Atomic Force Microscopy for Chromatin Structure Study

by

Qiang Fu

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2010 by the
Graduate Supervisory Committee:

Stuart Lindsay, Chair
Hao Yan
Giovanna Ghirlanda

ARIZONA STATE UNIVERSITY

December 2011

ABSTRACT

In eukaryotes, DNA is packed in a highly condensed and hierarchically organized structure called chromatin, in which DNA tightly wraps around the histone octamer consisting of one histone 3-histone 4 (H3-H4) tetramer and two histone 2A- histone 2B (H2A-H2B) dimers with 147 base pairs in an almost two left handed turns. Almost all DNA dependent cellular processes, such as DNA duplication, transcription, DNA repair and recombination, take place in the chromatin form. Based on the critical importance of appropriate chromatin condensation, this thesis focused on the folding behavior of the nucleosome array reconstituted using different templates with various controllable factors such as histone tail modification, linker DNA length, and DNA binding proteins. Firstly, the folding behaviors of wild type (WT) and nucleosome arrays reconstituted with acetylation on the histone H4 at lysine 16 (H4K16 (Ac)) were studied. In contrast to the sedimentation result, atomic force microscopy (AFM) measurements revealed no apparent difference in the compact nucleosome arrays between WT and H4K16 (Ac) and WT. Instead, an optimal loading of nucleosome along the template was found necessary for the Mg^{2+} induced nucleosome array compaction. This finding leads to the further study on the role of linker DNA in the nucleosome compaction. A method of constructing DNA templates with varied linker DNA lengths was developed, and uniformly and randomly spaced nucleosome arrays with average linker DNA lengths of 30 bp and 60 bp were constructed. After comprehensive analyses of the nucleosome arrays' structure in mica

surface, the lengths of the linker DNA were found playing an important role in controlling the structural geometries of nucleosome arrays in both their extended and compact forms. In addition, higher concentration of the DNA binding domain of the telomere repeat factor 2 (TRF2) was found to stimulate the compaction of the telomeric nucleosome array. Finally, AFM was successfully applied to investigate the nucleosome positioning behaviors on the Mouse Mammary Tumor Virus (MMTV) promoter region, and two highly positioned region corresponded to nucleosome A and B were identified by this method.

DEDICATION

To my beloved parents.

ACKNOWLEDGMENTS

I would like first thank my PhD advisor, Dr. Stuart Lindsay, who paved the way and gave me a lot of freedom for my research, as well as provided extensive opportunities to collaborate with and learn from world class scientists and researchers. I would also like to thank his patient guidance, encouragement and advice during my research time in his lab.

I want also thank Dr. Julian Chen, who generously let me use facilities and work in his lab for biochemistry experiments, Dr. Jin He for introducing me into Dr. Lindsay's lab, and also all kinds of helps during these years, Dr. Michael Poirior for showing me nucleosome reconstitution in his lab, Dr. Fletcher Terrace's discussion in the telomeric nucleosome studies, Dr. Hongda Wang's training in Atomic Force Microscopy, Dr. Dan Grilley for sharing with me his matlab program, Dr.Robert Ros for sharing the force distance analysis program and Alexander Fuhrmann for showing me his program. Thanks for our collaborators: Dr. Jon Widom, Dr. Philippe Georgel, Dr. Ausio Juan, Dr. Ghiovanna Ghirlanda, Dr. Hao Yan, and Dr. Frank Tsen, who provided me samples for measurements and also taught me the knowledge in their research fields.

I would also like to thank Xiaodong Qi and Chris Bley, who taught me a lot of biochemistry, the discussion with them helped me a lot in my experiments. I

am very grateful to all the fellow graduate students in Lindsay Laboratory: Liyun Lin, Paminder Kaur, Sen Peng for working together on the Chromatin project, Ashley Kibel , Shuai Chang, and Shuo Huang for sharing ideas in programming. Undergraduate students Trent Bowen and Aleasha Main helped me a lot in performing the experiments. I would also like to thank William Lee and Ryan Muller for their help during the summer high school inter program. Last but not least, my deepest thanks to all the people helped me through the lives, studies, and researches at Arizona State University.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Chapter 1

1 Introduction

1.1 Background in Chromatin and Nucleosome

1.1.1 Chromatin and Nucleosome

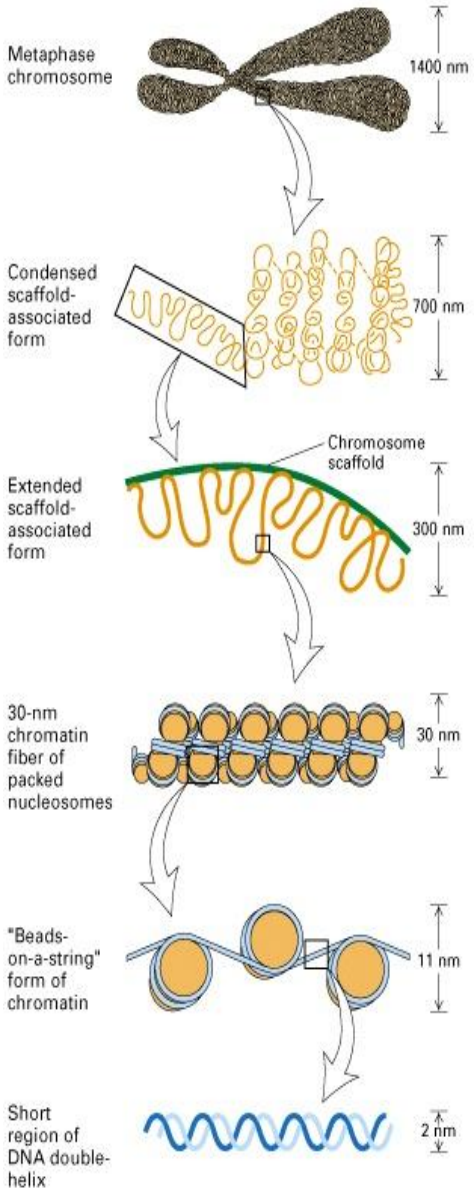

Figure 1-1 Different level of compaction of DNA in cell nuclei

Imaging taken from Reference [1]

Following a hierarchical packing pathway, the DNA, started from an extended 2 nm flexible polymer, first packs into an 11 nm 'beads-on-a-string' structure,

in which DNA wraps around the histone octamer in a left-handed fashion with approximately two turns. The basic repeating unit of this structure is called a nucleosome, which consists of around ~200 bp DNA wrapping around eight protein subunits: 2copies of each H3, H4, H2A and H2B[7]. The complete structure including the linker DNA and the linker histones, together with nucleosome was first characterized by Simpson as chromatosome[12]. In physiological conditions, through short range inter-nucleosomal interactions, the 'beads-on-string' structure automatically folds into a fiber structure with a diameter around 30 nm, which has been observed from *in vitro*[13] and *in vivo* experiments[14]. The '30 nm' structure could further associate with chromosome scaffold to form a '300 nm' structure. The '300 nm' structure can further fold into a metaphase chromosome with a size around 1400 nm as shown in Figure 1-1.

1.1.2 DNA structure/Nucleosome Positioning:

1.1.2.1 Basic Structure of DNA

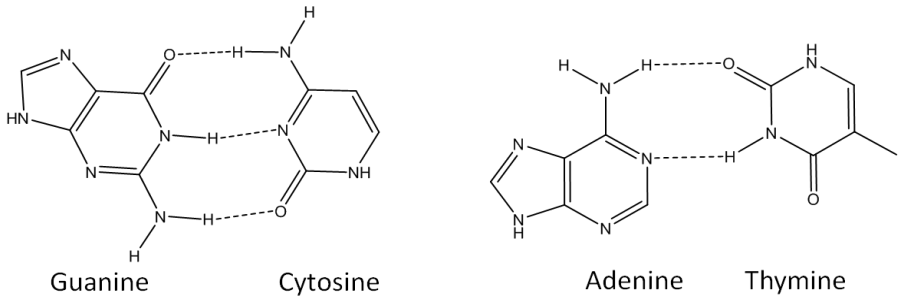

Figure 1-2 Four Deoxyribonucleotide bases

The molecule structures of the four DNA bases are shown here. Guanine is paired with Cytosine by three hydrogen bonds. Adenine is paired with Thymine by two hydrogen bonds.

The fundamental building units for DNA are the four deoxyribonucleotides: Guanine, Adenine, Cytosine, and Thymine (Figure 1-2). The DNA bases of Adenine and Cytosine can be modified by a methyl group both in prokaryotes and eukaryotes. Three types of DNA morphology structures are found in nature, namely, type A, B, and Z (Figure 1-3). In conditions with physiological ionic strength, a B-type structure, which has 10.5 bp for each turn, is preferred. In dehydration and high-salt conditions, the DNA fiber is adapted to two other types: A-DNA with 11 bp/turn, and Z-DNA with 12 bp/turn respectively[8]. Among these three forms, type A and B DNA are right-handed helices, and Z-DNA is a left-handed helix. In general, the structure of

DNA is an unbranched double helix. However, there are also double stranded DNA with junctions, such as the holiday junctions [15].

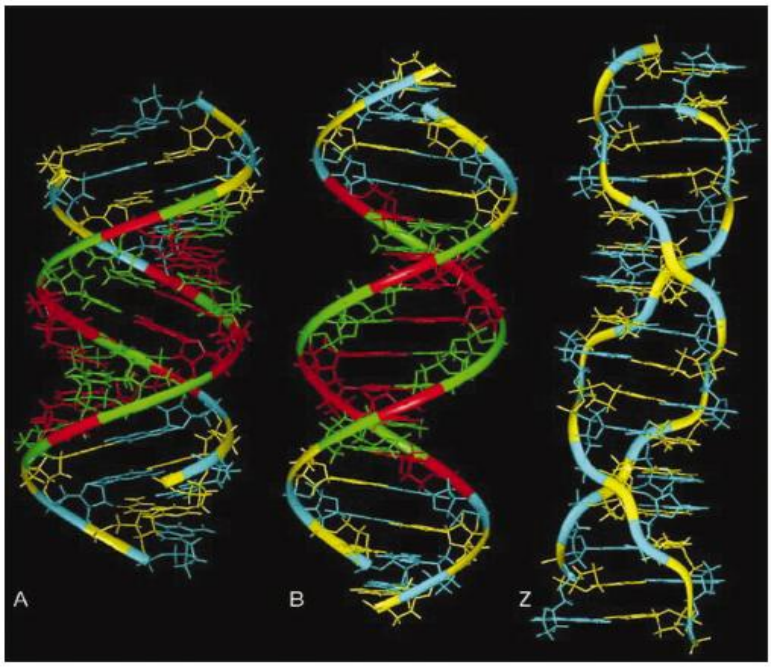

Figure 1-3 Three different types of DNA

The crystal structures of three different types of DNA, A, B, and Z are shown here. Image was taken from reference[8].

1.1.2.2 DNA Structure in the Nucleosome

Eukaryotic DNA possesses the type B DNA structure with a persistence length around 50 nm[16]. The rigid DNA is highly coiled and bended after wrapping around histone octamer[7]. The interaction between the DNA and the histone octamer mainly arises from the electrostatic force, which can be screened by high salt concentrations[17]. Moreover, the nucleosomal DNA,

which is strikingly different from the DNA present in other circumstances, has an average length of 10.17 bp per helical turn[18], while naked DNA only possesses 10.5 bp per helical turn[18] [19]. However, an early study of the hydroxyl radical foot printing demonstrated that DNA wraps around the nucleosome in a non-uniform bending distribution. The helix turns near the dyed positions have10.7 bp per turn, and the DNA near the entering and exiting sites is 10 bp per turn[20]. With the crystal structure of the nucleosome core at 1.9-Å-reolution, the details of how the DNA wraps around the nucleosome have been clearly revealed[18].

Since the initiation of gene expression needs unwrapping of the DNA away from the histone octamer, the positioning ability of the nucleosome at the upstream of the gene becomes extremely important[21-23]. Studies of the nucleosome positioning along the whole genome revealed that the positioning ability of the gene to the histone octamer is directly related to the transcription level[24]. Therefore, to study the interaction between DNA and the histone octamer, it is important to understand the gene activation process. Three stages are involved in the force-induced disassembling of each single 5S-nucleosome in a nucleosome array: first, 76 base pairs of DNA are unwrapped in low force, followed by the unwrapping of 80 base pairs at high force in the other two stages revealed by optical trap experiment[25]. While another experiment, which avoids inter-nucleosome interactions and only unwraps one 601-mononucleosomes, shows a two-stage unwrapping

process and that the edge of the DNA can be peeled off from the histone core with an energy cost of 0.6 kT/bp [26].

Not only the intrinsic DNA sequence pattern[27], other factors such as histone tail modifications, DNA methylation and histone variants can also affect the interaction between the DNA and the histone octamer. For example, post modifications on the histones can affect the binding between the DNA and the histone octamer. Hyperacetylation on histone tails can lead to a 1.1 to 1.8 fold increases in nucleosome DNA accessibility[28] and induces a more stabilized structure by about -1 Kcal/mol compared to unacetylated nucleosome[29]. Moreover, acetylation at H3 K115 and K22 can induce a lower histone octamer binding free energy, especially when the acetylation site is near the dyad position[30]. Additionally, DNA methylation at cytosine CpG dinucleotides can induce repression of gene expression[31], and this modification results in a reduction of bending flexibility[32] and affecting the nucleosome positioning[33]. Studies of the crystal structure show that methylation on the A-DNA induces a unfavorable bending structure under the presence of spermine[34]. Furthermore, histone variants, for example, CENP-A can induce unwrapping of the DNA [35], and H2A.Bbd replaced nucleosome only organizes 118 bp wrapping DNA[36].

1.1.3 Core Histones and Linker Histones

1.1.3.1 Core Histones

Histone proteins are proteins that closely bind to DNA in the chromatin. Histone proteins are persevered in the last step of genetic material preparation experiments in eukaryotic cells. Histones had been considered as the fundamental genetic material for carrying and passing genetic information before the discovery of the DNA structure. According to the hypothesis of 'histone code' [37, 38], the modifications of histone tails and replacements of histone variants may be controlling the states of chromatin and subsequently affecting the gene activation process. This novel view is now becoming more and more popular and extensive work has been done on this idea. Histone proteins possess very special structural properties, the C-terminal and N-terminal tails of the histones are flexibly unstructured peptides. The center of the protein possesses a folded alpha-helix structure, along with two beta sheets. The histone proteins are thought to be highly conserved. H3 and H4 are more conserved compared to H2A and H2B[39]. Nevertheless, a histone tail modification gives this conserved protein an extremely complex function in the gene regulation process.

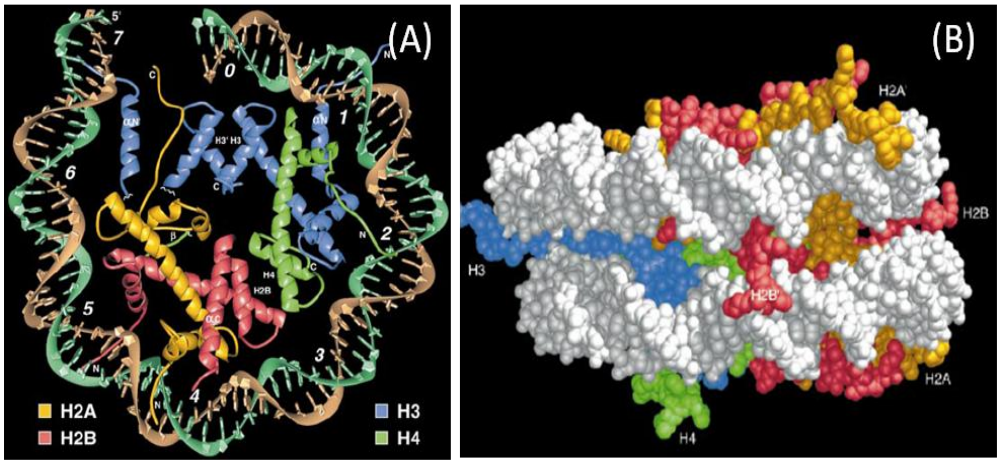

Figure 1-4 Crystal structure of the nucleosome core particle

(A) shows the crystal structure of the nucleosome core viewed from the face of the nucleosome disc; 73-bp DNA around 7 helix turns are required for one round of wrapping, (B) Side view of the nucleosome core shows the colored histone tails extending out through the minor groove between the two DNA super-helixes(white). Images were taken from Reference[7].

Figure 1.4 shows the crystal structure of the core nucleosome. Four histones H2A, H2B, H3 and H4 form a octamer located inside the DNA loops; histone tails are extending out from the histone octamer through the minor grooves between the DNA gap[7]. A detailed study on the salt dependent reconstitution on the nucleosome array shows that the H3/H4 forms a tetramer and binds to the DNA at 1.0M NaCl with a space of 207 bp intervals. In 0.8 M NaCl, H2A/H2B dimer start associating into the structure, the methidium propyl EDTA.Fe II digestion pattern looks almost the same as a fully reconstituted nucleosome array at this state; and in 0.6M NaCl,

nucleosome reconstitution is complete [40]. The arrangement of the core histone octamer is in a H2A/H2B-(H3/H4)2-H2A/H2B fashion. Most importantly, the wrapping DNA length is determined by the H3/H4 tetramer[41], and the DNA foot printing shows the same positioning signal between the H3/H4 tetramer and octamer nucleosome arrays[20].

Compared to the global domain of the core histones, the tails of the four core histones have gained more interest. Extensive modifications, such as methylation, acetylation, phosphorylation, and ubiquitination, can happen to the histone N-terminus and C-terminus. These modifications are directly involved in the gene regulation process[37, 42, 43]. A detailed introduction of the histone tail is given in chapter 2.

1.1.3.2 Linker Histone:

Between the connections of each nucleosome along the DNA templates, linker histones are identified to help stabilize the chromatin structure[44, 45]. Histones H1 and H5 are the two most common linker histones, the N-terminus ('nose') and C-terminus ('tail') of the linker histones are unstructured, flexible peptide chains consisting of strongly basic amino acids. The center of the linker histone is a non-polar central globular domain (head'), which consists of helix bundles, and beta-hairpins. This organization is defined as 'nose-head-tail' structure. The C-terminal domain is necessary for high binding affinity between linker Histone H1 to chromatin. Crystal

structure of the globular domain of histone H5 (GH5) illustrates that the DNA-binding domain is similar to the catabolite gene activator protein CAP[46]. Furthermore, two DNA-binding sites were found on the globular domain of H5 and the second binding site protects the structure of bulk or 5s nucleosome array [47]. Similarly, a systematic mutagenesis study reveals that there are two distinct DNA binding sites located in the linker histone H1's globular domain: one binding site specifically binds to the major groove close to the nucleosome dyad position, and the second one interacts with linker DNA close to the nucleosome core[48]. Linker histones can further compact the nucleosome array by affecting the entry and exit angle of the linker DNA[49].

Additionally, recent research shows that linker histones are depleted in the transcriptionally active domain of chromatin; the stoichiometry between linker histones and the nucleosome variation depends on the transcriptional level[50]. A closer investigation of the correlation between the amount of linker histones and the transcriptional level indicates that the linker histone might work as a repressor in gene regulation [50].

1.1.4 Chromatin Structure and Gene Expression

It has been found that the state of the chromatin condensation level was directly related to the transcriptional activity[51]. In order to access the DNA target that wrapped around the histone octamer to initiate the gene

transcription, a process called gene activation, which carries out a series of processes including unfolding of the chromatin and unwrapping of the DNA from the histone octamer, is required. Mouse Mammary Tumor Virus (MMTV) long terminal repeat (LTR) is a well established *in vitro* model for studying transcription activation. In the MMTV LTR region, there are six positioned nucleosome families, defined as Nucleosome A (Nuc-A) to (Nuc-F)[52] (Figure 1-5). The transcription of MMTV is triggered by the glucocorticoid hormone, which induces the binding of the glucocorticoid receptor (GR) to the MMTV promoter region located in the Nuc-B and C regions[53, 54]. Such binding changes the local environment of the nucleosome, and causes the DNA to be available for the ATP-dependent chromatin remodeling complex, such as, human Swi-Snf's accessing. HSwi-Snf complex is a multifunctional remodeling machine, which can change the chromatin structure for other transcription factors such as nucleic factor NF-1 and Oct-1 binding. Two unfolding processes are involved in the gene activation. The first one is the unwrapping of nucleosomal DNA from histone core. This process has been studied extensively by both traditional biochemistry methods and modern biophysics tools.

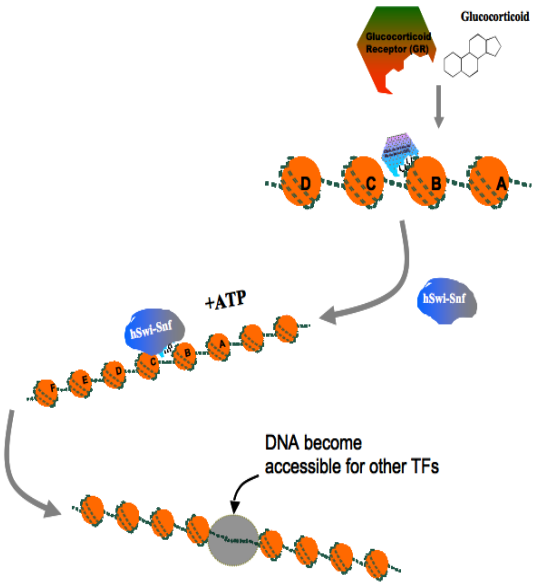

Figure 1-5 Gene activation process in the Mouse Mammary Tumor Virus (MMTV) promoter region

The second process involves the chromatin switching between the condensed state and the decondensed state, and this process has been studied in the tandem array of MMTV promoters by GFP-tagged glucocorticoid receptor (GR). The level of the decondensation of the MMTV promoter region has been shown to be directly related to the transcriptional level[55]. Additionally, the folded chromatin structure is also directly related to tumor cell progression according to a new theory on cancer pathogenesis proposed by Peter Duesberg[2]. In the highly compact form of chromatin, chromosome aberration directly relates to the canceration, Figure 1-6 shows the karyotype of chromosomes from tumor and normal cells. Normal cells show 23 pairs of standard chromosomes, while in tumor cells, the irregular aneuploid karyotype is presented[2]. The detailed mechanism behind the

nucleosome folding and the geometrical properties of the folded chromatin are still unclear.

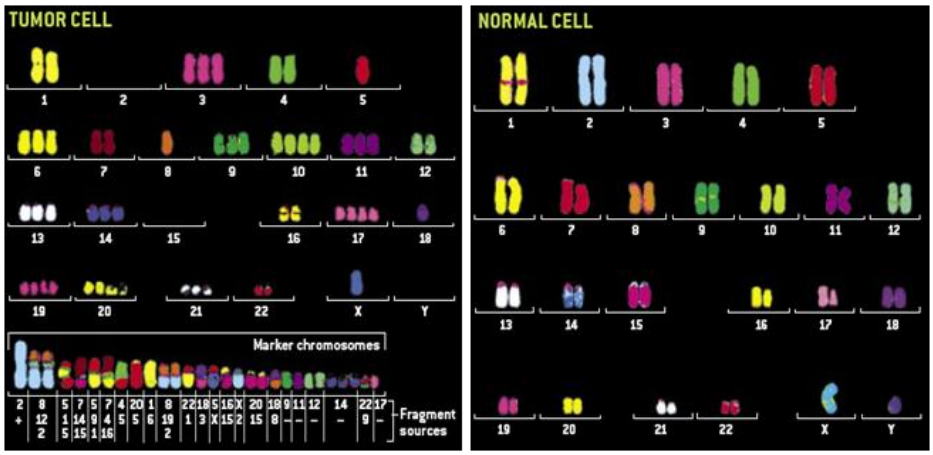

Figure 1-6 Karyotype of chromosome in tumor and normal cells

Images of the chromosomes in cancer and normal cell lines by fluorescence in situ hybridization (M-FISH) technique. Photo was taken from Peter Duesberg in reference[2].

1.1.5 Chromatin Higher Order Structures

Although the structure of the single nucleosome core has been discovered by crystallographic data at 1.7angstrom resolution[7], the structure of chromatin '30 nm' fiber is still an enigma. The '30-nm' fiber structure was first found in the early electron microscopy (EM) study of Hela metaphase chromosomes[13] and was also observed in chicken erythrocyte from an X-ray scattering *in vivo* experiment[14]. However, in situ observation of mitotic

chromosomes of the HeLa S3 cells by EM did not reveal any 30-nm fibers, but instead demonstrated a highly disordered and interdigitated structure of chromatin [56]. Therefore, the existence of '30' nm structure of chromatin in the live cell nuclei is still under extensive debate. Since the hierarchical level of chromatin structure is directly related to their function roles in the cell genetic and epigenetic regulation, it is important to understand its structure and how the structural changes correspond to its functions. Following extensive biophysical and biochemical experiments for the last two decades, two types of structures are confirmed so far: 'one-start' helix or solenoid structure, and 'two-start' helix model.

1.1.5.1 Type I : 'one-start' helix/solenoid structure

For the 'one-start' solenoid model, the helical turn of chromatin starts from the 10 nm 'beads-on-string' structures. By coiling this 10 nm fiber around an axle like a solenoid, a new folded structure with 6 nucleosomes per turn presents (Figure 1-7 A-a). Each nucleosome 'n' in this structure is directly related to its neighbor 'n+1' and 'n-1' nucleosomes as in the 10 nm fiber. 'One-start' indicates that the whole structure formed by twisting one helix ribbon by the beads-on-string structure, and the repeat unit for the 30 nm helix turn structure is composed from one ribbon. For the 'one-start' solenoid model, the linker DNA is merged inside of the 30 nm fiber[57]. The final diameter of the folded structure is independent of the linker DNA length[58,

59]. Even for a long nucleosomal repeat length ~240bp, 30 nm filaments of solenoid compaction were observed[60]. Since the structure of the final chromatin is independent of the linker DNA, a bending of the linker DNA is required for the highly compact organization[60].

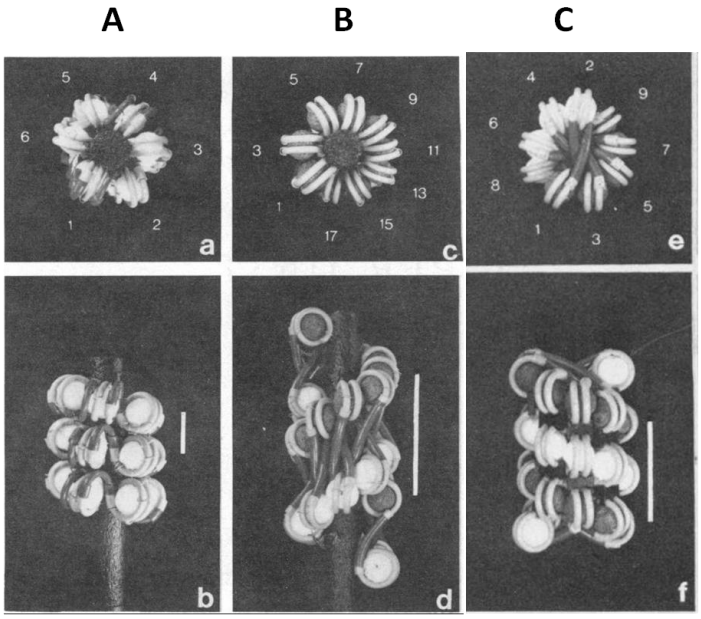

Figure 1-7 Different Models of ’30-nm’ chromatin fiber

The models of the '30 nm' chromatin fiber: (A) ‘one-start’ solenoid structure, (B) and (C) are the ‘two-start’ model, (B) is the 'twisted-ribbon' model, and (C) is the 'crossed-linker' model. Images were taken from reference[5].

Another structure model that supports a ‘one-start’ model comes from a recent study of the long nucleosome array templates with a linker DNA length ranging from 30 to 60 bp based on the 601 sequence[61]. A new model defined as interdigitation packing is proposed, which is shown in

Figure 1-7 [6]. This model is also in agreement with a recent Electron Microscopy (EM) observation of living eukaryotic cells, which shows interdigitated organizations of nucleosomes[56].

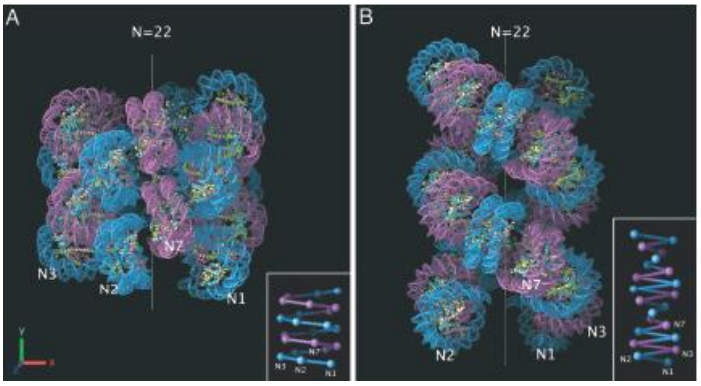

Figure 1-8 Model for the interdigitated compaction

Interdigitated 'one-start' model for the '30 nm' fiber, model constructed from the EM data with the presence of linker histone. (A) model for the short linker length, and (B) model for the long linker length. Image was taken from reference[6].

1.1.5.2 Type II: 'two-start' helix Model

For the 'two-start' helix model: The 10-nm nucleosome fiber first forms a zig-zag pattern, which usually shows in low-salt chromatin preparation experiments. This zig-zag structure looks like two separate nucleosome arrays packed together parallelly. That is why this model is defined as 'two-start' model. Following this arrangement, the zig-zag ribbon wraps around an axle and generates a 30 nm fiber structure. In the 'two-start' model, each nucleosome n in this structure is directly related to its 'n+2' and 'n-2'

neighbors as in the 10 nm fiber due to the zig-zag pattern. In other words, the difference between 'one-start' and 'two-start' is that 'two-start' structure has a zig-zag arrangement of nucleosome. Two subtypes of 'two-start' models have been proposed. The first one is the 'twisted-ribbon' model, in which the linker DNA is parallel to linker DNA [62] (Figure 1-7B). The other one is the 'crossed-linker' model, in which the linker DNA is perpendicular to the 30 nm fiber axis[5](Figure 1-7C).

Evidences have been found to support the 'two-start' model. For example, a compacted 12mer nucleosome array constructed with disulfide cross-linking between H2A, H2B and H4 shows a zig-zag arrangement of the nucleosome array from the Electron Microscopy (EM) study and support the 'two-start' model [63]. More strong evidences came from the same group by a X-ray crystallization study of a 'tetranucleosome' structure at 9 angstrom; a zig-zag pattern was clearly identified[64]. However, the tetranucleosome array is designed with a very short linker DNA length 167(20bp), and thus cannot represent longer and non-uniform linker DNA lengths construction found in nature[64]. The work on the tetranucleosome crystal structure was considered as a landmark for the further discovery of the 30 nm structure as a zigzag 'two-start' arrangement.

1.2 Biophysical Methods for Chromatin Study

1.2.1 Atomic Force Microscopy

1.2.1.1 Introduction of Atomic Force Microscopy

In 1981, Gerd Binning and Heinrich Rohrer developed a powerful instrument which could obtain atomic resolution of materials and was also capable of atomic scale manipulation on the sample surface. The instrument is called Scanning Tunneling Microscopy(STM)[65], which won the Noble Prize in Physics in 1986. Only five years after this invention, following the similar idea of STM, Binning, Quate and Gerber developed another scanning probe microscopy, which could operate on an insulator surface by measuring the force between the sample and the probe[66].

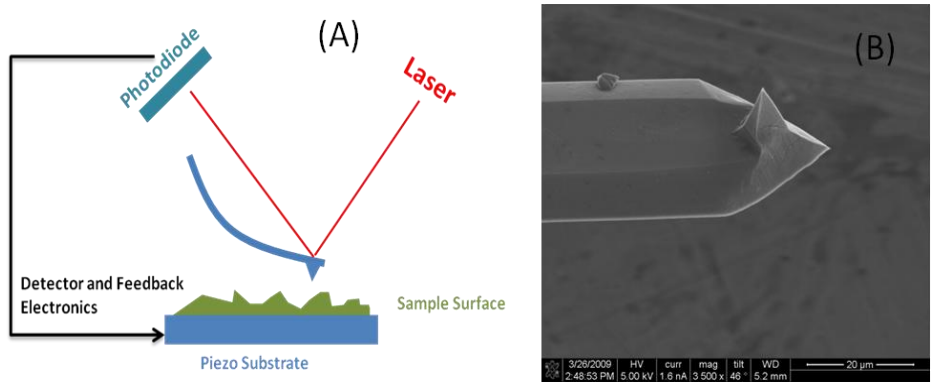

Figure 1-9 Schematic illustration of Atomic Force Microscopy (AFM) (A) Shows the scheme of the Atomic force microscopy, and (B) a typical SEM imaging of an AFM tip with a sharp edge. (Image A was taken from Askwmind at Wikipedia)

Figure 1-5 shows the scheme of the atomic force microscopy, a beam of laser directly points to the edge of the cantilever tip, a very sharp probe which has a diameter around 20 to 100 nm. The reflection signal is deflected to the photodiode detector. The bending of the tip is reflected as the laser spot position on the photodiode, and recorded by the detector. A feedback loop is applied to the sample piezo stage (or the tip piezo stage) to keep the deflection of the cantilever (position of the laser) constant. The movement of the piezo is recorded to reflect the sample surface topography information.

Based on the AFM tip's interaction with the surface, two major modes of AFM operation are defined: contact mode and non-contact mode. In contact mode, the AFM tip is gently contacted to the surface, the force between the tip and the sample surface induces a bending of the cantilever. Based on the Hook's principle, see equation 1-1, the bending z could be monitored by a beam of laser based on proper conversions. Atomic resolution can be obtained by the contact mode[67, 68]. However, the contact mode AFM can hardly be applied to soft biological samples, especially in physiological environments. In order to imaging soft samples and minimizing the interaction between the tip and sample, non-contact mode AFM was developed[69-71].

$$F = kz \qquad 1\text{-}1$$

For the non-contact mode AFM, two major sub-modes of operation were developed: amplitude modulated AFM (AM-AFM) [69]and frequency

modulated AFM (FM-AFM)[70]. Basically the AFM tip is excited to oscillate under an external force $F_0 \cos(\omega t)$ (equation 1-2). The tip motion can be modeled as a one-dimensional oscillated point-mass with damping under another tip-sample force F_{ts} when the tip is close to the surface. The dynamic properties of the tip vibration, such as amplitude, phase, and frequency, can be monitored. Based on the change of these parameters, the interactions between the tip and the sample surface can be derived.

$$m\frac{\partial^2 z}{\partial t^2} + \frac{m\omega_0}{Q}\frac{\partial z}{\partial t} + kz = F_{ts} + F_0\cos(\omega t) \qquad 1\text{-}2$$

In AM-AFM, the tip is vibrated at a fixed frequency. The amplitude of the tip vibration is measured by filtering and amplifying the selected frequency signal through a lock-in amplifier. The amplitude of the tip vibration is then fed into the proportional–integral–derivative (PID) controller. A setpoint of the amplitude decreasement is defined to characterize the tip and sample surface distance, by keeping the tip vibration amplitude the same through the PID controller, the piezo sample stage keeps adjusting the surface and tip distance. By tracing the piezo movement, the surface topography information can be obtained. A harsh interaction between the tip and sample surface can be avoided by using the amplitude modulated AFM. AM-AFM is now widely used in the research labs around the world for most imaging

purposes, especially for bio-molecule study. However, AM-AFM cannot obtain atomic scale resolution[72]. Compared to AM-AFM, frequency modulated AFM (FM-AFM) offers a high signal-to-noise ratio[70]. In FM-AFM, the tip is oscillated at its eigenfrequency with a high quality factor (Q). The amplitude of the oscillation is held at the same, and the shift of the frequency is monitored. The distance between the tip and the sample surface is changed according to the frequency shift, and topography information can be obtained by tracing the piezo positions. True atomic resolution of Si (111)-(7x7) was obtained by FM-AFM in 1994[73]. Several reviews on the Atomic Force microscopy can be found in these references[74-76].

1.2.1.2 Introduction to Recognition AFM

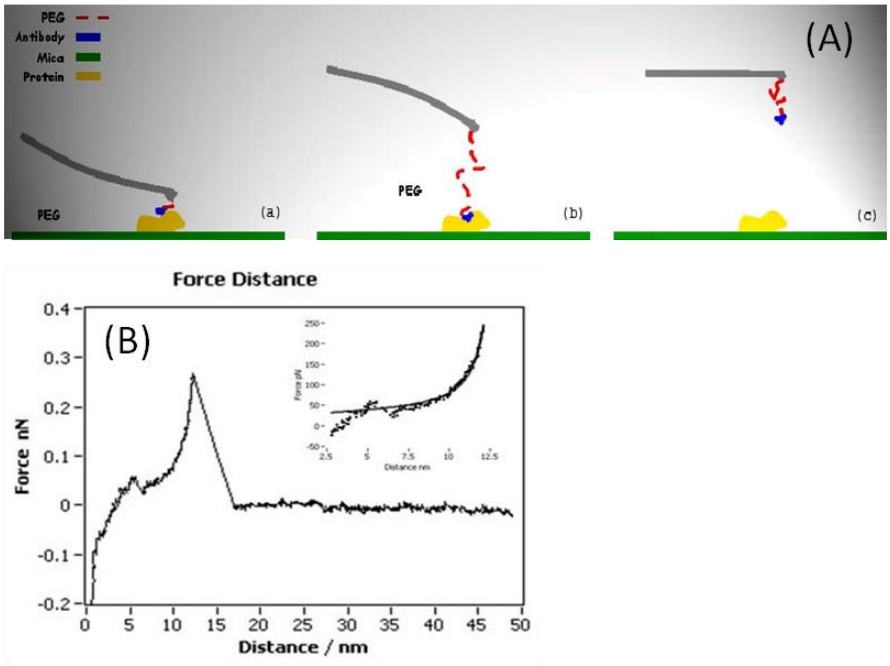

Figure 1-10 Intermediate States for Force Distance Experiments

(A)Shows typical force spectroscopy experiment: the antibody (in blue color) is covalently bound to the Polyethylene Glycol (PEG) linker, which is tethered to the AFM tip through the other end; the AFM tip is first moved closely to the sample surface, and the antibody takes the chance to bind to the protein (yellow) on the surface (a), the tip is then moved away from the surface, PEG is stretched during this process due to the antibody protein specific interaction(b), and when the force is acquired big enough, antibody and protein dissociate (c). (B) Typical single molecule force distance experiment result, which shows the extension of the PEG linker correspond to the force applied to it. The insert shows the fitting of the result to the worm like chain (WLC) model. Image B is the data obtained in the Lindsay Lab.

Atomic force microscopy has been widely used to characterize the geometrical properties of different biomolecules. However, despite the huge size differences, most proteins with spherical geometries displayed as globule particles in the topological image. It can be very valuable to identify the chemical Identities of individual proteins while the topography information is recorded. Several techniques have been developed to study the chemical information of the surface by AFM, for example, chemical force microscopy, which modifies the AFM tip with specific organic molecules, is exploited to study the specific chemical group interactions between the tip and surface by the adhesion and frictional forces[77]. Force-volume mapping, which uses the AFM tip to repeatedly indent soft samples, and the force distance curves are collected at each indentation. By repeating the process at different postions, a map of the sample elasticity map is reconstructed from these force distance curves[78]. Lastly, force distance curves, which a specific ligand or antibody through a Polyethylene glycol (PEG) linker is ligated to the AFM tip, can be used to study the surface chemical information. Force distance curves are then collected (Figure 1-8A), and specific interactions are identified by stretching of the PEG linkers[79]. All these methods are time-consuming and do not provide topography information at the same time[80].

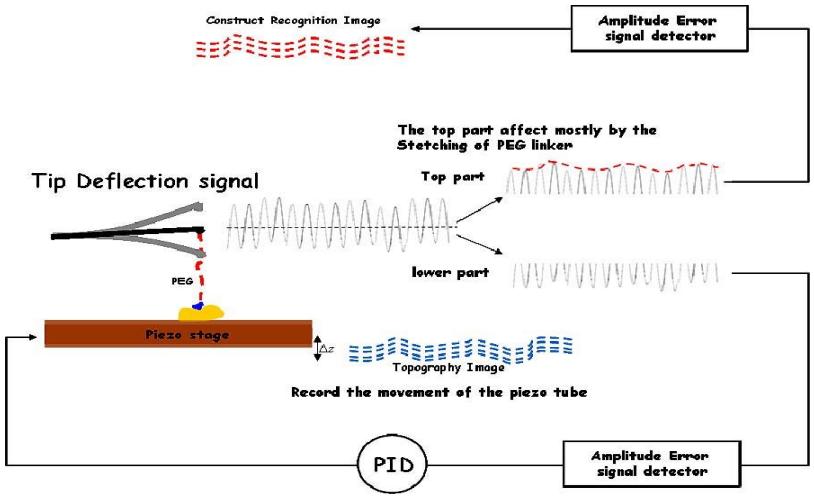

Figure 1-11 Schematic illustration of the Recognition AFM

The AFM tip vibration signal is monitored and separated to upper swing and downswing section, the down swing signal is fed into a PID controller for topological information, and the upswing is used for constructing the recognition imaging.

With the purpose of recording the surface chemical information while the topography image is constructed simultaneously, recognition AFM was developed. It tethered an antibody to the end of the AFM silicon tip through a PEG linker with a specific length (Figure 1-9). Then the AFM tip is coated with magnetic materials and driven by a magnetic field[81]. While the tip approached to the surface and oscillated on top of the sample, the downward swing of the tip can sense these short range tip-sample interactions and the corresponding amplitude change can be detected and trigger the response of the PID controller to keep a constant amplitude. Therefore, topography

information can be obtained. Meanwhile, when the tip-sample distance is optimized, the upward swing of the tip can only be affected mostly by the stretching of the PEG linker. The reduction of the amplitude change ΔA can be calculated with equation 1-3. In the equation, $S(z/L_0)$ is the stiffness of the PEG under fractional tension z/L_0, K is the spring constant of the AFM cantilever, A_Z is the Amplitude while the tip stretching the PEG, and A_0 is the amplitude with no stretching of PEG.

$$\frac{A_0}{A_z} = \frac{S(z/L_0)}{K} + 1 \qquad 1\text{-}3$$

The stretching of PEG shown in Figure 1-9 can be simply described by the Worm-Like Chain (WLC) model[82] (Equation 1-4).

$$F \approx \frac{k_B T}{L_p}\left(\frac{1}{4}\frac{1}{\left(1 - z/L_C\right)^2} - \frac{1}{4} + \frac{z}{L_C}\right) \qquad 1\text{-}4$$

Here L_p is the persistence length of PEG, which is ~0.38 nm[83], L_C is the contour length of PEG, and $k_B T$ is the thermal energy, which is 4.1 pN.nm at room temperature. When the extension z over the contour length L_C of PEG

is small, the force required to hold this extension is linearly depended on z/L_C with a slope $k_B T/L_p$.

The information derived from the upswing of the AFM cantilever was used to construct the recognition image to illustrate the specific interaction between the tip and the sample[84](Figure 1-10).

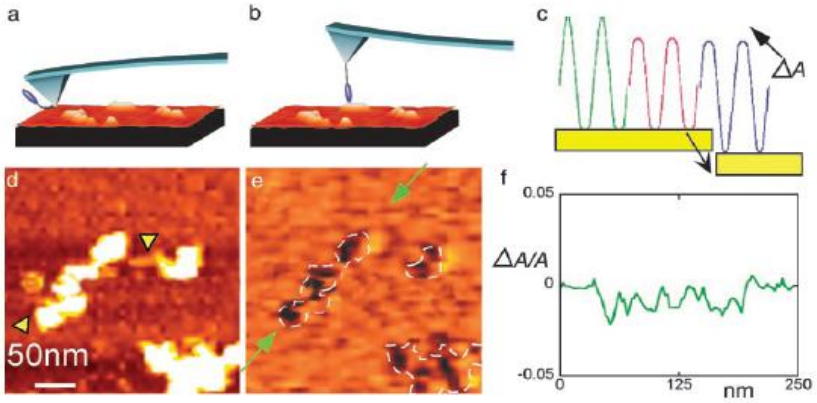

Figure 1-2 An example of studying nucleosome array by Recognition AFM Figure (c): the tip's vibration amplitude changed when the antibody bound with the antigen, green line, no bind, red the antibody bind to the surface protein, blue, when the servo pulled the tip away from the surface protein; (d, e) show the topography and recognition image of MMTV(Mouse mammary tumor virus) nucleosome array by anti-H3 tethered on the tip; the black spots in (e) correspond to the amplitude decrease, which caused by the stretching of the PEG linker (f) shows a plot of the recognition signal as a fraction of the full amplitude (ΔA/A) measured along the line joining the green arrows in (e).(1) The amplitude decrease can be calculated by equation (1-3)[9]. Image was taken from reference[9].

1.2.2 Hydrodynamic Methods for Chromatin Study

Two most common hydrodynamic methods: analytical ultracentrifugation and analytical agarose gel electrophoresis (AAGE) were used to study the

conformation changes of chromatin under different environments in the buffer [85, 86].

1.2.2.1 Analytical Ultracentrifuge Analysis

Ultracentrifugation technique is one of the most important techniques for biological sample purification in the early days. This technique was also developed to characterize the macromolecule's molecular weight and the geometrical conformations. This new technique is defined as Analytical Ultracentrifugation, and two subtypes of experimental methods have been developed: sedimentation velocity and sedimentation equilibrium. The theory for these two methods is the same: consider a macromolecule in a homogeneous solute as illustrated in Figure 1.11. Under high speed ultracentrifugation, the molecules experience three forces: friction force, buoyant force and gravitational force.

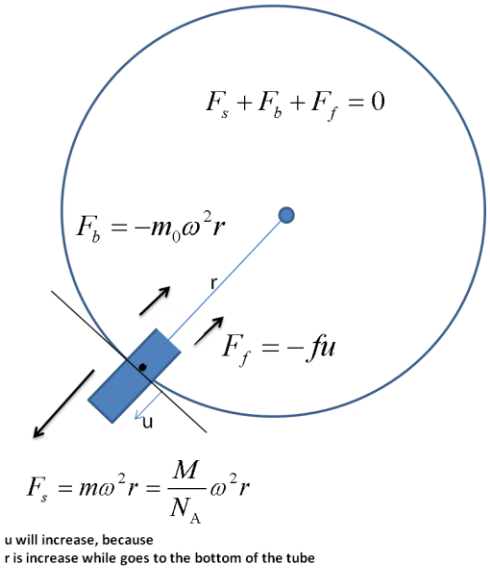

Figure 1-3 Forces for the macromolecules under ultracentrifuge

A comprehensive review of the principles and experimental approaches of analytical ultracentrifugation can be found in [87]. Since the molecule is moving towards to the opposite direction of the center through a viscous environment, the molecule experiences a friction force, which is proportional to the velocity:

$$F_f = -fu \qquad 1\text{-}5$$

Here f is the friction coefficient, a parameter that affected by the shape and the size of the molecule. Extended and flexible molecule with larger surface area has a high value of friction coefficient. Also, a buoyant force based on Archimedes' principle can be described as:

$$F_b = -m_0 \omega^2 r \qquad 1\text{-}6$$

Here, m_0 is the mass of fluid displaced by the particle, $\bar{v}$ is the volume in ml that each gram of the solute occupies in solution (partial specific volume; the inverse of its effective density), ρ is the density of the solvent with a unit of g/ml. Gravitational Force can be written as:

$$F_s = m\omega^2 r = \frac{M}{N_A} \omega^2 r \qquad 1\text{-}7$$

Here, ω is angular velocity, in radians per second, M is the molar weight of the solute in g/mol, N_A is the Avogadro's Number. Three forces come into balance:

$$F_s + F_b + F_f = 0 \qquad 1\text{-}8$$

$$\frac{M}{N_A} \omega^2 r - \frac{M}{N_A} \bar{v} \rho \omega^2 r - fu = 0 \qquad 1\text{-}9$$

$$\frac{M(1-\bar{v}\rho)}{N_A f} = \frac{u}{\omega^2 r} = s \qquad 1\text{-}10$$

Based on the above derivation, several parameters can affect the final sedimentation coefficient s : the molecular weight of the macromolecule, density of the macromolecule, and also the friction coefficient which reflects the geometrical flexibility of the macromolecule. In principle, sedimentation coefficient can reflect the characteristic properties of a structure. Assuming the same friction coefficient and buoyant density, a molecule with higher

molecular weight possesses a higher sedimentation coefficient. For the same macromolecule, such as a nucleosome array, its buoyant density and friction coefficient would not be the same in their extended and folded form. The extended form has a smaller sedimentation coefficient. However, only based on the sedimentation coefficient to predict the real characteristic properties of the macromolecules is not accurate. Considering a situation with similar molecular weight, but different friction coefficient or buoyant density, the sedimentation may vary but their geometrical structure may be the same, a detailed discussion can be found in chapter 2.3. The sedimentation coefficient has a unit of time, and denote as 'svedbergs', one 'S'=10^{-13} second. In nucleosome array folding experiment, a sedimentation coefficient of 55s corresponds to a highly compacted 30 nm structure.

1.2.2.2 Quantitative Agarose Gel Analysis

Agarose gel electrophoresis is one of the most popular techniques used in almost all the bio-related labs. This technique can be used to get macromolecule size in a short time and easy way. The principle of this technique is: the surface charge density of a macromolecule can affect the migration speed of the molecule in the agarose gel under electrical field [88]. The size, shape and geometrical flexibility of the macromolecule can affect the interaction between the sample and gel pores, this interaction can further change the migration speed of the molecule described in equation 1-11.

$$u / u_0^{'} = (1 - R_e / P_e)^2 \qquad 1\text{-}11$$

Here $u_0^{'}$ is the intrinsic gel-free mobility, which represents the electrostatic properties of the chromatin, R_e is the effective radius of the macromolecule and reflects the conformational properties of chromatin, P_e is the average radius of the gel pores, u is the measured mobility from the experiment. The unit for the gel mobility is $cm^2 \, v^{-1} s^{-1}$.

The term 'gel-free' mobility is defined as the mobility of a macromolecule without any interactions between the molecules and pores. In the early works on the chromatin structure, agarose gel becomes a useful tool to quantitatively characterize the chromatin folding level in buffer based on the surface charge density[86, 89-91]. In the quantitative Agarose gel analysis experiment, a reference such as T3 phage or charged microspheres with defined radius are introduced into the quantitative gel setup to derive for the gel pore sizes at different concentrations. Several different concentration agarose gel lanes are prepared in order to get the gel free mobility of the reference. Sometime, the gel free mobility need to be corrected due to the contribution from electrosmosis[85, 92]. With the known radius R_e and measured gel free mobility u_0^{T3}, the gel pore sizes P_e in different concentration agarose gels can be derived from equation 1-12:

$$P_e = \frac{R_e}{[1-\sqrt{\frac{u_{T3}}{u_0^{T3}}}} \qquad 1\text{-}12$$

After measuring the gel pore size, the effective radius of the macromolecules can be obtained by applying the measured mobility speed of the species in the gel. For example, in equation 1-13, $u_{chromatin}$ is the measured mobility of chromatin sample under different concentration gels, and $u^0_{chromatin}$ is the gel free mobility of the chromatin sample. The chromatin size R_e under different concentrations can be derived from equation 1-13.

$$R_e = P_e \times [1-\sqrt{\frac{u_{chromatin}}{u^0_{chromatin}}}] \qquad 1\text{-}13$$

Chapter 2

2 A Study of Histone Tail Modification by Atomic Force Microscopy (AFM)

2.1 Introduction

Epigenetic modifications on histone tails, such as methylation, acetylation, phosphorylation and ubiquitination play a critical role in the gene regulation process[37, 42, 43]. The tails of histones are flexible, unstructured and highly basic polypeptides, which carry about 30% mass of the whole histone proteins. These tails extend out from the histone octamer through the minor grooves between the DNA gap[7]. The functional role of the histone tail has been studied for a long time. Early hydroxyl radical foot printing result shows that the histone tail preferentially binds to the linker DNA[93] and doesn't affect the nucleosome positioning[20]. Furthermore, a dinucleosome modeling followed by cross-linking study reveals that the N-terminal tails of H2A and H2B contribute mostly to the inter-nucleosomal histone-DNA interactions[94]. However, single nucleosome unwrapping experiment suggests that H2A/H2B tails interact specifically ~36bp from the dyad location[41]. Those different observations indicate a multiple functional roles of the histone tails.

Histone tails have also been confirmed indispensable for high compaction of nucleosome array with the presence of Mg^{2+} [95] [96] [97]. Among all the histone tails, amino acids 14-19 of histone tail H4 are the most important for higher compacted nucleosome array [93, 98]. Recently it has been shown

that the histone tail modifications can also affect the Mg^{2+}-dependent nucleosome compaction. Among the variety of histone tail modifications, acetylation is one of the most important modifications in structure control. For example, acetylation on histone tail H4 inhibits the formation of 30 nm fiber both in short and long nucleosome template arrays[10, 99], while acetylation on histone H3 results in a lower histone octamer binding free energy[30].

Besides their primary function on the chromatin compaction control, histone tails are also directly involved in the transcription activation and initiation processes. The N-terminal tails of H3 and H4 are essential in the p300-dependent transcriptional activation[100]. In a repressible acid phosphatase (PHO5) promoter model, core histone acetylation is required for a transcriptionally activated state[101]. Additionally, H4-K16 acetylation inhibits the ATP-dependent chromatin-assembly factor (ACF)'s activity in the mono nucleosome's mobility experiments[10]. Acetylation of histone H3 at lysine 56 is directly related to DNA repair during S phase[102].

Hydrodynamic methods such as sedimentation coefficient and analytical agarose gel electrophoresis(AAGE) have been widely applied to characterize the histone tails' effects on the folding of nucleosome array in various buffer environments[86, 92]. The AAGE results have been well correlated to the sedimentation coefficient values[91]. In this section, we presented a study of

folding behavior of the nucleosome array reconstituted with various templates and H4-K16 acetylation in the presence of Mg^{2+} by Atomic force Microscopy (AFM). AFM imaging didn't show apparent difference between H4-K16 acetylated nucleosome array and wild-type (with no modification) control sample both in saturated nucleosome array reconstituted from 601 template (177-9) and unsaturated one from MMTV promoter sequence. We propose that the sedimentation experiment without confined buoyant density may not be enough for the interpretation of the geometric properties of the macromolecules. Moreover, the loading of histone octamer along the templates was found to primarily contribute to Mg^{2+} induced high compaction.

2.2 Materials and Methods

2.2.1 Materials

Three DNA templates were used for the experiment. Template I: MMTV promoter region (a generous gift of Gordon Hager, National Institutes of Health, Bethesda). Double digestion with Nco I and Hind III liberates the 1.9 kbp Mouse Mammary tumor virus (MMTV) promoter region. Temple II: 601 repeat sequence, by digesting the plasmid pM17 (a generous gift of Michael Poirior, Ohio State University) with EcoR I and Hind III. Three DNA fragments were produced: One 1454 bp EcoR I (2) ~Hind III (1455) with 8 -177-601 repeat, one 1588bp Hind III (1455) ~Hind III (3042) fragment with 9-177-601 repeats, and another piece of vector DNA. Temple III: by digesting

plasmid pM17 with EcoR I alone, a concatenated sequence with half 601 region with 17-177-601, and half non-601 region from the vector was obtained. The entire DNA fragments were purified by phenol-chloroform extraction and followed by ethanol precipitation.

2.2.2 Acetylation of H4 K16:

Fully unacetylated recombinant Xenopus laevis histone H4 was expressed and purified according to standard protocols [103] Chemical method was applied to the H4-K16 acetylated sample preparation: a synthetic peptide tail correspond to the H4 tail with acetylation at K16 was chemically ligated to the other portion of the recombinant histone H4 fragment lacking the tail (minus amino acids 1–22 with R23C mutation).[10] Histone H4 peptide of the N-terminal tail of histone H4: AGRGKGGKGLGKGGAK(Ac)RHRKVL(1-22) was chemically synthesized (Protein Chemistry Core Laboratory, Baylor College of Medicine, Houston, TX). The side-chain protected peptide (5 mM, final) was deprotected and cleaved from the resin, followed by C-terminal activation by DCC (100mM) in DMSO. After activation, the peptide was reacted with benzyl mercaptan (100 mM) at 25 °C for three hours. The cleaved and deprotected peptide was purified by C18 reverse phase high performance liquid chromatography and characterized and confirmed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectroscopy. The thioester peptide from the above step was then ligated to the recombinant H4 histone fragment under denaturing condition, the

products were purified by cation exchange chromatography. The purified chemically ligated histone fractions were confirmed by SDS-PAGE and MALDI-TOF mass spectrometry.

2.2.3 Nucleosome Reconstitution

Step salt dialysis method was used for the nucleosome array reconstitution. H4K16 (Ac) histone octamer and control sample with recombinant Xenopus histones (generously provided by Karolin Luger, Colorado State University) Also, histone octamer isolated from chicken erythrocyte was used for loading test. The DNA and histone octamers were incubated on ice for 30 min at a ratio of 1 to 1.5 (w/w) in 1xTE buffer (10 mM Tris, pH 7.5, 1 mM EDTA) with a final concentration of DNA at 100 ng/ul, DTT at 1 mM and NaCl at 1 M. The mixture was dialyzed stepwise with a 6–8 kDa MWCO membrane (Spectra/por) into 0.8 M, 0.6 M, and 0.15 M NaCl buffer under 1xTE. The final reconstituted sample was cross-linked by dialysis against 0.1% glutaraldehyde (v:v), the sample was dialysis in 1 mM EDTA pH7.5 overnight to get rid of excess glutaraldehyde.

2.2.4 Atomic Force Microscopy:

The mica was modified with the 3-aminopropyltriethoxysilane (APTES) described before[104]. Briefly, fresh cleaved mica was placed in a desiccator with 30 ul APTES (99%, Sigma-Aldrich, St. Louis, MO) and 10 ul N, N-diisopropylethylamine (99%, distilled, Sigma-Aldrich) in the bottom. The

desiccator was then purged with Argon for 3 mins; the mica was placed in the APTES vapor for 1 hour to get good modification. 100 ul 2 uM glutaradehyde (grade I, Sigma-Aldrich) were deposited onto the APTES mica surface for 10 mins, and the surface was then washed with distilled water gently; after that, 10 uL reconstituted samples with concentration ranging from 0.3 to 0.9 ng/uL (at A260) were pipetted onto the glutaradehyde treated mica surface and allowed to incubate for 40 mins to 80 mins; the mica surface with immobilized sample was then rinsed with distilled water gently and dried with nitrogen. The imaging were carried out with a PicoPlus 2500+ (Molecular Imaging, now 5500 AFM (N9410S) from Agilent) AFM equipped with a Si3N4 cantilever (AppNano SPM) with a spring constant ranging from 25-75 N/m and the resonance frequency around 300 kHz.

2.3 Results and Discussions

2.3.1 H4K16 and WT Histone Octamer Reconstituted on the MMTV Promoter Region DNA Template:

The MMTV promoter sequence was obtained by digesting the plasmid with Nco I (1798) and Hind III (3663) and followed by gel purification. The samples were reconstituted with H4K16 (Ac) and WT histone octamer on the MMTV promoter region, which has 6 nucleosome-positioned sequence identified as Nuc-A to Nuc-F on this template [105-107]. Since every 177 bp is required for one nucleosome, 10 should be the maximum loading number in this template. The samples were first diluted into folding buffer and HEPES

buffer, followed by cross-linking with 0.1% glutaraldehyde for 40 mins before immobilization on the APTES mica surface. The AFM images of nucleosome array reconstituted with the H4K16 (Ac) histone octamer and wild-type (WT) histone octamer are shown in Figure 2.1.

Imaging analyses have been performed on these two samples (Figure 2-2). The distributions of the nucleosome loading numbers were both centered at 8, indicating that the H4K16 (Ac) and WT MMTV nucleosome arrays have a similar loading number and acetylation on the H4-K16 does not affect the nucleosome loading preference[108]. Additionally, nucleosome-nucleosome center to center contour distances were measured, and the histograms of the distribution were plotted in Figure 2-2. From the histogram, the inter-nucleosome distance distributions between the H4K16 (Ac) and WT look comparable, showing a main peak of nucleosome distance at 25 nm, and several sub peaks at 50 nm, 70 nm, and 125nm. The similar distribution of the contour distance means that the favorable positioning of nucleosome was not changed by the H4-K16 acetylation.

Both WT and H4K16 (Ac) nucleosome arrays on the MMTV promoter template show an apparent change in compaction in the presence of 1 mM Mg^{2+}. However, the difference between the compaction of WT and H4K16 (Ac) nucleosome arrays is not significant. Since the loading on the MMTV promoter region is low and difficult to control, a highly defined template

constructed with high binding preference to histone octamer sequence (601)[61] is necessary for better characterizing the structure differences between H4K16 (Ac) and WT nucleosome arrays.

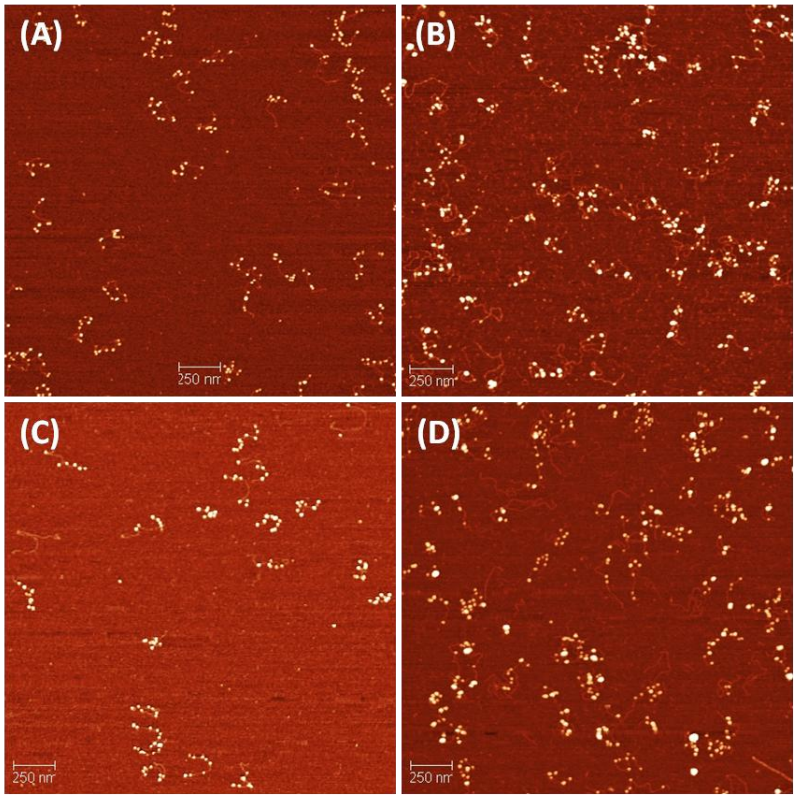

Figure 2-1 AFM images of histone H4 acetylated at lysine 16 (H4K16 (Ac)) and wild type (WT) nucleosome array on MMTV template.

Nucleosome array reconstituted with the MMTV promoter sequence, which has six positioned nucleosomes named as Nuc A to Nuc F. (A) H4K16 (Ac) in the HEPES buffer, (B) H4K16 (Ac) in the folding buffer, (C) WT in the HEPES buffer and (D) WT in the folding buffer. Nucleosome array shows extended beads-on-string structure at HEPES buffer at pH7.3, and a slightly compact structure at 1 mM Mg^{2+}.

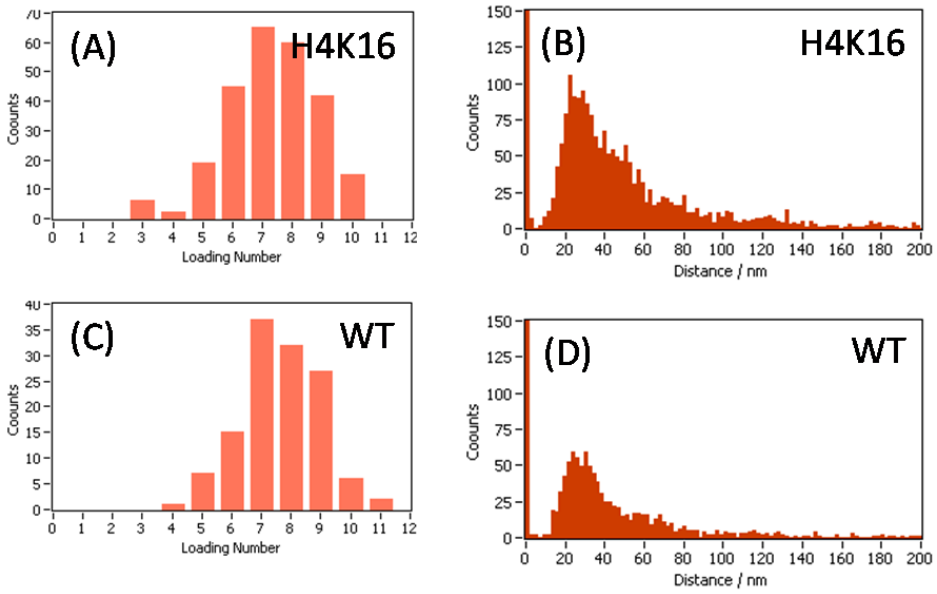

Figure 2-2 Loading number and inter-nucleosome contour distance analyses of the H4K16 (Ac) and WT nucleosome array on MMTV promoter sequence.

AFM image analysis shows that H4K16 (Ac) and WT nucleosome array have the same loading behavior in the MMTV promoter sequence. The distribution of the loading number are both centered at 8 for H4K16 (Ac) -MMTV nucleosome array (A) and WT-MMTV nucleosome array (C); and inter-nucleosome distance distribution shows the matching peak centered at 30 nm for H4K16 (Ac) -MMTV nucleosome array (B) and the WT-MMTV nucleosome array (D).

2.3.2 H4K16 (Ac) and WT Histone Octamer Reconstituted on the 601 Fragments

In order to study the structure differences between H4K16 (Ac) and WT nucleosome array in a more refined template, we used the 601 tandem repeat sequence for the following experiments. Plasmid pMP17[109] ,a gift from Michael Porior, was digested with EcoR I and Hind III. The digestion generated three templates, a non-601 sequence as buffer DNA, one 1454 bp EcoR I(2)-Hind III(1455) with eight 177-601 repeat, and 1588 bp Hind III(1455) –Hind III(3042) with nine 177-601 tandem repeat sequence. The nucleosome arrays reconstituted on the 601 sequence were mixed with another piece of vector sequence, which has a low binding affinity to histone octamer and acts as a buffer DNA. Saturated nucleosome array was obtained by using the buffer DNA from the AFM results shown in Figure 2.3. The loading number on the nucleosome array was analyzed both for nucleosome array assembled from the H4K16 (Ac) and WT histone octamer. Both H4K16 (Ac) and WT nucleosome array on the 601 template show a sharp peak distribution of loading centered at 9 nucleosomes. Since the reconstitution is in a mix of two different length templates with 8 and 9 maximum loading, we could also see another peak at around 8. Loading number of 10 might correspond to more saturation on the template. However, most of the samples are saturated at 9 nucleosomes, and no apparent difference on loading between the H4K16 (Ac) and WT nucleosome array reconstituted on the 601 template was observed. Additionally, the inter-nucleosome contour

distance distribution is almost the same with a distribution centered at 30 nm.

After exposed to 1 mM Mg^{2+}, both saturated H4K16 (Ac) and WT nucleosome array is compacted into a highly condensed particle. Since the measurement of the particle size distribution by AFM is not accurate due to the AFM tip broadening, the maximum height distributions of the condensed particle were measured instead. Both H4K16 (Ac) and WT nucleosome array folded particles have a maximum height distribution centered at around 7 to 9 nm. The images for the compacted H4K16 (Ac) and WT were acquired by the same tip using the same imaging parameter for each separate experiment to overcome the tip-tip variation effects in the AFM measurement. The variation of height measurement between each experiment is coming from the tip effect. However, comparison of the results within each independent experiment shows no apparent change in the maximum height distribution between the H4K16 (Ac) and WT nucleosome array.

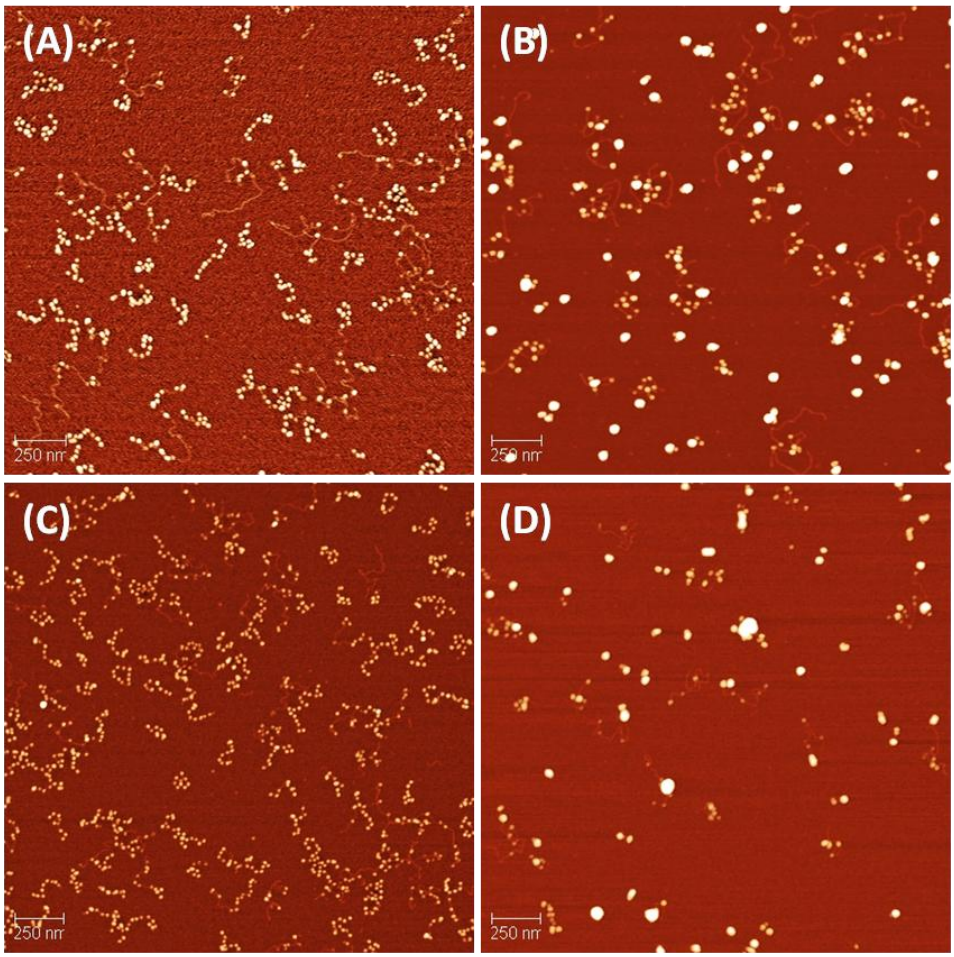

Figure 2-3 AFM images of the nucleosome array reconstituted with the DNA template II.

Nucleosome array reconstituted with the DNA template II, which has two pieces of 601 templates, one with 8 repeats and the other with 9 repeats. (A) H4-K16 in the HEPES buffer, (B) H4-K16 in the folding buffer, (C) WT in the HEPES buffer and (D) WT in the folding buffer. Nucleosome array shows extended beads-on-string structure in HEPES buffer at pH7.3, and folded to compact particle with 1 mM Mg^{2+} present.

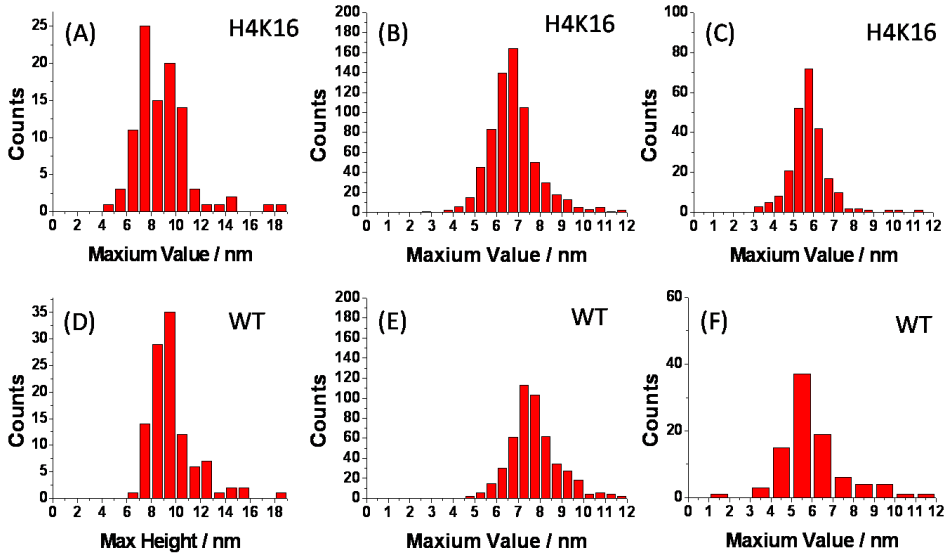

Figure 2-4 Height Analysis of the folded array with the H4K16 and WT histones on the 601 template.

The maximum height is measured from Gwyddion (www.gwyddion.net) for each folded particle shows in Fig. 2.3, three independent experiments were illustrated here, the histogram of the max height analysis shows that no apparent difference between H4K16 (A) and WT nucleosome array(D), this is the same for (B) and (E), (C) and (F). Here the upper panel (A), (B), (C) is the folded Nucleosome Array with H4K16 , the bottom panel (D), (E), (F) is the folded Nucleosome Array with WT-H.O. ; (A), (B), and (C) correspond to three independent experiments.

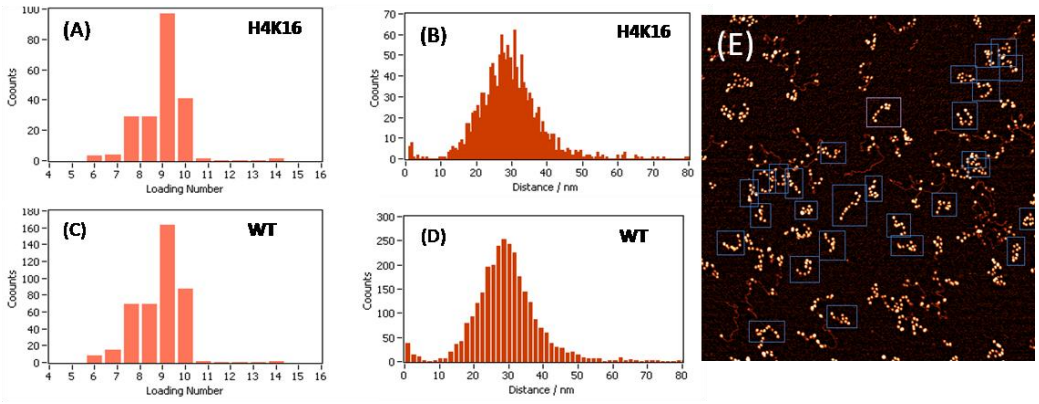

Figure 2-5 AFM image analyses of Saturated H4K16 and WT nucleosome array on the 601 template.

AFM image analysis shows that H4K16 and WT nucleosome array has the same loading behavior in the 601 repeat sequence. The distribution of the loading number for both centered at 9 for H4K16 (Ac) nucleosome array (A) and WT nucleosome array (C); and inter-nucleosome distance distribution shows the same peak centered at 30 nm for H4K16 (Ac) nucleosome array (B), and the WT-MMTV nucleosome array (D). (E) Shows the picked molecule in the AFM image.

2.3.3 Derive the Partial Specific Volume Information from the Sedimentation Coefficient Data:

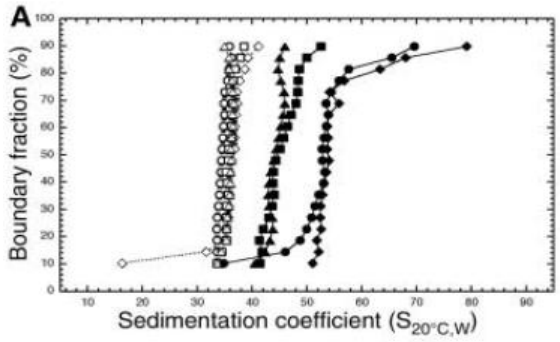

Figure 2-6 Sedimentation coefficient distribution of Nucleosome array in the absence or presence of 1.0 mM Mg^{2+}.

The figure shows the sedimentation coefficients of the nucleosome array on 177-12-601 templates reconstituted with WT: (diamonds), H4K16 (Ac): (Squares) and H4-tailess: (triangles) in the absence (open) or presence (close) of 1 mM Mg2+. S is corrected to water at 20 degree. The data is taken from [10]

AFM measurement didn't show apparent difference between compacted nucleosome array assembled with H4-K16 and WT histones. However, from the analytical ultracentrifugation experiment[10](Figure 2-6). Significant difference between the H4-K16 and WT nucleosome array was observed: the H4-K16 array was obtained with a sedimentation coefficient centered at 42.5s while WT with a measured sedimentation coefficient centered at 52s. Since the conclusion based on the result measured from AFM is different from the conclusion drawn from the sedimentation experiment, we think

that there are more parameters behind the condensed particle from the AFM results. Because AFM imaging shows that the compacted nucleosome array possesses a spherical structure; we treat the highly compacted nucleosome array as a smooth, compact spherical protein. Based on this assumption, we can derive the frictional coefficient of the spherical particle based on Stokes equation see Equation 2.4[110]. With the assumption that the sedimentation is measured in water at 20°C, the equation for partial specific volume can be obtained as Equation 2.2[110].

The following derivation for the final equation 2.6 is rewritten from reference [110] for consistence throughout the thesis. Based on the Stokes equation (Equation 2-1), the friction coefficient f_0 for a compact spherical particle can be obtained.

$$f_0 = 6\pi\eta R_0 \qquad 2\text{-}1$$

f_0: Frictional coefficient of the spherical particle

η: Viscosity of the solution

R_0: Radius of the sphere:

From the well known Svedberg equations:

$$s = \frac{u}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{N_A f} = \frac{MD(1-\bar{v}\rho)}{RT} \qquad 2\text{-}2$$

s: Sedimentation coefficient

u: Sedimentation speed

ω : Angular speed of the rotor

r : Distance from the sample to the center, which is keep increasing while sedimentation

N_A : Avagardoro constant

f : Friction coefficient

$\bar{v}$: Partial specific volume of the sample

ρ : Density of the solvent

M : Molecular weight of the Sample

D : Diffusion coefficient

R : Gas constant

The diffusion constant can be expressed as:

$$D = RT / N_A f \qquad 2\text{-}3$$

For smooth and compact spherical proteins:

$$R_0 = \left(\frac{3M\bar{v}}{4\pi N_A} \right)^{1/3} \qquad 2\text{-}4$$

Put equations 2-3 and 2-4 into 2-2, the sedimentation coefficient for a sphere sample can be obtained[110]:

$$s_{sphere} = \frac{M(1-\bar{v}\rho)}{N_A 6\pi\eta \left(\frac{3M\bar{v}}{4\pi N_A} \right)^{1/3}} \qquad 2\text{-}5$$

Correct the sedimentation coefficient to water at 20°C, we get:

$$s_{sphere} = 0.012 \frac{M^{2/3}(1-\bar{v}\rho)}{\bar{v}^{1/3}} \qquad 2\text{-}6$$

For 177-12-601 sequence, the molecular weight for both the H4K16 and H4-tailess and WT molecular weight were calculated as following and listed in table:

Table 2-1 Molecular Weight of each Histone for Sedimentation Calculation

	WT Xenopus Histone	H4 Tailless
H2A	13421.51	13421.51
H2B	13710.78	13710.78
H3	15044.54	15044.54
H4	11236.12	9389.992
12 octamer	1281910.8	1237603.728
DNA M.W.	1312508	1312508
Total M.W.	2594418.8	2550111.728

Assuming that the H4K16 has the same molecular weight as WT histones, by substituting the values for the M : molecular weight, ρ : Density of the solvent (here we use 1 g/ml), and the mid-point sedimentation coefficient read from figure 2-6 value for calculation which are $S_{WT} = 52.5\ s$ and $S_{H4K16} = 45\ s$ [10]. By solving the equation, we get: $\bar{v}_{WT} = 0.786\ ml/g$, $\bar{v}_{H4K16} = 0.815\ ml/g$, and $\bar{v}_{H4tailess} = 0.812\ ml/g$. The partial specific volume between H4K16 (Ac)-12-601(177) and WT-12-601(177) in the folding buffer presented here shows a difference of 5%. Since the nucleosome array is folded into a confined spherical particle structure based on the AFM

experiment, the expected radius for the spherical particle could be obtained from the measured partial specific volume based on equation 2-6[110]. The result of the expected radius of the compacted nucleosome fiber was listed in the Table 2-2. The calculated radius based on the sedimentation coefficient is 9.47 nm for the H4K16 (Ac), 9.33 nm for the WT, and 9.40nm for the tailless nucleosome array. Note that the maximum height measured by AFM is around 8 nm, which is smaller compared to16nm, the diameter of the spherical particle. This is partially due to the smaller template used in the experiment, and also can be explained by the fact that the sample is spread and deformed on the mica surface. In order to avoid the error coming from the sample immobilization process, we also acquired the AFM imaging of H4K16 (Ac) nucleosome array on the folding buffer, and no geometrical properties differences were observed between imaging operated in air and in buffer (Data not shown). From the radius data, the difference between H4K16 (Ac) and WT folded nucleosome array reduced to 1.2% based on the one dimensional geometrical feature, which is almost barely discernible in the AFM imaging measurement.

Table 2-2 Geometrical Parameters for Compact Nucleosome Array

	$\bar{v}\ ml/g$	M	$R\ nm$	difference compare to WT
H4K16	0.814	2594418.8	9.43	0.012
WT	0.786	2594418.8	9.32	0
tailess	0.813	2550111.7	9.37	0.01

2.3.4 A Proper Loading of Histone Octamer is Required for the Nucleosome Array Condensation under Mg^{2+}

Based on the previous section's result on the H4K16 (Ac) and WT nucleosome array, the single acetylation on the histone tail H4 –k16 did not show observable affects on the folded structure based on the AFM measurements, but from the analytical ultracentrifuge sedimentation coefficient measurements, a noticeable difference is easily identified. The problem coming from the partial specific volume of the sample is not solved, and a small change as low as 1% in the sample radius could induce a big difference in the sedimentation coefficient.

From the nucleosome array reconstituted on the MMTV and 601 sequence, we noticed that the loading of the nucleosome exerts more influence on the nucleosome array folding process and this has also been proved by analytical ultracentrifugation experiments[111]. In this section, we further studied the

folding behavior of nucleosome array under different loadings along the concatenated templates by AFM. The DNA template used for the experiment is from the single digestion of plasmid pMP17 at EcoR I (2) site. The plasmid is 5675 bp at full length and contains a 17-177 601 tandem repeat region. The histone octamers are first saturated in the 601 region and a couple other locations along the vector sequence section as shown in Figure 2-7. With Mg^{2+} presented in the system, the 601 region is initially folded into a compacted structure like a spherical particle as pointed out in the previous sections; the none-601 region with low loading of nucleosome is hanging around the compact particle and didn't involved in the compaction at all. With increased loading, the whole nucleosome array is folded into a compacted structure. Despite the huge effects on the compaction of nucleosome array by nucleosome loading, the 2D geometrical features of the nucleosome array in the HEPES buffer also shows a dramatic change with increased loadings. An identifiable extended 'beads-on-string' structure at low loadings was observed. When the array is saturated, 'beads-on-string' structure can still be observed in Fig 2-7B. However, extra free histones in the solution bound to the fiber and induced a twisted structure.

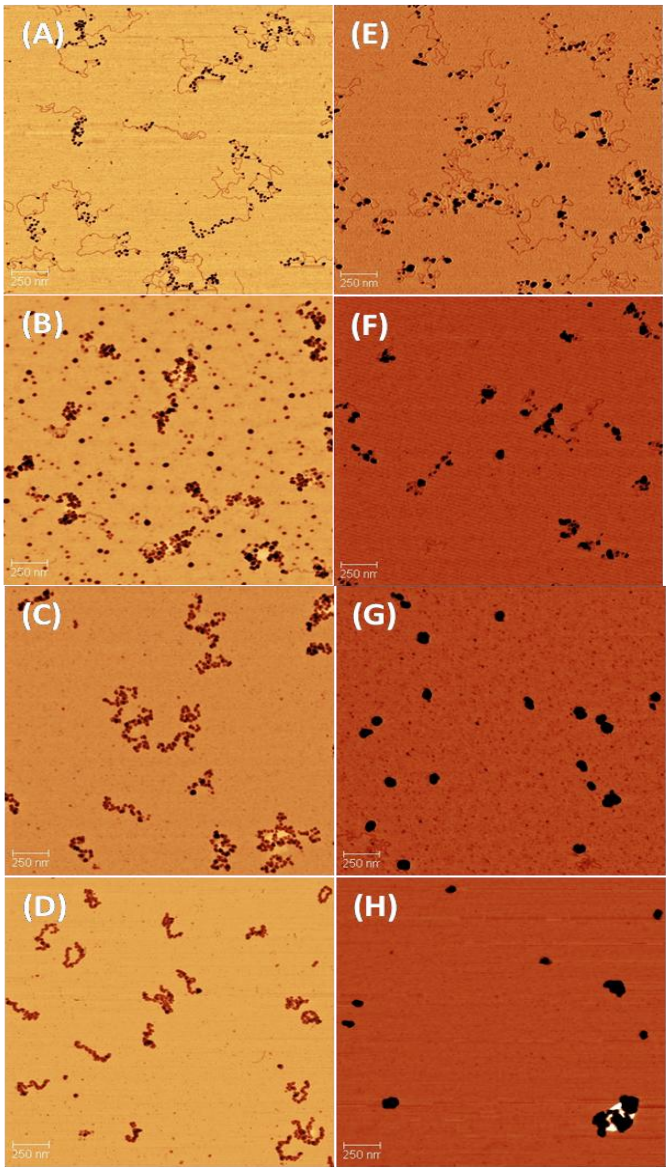

Figure 2-7 Increasing loading of histone in the concatenated sequence AFM images of the nucleosome array reconstituted on the template III, the sample were reconstituted under an increase of loading of histone octamers from (A) to (D); here (A), (B), (C), and (D) are in HEPES buffer; and (E), (F), (G) and (H) are in Folding buffer with 1 mM Mg^{2+}.

2.4 Conclusions

In conclusion, we studied the folding behavior of nucleosome array reconstituted with wild type (WT) and histone H4 acetylated at lysine 16 (H4K16 (Ac)). From the Atomic Force Microscopy (AFM) measurements, both WT and H4K16 (Ac) nucleosome arrays folded into compact sphere particle, and no apparent difference has been observed from these two particles. By assuming the folded particle as a compact spherical structure, we derived the structural parameters of these particles based on the sedimentation coefficients. Only 1.2% percent differences on radius between H4K16 (Ac) and WT nucleosome array were obtained, which is only 0.13 nm in radius. It is reasonable that we cannot see the differences from AFM measurement. Additionally, we found that a proper loading along the DNA template is required for maximal compaction under Mg^{2+}. The folding behavior along a concatenated sequence with uniform spaced 601 tandem repeat shows that 601 region saturated with nucleosomes can fold into compact structure independent of the free DNA on the non-601 region, a proper length of linker DNA is required for Mg^{2+} induced compaction.

Chapter 3

3 Construction and Structure Study of the Artificial Nucleosome Array

3.1 Experimental Design for Random Linker Length Library

3.1.1 Introduction

In order to get the transcriptional machinery into the gene target, a process called gene activation, in which a series of processes including unfolding the chromatin and unwrapping the DNA from the histone octamer is required. Therefore, studying the structural change of the chromatin is significant to understand the epigenetic and genetic control of the gene expression. However, due to the complexity of the nucleus, the huge size, and dynamics of the chromatin, the knowledge of the chromatin folding process is still very limited. Many factors are directly involved in the folding process of the chromatin. For example, the presentation of multivalent cations which can induce the chromatin condensation[112, 113]; a lower pH can induce aggregation of chromatin [114]. Nuclear protein such as MECP2 [115] and HP1[116], histone tails [44, 95, 97, 98, 117, 118] and histone tail modifications on the N-terminal [10, 119], linker histones[44, 119-122], and linker DNA [123, 124] have been shown involved in the regulation of the folding process of the ’30 nm’ chromatin fiber.

The efforts to elucidate the structure of the higher-order chromatin structure have never ceased, and remarkable progresses have been made in the past decades by taking advantage of new experimental designs, for instance,

developing more consistent DNA templates based on 601[61] sequence, and techniques such as high resolution Electron Microscopy(EM) and Atomic Force Microscopy (AFM). Currently, two models of the ’30 nm’ fiber has been revealed: the 'one-start' model including the solenoid and the interdigitated compaction [6, 125-128] and the 'two-start' model including the twisted-ribbon model and the crossed-linker model[63, 64, 129, 130]. For the ‘one-start’ model, the 30 nm fiber is constructed from the ‘one-start’ helices, the nucleosome is directly connected to its neighbor (n±1), and linker DNA is combined inside of the 30 nm fiber[57]. For the ‘two-start’ model, the ’30 nm fiber’ is folded from a zig-zag arrangement of nucleosomes, and with each nucleosome closing to the ‘n±2’ nucleosomes in the folded state.

The nucleosome templates used for the nucleosome array studies has been improved from the tandem repeat sequence based on the 5s RNA gene of sea urchin[131] to the 601[61] in recent years. The research on the folding of the nucleosome array was mostly based on short templates. Until recently, artificial nucleosome arrays with long and different linker lengths have been constructed. Defined ’30 nm’ construction has been obtained from uniform spaced nucleosome array *in vitro* [6, 132]. Moreover, since the spacing between each nucleosome is irregular in the natural chromatin, the linker lengths vary from species and tissues[39, 133, 134]. Nucleosome array with a variation of ±2 bp of linker DNA length has been tested on a 12 tandem repeats sequence[135]. Although there are all these efforts to understand the

mechanism of the folding process of the nucleosome array in nature, and the '30 nm' structure is still remaining in enigma.

In this section we address the need to mimic the natural system, which possesses irregular spacing of the nucleosome array. We introduced a systematic method to construct long and randomly spaced nucleosome array templates for improved understanding a series of problems in chromatin condensation. Such as the inter-nucleosome interactions, and the 2D geometrical structure of extended nucleosome arrays. We constructed sequence monomers based on the 601 sequence with different Linker DNA lengths: 30 bp, 40bp, 50bp and 60 bp. A linker length obeying 10n (n is a integer) has been widely used for nucleosome array study, while a recent study shows that linker length of yeast favored a 10n+5 periodicity[136]. For that reason, we constructed our monomer libraries with the left and right linker lengths varying in a step of 5 bp. The linker length between each nucleosome in the final construction obeys a form of 5n, while the average linker length keeps the same. For example in 30 bp monomer library: left linker length varies from 5 bp, 10 bp, 15 bp, 20 bp, 25 bp, and the right linker length vary from 25 bp to 5 bp. We successfully constructed uniformly and randomly spaced nucleosome array with the average linker DNA length of 30 bp, 40 bp, 50 bp, and 60 bp. This method could also be applied further to constructing any nucleosome array with distinct spacing.

3.1.2 Materials and methods

3.1.2.1 Core 601 Sequence Construction:

The core sequence of 601 is:

CACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGC

GGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCT

ACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCA

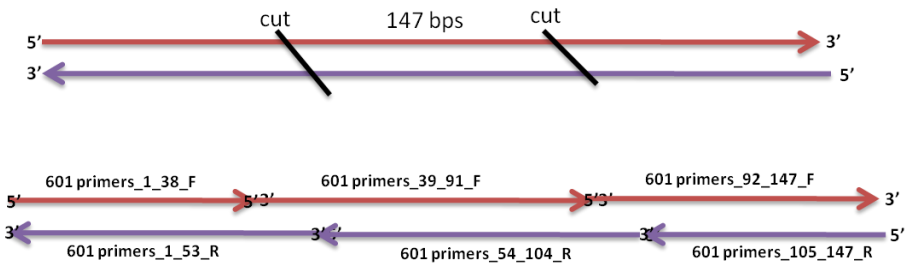

Figure 3-1 Scheme of the primers used for the 601 core sequence construction

Six primers were ordered from IDT (Integrated DNA Technologies). The sequences were first treated with kinase and heated up to 90°C to deactivate the enzyme. The six primers were then mixed at equal molarity and heated up to 90°C and annealed by slowly cooling down to room temperature. The mixture was then ligated with T4 DNA ligase (New England Biolabs) for 2 hours. The ligated products were then run through on a 3% Agarose gel, the 147bp fragment was cut out, purified, and used as the template for the following PCR experiment. 147bp 601 core sequence was PCR-amplified, gel

purified, and cloned into TOPO4-PCR vector. Plasmids were sequenced by the DNA core Lab (life Science ASU)

Table 3-1 Sequence of the Primers of the 601 Construction

601 primers_1_38_F	CACAGGATGTATATATCTGACACGTGCCTGGAGACTAG
601 primers_39_91_F	GGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTA
601 primers_92_147_F	AGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCA
601 primers_1_53_R	CAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTG
601 primers_54_104_R	CTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGC
601 primers_105_147_R	TGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAG

3.1.2.2 Random Linker Length Library:

Average Linker lengths with a range from 30 bp to 60 bp were constructed. For 30 bp linker length, the left linker DNA length was designed as 25, 20, 15, 10 and 5 bp and the right linker DNA length was designed as 5, 10, 15, 20 and 25 bp (Figure 3-2). This design gave a 30 bp total linker length for each monomer sequence. Ava I restriction sites were designed at both ends of the linker, and a non-palindromic sequence 'CCCGAG' pattern of Ava I was chosen as a directional ligation motif. For the 30 bp random library, the shortest linker length can be obtained was 10 bp and the longest linker length can be obtained was 50 bp. The Primers used for this construction are listed in appendix A, table A-1. The same strategy was used for the 40bp, 50 bp and 60 bp random linker length library construction and the primers can be found in appendix A, from table A-2 to 4. Consequently the 40 bp linker

length library has a left linker length range from 35 to 5 bp with an increment of 5 bp, and constitutes total 7 monomers; 50 bp linker length library has 9 monomers and 60 bp linker length library constitutes 11 monomers library. All the constructions were completed by PCR experiments with proper templates and primers and cloned into TOPO4-PCR vector and confirmed by sequencing. The template used for the 40bp linker length construction was based on the 30 bp linker length library; 50bp linker length construction was based on the 40 bp linker length library, comparably for the 60 bp length construction was based on the 50 bp. The sequences for all the monomer libraries are listed in appendix A-5. The sequence for the linker region was designed to avoid common restriction sites. Monomers with linker length from 31 to 39 were also constructed based on PCR experiments. The primers for this purpose are listed in appendix B table B-1, and the final sequences are listed in appendix B table B-2.

3.1.2.3 Long DNA Templates Construction:

Monomers with different linker lengths were digested with Ava I from the plasmids. Short UV light irradiation was avoided to give better ligation efficiency in the following experiments. The isolated monomers were mixed in an equal molar ratio to give a final DNA concentration at 100 ng/uL. 4 uL T4 DNA ligase (Invitrogen) were used for a fast ligation experiment. The mixture was ligated at 16°C for 30 mins, and then purified on a 1% agarose

gel. Four different fragments in different length ranges were isolated and purified with DNA gel purification kit (Promega). The purified long DNA fragments were then mixed with TOPO4-PCR vector with dephosphated Ava I sticky ends and ligated in room temperature for 30 mins. The ligation products were dried with vacuum centrifuge and 100 uL 70% percent ethanol was added to the dry product. Next, a strong vortex was used to get the salt to dissolve into the water; and the mixture was then centrifuged for 30mins at 16000g. The liquid was decanted and the ligated DNA product was left at the bottom of the tube, 25 ul DH10B competent *E coli* bacteria was added into each tube, following with gentle vortex. A transformation protocol was then applied to these samples.

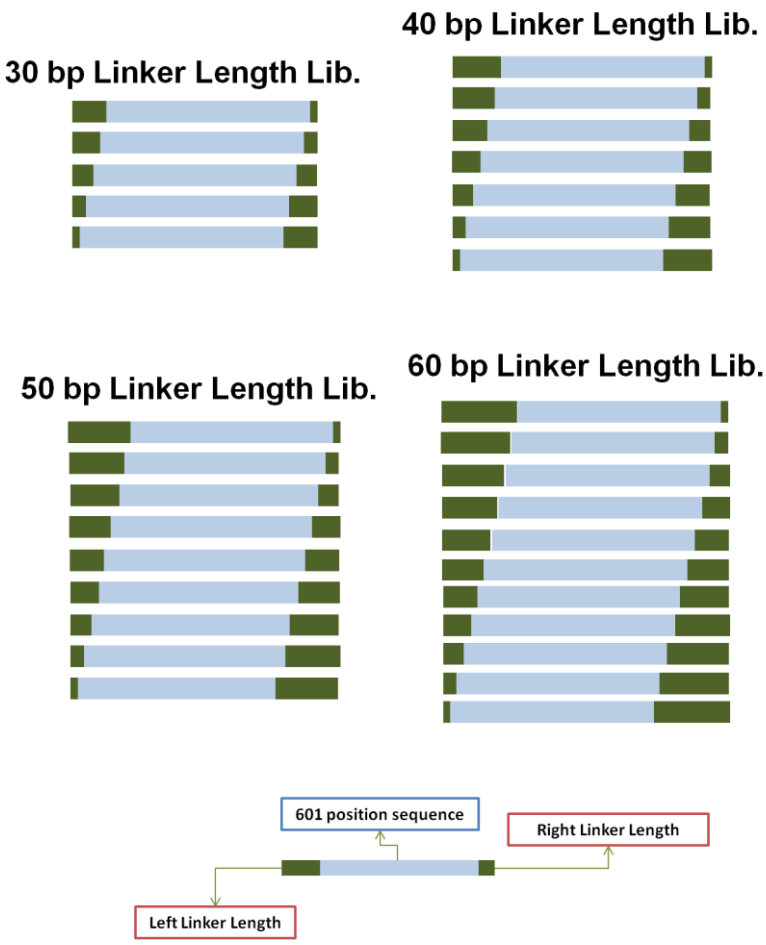

Figure 3-2 Scheme of the monomer library for random linker length Library

Monomer libraries with average linker lengths of 30 bp, 40 bp, 50 bp, and 60 bp were shown here. The light blue color represents the 601 region, and the dark green color represents the linker DNA length. For a 30 bp linker length library, 5 monomers were constructed; likewise, 7, 9 and 11 monomers were constructed for the other linker length libraries respectively.

3.1.2.4 Colony Screen for Long Repeats:

The DH10B *E coli* containing plasmids constructed from the TOPO4-PCR vector and repeats insert were selected by kanamycin agar plate, a colony screening protocol was used to select *E coli* colonies containing longer repeats. Clones with longer plasmids were amplified at a small scale and plasmids were isolated. The selected plasmids were tested by digesting with EcoR I and Ava I. The selected clones were further sequenced and confirmed with M13 forward and reverse primers at the two ends of the insert.

3.1.2.5 Advanced Vector Design For long Repeats Construction:

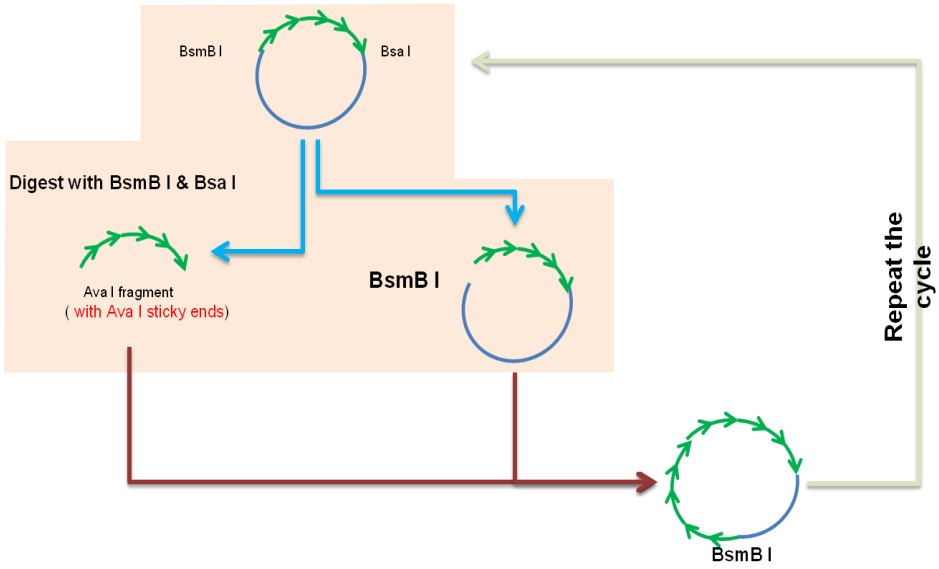

Figure 3-3 Scheme for advanced vector design for the long repeat construction

A model for repeating addition of the 601 tandem repeats, the vector is flanked by BsmB I and Bsa I. Single digestion of the plasmid with BsmB I produced Ava I sticky ends, and double digestion with BsmB I and Bsa I gave an insert with Ava I sticky ends, which can be ligated back to the vector.

Type IIs Restriction Enzyme recognition sites Bsmb I (-CGTCTCN'-) and Bsa I (-GGTCTCN'-) were introduced into the vector. Ava I recognition sequence pattern was designed in the 'N' region of the cutting sites to give Ava I sticky ends after being digested with Bsmb I and Bsa I. However, the Bsmb I and Bsa I cannot cut the Ava I sites in the middle along the insert. Hind III, EcoR V, Ava I and BamH I were also introduced into the new vector. Two primers

were ordered. BsmB I- Bsa I forward: 'AATTCGTCTCCCCGAGGAGACCGGATCCAGATATCAAGCTTG' and BsmB I- Bsa I Reverse: 'aattcaagcttgatatctggatccggtctcctcggggagacg', the primers were first treated with kinase and then annealed to room temperature, and used as inserts for the following experiments. TOPO4-PCR vector was digested with EcoR I and dephosphorylated with Alkaline Phosphatase Calf Intestinal (CIP) (New England Biolabs) to give two dephosphorylated EcoR I sticky ends and used as vector. The vectors and the insert were ligated together at 16°C overnight. The plasmid products were transformed into DB3.1 *E coli* competent bacteria. The new vectors with designed restriction sites were used for the following experiment.

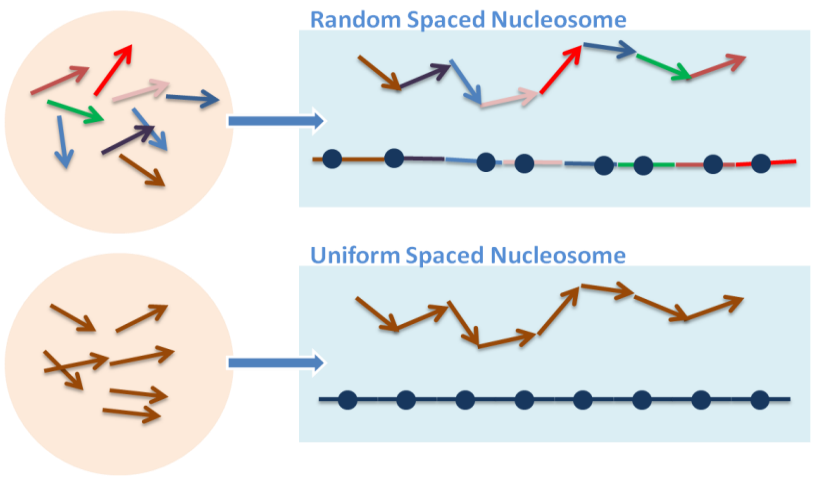

Figure 3-4 Schematic illustration of randomly spaced and uniformly spaced nucleosome array

3.1.3 Results and Discussions

3.1.3.1 Problems with Longer Repeat Inserts:

In order to construct a '30-nm' fiber structure, a decent length of a nucleosome array is required to show in AFM imaging, especially when they are highly folded. The *E coli* competent cells do not like longer repeat sequences. Several different engineered commercial competent cells, from SURE, Stbl2, Stlb3, Mach1, and DH10B were attempted in our Lab. The result showed that some of the competent *E coli* strains were not suitable for long repeat sequences. A recombination of plasmid was observed in our experiment. The second generation of the competent cell did not give exact same plasmid size; a deletion of repeat sequence was observed (Figure 3-5).

In the right gel image, the plasmids of 10 picked colonies were obtained and run through a 1% agarose gel to check the size. The gel result showed that the plasmids from the10 colonies possessed different sizes. Figure 3-6 shows another evidence for this problem. A long plasmid 159B3-2, containing around ~3000 bp 60 bp uniform construction give a strong smear after digestion with EcoR I or Sph I , which suggested that the insert lengths varied inside the plasmids. However, digestion of the plasmid with Ava I only gave two strong bands with fixed size. This suggested that the *E coli* Strain Mach1 only delete repeats from the 601repeat construction.

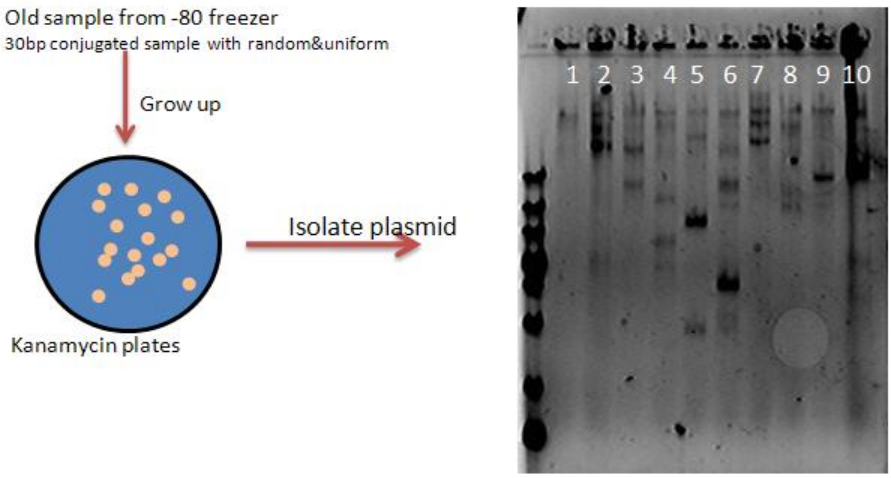

Figure 3-5 DNA recombination in *E coli* Mach1 Strain

DNA recombination problem for long repeat sequence happened in most *E coli* Strains. Here, the right 0.8% Agarose gel shows the plasmids isolated from 10 individual colonies.

Another problem came from the experiment itself, the ratio of longer sequence product is low among the self-ligated mixture. Several problems occurred here: First, when the length of the repeat sequences enlarged higher, the possibility of self ligation increased. The head and tail of the repeat sequence would self ligate together and inhibit combining into the vector. Second, the probability of bad sticky ends is higher for long repeat sequence compared to the short repeat. This is because the monomer library contains 5~10% of bad sticky end sequence caused by the enzyme digestion and DNA damage during the experiment. UV light, high salt, and high temperature in the DNA purification process can induce a damage to the sticky ends, and these bad sequences are accumulated at the end of the final monomer self ligation product. Third, a very long repeat sequence is difficult to ligate into the vector, or even to ligate with the vector, the probability of transforming large plasmid into *E coli* competent bacterial is fairly low.

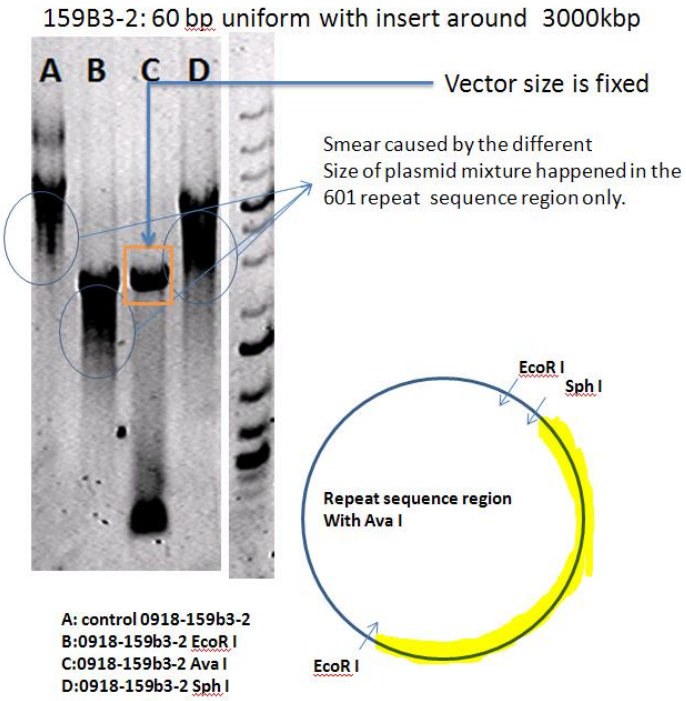

Figure 3-6 Recombination problem in the Mach1 *E coli* Strain

The plasmids were digested with EcoR I (lane B), Ava I (Lane C) and Sph I (lane D) and run through a 0.8% Agarose gel. Different sizes of the plasmid were shown in lane A as a smear. The fragments contained the inserts such as EcoR I (lane B) and Sph I (lane D) digestion shown a strong smear in the Agarose gel.

3.1.3.2 Advanced Linker Length Construction:

With the new designed routine for long repeat sequence construction, we could overcome several problems discussed in the previous section. By taking advantage of type IIs restriction endonuclease, which possesses

cutting sites are far away from their non-palindromic recognition sites, a method for repeat sequence construction was developed.

Figure 3-3 shows the detailed routine for the construction of longer repeat sequences, generally, Bsmb I and Bsa I were introduced to the new vector, the cutting sites of Bsmb I and Bsa I are at the same location of an Ava I restriction site. However, the recognition sites of Bsmb I and Bsa I were positioned left and right in several base pairs far away from Ava I site respectively. We started from a plasmid with an insert with two 601 repeat sequences, digesting the plasmid with Bsmb I was supposed to give a vector with Ava I sticky ends. Digesting the plasmid with Bsmb I and Bsa I was supposed to give an insert with flanked Ava I sticky ends. By combining the vectors and inserts together, we could get a new plasmid with 4 inserts, repeat this cycle, we could get 8 repeats, and then 16 repeats. This design could construct any kind of repeat sequence.

Figure 3-7 shows an example of 30 bp random linker length ligation to very long repeats

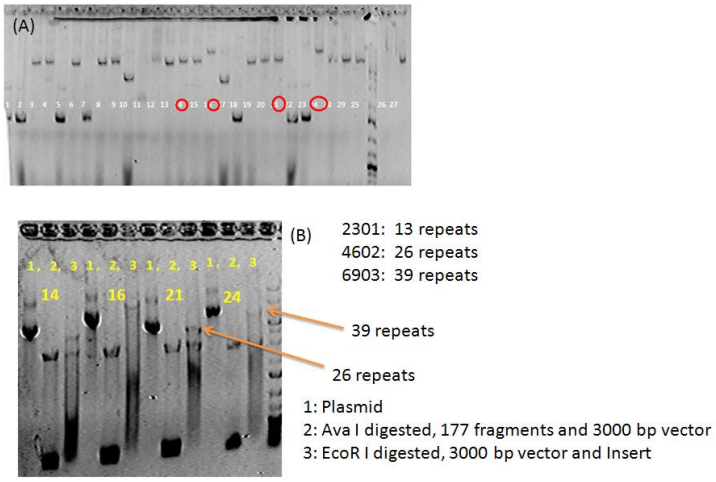

Figure 3-7 Longer sequence constructed from the short one

(A) Shows the colony screen of the constructed plasmids based from 13 repeats, most colonies have two copies of 13 inserts, and several have 3 copies of 13 inserts. (B) Endonuclease digestion confirms the insert is consisted of 601 repeat sequences by Ava I and EcoR I digestion. Clone 21 shows 26 repeats and clone 24 shows 39 repeats.

To order to verify that the method is suitable for very long repeat sequence construction, a construction of 1080-17-9 with 13 repeats of 30 bp random monomer library were chosen for the following experiments. The 1080-17-9 has been sequenced with M13 Forward and Reverse primers, the forward part gave a 20-10, 15-15, 10-20, 10-20, 25-5, 20-10 organization of linker length, the reverse part gave a 25-5, 15-15, and 25-5 linker length arrangement.

3.1.3.3 Construct Defined DNA Templates with 601 Positioning Sequence

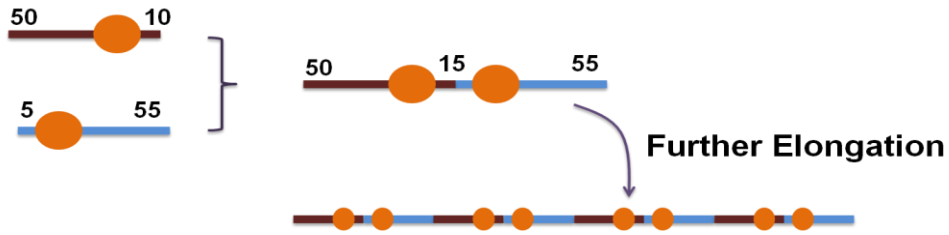

Figure 3-8 Scheme of defined nucleosome array organization construction The routine to construct a 110 bp-Nucleosome-15 bp is illustrated here, 50-10 and 5-55 monomers were chosen for this construction, a dimer was constructed first, and the dimer was further ligated to give the final construction containing longer repeats.

In order to study the role of linker DNA in the nucleosome array folding process, nucleosome array with constraint linker DNA pattern would be an interested object. Therefore, we also constructed a nucleosome array with a short and long linker DNA length pair based on the new method (see Figure 3-7 for the detailed routine). Two monomers, 50-10 and 5-55, were chosen for the construction. 50-10 monomer was ligated into the plasmid containing 15-55, and the dimer '50-601-15-601-55' was amplified, self-ligated, and further constructed. A sequence containing 3 copies of the dimers were obtained, which gives a structure with six 601 positions in a short-long fashion and we define this structure as 6-601 Histone octamer positioning behavior on this sequence were studied. In figure 3-9, the left part of the

graph shows the counting of the histone octamer positioned in the specifically defined 601 positions. Six 601 positions were shown clearly, and the binding intensity for each 601 location is approximately the same. However, interestingly, when the loading number is low at 1, a second site of the 601 location away from the DNA terminus is favored by the histone octamer. This observation is different from what we have observed before, which histone octamer favored a terminal position along the DNA templates[137]. Moreover, the inter-nucleosome distance was analyzed to further study the loading behavior of the histone octamer along this template. Since the two 601 positions are very close to each other with only 15 bp of linker DNA, it would be interesting to study the loading preferences for histone octamer [137]. From the inter-nucleosome distance histogram, (see Figure 3-10 A), a broad distribution of the inter-nucleosome distance was obtained, a peak located near 25 nm indicates a population of cooperative binding. However, peaks in 120 nm, 160 nm and 250 nm were also observed, these peaks correspond to the random locations of the 2nd nucleosome.

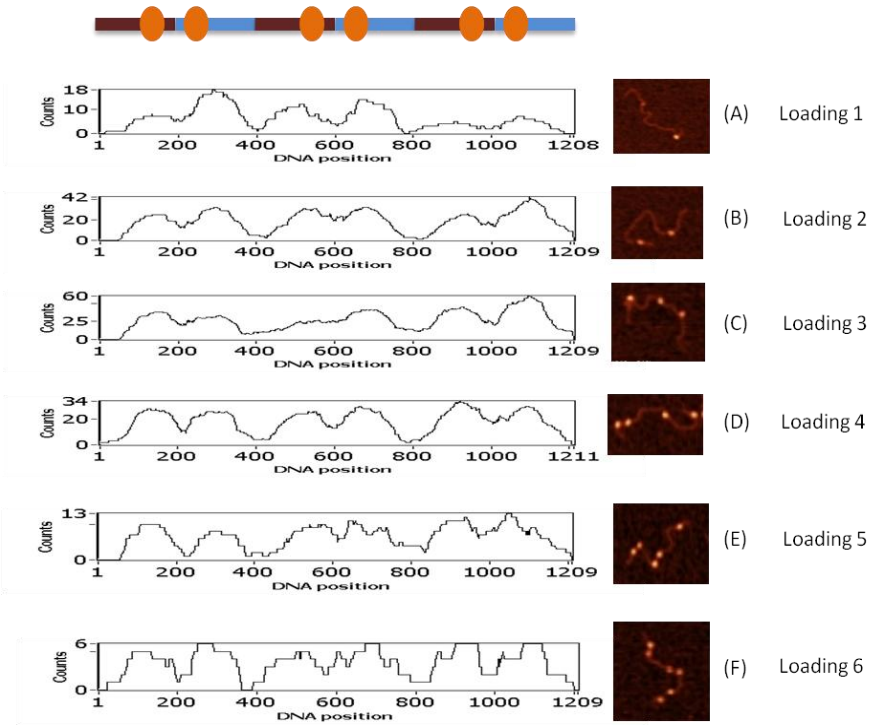

Figure 3-9 Nucleosome positioning behavior on the defined 601 templates

Here, the left plot shows the positioning ability at each 601 location from loading 1 to loading 6; and the right part of the figure shows AFM images of each individual nucleosome array with increasing loadings.

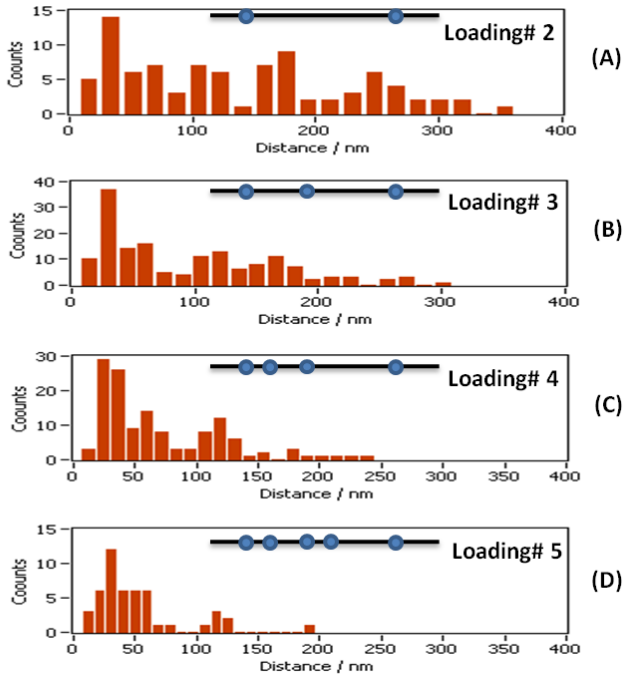

Figure 3-10 Inter-nucleosome distance distribution for the defined 601 template

Inter-Nucleosome distances of the nucleosome array under different loadings were analyzed here. The loading numbers are 2, 3, 4, and 5

3.1.4 Conclusions

In summary, we developed a system to construct tandem repeat sequences for nucleosome array structure study with a variable of linker DNA length obeying the form of 5n. Total of 42 monomers with varied linker lengths were constructed and cloned. 30 bp and 60 bp uniform and random constructions with varied repeat length were constructed. Additionally, we

extended our system to very long and defined repeat sequences construction by using the Type IIs restriction enzyme BsmB I and Bsa I. Using this method, we obtained 39 repeats of 177-601 sequence with defined sequence organization starting from a 13 repeats, and we constructed another tandem repeat sequence with defined organization of short and long linker DNA. We further test the nucleosome array loading behavior on the defined short and long linker DNA template. The result shows that the nucleosome positioning on the template is dominated by the sequence binding affinity mostly, and a cooperative way of loading histone octamers was observed. Furthermore, the positioning abilities of the six separate 601 locations are almost the same. However, when the loading is low, the nucleosome positioned more at the second 601 slot close to the DNA terminus compared to the terminal 601 position.

3.2 Randomly Spaced Compared to Uniformly Spaced Nucleosome Array

3.2.1 Introduction

The histone H1 depleted nucleosome filament shows an extended 'beads-on-string' extended structure in low salt from EM result[138]. This extended structure folds into a higher compacted '30-nm' fiber structure when cations present, such as 1-2 mM Mg^{2+} [139]and 50 mM NaCl[113]. Recent experiment shows that linker DNA played an important role in the compaction of the nucleosome array, for example, Linker DNA could destabilize the condensed chromatin by exerting a disruptive force on the

condensed structure[124] and different linker lengths induce different geometries of highly compacted nucleosome arrays[6].

The arrangement of the linker DNA in the compacted ’30 nm’ chromatin structure directly induced two major groups of modeling on the ’30-nm’ fiber structure: ‘one-start’ solenoid[6, 125-128] and ‘two-start’ zigzag model[63, 64, 129, 130]. In the solenoid model, the linker DNA merged inside the ’30-nm’ fiber interior region[57]; and the linker DNA’s length and variation are independent on the folding structure[58, 59].While for the zigzag ‘two-start’ model, the final diameter of the compact folded nucleosome fiber is strongly dependent on the linker DNA length. Since the model assumes a rigid and straight arrangement of linker DNA[5, 140] and is energetically more favorable. In addition, evidences have shown that linker DNA is bended in the higher order structure[124].

Additionally, the nucleosome repeat lengths have been shown related to the helical twist of DNA[133]. The length and flexibility of the linker DNA could be regulated by several factors for example, linker histones are one of the most important factors directly related to the nucleosome spacing, nucleosome spacing has been shown decreased when the histone H1 expression level is low[141]. with a repeat length decreased by a degree of ~15 bp, from ~189 to ~174 bp[142]. In addition, EM study shows that linker histones induce the linker DNA segments becoming juxtaposed ~8 nm and

also a zig-zag arrangement for higher order chromatin structure[121]. For the factors that affect linker DNA flexibility, ionic strength has been shown bending or kinking the linker DNA in a nucleosome dimer experiment[123], and histone tails can induce a bending of the linker DNA by electrostatic interaction[143]. In addition divalent ions induce a bending in the linker DNA, and leading sequential nucleosome interactions[135].

In vitro experiment focus on the linker DNA's effects has been done on regularly spaced nucleosome array with the introduction of linker histone at different linker length: 10 to 40 bp linker lengths give a 33 nm fiber and 50 to 70 bp linker lengths give a 44-nm wide fiber[6]. The linker DNA lengths affect the final structure, which fits the 'two-start' model's hypothesis. However, the EM data is inconsistent with the 'two-start' helix models[6]. Also a varied linker length construction with variation of ±2 bp were studied by EM, and no observable difference has been shown by this variation[135]. Additionally, computer modeling on the folding between uniform linker and variability linker nucleosome array has been done and both generate fibers resembling the native structures[144].

For a better understanding of series problems in chromatin condensation process, for example, inter-nucleosome interactions, and 2D geometrical structure of extended nucleosome array. In this section, we studied the behaviors of the nucleosome array in various buffer environments with

samples reconstituted on different templates: 30 bp uniform and 30 bp random, and 60 bp uniform and 60 bp random. The variation of the linker DNA lengths in our construction could go up to 110(55+55) bp and down to 10(5+5) bp but with a constrained average linker DNA length at 30 bp or 60 bp. We proposed that the study on the comparison between the uniformed and randomly spaced nucleosome array could give more understandings on the mechanism of the nucleosome folding. Based on the study of the relationship between the linker DNA variation and chromatin higher order structure, we demonstrated that the length of the linker DNA played an important role in the Mg^{2+} induced chromatin condensation.

3.2.2 Materials and Methods

3.2.2.1 Materials

The following plasmids: clone 0918-159B3-2 with ~3000 bp 601-207 60 bp uniform insert; clone 1080-29-94 with ~3000 bp 601-207 60 bp random insert; clone 0874-161G with ~8000 bp 601-177 30 bp uniform insert; and clone 1080-25-21 with ~3500 bp 601-177 30 bp random insert, were chose for the following experiments. The different templates with the 601 regions were liberated by EcoR I digestion. The DNA fragments were purified by phenol-chloroform extraction and ethanol precipitation. Buffer DNA of 147 bp was PCR amplified in the Puc19 with two primers near the vector region.

3.2.2.2 Nucleosome Reconstitution:

Nucleosome reconstitution was obtained by mixing the buffer DNA/sample DNA at weight ratio 1:1 and with a final DNA concentration at 100 ng/ul, histone octamer, 1x TE (10 mM Tris-HCl pH 7.5, 1 mM Na2EDTA), 1 mM DTT and 1 M NaCl. The following dialysis process has been described in chapter 2.

3.2.2.3 Imaging Analysis:

Atomic force microscopy images were first analyzed with Gwiddion (www.gwiddion.net). The noise filtered images were stored as 'tif' type, analyzed by custom written Matlab programs. Identified nucleosome arrays were picked for further in depth analysis. The contours of the molecule were traced with mouse, and the center positions of nucleosomes were marked as well as the DNA entering and exiting locations. Center to center distances of nucleosomes were calculated. DNA entering and exiting angles were calculated for each non-terminal nucleosome. Complete information of the Matlab programs are given in Chapter 6.

3.2.3 Results and Discussions

In order to study the structure and folding process of nucleosome array with uniform and random linker length under Mg^{2+}, the nucleosome arrays were reconstituted with the different templates by using step salt dialysis as described [113]. Buffer DNA with a length of147 bp was added into the

reconstitution process to avoid aggregation of the histone octamers on the 601 templates. Based on this method, the nucleosome reconstitution is dominated by thermodynamics[61], and histone octamers bind specifically to the 601 region even under a very short linker length [145]. Several ratios of DNA and histone octamer were used for reconstitution, and atomic force microscopy (AFM) was applied to study the geometrical change of the uniformly and randomly spaced nucleosome array.

3.2.3.1 2-Dimentional Geometry Differences between Uniform and Random Linker Construction

30 bp uniformly and randomly, as well as 60 bp uniformly and randomly spaced nucleosome array were reconstituted and gently crosslinked with 0.5% glutaraldehyde in 10 mM HEPES buffer (pH 7.3) and deposited on the APTES mica. AFM images of these samples were then obtained. Identifiable molecules were chosen for the imaging analysis. In order to study the geometrical features of the nucleosome array in two dimensions on the mica surface, we analyzed the DNA entering and exiting angles (EEAs) as well as the inter-nucleosome contour distances (Figure 3-10). For every non-terminal nucleosome, three parameters were measured: the contour distance to the left nucleosome center, the contour distance to the right nucleosome center, and the DNA entering and exiting angles. Note that, in some case, the nucleosome was in close contact to its neighbors, and this induced the linker

DNA being hardly identifiable in the image, at this case, the DNA entering and exiting angles were replaced by the center to center angles, and the inter nucleosome contour distances were replaced by center-center distances as shown in Figure 3-10.

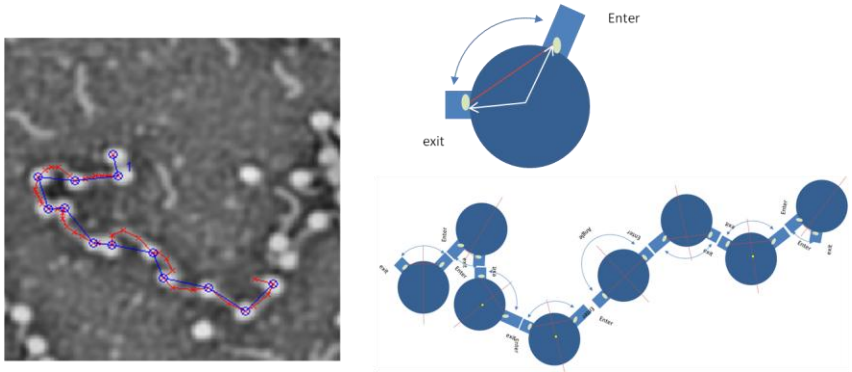

Figure 3-11 Illustration of the Matlab measurements along the nucleosome array from the AFM images

The left part of the graph shows an example of the nucleosome array analyzed by the MatLab program. Red line indicates the contour trace along the molecule, blue line indicates the center to center tracing, and blue circle indicates the center of each nucleosome. A cartoon illustrates the 2D geometrical features for the nucleosome array measured in the experiments (right part).

3.2.3.2 Inter-Nucleosome Distance

The inter-nucleosome distances distribution of 30 bp uniform and random constructions were both centered around 30 nm (Figure 3-12). The 30 nm inter nucleosome contour distance consists of the nucleosome diameter of 11

nm, linker DNA length of 10.2 nm (30 * 0.34 nm), and an additional 8.8 nm around 26 bp came from the unpeeled DNA around the histone core. DNA unwrapping from the core has been observed in our previous experiments[137], and this might be caused by the positive charged APTES mica. For 60 bp construction, a center of 40 nm was identified; this distance contains 8.6 nm of the DNA coming from unwrapping of the histone octamer as measured in the 30 bp construction. Additionally, from the histogram of the 60 bp random, despite the center peak at 44 nm, another peak centered at 22 nm and a wide tail in the 60 to 80 nm region were identified. This broad distribution corresponds to a random behavior of the nucleosome loading.

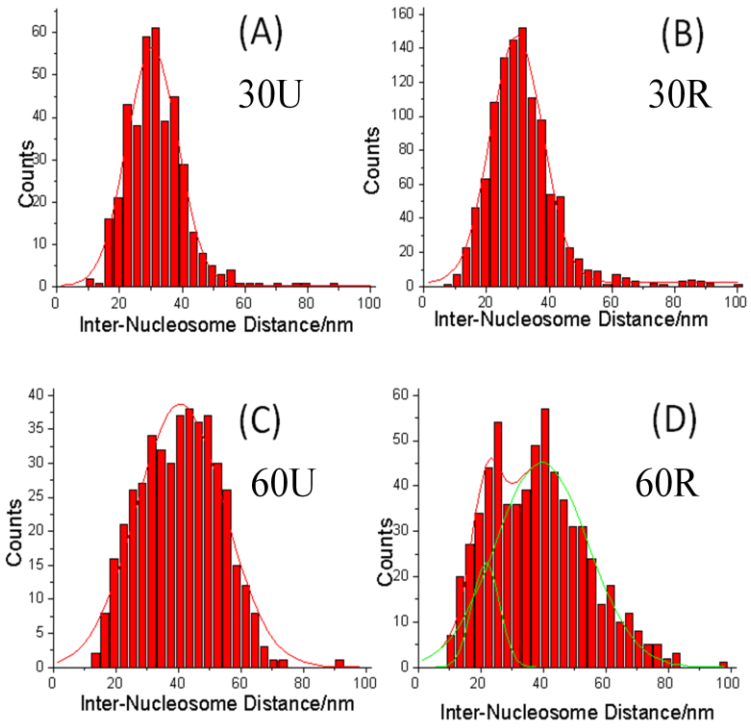

Figure 3-12 Histogram of the Inter-nucleosome distances for the 30 bp and 60 bp constructions

Inter-nucleosome distances were measured for all the samples, here shows histogram of the distributions. (A) 30 bp uniform centered at 30 nm, (B) 30 bp random centered at 30 nm, (C) 60 bp uniform centered at 40 nm, and (D) 60 bp random centered at 20nm and 40 nm.

3.2.3.3 DNA Entering and Exiting Angles

The DNA entering and exiting angles (EEAs) were also measured for each nucleosome. As the resolution of the AFM images were low caused by the tip broadening problem, in some case, the linker DNA was hardly identifiable.

DNA entering and exiting angles were approximate with the nucleosome center-center angles. Therefore, the EEAs measured here cannot precisely reflect the DNA wrapping behavior around the histone octamer. However, this parameter can reveal the nucleosome array's arrangement on the mica surface. The histograms of the EEAs distribution were plot for all the samples (Figure 3-12). For the 30 bp construction, 30bp random had a more intensive distribution at 120° and also at the small angle region compared to the 30 bp uniform. This observation represents a more twisted organization for 30 bp random construction. However, compared to the 60bp random construction, 60 bp uniform one showed a more twisted organization by showing a peak around 120°. Moreover, 60 bp linker constructions presented a more twisted organization on the mica surface compared to the 30 bp construction.

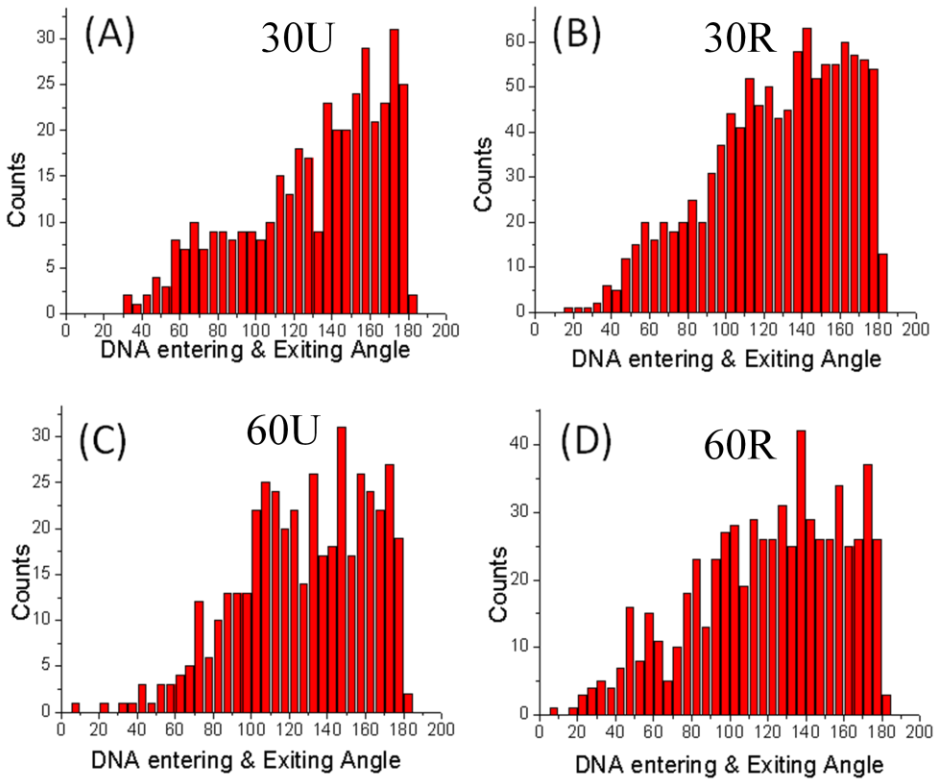

Figure 3-13 Histogram of the DNA Entering &Exiting Angles (EEA) on the 30 bp and 60 bp constructions

DNA entering and exiting angles were measured for all the samples, the distributions were plot here: (A) 30 bp uniform,(B) 30 bp random, (C)

3.2.3.4 Specific Unwrapping Patterns

In order to study the relationship between linker DNA lengths and the DNA entering and exiting angles, the data for each nucleosome were plot in a 2 dimensional way (Figure 3-13). Since for each nucleosome, the linker DNA lengths to the nucleosome‘s left and right neighbors were measured. In Here, we used the average linker length to reflect the neighbor nucleosomes’

effects on the DNA entering and exiting angles. In Figure 3-14, the x axel corresponds to the linker DNA length, and the y axel corresponds to the DNA entering and exiting angles. Compared to the randomly constructions, the uniformly constructed nucleosome arrays presented a more defined average linker length both for the 30 bp and 60 bp linker length construction. No direct relations were found between the linker DNA length and the DNA entering and exiting angles (EEA) in each independent sample. However, compared to the 30 bp construction, the 60 bp linker length construction had more wide distributions of both Linker DNA length and the EEAs. This distribution behavior suggests that longer linker DNA gives more flexibility on the structure of the nucleosome array. Additionally, the distributions between 30 bp randomly and uniformly spaced nucleosome array showed that the EEA of 30 bp random favored more at an angle of 140°, which corresponds to a more twisted structures. In contrast to the 30 bp construction, the 60 bp uniform construction gave lower EEAs population at 120°, and the 60 bp random had a broad distribution of EEAs, but with a center around 160°. We propose that the DNA entering and exiting angles can be affected by their neighbor nucleosomes, and the nucleosome-nucleosome interaction is dominated by the linker DNA length, which might further affect the compaction of the nucleosome array. A short linker length can induce a more rigid fiber structure compared to the long linker length, and uniform spaced nucleosome array possesses a more constrained organization.

Specific Patterns for the nucleosome array's EEAs were identified at 120°,140°, 160°, 180° with a step of 20°. Based on the crystal structure of mononucleosome, 20° corresponds to about1.57 nm (3.14*9nm *20/360) length of nucleosomal DNA, which is around ~4.6 bp. This number is close to the half of the DNA helix turn number, which is 10.17 bp per turn for the nucleosomal DNA[19]. It might suggest that, the unpeeling of the DNA away from the histone octamer is in a favorable step of 5 bp.

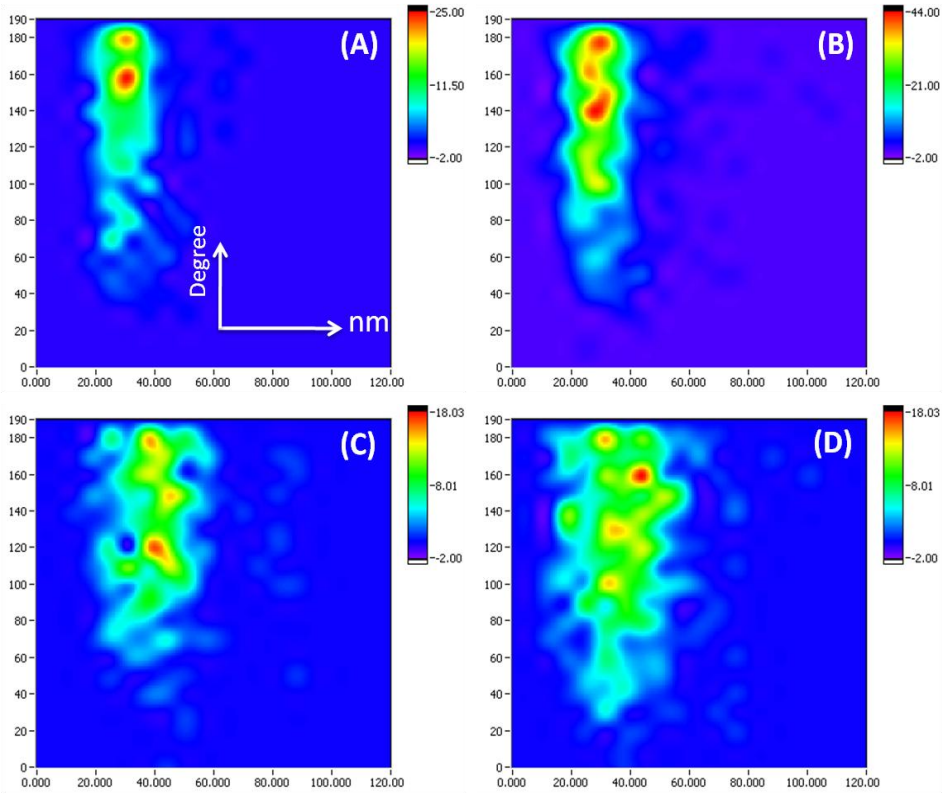

Figure 3-14 2D histogram of the linker DNA lengths and the EEAs for the 30 bp and 60 bp construction

2D histogram of the distribution of nucleosome linker lengths and the EEAs were plot here, the x axel corresponds to the linker length with unit of nm, and y axel corresponds to the angle measurement with unit of degree. (A) 30 bp uniform, (B) 30 bp random, (C) 60 bp uniform, and (D) 60 bp random.

3.2.3.5 Mg^{2+} Induced Compaction of Nucleosome Array

In the previous section, we have studied the effects of linker DNA's length on the geometrical features of extended nucleosome array at 10 mM HEPES

buffer at pH7.3 after deposition to the mica surface. To elucidate the role of the linker DNA in the chromatin folding process and also in the higher-ordered chromatin structure, the nucleosome array folding behavior under 1 mM Mg^{2+} folding buffer was studied[10]. Figure 3-15 shows the AFM results for the 30 bp uniform and random construction of the nucleosome array, the left images (A) and (C) were extended nucleosome array fiber at 10 mM HEPES buffer pH7.3. The right one (B) and (D) are the same sample in the folding buffer, which has additional 1 mM Mg^{2+}. Both of the 30 bp random and uniform nucleosome arrays were folded in the 1mM Mg^{2+}. However, the 30 bp random construction presented more highly condensed particles compared to the 30 bp uniform. This result can be explained from the 2D geometry measurements, since the 30 bp random structure was presented as a more twisted structure compared to the 30 bp uniform one in HEPES buffer.

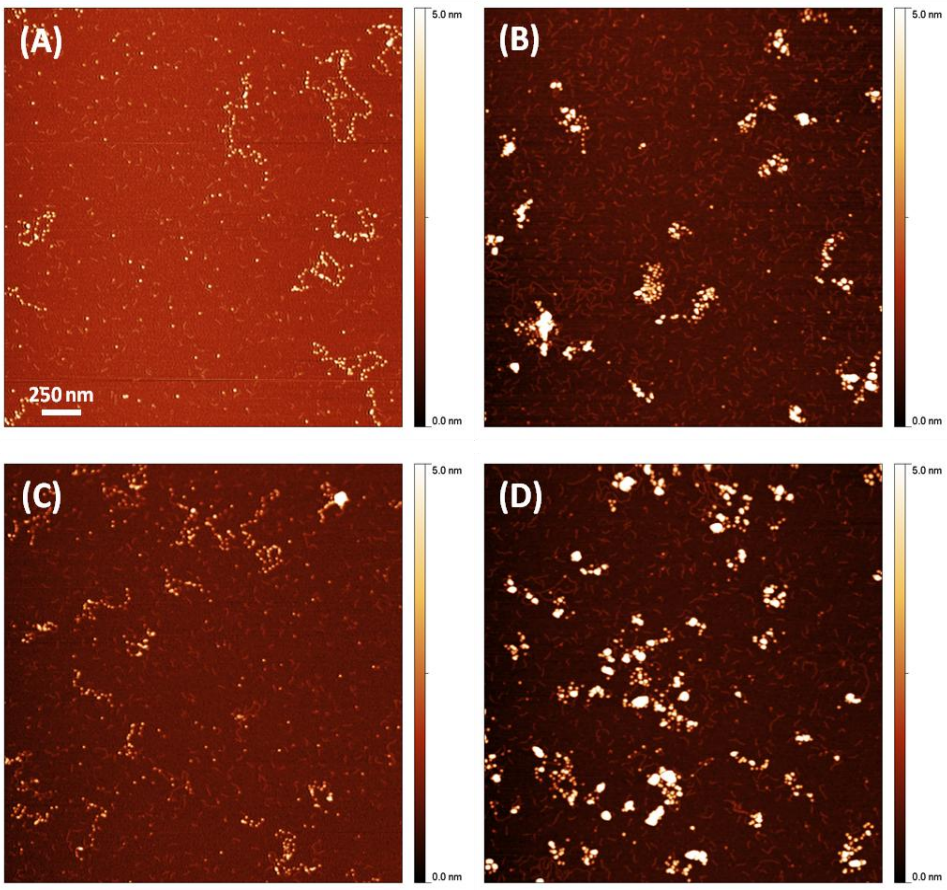

Figure 3-15 AFM images of 30 bp uniform and random nucleosome array AFM images of the nucleosome array immobilized on the mica surface in the absence or presence of 1 mM Mg^{2+} (folding buffer) in the 10 mM HEPES buffer. (A) 30 bp uniform in HEPES, (B) 30 bp uniform in the folding buffer, (C) 30 bp random in HEPES, and (D) 30 bp random in folding buffer. All the AFM Images were taken in air.

The same results were found in the 60 bp construction. 60 bp uniform constructions presented a more compact particle compared to the 60 bp

random (Figure 3-16 B (uniform) and D (random)). Both templates had a comparable length around 3000 bp. A more twisted structure of 60 bp uniform construction was revealed in the linker distance and EEAs distribution in the previous section.

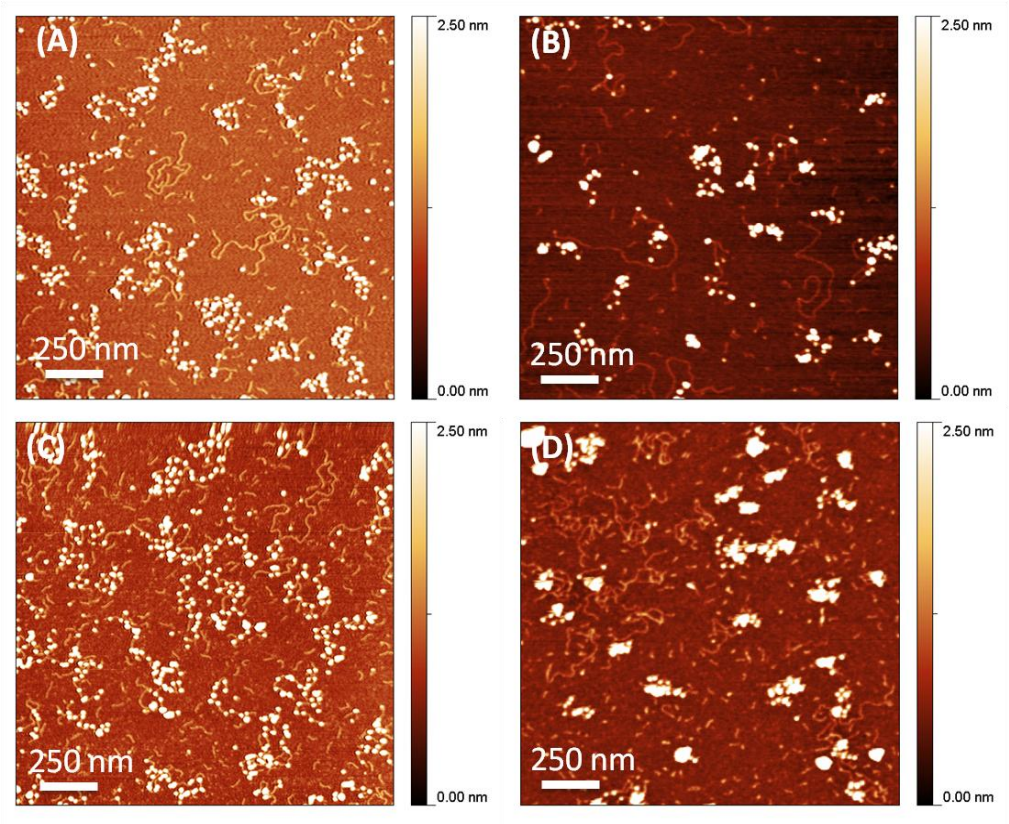

Figure 3-16 AFM images of 60 bp uniform and random nucleosome array AFM images of the nucleosome array immobilized on the mica surface in the absence or presence of 1 mM Mg^{2+} (folding buffer) in the 10 mM HEPES buffer, (A) 60 bp uniform in HEPES (B) 60 bp uniform in the folding buffer; (C) 60 bp random in HEPES, and (D) 60 bp random in folding buffer. All the AFM Images were taken in air.

3.2.4 Conclusions

In summary, based on the atomic force microscopy, we studied the folding behavior of uniformly and randomly constructed nucleosome array at single molecule level. From our study, the linker DNA length directly involved in the organization of extended nucleosome array in 10 mM HEPES buffer, and this organization further influenced the folding behavior of the nucleosome array. For example, a more twisted organization of the nucleosome array in the extended form could induce a more compact structure after folding under Mg^{2+}. However, results based on the AFM imaging alone is not enough to illustrate the compact particle structural properties. Quantitative Agarose gel electrophoresis or analytical ultracentrifugation is required for better characterization of the surface charge radius, as well as the folded particle's density, which are independent of the nucleosome array size.

Chapter 4

4 Compaction of Telomere Nucleosome Array by Telomeric Repeat Factor 2 (TRF-2)

4.1 Introduction

The telomere is defined as the region located at the both ends of each chromosome and was first observed by Hermann J. Müller[146]. Since the DNA in eukaryotic cells is not circular, in order to prevent the linear DNA from being recognized and processed as DNA damage, both ends of the DNA for each chromosome are protected by specialized telomere-specific nucleoprotein complexes defined as the Shelterin Complex[147]. Additionally, when the linear DNA replicates in each cycle, the ends of the DNA would be truncated off a certain number of base pairs, which serve as binding sites for the DNA polymerase[148]. For this reason, the telomere length would be decreased after every cycle of most somatic cell duplication, and the length of the telomere limits the life time of the somatic cell[149, 150]. However, for cancer and germ line cells, the length of their telomeres can be preserved to some extent by expressing an enzyme called telomerase[151], which adds specific guanine-rich DNA sequence (TTAGGG) pattern to the ends of telomeres.

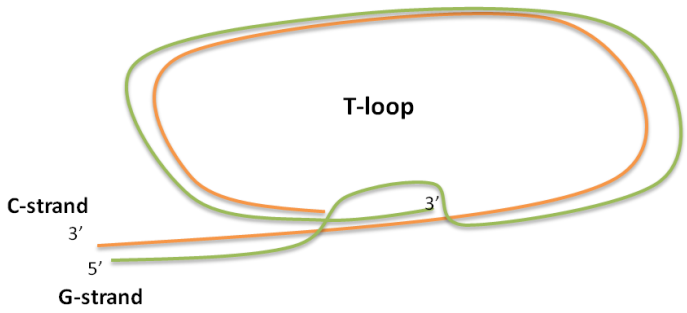

Figure 4-1 T-loop structure of the telomeric DNA

In higher eukaryotes like mammals, the DNA in the telomere region is also packed with core histones and linker histones, except that the chromatin in the telomere region is more compact than bulk chromatin and contains a short DNA repeat length~157 bp[152]. However, there is no evidence demonstrating a nucleosomal organization in the telomere of the lower eukaryotes[153]. Electron microscopy shows telomere chromatin fibers isolated from chicken erythrocyte and mouse lymphocyte nuclei as a 't-loop' structure with a 'bead-on-string' structure in the open form[11] . Similar to other regions of chromatin, the telomere chromatin obtains the same linker/core histone stoichiometry. However, the arrangement of core and linker histones in the telomere region is still not clear. Moreover, a depletion of H1 in mice could induce a increase the length of telomere[154]. Epigenetic modifications on H3 (trimethylated lysine 9) and H4 (trimethylated lysine 20) show a longer telomere compared to the wild-type[155] .

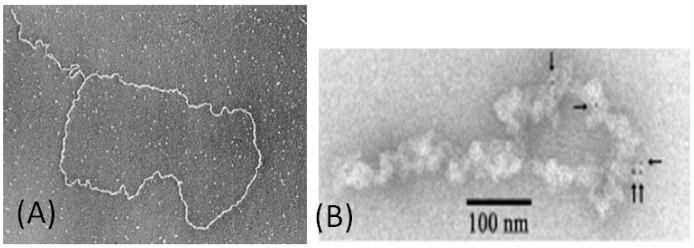

Figure 4-2 T-loop structure of telomere DNA (A), and Nucleosome array (B)

Image A is taken from [3]Image B is taken from [11]

After two years of the discovery of the telomere, chromosome adhesion and fusion with the ends was discovered by Barbara McClintock. She proposed that the telomere cap was required for the integrity of the chromosomes[156].Thus the stability of the telomere is directly related to the cell's life cycle[157]. The organization and stability of telomeres become extremely important in understanding the cell apoptosis and canceration process. EM study of telomeric DNA purified from human cells show large t-loops (Figure 4-2). The size of the t-loop is consistent with the telomeric origin [3]. The long 3' single-strand overhang of the telomeric sequence is merged into the double strand region by strand invasion and a G-quadruplex is formed (Figure 4-1). The protection of T-loop might play a critical role in the telomeric DNA maintenance and replication[3]. The shelterin complex that protects and adjusts the function of the telomere is composed of six subunits: TRF1, TRF2, TIN2, Rap1, TPP1 and POT1[147].

The TTAGGG repeat factor 2 (TRF2) is one of the key players in the shelterin complex that helps to stabilize and protect the t-loop structure of the telomere. TRF2 is required for maintaining the telomere termini structure, which is important for protection of telomeres from end-to-end fusions.[158] Additionally, TRF2's effect on the structure of the telomere has been studied by dominant negative expression and over expression of TRF2. Dominant negative expression TRF2 induces cellular senescence or apoptosis when mediated by p53/ATM[159], and also causes chromosome end-to-end fusions as well as shortening the G-strand overhang. While over expression of TRF2 induces an increase of the telomere shortening rate. However, this process does not accelerate the senescence speed; and the setpoint defined by the length of telomere decreased from 7 kbp to 4 kbp[160].

In vitro experiments demonstrate that TRF2 can induce the liner telomeric DNA to form into a T-loop structure [3]. TRF2 stabilizes the T-loop by specifically binding at the junction located at the duplex repeats and single-stranded overhang. At least six nucleotides overhang is necessary for the t-loop formation[161]. TRF2 molecules specifically bind to the telomeric DNA region as a dimer. A tetramer form of TRF2 was found at the root of the looped structure[162].

Both TRF2 and TRF1 protein binds to the double strand telomere DNA with a myb/SANT DNA binding motif[163]. The Myb-DNA binding domain is critically important for the interactions between TRF2 and telomeric chromatin. An *in vivo* experiment showed that temperature sensitive mutation of the binding domain induced a disruption in the sheltin complex and further caused the telomere dysfunction[164]. The Myb/SANT DNA binding domain is a helix-turn-helix DNA binding domain (see Figure 4-3). The C-terminal helix of the TRF2 interact within the DNA major groove, and the N-terminal interacts with the minor groove[163].

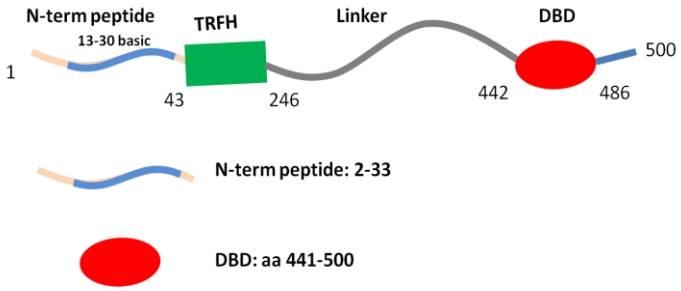

Figure 4-3 Schematic illustration of the amino acids' organization in the telomere repeat factor 2 (TRF2)

The interaction between TRF1 and telomeric chromatin has been observed by EM[11]. However, the detailed mechanism study of how the TRF1 and TRF2 interact with the telomeric chromatin has not been published yet. The structure of the telomere is strongly related to its function; an unwrapping of the shelterin complex is needed for DNA replication at the telomere region, telomerase function, telomere fusion, etc. Due to the special

arrangement of the nucleosome DNA and the sequence of the telomere, research has been done to demonstrate that the nucleosome structure affects the efficiency of binding factors. In this chapter, we focus on the structural characterization of the TRF2-DBD's interaction with the telomeric nucleosome array and how TRF1 and TRF2 can potentially affect the higher order structure of the telomeric nucleosome fiber. We characterized the TRF2-DBD and telomeric nucleosome complex with analytical agarose gel electrophoretic (AAGE) technique as well as the Atomic force microscopy.

4.2 Materials and Methods

4.2.1 Materials

The 3.5-kb pRST5 plasmid ([161] contains 96 TTAGGG DNA repeats was digested with Pvu II to cut out the 1-kb fragment containing the telomeric DNA and 2.5-kb non-telomeric DNA (Figure 4-5 A). Digesting the plasmid by SfaN I, a 2-kb fragment containing the telomeric DNA located at the center region and 1 kb and two ~220 bp fragments of non-telomeric DNA will be produced (Figure 4-5 A). The 208-12 5s rDNA was obtained by Hha I digestion of the p208-12 plasmid[113]. Telomeric DNA fragments were purified with the Qiagen QIAXII gel extraction kit. For atomic force microscopy (AFM) studies, 2-kb telomeric fragments was liberated by the SfaN I digestion, and the mixture was run through a 0.8% agarose gel. The 2Kbp fragment was excised, electroeluted and concentrated with an 8000-kDa filter (Amicon, Millipore). The sample was filtered again to remove

small amounts of agarose with a 2 um filter, and then followed with phenol-chloroform extraction and ethanol precipitation. Recombinant TRF2-DBD was provided from Dr. Fletcher's Lab, A complete protocol for the recombinant his6-tagged TRF2-DBD can be found[4].

4.2.2 Nucleosome Reconstitution in AFM Part

For AFM studies, 3 mg of 2 kb SfaN I fragments[161] were used for each reconstitution. The ratio between the DNA and chicken erythrocyte histone octamer were adjusted to 1.3 or 1 (histone/DNA, mass ratio). The histone octamer and DNA were mixed to achieve final concentrations of 0.1 mg/ml DNA and histone octamer, 1x TE (10 mM Tris-HCl pH 7.5, 1 mM Na2EDTA), 1 mM DTT and 1 M NaCl. The mixture was placed on ice for 30 min before stepwise salt dialysis against 0.8 M NaCl, 0.6 M NaCl and 0.15 M NaCl with 1x TE (pH 8.0) buffer for 1.5 hours each at room temperature. The sample was finally dialyzed against 1 mM Na2EDTA (pH 8.0) overnight at 4°C.Reconstituted nucleosome arrays were cross linked by dialysis against 0.1% Glutaraldehyde in 1 mM Na2EDTA (pH 8.0) for 6 hours at 4°C. All samples were first imaged under AFM to check the loading of histone octamer, Samples with proper loading were chosen for the following experiments[4].

4.2.3 Atomic Force Microscopy (AFM)

Histone-free DNA or nucleosomal arrays reconstituted with chicken erythrocyte histones were incubated with TRF2B in EMSA buffer lacking Mg^{2+}, at concentrations indicated in Figure 4-7, for 30 min at room temperature. Nucleosome arrays were added during last step to prevent dissociation in the 10x EMSA buffer. The resulting complexes were then crosslinked with 0.1% glutaradehyde for an additional 30 min, and diluted with 1x EMSA buffer (lacking Mg^{2+}) to 0.3 ng/uL (in DNA) for imaging. A 5 mL aliquot of each sample was deposited on APTES-mica (40), and incubated for 40 min, followed by rinsing with distilled water and drying with nitrogen. The imaging was carried out with a PicoPlus 2500+ (Molecular Imaging, 5500 AFM (N9410S) from Agilent) AFM equipped with a Si3N4 cantilever (AppNano SPM) with a spring constant ranging from 25-75 N/m. The resonance frequency was around 300 kHz; and the scan rate was 1.71 Hz. Gwyddion (www.gwyddion.net) and Chromatin Analysis 1.1.7 software were both used for image analysis.

4.3 Results and Discussion

4.3.1 Reconstitution of Telomeric Nucleosomal Array

Table 4-1 Analytical Agarose Gel Electrophoretic (AAGE) Analysis of Reconstitutes Nucleosome Array

Samples	u_0' (x10^{-4} cm^2.volt^{-1}.sec^{-1}) Charge	**Radius** (Re, nm). from dilute gels
208-12 DNA	-3.1	61.9
208-12 NA	-2.3	31.9
Telomere DNA 1kb	-3.3	27.0
Telomere NA 1 kb	-2.5	23.0
Telomere DNA 2 kb	-3.2	49.6
Telomere NA 2kb	-2.5	29.3

Table was taken from reference [4]

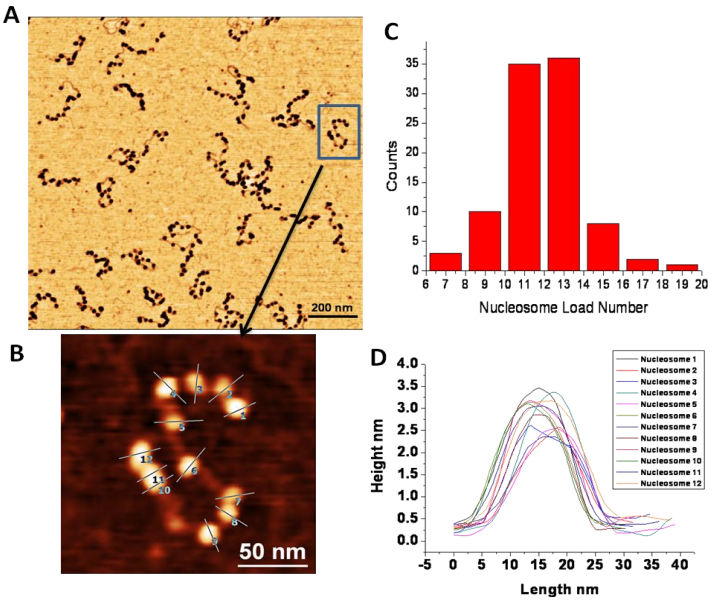

Figure 4-4 AFM analysis of reconstituted telomeric nucleosome array

(A) Nucleosome array reconstituted with the 2 kbp telomeric DNA fragments, (D) shows the height measurements for each nucleosome Indicated in (B), and (C) shows the nucleosome

In order to characterize the structural change after the binding of TRF2-DBD on the telomeric nucleosome array, we first reconstituted the nucleosome array on a DNA template with a ~580 bp 5'-TTAGGG-3' repeat sequence. Atomic force microscopy, micrococcal nuclease digestion and analytical agarose gel electrophoresis[90, 91] (AAGE) were applied to study the structural conformation change of the telomeric nucleosome array with different concentration of TRF2-DBD.

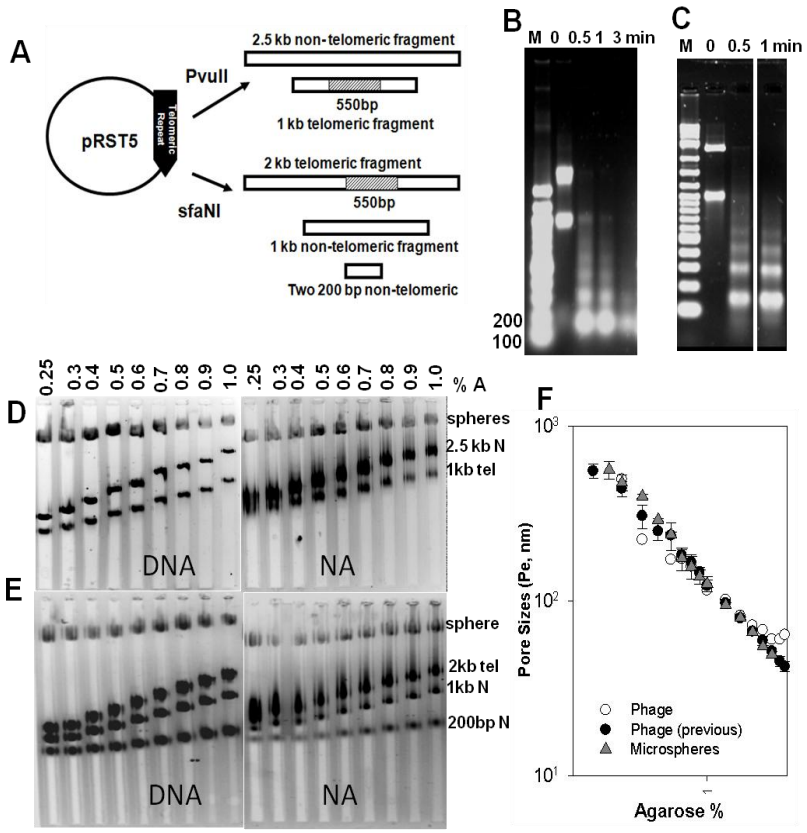

Figure 4-5 Agarose gel electrophoresis analysis on Telomeric Nucleosome Array

Here, (A) Shows the two different digestion products by Pvul I and sfaN I, (B) and (C) show the micrococcal nuclease digestion of the reconstituted nucleosome array in indicated time, (D) and (E) show the multi-gel result of histone free DNA (DNA) and nucleosome array fiber (NA), and (F) gel pore size measurement result by bacteriophage T3 (Phage) and carboxylate-coated microspheres (Microspheres)[4].

To avoid the interruption of the fragment size in the data interpretation, we studied the nucleosome reconstitution on both 1-kb and 2-kb telomeric fragments (Figure 4-5) and their binding to TRF2-DBD. A proper loading of nucleosome array was used for our study. From the micrococcal nuclease digestion results, the nucleosomes were properly reconstituted with periodic spacing. The 580 bp of telomeric DNA was also positioned with nucleosomes (AFM images). The nucleosome saturation levels were obtained from the AAGE gel. The gel-free mobility $u_0^{'}$ directly related to the electrical surface charge density of the molecule[90, 91] . Different DNA fragments should give a similar charge density as shown in table 1; the charge densities for 208-12 and telomeric DNA are the same. When DNA templates bind with positive charged histone octamers, the surface charge density should decrease [7, 165]. The negative surface charge density should be proportional to the number of nucleosomes per DNA base pair[91].

The $u_0^{'}$ can be directly related to the sedimentation coefficient since a decrease in negative surface charge density correspond to an increase in S20,w [91]. Every ~20% drop of negative charge relates to ~ 1 nucleosome/208 bp of DNA. In table 4-1, ~22-25% drop of $u_0^{'}$ was obtained when the telomeric DNA sequence was saturated with histones. The loading level of the nucleosome array can also be obtained from the effective radius

(R_e)[91]. We used to use bacteriophage T3 with a 30.1 nm radius as our standard for the gel pore sizes (P_e) extrapolations[91]. Instead of bacteriophage T3, we use commercially available, carboxylate-coated microspheres as our standard to get P_e. The microspheres yielded the same result as the T3 (see Figure 4-5F). In dilute gels (0.25% to 0.6% gels), the effective radius of the molecule was constant and the radius obtained from this case was more like the hydrodynamic radius of the molecule in solution. This is confirmed by comparing to the sedimentation coefficient; when increasing the loading of histone octamer on the 208-12 DNA template, the Re decreases[91]. For the 208-12 and 2-kb telomeric fragments, the Re was reduced by 40-50% after being assembled with histones (Table 4-1). Based on the $u_0^{'}$ and the R_e data, it is assured that the density of loading for the telomeric nucleosome array is 11-12 nucleosomes on the 2-kb fragment, and this loading was utilized for the following experiments. The loading was also confirmed by atomic force microscopy; the nucleosome array fiber on the 2-kb fragments gave an average loading of 12 nucleosomes (Figure 4-4 A and C), which is the same as the AAGE gel result. The maximum height distribution of nucleosomes measured from the image is ranged from 2.5 to 3.5 nm for the nucleosome which is in agreement with our previous result[166].

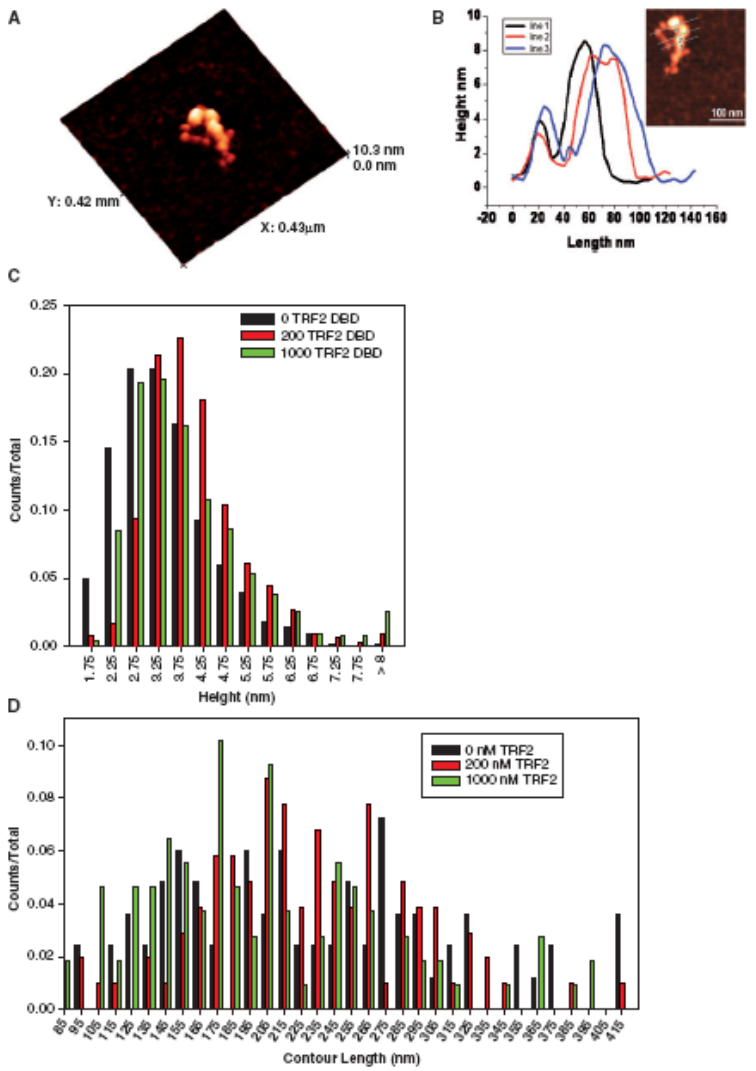

Figure 4-6 TRF2 DNA binding domain (TRF2-DBD) and telomeric nucleosome array fiber structure analysis by Atomic Force Microscopy. (A) 200 nM TRF2-DBD binding at the center of the telomeric nucleosome array, (B) Height analysis cross the molecule shows the binding of DBD giving a height around 10 nm, (C) normalized TRF2-DBD dependent fiber height distribution, and (D) normalized TRF2-DBD dependent fiber contour length distribution [4].

4.3.2 AFM of TRF2-DBD Complexed with DNA and Nucleosomal Array Fibers

The AAGE data shows that the addition of TRF2-DBD does not increase the radius R_e (not shown here). When the concentration of TRF2-DBD rose higher than 400 nM, the negative charge continued to decrease when the concentration reached 1000 nM. Micrococcal nuclease digestion shows that the accessibility of the DNA was dramatically reduced when the TRF2-DBD concentration rose higher than 500 nM. From these result, we can conclude that the TRF2-DBD keeps binding to the nucleosome when concentration rises higher than 400 nM. However, a compaction of the telomeric nucleosome array was induced. Single-molecule visualization of the process will be very informative. A proper loading~8 of nucleosome arrays was used for the TRF2-DBD interaction. The TRF2-DBD was mixed with the nucleosome array in several different concentration ranges and fixed with glutaraldehyde. AFM images were obtained over a range of TRF2-DBD with the concentration from 0 to 1000nM. From the AFM image of the single nucleosome array and TRF2-DBD complex, a preferential binding of TRF2-DBD to the center of the nucleosome array where the telomeric sequence is located was observed (Figure 4-6A). Based only on the AFM image, the role of TRF2-DBD binding to the telomeric nucleosome array is not clear. Whether TRF2-DBD causes nucleosome displacement or the TRF2-DBD

induce a compaction of the nucleosome array is not clear. The results from the micrococcal nuclease digestion didn't show any nucleosome displacement (Figure 4). Additionally, the increase of Re at low TRF2-DBD concentration was not in agreement with significant compaction. The interaction between TRF2-DBD and Histone-free DNA (with telomeric Sequence) were also studied (Figure 4-8F).

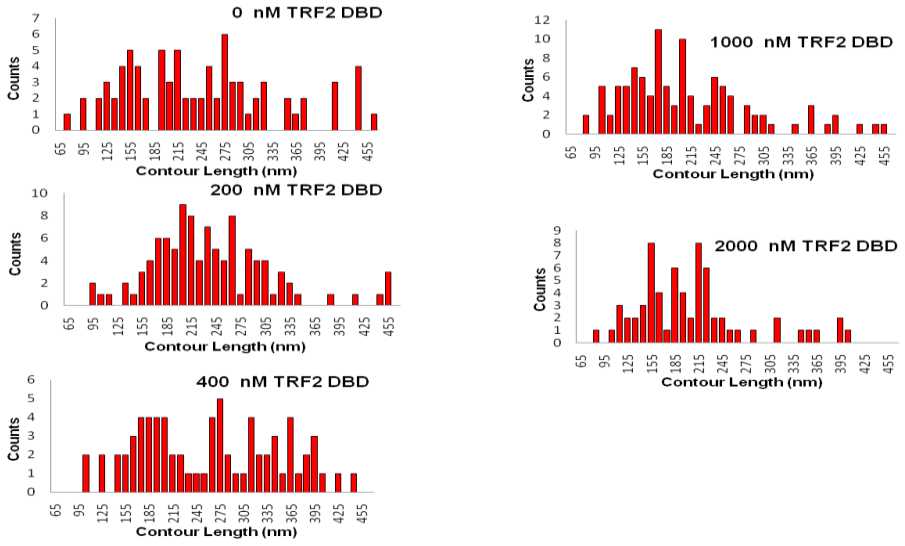

Figure 4-7 Histogram analysis of the contour length of telomeric nucleosome array-DBD complex at different DBD ratios.
Different concentrations of DBD, ranging from 1 to 2000 nM, were mixed with telomeric nucleosome array, contour lengths distribution were measured from AFM images [4].

Quantitative characterization of the TRF2-DBD-dependent variations of the telomeric nucleosome array structure was done by analyzing the fiber contour length and heights as a function of TRF2-DBD concentration. (Figure 4-6D, 4-7, and 4-9).

The nucleosomes height from AFM images usually varied from 2.5 to 3.0 nm (Figure 4-4D). Figure 4-6A and B illustrate an example of the height distribution of a nucleosomal fiber when binding with TRF2-DBD. The height in the nucleosome range is ~3nm, an intermediate size of 4-6 nm, and the highest range at around ~6 nm. Histogram analysis of the molecule heights for each TRF2-DBD concentration was studied. (Figure 4-7). All the histogram data was normalized to 1 for comparison between TRF2-DBD concentrations (Figure 4-6C).

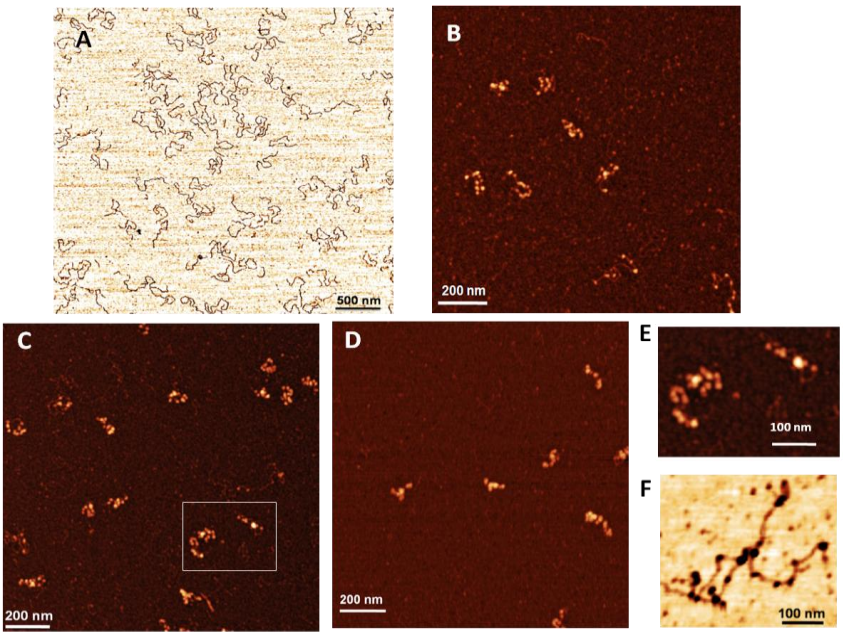

Figure 4-8 Examples of the AFM imaging of telomeric DNA, TRF2-DBD nucleosome array complex, and DBD-DNA complex

(A) Free 2 kbp telomeric DNA, (B, C, D, and E) TRF2-DBD nucleosome array fiber, and (F) TRF2-DBD binding with DNA[4].

When the TRF2-DBD concentration went to 200 mM, the population of the fiber heights shifted from around 3 nm to nearly 4 nm, and an increase of the population of fibers from a height of 4 nm to 8 nm was also demonstrated. Increasing TRF2-DBD concentration from 400 to 1000 nM did not induce an enhancement of the number of fibers with heights higher than 6 nm. The increase in heights of the complex could be induced by the TRF2-DBD-nucleosmoe complexes, fiber compaction, or both. To understand whether the height change was due to fibers compaction or not,

the relationship between contour lengths and TRF2-DBD concentration was studied. (Figure 4-6D and Figure 4-9).

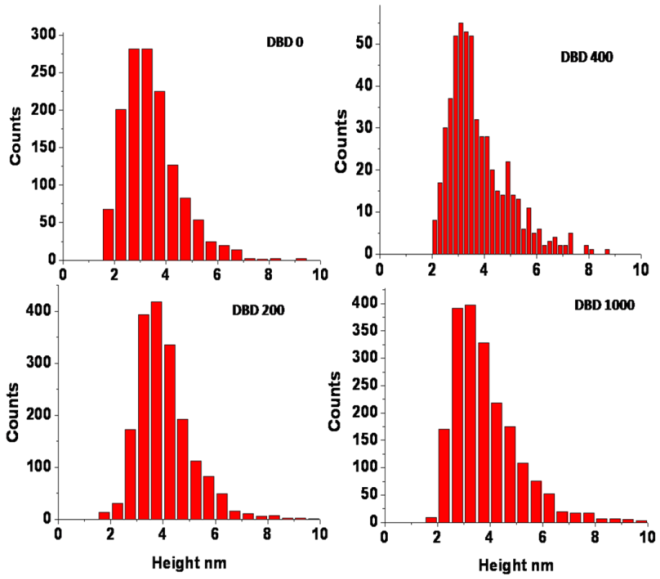

Figure 4-9 Histogram of the height of TRF2-DBD-Nucleosome complex at different DBD concentration

Different concentrations of DBD, ranging from 1 to 2000 nM, were mixed with telomeric nucleosome array, the height of the particles were measured from AFM images, and distributions were plotted here [4].

The distribution of the telomeric nucleosome array with no TRF2-DBD presentation was very broad with the center around 240 nm. When the TRF2-DBD went to 200 nM and 400 nM, a more narrow length distribution was shown; this means that the DNA wrapping around the histone is more confined after binding with TRF2-DBD, and the distribution center is still

around 240 nm. (Figure 4-9). However, when the TRF2-DBD concentration rose to 1000 and 2000 nM, the contour length of the molecule decreased dramatically to less than 200 nm. This suggests that when the TRF2-DBD concentration goes higher than 400 nM, a compaction of nucleosome fiber will happen.

4.4 Conclusions

In summary, we characterized the structural properties of the TRF2-DBD and telomeric nucleosome array complex fiber by AAGE and AFM. From the AFM observations, TRF2-DBD bound specifically to the telomeric region of the nucleosome array. Lower concentration of TRF2-DBD, such as 200 nM, did not induce obvious change from AFM measurement, however, a more flexible nucleosome array structure was revealed by the Analytical agarose gel electrophoresis (AAGE). The increased flexibility might be contributed by the binding of TRF2-DBD to the telomeric nucleosomal DNA. In comparison, higher concentrations of the TRF2-DBD induced a more compacted structure from the contour length measurements of AFM images. This was also further confirmed by the AAGE measurements, which demonstrated a fiber complex with reduced negative surface charge and flexibility caused by high concentration of TRF2-DBD.

Chapter 5

5 Nucleosome Positioning on a Mouse Mammary Tumor Virus (MMTV) Promoter

5.1 Introduction

To study the organization of the nucleosome along the DNA becomes extremely desirable for a better understanding of the gene transcription and activation process, especially for the nucleosome positioning at the upstream of the gene; accumulating evidence shows that the positioning ability of the gene to the histone octamer is directly related to the transcription level[24]. The binding affinity landscape between the regulatory DNA sequence and the histone octamer could be changed directly by several different ways [167].

Techniques currently used to study the *in vitro* nucleosome positioning have been largely based on DNA foot printing related methods. The first high quality mapping of nucleosome position down to base-pair resolution was achieved by site-directed hydroxyl radicals digestion [107]. This method modifies the Histone H4 to carry an EDTA-derived reagent to tag Fe^{3+} for the following hydrogen peroxide treatment. The base pair resolution was limited by the length of the DNA template. However, large scale nucleosome mapping for a whole genome was achieved by DNA array technology [168], where a large genome of chromatin was first digested into mononucleosomes.DNA fragments were then isolated from the

mononucleosomes and hybridized into a DNA microarray chip for further analysis[169].

Atomic force microscopy has been used as a powerful and versatile tool for biochemical research, especially for studies in physiological environments since the AFM probe can operate under buffer conditions. Taking advantage of this property, we have applied this technique to study the statistics of the distribution of both the histone octamer along the Mouse Mammary Tumor Virus (MMTV) promoter region and *in vitro* chromatin remodeling by SWI/SNF[170]. The MMTV promoter region is a well established model system for studying a steroid induced transcription activation process. Additionally, the histone octamer positioning ability on the MMTV promoter region has been both studied *in vivo*[52] and *in vitro*[171], and six positioned nucleosome families have been identified in this region, which was defined as nucleosome A (Nuc-A) to Nuc-F[23, 105, 107, 171]. However, all the AFM work on this template is limited by identifying the exact nucleosome position along the DNA template. The orientation problem of molecules in the AFM images is caused by the identity of the two ends of the DNA or nucleosome array.

To overcome this problem, we constructed a DNA template consisting of a 9 tandem repeats (177x9-601) positioning sequence[109] and the MMTV promoter sequence; the 601 sequence is a non natural histone binding

sequence that has the strongest binding affinity to the histone octamer obtained through SELEX[61]. This hybrid sequence was reconstituted with a histone octamer to give a substantiated nucleosome array. By using the highly positioned 601 sequence as a built-in reference, the nucleosome positions along the MMTV promoter were studied. The result from the AFM mapping experiment showed a nucleosome favorable position around the Nuc-A and B region when loadings of the histone octamer were around 11 to 13, which has been pointed out by others (Richmond Group) from biochemistry experiments[23]. With an increased loading, the nucleosome positioning ability was lost, but a region of favorable position was still observable in the A to D region. Strong terminal effects were observed in all of the different loadings.

5.2 Material and Method

Plasmid Construction: The MMTV plasmid was digested with Pst I, the sticky ends were then blunted with T4 DNA polymerase (New England Biolabs). Pst I digestion produced a 3' overhang, thus 4 nucleotides were lost after this treatment. This product works as a vector for the following experiment, for the 601 insert, 17mer plasmids were digested with Hind III. This would give us a 1587 bp with 9 601 sequence repeats. Hind III is 5' overhang, 4 nucleotides will added after blunt with T4 DNA polymerase. This insert was blunted with T4 DNA polymerase. So a 9 repeat 601 sequence is attached to the MMTV sequence. The 601-MMTV promoter sequence was cut out from

the new plasmid with Hind III and Nco I, the full length for this DNA template is 3452 bp.

Reconstitution of Nucleosomal Arrays: Histone octamers were purified from chicken erythrocytes [113]. Nucleosomal arrays were reconstituted by stepwise salt dilution[113]. Briefly, 3 ug DNA template purified from the plasmid was used for each reconstitution, the DNA and histone ratios were adjusted to get sub saturated nucleosome array. The histone octamer and DNA were mixed to achieve final concentrations of 0.1 ug/ul DNA, 1 mM DTT, 1xTE, 1M NaCl. The mixture was allowed to stay on the bench for 30 mins at room temperature, and then transferred to dialysis tubing and followed by step salt dialysis against 0.8 M NaCl, 0.6 M NaCl and 0.15 M NaCl with 1x TE (pH 8.0) buffer for 2 hours respectively at room temperature. The sample was finally dialyzed against 1 mM Na2.EDTA (pH 7.5) overnight at 4°C. Salt reconstitution produced an equilibrium state conditions for the sample[61].

Atomic Force Microscopy: See chapter 2 Material and Method section

5.3 Results and Discussion

5.3.1 Nucleosome Reconstitution:

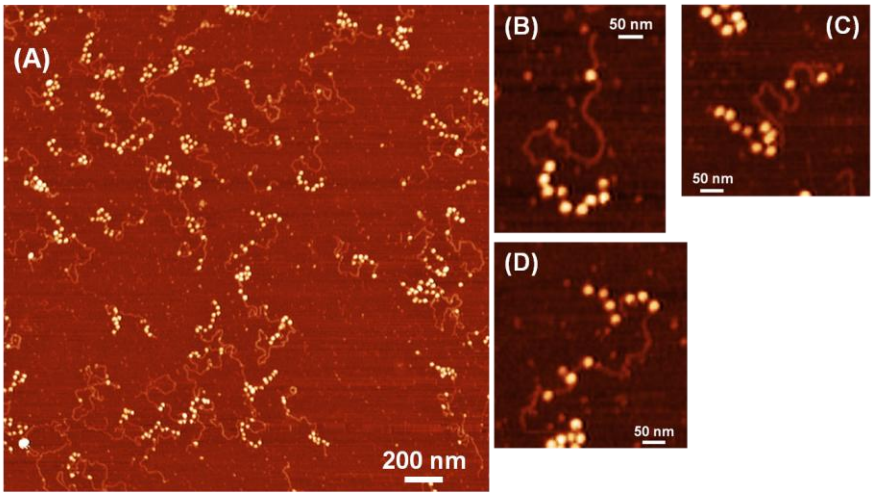

Figure 5-1 AFM images of the nucleosome array reconstituted on the concatenated 601-MMTV sequence.

Nucleosomes were reconstituted on the concatenated sequences by step salt dialysis method, and this method induce a binding of histone octamer in a thermodynamic favorable fashion: histone octamers saturate the 601sequence first and then binds to other sequence's preferred locations on the MMTV promoter region[61]. A lower ratio between DNA and histone octamer was used in the experiment for a subsaturated nucleosome array, see Figure 5-1, a zoom in image in Figure 5-1 B shows an example of the saturated 601 region and another single nucleosome located in the MMTV region. By using the 601 as a build-in reference marker, we could identify the exact position of the nucleosome located on the MMTV region, with several process of imaging analysis, the exact position of the nucleosome

located on the MMTV promoter region could be identified in almost base pair resolution.

5.3.2 Image Analysis:

In order to obtain the relative location of the nucleosome on the MMTV region, the path of each nucleosome molecule in the image was analyzed by the chromatin analysis platform. The contour distances between the centers of each nucleosome were measured. With increased loading, the end-to-end contour lengths decreased. Therefore, in order to obtain the wrapping length, the contour length as a function of loading was studied (Figure 5-2). Since the molecules were traced by starting from the center of the terminal nucleosome, there was a missing part of the molecule's contour length that came from the unmeasured half part of the terminal nucleosome. The contour lengths of the molecule were corrected for the terminal nucleosome's effect by adding an offset term as shown in equation 5-1 and Figure 5-5.

The diameter of each nucleosome was taken from the crystal structure, which is 11 nm[7]. For a given DNA wrapping length w assembled to a nucleosome, the contour length of the sample would decrease by a factor of $w-d$. Additionally, the terminal nucleosomes were corrected in Equation 5.4, here ne is the number of terminal nucleosome, L is the free DNA length,

n is the number of loading, and L_{tip} is the broadening effect coming from the AFM tip. By plotting the loading with the contour length, the slope will be the $w-d$, where d is 11 nm, and wrapping length w can be measured. The theoretical length of the DNA was ~1174(3452*0.34) nm [172] . The measured length from the AFM images gave a value of 1214 ±2 nm. The value from our data analysis is approximately 1127nm, which was estimated based on the interception of the plot of the loading vs. contour at zero loading. The contour length 1127 nm of the free loading from the plot is smaller compare to the theoretical DNA contour length; this might caused by the linker DNA was hardly being identified in the AFM image for these closely contacted nucleosomes. The slope obtained is 30.48 nm, correspond to 41.48 nm and 122 bp of wrapping length, the theoretical wrapping length for nucleosome is 147*0.34=49.9 nm[7]. The measured wrapping length is around 10 nm less compared to the theoretical wrapping length; the decrease of wrapping length is due to the unwrapping of the nucleosome on the positive charged APTES modified mica surface and is in reasonable range. Moreover, from early hydroxyl radical foot printing study, 120 bp DNA were shown directly binding to the H3/H4 tetramer[20]. We can use the 122 bp of wrapping length for the further nucleosome positioning analysis.

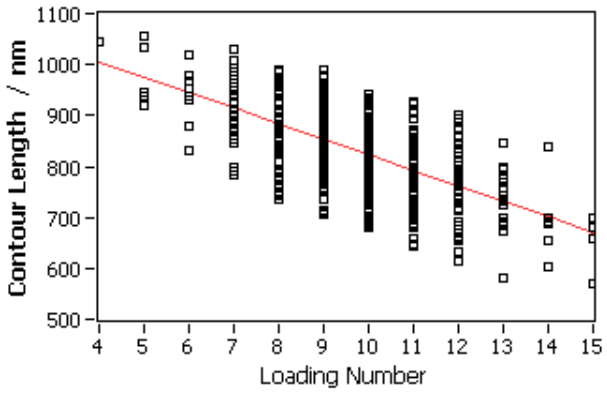

Figure 5-2 Nucleosome array contour length as a function of loading number

Data obtained from approximately 520 molecules were plotted here; each open square represents one molecule with specified loading number and contour length. The red line is the linear fitting of all the data. The slope of the red line, 30.48 nm, gave the wrapping length of 41.48 nm, and the interception of this line at zero loading was 1127 nm.

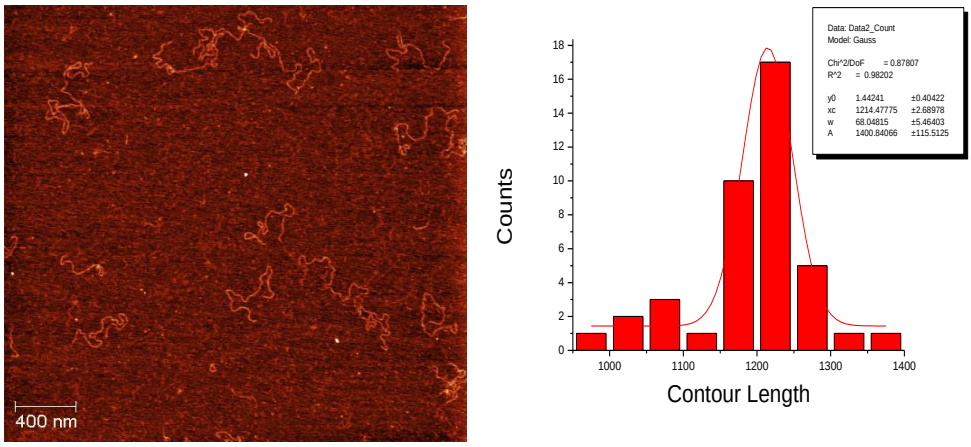

Figure 5-3 AFM imaging of free MMTV-601 DNA on APTES mica

Left part of the graph shows an atomic force microscopy image of free MMTV-601 DNA, right part of the graph shows the contour length distribution of this template.

5.3.3 Wrapping Length Analysis:

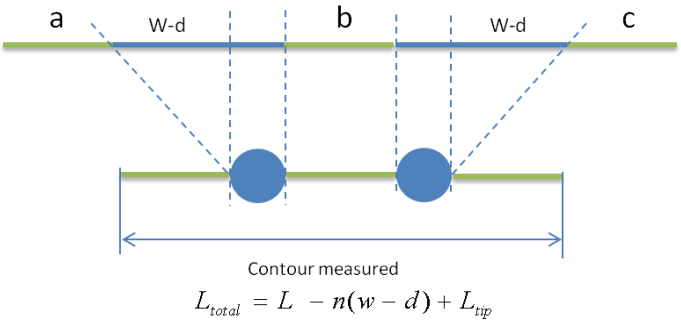

Figure 5-4 Illustration of Non-terminal Nucleosome analysis

The free DNA length is a constant denoted as L . Each nucleosome consists of 147 bp DNA wrapped around the histone octamer, and the length for the

wrapping length was denoted as w, and the theoretical wrapping length is 50 nm. Each nucleosome width was denoted as $d = 11\,nm$ obtained from the crystal structure, and the tip broadening effect was denoted as L_{tip}. The relationship between the nucleosome loading n and the nucleosome array contour length L_{total} was shown in figure 5-4 for non terminal nucleosomes. However, for terminal nucleosomes, the equation was corrected to equation 5-1 (Figure 5-5). n_e is the number of terminal nucleosomes. The wrapping length for each molecule could be described by equation 5-2.

$$L_{total} = L - n(w-d) - n_e d/2 + L_{tip} \qquad 5\text{-}1$$

$$w = \frac{L - n_e d/2 + L_{tip} - L_{total} + dn}{n} = \frac{L - L_{total} + d(n-n_e) + n_e d/2 + L_{tip}}{n} \qquad 5\text{-}2$$

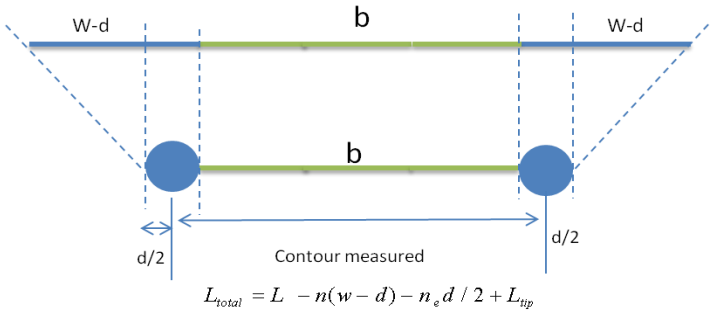

Figure 5-5 Illustration of terminal nucleosome analysis

With the measured wrapping length of ~122 bp, the dyad location of each nucleosome on the template can be studied by unwrapping the nucleosome array and remapping it to DNA in base pairs (Figure 5-6). The DNA that

wrapped around the nucleosome was coded to 1, and the length of the histone octamer positioning region was determined by the wrapping length, which was 122 bp. The intermediate DNA that connects the nucleosome was coded as 0, since the contour length between each nucleosome measured is center-to-center distance, so the real linker DNA length need to minus the diameter of the nucleosome, which is 11 nm[7] (Figure 5-6).

The remapped data appeared a connection of different lengths of color coded DNA, 1 in red color and were referred to nucleosome positioning sequence, 0 in light blue and refers to linker DNA. By normalizing the length of the DNA to the base pair length of the DNA, which is 3452 bp, we could achieve a base--pair resolution of the position of each nucleosome (Figure 5-7).

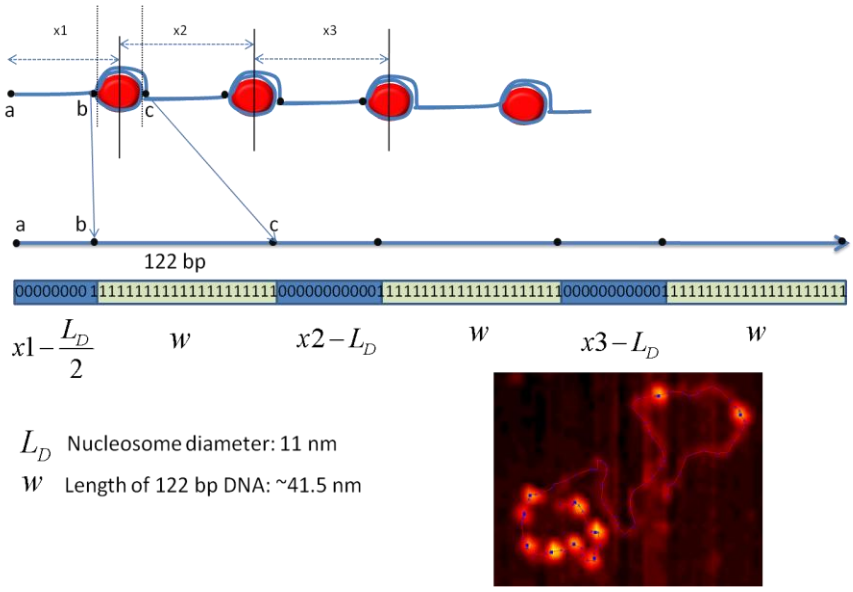

Figure 5-6 Schematic illustration of the remapping process

The dyad position of each nucleosome was identified. With a given wrapping length of 122 bp, the sequences that wrapped around histone core could be identified and was coded as 1(light green), the linker DNA region was coded as 0 (blue).

5.3.4 Nucleosome Positioning on MMTV Promoter

After mapping near 520 concatenated nucleosome arrays with different loadings, a nucleosome favorable positioning pattern was plotted (Figure 5-7). Histone favorable 601 positions were shown on the left of the graph as 9 sine wave peaks, indicate 9 nucleosome preferred sites. The right side of the graph randomly shows nucleosome favorable sites on the MMTV promoter region accumulated over all the molecules. In order to identify the one terminal of the nucleosome array, the molecules with an identifiable 601 region were chosen for this positioning analysis, which means subsaturated

arrays. Therefore, based on our result, the 601 region has a higher count for positioning compared to the MMTV promoter region (Figure 5-7). Three Endonuclease restriction digestion sites were indicated in Figure 5-8: AlwN I at 2490, and Sac I at 2675, EcoR I at 3148 to work as a reference for the 6 positioned nucleosome position with a biochemistry method[105, 171]. The locations of the AlwN I and Sac I were labeled in the relative positions compared to specific nucleosomes. Restriction site AiwN I is located near the ends of Nuc-C, and Sac I site is located at the ends of Nuc-B.

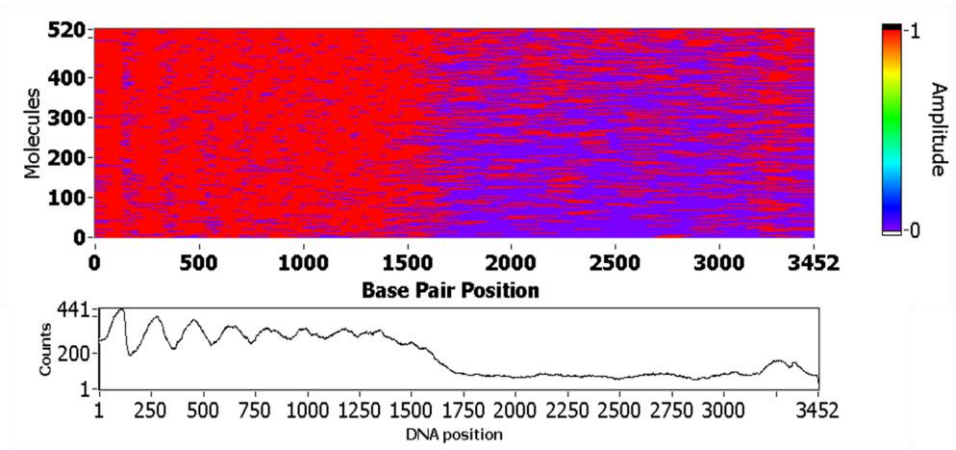

Figure 5-7 Nucleosome position over 520 molecules on the MMTV-601 sequence

The top part of the graph illustrates the nucleosome positioning along this templates for every single molecule. Red line means a nucleosome is positioned there and denoted as 1. Purple is free DNA and denote as 0. The bottom part illustrates the plot of accumulation of all the counting on the nucleosome positioning along the template.

The nucleosome positioning as a function of different loadings was also plotted; the nine 601 positioning regions were observed in all the loadings, and a strong terminal nucleosome effect was found in all the loadings for both the MMTV promoter region and the 601 region. A nucleosome favorable position range over the Nuc-B and Nuc-A region was also observed at low loading number ranging from 7 to 11. Since the positioning ability plot shows in Figure 5-7 is an accumulated result of all the possibilities of the nucleosome positions on the DNA template, it shows a dramatic peak difference between the 601 region and the MMTV promoter region. As the loading numbers reaches 14, two distinct peaks with almost similar counts as the 601 nucleosome were identified. They correspond to the Nuc-B and Nuc-A correctly based on the restriction sites marker; both Nuc-A and Nuc-B show a similar count, which suggests an equivalent binding preference for histone octamer at higher loadings.

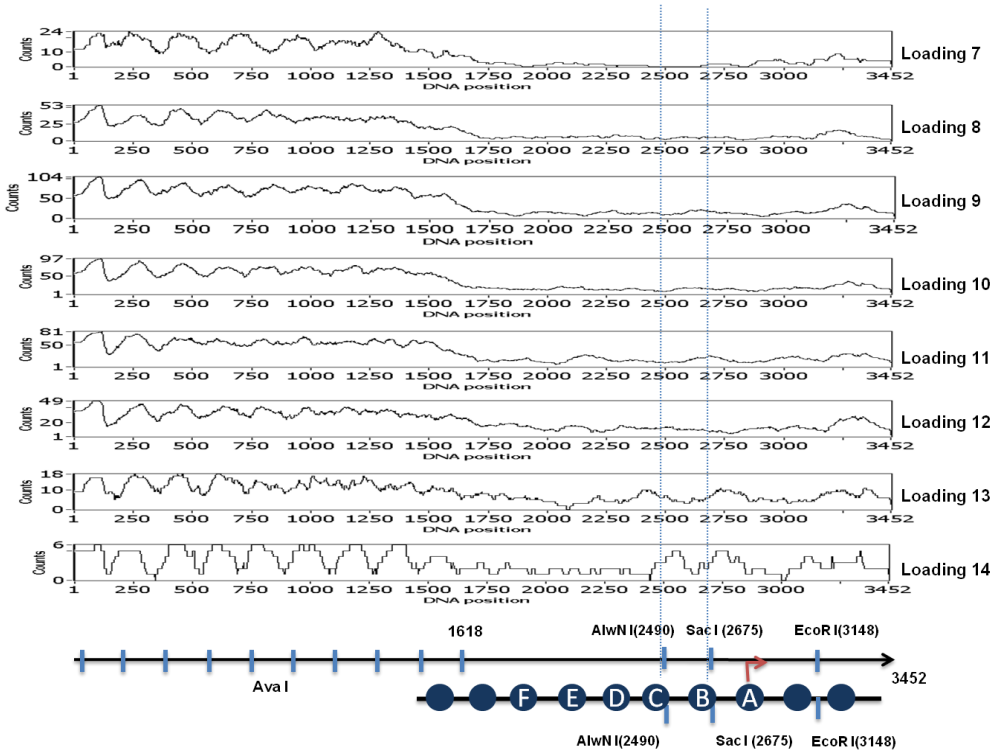

Figure 5-8 Nucleosome position at different loadings on the MMTV-601 sequence

The nucleosome positioning behavior along the MMTV-601 template at different loadings was shown here. At loading 14, two highly positioned regions corresponded to the Nuc B and Nuc A were indicated in the graph. The bottom part of the graph illustrates the relative position of the six positioned nucleosome along the MMTV promoter.

5.4 Conclusion

In summary, we studied the nucleosome positioning behavior on a target sequence *in vitro* by introducing a nucleosome highly favorable sequence 601 next to the MMTV promoter sequence. We also investigated the

nucleosome positioning behaviors on the MMTV promoter as a function of loading and identified two highly positioned locations on the MMTV promoter region, which correspond to Nuc-A and Nuc-B. The obtained result herein agrees well with those previously reported using biochemistry mapping methods[23, 105, 107, 171]. This developed method has the promise to study nucleosome positioning behavior along the DNA template with any length.

Chapter 6

6. Conclusions and Outlook

In this thesis, we focused on the structural properties of chromatin in different states, especially in their condensed state with the method of atomic force microscopy. Atomic force microscopy has been developed as a versatile tool for analyzing structural properties as well as chemical information of biomaterials. Different from electron microscopy (EM), AFM could operate under various environments, such as in physiological condition. Despite the two dimensional size information, AFM could also provide the height, stiffness, and chemical information of the sample based on different experiment schemes. The size of chromatin is in a range from 2 nm (DNA) to 1000 nm (chromosome). This broad distribution falls into the range which can be handled by AFM.

In chapter 2 of this thesis, the condensation of an artificial nucleosome array reconstituted from histone octamers with single acetylation at histone H4 lysine 16 was studied by AFM. Amino acids 16 to 20 of histone tail H4 are the most important ones in facilitating the interaction between nucleosomes based on the crystal structure[7] . Particularly, acetylation at lysine 16 can potentially break this specific interaction[173]. From our AFM observations, both WT and H4K16 (Ac) nucleosome arrays folded into compact spherical particles, and no apparent difference was found. Geometrical properties of compact H4K16 (Ac) and WT nucleosome arrays were derived from the AFM

data, and a difference of 0.46 nm in radius was obtained. Since the nucleosome array folded into spherical particles, the Stokes equation was applied to approximate these particles' hydrodynamic friction coefficients. By doing this way, a difference of 0.11 nm in radius was obtained from the published sedimentation coefficient data[10]. Compared to hydrodynamic methods, AFM measurements provided a direct visualization of compaction of nucleosome arrays for both H4K16 (Ac) and WT. In comparison to the small effects of histone tail H4 K16 acetylation in the compaction of nucleosome array, a proper loading along the DNA template played a more progressive role in compaction of the nucleosome array under Mg^{2+}. The folding behavior of a concatenated sequence showed that highly positioned 601 region with uniformly spaced nucleosomes folding into compact structure alone independent of the low loading region. In addition, the compaction of nucleosome array started from a close compaction between its nearest neighbors. This close interaction between closely contact nucleosome induced a dinucleosome structure, which collapsed in a face to face fashion revealed by the AFM. This might suggest a one-start organization of the nucleosome in higher ordered structures.

In order to better understand the role of linker DNA in controlling the compaction of the nucleosome array, a method to construct very long artificial nucleosome array based on the highly positioned 601 sequence was developed in chapter 3.1. In the first part of chapter 3, total 42 monomers

with average linker lengths of 30bp, 40bp, 50bp and 60bp were constructed. Each monomer consists of a 147 bp of 601 highly positioned sequence located in the center, and a varied left and right linker DNA located in the two ends. The final constructed long tandem repeat sequences possess a distribution of linker DNA with lengths obeying the form of 5n. DNA templates containing tandem 601 repeats with average linker length of 30 bp and 60bp in both uniformly and randomly spaced forms were constructed. Additionally, by introducing the type IIs restriction enzyme BsmB I and Bsa I into the vector, very long and defined repeat sequences can be constructed. For example, 39 repeats of 177-601 sequence were constructed based on a short template which contains 13 repeats. Furthermore, a specific defined DNA template with constrained organization of linker length was constructed. For example, in the last part of chapter 3.1, a DNA template was constructed with a short -long organized fashion of linker DNA while keeping the average linker length of 60bp. The loading behavior and nucleosome positioning were studied on this template. The nucleosome positioning on the template is dominated by the sequence binding affinity mostly and a cooperative loading of nucleosomes was found from AFM result. Interestingly, in low loading, the nucleosome positioned more at the second 601 slot, which is close to the DNA terminus, compared to the terminal 601 position. These studies on the nucleosome array with extremely positioned situation can help us better understanding the effects of nucleosome positioning on their higher order structure.

In chapter 3.2, the 2D geometrical features of the 30 pb as well as 60 bp uniformly and randomly spaced nucleosome array were studied. Combing with custom written matlab program, the DNA entering and exiting angles as well as the inter-nucleosome distances were analyzed. The arrangement of the DNA entering and exiting angle could partially help us to understand the nucleosome arrays' geometries before their compaction in buffered environment. A uniformly spaced nucleosome array in a short linker length of 30bp provided a more straight structure compared to the randomly spaced nucleosome array. This was illustrated from the 2D histogram of the entry and exit angles. 2D histogram map is useful in interpreting result when the data contain multiple features. In comparison to the 30bp construction, 60 bp uniformly spaced nucleosome array presented a higher population of small angles compared to the 60 bp randomly spaced nucleosome array. In addition, 60bp linker DNA length gave a broader distribution of the DNA entry-exit angles suggesting that longer linker DNA length facilitated a more flexible organization of nucleosomes along the DNA template. We propose that the orientation of each nucleosome in the extended form can provide insights into their pathway of compaction when Mg^{2+} present. Therefore, the folding behavior between uniformly and randomly spaced nucleosome array were further studied in 1 mM Mg^{2+} folding buffer. From the AFM result, the linker DNA length seems playing an important role in controlling the Mg^{2+} induced compaction of nucleosome array. However, results based on the AFM imaging alone is not enough in understanding the compact particle's

structural properties. Quantitative agarose gel electrophoresis or analytical ultracentrifugation is required for better characterization of the surface charge radius, as well as the folded particle's partial specific volume, which are independent of the nucleosome arrays' sizes.

Nuclear factors play an important role in maintaining chromatin structures. Therefore, in chapter 4 of this thesis, the compaction of telomeric nucleosome array under telomeric repeat factor 2's (TRF2) DNA binding domain (DBD) was studied by AFM. TRF2 specifically binds to double stranded - (TTAGGG) n- sequence located in the edge of the T-loop structure, and protects the 3' single stranded overhang in the G-quadruplex structure. From the AFM imaging, TRF2-DBD bound specifically to the telomeric region of the nucleosome array and gave a complex with a maximum height around 8 nm. In low concentration of TRF2-DBD, such as 200 nM, no apparent difference was found from the AFM measurement between 0 nM and 200 nM. However, analytical gel electrophoretic (AAGE) revealed that the complex possessed a more flexible structure at 200 nM, which might be caused by the binding of TRF2-DBD to the nucleosomal DNA. In contrast, at higher concentration of TRF2-DBD, such as 1000 nM, a decrease of the contour length was revealed from the AFM measurement suggesting that a compaction of the nucleosome array was caused by the binding of TRF2-DBD. This observation was also confirmed by the reduced negative surface charge and flexibility from the AAGE data.

In chapter 5, AFM was applied to study the nucleosome positioning along the Mouse Mammary Tumor Virus (MMTV) promoter region. The orientations of DNA and nucleosome array are usually difficult to be identified in AFM images. A novel method was developed in this chapter for locating exact nucleosome position along the template by introducing an internal marker, such as 601 tandem repeats. The wrapping length of the nucleosome was obtained by plotting the contour length as a function of loading. With the wrapping length for each nucleosome, the exact location of nucleosome positioning along the MMTV promoter region was identified from AFM images by the custom written nucleosome positioning analysis program. Two highly positioned locations corresponded to Nuc A and Nuc B on the MMTV promoter region were revealed from our analysis. This result is in well agreement with those previously reported by using biochemistry method[23, 105, 107, 171]. Moreover, direct visualization of nucleosome positioning in preferential locations can be done by AFM. This developed method has the promise to study nucleosome positioning behavior along the DNA template with any length.

In the near future, we believe that the studies of chromatin sample directly isolated from cells by AFM are the most exciting work to be done. Recent study of nucleosome isolated from the centromere of chromatin is a very good example. The AFM revealed that the height of the nucleosome isolated

from the centromere is different from nucleosomes obtained in other circumstances[174]. Centromere nucleosome possesses a tetrameric structure containing one copy of CenH3, H2A, H2B, and H4 each, and with a half height of normal nucleosome from AFM[174, 175]. Advanced techniques in chromatin sample isolation and preparation is in need for fully utilizing the AFM approach. Comprehensive studies and understandings of the chromatin structure and composition during different states and locations will be eager to be revealed.

REFERENCES

1. Harvey Lodish, A.B., S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell, *Molecular Cell Biology, 4th edition.* 2000.

2. Duesberg, P., *Chromosomal chaos and Cancer.* Scientific American, 2007.

3. Griffith, J.D., et al., *de Lange, Mammalian telomeres end in a large duplex loop.* Cell, 1999. **97**: p. 503e514.

4. Baker, A.M., et al., *The Myb/SANT domain of the telomere-binding protein TRF2 alters chromatin structure.* Nucleic Acids Res, 2009. **37**: p. 5019-5031.

5. Williams, S., *Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length.* Biophysical Journal, 1986. **49**: p. 233-248.

6. Robinson, P.J., et al., *measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure, Proc.* Natl. Acad. Sci. USA, 2006. **103**: p. 6506-6511.

7. Luger, K., et al., *Crystal structure of the nucleosome ˚ resolution core particle at 2 . 8 A.* Nature, 1997. **7**.

8. Ghosh, A. and M. Bansal, *A glossary of DNA structures from A to Z.* Acta Crystallographica Section D-Biological Crystallography, 2003. **59**: p. 620-626.

9. Liu, Y.Z., S.H. Leuba, and S.M. Lindsay, *Relationship between stiffness and force in single molecule pulling experiments.* Langmuir, 1999. **15**(24): p. 8547-8548.

10. Shogren-Knaak, M., et al., *Histone H4-K16 acetylation controls chromatin structure and protein interactions.* Science, 2006. **311**(5762): p. 844-847.

11. Nikitina, T. and C.L. Woodcock, *Closed chromatin loops at the ends of chromosomes.* The Journal of cell biology, 2004. **166**: p. 161-5.

12. Simpson, R.T., *Structure of Chromatosome, a Chromatin Particle Containing 160-Base Pairs of DNA and All Histones.* Biochemistry, 1978. **17**(25): p. 5524-5531.

13. Marsden MP, L.U., *Metaphase chromosome structure: evidence for a radial loop model.* Cell, 1979. **17**((4)): p. 849-58.

14. Langmore, J.P. and C. Schutt, *The higher order structure of chicken erythrocyte chromosomes in vivo.* Nature, 1980. **288**: p. 620-622.

15. Hays, F.A., J. Watson, and P.S. Ho, *Caution! DNA crossing: crystal structures of Holliday junctions.* Journal of Biological Chemistry, 2003. **278**(50): p. 49663-49666.

16. Smith, S.B., Y.J. Cui, and C. Bustamante, *Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules.* Science, 1996. **271**(5250): p. 795-799.

17. Yager, T.D., C.T. Mcmurray, and K.E. Vanholde, *Salt-Induced Release of DNA from Nucleosome Core Particles.* Biochemistry, 1989. **28**(5): p. 2271-2281.

18. Richmond, T.J. and C.A. Davey, *The structure of DNA in the nucleosome core.* Nature, 2003. **423**: p. 145-150.

19. Travers, A.A. and A. Klug, *The bending of DNA in nucleosomes and its wider implications.* Phil. Trans, 1987. **317**: p. 537-561.

20. Hayes, J.J., D.J. Clark, and A.P. Wolffe, *Histone Contributions to the Structure of DNA in the Nucleosome.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(15): p. 6829-6833.

21. Kornberg, R.D. and Y. Lorch, *Chromatin structure and transcription.* Annual review of cell biology, 1992. **8**: p. 563-87.

22. Shimizu, M., et al., *Nucleosomes Are Positioned with Base Pair Precision Adjacent to the Alpha-2 Operator in Saccharomyces-Cerevisiae.* Embo Journal, 1991. **10**(10): p. 3033-3041.

23. Flaus, A. and T.J. Richmond, *Positioning and stability of nucleosomes on MMTV 3JLTR sequences.* J. Mol. Biol, 1998. **275**: p. 427-441.

24. Segal, E. and J. Widom, *From DNA sequence to transcriptional behaviour: a quantitative approach.* Nature reviews. Genetics, 2009. **10**: p. 443-56.

25. Brower-toland, B.D., et al., *Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA.* Proc. Natl. Acad. Sci. USA, 2002. **99**: p. 1960-1965.

26. Mihardja, S., et al., *Effect of force on mononucleosomal dynamics.* Proc. Natl. Acad. Sci. USA, 2006. **103**: p. 15871-15876.

27. Anderson, J.D. and J. Widom, *Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites.* Journal of Molecular Biology, 2000. **296**(4): p. 979-987.

28. Anderson, J.D., P.T. Lowary, and J. Widom, *Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites.* Journal of Molecular Biology, 2001. **307**(4): p. 977-985.

29. Widlund, H.R., et al., *DNA sequence-dependent contributions of core histone tails to nucleosome stability: Differential effects of acetylation and proteolytic tail removal.* Biochemistry, 2000. **39**(13): p. 3835-3841.

30. Manohar, M., et al., *Acetylation of Histone H3 at the Nucleosome Dyad Alters DNA-Histone Binding.* Journal of Biological Chemistry, 2009. **284**(35): p. 23312-23321.

31. Bird, A.P. and A.P. Wolffe, *Methylation-induced repression - Belts, braces, and chromatin.* Cell, 1999. **99**(5): p. 451-454.

32. Nathan, D. and D.M. Crothers, *Bending and flexibility of methylated and unmethylated EcoRI DNA.* Journal of Molecular Biology, 2002. **316**(1): p. 7-17.

33. Pennings, S., J. Allan, and C.S. Davey, *DNA methylation, nucleosome formation and positioning.* Brief. Funct. Genomic. Proteomic., 2005. **3**: p. 351-361.

34. Tippin, D.B. and M. Sundaralingam, *Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm(5)CGG), d(Cm(5)CGGGCCm(5)CGG) and d(CCGGGCC(Br)(5)CGG) in three different conformations: Effects of spermine binding and methylation on the bending and condensation of A-DNA.* Journal of Molecular Biology, 1997. **267**(5): p. 1171-1185.

35. Silva, N.C.E., et al., *CENP-A-containing nucleosomes: Easier disassembly versus exclusive centromeric localization.* Journal of Molecular Biology, 2007. **370**(3): p. 555-573.

36. Bao, Y.H., et al., *Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA.* Embo Journal, 2004. **23**(16): p. 3314-3324.

37. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science (New York, N.Y.), 2001. **293**: p. 1074-80.

38. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-45.

39. KE, v.H., *Chromatin.* New York: Springer Verlag, 1989.

40. Hansen, J.C., K.E. Vanholde, and D. Lohr, *The Mechanism of Nucleosome Assembly onto Oligomers of the Sea-Urchin 5-S DNA Positioning Sequence.* Journal of Biological Chemistry, 1991. **266**(7): p. 4276-4282.

41. Brower-toland, B., et al., *Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes.* J. Mol. Biol., 2005. **346**: p. 135-146.

42. Berger, S.L., *Histone modifications in transcriptional regulation.* Current Opinion in Genetics & Development, 2002. **12**(2): p. 142-148.

43. Hake, S.B., A. Xiao, and C.D. Allis, *Linking the epigenetic 'language' of covalent histone modifications to cancer.* British Journal of Cancer, 2004. **90**(4): p. 761-769.

44. Carruthers, L.M., et al., *Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding.* Biochemistry, 1998. **37**: p. 14776-14787.

45. Thoma, F., T. Koller, and A. Klug, *Involvement histone Hl in the organization of the nucleosome and of the salt-dependent superstructures of chromatin.* 1979.

46. Ramakrishnan, V., et al., *Crystal-Structure of Globular Domain of Histone H5 and Its Implications for Nucleosome Binding.* Nature, 1993. **362**(6417): p. 219-223.

47. Duggan, M.M. and J.O. Thomas, *Two DNA-binding sites on the globular domain of histone H5 are required for binding to both bulk and 5 S reconstituted nucleosomes.* Journal of Molecular Biology, 2000. **304**(1): p. 21-33.

48. Brown, D.T., T. Izard, and T. Misteli, *Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo.* Nature Structural & Molecular Biology, 2006. **13**(3): p. 250-255.

49. Furrer, P., et al., *DNA at the Entry-Exit of the Nucleosome Observed by Cryoelectron Microscopy.* Journal of Structural Biology, 1995. **114**(3): p. 177-183.

50. Zlatanova, J. and K. Vanholde, *Histone H1 and Transcription - Still an Enigma.* Journal of Cell Science, 1992. **103**: p. 889-895.

51. Hansent, J.C. and A.P. Wolffe, *Influence of Chromatin Folding on Transcription Initiation and Elongation by.* DNA Sequence, 1992: p. 7977-7988.

52. Richardfoy, H. and G.L. Hager, *Sequence-Specific Positioning of Nucleosomes over the Steroid-Inducible Mmtv Promoter.* Embo Journal, 1987. **6**(8): p. 2321-2328.

53. Keeton, E.K., et al., *Glucocorticoid receptor domain requirements for chromatin remodeling and transcriptional activation of the mouse mammary tumor virus promoter in different nucleoprotein contexts.* Journal of Biological Chemistry, 2002. **277**(31): p. 28247-28255.

54. Truss, M., et al., *Hormone Induces Binding of Receptors and Transcription Factors to a Rearranged Nucleosome on the Mmtv Promoter in-Vivo.* Embo Journal, 1995. **14**(8): p. 1737-1751.

55. Müller, W.G., et al., *Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter.* Cell, 2007. **177**: p. 957-967.

56. Eltsov, M., et al., *Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(50): p. 19732-19737.

57. Graziano, V., et al., *Histone H1 Is Located in the Interior of the Chromatin 30-Nm Filament.* Nature, 1994. **368**(6469): p. 351-354.

58. Mcghee, J.D., et al., *Higher-Order Structure of Chromatin - Orientation of Nucleosomes within the 30 Nm Chromatin Solenoid Is Independent of Species and Spacer Length.* Cell, 1983. **33**(3): p. 831-841.

59. Butler, P.J., *A defined structure of the 30 nm chromatin fibre which accommodates different nucleosomal repeat lengths.* The EMBO journal, 1984. **3**: p. 2599-604.

60. Widom, J., J.T. Finch, and J.O. Thomas, *Higher-Order Structure of Long Repeat Chromatin.* Embo Journal, 1985. **4**(12): p. 3189-3194.

61. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.* Journal of Molecular Biology, 1998. **276**(1): p. 19-42.

62. Worcel, A., S. Strogatzt, and D. Rileyt, *AndbW *.* Biochemistry, 1981. **78**: p. 1461-1465.

63. Dorigo, B., et al., *Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber.* Science, 2004. **1571**: p. 10-13.

64. Schalch, T., *X-ray structure of a tetranucleosome and its implications for the chromatin fibre.* Nature, 2005. **436**: p. 138-141.

65. Binning, G., et al., *Surface Studies by Scanning Tunneling Microscopy.* Physical Review Letters, 1982. **49**(1): p. 57-61.

66. Binnig, G., C.F. Quate, and C. Gerber, *Atomic Force Microscope.* Physical Review Letters, 1986. **56**(9): p. 930-933.

67. Binnig, G., et al., *Atomic Resolution with Atomic Force Microscope.* Surface Science, 1987. **189**: p. 1-6.

68. Albrecht, T.R. and C.F. Quate, *Atomic Resolution Imaging of a Nonconductor by Atomic Force Microscopy.* Journal of Applied Physics, 1987. **62**(7): p. 2599-2602.

69. Martin, Y., C.C. Williams, and H.K. Wickramasinghe, *Atomic Force Microscope Force Mapping and Profiling on a Sub 100-a Scale.* Journal of Applied Physics, 1987. **61**(10): p. 4723-4729.

70. Albrecht, T.R., et al., *Frequency-Modulation Detection Using High-Q Cantilevers for Enhanced Force Microscope Sensitivity.* Journal of Applied Physics, 1991. **69**(2): p. 668-673.

71. Durig, U., O. Zuger, and A. Stalder, *Interaction Force Detection in Scanning Probe Microscopy - Methods and Applications.* Journal of Applied Physics, 1992. **72**(5): p. 1778-1798.

72. Garcia, R. and R. Perez, *Dynamic atomic force microscopy methods.* Surface Science Reports, 2002. **47**(6-8): p. 197-301.

73. Giessibl, F.J., *Atomic-Resolution of the Silicon (111)-(7X7) Surface by Atomic-Force Microscopy.* Science, 1995. **267**(5194): p. 68-71.

74. Cappella, B. and G. Dietler, *Force-distance curves by atomic force microscopy.* Surface Science Reports, 1999. **34**(1-3): p. 1-+.

75. Pe, Â. and R. Garcõ, *Dynamic atomic force microscopy methods.* Surface Science Reports, 2002. **47**.

76. Butt, H.J., B. Cappella, and M. Kappl, *Force measurements with the atomic force microscope: Technique, interpretation and applications.* Surface Science Reports, 2005. **59**(1-6): p. 1-152.

77. Noy, A., D.V. Vezenov, and C.M. Lieber, *Chemical force microscopy.* Annual Review of Materials Science, 1997. **27**: p. 381-421.

78. A-Hassan, E., et al., *Relative microelastic mapping of living cells by atomic force microscopy.* Biophysical Journal, 1998. **74**(3): p. 1564-1578.

79. Hinterdorfer, P., et al., *Detection and localization of individual antibody-antigen recognition events by atomic force microscopy.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(8): p. 3477-3481.

80. Stroh, C., et al., *Single molecule recognition imaging microscopy.* Proc. Natl. Acad. Sci, 2004. **101**: p. 12503-12507.

81. Han, W.H., S.M. Lindsay, and T.W. Jing, *A magnetically driven oscillating probe microscope for operation in liquids.* Applied Physics Letters, 1996. **69**(26): p. 4111-4113.

82. Bustamante, C., et al., *Entropic Elasticity of Lambda-Phage DNA.* Science, 1994. **265**(5178): p. 1599-1600.

83. Kienberger, F., et al., *Contact and macmode molecular recognition force microscopy (MRFM): Force spectroscopy of polyethylene glycol.* Biophysical Journal, 2000. **78**(1): p. 381A-381A.

84. Stroh, C., et al., *Single-molecule recognition imaging-microscopy.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(34): p. 12503-12507.

85. Fletcher, T.M., P. Serwer, and J.C. Hansen, *b. Quantitative analysis of macromolecular conformational changes using agarose gel electrophoresis: Application to chromatin folding.* Biochemistry, 1994. **33**: p. 10859-10863.

86. Hansen, J.C., et al., *Analytical ultracentrifugation and agarose gel electrophoresis as tools for studying chromatin folding in solution.* Methods-a Companion to Methods in Enzymology, 1997. **12**(1): p. 62-72.

87. Ralston, G., *Introduction to Analytical Ultracentrifugation.* BECKMAN.

88. Serwer, P. and J.L. Allen, *Agarose-Gel Electrophoresis of Bacteriophages and Related Particles .4. An Improved Procedure for Determining the Size of Spherical-Particles.* Electrophoresis, 1983. **4**(4): p. 273-276.

89. Fletcher, T.M. and J.C. Hansen, *Folding and Nucleosomal DNA Organization through Distinct Molecular Mechanisms *.* Biochemistry, 1995. **8**: p. 25359-25362.

90. Griess, G.A., et al., *The Sieving of Spheres during Agarose-Gel Electrophoresis - Quantitation and Modeling.* Biopolymers, 1989. **28**(8): p. 1475-1484.

91. Fletcher, T.M., et al., *Quantitative Agarose-Gel Electrophoresis of Chromatin - Nucleosome-Dependent Changes in Charge, Shape, and*

Deformability at Low Ionic-Strength. Biochemistry, 1994. **33**(8): p. 2226-2233.

92. Fletcher, T.M., P.S. Serwer, and J.C. Hansen, *Probing the Conformational Dynamics of Nucleosomal Arrays Using Quantitative Agarose-Gel Electrophoresis.* Faseb Journal, 1994. **8**(7): p. A1271-a1271.

93. Arya, G. and T. Schlick, *A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments.* The journal of physical chemistry. A, 2009. **113**: p. 4045-59.

94. Zheng, C.Y. and J.J. Hayes, *Intra- and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system.* Journal of Biological Chemistry, 2003. **278**(26): p. 24217-24224.

95. Schwarz, P.M., et al., *Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains.* Biochemistry, 1996. **35**: p. 4009-4015.

96. Carruthers, L.M. and J.C. Hansen, *The core histone N termini function independently of linker histones during chromatin condensation.* J. Biol. Chem., 2000. **275**: p. 37285-37290.

97. Hizume, K., et al., *Ultramicroscopy Removal of histone tails from nucleosome dissects the physical mechanisms of salt-induced aggregation , linker histone H1-induced compaction , and 30-nm fiber formation of the nucleosome array.* Ultramicroscopy, 2009: p. 1-6.

98. Dorigo, B., *Chromatin Fiber Folding: Requirement for the Histone H4 N-terminal Tail.* Journal of Molecular Biology, 2003. **327**: p. 85-96.

99. Robinson, J.P., et al., *30nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction.* J. Mol. Biol., 2008. **381**: p. 816-825.

100. An, W.J., et al., *Selective requirements for histone H3 and H4N termini in p300-dependent transcriptional activation from chromatin.* Molecular Cell, 2002. **9**(4): p. 811-821.

101. Terrell, A.R., et al., *Reconstitution of nucleosome positioning, remodeling, histone acetylation, and transcriptional activation on the PHO5 promoter.* Journal of Biological Chemistry, 2002. **277**(34): p. 31038-31047.

102. Downs, J.a., *Histone H3 K56 acetylation, chromatin assembly, and the DNA damage checkpoint.* DNA repair, 2008. **7**: p. 2020-4.

103. Luger, K., T.J. Rechsteiner, and T.J. Richmond, *Preparation of nucleosome core particle from recombinant histones.* Methods Enzymol., 1999. **304**: p. 3-19.

104. Wang, H.D., et al., *Glutaraldehyde modified mica: A new surface for atomic force microscopy of chromatin.* Biophysical Journal, 2002. **83**(6): p. 3619-3625.

105. Fragoso, G., et al., *Nucleosome positioning on the MMTV LTR results from the frequency-biased occupancy of multiple frames.* Genes Dev, 1995. **9**: p. 1933-1947.

106. Fragoso, G., et al., *Nucleosome Positioning on the Mmtv Ltr Results from the Frequency-Biased Occupancy of Multiple Frames.* Genes & Development, 1995. **9**(15): p. 1933-1947.

107. Flaus, A., et al., *Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals.* Proc. Natl. Acad. Sci. USA, 1996. **93**: p. 1370-1375.

108. Solis, F.J., et al., *Articles Properties of Nucleosomes in Acetylated Mouse Mammary Tumor Virus versus.* Society, 2007. **46**.

109. Poirier, M.G., et al., *Spontaneous access to DNA target sites in folded chromatin fibers.* Journal of molecular biology, 2008. **379**: p. 772-86.

110. Lebowitz, J., M.S. Lewis, and P. Schuck, *Modern analytical ultracentrifugation in protein science: A tutorial review.* Protein Science, 2002. **11**(9): p. 2067-2079.

111. Schwarz, P.M. and J.C. Hansen, *Formation and stability of higher order chromatin structures.* 1994.

112. Sen, D. and D.M. Crothers, *Condensation of Chromatin - Role of Multivalent Cations.* Biochemistry, 1986. **25**(7): p. 1495-1503.

113. Hansen, J.C., et al., *Homogeneous Reconstituted Oligonucleosomes, Evidence for Salt-Dependent Folding in the Absence of Histone-H1.* Biochemistry, 1989. **28**(23): p. 9129-9136.

114. Guo, X.W. and R.D. Cole, *Chromatin Aggregation Changes Substantially as Ph Varies within the Physiological Range.* Journal of Biological Chemistry, 1989. **264**(20): p. 11653-11657.

115. Georgel, P.T., et al., *Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation.* J. Biol. Chem., 2003. **278**: p. 32181-32188.

116. Verschure, P.J., et al., *In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation.* Molecular and Cellular Biology, 2005. **25**(11): p. 4552-4564.

117. Bertin, A., et al., *H3 and H4 histone tails play a central role in the interactions of recombinant NCPs.* Biophysical journal, 2007. **92**: p. 2633-45.

118. Allan, J., et al., *Participation of Core Histone " Tails " in the Stabilization of the Chromatin Solenoid Characterization of Trypsin-treated SPN.* The Journal of Cell Biology, 1982.

119. Robinson, P.J.J., et al., *30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction.* Journal of Molecular Biology, 2008. **381**(4): p. 816-825.

120. Bednar, J., et al., *Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin.* Proceedings of the National Academy of

Sciences of the United States of America, 1998. **95**(24): p. 14173-14178.

121. Biology, C., *Nucleosomes , linker DNA , and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin.* Cell, 1998. **95**: p. 14173-14178.

122. Robinson, P.J.J. and D. Rhodes, *Structure of the '30 nm' chromatin fibre: A key role for the linker histone.* Current Opinion in Structural Biology, 2006. **16**(3): p. 336-343.

123. Butler, P.J.G. and J.O. Thomas, *Dinucleosomes show compaction by ionic strength, consistent with bending of linker DNA.* Journal of Molecular Biology, 1998. **281**(3): p. 401-407.

124. Green, G.R., et al., *Linker DNA destabilizes condensed chromatin.* Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire, 2001. **79**(3): p. 349-363.

125. Thoma, F., T. Koller, and A. Klug, *Involvement of Histone-H1 in the Organization of the Nucleosome and of the Salt-Dependent Superstructures of Chromatin.* Journal of Cell Biology, 1979. **83**(2): p. 403-427.

126. Finch, J.T. and A. Klug, *Solenoidal model for superstructure in chromatin.* Proc. Nod . Acad. Sci. U S. A, 1976. **73**: p. 1897-190113.

127. Felsen, G., *Nucleosomestructure +12067.* Structure, 1980.

128. Mcghee, J.D., et al., *Orientation of the Nucleosome within the Higher-Order Structure of Chromatin.* Cell, 1980. **22**(1): p. 87-96.

129. Worcel, A., S. Strogatz, and D. Riley, *Structure of Chromatin and the Linking Number of DNA.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1981. **78**(3): p. 1461-1465.

130. Woodcock, C.L.F., L.Y. Frado, and J.B. Rattner, *The Higher-order Structure of Chromatin : Evidence for a Helical Ribbon Arrangement Fibers from Interphase Nuclei and Chromosome.* The Journal of Cell Biology, 1984. **99**.

131. Simpson, R.T., F. Thoma, and J.M. Brubaker, *Chromatin Reconstituted from Tandemly Repeated Cloned DNA Fragments and Core Histones - a Model System for Study of Higher-Order Structure.* Cell, 1985. **42**(3): p. 799-808.

132. Huynh, V.a.T., P.J.J. Robinson, and D. Rhodes, *A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone.* Journal of molecular biology, 2005. **345**: p. 957-68.

133. Widom, J., *A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**: p. 1095-9.

134. Woodcock, C.L., et al., *A Chromatin Folding Model That Incorporates Linker Variability Generates Fibers Resembling the Native Structures.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(19): p. 9021-9025.

135. Grigoryev, S.A., et al., *Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(32): p. 13317-13322.

136. Wang, J.P., et al., *Preferentially Quantized Linker DNA Lengths in Saccharomyces cerevisiae.* Plos Computational Biology, 2008. **4**(9): p. -.

137. Bash, R., et al., *Nucleosomal arrays can be salt-reconstituted on a single-copy MMTV promoter DNA template: Their properties differ in several ways from those of comparable 5S concatameric arrays.* Biochemistry, 2003. **42**(16): p. 4681-4690.

138. Olins, A.L. and D.E. Olins, *Spheroid Chromatin Units (Ru Bodies).* Science, 1974. **183**(4122): p. 330-332.

139. Jerzmanowski, A. and K. Staron, *Mg-2+ as a Trigger of Condensation-Decondensation Transition of Chromatin during Mitosis.* Journal of Theoretical Biology, 1980. **82**(1): p. 41-46.

140. Smith, M.F., et al., *Radial Density Distribution of Chromatin - Evidence That Chromatin Fibers Have Solid Centers.* Journal of Cell Biology, 1990. **110**(2): p. 245-254.

141. Fan, Y., et al., *H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo.* Mol. Cell. Biol, 2003. **23**: p. 4559-4572.

142. Fan, Y., *Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation.* Cell, 2005. **123**: p. 1199-1212.

143. Perico, A., G. La Penna, and L. Arcesi, *Electrostatic interactions with histone tails may bend linker DNA in chromatin.* Biopolymers, 2006. **81**(1): p. 20-28.

144. Woodcock, C.L., et al., *A chromatin folding model that incorporates linker variability generates fibers resembling the native structures.* Proc. Natl. Acad. Sci. USA, 1993. **90**: p. 9021-9025.

145. Engeholm, M., et al., *Nucleosomes can invade DNA territories occupied by their neighbors.* Nature Structural & Molecular Biology, 2009. **16**: p. 151-158.

146. Müller, H.J., *The remaking of chromosomes.* Collecting Net, 1938. **13**: p. 181-198.

147. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes & development, 2005. **19**: p. 2100-10.

148. Watson, J.D., *Origin of Concatemeric T7 DNA.* Nature-New Biology, 1972. **239**(94): p. 197-&.

149. Olovniko.Am, *Principle of Marginotomy in Template Synthesis of Polynucleotides.* Doklady Akademii Nauk Sssr, 1971. **201**(6): p. 1496-&.

150. Olovniko.Am, *Theory of Marginotomy - Incomplete Copying of Template Margin in Enzymic-Synthesis of Polynucleotides and Biological Significance of Phenomenon.* Journal of Theoretical Biology, 1973. **41**(1): p. 181-190.

151. Greider, C.W. and E.H. Blackburn, *Identification of a Specific Telomere Terminal Transferase-Activity in Tetrahymena Extracts.* Cell, 1985. **43**(2): p. 405-413.

152. Bedoyan, J.K., et al., *Condensation of rat telomere-specific nucleosomal arrays containing unusually short DNA repeats and histone.* H1, J. Biol. Chem., 1996. **271**: p. 18485-18493.

153. Makarov, V.L., et al., *Nucleosomal Organization of Telomere-Specific Chromatin in Rat.* Cell, 1993. **73**(4): p. 775-787.

154. Murga, M., et al., *Global chromatin compaction limits the strength of the DNA damage response.* The Journal of cell biology, 2007. **178**: p. 1101-8.

155. García-Cao, M., et al., *Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases.* Nature genetics, 2004. **36**: p. 94-9.

156. McClintock, B., *The stability of broken ends of chromosomes in Zea mays.* Genetics, 1941. **26**: p. 234-282.

157. Karlseder, J., A. Smogorzewska, and T. de Lange, *Senescence induced by altered telomere state, not telomere loss.* Science, 2002. **295**(5564): p. 2446-2449.

158. van Steensel, B., A. Smogorzewska, and T. de Lange, *TRF2 protects human telomeres from end-to-end fusions.* Cell, 1998. **92**(3): p. 401-413.

159. Karlseder, J., *p53- and ATM-Dependent Apoptosis Induced by Telomeres Lacking TRF2.* Science, 1999. **283**: p. 1321-1325.

160. Karlseder, J., A. Smogorzewska, and T. de Lange, *Senescence induced by altered telomere state, not telomere loss.* Science (New York, N.Y.), 2002. **295**: p. 2446-9.

161. Stansel, R.M., T.D. Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3 0 telomeric overhang.* EMBO J, 2001. **20**: p. 5532-5540.

162. Yoshimura, S.H., et al., *Molecular mechanisms of DNA end-loop formation by.* TRF2, Genes Cells, 2004. **9**: p. 205-218.

163. Court, R., et al., *How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures (vol 6, pg 39, 2005).* Embo Reports, 2005. **6**(2): p. 191-191.

164. Konishi, A. and T.D. Lange, *Cell cycle control of telomere protection and NHEJ revealed by a ts mutation in the DNA-binding domain of TRF2.* Genes & Development, 2008: p. 1221-1230.

165. Arents, G. and E.N. Moudrianakis, *Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA.* 1993.

166. Lohr, D., et al., *Using atomic force microscopy to study chromatin structure and nucleosome remodeling.* Methods, 2007. **41**: p. 333-341.

167. Segal, E. and J. Widom, *What controls nucleosome positions?* Trends in genetics : TIG, 2009. **25**: p. 335-43.

168. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast.* Cell, 2005. **122**: p. 517-527.

169. Lantermann, A., et al., *Genome-wide mapping of nucleosome positions in Schizosaccharomyces pombe.* Methods (San Diego, Calif.), 2009. **48**: p. 218-25.

170. Bash, R., et al., *AFM imaging of protein movements: histone H2A-H2B release during nucleosome remodeling.* FEBS letters, 2006. **580**: p. 4757-61.

171. Roberts, M.S., G. Fragoso, and G.L. Hager, *Nucleosomes Reconstituted in Vitro on Mouse Mammary Tumor Virus B Region DNA Occupy Multiple Translational and Rotational Frames.* Cell, 1995: p. 12470-12480.

172. Mandelkern, M., et al., *The Dimensions of DNA in Solution.* Journal of Molecular Biology, 1981. **152**(1): p. 153-161.

173. Shogren-Knaak, M. and C.L. Peterson, *Switching on chromatin - Mechanistic role of histone H4-K16 acetylation.* Cell Cycle, 2006. **5**(13): p. 1361-1365.

174. Dalal, Y., et al., *Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells.* Plos Biology, 2007. **5**(8): p. 1798-1809.

175. Wang, H.D., et al., *Single-epitope recognition imaging of native chromatin.* Epigenetics & Chromatin, 2008. **1**: p. -.

APPENDIX A

PRIMERS FOR MONOMER LIBRARIES' CONSTRUCITON

Table A-1 Forward and Reverse Primers for 30 bp library construction

Primers Forward	
177-30-27-11_F	tcccgagttcaatacatggggcgggatcacaggatgtatatatctgacac
177-30-22-16_F	tcccgagttcaatacatggggccacaggatgtatatatctgacac
177-30-17-21_F	tcccgagttcaatacatcacaggatgtatatatctgacac
177-30-12-26_F	tcccgagttcaacacaggatgtatatatctgacac
177-30-7-31_F	tcccgagcacaggatgtatatatctgacac
Primers Reverse	
177-30-27-11_R	cctcgggatgctggagaatcccggtgccga
177-30-22-16_R	cctcgggatgcatccctggagaatcccggtgccga
177-30-17-21_R	cctcgggatgcatcccgccccctggagaatcccggtgccga
177-30-12-26_R	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
177-30-7-31_R	cctcgggatgcatcccgccccatgtattgaatggagaatcccggtgccga

Table A-2 Forward and Reverse Primers for 40 bp library construction

Forward Primer	
187-40-35-5-F	tcccgagtcgctgttcattcaatacatggggcgggat
187-40-30-10-F	tcccgagtcgcttttcaatacatggggcgggat
177-30-27-11_F	tcccgagttcaatacatggggcgggatcacaggatgtatatatctgacac
177-30-22-16_F	tcccgagttcaatacatggggccacaggatgtatatatctgacac
177-30-17-21_F	tcccgagttcaatacatcacaggatgtatatatctgacac
177-30-12-26_F	tcccgagttcaacacaggatgtatatatctgacac
177-30-7-31_F	tcccgagcacaggatgtatatatctgacac
Reverse Primer	
177-30-27-11_R	cctcgggatgctggagaatcccggtgccga
177-30-22-16_R	cctcgggatgcatccctggagaatcccggtgccga
177-30-17-21_R	cctcgggatgcatcccgccccctggagaatcccggtgccga
177-30-12-26_R	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
177-30-7-31_R	cctcgggatgcatcccgccccatgtattgaatggagaatcccggtgccga
187-40-10-30-R	cctcgggagcgaatgcatcccgccccatgta
187-40-5-35-R	cctcgggtgaacagcgaatgcatcccgccccatgta

Table A-3 Forward and Reverse Primers for 50 bp library construction

Forward Primer	
197-50-45-5-F	tcccgagtatagggtcctcgctgttcattcaatacat
197-50-40-10-F	tcccgagggtcctcgctgttcattcaatacat
187-40-35-5-F	tcccgagtcgctgttcattcaatacatggggcgggat
187-40-30-10-F	tcccgagtcgcttcaatacatggggcgggat
177-30-27-11_F	tcccgagttcaatacatggggcgggatcacaggatgtatatatctgacac
177-30-22-16_F	tcccgagttcaatacatggggccacaggatgtatatatctgacac
177-30-17-21_F	tcccgagttcaatacatcacaggatgtatatatctgacac
177-30-12-26_F	tcccgagttcaacacaggatgtatatatctgacac
177-30-7-31_F	tcccgagcacaggatgtatatatctgacac
Reverse Primer	
177-30-27-11_R	cctcgggatgctggagaatcccggtgccga
177-30-22-16_R	cctcgggatgcatccctggagaatcccggtgccga
177-30-17-21_R	cctcgggatgcatcccgccccctggagaatcccggtgccga
177-30-12-26_R	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
177-30-7-31_R	cctcgggatgcatcccgccccatgtattgaatggagaatcccggtgccga
187-40-10-30-R	cctcgggagcgaatgcatcccgccccatgta
187-40-5-35-R	cctcgggtgaacagcgaatgcatcccgccccatgta
197-50-10-40-R	cctcgggctatatgaacagcgaatgcat
197-50-5-45-R	cctcgggggaccctatatgaacagcgaatgcatcccg

Table A-4 Forward and Reverse Primers for 60 bp library construction

Forward Primer	
207-60-55-5-F	tcccgagatcacataactatagggtcctcgctgttca
207-60-50-10-F	tcccgagataactatagggtcctcgctgttca
197-50-45-5-F	tcccgagtatagggtcctcgctgttcattcaatacat
197-50-40-10-F	tcccgaggtcctcgctgttcattcaatacat
187-40-35-5-F	tcccgagtcgctgttcattcaatacatggggcgggat
187-40-30-10-F	tcccgagtcgcttcaatacatggggcgggat
177-30-27-11_F	tcccgagttcaatacatggggcgggatcacaggatgtatatatctgacac
177-30-22-16_F	tcccgagttcaatacatggggccacaggatgtatatatctgacac
177-30-17-21_F	tcccgagttcaatacatcacaggatgtatatatctgacac
177-30-12-26_F	tcccgagttcaacacaggatgtatatatctgacac
177-30-7-31_F	tcccgagcacaggatgtatatatctgacac
Reverse Primer	
177-30-27-11_R	cctcgggatgctggagaatcccggtgccga
177-30-22-16_R	cctcgggatgcatcccctggagaatcccggtgccga
177-30-17-21_R	cctcgggatgcatcccgccccctggagaatcccggtgccga
177-30-12-26_R	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
177-30-7-31_R	cctcgggatgcatcccgccccatgtattgaatggagaatcccggtgccga
187-40-10-30-R	cctcgggagcgaatgcatcccgccccatgta
187-40-5-35-R	cctcgggtgaacagcgaatgcatcccgccccatgta
197-50-10-40-R	cctcgggctatatgaacagcgaatgcat
197-50-5-45-R	cctcgggggaccctatatgaacagcgaatgcatcccg
207-60-10-50-R	cctcggggttatggaccctatatgaacagcga
207-60-5-55-R	cctcggggttatgtgatggaccctatatgaacagcga

APPENDIX B

PRIMERS FOR MONOMER CONSTRUCTION WITH LINKER LENGTH OF 31 TO 39 BASE PAIRS

Table B-1 Primers for all the 31 to 39 linker DNA length Library

	Forward Primers	
1	16-15	tcccgagttcaatacatg **cacaggatgtatatatctg**
2	17-15	tcccgagttcaatacatgg **cacaggatgtatatatctg**
3	18-15	tcccgagttcaatacatggg **cacaggatgtatatatctg**
4	19-15	tcccgagttcaatacatgggg **cacaggatgtatatatctg**
5	20-15	tcccgagttcaatacatggggc **cacaggatgtatatatctg**
6	16-20	tcccgagttcaatacatg **cacaggatgtatatatctg**
7	17-20	tcccgagttcaatacatgg **cacaggatgtatatatctg**
8	18-20	tcccgagttcaatacatggg **cacaggatgtatatatctg**
9	19-20	tcccgagttcaatacatgggg **cacaggatgtatatatctg**
	Reverse Primers	
1	15	cctcgggatgcatcccgccccctggagaatcccggtgccga
2	15	cctcgggatgcatcccgccccctggagaatcccggtgccga
3	15	cctcgggatgcatcccgccccctggagaatcccggtgccga
4	15	cctcgggatgcatcccgccccctggagaatcccggtgccga
5	15	cctcgggatgcatcccgccccctggagaatcccggtgccga
6	20	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
7	20	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
8	20	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga
9	20	cctcgggatgcatcccgccccatgtatggagaatcccggtgccga

Table B-2 The designed sequences for 31 to 39 linker DNA library

linker	**sequence**
31	tcccgagttcaatacatg -601-ggggcgggatgcatcccgagg
32	tcccgagttcaatacatgg -601-ggggcgggatgcatcccgagg
33	tcccgagttcaatacatggg -601-ggggcgggatgcatcccgagg
34	tcccgagttcaatacatgggg -601-ggggcgggatgcatcccgagg
35	tcccgagttcaatacatggggc -601-ggggcgggatgcatcccgagg
36	tcccgagttcaatacatg -601-tacatggggcgggatgcatcccgagg
37	tcccgagttcaatacatgg -601-tacatggggcgggatgcatcccgagg
38	tcccgagttcaatacatggg -601-tacatggggcgggatgcatcccgagg
39	tcccgagttcaatacatgggg -601-tacatggggcgggatgcatcccgagg

APPENDIX C

MATLAB AND LABVIEW PROGRAM FOR CHROMATIN ANALYSIS

C-1. MatLab Programming for Nucleosome Array Image Analysis

In order to study the 2D features of the nucleosome array in the mica surface, for example, how the nucleosome interact with its neighbors, how the linker DNA length affects the twisting angles of the nucleosome, the contour length and persistence length of the nucleosome array. It will be informative to get the contour trace of the nucleosome array and analyze the inter-nucleosome distances as well as the DNA entering and exiting angles in the nucleosome.

We used to use the chromatin analysis platform for the nucleosome array data analysis. However, the chromatin analysis platform is limited in several ways, for example, no digitalized path information could be obtained, angle calculation is not convenient, no tool boxes in imaging analysis, and etc. Instead, we wrote another program based on MatLab for the 2D nucleosome array data analysis with the help of Dr. Dan Grilley (Northwestern University). In this chapter, I will briefly introduce how to use the program for the data analysis.

The whole program is based on Matlab object-oriented programming, and all script files are stored in a folder called @afm. '@' is a sign for object programming in Matlab, so it is required to present in front of the folder name. A main file called 'afm.m' file is inside the '@afm' folder, this file defines the new object 'afm', which is corresponded to the afm image file. Several properties were defined in this file, which can be further modified.

Step1: Read in files.

To start analyze AFM results, an image file with 'TIF' type is required, other image formats will also be supported, but need tiny modifications in the program. An ‘afm’ object is created by reading in an image file with command: ‘dna=afm (filename) ’. Here, the newly created object has a name of ‘dna’ and a type as ‘afm’. A message asking for the scan size, which is used for the image calibration, should be displayed out. The following things are to do some basic image treatment and pre-analysis. In our case, if all the afm images have already been processed with Gwyddion (www.gwyddion.net), the image filtering process is unnecessary. Otherwise, the image is required to trim out some background noise. In order to analyze the imaging in a more accurate way, the size of the image is required to be sized by a bicubic interpolation to 2048x2048 pixels or even larger, which depends on the resolution of the images.

Step2: Tracing the contour along the nucleosome array.

The script is stored in the 'getStarts_ma.m' file. With the newly created ‘afm’ object, the 'getStarts_ma' file analyzes the contour tracing along the nucleosome molecule by the following steps:

1. Press ‘N’ to start a new molecule. Start clicking mouse to pick points at the end of the nucleosome array. If the nucleosome is terminated with a nucleosome, the start point is defined in the center of the terminal nucleosome; otherwise the points start from the ends of the identifiable DNA.

Nucleosome centers were identified by press ‘space key’ in the keyboard, which marked as a blue circle (Figure 6-1).

2. The right button of the mouse is used for unselect the last point;

3. After picking the last point along the molecule, you can finish the selection of this molecule by pressing the middle button of the mouse. A new molecule with path information will be recorded.

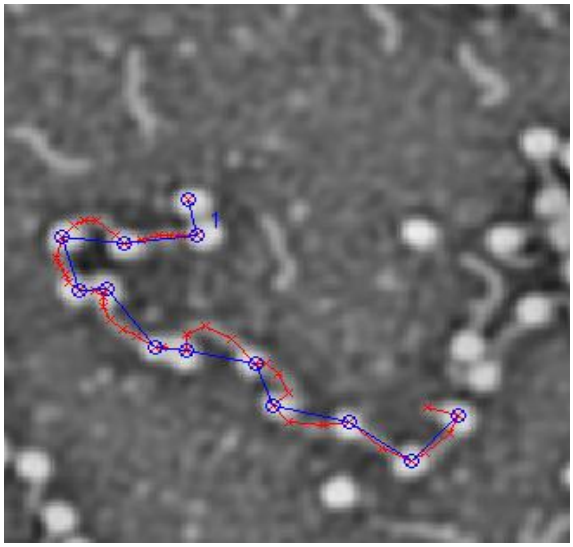

Figure C-9 An example of imaging analysis by MatLab

Red line indicates the contour trace along the molecule, blue line indicates the center to center tracing, and blue circle indicates the center of each nucleosome.

All the points along the molecule (Figure 6-1) are stored in the 'dna.starts.trackPntsWithTag' structure. Since we would like to measure the DNA entering and exiting angles along the molecule, identifiable position of linker DNA entering and exiting sites are picked by mouse. If two nucleosomes are in close contact to each other and the linker DNA cannot be

identified in the image, the nucleosome center to center angles would be calculated instead of DNA entering and exiting angles (Figure 6-2).

Step3: Data Analysis.

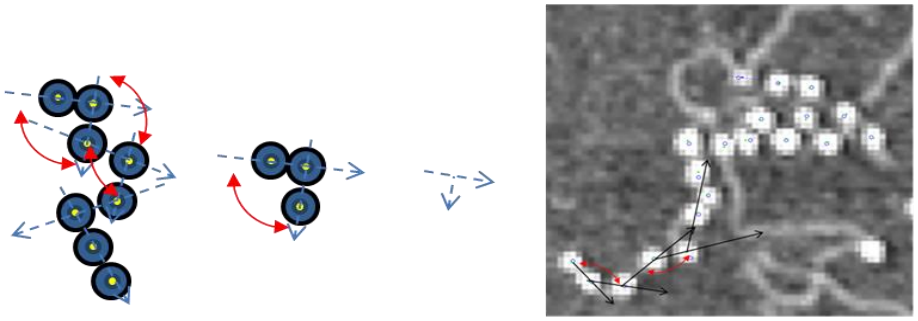

Figure C-10 Schematic illustration of the angle calculation in the MatLab program

Here, the figure shows the definition of nucleosome center to center angles: (blue) dash lines indicate the nucleosome array's direction and red lines with arrows indicate the angle measured in the program. The angles measured here have a range from 0° to 180°.

For every non-terminal nucleosome, three features are measured: the contour length to the left neighbor nucleosome; the contour length to the right neighbor nucleosome; and the angle between the DNA entering and exiting sites. The scheme of the angle calculation is shown in Figure 6-2. Due to the resolution of the AFM imaging, the angle calculated here cannot exactly reveal the DNA wrapping behavior along the nucleosome template. However, this feature is informative in understanding the arrangement of the

nucleosomes along the templates in HEPES buffer. The command for this step is: ‘dna=dna.angleDistanceCal’.

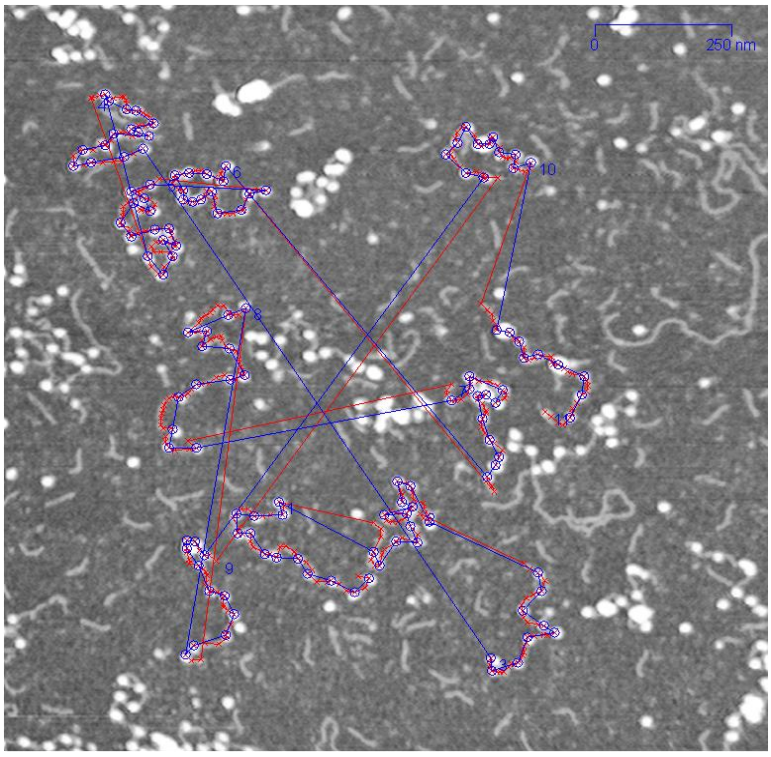

Figure C-11 An example of the selected molecules by the Matlab program The red line indicates the contour trace along the molecule, the blue line indicates the center to center tracing, and the blue circle indicates the center of each nucleosome.

Step 4 Data presentation

In order to study the distribution of the inter-nucleosome distances and also the DNA entering and exiting angles, the 1D and 2D histogram analyses are

included in the 'plotDistAngle.m' file. All the nucleosome contour data and nucleosome center to center data can be analyzed (Figure 6-4). The command for this step is: 'dna=dna.plotDistAngle'.

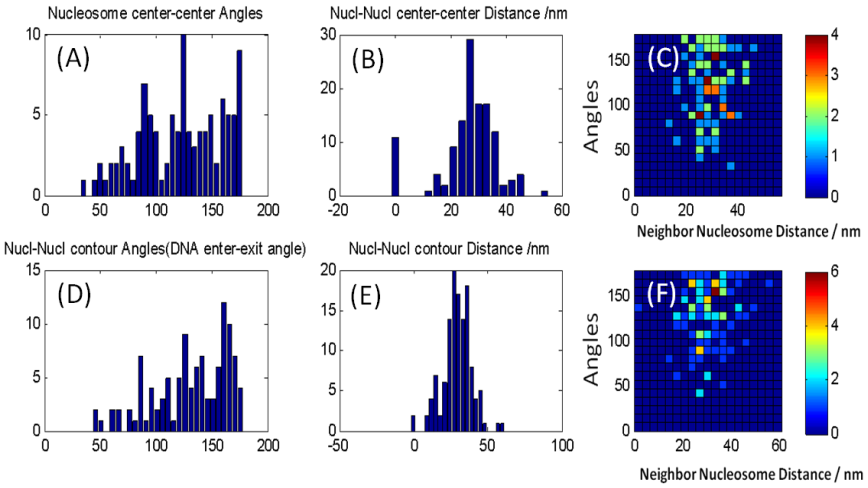

Figure C-12 Nucleosome array data analysed by the MatLab program Histogram of nucleosome center-center angles (A), DNA entering-exiting angles (D), center to center distances (B), and contour distances (E) are shown here. (C) and (F) are the 2D histogram of the center-center measurements and contour measurements of linker DNA length and nucleosome-nucleosome angles.

C-2. Labview Program for Nucleosome Positioning Analysis

In order to study the positioning ability of the histone octamer along the DNA templates, a labview program was written to analyzing the data generated from the chromatin analysis platform or from the matlab. The data type for the Labview program analysis is based on the chromatin analysis platform.

Basically, the contour information of the nucleosome array was organized as: number of loading, end-end contour distance, and inter nucleosome distances (Table 6-1). The steps to analyze the nucleosome positioning are listed in the following.

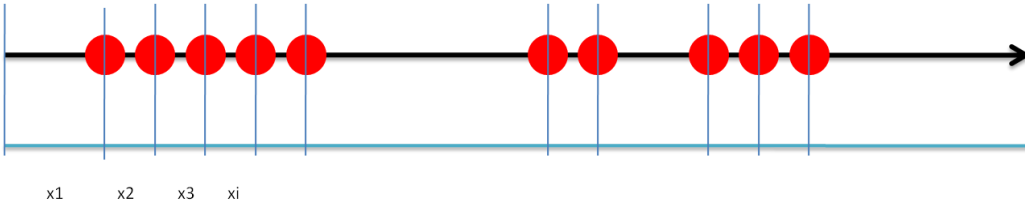

Figure C-13 Illustration of the nucleosome spacing data by chromatin analysis platform

The cartoon here represents the chromatin analysis platform data type, the program traces the start of each molecule, and get all the contour distances between each nucleosome, and also the distances between the first and last nucleosome to the start or the end of the molecules. Here, such as x1, x2 and x3.

To start using the labview program, the contour information of each nucleosome array should saved as a 'txt' file, the data should be organized as a 2D matrix shown in table 6-1. Start running the labview program by short cut 'Ctrl+r', and press the 'read file' button to read in selected '*.txt' file.

Table C-1 Data Saved Format Example

loading	end-end contour	interdistance1	interdistance2	…
2	2330	…	…	…

After finishing reading in the data, the program firstly searches for all the terminal nucleosomes, and the number of the terminal nucleosome is stored for each individual molecule. The input data is sorted as: loading number, contourlength and SegArray; the output data is sorted as: loading number, contourlength, segArrays and number of terminal nucleosomes in the labview program. Next, the contour lengths were corrected based on the number of terminal nucleosomes and the wrapping length can be derived from the plot of contour distance vs. the loadings (Shown as in equation 5-1). With the assumption that the dyad position of each nucleosome along the DNA template doesn't change, and the DNA wrapping length is evenly contributed from the nucleosomal DNA around the dyad location, the locations of each nucleosome's dyad positions along the template can be identified by using the calculated wrapping length and the measured contour information. Figure 6-6 shows the options for the positioning analysis. The positioning behavior under different loadings can be analyzed separately, and the data can be normalized to the DNA length in the unit of base pair.

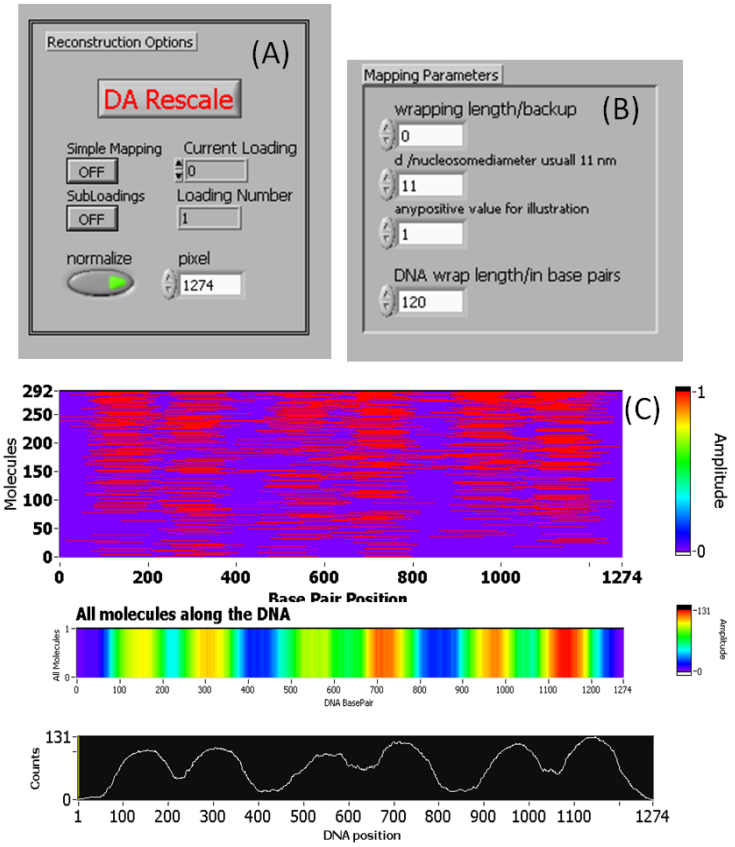

Figure C-14 Nucleosome positioning analysis by Labview

(A) Options for the data analysis, (B), Parameters for the nucleosome positioning analysis; (C) shows the mapping result of the positioning analysis along the templates, the upper panel shows the distribution of each individual nucleosome, and bottom shows the overall positioning counts along the templates.

BIOGRAPHICAL SKETCH

Education

Ph.D. in Chemistry (2005-2010), Arizona State University, Tempe, AZ
Area focused: Chromatin structure
Advisor: Prof. Stuart Lindsay

B.S. in Chemistry (2001-2005), Fudan University, Shanghai, China,
Area focused: Surface chemistry & Catalysis
Research Advisor: Prof. Minghua Qiao

Publications

A non-noble amorphous Co–Fe–B catalyst highly selective in liquid phase hydrogenation of crotonaldehyde to crotyl alcohol. Yan Pei, Jianqiang Wang, Qiang Fu, Pingjun Guo, Minghua Qiao, Shirun Yan and Kangnian Fan, New J. Chem., 2005, **29**(8), 992

Electrochemical origin of voltage-controlled molecular conductance switching He, J., Fu, Q., Lindsay, S., Ciszek, J. W. & Tour, J. M. Journal of the American Chemical Society 128, 14828-14835 (2006)

Transverse tunneling through DNA hydrogen bonded to an electrode. He, J. Lin, L. S. Zhang, P. M. Spadola, Q. Xi, Z. Q. Fu, Q. Lindsay, S. Nano Letters 8, 2530-2534 (2008).

Chromatin Stability at Low Concentration Depends on Histone Octamer Saturation Levels. Hagerman, T. A. Fu, Q. Molinie, B.Denvir, J.Lindsay, S. Georgel, P. T. Biophysical Journal 96, 1944-1951 (2009).

The Myb/SANT domain of the telomere-binding protein TRF2 alters chromatin structure. Baker, A. M., Fu, Q., Hayward, W., Lindsay, S. M. & Fletcher, T. M. Nucleic Acids Research 37, 5019-5031 (2009).

Recognition Imaging of Acetylated Chromatin Using a DNA Aptamer. Lin, L. Y. Fu, Q. Williams, B. A. R.Azzaz, A. M.Shogren-Knaak, M. A. Chaput, J. C. Lindsay, S. Biophysical Journal 97, 1804-1807 (2009).

Photonic approach to the selective inactivation of viruses with a near-infrared subpicosecond fiber laser Tsen, K. T., Tsen, S. W., Fu, Q., Lindsay, S. M.,Kibler, K.,Jacobs, B.,Wu, T. C.,Karanam, B.,Jagu, S.,Roden, R. B. ,S.Hung, C. F.,Sankey, O. F.,Ramakrishna, B.,Kiang, J. G. Journal of Biomedical Optics 14,6 2009 (Highlighted in Nature hotonics news & views, vol4 March 2010)