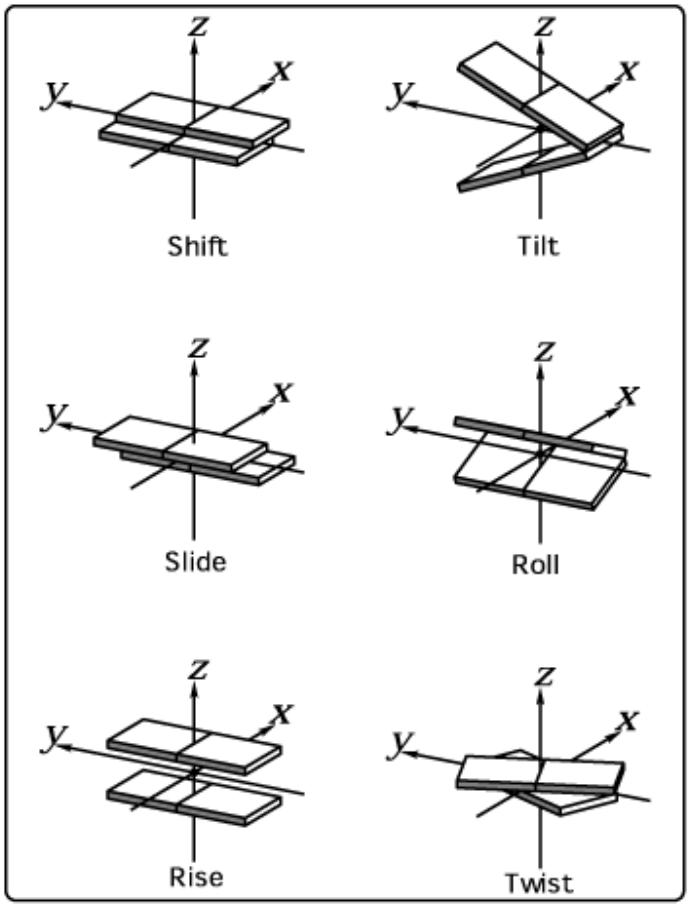


Figure 24. **Helical parameters**

Helical Parameter	<avg> 1KX5,20ns	<avg> 1KX5,10ns	<avg> 601,10ns	<avg> Bishop 1KX3,10ns	<avg> Zacharias 1KX5, 20ns	<avg> B-DNA
Roll(°)	3.5430	3.7600	4.2432	2.43	-	1.7
Tilt(°)	-0.1703	-0.2889	0.0411	0	-	0
Twist(°)	34.3509	34.1096	34.3157	34.3	34.5	36
Bend(°)	6.3956	6.0846	6.0966	7.01	-	-
Shift(Å)	-0.0179	-0.0246	0.0083	0	-	0
Slide(Å)	0.2714	0.2427	0.3565	-0.25	-	0.45
Rise(Å)	3.3879	3.3957	3.3840	3.32	-	3.36

Figure 25. **Tabulated time averaged base pair parameters for 1KX5 (20ns, 10ns), 601 (10ns), the Bishop 1KX3 simulation (10ns), the Zacharias 1KX5 simulation (20ns), and B-DNA (calf thymus). Interestingly the calf thymus DNA is what was used as the competitor in the reconstitution experiments. Standard deviations were computed but not listed.**

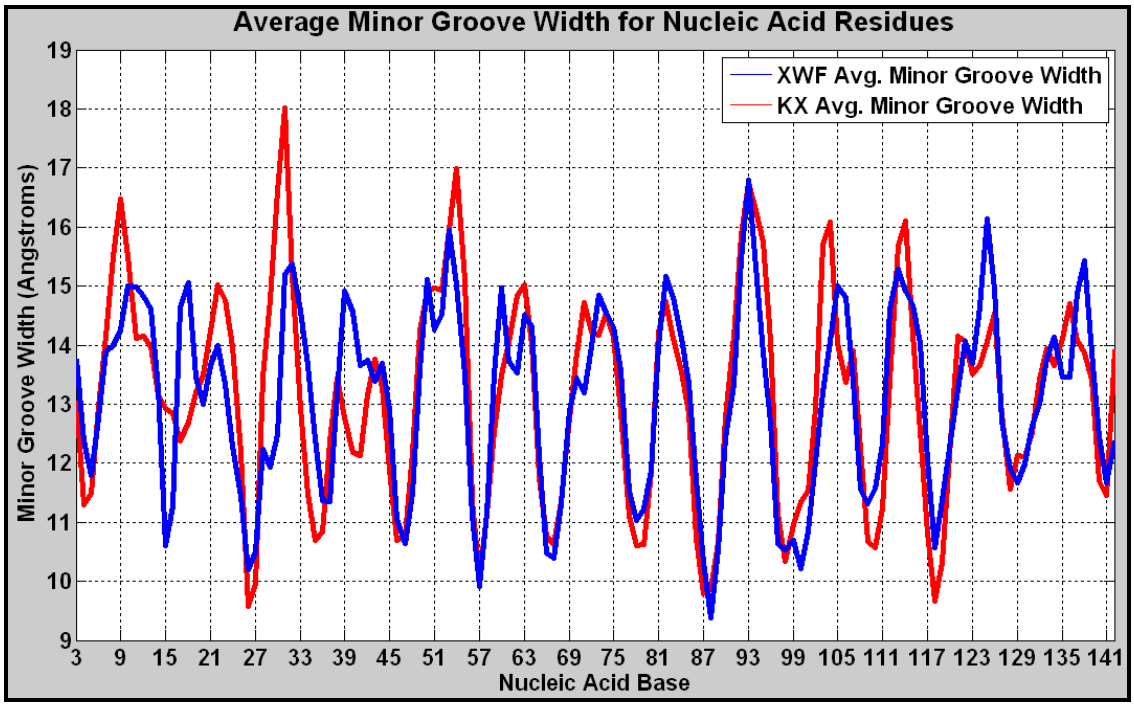


Figure 26. **Minor grove widths of DNA collected from X3DNA averaged over time. The large changes in minor groove widths indicate kinked regions of DNA. These regions occur more frequently in the KX structure providing further evidence of structurally unstable DNA.**

4.5 Dynamic Cross-Correlations

The covariance can address the dynamical situation of the structural tendencies of the DNA [49]. It is suspected that low-stability nucleosomes exhibit less covariance based on Ramaswamy's elastic network dynamics [44]. Dynamic cross-correlations (DCC) of the displacements from the mean position of selected carbon α atoms in the protein and phosphate atoms in the DNA backbone are computed. The dynamic cross correlation map reveals coordinated motions of the selected atoms in a simulation [33][34]. The elements of the DCC map (C_{ij}) were computed as:

$$\mathbf{C_{ij} = < \Delta r_i \cdot \Delta r_j > / [sqrt(< \Delta r^2_i >) \cdot sqrt(< \Delta r^2_j >)]}$$

Where Δr_i is the displacement from the mean position of the *i*th atom and the < > represent the time average over the whole trajectory. Positive values of C_{ij} refer to correlated motions between residues *i* and *j*, i.e the residues move in the same direction. Negative values of C_{ij} represent an anti-correlated motion between residues *i* and *j*, i.e. they move in opposite directions.

In this analysis the Carbon-α atoms in the protein backbone are selected along with the phosphate atoms in the DNA backbone (both I and J chains are mapped). Only the two H3 histone proteins are considered here. The tails of the two histone proteins fall in the regions between {1-75} and {136-211} and can be seen strongly correlated with the ends of the DNA {325-350}, {400-435}, and {475-515}. This illustrates the cross talk between the proteins and DNA facilitating the attachment of the ends to the core. We also note the strong correlations within the core of the proteins due to relatively little

motion within the core of the nucleosome. Comparing the two DCC maps of 1KX5 and 601 at 10ns stronger correlations can be seen at the ends of the DNA in the 1KX5 simulation. This is in agreement with the large RMSD values found at the ends of the DNA chains. We also can observe correlations grow more defined after taking a longer time average of 1KX5 (fig. 29). The histone protein H3 (A) can be seen gaining more correlations when comparing figure 28 and figure 29.

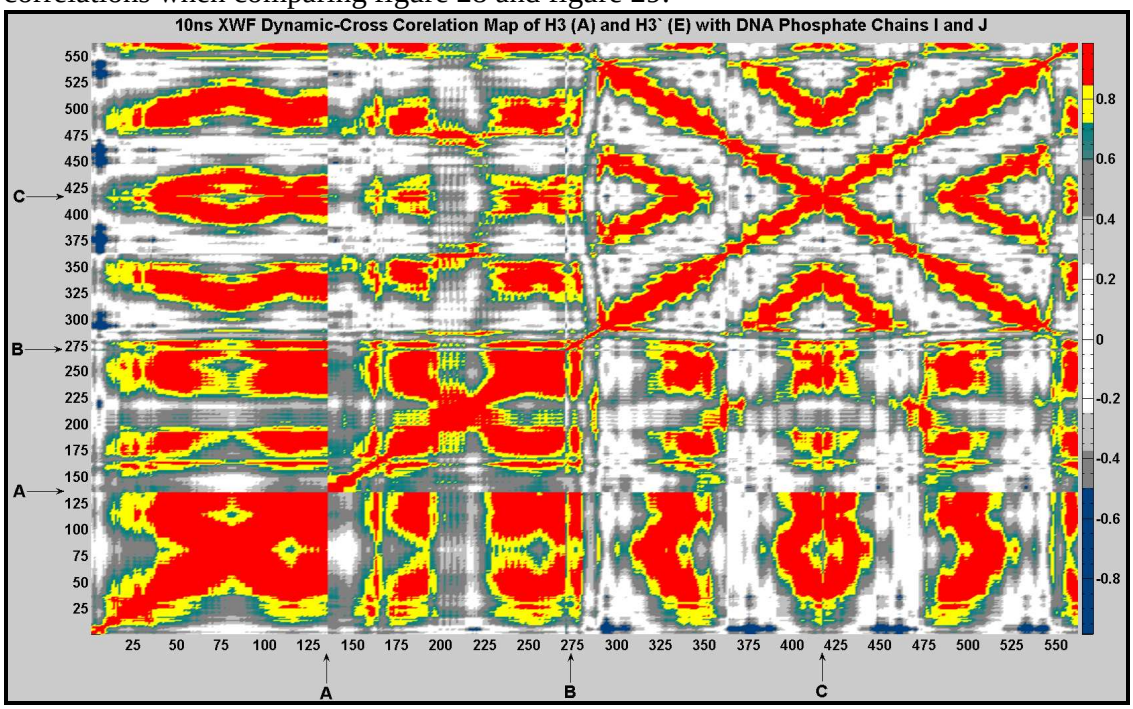


Figure 27. **The region 1 to 135 are the protein residues found in histone H3 (A) the start of the next 135 residues of histone H3`(E) is demarked with the letter A and the end is demarked with the letter B. The next 146 residues are the phosphates found in the DNA backbone on chain I and begin at letter B and end at letter C. The remaining 146 residues are the phosphates found in chain J.**

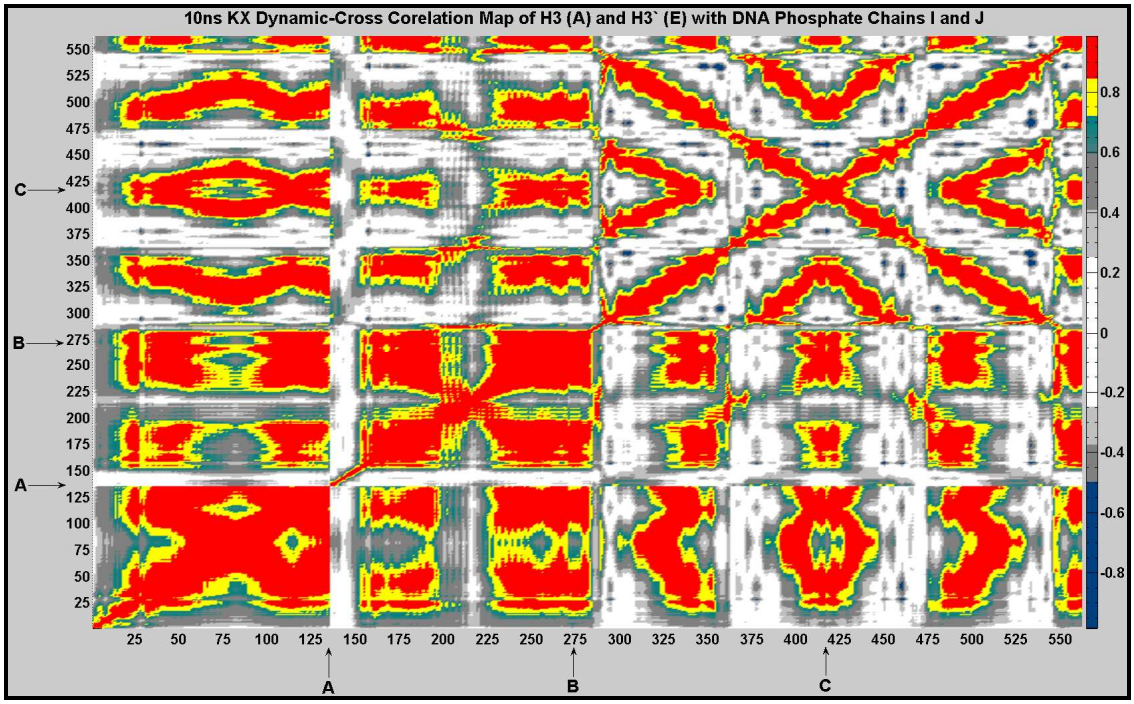


Figure 28. **The region 1 to 135 are the protein residues found in histone H3 (A) the start of the next 135 residues of histone H3`(E) is demarked with the letter A and the end is demarked with the letter B. The next 146 residues are the phosphates found in the DNA backbone on chain I and begin at letter B and end at letter C. The remaining 146 residues are the phosphates found in chain J.**

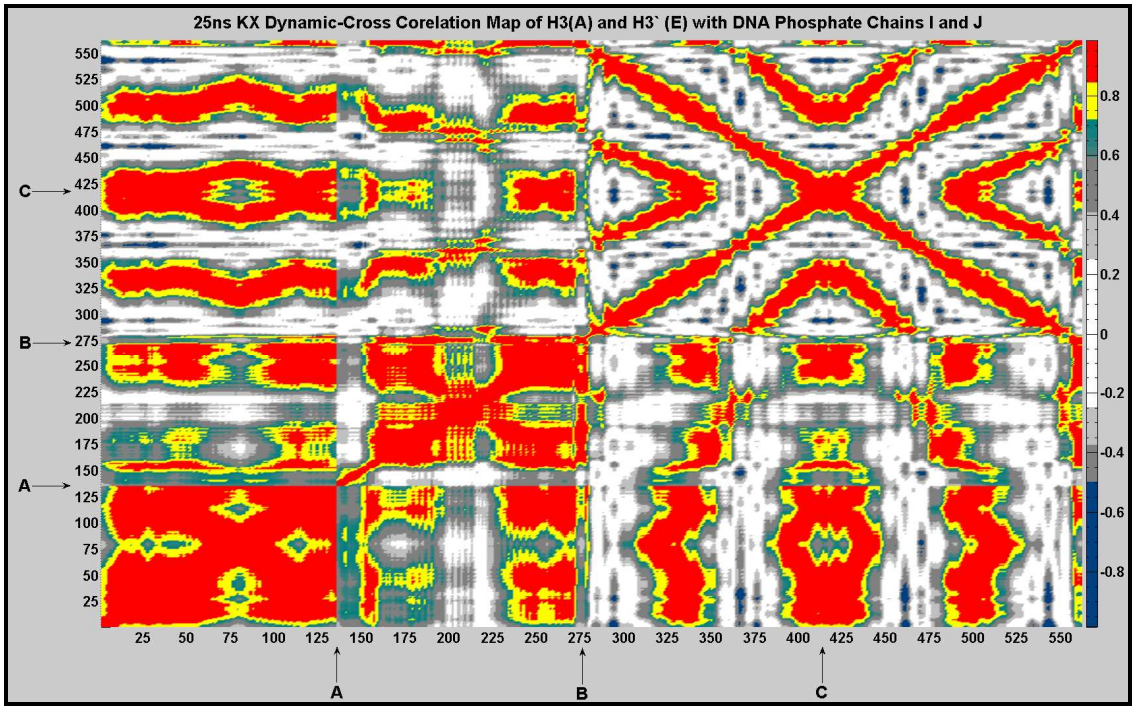


Figure 29. **The region 1 to 135 are the protein residues found in histone H3 (A) the start of the next 135 residues of histone H3`(E) is demarked with the letter A and the end is demarked with the letter B. The next 146 residues are the phosphates found in the DNA backbone on chain I and begin at letter B and end at letter C. The remaining 146 residues are the phosphates found in chain J. This map illustrates averaging over a greater time length.**

5 Rotational Conformations of Substituted Sequences

Substitution of new sequences in place of the palindromic sequence on the nucleosome core protein provides a means to explore structural properties of new sequences bound to the nucleosome. This method by itself is valuable due to the difficulty of producing high resolution crystal structures with sequences bound other than the palindromic sequence. Moreover, it is costly and time consuming to produce a crystal structure. The substitution method makes the assumption that the new sequence is positioned in the same position as the palindromic sequence. However, experimental evidence shows that DNA sequences do have a preferred orientation when bound to the nucleosome. The following describes the new method developed for generating and determining the most favorable orientation of substituted DNA on the nucleosome core protein by energetically sampling rotational conformations.

5.1 Experimental Evidence for Rotational Preferences

In a study carried out by Johnathan Widom he showed that there is a single preferential rotational state for the clone 601 shown in lanes 3 and 4 of figure 12. In contrast the 5S rRNA sequence has several rotational conformations exhibited by the numerous dark bands in that assay. This evidence motivated the search for the preferred rotational conformation of the substituted 601 sequence.

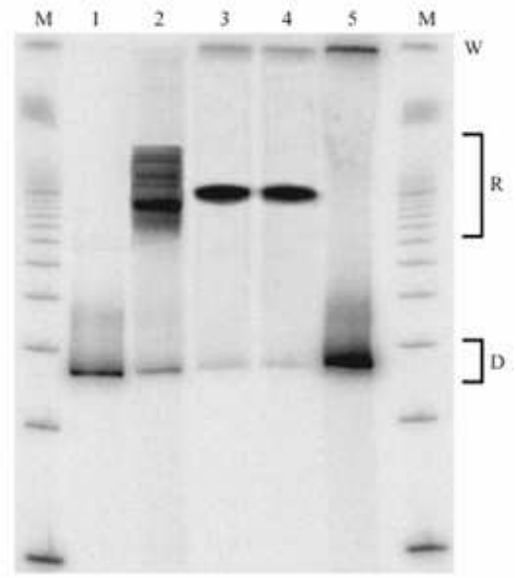


Figure 30. **Rotational Positioning Preference of DNA. Lanes M are 100bp markers, Lanes 1 and 5 are the 5s rRNA and 601.3b sequence naked DNA without histone protein. Lanes 3 and 4 are the 601.3b sequence with histone protein exhibiting a single preferred rotational position.**

5.2 Development of a Computational Method

The solvated system after substitution of the clone 601 sequence was used as the starting point for the rotation algorithm. The rotation was performed iteratively. Each iteration cycle begins by computing the curvilinear axis for the DNA. This was done by first separating the 3-dimensional DNA helix from the nucleosome in preparation for input to Curves [22]. Curves computes the curvilinear helical axis as a least squares fit to 732 points chosen from the double helix. In this process, the helical axis is weighted by mass; as a result, the curvilinear axis favors a position closer to the purines. Next, the software package Visual Molecular Dynamics (VMD) is used to rotate each nucleotide pair (complementary nucleotides) an increment of degrees around the curvilinear axis. In each cycle of the iteration all nucleotides are rotated in the same direction by a constant, small increment. This was done as follows.

A *tcl* script driving VMD loads the curvilinear helical axis data and generates 147 vectors. The vectors are formed for each base pair by selecting two points from the helical axis data. An index into the helical axis data is computed based on the residue number *r*, the details are listed in Appendix 7.2. Figure 1 illustrates the final vector formed, highlighted in red, for the last base pair in the sequence to be rotated. VMD defines a nucleic residue as the nucleic acid base (shown in blue), the deoxy-ribose sugar (shown in green) and the backbone linked to the sugar. The vectors serve as a local reference for the VMD rotate function which performs the rotation of each base pair. All base pairs are rotated simultaneously in a single iteration of the algorithm. The base pairs are rotated clockwise in increments of 3 degrees. Larger increments were found to cause breaks in the phosphate backbone due to excessive translational distance between neighboring phosphate and oxygen atoms.

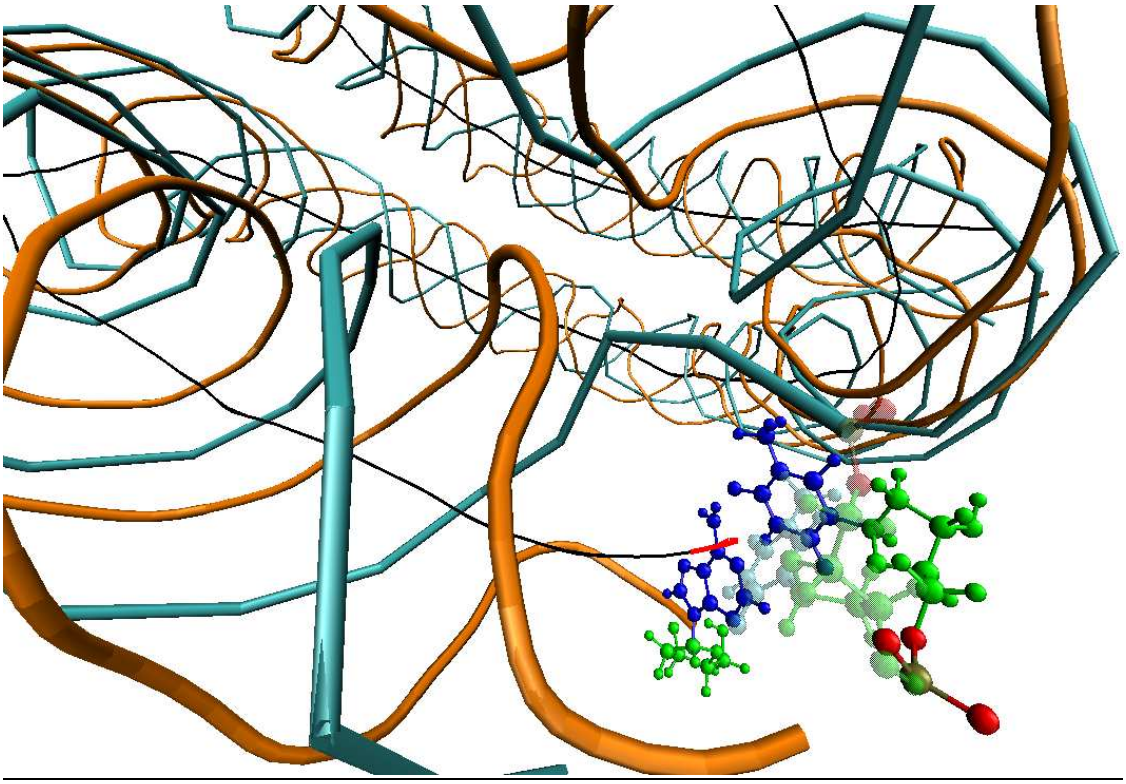


Figure 31. **Using VMD a rotational conformation of 111 degrees is shown in orange and is compared to the initial position of the structure shown in cyan. The curvilinear helical axis computed by Curves is displayed in black along with a rotation vector highlighted in red. The base pair that is rotated around the red vector is broken down into three components, in blue is the nucleic acid base, in green are the deoxyribose sugar and backbone components. The phosphate and oxygen atoms O1P', O2P, and O5' are shown for the current 111 degree rotational conformation and a translucent representation of the same atoms in the initial structure is displayed providing further reference. Rotations are performed in a clockwise manner.**

Backbone breaks at covalent bonds are the result of the backbone not adjusting to the new curvature imposed by the rotation. Moreover, the rotation method has to accommodate the reorientation of the helix curvature to the new rotational position. Generally, the preferential changes to the DNA double helix in adopting new curvature and the effects of sequence are not known, although the studies are numerous [ref]. As the DNA is rotated around the curvilinear helical axis, the super helix begins to bend back on itself forming an unfavorable kinked structure. To overcome both the breakage in the backbone and the improper bending direction of the super helix, molecular dynamics was applied to allow the DNA to settle into new energetically favorable orientations after each rotation increment; this is performed in each rotation cycle. Since each rotated structure perturbs interactions with solute, protein, DNA and ions, the rotated structure is re-thermalized, equilibrated, and run for a short number of molecular dynamics steps to allow the structure to adjust. In this step, explicit solvent and particle mesh Ewald electrostatics are used.

First, the rotated structure is energy minimized for 3000 steps, followed by a thermalization routine. The protein and DNA are held fixed for 10000 steps while ions and water are heated from 0 to 100K in 10K increments. The protein and DNA are then

allowed to move keeping the water and ions fixed for 10000 steps, while increasing the temperature of the system to 100K in increments of 10K. These two runs are repeated for 20000 steps bringing the entire system up to 300K. Finally, 40,000 steps of full MD are run using 2fs time steps effectively generating an 80ps trajectory for every 3 degrees of rotation.

Although the above protocol was adaptive, another adjustment was made to overcome the system's tendency to return to its pre-rotated conformation. This problem was addressed by fixing the nucleic acid side chain atoms during the molecular dynamics steps. This was sufficient to keep the base pairs in the rotated position, while the remainder of the DNA backbone adapted to the new rotated positions of the base pairs. A ball and chain analogy can be applied where picking up the ball and repositioning it forces the chain to move along with it to the new position.

The refined rotation algorithm ran iteratively producing 15 structures totaling 45 degrees of rotation before NAMD failed due to unfavorable atom positions reported as shake errors. As expected, after accumulated rotations and fixed nucleic acid bases, unfavorable kinking occurred due to the bases being held fixed during the molecular dynamics. Bases located initially closer to the protein core have narrow groves which need to widen as the helix is rotated. Similarly, the bases located on the outer surface of the helix had wide grooves which were in need of narrowing as they were rotated inwards closer to the protein core. It was clear that the nucleic acid bases needed to be allowed to move unconstrained periodically during the rotation process allowing them to adjust to the sugar phosphate backbone atoms. In order to determine when fully unconstrained normalization MD needed to be run on the structure, the non-bonded interactions were monitored using MDenergy, using a heuristic assessment of the energy to trigger normalizing the entire structure.

MDenergy is an analysis package accompanying NAMD that can be used on a selection of atoms to compute energy components such as, conformational energy, van der Waals (VDW), and cutoff-based coulombic electrostatics. MDenergy was run on the initial collection of 15 structures (3 degrees of rotation per structure) and an exponential increase in energy was seen in the VDW energy (fig. 15). Performing a fully unconstrained molecular dynamics run on the energetically compromised structure effectively stabilized them. The normalization MD run consists of running the same energy minimization, thermalization, and MD routine as used after rotation, but, releasing the previously fixed nucleic acid bases.

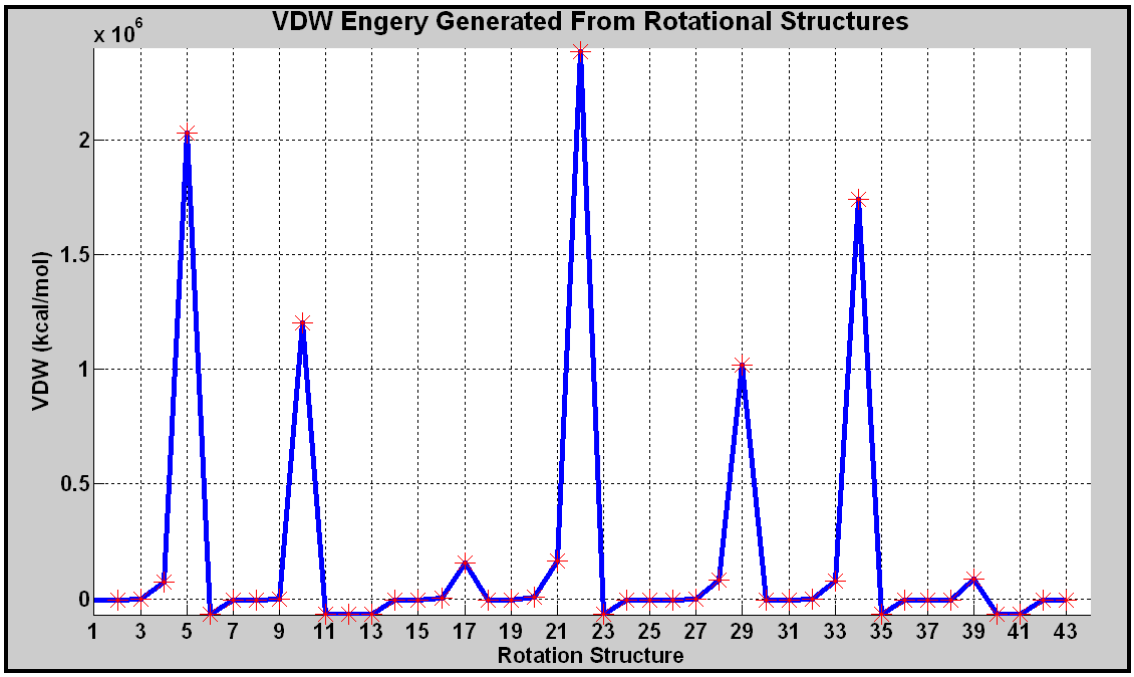


Figure 32. **Computed VDW energy for 43 structures. The sharp drop in energy after a spike is the result of the normalization run.**

The normalization approach was effective at producing an energetically stable structure and allowed the rotation to proceed. However, another problem was discovered after rotating the DNA through 180 degrees. The protein was allowed to move throughout all the MD runs except where noted. As a consequence the protein was rolled out of the center. Figure 16 displays the observed phenomenon showing snap shots of the protein at 0 degrees rotation 90 degrees and 180 degrees. This problem can be conceptualized by imagining rolling a rubber band off of a bundle of pencils. To solve this problem, we fixed the protein motion during all MD runs except the normalization runs, where it is only fixed during the thermalization steps.

Figure 33. **Protein displaced after 180 degrees of rotation. The image at the far left is at 0 degrees, the middle image is at 90 degrees, and the far right image is at 180 degrees.**

To summarize the current protocol, MDenergy is run after each rotation with fixed side chains and protein. The computed energy components for the current structure are compared to those of the previous structure. If a large deviation is found in the van der Waals energy above an empirically determined threshold, then a normalization MD run is performed to allow the nucleic acid bases to adjust to the current rotational state. The present heuristic used is +7500 kcal/mol which triggers normalization every 12 degrees of rotation. Shown in figure 34 is a flow diagram of the rotation algorithm.

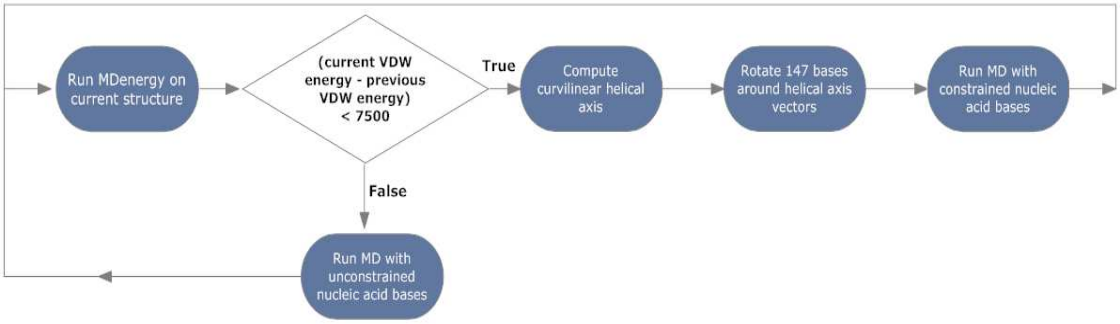


Figure 34. **Rotation algorithm decision process.**

5.3 Applying the Rotational Method

The clone601 sequence has been rotated 111 degrees producing 37 unique structures. Preliminary analysis of these structures has consisted of looking at the hydrogen bonding network within the helix. The counts of the hydrogen bonds in the rotational structures can be incorporated into the rotational algorithm as another metric used to assess the quality of the rotations. The hydrogen bonds were assessed using a cutoff distance of 3.0 Å and co-linearity of 30.0 degrees for determining a single hydrogen bond. The first rotated structure had 95 hydrogen bonds between DNA base pairs and the (current) last 111 degree structure contains 117 hydrogen bonds. Rotation structure number 20 (60 degrees) can be seen in figure 3 to contain considerably less hydrogen bonds compared to the other structures; it is also has a large deviation in the van der Waal energy which triggered a normalization run. For comparison the number of hydrogen bonds at 1ns and 10ns in the 601 non-rotated structures are both 116.

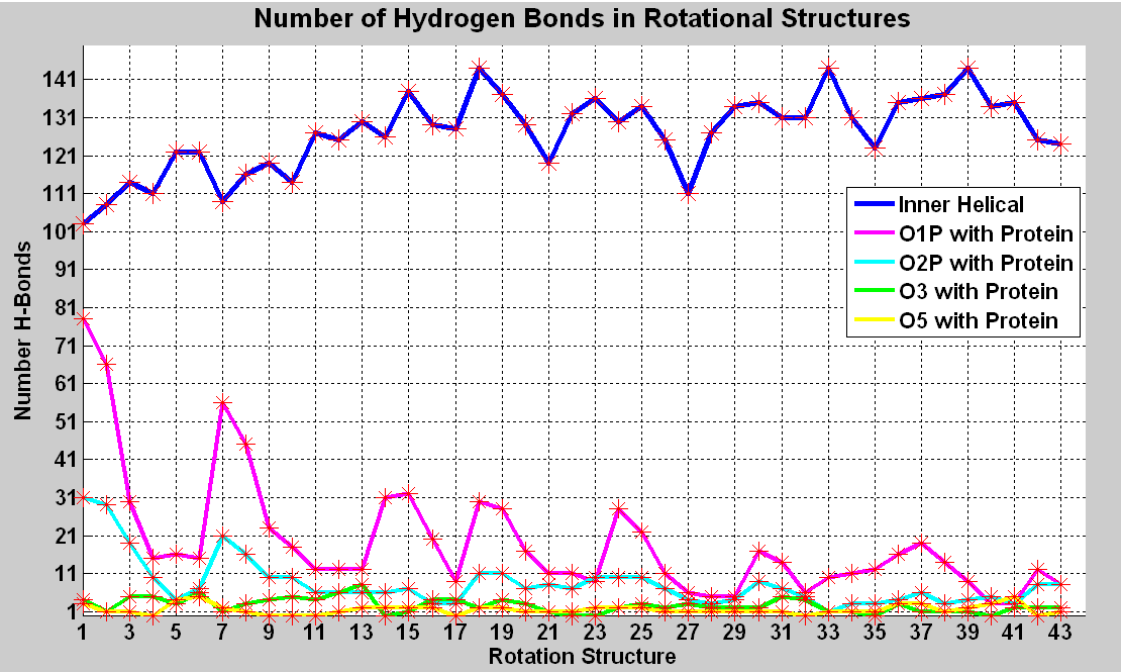


***Figure 35*. Number of hydrogen bonds between nucleic acid base pairs in each rotated structure.**

Contributing to the stability of the overall structure are contacts between the negatively charged phosphate backbone and positively charged amino acids in the histone proteins. Commonly identified contact points in the 1KX5 crystal structure are monitored but it is hypothesized that new contacts will be explored as the sugar-phosphate backbone moves along the surface of the protein, parallel to the curvilinear helical axis. In essence, the rotation is a search algorithm, exploring the energetically charged surface of the protein with the negatively charged phosphates. Hydrogen bonds formed between the O1P, O2P, O3', and O5' oxygen atoms are monitored during the course of the rotation. For reference the number of hydrogen bonds between the O1P, O2P, O3', and O5' oxygen atoms and the protein at 1ns and 10ns in the 601 non-rotated structure are {35, 73, 0, 7} and {40,77,2,3}, respectively.

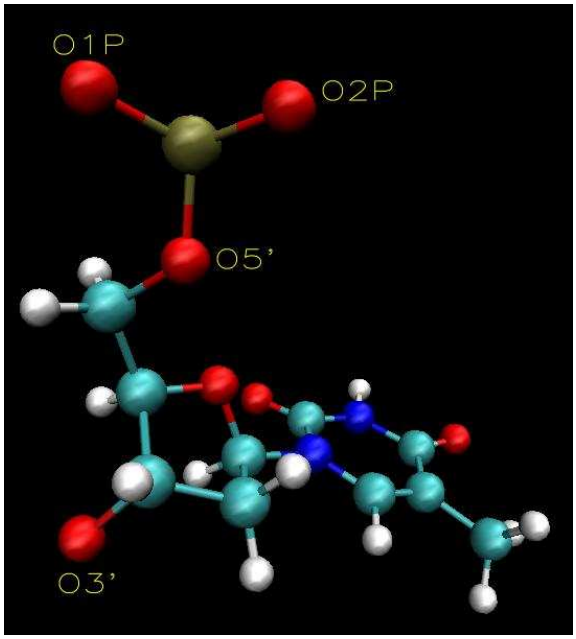


Figure 36. **Map of oxygen atoms that can form hydrogen bonds within the DNA backbone.**

5.4 Geometric Analysis of Method

The necessary base-pair-step curvature to produce the superhelix path was computed to be 600.0º bending the 147 base pairs through 1.84 superhelical turns [8]. The actual curvature found in the 1KX5 crystal structure was 1333.3 º, more than double the theoretical curvature. Using the same software, Curves, the curvature computed was 989.3º and 1155.7º for the superhelix path of the 1st structure and the 37th structure, respectively. The superhelix curvature of the initial solvated 1KX5 system was computed and found it to be 1107.8 º. Comparing the curvature of the current rotated structure to that of the solvated 1KX5 system slightly more curvature is observed for the current rotated structure. The global curvature is another metric which is used to monitor the rotations.

The average lateral translation parallel to the curvilinear helical axis of a residue is 0.06 Angstroms per 3 degrees of rotation. The total lateral translation after 180 degrees of rotation is 3.9 angstroms. The rise of an individual residue in B-form DNA corresponds to 3.32 angstroms. Therefore after 180 degrees of rotation the entire sequence has been translated by a full base pair.

5.5 Selecting Favorable Rotational Conformations

Selected structures from the suite of rotations were run for an additional 160ps with no constraints applied using the same normalization protocol. The selection process was based on sampling coarse grain rotations of ~9 degree increments and high VDW energy. Fourteen structures within the suite of 43 have been normalized an additional 160ps. When comparing the hydrogen bonds in a given structure it is important to keep in mind the total amount of simulated time accumulated during the rotation process. For

example, structure 4 has undergone a total of 240ps of MD after a 80ps normalization run and 4 previous 40ps rotation runs. To perform a more reliable selection a full ns of unconstrained simulation time should be run on a given structure. After only 80ps of unconstrained MD structure 4 gains 56 (15 to 71) O1P hydrogen bonds with protein and 20 more hydrogen bonds with in the helix. A reasonable amount of time can be deduced by looking at when the RMSD flattens out in the non-rotated simulation. This occurs after ~1ns (fig. 12) therefore an appropriate amount of time would be at least 1ns.

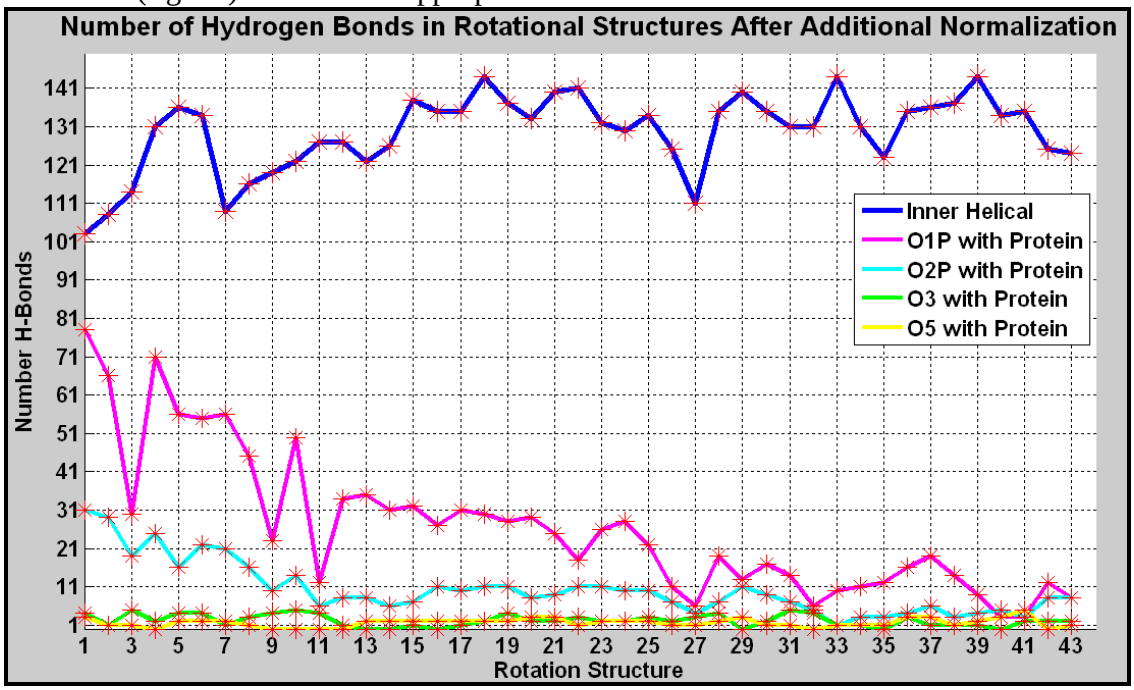


Figure 37. **Number of hydrogen bonds between nucleic acid base pairs in each rotated structure after normalization of selected structures {4,5,6,10,12,13,16,17,20,21,22,23,28,29,33,34}.**

Using the discussed metrics to select several favorable rotational conformations the next step is to perform a long range MD simulation on each of the selected structures. A non-rotated 601 clone simulation is currently underway just passing the 10ns mark which will be used to compare the long range simulations of the selected rotated structures. These simulations will enable us to elucidate the structural properties of the nucleosome that confer favorable positioning of the 601 clone sequence around the core histone proteins.

5.6 Experimental Methods for Verification of Favorable Conformations

Lugers lab has been developing synthetic linkers to fix DNA in a particular rotational conformation [45]. Once a predicted rotational conformation is determined this is one method that could be used to experimentally verify that the DNA is bound in a certain orientation. The PEG linker (fig. 40) is highly specific to the DNA in the minor groove and can be modified to recognize a particular sequence. Figure 38 shows the general idea which would be to employ the linker to capture a theoretical rotational conformation. Figure 39 is a model of the PEG linker attached to the minor groves of the super helix.

Another possibility for experimental verification would be to attach florescent probes to particular base pairs that would come into close contact given a predicted rotational conformation. The intensity of fluorescence would allow a quantitative measurement to be made. Either of these methods is feasible to do and would provide verification of a selected rotational conformation.

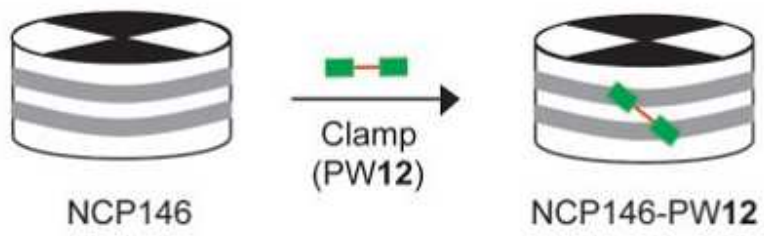


Figure 38. **Development of the clamp.**

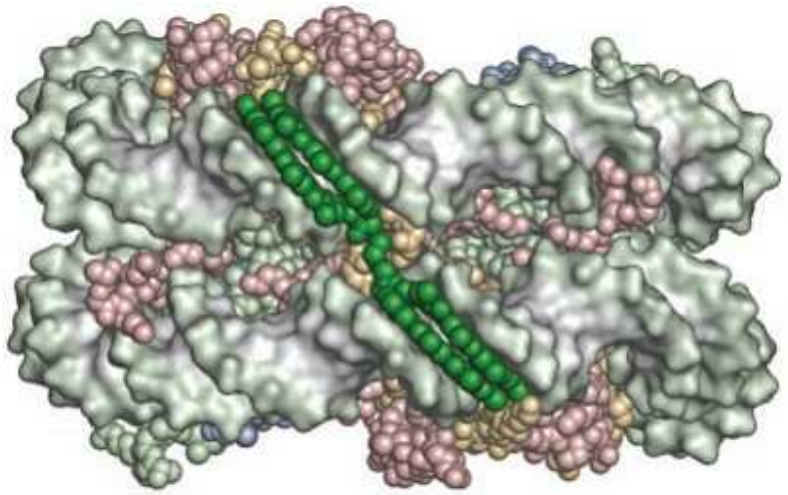


Figure 39. **Illustration of the PEG linker bound in the minor groves of the superhelix.**

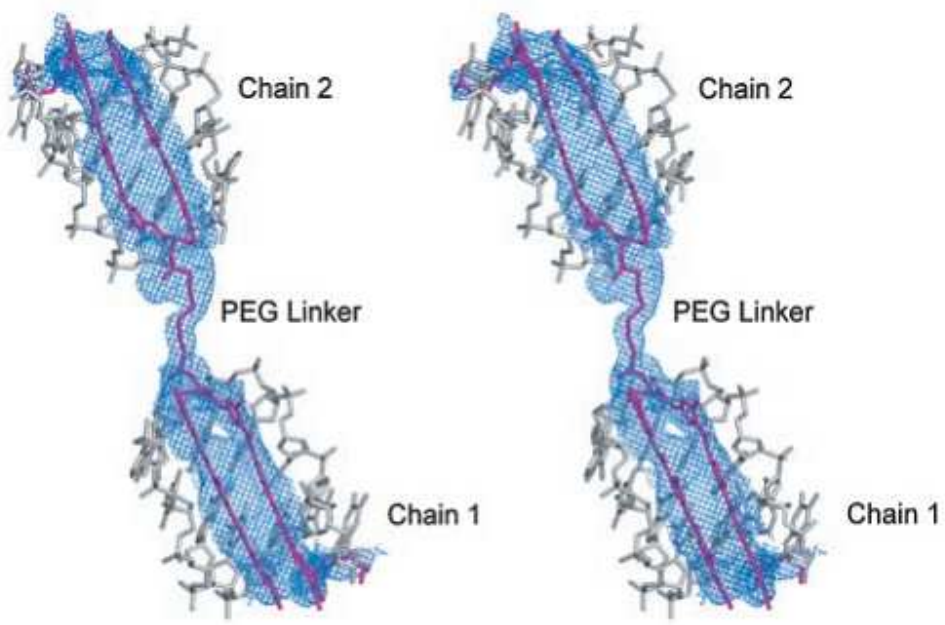


Figure 40. **Synthetic PEG linkers.**

6 Discussion of Results

The development and application of a sound model for examining sequences effects of DNA bound to the nucleosome has been established. The H3 histone tails must be included when modeling the nucleosome. The experimental data collected provided a ranking for the 1KX5 palindrome sequence and the 601 sequence affinity for the nucleosome. The ranking served as a guide which was used when analyzing the trajectory data produced by the two simulations. The 601 sequence was found to be more stable based on its average RMSD, RMSF, and correlated motions. The computational findings match what was expected from the experimental guide.

Experimental evidence also shows that the 601 sequence has a rotational preference when bound to the nucleosome motivating the development of a way to explore such conformations. Through the use of geometric manipulation and MD rotational conformations were generated for the 601 sequence bound to the histone protein core. These conformations were then analyzed based on their hydrogen bonding within the helix to confirm the fidelity of the helix as it was being rotated. An increase in the number of hydrogen bonds has been observed verifying that the helix is remaining intact throughout the rotation process. Searching for a preferred rotational conformation is being carried out by looking at the hydrogen bonding network between the oxygen atoms in the backbone and neighboring protein residues. An increase in the number of bonds is found in normalized structures prompting the selection of the 24^{th} rotational conformation for a long range simulation. The long range simulation will be compared to a non-rotated structure in the future to determine how the dynamics of the simulations are performing.

7 Appendix

7.1 Experimental Protocols

The primers used in the 1KX5 experiments are shown below giving their length, melting temperature, GC content, self dimerizing tendency (PD), and concentration of received samples.

1KX5-1 49bp T_m:78.7ºC GC:38.78% PD:No OD@260=0.304 OD@280=0.156 [DNA]=1216ng/µl

5'-ATCAATATCCACCTGCAGATACTACCAAAAGTGTATTTGGAAACTGCTC-3'

1KX5-2 46bp T_m:86.98ºC GC:47.83% PD:No OD@260=0.319 OD@280=0.173 [DNA]=1276ng/µl

5'-GGATCCGATTCCAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAA-3'

1KX5-3 46bp T_m:86.98ºC GC:47.83% PD:No OD@260=0.363 OD@280=0.194 [DNA]=1452ng/µl

5'-GGATCCGAATCCAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAA-3'

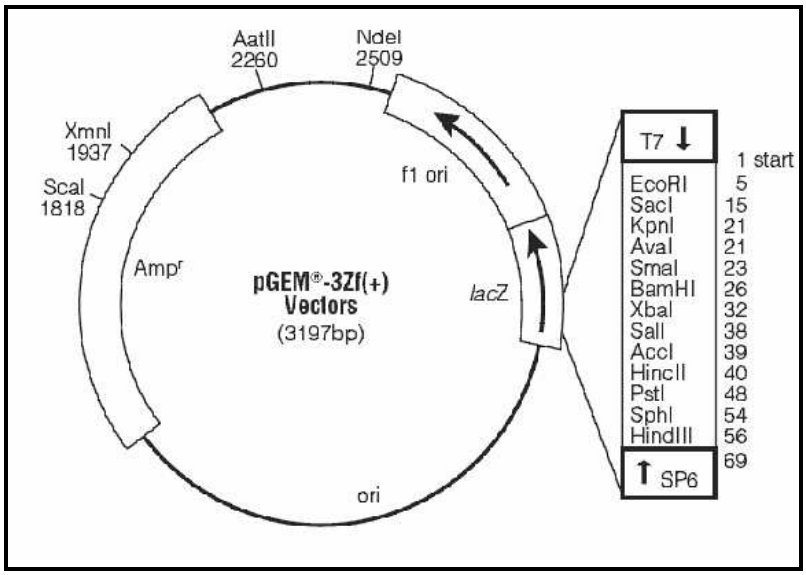


Figure 41. **pGEM-3Zf(+) Vector used for half clones of 1KX5.**

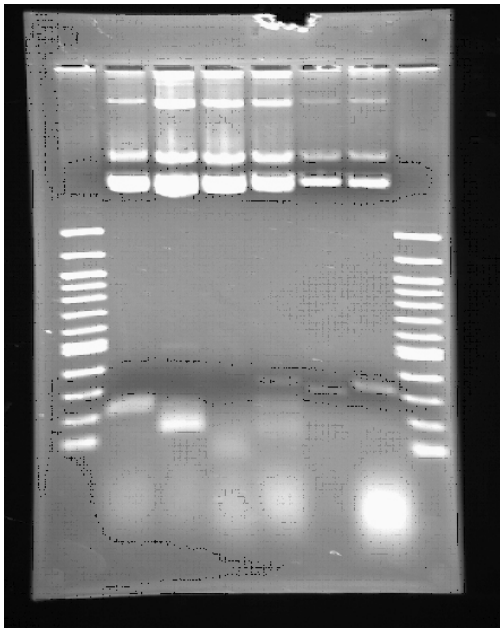


Figure 42. **Checking for insertion of the half clone 1KX5-1/2 in the plasmid vector using restriction enzymes. Lanes 1 and 8 contain 100bp markers.**

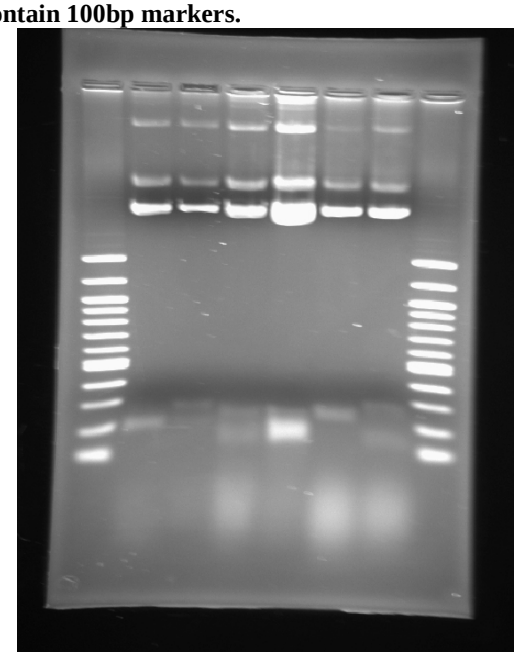


Figure 43. **Checking for insertion of the half clone 1KX5-1/3 in the plasmid vector using restriction enzymes. Lanes 1 and 8 contain 100bp markers**

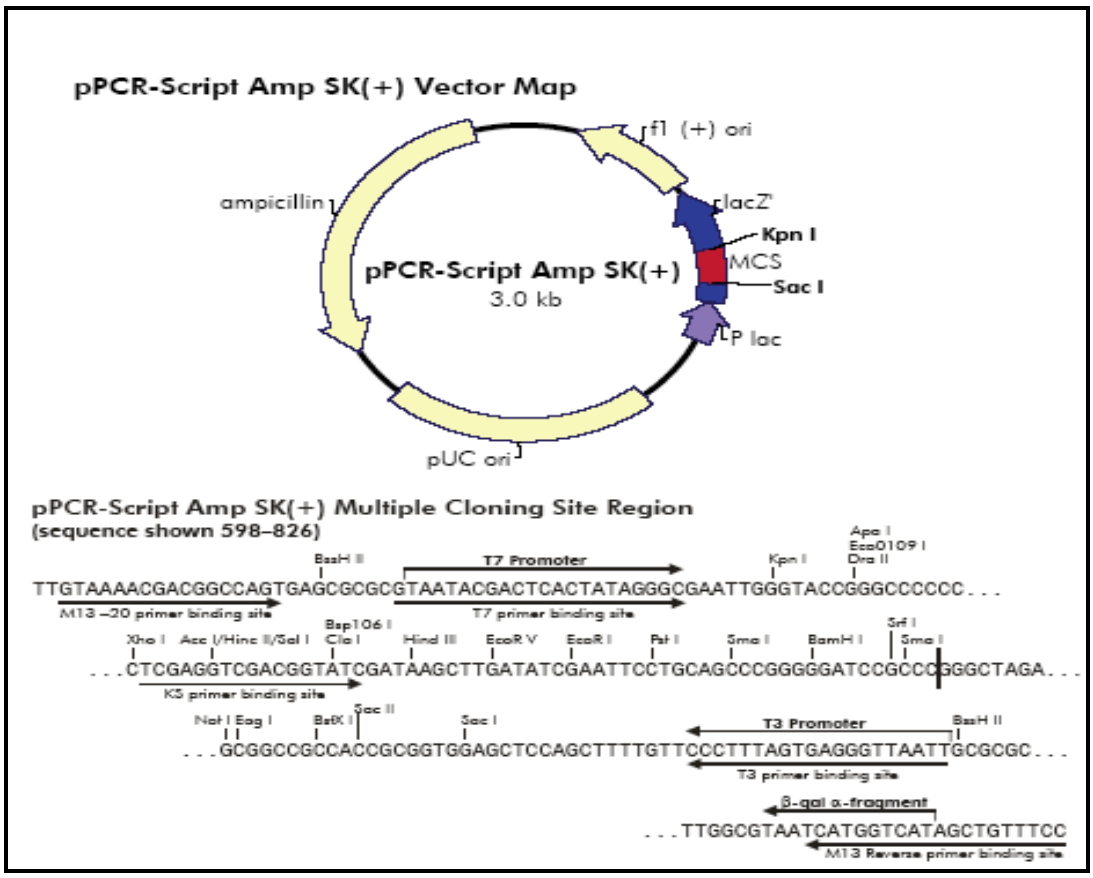


Figure 44. **Vector Map displaying where CGG is inserted into the plasmid.**

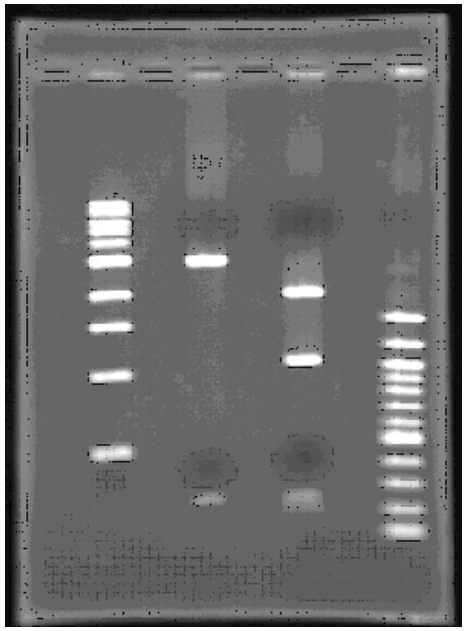


Figure 45. **The CGG restriction experiment was run on a 1% gel at 100V for 1hour. Counting the lanes from left to right the 1000bp ladder is in lane 2 and the 100bp ladder is in lane 8. pFXA53 is in lane 6. The restriction sites within the vector created 4 fragments.**

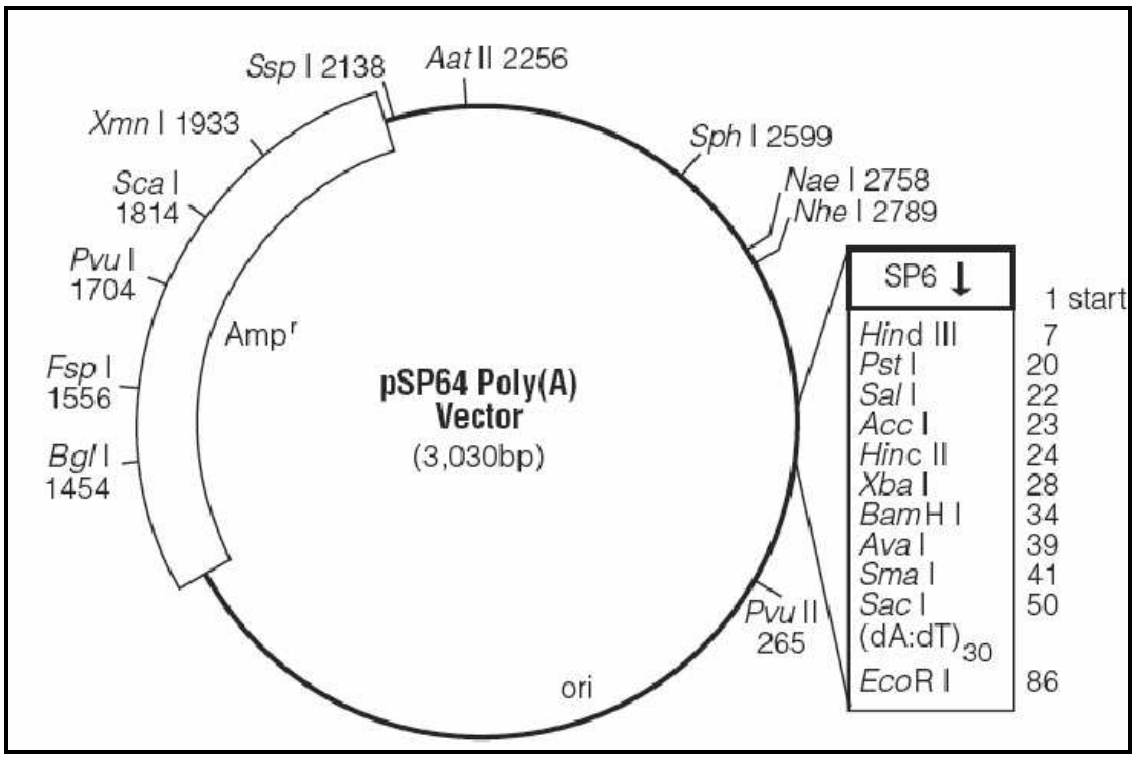


Figure 46. **5S-RNA vector map.**

Two PCR reactions were run using the 5S1/5S3 primers and the 5S2/5S3 primers The gel in figure 23 shows the results of the 5S1/5S2 PCR reaction expected length of 180bp. Lanes counted from left to right lane 2 and 7 loaded with 100bp ladder. The same heating schedule was used varying the concentrations of $MgCl_2$. Lane 5 showed the best results using 1.0mM $MgCl_2$.

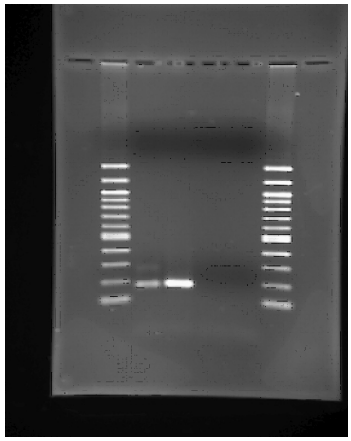


Figure 47. **PCR reaction results for 5S1/5S2.**

Counting left to right lanes 2 and 7 are loaded with 100bp ladder.
5S2/5S3 PCR products did not show up. The expected length was 147bp.
The 20bp primers are what have run past the 100bp marker at the end.
The 5S3 primer worked with 5S1 so looking back at the primers 5S2 can form dimers and is GC rich (70%).

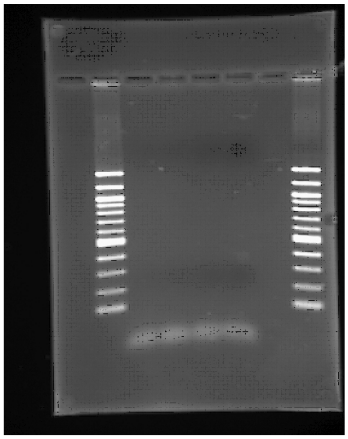


Figure 48. **PCR reaction results for 5S2/5S3.**

Two separate experiments were run, one using a lower temperature and DMSO the other using a higher temperature. The expected length was 147bp. The first gel shows the results of adding the DMSO and running at a lower temperature. The 1.5mM $MgCl_2$ concentration in lane 5 showed the best results. The second gel run with higher temperature also proved successful at a concentration of 1.5mM $MgCl_2$.

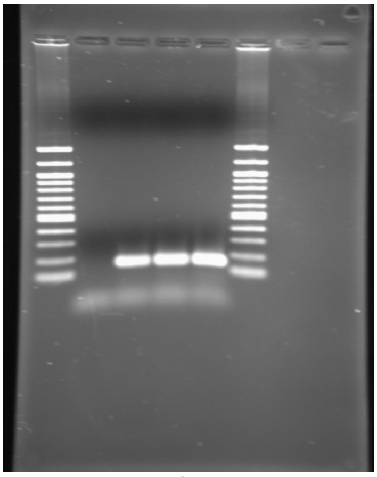


Figure 49. **5SRNA Modifications to PCR experiment for 5S2/5S3 using lower temperatures and DMSO.**

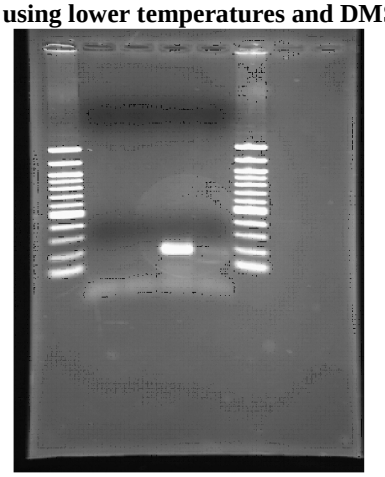


Figure 50**. Modification to PCR experiment 5S2/5S3 run with higher temperatures.**

7.2 Computational Methods

NAMD restraining forces

```
freeEnergy on
freeEnergyConfig {
  urestraint {
    dist (N01J, 146, P) (N01J, 67, P) kf=20 ref=8.5
    dist (N01I, 145, P) (N01I, 68, P) kf=20 ref=8.5
    dist (N01J, 147, P) (N01I, 4, P) kf=20 ref=15.0
    dist (N01I, 147, P) (N01J, 4, P) kf=20 ref=15.0
  }
  pmf {
    task=stop
    time=49999fs
  }
}
```

Rotation *tcl* function

```
proc rotateAllHelix {pdbInput degrees hTraceIn pdbOutput fileID} {
    #Input PDB to roatate
    mol load pdb $pdbInput
    set nucleic [molinfo index 0]
    #Contains the curvilinear helica axis
    mol load pdb $hTraceIn
    set hTrace [molinfo index 1]
    for { set i 1 } { $i < 148 } { incr i } {
            set cen1index [expr $i - 1]
            set cen1indexN [expr $cen1index*5]
            set cen1 [lindex [[atomselect $hTrace "index $cen1indexN"] get {x y z}]
0]
            set j $i
            if { $i==147 } {
                    set cen2index 731
            } else {
                    set cen2index [expr $j*5]
            }
            #Get second point (Head of vector)
            set cen2 [lindex [[atomselect $hTrace "index $cen2index"] get {x y z}] 0]

            set b1 [atomselect $nucleic "resid $i and chain I"]
            #Get position of center of mass before rotation
            set coords1 [measure center $b1 weight mass]
            set comp [expr 148 - $i]
            set b2 [atomselect $nucleic "resid $comp and chain J"]
            set coords2 [measure center $b2 weight mass]
            #Get positino of center of mass after rotation
            $b1 move [trans bond $cen1 $cen2 $degrees deg]
            $b2 move [trans bond $cen1 $cen2 $degrees deg]
            set b3 [atomselect $nucleic "resid $i and chain I"]
            set b4 [atomselect $nucleic "resid $comp and chain J"]
            set coords3 [measure center $b3 weight mass]
            set coords4 [measure center $b4 weight mass]
            set nucRotated [atomselect $nucleic "all"]
            if {$i == 147} {
                    $nucRotated writepdb $pdbOutput
            }
    }
    mol delete $nucleic
    mol delete $hTrace
    return
}
```

Generation of potential maps using APBS and PDB2PQR

python pdb2pqr.py --assign-only --ff=CHARMM nucleicKX.pdb nucleicKX.pqr

7.3 Xenopus Laevis and Human Histone Protein Homology Analysis

Histone proteins H3, H4, H2A, and H2B were analyzed by comparing the crystal structures 1KX5 (*Xenopus laevis* 1.94Å) and 2CV5 (Homo Sapien 2.5Å); the histones are labeled A, B, C and D, respectively [36]. Analysis was performed by loading the structures in VMD and using the MultiSeq plugin tool. The plugin combines clustalW and basic plotting functionality to present the results. The most obvious difference found are the higher resolution structure (*Xenopus laevis*) contains longer histone tails which are evident in the alignment plots as large regions of 0% conservation. The proteins are highly conserved and the analysis has proved that the non-conserved residues are not in contact with the DNA backbone. The one exception is the non-conserved region 30-37 found in the H4 histone protein. Why the tails in the Homo Sapien protein structures are shorter than those in the *Xenopus laevis* structures is due to the difference in resolution of the structures. Examining the histone proteins in the SwissProt database shows they are indeed the same length as the *Xenopus laevis* proteins.

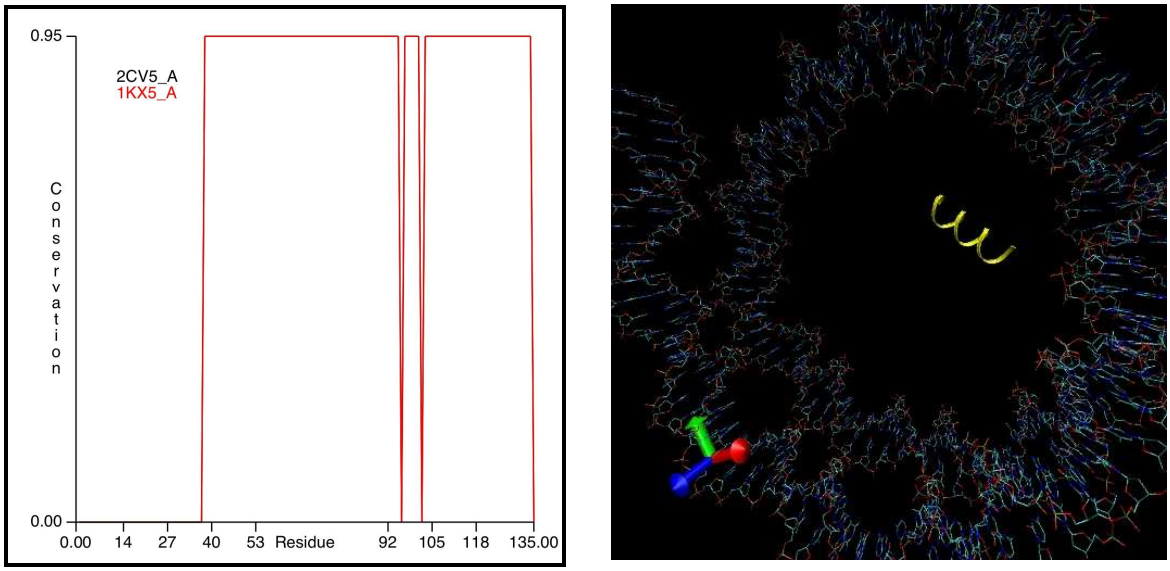


Figure 51. **H3 Human and *Xenopus* sequence alignment. The region of the two non-conserved residues found within the H3 protein is highlighted in yellow.**

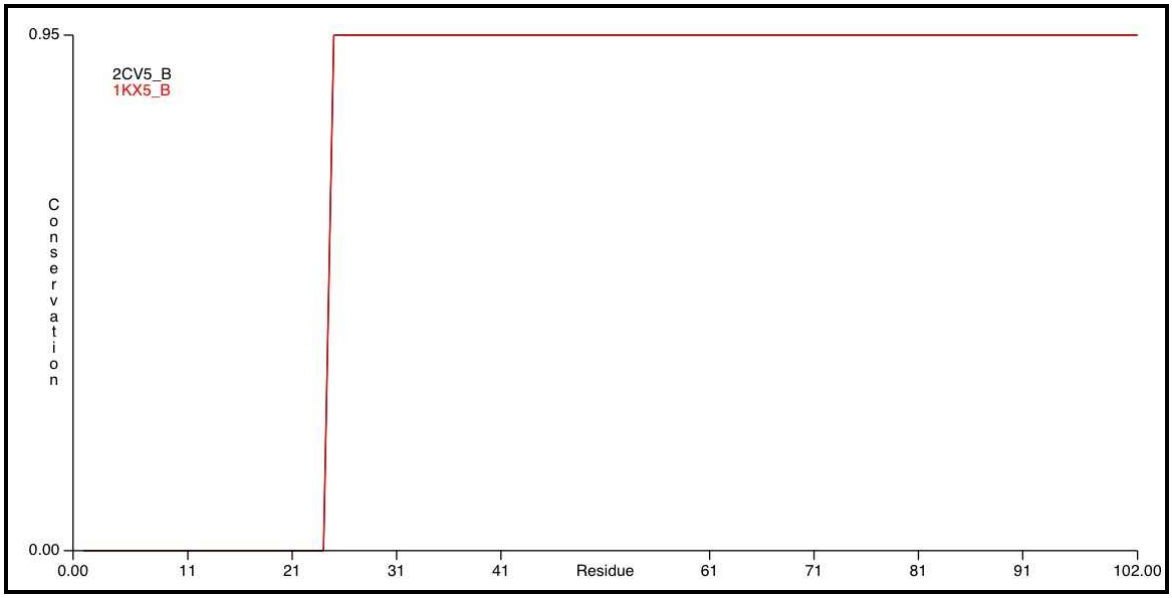
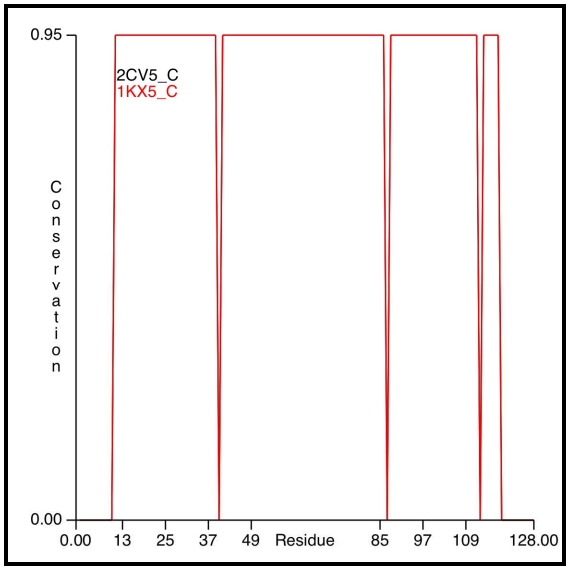


Figure 52. **H4 Human and Xenopus sequence alignment. Note the 0% conserved region representing the longer tail in the higher resolution structure**



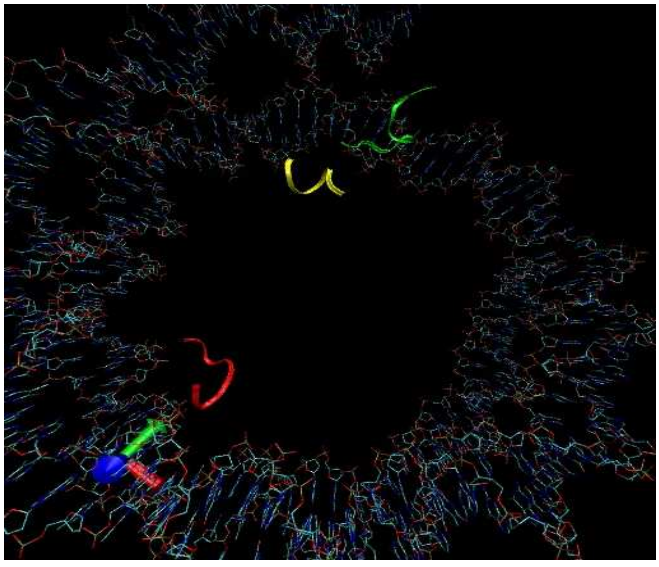


Figure 53. **H2A Human and Xenopus sequence alignment showed 3 residues that were non conserved. Region 37-49 is highlighted in red, region 85-97 is highlighted in green, and region 109-120 is highlighted in yellow. Spatial analysis reveals that these regions are distant from the DNA backbone and would not directly interact with the DNA.**

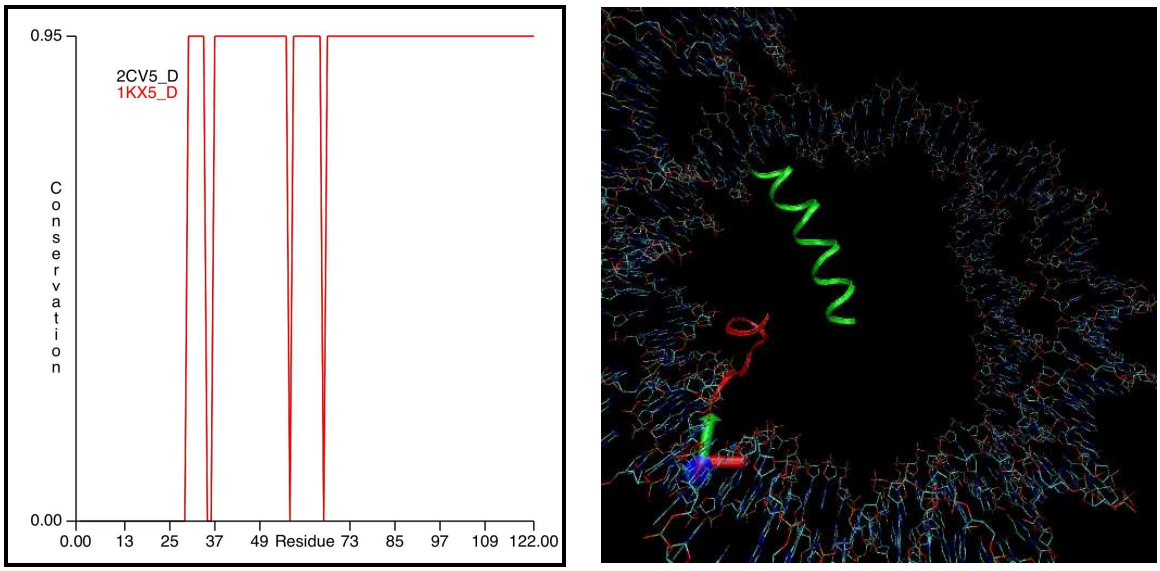


Figure 54. **H2B Human and Xenopus sequence alignment showed 4 residues that were non-conserved. Two of the four residues occur consecutively and are found in the region 30-37 highlighted in red. The other residues are found in the region 49-73 and are highlighted in green. The red region is close to the DNA backbone.**

7.4 Sequences Studied

The first 72 bases (1-72) are shown in red and the last 72 bases (74-147) are shown in blue with a single base at position 73 representing the center of the two flanks. The palindromic sequence, 1KX5, can be observed by reading a single base at a time on each flank moving away from the central black base.

> I1 **** oriented 5' to 3' ******

5'-AAATGGGCTGAAATGGGCTGTAAATGGGCTGAAATGGGCTGTAAATGGGCTGAAATGGGCTGTAAATGGGCTGA
AATGGGCTGTAAATGGGCTGAAATGGGCTGTAAATGGGCTGAAATGGGCTGAAATGGGCTGTAAATGGGCTGT-3'

> A7 **** oriented 5' to 3' ******

5'-AAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAAC
CCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCT-3'

> clone 601 **** oriented 5' to 3' ******

5'-ATGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGC
TGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT-3'

> 1KX5 **** oriented 5' to 3' *****

5'-ATCAATATCCACCTGCAGATACTACCAAAAGTGTATTTGGAAACTGCTCCATCAAAAGGCATGTTCAGCTGGAA
TCCAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAAATACACTTTTGGTAGTATCTGCAGGTGGATATTGAT-3'

> CGG **** oriented 5' to 3' *****

5'-CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC-3'

8 Acknowledgements

Yuh-Hwa

Lindsay Holder

Adam Hock

9 References

[1] I. Albert, TN Mavrich, LP Tomsho, J Qi, SJ Zanton, SC Schuster, and BF Pugh," Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome," *Nature*, vol. 446, pp. 572-576, 2007.
[2] E. Segal, *et al.*, "A Genomic Code for Nucleosome Positioning," *Nature*, vol. 442, pp. 772–778, 2006.
[3] M. Ganapathi, *et al.*, "Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes," *BMC Bioinformatics*, vol. 6, no. 126, 2005.
[4] JD Anderson and J Widom, "Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites," *Molecular and Cellular Biology*, vol. 21 no. 11 pp. 3830-3839, 2001.
[5] V. Iyer and K. Struhl, "Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure," *Embo J*., vol. 14, pp. 2570–2579, 1995.
[6] Y-H Wang, R. Gellibolian, M. Shimizu, RD Wells and JD Griffith, "Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes," *J. Mol. Biol.*, vol. 263, pp. 511-516, 1996.
[7] S. Minucci and P. G. Pelicci, "Histone Deacetylase Inhibitors and the Promise of Epigenetic (and more) Treatments for Cancer," *Nat. Rev. Cancer*., vol. 6, no. 1**,** pp. 38-51, 2006.
[8] T.J. Richmond and C.A. Davey, "The structure of DNA in the nucleosome core," *Nature*, vol. 423, pp. 145–150, 2003.
[9] K. Luger, "Structure and Dynamic behavior of nucleosomes," *Curr Opin Genet Dev.*, vol. 13, no. 2, pp. 127-135, Apr. 2003.
[10] PT Lowary and J Widom, "Nucleosome packaging and nucleosome positioning of genomic DNA," *Proc. Natl. Acad. Sci.*, vol. 94 pp. 1183-1188, 1997.
[11] SC Satchwell, HR Drew, and AA Travers, "Sequence periodicities in chicken nucleosome core DNA," *J. Mol. Biol*., vol. 191, pp. 659-675, 1986.
[12] TE Shrader and DM Crothers, "Artificial nucleosome positioning sequence," *PNAS*, vol. 86, pp. 7418-7422, 1989.
[13] JM Gale and MJ Smerdon, "Photofootprint of nucleosome core DNA in intact chromatin having different structural states," *J. Mol. Biol.*, vol. 204, pp. 949-958, 1988.
[14] A Thastrom, PT Lowary, and J Widom, "Measurement of histone-DNA interaction free energy in nucleosomes," *Methods*, vol. 33, pp. 33-44, 2004a.
[15] C. Zerylnick, A Torroni, SL Sherman, and ST Warren, "Normal variation at the myotonic dystrophy locus in global human populations," *Am. J. Hum. Genet.*, vol. 56 no. 1, pp. 123–130, 1995.
[16] WJAA van den Broek, DG Wansink, and B Wieringa, "Somatic CTG•CAG repeat instability in a mouse model for myotonic dystrophy type 1 is associated with changes in cell nuclearity and DNA ploidy," *BMC Molecular Biology*, vol. 8 no. 61, 2007.

[17] Y-H Wang, "Chromatin structure of repeating CTG/CAG and CGG/CCG sequences in human disease," *Front Biosci.*, vol. 12, pp.4731-4741, 2007.
[18] Y-H Wang, S Amirhaeri, S Kang, RD Wells and JD Griffith, "Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene," *Science*, vol. 265, pp. 669-671, 1994.
[19] D. Roccatano, A. Barthel, and M. Zacharias, "Structural Flexibility of the Nucleosome Core Particle at Atomic Resolution Studied by Molecular Dynamics Simulation," *Biopolymers*, in press.
[20] T.C. Bishop, "Molecular Dynamics Simulations of a Nucleosome and Free DNA," *J Biomol Struct Dyn.*, vol. 22, no. 6, pp. 673-686, June 2005.
[21] CA Davey and TJ Richmond, "DNA-dependent divalent cation binding in the nucleosome core particle," *PNAS*, vol. 99, no. 17 pp. 11169-11174, 2002.
[22] R. Lavery and H. Sklenar, "The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids," *J. Biomol. Struct. Dynam.*, vol. 6, pp. 63-91, 1988.
[23] R. D. Kornberg and Y. Lorch, "Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome," *Cell.*, vol. 98, pp. 285-294, August 6, 1999.
[24] T. Castrignano, G. Chillemi, A. Desideri, "Structuer and hydration of BamHI DNA recognition site: a molecular dynamics investigation," *Biophys.*, vol. 79, pp. 1263-1272, 2000.
[25] T.E. Cheatham, and P. A. Kollman, "Molecular dynamics simulation of nucleic acids," *Annu. Rev. Phys. Chem.*, vol. 51, pp. 435-471, 2000.
[26] J. Norberg, and L. Nilsson, "Molecular dynamics applied to nucleic acids," *Acc. Chem. Res.*, vol. 35, pp. 465-472, 2002.
[27] T. Castrignano, G. Chillemi, G. Varani, and A. Desideri, "Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reels dynamic features of the intermolecular interface and its hydration," *Biophys.*, vol. 83, pp. 3542-3552, 2002.
[28] G. Chillemi, T. Castrignano, and A. Desideri, "Structure and hydration of the DNA-human topoisomerase I covalent complex," *Biophys.*, vol. 81, pp. 490-500, 2001.
[29] C. M. Reyes and P.A. Kollman, "Molecular dynamics studies of U1A-RNA complexes," *RNA.*, vol. 5, 235-244, 1999.
[30] E. Guidice and R. Lavery, "Simulations of nucleic acids and their complexes," *Acc. Chem. Res.*, vol. 35, pp. 350-357, 2002.
[31] J. C. Phillips, *et al.*, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*, vol 26, pp. 1781-1802, 2005.
[32] D. S. Goodsell and R. E. Dickerson, "Bending and curvature calculations in B-DNA," *Nucleic Acids Research*, vol. 22, no. 24, pp. 5497-5503, 1994.
[33] J. A. McCammon, S.C. Harvey, "Dynamics of Proteins and Nucleic Acids," *Cambridge University Press, London, UK.*
[34] G. Chillemi, P. Fiorani, P. Benedetti, and A. Desideri, "Protein concerted motions in the DNA-human topoisomerase I complex," *Nucleic Acids Research*, vol. 31, no. 5, pp. 1525-1535, 2003.
[35] Y. Bao, C. L. White, and K. Luger, "Nucleosome Core Particles Containing a Poly(dA·dT) Sequence Element Exhibit a Locally Distorted DNA Structure," *J. Mol. Biol.*, vol. 361, pp. 617-624, 2006.

[36] Y. Tsunaka, N. Kajimura, S. Tate, and K. Morikawa, "Alteration of the nucleosome DNA path in the crystal structure of a human nucleosome core particle," *Nucleic Acids Research,* vol. 33, no. 10, pp. 3424-3434, 2005.
[37] I. Albert, *et al.,* "Translational and rotational settings of H2A.Z nucleosome across the Saccharomyces cerevisiae genome, " *Nature,* vol. 446, pp. 572-576, 2007.
[38] A. Barbic and D. M. Crothers, "Comparison of Analyses of DNA Curvature," *Journal of Biomolecular Structure & Dynamics,* vol. 21, no. 1, pp. 89-97, 2003.
[39] D. Strahs and T. Schlick, "Analysis of A-tract bending: Insights into experimental structues by molecular dynamics simulations," *J. Mol. Biol.*, vol. 301, pp. 643-663, 2000.
[40] F. Lankas, J. Sponer, J. Langowski, and T. E. Cheatham III, "DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations," *Biophysical Journal,* vol. 85, pp. 2872-2883, 2003.
[41] X. Lu, and W. K. Olson, "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures," *Nucleic Acids Research,* vol. 31, no. 17, pp. 5108-5121, 2003.
[42] L. Sutto, J. Latzer, J. A. Hegler, D. U. Ferreiro, and P. G. Wolynes, "Consequences of localized frustration for the folding mechanism of the IM7 protein," *PNAS,* vol. 104, no. 50, pp. 19825-19830, 2007.
[43] J. Widom, "Role of DNA sequence in nucleosome stability and dynamics," *Quarterly Reviews of Biophysics,* vol. 34, no. 3, pp. 269-324, 2001.
[44] A. Ramaswamy, I. Bahar, and I. Ioshikhes, "Structural Dynamics of Nucleosome Core Particle: Comparison with Nucleosomes Containing Histone Variants," *PROTEINS: Structure, Function, and Bioinformatics,* vol. 58, pp. 683-696, 2005.
[45] R. S. Edayathumangalam, P. Weyermann, J. M. Gottesfeld, P. B. Dervan and K. Luger, "Molecular Recognition of the nucleosome "supergrove"," *PNAS,* vol. 101, no. 18, pp. 6864-6869, 2004.
[46] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, "Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 A Resolution," *J. Mol. Biol.*, vol. 319, pp. 1097-1113, 2002.
[47] J. Z. Ruscio and A. Onufriev, "A Computational Study of Nucleosome DNA Flexibility," *Biophysical Journal,* vol. 91, pp. 4121-4132.
[48] S. Y. Ponomarev, K. M. Thayer, and D. L. Beveridge, "Ion motions in molecular dynamics simulations on DNA," *PNAS,* vol. 101, no. 41, pp. 14771-14775.
[49] S. T. Wlodek, T. W. Clark, L. R. Scott, and J. A. McCammon, "Molecular dynamics of acetylcholinesterase dimer complexed with tacrine," *J. Am. Chem. Soc.*, vol. 119, pp. 9513-9522.
[50] D. J. Mulvihill and Y-H Wang, "Two Breakpoint Clusters at Fragile Site FRA3B Form Phased Nucleosomes," *Genome Research,* no. 14, pp. 1350-1357, 2004.
[51] T. J. Richmond, M. A. Searles, and R. T. Simpson, "Crystals of a Nucleosome Core Particle Containing Defined Sequence DNA," *J. Mol. Biol.*, no. 199, pp. 161-170, 1988.
[52] D. J. Mulvihill, K. N. Edamura, K. A. Hagerman, C. E. Pearson, and Y-H Wang, "Effect of CAT and AGG Interruptions and CpG Methylation on Nucleosome Assembly upon Trinucleotide Repeats on Spinocerebellar Ataxia, Type 1 and Fragile X Syndrome," *The Journal of Biological Chemistry,* vol. 280, no. 6, pp. 4498-4503, 2005.
[53] J. S. Godde and A. P. Wolffe, "Nucleosome Assembly on CTG Triplet Repeats," *The Journal of Biological Chemistry,* vol. 271, no. 25, pp. 15222-15229, 1996.