DISSERTATION

NUCLEOSOMAL ARRAY CONDENSATION: NEW INSIGHTS INTO AN OLD "TAIL"

Submitted by

Troy C. Sorensen

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2011

Doctoral Committee:

Advisor: Jeffrey C. Hansen

Laurie Stargell
Karolin Luger
Susan Bailey

ABSTRACT

NUCLEOSOMAL ARRAY CONDENSATION: NEW INSIGHTS INTO AN OLD "TAIL"

The DNA present within the nucleus of each human somatic cell, when extended end to end, would span a distance of about one meter. The first level of compaction critical to fitting the entire genome into the nucleus is the nucleosome, consisting of ~147 base pairs of DNA wrapped 1.7 times around an octameric structure composed of the four core histones H2A, H2B, H3 and H4. Nucleosomes separated by up to 80 base pairs of linker DNA called nucleosomal arrays compact the DNA further through short range intra-array and long range inter-array contacts that generate different levels of higher order condensed structures. This dissertation investigates the involvement of the core histone "tail" domains as well as the influence of the H3 centromeric variant CENP-A in nucleosomal array condensation events.

In vitro, 12-mer nucleosomal arrays condense intra- and inter-molecularly through nucleosome-nucleosome interactions driven primarily by the core histone tail. This dissertation details the contributions and the molecular determinants of the histone tail domains to the condensation processes. Importantly, we found that the H3 and H4 tail domains were the largest contributors to array condensation. The mode of action used by the H4 tail domain in intra- and intermolecular condensation centered on the following

determinants: 1) position of the H4 tail, 2) amino acid composition, 3) positive charge density and 4) tail domain length. Importantly, the primary sequence of the H4 tail was found to not be an important molecular determinant.

To date no study has been performed to determine short-range compaction between "bulk" H3 containing and H3 centromeric specific variant, CENP-A chromatin. 12-mer nucleosomal arrays containing either H3 or CENP-A histones were reconstituted and tested for their ability to fold intra-molecularly. Major finding include that CENP-A containing nucleosomal arrays assemble in the same stepwise manner as conical arrays and were always more compact than H3 containing arrays at every salt concentration tested. The increased compaction was found to be in part due to a lysine to arginine mutation at position 49 of CENP-A.

ACKNOWLEDGEMENTS

First I would like to thank my advisor, Jeff Hansen. Jeff has continually encouraged me to become a “thinker” and not just a “doer.” Secondly, I would like to thank my committee for keeping me focused and on the straight and narrow throughout my doctoral work. I want to especially thank my wife, Billi, who at times endured more hardships due to my doctoral research then when I was in the deserts of Iraq. Thanks to my son, Trydon, whose constant and infectious smiles and giggles gave me the inspiration to finish this dissertation. I love you Buddy! I would like to thank my mother, Linda, for giving me the opportunity to succeed in life, for teaching me to always be a better person and to always lend a helping hand to those in need. This work is dedicated to my late father, Kevin C. Sorensen. He missed out on almost my entire life, yet I have felt his guidance and presence every step of the way.

TABLE OF CONTENTS

Chapter 1

Literature Review

1.1 Introduction

The DNA present within the nucleus of each human somatic cell, when extended end to end, would span a distance of about one meter. The first level of compaction critical to fitting the entire genome into the nucleus is the nucleosome, consisting of ~147 base pairs of DNA wrapped 1.7 times around an octameric structure composed of the four core histones H2A, H2B, H3 and H4 (Luger *et al*, 1997a,b; Davey *et al*, 2002; Vasudevan *et al*, 2010). Stretches of polynucleosomes separated by up to 80 base pairs of linker DNA called nucleosomal arrays are critical components for further compaction. Short range intra-array and long range inter-array contacts that generate different levels of higher order condensed structures (Fig. 1.1 and Hansen, 2002). Linker histones bound to the nucleosome form what is commonly referred to as the chromatin fiber. The second level of compaction is driven by short range nucleosome-nucleosome interactions with or without linker histones and is commonly called the 30 nanometer diameter fiber. The structure (and in fact the in vivo relevevance) of the 30 nm fiber is very dynamic and highly contested (Schalch *et al*, 2005; Tremethick, 2007; Eltsov *et al*, 2008; Grigoryev *et al*, 2009). Finally, with the assistance of chromatin associated proteins, scaffolding proteins, multi-valent cations and/or polyamines, long range fiber-fiber interactions generate higher order tertiary chromatin structures (Hansen, 2002; Szerlong and Hansen, 2011). The size of these fibers can range from

100 to 300 nanometer diameter “chromonema” fibers to the metaphase chromosome. Although chromatin higher order structures are necessary to compact the genome into the nucleus, DNA accessibility is paramount in maintaining normal cellular functions.

Understanding the underlying mechanisms that drive chromatin condensation is fundamental to how we think about gene expression, DNA repair, cell division and virtually every aspect of the eukaryotic genomic environment including genetic diseases and cancers. This dissertation as a whole focuses on the N-terminal “tail” domains of the four core histones, some of the simplest yet critical components of the condensation process. Importantly, my doctoral research sought to first understand the individual contributions of the core histone tail domains. Secondly, I used specific H4 tail domain mutants to better understand the mode of action of the H4 tail to induce nucleosomal array condensation. I report here that each of the histone tail domains is involved in nucleosomal array condensation to varying degrees, with the H3 and H4 tails being the primary contributors. However, tail position, length, positive charge density and amino acid composition do play critical roles.

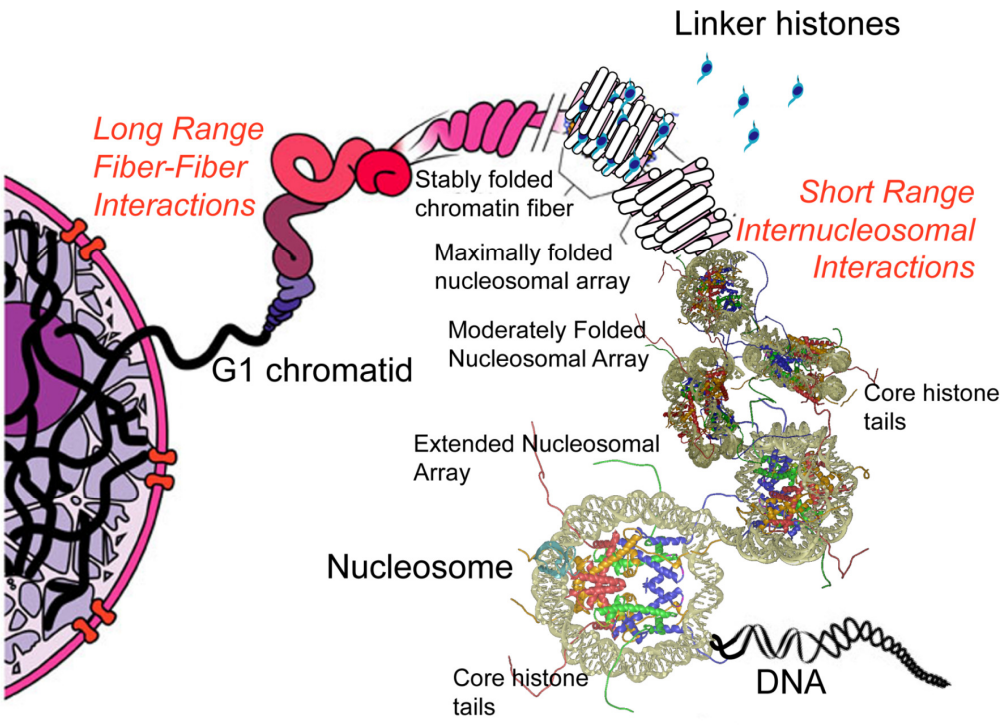

Figure 1.1 Higher order chromatin structures. Schematic detailing the levels of chromatin higher order structures and the known resolution of their structures. Starting from the bottom right of figure, the crystal structure of the nucleosome, DNA = silver, H4 = green, H3 = blue, H2A = yellow, and H2B = red. Eight core histone N-terminal tail domains extend from each nucleosome, and with the aid of cations participate in short range nucleosome-nucleosome interactions that lead to folded nucleosome arrays. One linker histone binds to an individual nucleosome to stabilize the folded chromatin fibers. With or without the linker histones, folded fibers can participate in long-range fiber-fiber interactions leading to higher order chromatin structures. Importantly, higher order chromatin structures are dependent upon the core histone tail domains. (Adapted from Hansen, 2002)

1.2 Chromatin Structure Hierarchy

Starting with the 30 nm chromatin fiber, the structures of higher order chromatin assemblages are unknown, highly contested and probably highly heterogeneous. Over the years numerous models have been proposed for the 30 nm fiber base on different lines of research. Purified nuclear chromatin and *in vitro* reconstituted nucleosomal arrays are highly condensed in ionic strengths similar to what is found in the nucleus (Hansen *et al*, 1989; Garcia-Rameriz, *et al*, 1992; Schwarz *et al*, 1996). Due to this observation, the 30 nm fiber is likely the smallest condensed chromatin structure that trans acting factors have access to perform DNA replication, repair, transcription of other DNA metabolic processes. Because of this connection, a vast amount of time and effort has been put into determining the structure of the 30 nm fiber.

Two competing models for the 30-nm fiber have dominated all other models in the past decades. A 30-nm one-start solenoid helix where nucleosomes interact with their nearest neighbor was first proposed in 1976. (Finch and Klug, 1976). More recent data support an interdigitated solenoid model where nucleosomes interact with their nearest neighbor and 5 and 6 nucleosomes away leading to bent linker DNA at the center of the fiber (Fig. 1.2; Robinson *et al*, 2006 and Kruithof *et al*, 2009). In comparison to the one-start solenoid models, the two start helical model predicts that nucleosomes n and n+/-2 interact with one another leading to a zig-zag conformation and straight linker DNA. Perhaps the most striking evidence in favor of the two-start helical model comes from cross-linking studies of 12-mer nucleosomal arrays and crystallization of the tetranucleosome (Fig. 1.2; Dorigo *et al*, 2004 and Schalch *et al*, 2005). Recently, a new "heteromorphic" chromatin fiber model has been proposed that incorporates elements of both models (Grigoryev *et al*, 2009). The results of this publication show that straight (n and n+/-2) linker DNA dominates a compacted fiber, but that bent (n and n+/-1, 5, 6)

linker DNA is present in the 30-nm fiber (Grigoryev *et al*, 2009). Furthermore, computer modeling supports this *in vitro* data and predicts that fibers containing both "types" of nucleosome-nucleosome interaction generate a more condensed structure (Grigoryev *et al*, 2009).

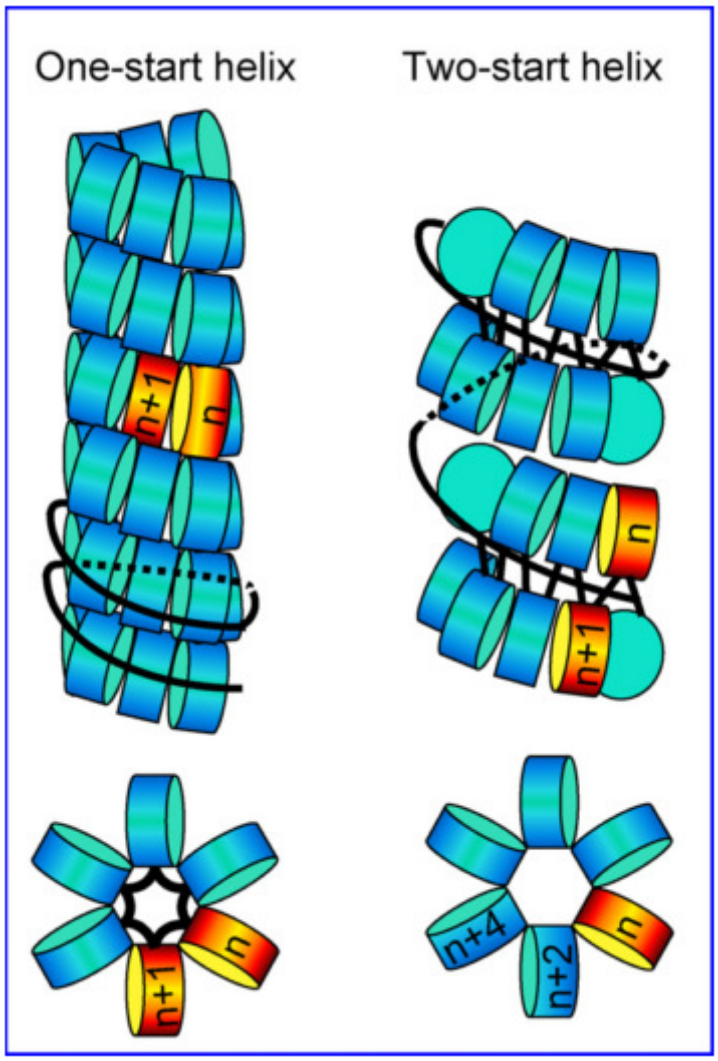

Figure 1.2. Two competing models for the 30 nm chromatin fiber. Top left, one-start helical model depicting the probable nearest neighbor nucleosomes within a folded fiber. The two-start helical model of the 30 nm chromatin fiber consists of straight linker DNA at the interior of the fiber and nucleosomes n and n+/- 2 interacting with one another. Image adapted from van Holde and Zlatanova, 2007.

However, many of these studies have used a high affinity nucleosome positioning sequence that generates a homogenous nucleosome repeat length that is likely not found in nature. An array of variably spaced nucleosomes compacted into a heterogeneous higher order chromatin structures is the most likely scenario found within the nucleus.

Chromosomes were first described in the literature in the early 19th century using light microscopy. By the 1950s it was understood that the condensed state of chromatin within the nucleus was readily dependent upon the ionic environment including the pH of the solution. With the use of phase-light microscopy, two states of chromatin were readily observed in isolated eukaryotic nuclei depending on the ionic strength of the solution (Olins and Olins, 1972). Nuclei suspended in a neutral pH buffer without divalent cations appeared swollen and homogenous, whereas nuclei suspended in the presence of sufficient mono and divalent cations or at a low pH appeared granulated and compacted due to chromatin condensation (Olins and Olins, 1972). Initial chromatin purification protocols easily separated the soluble chromatin fibers from the granulated "precipitated" chromatin, and therefore the soluble fraction was more readily studied.

The late 1960s and 1970s not only brought a clearer image of the repeating unit of chromatin (the nucleosome), but also a more basic understanding of the determinants that drive chromatin condensation into higher order structures. Electron micrographs of purified nuclei and chromatin provided the first "look" at the sub-chromosomal chromatin structures that are more readily found throughout the cellular life cycle. These images showed that linker histones stabilized the soon to be called "30 nm fiber" but were not required, that condensed chromatin was highly dependent on charge neutralization from mono and divalent cations, and removal of cations would reverse the condensation leading to a swollen and homogenous looking nucleus (Olins and Olins, 1972; Thoma and Koller, 1977, Thoma *et al*, 1979; Butler and Thomas, 1980; Woodcock et al, 1984;

van Holde, 1988; Lu *et al*, 2006). In hindsight these studies showed that the differences between soluble and precipitated forms of chromatin were actually differences in soluble chromatin and very large soluble chromatin that pelleted easily under centrifugal force.

1.3 The Nucleosome: The repeating unit of Chromatin.

Only in the past fifteen years has a detailed structure of the nucleosome, the basic repeating unit of chromatin been known (Luger *et al*, 1997a). Almost forty years ago the first of a repeating, nucleoprotein subunit of chromatin was published (Hewish and Burgoyne, 1973; Kornberg, 1974). Roger Kornberg first proposed a H3/H4 tetramer as the central repeating unit of chromatin, eventually different research groups have crystalized progressively higher resolution structures of the histone octamer and the nucleosome (Kornberg and Thomas, 1974; Kornberg 1974; Richmond *et al*, 1984; Arents and Moudrianakis, 1991). However, it wasn't until 1997 that a high resolution x-ray crystallographic structure of the nucleosome was resolved, providing a wealth of insight into the basic repeating unit of chromatin unit and a whole new era in chromatin research (Luger *et al*, 1997a).

The canonical nucleosome consists of 147 base pairs of DNA wrapped 1.7 times around a histone octamer consisting of two H2A/H2B heterodimers positioned on either side of a H3/H4 heterotetramer (Luger *et al*, 1997a). Nucleosomal DNA wraps around the circumference of the highly basic core histone octamer. Rich in arginines and lysines, the histone octamer provides multiple contacts with the DNA in which an arginine intercalates into the nucleosomal DNA minor groove to help stabilize the superhelical bend in its path around the histone octamer (Luger *et al*, 1997a,b; Luger and Richmond, 1998). The surface of the nucleosome consists of basic, acidic and hydrophobic patches. A topological view of the nucleosome surface shows regions of valleys and peaks, providing attractive protein-protein and protein-DNA interaction sites

in cis and trans (Luger *et al*, 1997a). However, it is what is not seen in the crystal structure of the nucleosome (the histone N-terminal tail domains) that is the primary determinant of chromatin condensation and undoubtedly remains perplexing and misunderstood. The core histone N-terminal "tail" domains are highly basic, disordered and dynamic condensation forces. The research presented here seeks to better understand the contributions of the N-terminal tail domains to higher order chromatin formation.

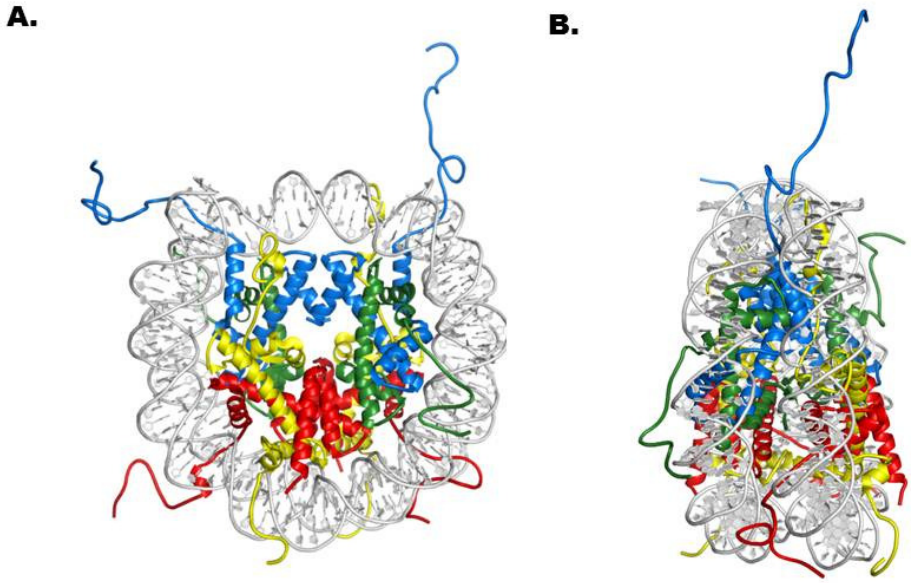

Figure 1.3. Crystal structure of the conical nucleosome core particle. Histones H2A, H2B, H3 and H4 are shown in yellow, red, blue and green, respectively, DNA is shown in silver. A, view of the nucleosome down the superhelical axis and depicting the nucleosome surface. B, view of A rotated 90 degrees horizontally. **Image courtesy of Dr. Uma Mutharajan (Davey *et al*, 2002).**

The core histone tail domains are intrinsically disordered, basic polypeptides that extend out and away from the nucleosome (Fig. 1.3). Each of the four tail domains are highly conserved amino acid primary sequence between species. The H4 and H2B tails are each composed of 27 and 24 amino acids respectively. The H2A tail is the shortest at 13 amino acids, and H3 the longest at 37 amino acids. Although there is little sequence homology between the four tail domains, each share striking similarities in amino acid composition (~33% lysines and arginines) and intrinsic disorder. The only exception being the high sequence homology (98%) between the first 13 amino acids of the H4 tail and the 13 amino acids of the H2A tail (see figure 2.1). The tail domains have not only been implicated as crucial to chromatin condensation *in vitro*, but are important protein-protein interactions critical for cell signaling, gene activation and silencing, DNA repair and replication. Of the four histone tail domains, the H4 tail has previously been implicated as most important contributor to nucleosomal array condensation *in vitro* (Dorigo *et al*, 2003; Gordon *et al*, 2005; Kan *et al*, 2009; McBrant *et al*, 2009). However, several lines of research have implied both direct and indirect roles for all four tails in different chromatin condensation events (Dorigo *et al*, 2003; Gordon *et al*, 2005; Zheng *et al*, 2005; Kan *et al*, 2007; Wang *et al*., 2008; Kan *et al*, 2009; McBrant *et al*, 2009).

Extending from the nucleosome core particle, each tail domain has two specific exit sites as it protrudes from the nucleosome. Importantly, the spatial differences in tail domain exit sites potentially provide each of the four tails with different opportunities to participate in nucleosome-nucleosome interactions both in cis and trans (Fig. 1.3). The H4 tail has been shown to participate in multiple interactions with DNA and protein in cis and trans during array condensation events, and its position within the nucleosome has already been implicated as an important molecular determinant of nucleosome array

oligomerization (Dorigo *et al*, 2004; Shogren-Knaack *et al*, 2006; Wang *et al*, 2008; Kan *et al*, 2009; McBryant *et al*, 2009).

Of the four core histone N-terminal tails, the H3 tail is the only tail positioned to interact with the terminal nucleosomal DNA at the entry/exit site (the H2A C-terminal tail is also located in this region, Fig 1.3) (Luger *et al*, 1997a). Not surprisingly the H3 tail has been shown to interact with DNA both in cis and trans during condensation events (Kan *et al*, 2007). Both the H3 and the H2B tails exit the nucleosome through the gyres of the DNA superhelix and together create a unique 20-base pair periodicity of tails spanning the circumference of the nucleosome (Luger *et al*, 1997a). Like the H4 tail, the H2A tail also exits the core particle from the nucleosome surface (Fig 1.3) (Luger *et al*, 1997a). However the H2A tail exits the nucleosome from the opposite end of the H3 tail and the DNA entry exit site (Luger *et al*, 1997a).

1.4 Defined 12-mer nucleosomal arrays as *in vitro* chromatin models

Chromatin condensation has been well studied using a defined 12-mer nucleosomal array model system (Hansen, 2002). Over the past 20 years, research using different variations of the 12-mer nucleosomal array has led to a better understanding of the basic functional requirements needed to condense an array of nucleosomes. The 5S rDNA 208-12 nucleosomal array model system was the original in vitro nucleosomal array model system (Simpson *et al*, 1985). First reconstituted in 1985 by Simpson and colleagues (Simpson *et al*, 1985), the 5S 208-12 nucleosomal array has been used in numerous studies that have provided the field with a wealth of information (Hansen, 2002). The 208-12 DNA template consists of twelve 208 base pair repeats of a nucleosome positioning sequence from the 5S rDNA gene of sea urchin (Simpson et al, 1985). Initially, this template was reconstituted with endogenous core histone octamers and linker histones purified from chicken erythrocytes (Hansen and van Holde, 1989;

Garcia-Rameriz *et al*, 1989, Schwarz and Hansen, 1994; Schwarz *et al*, 1996; Tse and Hansen, 1997; Carruthers and Hansen, 2000). However, with the advent of recombinant technology, core histones from *Xenopus laevis* have been used to generate a completely recombinant model system to study chromatin condensation (Luger *et al*, 1997a,b).

Many biophysical methods used to study nucleosomal array condensation have yielded little structural information of 12-mer nucleosomal arrays beyond the extended beads on a string conformation. However, hydrodynamic studies have defined three solution-state conformations of the 208-12 nucleosomal array. Sedimentation velocity experiments in an analytical ultracentrifuge provide an unparalleled method to study chromatin condensation. In sedimentation velocity experiments, the 208-12 nucleosomal array in low salt (2.5mM NaCl) is extended and sediments at 29S, resembling “beads on a string” in electron micrographs (Fig. 1.4) (Thoma and Koller, 1977; Hansen and van Holde, 1989; Hansen, 2002 Grigroreyv *et al*, 2009).

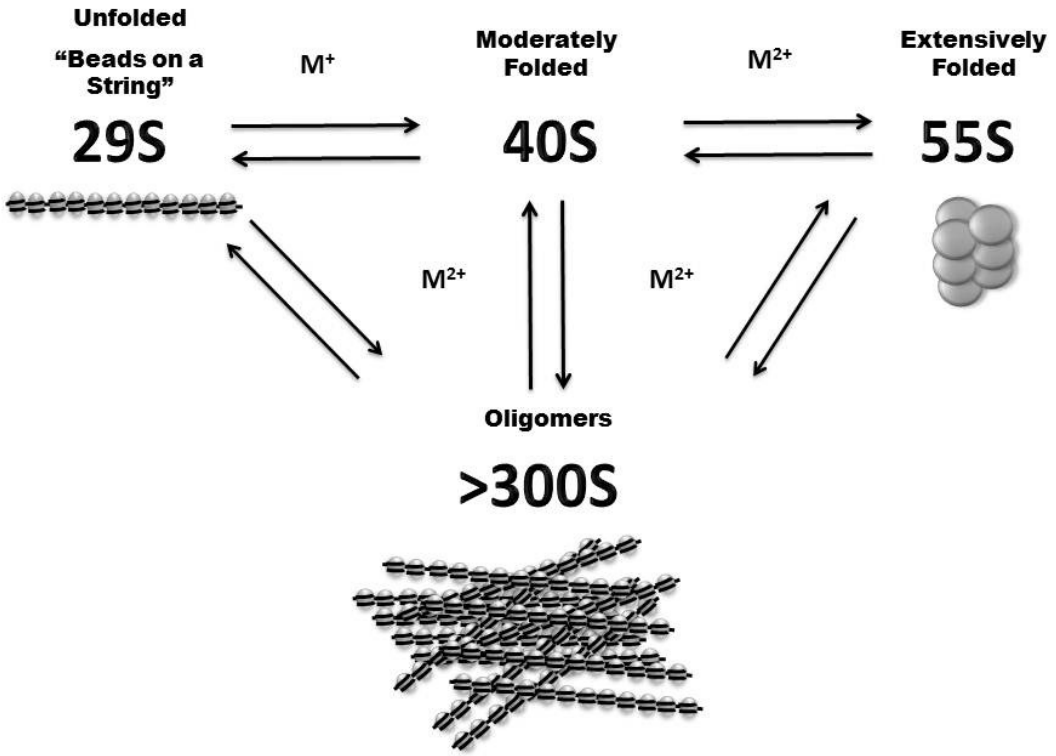

Figure 1.4 Salt dependent transitions of nucleosomal arrays. Model depicting the solution-state sedimentation coefficient of nucleosomal arrays in the presence of mono and divalent cations. Folding of the extended 29S nucleosomal array into moderately folded 40S and extensively folded 55S fibers are intra-array interactions driven by nucleosome-nucleosome interactions in cis. Oligomerization occurs at higher $MgCl_2$ concentrations, self-association between arrays in trans lead to large oligomers in excess of 300S. Cartoons are for illustrative purposes only, and are not intended to represent actual structures. Schematic adapted from Hansen, 2002.

With the addition of cations, the 12-mer array transitions into both condensed complexes mediated by intra- and inter molecular nucleosome-nucleosome interactions. Cation concentrations typically above 50 mM NaCl and below 1.0 mM $MgCl_2$ shield the negative charge of the linker DNA generating nucleosome-nucleosome interactions in cis leading to moderately "folded" nucleosomal arrays (Fig. 1.4) (Hansen, 2002). Moderate folding of nucleosomal arrays in increasing amounts of cations causes a progressive increase in the sedimentation coefficient from 29 to 40S, and is visualized in electron micrographs as a contacting zig-zag helix (Bednar *et al*, 1998; Schwarz and Hansen, 1994). Between 1-2 mM $MgCl_2$, further intra-molecular compaction generates an extensively folded nucleosomal array sedimenting at 55S (Fig. 1.4) (Schwarz and Hansen, 1994; Tse and Hansen, 1997; Carruther *et al*, 2000). EM measurements of the extensively folded array average 30 nm in diameter, consequently this structure is thought to correlate with the *in vivo* 30 nm chromatin fiber (Hansen *et al*, 1989; Schwarz and Hansen, 1994; Bednar *et al*, 1998; Routh *et al*, 2008). Nucleosomal array folding is highly reversible, as removal of cations unfolds the array into its original 29S bead on a string conformation (Schwarz and Hansen, 1994).

Independent of folding, the 12-mer nucleosomal array also oligomerizes in divalent cation concentrations above 2mM $MgCl_2$. Nucleosomal array oligomerization is a cooperative transition generating large, extensively self-associated and soluble oligomers sedimenting above 100S (Fig. 1.4) (Hansen, 2002). The oligomerization transition is independent of the folded state of the nucleosomal arrays (Schwarz and Hansen 1994). Similar to early experiments with purified eukaryotic nuclei, oligomerization of nucleosomal arrays *in vitro* has incorrectly been referred to as precipitation because the large oligomers easily pellet under a centrifugal force. However, as was observed with purified nuclei, this transition is highly cooperative and reversible. Removal of divalent cations causes dissociation of oligomers into monomeric

and extended nucleosomal arrays (Shwartz *et al*, 1996). Because of these similarities to purified eukaryotic nuclei, oligomerization assays of 12-mer arrays are thought to recapitulate the long-range fiber-fiber interactions that generate large chromonema and chromosome structures found *in vivo* (Dorigo *et al*, 2003; Lu *et al*, 2006 and McBryant *et al*, 2009).

Importantly, both condensation transitions occur without linker histones bound to nucleosomes. However, the presence of linker histones helps to stabilize the condensed states of the nucleosomal arrays. Only with the addition of one linker histone per nucleosome do nucleosomal arrays extensively fold and oligomerize in the presence of monovalent cations (Carruthers *et al*, 1998).

1.5 Chromatin Condensation Dynamics

Many factors contribute to chromatin condensation and decondensation events. The source of DNA, linker DNA length, histone variants and post-translation modifications, and especially core histone N-terminal tail domains all play a role in chromatin condensation and/or decondensation known as chromatin dynamics.

There have been several different nucleosome array templates used to generate arrays from both endogenous and synthetic sources. As discussed above the endogenous 5S rDNA nucleosome positioning sequence has been widely used to study *in vitro* nucleosomal arrays. However, in recent years new DNA sequences have been generated with nucleosome positioning sequences with higher affinity than that found in endogenous sequences. Most notably the 601 nucleosome positioning sequence was selected from a SELEX experiment and under specific experimental conditions was found to have the highest known affinity for nucleosome formation (Lowary and Widom, 1998). In recent years the 601 nucleosome position sequence has become widely used to reconstitute and study nucleosomes and nucleosomal arrays *in vitro* (Dorigo *et al*,

2003; Dorigo *et al*, 2004; Schalch *et al*, 2005; Shogren-Knaack *et al*, 2006; Lu *et al*, 2008). One benefit of using the 601 sequence is the ability to assemble nucleosomal arrays with nucleosome positioned along the template with base pair resolution, generating arrays with homogenous nucleosome repeat lengths (Lowary and Widom, 1998 and Anderson; Widom, 2000). This is unique to the synthetic sequence versus the endogenous sequences used in the past, and allows for more homogenous nucleosomes arrays after reconstitution (Dorigo *et al*, 2003 and Shogren-Knaak *et al*, 2006).

The linker DNA that separates nucleosomes can vary depending on the organism, cell type and current state of transcription or replication (van Holde, 1988). For instance, linker DNA in budding yeast has been shown to have nucleosomes spaced on average 167 base pairs apart giving an average of 20 bp of linker DNA throughout the genome. In constrast, nucleosomes in chicken erythrocyte nuclei are spaced an average of 200 bp apart yielding over 50 bp of linker DNA (van Holde, 1988; Hansen, 2002). Given the electrostatic repulsion from the negatively charged phosphate backbone of DNA, it has been shown that condensation properties will differ between nucleosomal arrays with different lengths of linker DNA. Studies using different variations of 12-mer and 35-mer nucleosomal array model systems have detailed just that (Routh *et al*, 2008; Wang and Hayes, 2008; Szerlong and Hansen, 2011). *In vitro,* nucleosomal arrays show a tendency to compact intra- and inter-molecular in lower $MgCl_2$ concentrations with shorter distances between each nucleosome (Schwarz *et al*, 1996; Dorigo *et al*, 2003; Wang and Hayes, 2008). Of note, sub-saturated arrays (less than 12 nucleosomes out of 12 possible) *in vitro* require higher concentrations of $MgCl_2$ to oligomerize and are unable to fold (due to increased linker DNA separating nucleosomes) (Hansen and Lohr, 1993; Schwarz *et al*, 1996; Wang and Hayes, 2008; McBryant *et al*, 2009).

There are two nucleosomal array templates that have been studied in great detail for their ability to condense, each containing different linker DNA lengths. First, as detailed in section 1.4, the 5S rDNA 208-12 nucleosome array folds intra-molecularly into a moderately folded 40S and extensively folded 55S conformation in low $MgCl_2$ concentrations (Schwarz and Hansen, 1994). The second nucleosome array using 601_177-12 template, however, does not fold into two states of folded nucleosome arrays. Instead the 177-12 array progressively compacts into faster sedimenting conformation with increasing $MgCl_2$ concentrations until it reaches as maximum of approximately 55-56S (Dorigo *et al*, 2003). At 1.0 mM $MgCl_2$, the 208-12 array shows a broad distribution of arrays sedimenting between 29 and 40S and 40 and 55S, whereas the 177-12 array maintains a very homogenous distribution sedimenting between approximately 50-55S (Schwarz; Tse and Hansen, 1997; Carruthers *et al*, 1998; Dorigo et al, 2005; Shogren-Knaak *et al*, 2006). It remains unknown whether the difference in sedimentation coefficient distributions between these two model systems is due to the difference in sequence or the difference in linker DNA length. Importantly, the nucleosomal arrays reconstituted in this study use the 601_207-12 template with a similar linker length as the 5S_208-12 array and will help to clarify any discrepancies in the literature.

Another important aspect involved in chromatin dynamics are the variants of the core histones. Histone variants have been shown to be causal agents of a wide variety of genomic processes as well as marking chromosomal domains. For instance, the H2A variant H2A.Z is enriched at the 5' ends of active and inactive genes and *in vitro* the H2A.Z variant facilitates nucleosomal arrays folding but disrupts array oligomerization (Fan *et al*, 2002 and Raisner *et al*, 2005). Another important histone variant, the H3 variant CENP-A is localized to the eukaryotic centromeric chromatin and is important for the faithful inheritance of chromosomes (Cleveland *et al*, 2003 and Allshire and Karpen,

2008). CENP-A and H3 share only about 60% sequence homology in the C-terminal two-thirds but shares no homology at the N-terminal third of the proteins. Despite divergent sequences, CENP-A is capable of assembling into a nucleosome with the H4, H2A and H2B and a crystal structure of the CENP-A nucleosome has recently been resolved (Tachiwana, *et al*, 2011). However, it is unknown if CENP-A assembles into nucleosomal arrays in the same step-wise manner as the conanical H3 containing arrays. Furthermore, the solution-state dynamics of CENP-A containing nucleosomal arrays is unexplored.

Another important aspect involved in chromatin dynamics are the variants of the core histones. Histone variants have been shown to be causal agents of a wide variety of genomic processes as well as marking chromosomal domains. For instance, the H2A variant H2A.Z is enriched at the 5' ends of active and inactive genes and *in vitro* the H2A.Z variant facilitates nucleosomal arrays folding but disrupts array oligomerization (Fan *et al*, 2002 and Raisner *et al*, 2005). Another important histone variant, the H3 variant CENP-A is localized to the eukaryotic centromeric chromatin and is important for the faithful inheritance of chromosomes (Cleveland *et al*, 2003 and Allshire and Karpen, 2008). CENP-A and H3 share only about 60% sequence homology in the C-terminal two-thirds but shares no homology at the N-terminal third of the proteins. Despite divergent sequences, CENP-A is capable of assembling into a nucleosome with the H4, H2A and H2B and a crystal structure of the CENP-A nucleosome has recently been resolved (Tachiwana, *et al*, 2011). However, it is unknown if CENP-A assembles into nucleosomal arrays in the same step-wise manner as the conanical H3 containing arrays. Furthermore, the solution-state dynamics of CENP-A containing nucleosomal arrays is unexplored.

1.6 Sedimentation Velocity: A unique technique to study nucleosomal array folding

Sedimentation velocity is a sensitive technique that measures the sedimentation coefficient of a macromolecule and can be used to understand a wide range of parameters (Hansen *et al*, 1997; Cole *et al*, 2008). A sedimentation velocity experiment is performed using an analytical ultracentrifuge equipped with absorbance optics. As such, the analytical centrifuge measures the absorbance of a sedimenting sample under a centrifugal force (Hansen *et al*, 1997; Cole *et al*, 2008). For any macromolecule, the sedimentation coefficient is proportion to the molecular weight of the molecule and inversely proportional to the frictional coefficient of the molecule, or $s \approx M/f$ (Demeler and van Holde, 2004). This simplified equation allows one to easily understand how nucleosomal array folding is studied with sedimentation velocity experiments. First, nucleosomal array folding is performed in low $MgCl_2$ concentrations that only induce intra-array compaction and therefore the molecular weight of the sedimenting species does not change. The frictional coefficient, f, is a sensitive measure of a macromolecule's shape. Therefore, long extended species (beads on a string nucleosomal array sediments at 29S) have a larger frictional coefficient than that of a more compact species (extensively folded array sediments at 55S) of the same molecular weight.

The raw data generated from a sedimentation velocity experiment is dependent upon many parameters, but can be visualized in a graph of radial scans depicting the absorbance of the sedimenting sample versus the radial position of the sample as it sediments to the bottom of the cell. As a consequence, over time each radial scan depicts a moving boundary. Each particle within a sample is not only subjected to sedimentation, but diffusion as well. Over time, the moving boundary broadens due to

the heterogeneity of the sample and the diffusion of the particles (Demeler and van Holde, 2004). Macromolecular diffusion therefore can potentially obscure sample separation due to heterogeneity making it impossible to determine if multiple species are present in a given sample (Demeler and van Holde, 2004). The Demeler and van Holde method, formally known as the van Holde-Weischet method makes use of the fact that s is proportional to time, whereas diffusion in inversely proportional to time, so that at $t = \infty$, diffusion becomes negligible (Demeler and van Holde, 2004). As a result, the Demeler and van Holde method produces an integral distribution (G(s)) of sedimentation coefficients corrected for diffusion and has become a standard method used to analyze nucleosomal array folding (Demeler and van Holde, 2004).

Chapter 2

Contributions of the core histone N-terminal tail domains to nucleosomal array oligomerization

Chapter 2 details the model system (601_207-12 nucleosomal array) and histone tail mutants used in this dissertation to study nucleosomal array condensation. To study array folding, one needs to use divalent cation concentrations high enough to induce maximal folding but low enough to not induce oligomerization. The bulk of this chapter details the oligomerization profiles from 26 different combinations of wild type and mutant tail containing arrays. I demonstrate here that the wild type 601_207-12 array behaves similarly to previous oligomerization studies using different templates, and that the H3 and H4 tail domains oligomerize to similar extents, but that loss of the H4 tail leads to the largest loss in oligomerization. Importantly, this chapter extends our 2005 and 2009 studies by detailing the individual contribution of each WT and mutant histone tail domain to array oligomerization.

2.1 Introduction

The study of nucleosomal array condensation *in vitro* dates back over two decades with the advent of the 5S rDNA 208-12 nucleosomal array model system in 1985 (Simpson *et al*, 1985). Since that time several different nucleosomal array model systems have been generated, ranging from differences in the nucleosome positioning sequence and length of linker DNA, to the number of nucleosomes per template. Herein

I used the high affinity 601 nucleosome positioning sequence consisting of 12 repeats of 207 base pairs of DNA (Lowary and Widom, 1998). However, a detailed study of array self-association using the 601_207-12 nucleosomal array template has not been done. Therefore it was important to document the hydrodynamic and biochemical similarities and differences between this template and those used in the past.

There is strong evidence indicating that *in vitro* nucleosomal array oligomerization is a recapitulation of the long range fiber-fiber interactions found in the interphase chromosome (Lu *et al*, 2006). The role of the core histone tails in nucleosomal array oligomerization first became apparent in the 1990s when trypsinized (and therefore tailless) arrays were unable to oligomerize at any salt concentration with or without linker histones (Schwarz *et al*, 1996; Carruthers and Hansen, 2000). The role of the histone tails was further refined to the H3 and H4 tails as being the most important contributors with the H2A and H2B dimer tails also contributing, but to a lesser extent (Tse and Hansen, 1997). Two recent studies published from the Hansen lab provide further resolution to the histone tails role in array oligomerization (Gordon *et al*, 2005; McBryant *et al*, 2009). The first study conclusively showed that each of the tail domains independently and additively contribute to nucleosomal array oligomerization (Gordon *et al*, 2005). Not surprisingly, the H4 tail was the most important contributor to oligomerization with the contributions of the remaining tails as follows: H3>H2B>H2A. We followed this study in 2009, by detailing the molecular determinants of the H4 tail that enables it to be the largest contributor. Using novel "tail-swap" H4 mutants we determined that the mode of action of the H4 tail domain centered on four determinants: 1) amino acid composition, not primary sequence, 2) tail domain length, 3) charge density and 4) tail position within the nucleosome (McBryant *et al,* 2009). All of these studies from our lab provided interesting and novel findings and predominantly used the same 5S rDNA 208-12 model system with undersaturated (8 nucleosomes out of 12

possible) nucleosomal arrays. The research reported in this study uses a fully saturated (29-30S) Widom 601_207-12 nucleosomal array instead of the 5S_208-12 nucleosome array to reconstitute a more homogenous population of arrays in which to study array condensation. Furthermore, with the use of tailless histones, nucleosomal arrays where reconstituted containing "single-tail" nucleosomes. These important distinctions from our 2009 study allowed us to determine the individual contributions of the wild type and mutant tail domains providing a finer resolution of the role of the tail domains in array condensation (McBryant *et al*, 2009).

2.2 Materials and Methods

2.2.1 DNA purification

The Widom 601 207-12 nucleosomal array DNA template was prepared as previously described (Grigoryev *et al*, 2009; McBryant *et al*, 2009). Briefly, *Escherichia coli* DH5α were transformed with a pUC19 vector containing twelve, 207 base pair repeats of the 601 nucleosome positioning sequence. Six liters of culture were grown overnight in the presence of chloramphenicol and harvested the next morning. A Qiagen Giga-Prep kit was then used to purify the pUC19 plasmid. Generally, eight micrograms of plasmid was digested with HindIII, XbaI, DraI and HaeII using New England Biolabs buffer II for twenty hours at 37ºC. Once complete, digested DNA was put through a round of chloroform extraction and ethanol precipitation and resuspended to 2 mg/ml in 1xTE, 50mM NaCl. The concentrated DNA was then added to a Sephacryl 1000 column equilibrated in 1xTE, 50mM NaCl. The column was run at 0.5 ml/min overnight and 10 microliters of fractions were collected. A ten microliter aliquot of each fraction was run on a 1% agarose gel. Fractions corresponding to the 207-12 template were pooled and concentrated to >1mg/ml using ethanol precipitation.

2.2.2 Assembly of histone octamers

Wild type *X. laevis* histone octamers have been extensively used for nucleosomal crystallography and for studies of *in vitro* nucleosomal array condensation as well. Here we make use of these same *X. laevis* histones for construction, expression, purification and characterization of H4 tail domain mutants. Histone expression and purification protocols have been previously optimized and published and are well established techniques in our lab (Dyer *et al*, 2004; McBryant *et al*, 2009). Using a pET-based expression vector, H4 NTD mutants were expressed in, and extracted from, *E. coli* and further purified using S200 size-exclusion and cation exchange chromatography. Wild type and tailless H2A, H2B, H3 and H4 histones (as well as the H3KR and H4KR mutant histones) were purified by the CSU Protein Expression and Purification Facility, and were available in frozen, lyophilized ready to use aliquots. Individually purified histones were then mixed into equimolar ratios of H3-H4 tetramer and slight stoichiometric excess of the H2A-H2B dimer and dialyzed into TE buffer containing 2M NaCl and 1 mM 2-mercaptoethanol. Assembled histone octamers were purified from nonspecific aggregates and excess H2A-H2B dimers using a Superdex 200 gel filtration column. Histone purification and octamer assembly protocols have been previously optimized and were performed as previously described (Dyer *et al*, 2004; and McBryant *et al*, 2009).

2.2.3 Reconstitution and characterization of biochemically defined nucleosomal array model systems.

Nucleosomal arrays are assembled with purified recombinant histone octamers and a defined nucleosome positioning sequence *in vitro* using a salt-dialysis reconstitution protocol (Schwarz and Hansen, 1994; Gordon *et al*, 2005; McBryant *et al*, 2009; Panchenko *et al*, 2011). The 601 DNA template has been previously shown to yield equally spaced nucleosomes and a highly homogenous population of nucleosomal arrays (Dorigo *et al*, 2003). Assembly begins by mixing wild type or mutant histone

octamers and template DNA at a ratio of 1.2 mol octamer to 1 mol DNA repeat. Salt-dialysis reconstitutions begin in TE buffer containing 2M NaCl and are successively dialyzed into TE buffer containing 1M, 0.75M and 2.5mM NaCl over the course of two days.

Early work in our lab has established that oligomerization and folding is highly dependent upon nucleosome saturation. Therefore, to determine levels of octamer loading on template DNA, sedimentation velocity experiments were performed in a Beckman XL-A or XL-I analytical ultracentrifuge. Sedimentation velocity data was analyzed using UltraScan v.9.9 software and the van Holde-Weischet Method, which removes diffusion contributions from the measured sedimentation coefficient (Demeler and van Holde, 2004). Previous characterization of 12-mer nucleosomal arrays by this method have established that a fully saturated wild type 12-mer array sediments at 29-30S in low salt TE buffer. Thus, all nucleosomal arrays were reconstituted such that they sedimented at 29-30S in low salt 1xTEN buffer (10mM Tris, pH 7.8; 0.25mM EDTA; 2.5mM NaCl). This allows accurate folding comparisons between wild type and mutant arrays.

2.2.4 EcoRI digest

As a second method to determine if arrays were fully saturated, 1 ug of each array was digested with the EcoRI restriction enzyme in a 10 ul final reaction for either 1 hour at 37ºC or 16 hours at room temperature, with enzyme specific buffers from New England Biolabs. Mononucleosomes and free DNA were visualized on a 1% agarose gel, stained with ethidium bromide and imaged with a Kodiak Gel Logic 200 imager. Past literature documenting endonuclease digestions of 12-mer nucleosomal arrays showed that a fully saturated array maintains 0-3% free DNA versus mononucleosomal DNA after digestion (Tse and Hansen, 1997; Dorigo *et al*, 2003; Gordon *et al*, 2005).

2.2.5 Differential centrifugation assay

To test the nucleosomal arrays ability to oligomerize and form large >300S oligomers with increasing amounts of $MgCl_2$ the differential centrifugation assay was used (Schwarz and Hansen, 1994; Schwarz *et al*., 1996; McBryant *et al*, 2009). Briefly, 1.2 A_{260} units of nucleosomal arrays were mixed with increasing amounts of $MgCl_2$, incubated at room temperature for 10 minutes and then centrifuged on a table top centrifuge for 10 minutes. As the concentration of $MgCl_2$ increases nucleosomal arrays begin to oligomerize and form large insoluble material that pellet under moderate centrifugal force. This oligomerization is measured as a decrease in absorbance of the supernatant at 260 nm (Schwarz and Hansen, 1994 and Schwarz *et al*, 1996; McBryant *et al*, 2009). When plotted as the percent remaining in the supernatant versus the $MgCl_2$ concentration, self-association curves of nucleosomal arrays are sigmoidal and show that the oligomerization process is cooperative (Schwarz and Hansen, 1994 and Schwarz *et al*., 1996). Mg50 plots were generated depicting the $MgCl_2$ concentration at which 50% of the nucleosomal array was oligomerized (Gordon *et al*, 2005 and McBryant *et al*, 2009).

2.3 Results

2.3.1 Fully saturated nucleosomal arrays with different combinations of full length and "tailless" core histones maintain different sedimentation coefficients.

A systematic study of the role of the histone tails in nucleosomal array oligomerization has not been reported with the 601_207-12 nucleosomal array. Therefore, it was important to establish any differences between this template, and the 5S_208-12 template characterized previously (see section 1.5). Core histone octamers containing different combinations of wild-type and mutant N-terminal tail domains were mixed with the Widom 601_207-12 DNA at r-values ranging from 0.8 to 1.8 (r-value is the molar ratio of histone octamers to 207bp repeat) (see figure 2.1 for sequences of

wild type, tailless and mutant histones). After salt-dialysis, freshly reconstituted nucleosomal arrays were analyzed for the degree of octamer loading using sedimentation velocity analysis with an analytical ultracentrifuge, described in section 2.2.5. Nucleosomal array intra- and intermolecular condensation events are very sensitive to the degree of octamer loading. Therefore, it was critical to ensure that each set of arrays contained a majority of fully saturated 12-mer nucleosomal arrays. Fully saturated wild type nucleosomal arrays (WT) containing all four full length core histones exhibited an average sedimentation coefficient (s_{ave}) of 29.5S in low salt 1xTEN buffer (Fig. 2.2). Conversely, fully saturated "tailless" nucleosomal arrays (TL) exhibited a s_{ave} of 25S (Fig. 2.2). The difference in the sedimentation coefficient between WT and TL

WT H4: SGRGKGGKGLGKGGAKRHR**K**VLRDNIQ-H4 HF
WT H3: ARTKQTARKSTGGKAPRKQLATKAAR**K**SAPATGGVKK-H3 HF
WT H2A: SGRGKQGGKTRA**K**-H2A HF
WT H2B: AKSAPAPKKGSKKAVTKTQK**K**DGRP-H2B HF
^{T}3^{HF}4: ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKK-H4 HF
^{T}2A^{HF}4: SGRGKQGGKTRAK-H4 HF
^{T}2B^{HF}4: AKSAPAPKKGSKKAVTKTZKKDGRP-H4 HF
^{T}4^{HF}2B: SGRGKGGKGLGKGGAKRHRKVLRDNIQ-H2B HF
^{T}[4]$_{R1}$^{HF}4: KRGRVRGDSLIGLGGQANHGKKKGGKR-H4 HF
^{T}[4]$_{2}$^{HF}4: SGRGKGGKGLGKGGAKRHRKVLRDNIQ·***GA***SGRGKGGKGLGKGGAKR HRKVLRDNIQ-H4 HF
^{T}[4]$_{3}$^{HF}4: SGRGKGGKGLGKGGAKRHRKVLRDNIQ·SGRGKGGKGLGKGGAKRH RKVLRDNIQ·***GA***SGRGKGGKGLGKGGAKRHRKVLRDNIQ-H4 HF
^{T}[4]$_{CD}$^{HF}4: SGQGKGGAGLGKGGAKGHRSVLRDNIQ-H4 HF
^{T}[2A]$_{CD}$^{HF}4: SKRGKQRGKTRKK-H4 HF
H4K→R: SGRGRGGRGLGRGGARRHRKVLRDNIQ-H4 HF
H3K→R: ARTKQTARRSTGGRAPRRQLATRAARKSAPATGGVKK-H3 HF
CENP-A: GPRRRSRKPEAPRRRSPSPTPTPGPSRRGPSLGASSH-CENP-A HF

Figure 2.1 Amino acid sequence of wild type and mutant core histone tail domains. Bold and underlined lysines represent the first amino acid of the tailless core histones. Dots and bold/italicized GA's indicate links between tandemly repeated tail domains. Underlined residues of charge density mutants indicate mutated residues. WT = Wild Type, HF = Histone Fold domain, T = Tail domain, CD = Charge Density.

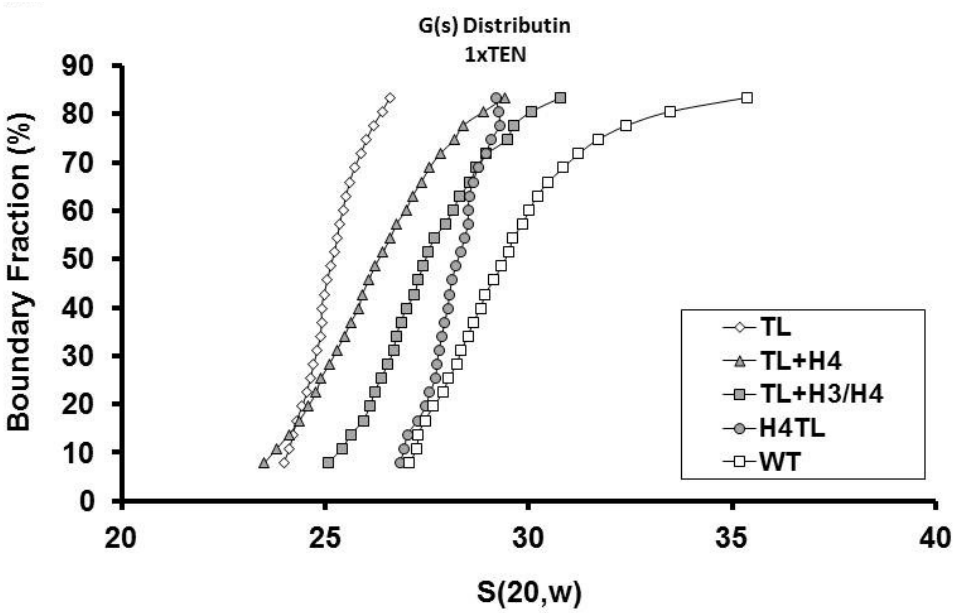

Figure 2.2 Properties of fully saturated nucleosome containing different combinations of N-terminal tail domains. WT and TL nucleosomal arrays maintain s_{ave} of 29.5 and 26S respectively. Nucleosome arrays with one, two and three core histone tails maintain a s_{ave} of 26.5, 27.5 and 28.5S respectively. Integral distribution of fully saturated arrays in low salt TEN buffer.

arrays has been previously observed and is due to a more extended and consequently slower sedimenting structure of the TL arrays due to less DNA being wrapped around each nucleosome (Fletcher and Hansen, 1995; Tse and Hansen, 1997; Carruthers and Hansen, 2000). Nucleosomal arrays containing 1, 2 and 3 full length histones were reconstituted to a S_{ave} of 26.5, 27.5S and 28.5S, respectively (Fig. 2.2A).

After the correct octamer loading was accomplished, EcoRI digests were performed as a secondary method to determine the extent of octamer loading. EcoRI is an endonuclease that targets a specific 5 base pair sequence in the linker DNA region of each 207 base pair repeats. EcoRI therefore digests the nucleosomal array into individual mononucleosomes that migrate in an agarose gel at approximately 250bp. Any 207 base pair repeat not bound by a histone octamer migrates as free DNA at approximately 200bp. Therefore, it is possible to determine the average number of nucleosomes per template based off of the proportion of free DNA to mononucleosomes. Generally, 12-mer 208-12 arrays have on 0-3% free DNA after EcoRI digests (Tse and Hansen, 1997; Dorigo *et al*, 2003; Gordon *et al*, 2005). However, no EcoRI digests of fully saturated 207-12 arrays in this study showed any free DNA. Fully saturated 601_177-12 arrays have previously been shown to exhibit 0% free DNA after endonuclease digestion (Dorigo *et al*, 2003).

2.3.2 Effect of the core histone tails on nucleosomal array oligomerization.

To begin to establish any differences in the 601_207-12 nucleosomal array from previously published nucleosomal arrays (601_177-12 and 5S_208-12) I reconstituted both full length wild type and tailless nucleosomal arrays to 29.5S and 25S, respectively. To determine how each array oligomerizes in the presence of $MgCl_2$, the differential centrifugation assay was performed, see section 2.2.7. Similarly to what has previously been published, the wild type 207-12 array oligomerizes cooperatively, achieving 50% oligomerization (Mg50) at 1.89 mM $MgCl_2$, and are completely oligomerized by 3.0 mM

$MgCl_2$ (Fig. 2.3 and Table1). In contrast to WT arrays and in good agreement with published results, the TL array does not appreciably oligomerize at any salt concentrations tested (up to 125mM $MgCl_2$). This lends further evidence that the core histone tails are primary determinant in nucleosomal array condensation.

2.3.3 Deleting the H4 tail causes the largest deficiency in oligomerization

To determine how the loss of each of the core histone tails individually would affect the arrays ability to oligomerize, four more nucleosomal arrays were generated each containing three full length histones and one tailless histone. Each array was named based on the tailless histone it contained (H4TL: tailless histone H4 and full length histones H3, H2A and H2B). Results of the differential centrifugation assay are similar to what has been previously published in our lab and others (Gordon *et al*, 2003; *Sinha et al*, 2010, McBryant *et al*, 2009). Deletion of the H4 tail leads to the largest loss in oligomerization, reproducibly generating a Mg50 of 4.03 mM $MgCl_2$, whereas loss of the H3, H2A and H2B tail have intermediate Mg50s to that of the WT and H4TL. Interestingly, loss of the H3 and H2B tail lead to similar Mg50s (2.90 and 2.96 mM $MgCl_2$) whereas H2A tail deficient arrays (H2ATL) produce the smallest Mg50 at 2.37mM $MgCl_2$ and are the least affected in their ability to oligomerize (Fig. 2.4).

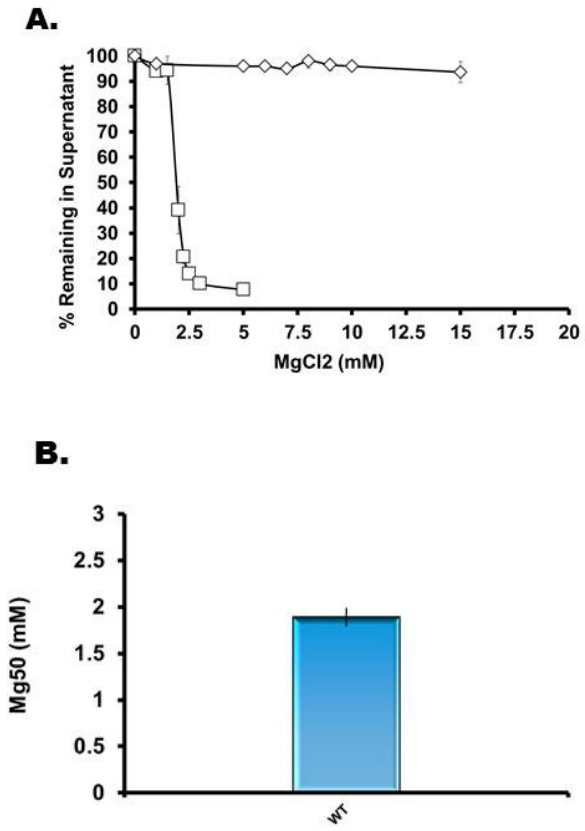

Fig 2.3 Oligomerization of WT and TL nucleosomal arrays. Fully saturated WT nucleosomal arrays are able to cooperatively oligomerization, TL arrays do not appreciably oligomerize. A, oligomerization as a function of $MgCl_2$ concentration as determined by the differential centrifugation assay (see section 2.3.) for fully saturated nucleosome arrays containing all core histone N-terminal tail domains (WT, -□-) and nucleosome arrays deficient of all tail domains (TL, -◊-). B, Mg50 value derived from the data in A. Error bars represent the S.D. of 3 or more independent oligomerization assays, see Materials and Methods. TL arrays do not oligomerize and therefore do not have Mg50 values.

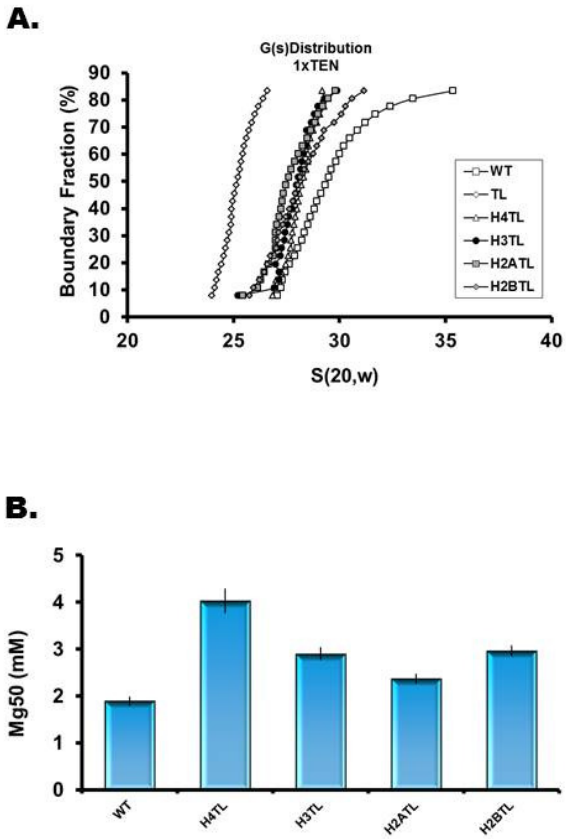

Fig 2.4 Oligomerization of nucleosomal arrays containing a single tail deletion. Fully saturated nucleosomal arrays (A) containing octamers with three full length and one tailless histone. Nucleosomal arrays were assayed for oligomerization using the differential centrifugation assay. Deletion of the H4 tail leads to the largest increase in Mg50 (B). A, Integral distribution of fully saturated arrays TEN buffer containing a single tailless histone. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the MgCl2 concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays.

2.3.4 Contributions of the tail domains from the H3-H4 heterotetramer versus the H2A-H2B dimer tails towards oligomerization

Previously it has been shown in our lab that the tail domains of the H3-H4 tetramer are more potent inducers of oligomerization than the H2A-H2B dimer tails (Tse and Hansen, 1997 and Gordon *et al*, 2005). To determine if this holds true for fully saturated 207-12 arrays, I reconstituted two additional nucleosomal arrays, one containing only the full length H3 and H4 histones, the other containing only the full length H2A and H2B histones. Both arrays (TL+H3/H4 and TL+H2A/H2B) were reconstituted to a s_{ave} of 27.5S (Fig. 2.5A). Oligomerization results show that arrays containing tailless dimer and full length tetramer histones (TL+H3/H4) were able to induce 50 percent oligomerization at considerably lower $MgCl_2$ concentrations than the TL+H2A/H2B nucleosomal array, 3.15 mM versus 7.20 mM $MgCl_2$, respectively (Fig. 2.5). However, arrays containing the H3 and H4 tails alone are unable to oligomerize at the same salt concentrations as that of the WT array.

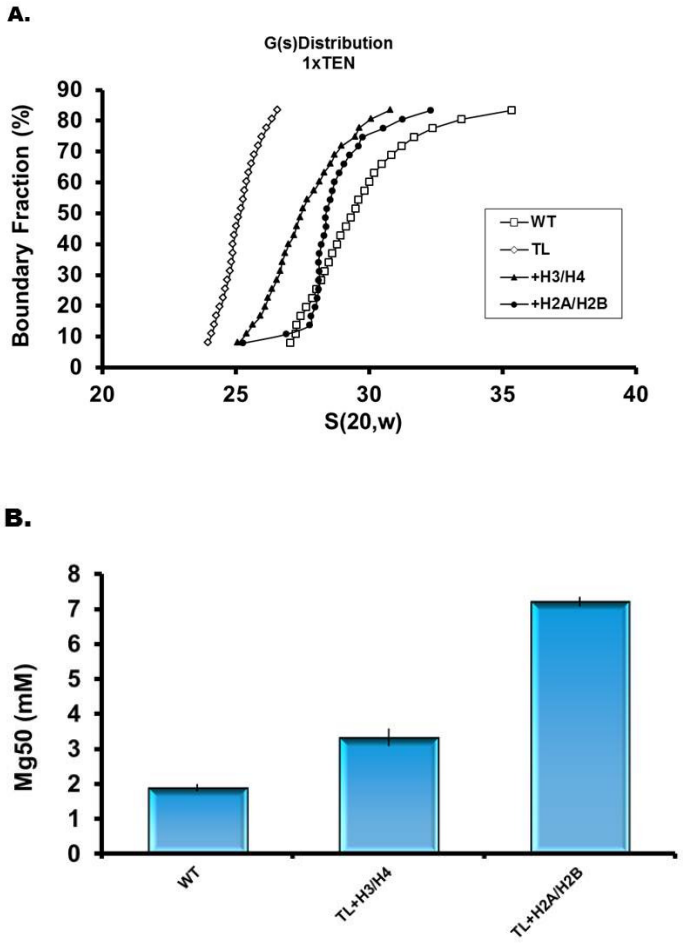

Fig 2.5 Contributions from the H3-H4 tetramer or H2A-H2B dimer tails in oligomerization. Fully saturated nucleosomal arrays (A) containing octamers with either full length H3 and H4 and TL H2A and H2B histones or full length H2A and H2B and TL H3 and H4 histones. Nucleosomal arrays were assayed for oligomerization using the differential centrifugation assay. The core histone tail domains that make up the H3/H4 tetramer induce oligomerization at lower $MgCl_2$ concentrations than that of the H2A/ H2B dimer tail domains. (B). A, Integral distribution of fully saturated arrays in TEN buffer. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays.

2.3.5 Individual contributions of the H3, H4, H2B and H2A tails during nucleosomal array oligomerization

To determine how the four core histone tails independently contribute to oligomerization, four single-tail containing (one full length and three tailless histones) nucleosomal arrays were reconstituted to a S_{ave} of 26.5S (Fig. 2.6A). The differential centrifugation assay was again used to determine oligomerization profiles and Mg50s. Arrays containing only the H3 or H4 tail achieved similar Mg50s at 7.27 and 8.47 mM $MgCl_2$, respectively (Fig. 2.6 and Table 1). Interestingly, the H2B tail by itself was able to induce oligomerization of nucleosomal arrays, achieving a Mg50 of 15.1 mM $MgCl_2$ (Fig. 2.6B), whereas sub-saturated 5S_208-12 arrays containing the same octamer where unable to oligomerize, as previously reported (Gordon *et al*, 2005). Arrays containing only the H2A tail were unable to appreciably oligomerize at all $MgCl_2$ concentrations tested (up to 30 mM $MgCl_2$). Interestingly, no combination of different tailless histone containing arrays tested was able to induce WT-like oligomerization. The independent actions of the four tails in oligomerization have been reportedly previously (Gordon *et al*, 2005).

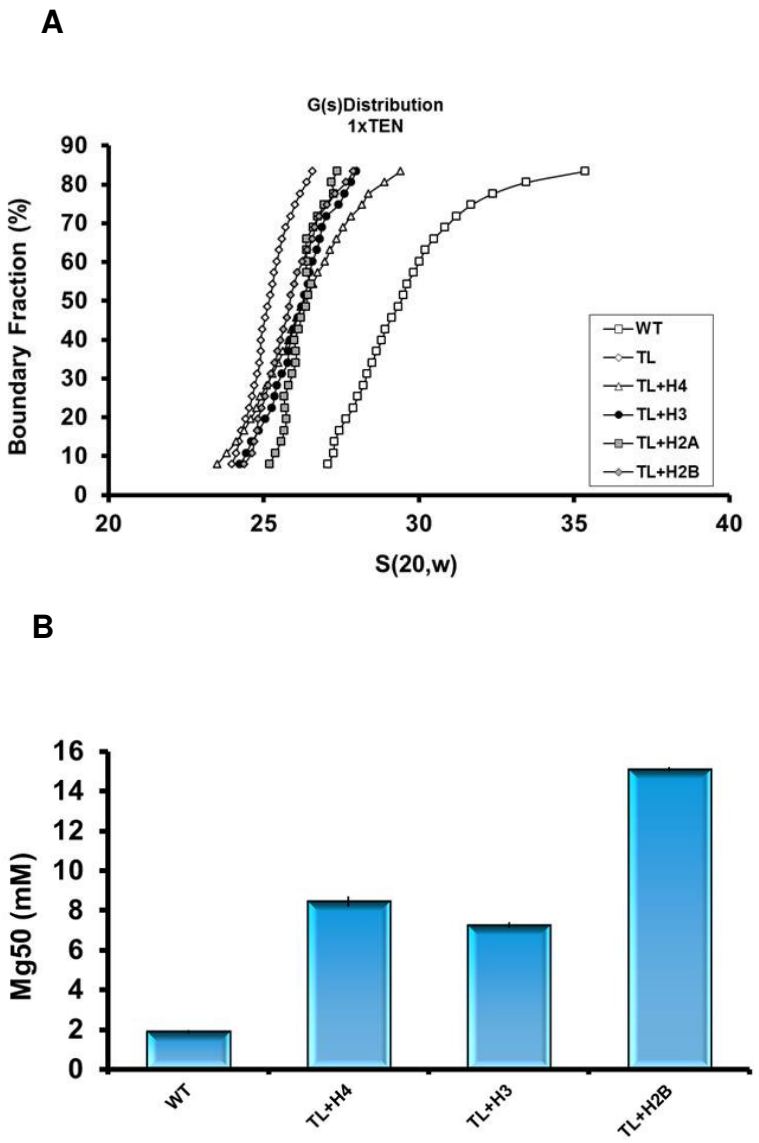

Fig 2.6 Oligomerization of nucleosomal arrays containing a single histone tail domain. Fully saturated nucleosomal arrays (A) containing octamers with one full length and three TL histones. The H3 and H4 tail domains independenlty induce oligomerization to similar Mg50s, whereas the H2B tail domain independently requires almost two times as much $MgCl_2$ to oligomerize. The H2A tail is unable to oligomerize. A, Integral distribution of fully saturated arrays in low salt TEN buffer. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays. No Mg50 is reported for TL+H2A due to its inability to oligomerize nucleosome arrays.

2.3.6 Effect on Mg50 after replacement of the H4 tail with the H3, H2A and H2B tail

It has been well established in the literature that the H4 tail is the primary determinant of nucleosomal array condensation, and the data I have represented here suggests in part that the same holds true for the 601_207-12 nucleosomal arrays as well (Dorigo *et al*, 2003; Gordon *et al*, 2005; McBryant *et al*, 2009). To determine the mode of action used by the H4 tail to direct array compaction I started by using mutant H4 histones in which the native H4 tail was replaced with each of the other three histone tails. Each mutant "tail-swapped" array contains tailless H3, H2A and H2B histones and the full length mutant H4 histone and are designated by the tail (T) they contain and the histone fold (HF) the tail is attached to (TL+$^{T}3^{HF}4$, tail of H3 fused to the histone fold domain of H4). Previously, we have used these H4 "tail-swapped" mutant histones to characterize the molecular determinants of the H4 tail during oligomerization using the lower affinity nucleosome positioning sequence (5S_208-12), under saturated arrays and the full length H3, H2A and H2B histones were present (McBryant *et al*, 2009). We found that oligomerization induced by the H4 tail is not primary sequence dependent, but instead driven by the intrinsic disorder and amino acid composition (high lysine/arginine content) inherent to all four core histones.

Similar to what we published in 2009, in a tailless background the H3 and H2B tails (TL+$^{T}3^{HF}4$ and TL+$^{T}2B^{HF}4$) are able to replace the H4 tail and oligomerize fully saturated arrays at 7.17 to 7.77 mM $MgCl_2$, respectively versus 8.47mM $MgCl_2$ with the wild type TL+H4 array (compare Figs. 2.6 and 2.7). This is in contrast to when the shorter H2A tail replaces the H4 tail (TL+$^{T}2A^{HF}4$). In this case, the nucleosomal array is unable to oligomerize similar to what was observed with TL+H2A. This is in good agreement with our 2009 study in which the $^{T}2A^{HF}4$ mutant histone was unable to replace the wild type H4 histone during the oligomerization process and maintained a Mg50 intermediate to

the wild type and H4 TL arrays (McBryant *et al*, 2009). Again as we reported in 2009, it appears that the position of the tail may contribute to oligomerization (TL+H2B maintains a Mg50 of 15.1mM versus an Mg50 of 7.77mM for TL+^{T}2B^{HF}4).

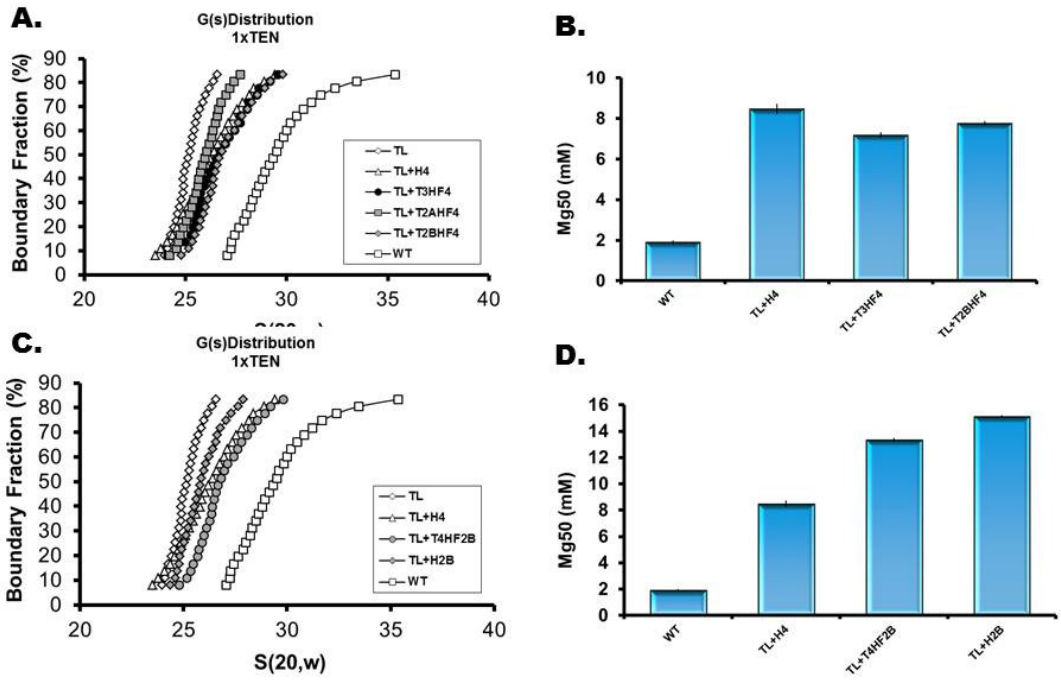

Fig 2.7 Oligomerization of nucleosomal arrays containing a single "tail" swap histone tail domain. Fully saturated nucleosomal arrays (A and C) containing octamers with one full length "tail" swap and three TL histones. The H3 and H2B tail can replace the H4 tail domain during oligomerization, the H2A tail domain cannot. However, the H4 tail requires more $MgCl_2$ to induce oligomerization when relocated to the H2B tail's position. A and C, Integral distribution of fully saturated arrays in low salt 1xTEN buffer. Only boundary fractions 8-83% are shown. B and D, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays. No Mg50 is reported for TL+H2A due to its inability to oligomerize nucleosome arrays.

In figure 2.6, the H2B tail alone (TL+H2B) was able to induce oligomerization of an otherwise tailless array, albeit at almost twice the $MgCl_2$ concentration as the H3 and H4 tails. However when the H2B tail was relocated to the H4 tail position (TL+^{T}2B^{HF}4) it was able to achieve an Mg50 of almost half that of the wild type H2B. I was curious as to whether the H4 tail extending from the H2B histone fold (^{T}4^{HF}2B) could generate lower Mg50s than the wild type H2B. Based off of our results in 2009, the TL+^{T}4^{HF}2B array should oligomerize similar to that of the wild type array (TL+H2B). A tailless array containing the ^{T}4^{HF}2B histone was reconstituted to a S_{ave} of 26.5S and analyzed for its ability to oligomerize. Not surprisingly, the H4 tail is only marginally better at inducing oligomerization of arrays than the wild type H2B, achieving an Mg50 of 13.33mM vs 15.1mM, respectively (Fig. 2.7 B and D). These studies lend further evidence to the H4 tail's position within the nucleosome and not the H4 tail itself as a prime contributor to oligomerization.

2.3.7 Consequence of scrambling the H4 tails primary sequence during oligomerization

Past studies have suggested that the primary sequence of the H4 tail, in particular amino acids 14-19 is the primary molecular determinant driving nucleosomal array condensation (Dorigo *et al*, 2003; Shogren-Knaak *et al*, 2006). My research to this point has indicated that the position of the H4 tail within the nucleosome and the amino acid composition inherent in all core histones plays a more significant role. To directly test the requirement that the H4 tail's primary sequence is not a contributing factor, a full length H4 histone was generated in which the first 27 amino acids were randomized (Fig. 2.2). As before, a single-tail nucleosomal array containing only the ^{T}4R^{HF}4 mutant histone was reconstituted to a s_{ave} of 26.5S. Oddly and in contrast to my original hypothesis, scrambling the H4 tails amino acid sequence has an adverse effect on the oligomerization. Whereas the wild type H4 tail is able to induce 50% oligomerization at

8.47 mM Mg, scrambling its amino acid sequence leads to an increase in the amount of $MgCl_2$ needed to induce the same 50% oligomerization (13.10mM, Fig. 2.8), indicating that the H4 tails primary sequence does play at least a partial role in oligomerization.

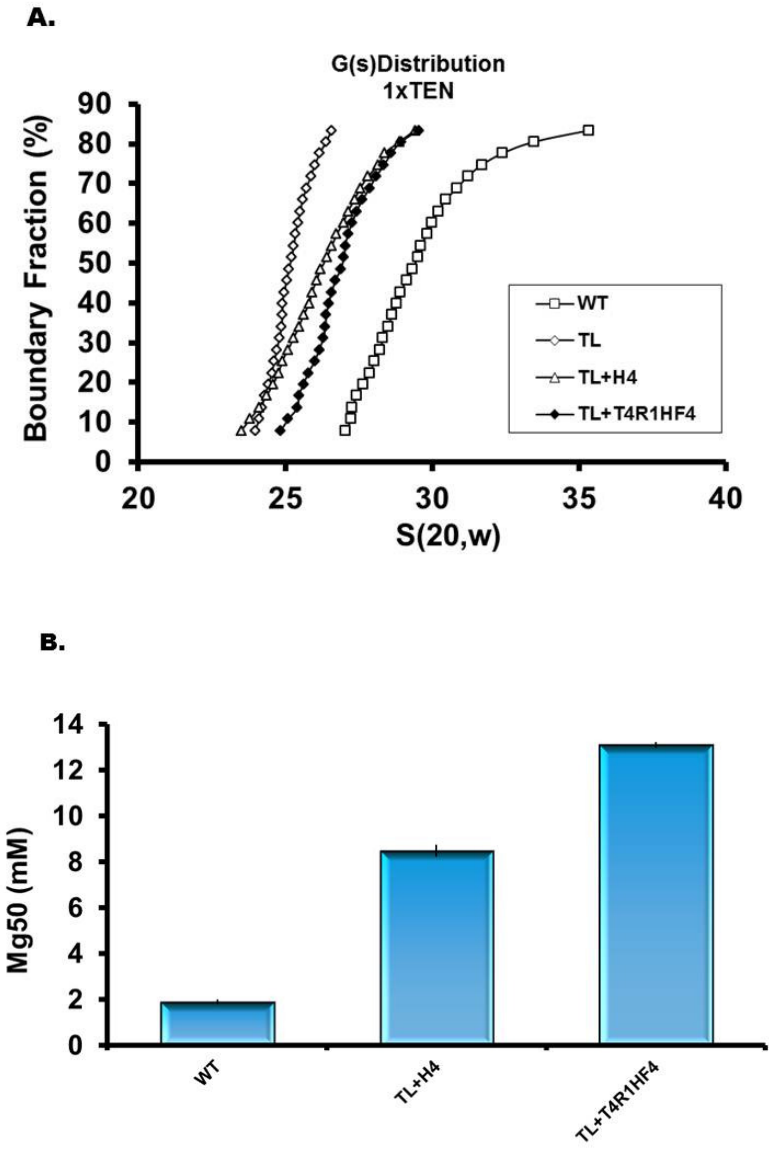

Figure 2.8 Oligomerization of nucleosomal arrays containing a single "randomized" H4 tail domain. Fully saturated nucleosomal arrays (A) containing octamers with one full length scrambled H4 tail and three TL histones. Randomizing the H4 tail's primary sequence decreases the arrays ability to oligomerize (B). A, Integral distribution of fully saturated arrays in low salt 1xTEN. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays.

2.3.8 Effect of increasing the length of the H4 tail on oligomerization

In 2009, we established that the length of the H4 tail was inversely proportional to the Mg50 of an array (McBryant *et al*, 2009). Originally, we hypothesized that a longer H4 tail would inhibit oligomerization due to steric hindrance. Yet, we found that the sub-saturated WT nucleosomal arrays required progressively less $MgCl_2$ to achieve 50% oligomerization the longer we made the H4 tail (McBryant *et al*, 2009). However, any potential steric hindrance between interacting nucleosomes may have been overcome due to a sub-saturated template. The nucleosomal array system reconstituted here contains on average four more nucleosome per template (compared to the 2009 study), which could potentially lead to a steric clash between neighboring nucleosomes (McBryant *et al*, 2009). However, I was still interested to determine if lengthening the H4 tail would decrease the arrays Mg50 when there is no other tail present.

I generated two more single-tail containing arrays to an average sedimentation coefficient of 26.5S. The first array doubled the first 27 N-terminal amino acids to generate a 54 amino acid tail domain ($^{T}[4]_2^{HF}4$) (Fig. 2.2 and Table 1). The second array tripled the first 27 amino acids to generate an 82 amino acid tail domain ($^{T}[4]_3^{HF}4$) , (Fig. 2.2 and Table 1). As we reported in 2009, the Mg50 of an array is inversely proportional to the length of the tail (McBryant *et al*, 2009). Whereas, the WT 27 amino acid length H4 array (TL+H4) maintained an Mg50 of 8.47mM $MgCl_2$, doubling and tripling (TL+$^{T}[4]_2^{HF}4$ and TL+$^{T}[4]_3^{HF}4$) the length of the N-terminal tail domain successively decreased the Mg50 to 3.57 and 1.84mM $MgCl_2$, respectively (Fig. 2.9 and Table 1). Interestingly, the 82 amino acid H4 tail in an otherwise tailless array is able to induce WT-like Mg50s. However, as we noted in 2009, lengthening the H4 tail also increases the amount of positive charge available to neutralize linker DNA (McBryant *et al*, 2009). Therefore, it may not necessarily be the length of the H4 tail, but the increased positive charge driving oligomerization.

2.3.9 Effect of changing the H4 tail's positive charge density on oligomerization

To test whether the increased length or positive charge density of the H4 tail is driving the oligomerization detailed in section 2.3.9, I generated two new H4 tail mutants. The first mutant used the wild type H4 histone as a template to mutate five of the nine lysines and/or arginines to amino acids similar to the overall amino acid composition of the histone tails (Fig. 2.2). The second mutant uses the shorter, tail-swapped histone $^{T}2A^{HF}4$ as a template and mutated 5 of the 13 total N-terminal domain amino acids to lysines or arginines, see figure 2.2. The wild type H4 and $^{T}2A^{HF}4$ tail domains both have a similar positive charge density of approximately 33%. The new charge density mutants now maintain either a wild type length but a reduced positive charge density of 19% ($^{T}4_{CD}{}^{HF}4$) or a shorter tail domain and an increased positive charge density of 62% ($^{T}2A_{CD}{}^{HF}4$). Each single tail containing mutant was reconstituted to an average sedimentation coefficient of 26.5S (Fig. 2.10 and Table 1).

Decreasing the positive charge density of the wild H4 histone causes complete disruption of oligomerization as the TL+$^{T}4_{CD}$HF4 array failed to oligomerize. This is very similar to what was observed in the shorter tail containing array, TL+ $^{T}2A^{HF}4$ and the completely tailless array (TL) which were also unable to oligomerize (Fig. 2.1A). However, when the positive charge density was almost doubled with half the wild type H4 tail length (TL+ $^{T}2A_{CD}{}^{HF}4$) the array is able to cooperatively oligomerize achieving an Mg50 of 10.44mM $MgCl_2$ (Fig. 2.10, Table 1, data not shown). Increasing the charge density of the shorter tail did not lead to TL+H4 oligomerization levels (10.44mM versus 8.47mM). It is clear that charge neutralization is a significant determinant of the H4 tail during the oligomerization transition.

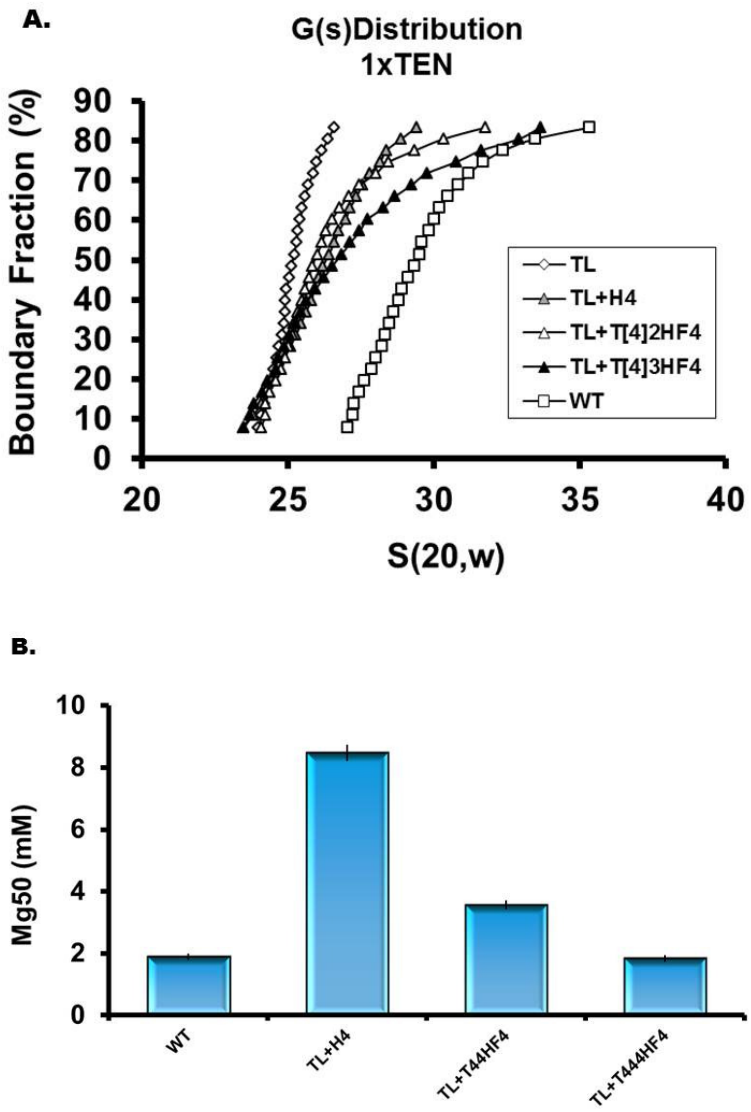

Figure 2.9 Oligomerization of nucleosomal arrays containing single tail "length" mutants. Fully saturated nucleosomal arrays (A) containing octamers with one H4 "length" mutant and three TL histones. Increasing the length of the H4 tail increases the arrays propensity to oligomerize (B). A, Integral distribution of fully saturated arrays in low salt 1xTEN buffer. A, Integral distribution of fully saturated arrays in low salt 1xTEN. Buffer. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays.

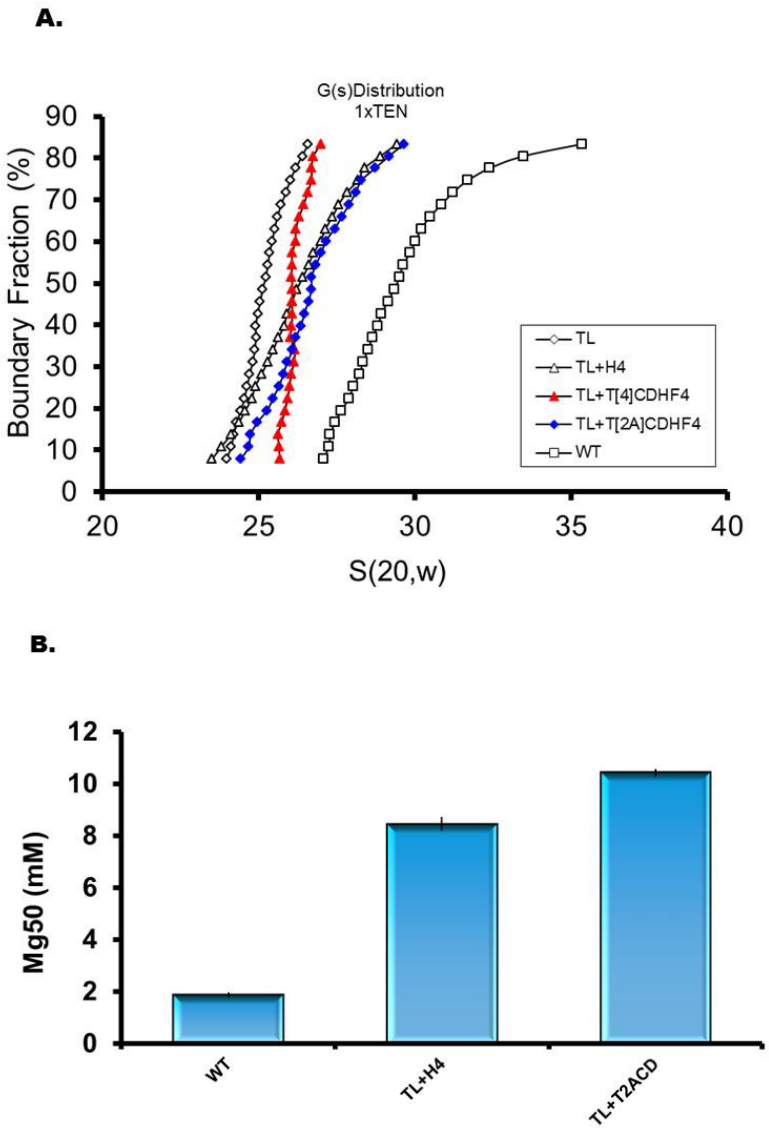

Figure 2.10 Oligomerization of nucleosomal arrays containing single tail "charge density" mutants. Fully saturated nucleosomal arrays (A) containing octamers with one H4 "charge density" mutant and three TL histones. Increasing the charge density of a shorter H4 tail induces oligomerization, decreasing the charge density of a wild type length H4 tail disrupts oligomerization (B). A, Integral distribution of fully saturated arrays in low salt 1xTEN Buffer. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays.

2.3.10 Consequence of replacing the lysines for arginines in the H3 and H4 tail on the Mg50

Each of the tail domains is rich in the basic amino acids lysine and arginine. As a whole, lysine residues outnumber the arginine residues by as much as 2 to 1. Because arginine is known to intercalate the DNA in many important protein-DNA interactions (histone octamers bound to nucleosomal DNA and TBP binding to the minor groove of promoters), I was curious as to whether exchanging the lysine residues in the histone tails with arginines would increase the arrays ability to oligomerize. I originally hypothesized that the tail domains function during oligomerization was driven mainly by charge neutralization, and therefore a conserved lysine to arginine mutation would have little effect in oligomerization profiles.

To test this hypothesis, I generated a single tail containing array and a double tail containing array. The first array (TL+H4KR) used a wild-type length H4 protein with K5,8,12,16R mutations (Fig. 2.2). Whereas the wild type H4 tail contained five lysines and four arginines, the H4KR mutant contains eight arginines and one lysine. The second array, (TL+H3/H4KR) added the H3KR tail mutant to the TL+H4KR array (H3KR mutant contains K9,14,18,23R. I chose these two variations because these arrays in their wild type form were the simplest arrays that generated the maximum amount of folding. TL+H4KR was reconstituted to an S_{ave} of 26.5S, similar to the rest of the single tail containing arrays, and TL+H3/H4KR was reconstituted to an S_{ave} of 28.5S similar to its wild type counterpart TL+H3/H4 (Fig. 2.11A and C).

The TL+H4KR slightly increased the arrays ability to oligomerize, decreasing the Mg50 from 8.47mM to 7.72mM (Fig. 2.11B). Unexpectedly, TL+H3/H4KR almost required double the amount of Mg to achieve 50% oligomerization than the TL+H3/H4 array, 3.15mM versus 5.63mM, respectively (Fig. 2.11D).

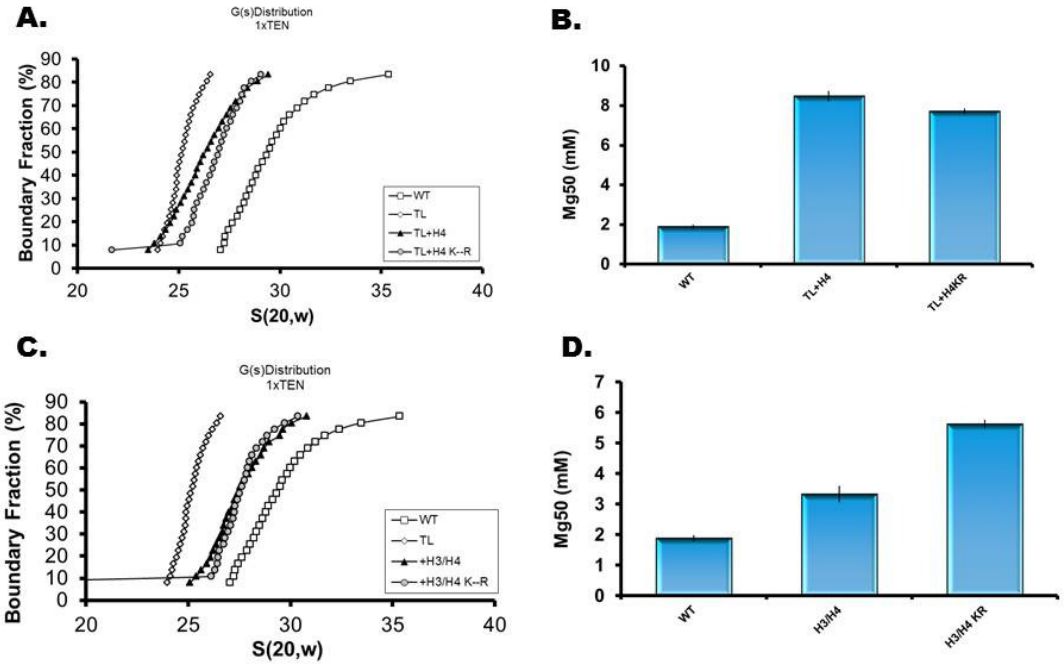

Figure 2.11 Oligomerization of nucleosomal arrays containing K→R mutants. Fully saturated nucleosomal arrays (A and C) containing octamers with H3 and/or H4KR mutations. Increasing the arginine content of the H4 tail induces similar Mg50s as the WT tail, increasing the arginine content of both the H3 and H4 tails disrupts oligomerization (B and D). A, Integral distribution of fully saturated arrays in low salt 1xTEN Buffer. Only boundary fractions 8-83% are shown. B, Mg50 values derived from the $MgCl_2$ concentration at which 50% oligomerization occurs. Error bars represent the S.D. of 3 or more independent oligomerization assays.

2.4 Discussion

2.4.1 Requirement of the core histone N-terminal "tail" domain during nucleosomal array oligomerization

There is strong evidence indicating that *in vitro* nucleosomal array oligomerization is a manifestation of the long range fiber-fiber interactions that lead to higher order chromatin structures (Lu *et al*, 2006; McBryant *et al*, 200). There have been several studies detailing the role of the histone tails in nucleosomal array oligomerization using the 5S rDNA 208-12 nucleosomal array model system. The Widom 601_207-12 nucleosomal array model system has not been extensively used in past oligomerization studies. The advantage of using the Widom 601 sequence comes from using a tandemly repeated single high affinity nucleosome position that consequently leads to more homogenous linker DNA lengths generating therefore a more homogenous population of nucleosomal arrays. The disadvantage is that, to date it is unknown whether the 601_207-12 array maintains similar solution-state behaviors as that of the 5S_208-12 nucleosomal array. Therefore, it was necessary to repeat many of the early experiments that have provided us with our current knowledge of the role of the core histone tails in array oligomerization to establish the basic solution state properties of the fully saturated WT and TL 207-12 nucleosomal arrays.

We found that both the WT and TL 207-12 arrays behave very similar to what has been previously published with their 208-12 counterparts. Figure 2.1A shows that the WT 12-mer array maintains a S_{ave} of 29.5S in low salt 1xTEN buffer, whereas a fully saturated completely tailless array sediments at 25S. This is in good agreement with past literature using the 5S_208-12 model system (Tse and Hansen, 1997; Carruther *et al*, 1998). Furthermore, the WT array cooperatively oligomerizes with increasing $MgCl_2$ concentrations, whereas the TL array is unable to oligomerize at any $MgCl_2$ concentration tested, figure 2.3, data not shown. Past studies have shown that the

arrays containing the 601 sequence induce array compaction at lower $MgCl_2$ concentrations than that of their 5S rDNA counterparts (Tse and Hansen, 1997; Dorigo *et al*, 2003). Fully saturated wild type 5S_208-12 arrays have previously been reported to maintain Mg50 values between 2.5 and 3.0 mM $MgCl_2$ (Schwarz and Hansen, 1994; Tse and Hansen; 1997; Wang and Hayes, 2008). Here, I report that the wild type 207-12 array maintains an Mg50 of 1.98mM $MgCl_2$, again showing that the 601 sequence containing arrays oligomerize at lower $MgCl_2$ concentrations.

2.4.2 Independent and additive contributions of the histone tails in nucleosomal array oligomerization

Previous research reports on the role of the histone N-terminal tail domains using sub-saturated 5S_208-12 nucleosomal arrays describe how each of the tails contribute independently and additively to array oligomerization, but not equally (Gordon *et al*, 2005). The current research detailed how the H4 tail is the most important contributor to oligomerization, followed by the H3 tail and the H2A and H2B tails contributing equally. Importantly, neither dimer tail was shown to induce oligomerization independently and needs at least one of the other three histone tails (Gordon *et al*, 2005). Similar to 2005, many of the same observation reported in 2005 hold true for the 601_207-12 nucleosomal array (Gordon *et al*, 2005). For example, deletion of the H4 tail leads to the largest increase in Mg50 value (Fig 2.4), the H3 and H4 tails together achieve a lower Mg50 than that of the H2A and H2B tails together (Fig 2.5) and the H2A tail cannot independently induce oligomerization (Fig 2.6) (Gordon *et al*, 2005).

However, contrary to previous reports (2005), deletion of the H3, H2A and H2B tail lead to relatively similar increases in Mg50 values compared to WT arrays (Fig 2.4) and the H2B tail was able to independently induce oligomerization (Fig 2.6) (Gordon *et al*, 2005). Another important difference was that the H3 tail individually was able to induce oligomerization at lower $MgCl_2$ concentrations than the H4 tail, indicating an increased

role for the H3 tail in oligomerization. The H2B tail also takes on an increased role in oligomerization in the model system used in this study.

Probable differences in oligomerization between these two studies stem from the differences in the model systems. The fully saturated 601_207-12 nucleosomal array used here maintains a very homogenous population of arrays each of which maintain an even distribution of nucleosomes per template relative to the sub-saturated 5S_208-12 nucleosomal array. Notably, the fully saturated array would be more condensed intra-molecularly at any $MgCl_2$ concentration that induces oligomerization than that of the sub-saturated arrays. By averaging four more nucleosomes per template the fully saturated 601_207-12 array by design also adds 16 more tails per array. The increased proximity of nucleosomes within each fully saturated array in addition to an increased number of tails would potentially increase the number of interactions that the tail domains are involved in.

2.4.3 No single physical characteristic of the H4 tail determines its role in nucleosomal array oligomerization.

In 2009, using "tail-swap" H4 mutants, we found that the H4 tail's mode of action centered on four determinants: 1) amino acid composition, not primary sequence, 2) tail domain length, 3) charge density and 4) tail position within the nucleosome. This study predominantly used the 5S rDNA 208-12 model system with sub-saturated (8 nucleosomes out of 12 possible) nucleosomal arrays. The research reported here, using a fully saturated 601_207-12 nucleosomal array and only single tail containing nucleosomes has provided valuable supporting evidence and further resolution into the role of the H4 tail during oligomerization. For instance, with the exception of the shorter H2A tail, the H3 and H2B tail domains were able to replace the H4 tail achieving similar and sometimes lower Mg50 values than the WT H4 tail (Fig 2.7b). Furthermore, repositioning the H4 tail to the H2B tail's location within the nucleosome increases the

Mg50 almost 5.0 mM drastically decreasing the arrays ability to oligomerization when no other tail is present (Fig 2.7d). Together, these results indicate that the amino acid composition and position of the H4 tail are important molecular determinants to nucleosomal array regardless of the nucleosomal array model system (Fig 2.7d). We gain further insight into the role of the tails during oligomerization, because we now know the direct contribution of the individual tails from their native location and the H4 tail's position. For instance, TL+H2B is only able to achieve an Mg50 of 15.1 mM $MgCl_2$, however, once repositioned to the H4 location within the nucleosome (TL+^{T}2B^{HF}4), the Mg50 decreases by more than half to 7.77 mM (Table 1, Figs, 2.5, 2.6).

As with amino acid composition and tail position, the H4 tail length and positive charge density have been shown to be important molecular determinants for nucleosomal array oligomerization (McBryant *et al*, 2009). Here we show that in a single tail nucleosomal array model system, increasing the tail length increases the ability of the array to oligomerize. In fact, tripling the length of the wild type H4 tail from 27 to 84 amino acids enabled arrays to oligomerize at similar to even lower Mg50 values than when the full complement of histone tails were present in the wild type array. Increasing the positive charge density of a shorter H4 tail (TL+^{T}[2A]$_{CD}$^{HF}4) overcomes the inability of the array to oligomerize, whereas decreasing positive charge density of the WT H4 tail completely abolishes the arrays ability to oligomerize similar to shortening the length of the tail. Together these data provide strong evidence that the charge neutralization of the DNA phosphate backbone is a key molecular determinant of the H4 tail during nucleosomal array oligomerization. As discussed in our 2009 oligomerization study, despite the highly conserved amino acid sequence of the H4 tail, the amino acid composition, positive charge density and length of the H4 tail has evolved in such a way as to balance the state of chromatin compaction (McBryant *et al*, 2009).

As detailed in chapter 2.3.5, the primary sequence of the H4 tail not necessarily a contributing factor to oligomerization (H3 and H2B tails can replace the H4 tail). However, when we scramble the primary sequence of the H4 tail to truly test for an amino acid composition prerequisite versus primary sequence the array loses oligomerization efficiency. The same phenomenon occurs with the TL+H3/H4KR nucleosomal array, almost doubling the Mg50 compared to the wild type counterpart. Interestingly then, there are in certain contexts, a yet to be explained primary sequence component to nucleosomal array oligomerization. One could speculate that the core histone tail domains have evolved to maintain a yet to be defined and conserved sequence motif that is disrupted after the sequence is scrambled. However, one explanation for the TL+H3/H4KR arrays increase in Mg50 was probably due to its increased intra-array folding detailed in section 3.4.5. The K→R mutations actually dramatically increase an arrays ability to stably fold as judged by a homogenous distribution of 40-55S sedimentation coefficients compared to a WT array (29-55S, Fig 3.10). The increased arginine content of the H3 and H4 tails may then be stably bound to their intra-array partners, and so are not readily available to participate in inter-array oligomerization interactions.

Chapter 3

Nucleosomal Array Folding: Core histone tail contributions are position and context dependent, but primary sequence independent.

Previous studies detailing contributions of the core histone tails towards nucleosomal array folding have indicated that the H4 tail is the only contributing factor (Dorigo *et al*, 2003). In this chapter I use the same nucleosomal arrays used to study oligomerization in chapter two to better understand the H4 tail's mode of action that drives nucleosomal array folding. I found that the H4 and the H3 tails are the most important contributors to folding. Furthermore, the mode of action for the H4 tail driving nucleosomal array folding is not dependent on its primary sequence, but instead depends on intrinsic disorder of its amino acid sequence and on positive charge density. The length and position of the H4 tail within the nucleosome are also important molecular determinants for nucleosomal array folding.

3.1 Introduction

The core histone N-terminal "tail" domains consist of 13-37 unstructured amino acids extending out from the nucleosome (Luger *et al*, 1997a). About 33% of the amino acids in each tail are lysines or arginines providing eight dynamic and highly positive charged tails around the exterior of each nucleosome required for nucleosome-nucleosome interactions. Several research papers have shown that the histone tails are responsible for *in vitro* nucleosomal array condensation (Garcia-Rameriz *et al*, 1992; Tse and

Hansen, 1997; Carruthers and Hansen, 2000; Dorigo *et al*, 2003; Hizume *et al*, 2009). For instance, the tail domains are major contributors to array folding in cis, as completely tailless arrays with and without linker histones are only able to moderately fold at the $MgCl_2$ concentrations tested (Fletcher and Hansen, 1995; Schwartz *et al*, 1996; Tse and Hansen, 1997; Carruthers and Hansen, 2000; Dorigo *et al*, 2003). The tail domains are thought to act primarily through electrostatic mechanisms during array condensation, although a simple charge neutralization model does not sufficiently describe the mode of action of the tails (Polach *et al*, 2000; Hansen, 2002; Shogren-Knaak *et al*, 2006; Robinson *et al*, 2008; McBryant *et al*, 2009; Korolev *et al*, 2010; Sinha and Shogren-Knaak, 2010). Disulfide-bond and photo-affinity cross-linking studies have captured the dynamic nature of the tail domains interacting with both DNA as well as other proteins during intra-molucular folding events (Dorigo *et al*, 2004; Kan *et al*, 2007; Kan *et al*, 2009; Sinha and Shogren-Knaak, 2010). Although the contribution of the histone tail domains is well-defined for nucleosomal array oligomerization, it remains unclear as to how the enigmatic histone tails function to induce extensive nucleosomal array folding.

Of the four histone tails, the H4 tail has received the most attention in recent years with regards to its role in chromatin condensation (Dorigo *et al*, 2003; Gordon *et al*, 2005; Shogren-Knaak *et al*, 2006; Robinson *et al*, 2008; Kan *et al*, 2009; McBryant *et al*, 2009; Blacketer *et al*, 2010; Korolev *et al*, 2010; Sinha and Shogren-Knaak, 2010). The crystal structure of the nucleosome displays an H4 tail in direct contact with the H2A-H2B acidic patch of a neighboring nucleosome within the crystal lattice (Luger *et al*, 1997a). Furthermore, the primary sequence of the H4 tail (in particular amino acids 14-19) has been directly implicated in driving chromatin condensation (Dorigo *et al*, 2003; Dorigo *et al*, 2004; Kan et al, 2009). Within this basic motif (residues 14-19), H4 K16 acetylation disruptx chromatin condensation *in vitro* and *in vivo* to the same extent as deletion of the entire tail (Shogren-Knaak *et al,* 2006; Robinson *et al*, 2008; Wang *et al*,

2008; Korolev *et al*, 2010). Although the evidence clearly points towards the H4 tail's dominant role in chromatin condensation, the H4 tail's mechanism of action during nucleosomal array folding is still undefined.

We have previously documented the molecular determinants of the H4 tail domain that facilitate nucleosome-nucleosome interactions in trans (Chapter 2; McBryant *et al*, 2009). In this study, we established that the H4 tail mediates oligomerization via four important determinants: amino acid composition, positive charge density, length and most profoundly its position within the nucleosome specified by the H4 N-terminal exit orientation (McBryant *et al*, 2009). However, oligomerization and folding are two separate condensation events and the contributions of the histone tails has previously been shown to be different in each transition (Dorigo *et al*, 2003; Gordon *et al*, 2005). In this chapter I extended our previous study to focus on the H4 tail's molecular determinants that drive nucleosomal array folding. Despite the mounting evidence of the importance of the H4 tail's primary sequence, my data indicate that analogous to oligomerization, the H4 tail acts through its position within the nucleosome, as well as the amino acid composition, intrinsic disorder and length of the tail domain. My data clearly indicates that the primary sequence of the H4 tail is not a driving factor in array folding.

3.2 Materials and Methods

3.2.1 Sedimentation velocity of folded nucleosomal arrays

Sedimentation velocity provides a sensitive measurement of the degree to which a nucleosomal array compacts intra-molecularly in the presence of cationic metals (Mg^{2+}) (Schwarz and Hansen, 1994; Schwarz *et al*., 1996; Tse and Hansen, 1997; Carruthers and Hansen, 2000; Dorigo *et al*, 2003; Shogren-Knaak *et al*, 2006; Grigoreyv *et al*, 2009; Korolev *et al*, 2010; Sinha and Shogren-Knaak, 2010). Folding studies were performed using sedimentation velocity in Beckman Xl-A and I analytical ultracentrifuges equipped

with scanner optics, 12-mm double sector cells and four or eight-hole rotors. Sedimentation velocity was performed at 12,000 to 30,000 rpms at room temperature for two to four hours. Boundaries were analyzed by the method of van Holde and Weischet using Ultrascan 2.0 v9.9. This method and software determines the diffusion corrected sedimentation coefficients corrected for water at 20ºC ($S_{20,w}$) using a partial specific volume of 0.622 ml/g and density and viscosity parameters specific to the buffer solution (Schwarz and Hansen 1994; Tse and Hansen, 1997; Demeler and van Holde, 2004). Folding was examined in the $MgCl_2$ range that precedes the onset of oligomerization (typically 0-1.75mM $MgCl_2$). At least two independent 12-mer nucleosomal arrays were reconstituted and assayed for folding two or more times at each respective $MgCl_2$ concentration, unless otherwise noted. The most representative integral distribution at each $MgCl_2$ concentrations is reported.

3.3 Results

3.3.1 Effect of the core histone tails toward nucleosomal array folding

Sedimentation velocity analysis was used to study nucleosomal array folding on arrays containing an average of 12 nucleosomes per template over a range of $MgCl_2$ concentrations. I used the same WT and TL arrays described previously (Chapter 2) for their ability to oligomerize, to study their ability to fold intra-molecularly in the presences of low concentrations of $MgCl_2$. In low salt buffer 1xTEN, the fully saturated WT and TL arrays sediment at 29 and 25S, respectively. Under conditions in which divalent cations are present (in this case Mg^{2+}) both types of nucleosomal arrays condense into progressively more compact structures with increasing salt concentrations (Fig. 3.1). However, the extent of compaction is drastically different for arrays with and without the histone tails. Nucleosomal arrays containing all full length histone tails exhibit two different populations of condensed arrays. The first and slower sedimenting population of nucleosomal arrays sediments between the extended 29S beads on a string

conformation and a moderately folded 40S conformation. A second population of arrays is observed sedimenting between the 40S conformation and an extensively folded 55S conformation. Conversely, arrays containing no histone tails are only able to condense into the moderately folded 40S conformation. Importantly, the TL array in the presence of $MgCl_2$ only sediments between the 29 and 40S conformations (not 25S), indicating that $MgCl_2$ is able to replace the tail domains to form a 29S extended structure versus the 25S conformation without $MgCl_2$ (Fig. 3.1A). Figure 3.1A and B, depicts how each type of array progressively condenses with increasing $MgCl_2$ concentrations. Figure 3.1A and B, also intuitively show how the histone tails are able to increase the state of compaction of the nucleosomal array to a maximum of approximately 55S.

3.3.2 Effect of the loss of a single tail on array folding

I previously showed that tailless nucleosomal arrays are unable to fold past the 40S state (Fig. 3.1; Tse and Hansen, 1997; Dorigo *et al*, 2003). I was curious to see how many and which tails might be needed to regain WT-like folding capacities. Four nucleosomal arrays were reconstituted to an average of 28.5S and contained three WT histones and one tailless histone (Fig. 2.4). Dramatic differences in the extent of compaction were observed when a tail from the tetramer position was missing compared to when either dimer tail is missing. Deleting the H2A or H2B tail did not significantly change the folding profile compared to the WT arrays. However, deleting either the H3 or H4 tail dramatically decreased the ability of the arrays to fold, generating folding states with only marginally higher S values than the completely tailless array (Fig. 3.2A and B). Below 1.25 mM $MgCl_2$, H3TL and H4TL fold to similar extents, not until 1.75mM Mg does one see appreciable difference (Fig. 3.2B). When the H4 tail is present with the dimer tails (H3TL) at high enough $MgCl_2$ concentrations (1.75mM) the array

compacts into extensively folded arrays (Fig. 3.2B). The H4TL array (H3, H2A and H2B tails) was unable to compact to the same degree, indicating that the H4 tail domain can induce nucleosomal array folding to a higher extent than the H3 tail (Fig. 3.2B).

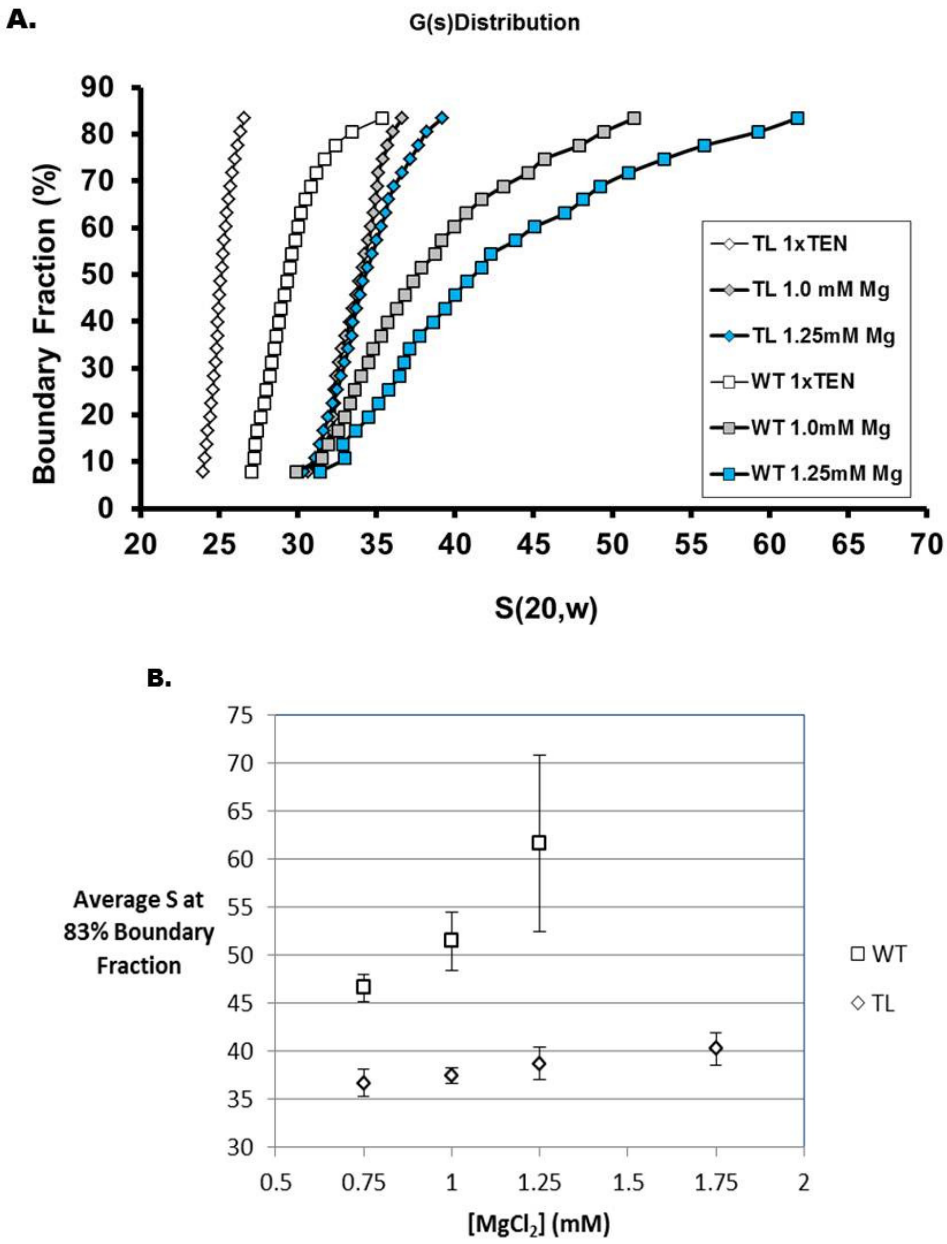

Figure 3.1 Folding of WT and TL nucleosomal arrays with increasing amounts of $MgCl_2$. Fully saturated nucleosomal arrays containing full length (WT) histone octamers progressively compact with increasing concentrations of $MgCl_2$. WT arrays can compact into extensively folded fibers; TL arrays are only able to moderately fold at the same $MgCl_2$ concentrations. A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) at 0 mM $MgCl_2$ (1xTEN), 1.0 mM and 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. B, Maximum sedimentation coefficients of WT and TL nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of 3 or more independent oligomerization assays.

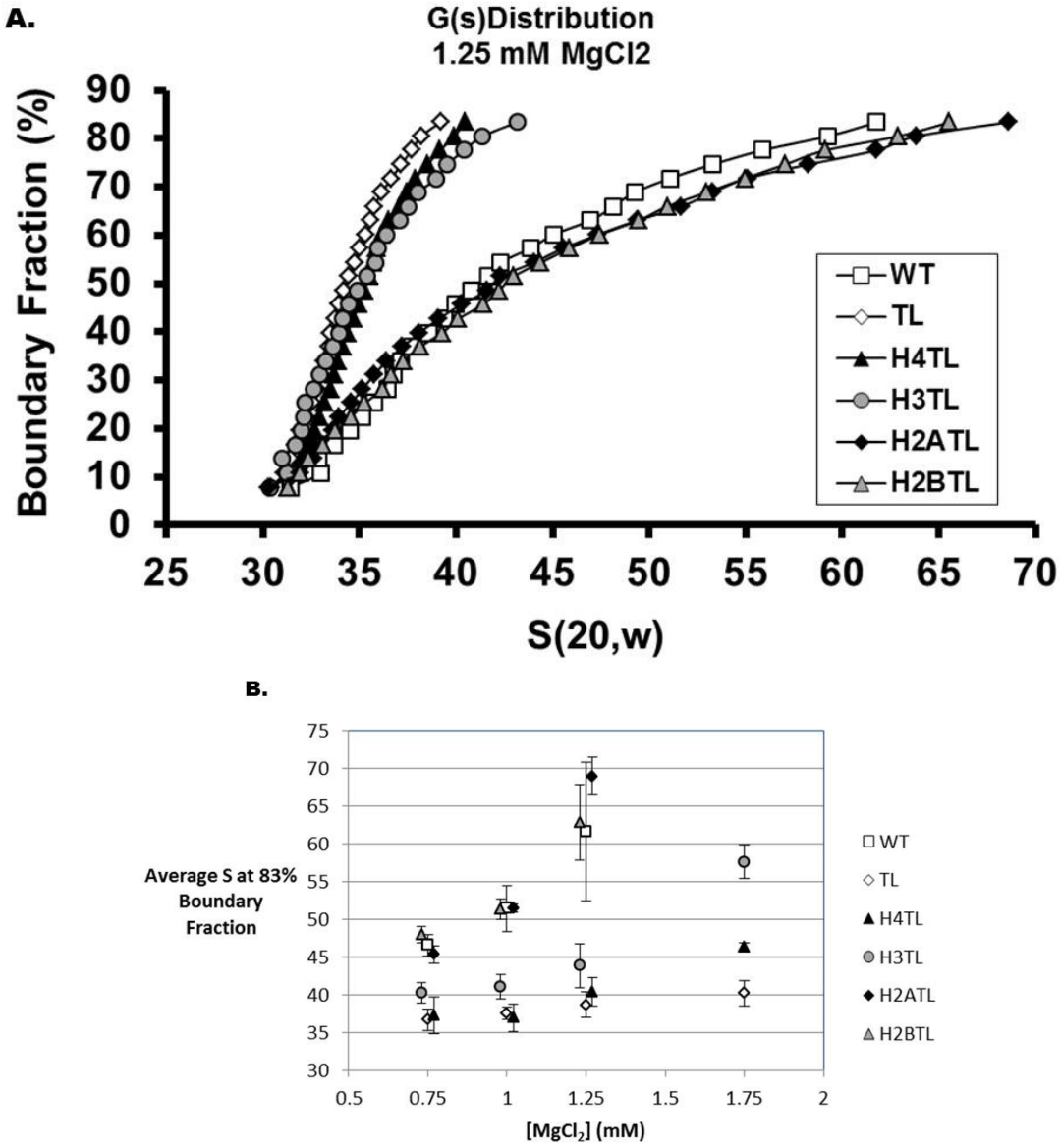

Figure 3.2 Folding of nucleosomal arrays containing a single tail deletion. Fully saturated nucleosomal arrays containing a single tail deletion progressively compact with increasing concentrations of $MgCl_2$. Deletion of the H3 or H4 tail leads to the largest decrease in array folding A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter two for corresponding G(s) Distributions in 1xTEN B, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays.

3.3.3 Minimal tail "requirement" for maximal nucleosomal array folding

In the previous section I described how deletion of either the H2A or H2B tail did not affect folding and deleting either core histone tail from the tetramer position leads to a drastic decrease in the extent of compaction. To further investigate the differences in the contributions to folding from the tetramer and dimer tails, two additional arrays were generated. The first array contained the wild type H3 and H4 and tailless H2A and H2B histone designated TL+H3/H4. The second array contained wild type dimer histones and tailless tetramer histones designated TL+H2A/H2B. The TL+H3/H4 array was able to compact to the same extent as the WT nucleosomal array, indicating that neither dimer tail is required for extensive folding (Fig. 3.3A and B). Below 50% of the boundary, the TL+H3/H4 was less compact compared to the WT array as judged by the left-shifted distribution (Fig. 3.3A). When only the two dimer tails (TL+H2A/H2B) were present, minimal compaction was observed over that of the completely tailless array (Fig. 3.3A and B). Again it appears that the H2A and H2B tails are minimally participating in array folding.

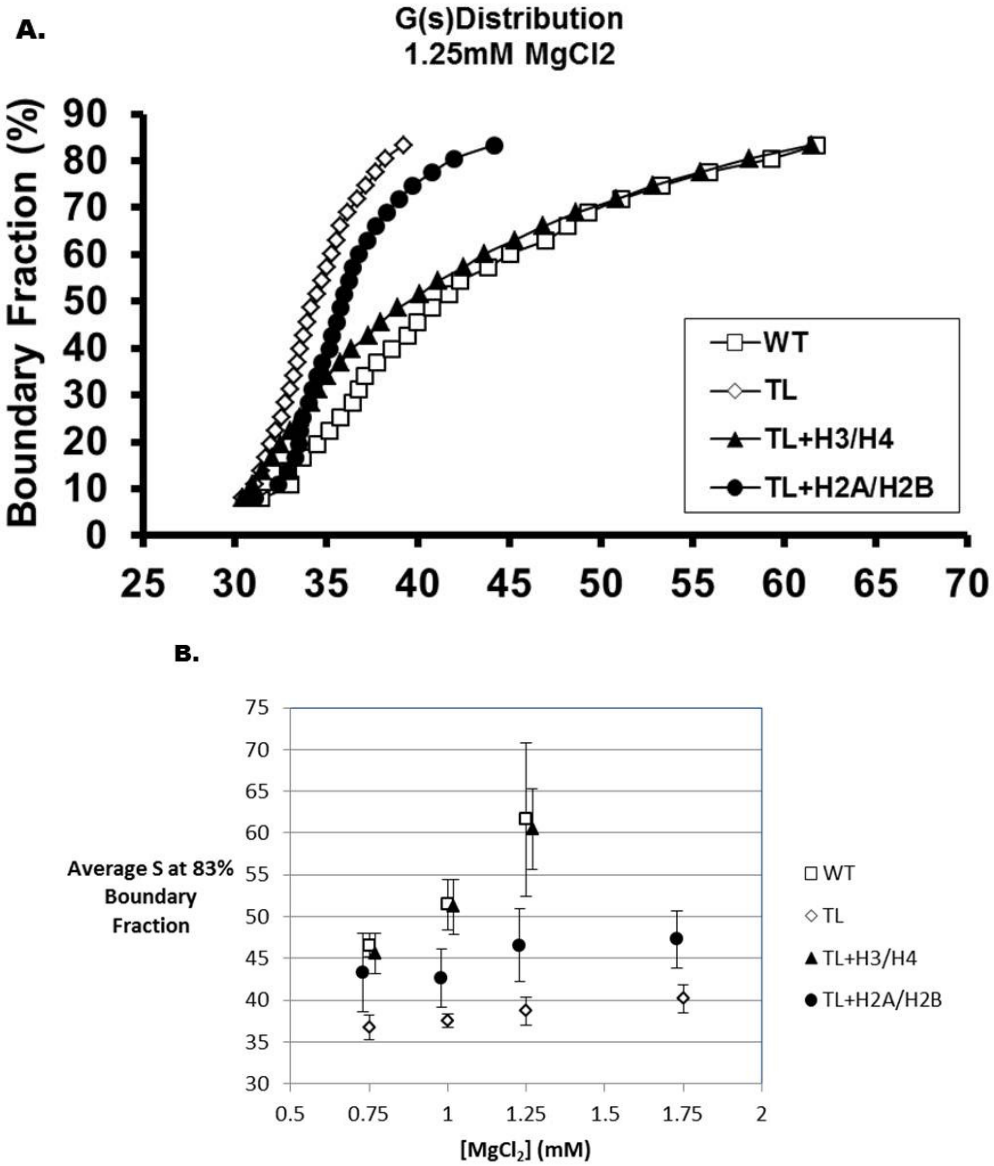

Figure 3.3 Folding of nucleosomal arrays containing only the H3 and H4 or H2A and H2B tails. Fully saturated nucleosomal arrays containing the full length H3 and H4 and TL H2A or H2B or full length H2A and H2B and TL H3 and H4 histones progressively compact with increasing concentrations of $MgCl_2$. The H3 and H4 tail combine can extensively fold to WT levels. A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter two for corresponding G(s) Distributions in 1xTEN B, Maximum sedimentation

3.3.4 Individual contributions of the H4, H3, H2A and H2B tails to folding

Next I sought to determine if any one histone tail could compact arrays to the same extent as when all four histone tails are present. The same four single tail containing nucleosomal arrays from Section 2.3.4 were assessed for their ability to induce array compaction intra-molecularly. Each of the four "single-tail" arrays progressively condensed into more compact conformations with increasing Mg concentrations (Fig. 3.4A and B). However, no tail by itself was able to compact to the same extent as that of WT arrays (Fig. 3.4A and B). The H2A and H2B tails individually are unable to condense nucleosomal arrays more than a completely tailless array, (Fig. 3.4A and B). However, when either of the H2A or H2B tails were present in a tailless background the entire G(s) distribution was slightly, but reproducibly right-shifted compared to the TL array (Fig. 3.4A). Although the H2A and H2B tails contribute little to folding, the tails extending from the tetramer position (H3 or H4) were able to compact arrays considerably more than that of completely tailless array. Figure 3.4A and B, shows how both the H3 and H4 tails alone were able to condense into folded conformations that sediment above 40S. At 1.25mM and 1.75mM $MgCl_2$, both the TL+H3 and TL+H4 arrays were more significantly compacted than the TL arrays (Fig. 3.4B), but were obviously deficient in folding capabilities compared to when both tails were present (compare Fig. 3.4A and 3.3A).

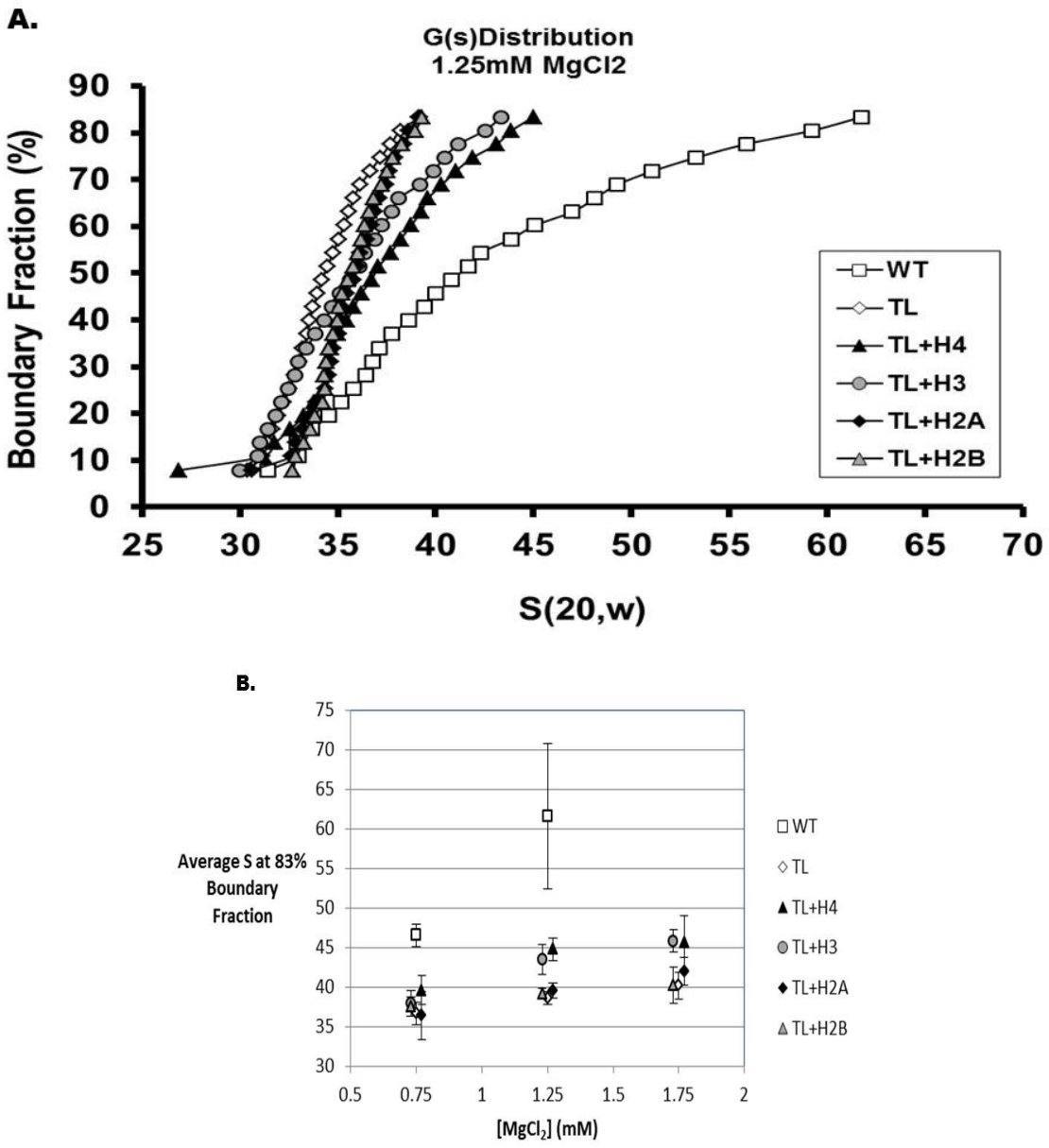

Figure 3.4 Folding of nucleosomal arrays containing a single full length histone. Fully saturated nucleosomal arrays containing a full length histone and three TL histones only minimally fold to a higher degree than the completely TL array. No tail alone is able to compact to the same degree as WT arrays. The H3 or H4 tails do fold to a greater degree than the H2A or H2B tails. A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter 2.4 for corresponding G(s) Distributions in 1xTEN B, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays.

To compare differences in folding between arrays containing different tails, I combined different distributions from the two sets of arrays (single-tail containing versus single-tail deletions) (Fig. 3.5). Arrays containing only the H4 tail (TL+H4) or the H3 tail (not shown) compacted similarly as that the H4TL array (containing the H3, H2A and H2B tails) in 1.75mM $MgCl_2$. This is in contrast to arrays containing the H4, H2A and H2B tails (H3TL) in 1.75mM $MgCl_2$ which compacted to the same extent as WT arrays in 1.25mM $MgCl_2$ (Fig. 3.5A). At high enough $MgCl_2$ concentrations, the H4 (or H3) tail individually folded arrays to the same extent as the H3, H2A and H2B tails combined. Also at high enough $MgCl_2$ concentrations, the H2A and H2B tails were able to contribute to folding, but only if the H4 tail was present (H3TL) and not the H3 tail (H4TL).

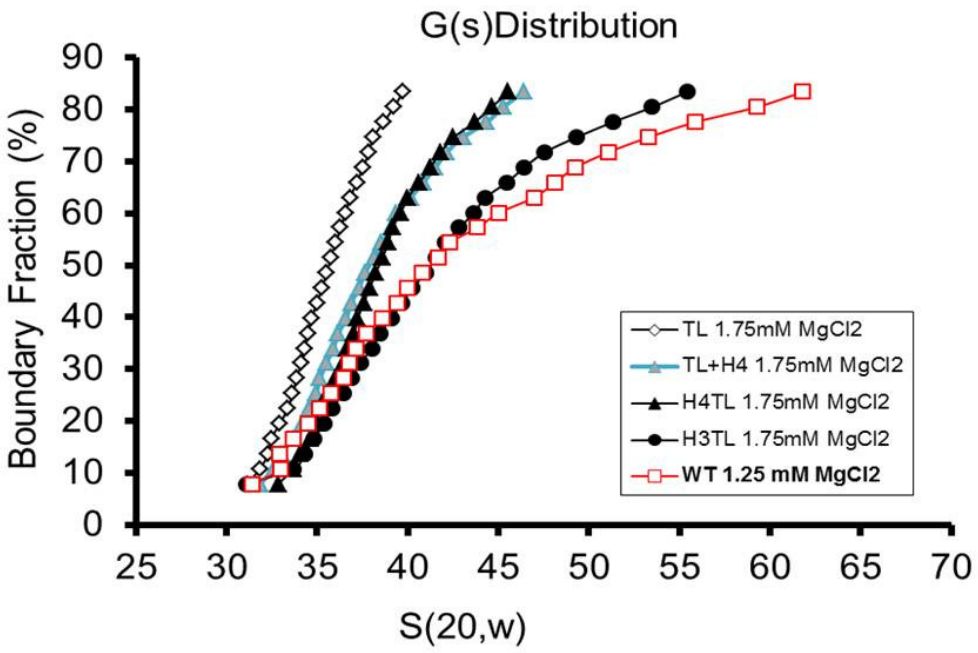

Figure 3.5 Folding comparisons of TL+H4, H4TL and H3TL nucleosomal arrays. The H4 tail (TL+H4) independently folds to the same degree as the H3, H2A and H2B tail combined (H4TL) at 1.75mM $MgCl_2$. The H4 tail combined with the H2A and H2B tails (H3TL) and increased $MgCl_2$ can reach WT folding levels. Note, the WT array is in 1.25mM $MgCl_2$, TL, TL+H4, H4TL and H3TL arrays are in 1.75mM $MgCl_2$. Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter 2.3 for corresponding G(s) Distributions in 1xTEN.

3.3.5 Effects of tail "swapped" histones on nucleosomal array folding

Past literature and the data I have presented so far suggest that the H4 tail is the primary determinant in nucleosomal array condensation (Chapter 2; Fig. 3.5; Dorigo *et al*, 2003; Gordon *et al*, 2005; McBryant *et al*, 2009). To determine the mode of action used by the H4 tail to direct compaction, used mutant H4 histones that replaced the native H4 tail with the other three histone tails (see figure 2.1 and section 2.10 for mutant descriptions). I examined each of the arrays ability to fold in the presence of 1.75 mM $MgCl_2$ or lower using the same arrays characterized for oligomerization in chapter 2. Similar to oligomerization, "tail-swapped" mutant arrays tested for the following determinants in folding: primary sequence, intrinsic disorder and/or amino acid composition, and if the position of the H4 tail within each nucleosome was a contributing factor.

When the H3 tail is moved to the H4 tail's position within the nucleosome (TL+$^{T}3^{HF}4$) there is relatively no change in the distribution at 1.25mM Mg, see figure 3.6A and B. In contrast, the two dimer tails (TL+ $^{T}2A^{HF}4$ and TL+$^{T}2B^{HF}4$) as a whole fold differently than that of the H4 tail (Fig. 3.6A). The shorter tail of the two (TL+$^{T}2A^{HF}4$) showed a highly reproducible, homogenous and right shifted distribution of sedimentation coefficients below 40S compared to TL+H4 and TL+$^{T}3^{HF}4$, figure 3.6a. Conversely, the TL+$^{T}2B^{HF}4$ nucleosomal array folds similarly to that of the completely tailless array (not the TL+H4 array) in the bottom portion of the distribution, and only at the top 60 to 80 % of the boundary is it able to compact to a higher degree than the TL array (Fig. 3.6A). However by 1.75mM $MgCl_2$, the TL+$^{T}2B^{HF}4$ array is able to fold similarly to that of the TL+H4 array across the entire distribution (data not shown). Importantly, all three H4 tail-swap mutants achieved similar maximum sedimentation coefficients measured at 83% of the boundary fraction (Fig. 3.6B). Despite differences in the level of moderate folding, all three tails can compact arrays to the same extent as the TL+H4 array,

indicating that the primary sequence of the H4 tail is not a molecular determinant of extensive nucleosomal array folding.

In 2009 we hypothesized that if the H4 tails molecular determinant that induced oligomerization lies in its amino acid sequence, then adding additional H4 tails within the nucleosome would create a super condensing array (McBryant *et al*, 2009). We reconstituted a WT arrays that contained the mutant "tail-swapped" H2B histone (the H4 tail replaced the H2B tail, $^{T}4^{HF}$2B) so that the array contained four total H4 tails instead of two tails. Oligomerization of the $^{T}4^{HF}$2B arrays was similar to the WT array, indicating that the position of the H4 tail within the nucleosome is a major determinant to the H4 tails contribution to oligomerization. To determine whether the H4 tails position within the nucleosome was also a major determinant for array folding, I reconstituted a tailless array containing the full length $^{T}4^{HF}$2B tail swap mutant to a s_{ave} of 26.5S. As observed for oligomerization, the position of the H4 tail in an important determinant of folding. The TL+$^{T}4^{HF}$2B array folds almost identically to that of the TL+H2B nucleosomal array, and is unable to compact arrays past the 40S conformation (Fig. 3.6C and D).

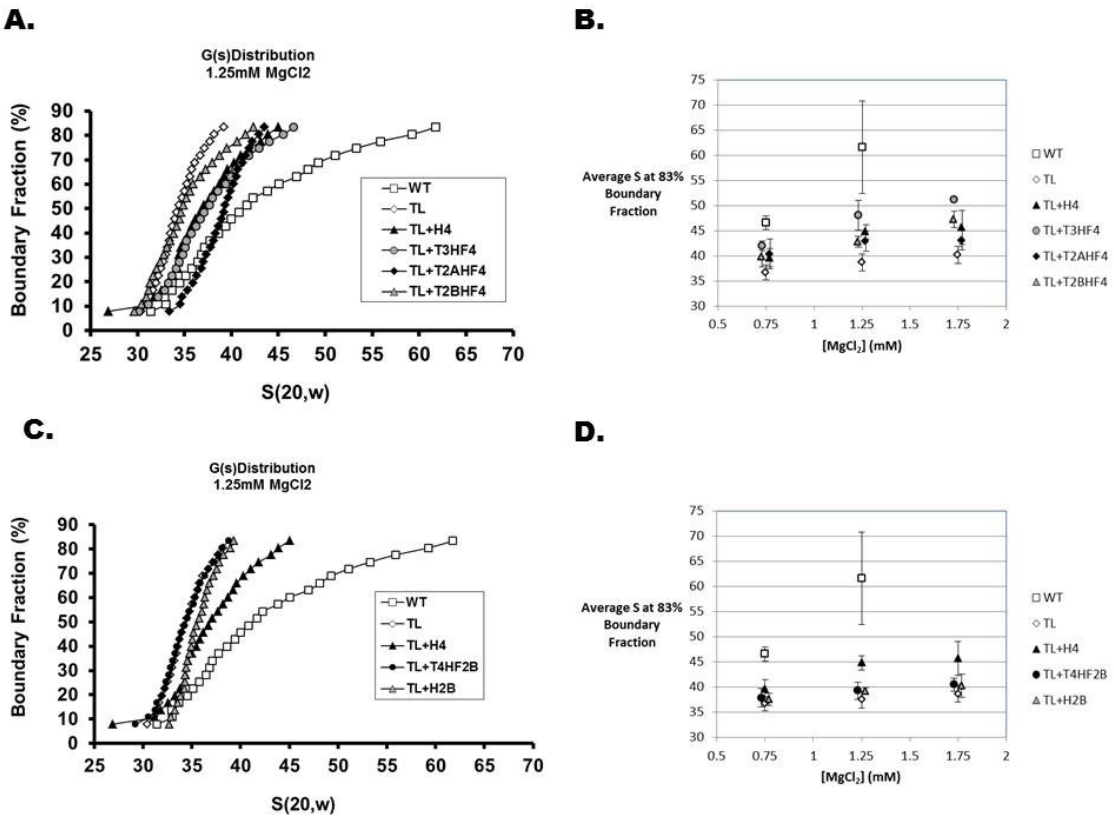

Figure 3.6 Folding of nucleosomal arrays containing a single "tail" swap histone. Fully saturated nucleosomal arrays containing a single "tail" swap histone and three TL histones only minimally fold to a higher degree than the completely TL array. The H3, H2A and H2B tails are able to reach a similar maximum s_{ave} as the WT H4 tail (B). A and C, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter two for corresponding G(s) Distributions in 1xTEN B and D, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays.

3.3.6 Consequence of scrambling the primary sequence of the H4 tails on folding

Figure 3.6 showed that the primary sequence of the H4 tail is not necessarily a contributing factor to nucleosomal array folding (TL+$^{T}3^{HF}$4 array compacts to the same extent as that of the TL+H4 array). To directly test whether the H4 tail's primary sequence plays a direct role in array folding I reconstituted a tailless array which contained an H4 tail mutant with the first 27 amino acids randomized (TL+$^{T}[4]_{R1}{}^{HF}$4) (Fig. 2.2). At every $MgCl_2$ tested, the TL+$^{T}[4]_{R1}{}^{HF}$4 array was able to compact to the same extent as the wild type, TL+H4 array, indicating that the H4 tail's primary sequence is not a major determinant to array folding (Fig. 3.7).

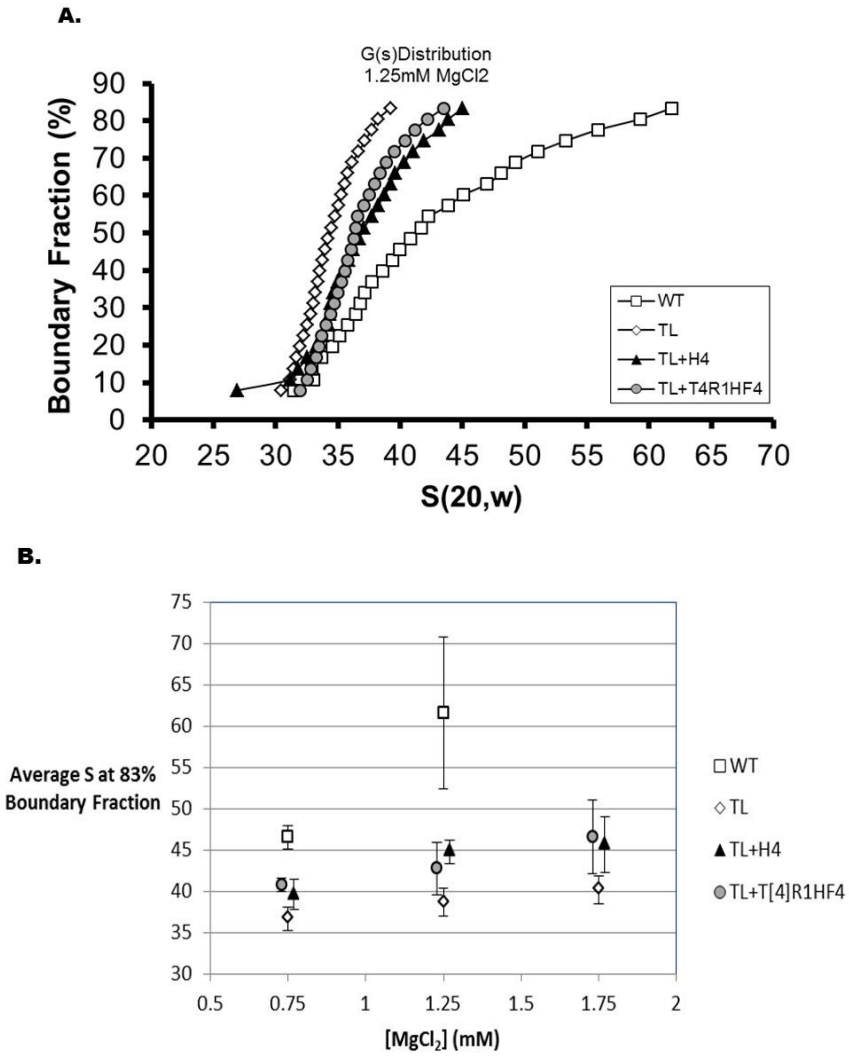

Figure 3.7 Folding of nucleosomal arrays containing a single "randomized" H4 tail domain. Fully saturated nucleosomal arrays containing a randomized H4 tail domain and three TL histones only minimally fold to the same extent as the WT TL+H4 array. A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter 2.4 for corresponding G(s) Distributions in 1xTEN B, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays.

3.3.7 Consequence of increasing the length of the H4 tail on moderate and extensively folded nucleosomal arrays

We have previously published, and I have shown in figure 2.9, that the length of the H4 tail is a primary molecular determinant for array oligomerization (McBryant *et al*, 2009). In the case of oligomerization, increasing the length of the H4 tail increased its ability to oligomerize (i.e., decreased the amount of Mg needed to induce 50% oligomerization) (Fig. 2.9B). It is important to note that oligomerization and folding are independent condensation transitions. To determine if the length of the H4 tail also plays a role in array folding I compared the TL+$^{T}[4]_{2}^{HF}$4 (twice the length of the H4 tail) with the wild type TL+H4 array. The longer H4 tail was able to compact intra-molecularly to a higher degree than that of the wild type H4, reaching a maximum of 51S in 1.25mM $MgCl_2$ versus 45S, respectively (Fig. 3.8A). By 1.75mM $MgCl_2$, the TL+$^{T}[4]_{2}^{HF}$4 array was substantially more folded than the TL+H4 array, achieving almost 30% of the boundary sedimenting between 40-57S, whereas the TL+H4 array only reached 46S (Fig. 3.8B and data not shown. It is interesting to note that the TL+$^{T}[4]_{2}^{HF}$4 array was consistently less folded (left shifted distribution) in the bottom half of the boundary at every $MgCl_2$ concentration tested, possibly indicating that the longer tail was inhibiting moderate folding (figure 3.8A, data not shown). However, it is clear that at least for extensive folding, the length of the H4 tail is a major determinant.

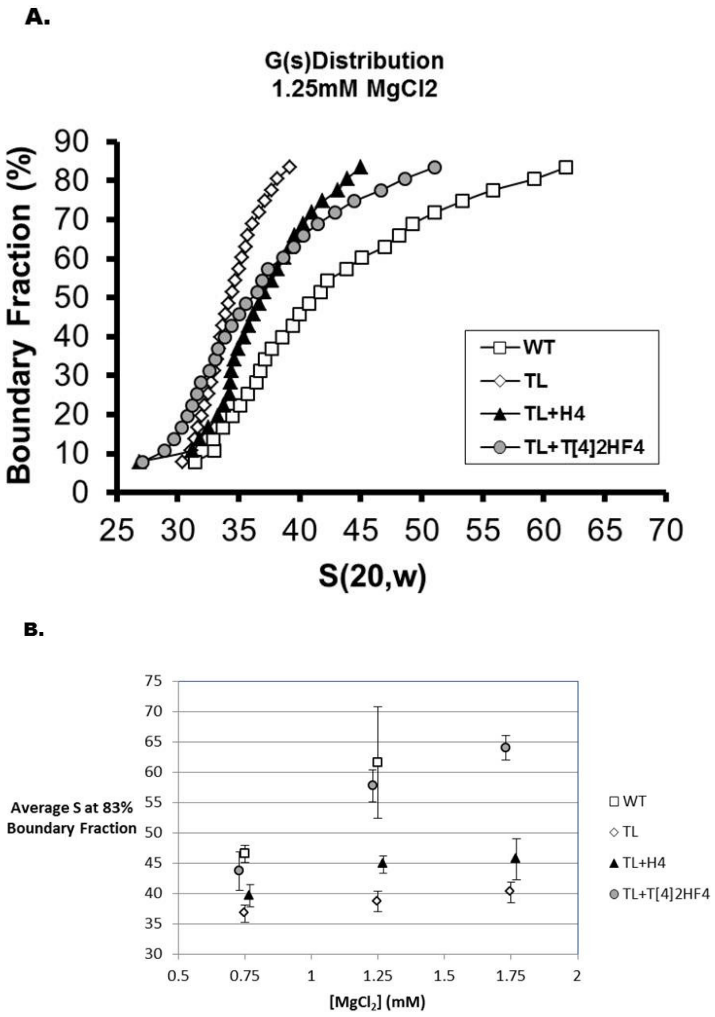

Figure 3.8 Folding of nucleosomal arrays containing a single "length" H4 tail mutant histone. Fully saturated nucleosomal arrays containing a H4 tail domain two time the length of the WT H4 tail is able to substantially compact to a greater degree. A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter 2.4 for corresponding G(s) Distributions in 1xTEN B, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays.

3.3.8 Influence of the positive charge density of the H4 tail on moderately folded nucleosomal arrays

As discussed in section 2.3.9, increasing the length of the H4 tail also increases the amount of positively charged amino acids available to help neutralize the negative charge of the linker DNA. I used the TL+^{T}[4]$_{CD}{}^{HF}$4 and TL+^{T}[2A]$_{CD}{}^{HF}$4 arrays (Chapter 2.3.9) to test how the positive charge of the H4 tail is involved in array folding. The TL+^{T}[4]$_{CD}{}^{HF}$4 array maintains a wild type length H4 tail, but decreases the positive charge density from 33% to 19%. Conversely, the TL+^{T}[2A]$_{CD}{}^{HF}$4 array maintains a tail half the length of the wild type H4 tail, yet increases the positive charge density from 33% to 62%. For a detailed description see section 2.3.9 and figure 2.2.

Folding experiments were performed as before. Interestingly there is really no significant difference in the amount of maximum compaction between the TL+H4, TL+^{T}[2A]^{HF}4,TL+^{T}[4]$_{CD}{}^{HF}$4 and TL+^{T}[2A]$_{CD}{}^{HF}$4 arrays, each reaching a maximum S value of 45, 44, 45 and 43S, respectively at 1.25mM $MgCl_2$ (Fig. 3.9A and B). However, the G(s) distribution below 40S of the TL+^{T}[4]$_{CD}{}^{HF}$4 array resembled that of the shorter tail, TL+^{T}[2A]^{HF}4 array more so than that of the wild type TL+H4 array. In contrast, the shorter tail charge density mutant array, TL+^{T}[2A]$_{CD}{}^{HF}$4 reproducibly maintained a distribution similar to that of the TL+H4 array (Fig. 3.9A). Curiously, a wild type length and positive charge density, or a shorter length and increased positive charge gave a more heterogeneous distribution, whereas a shorter tail with wild type charge density or wild type length and decreased charge density, generated a more homogenous distribution of folded arrays.

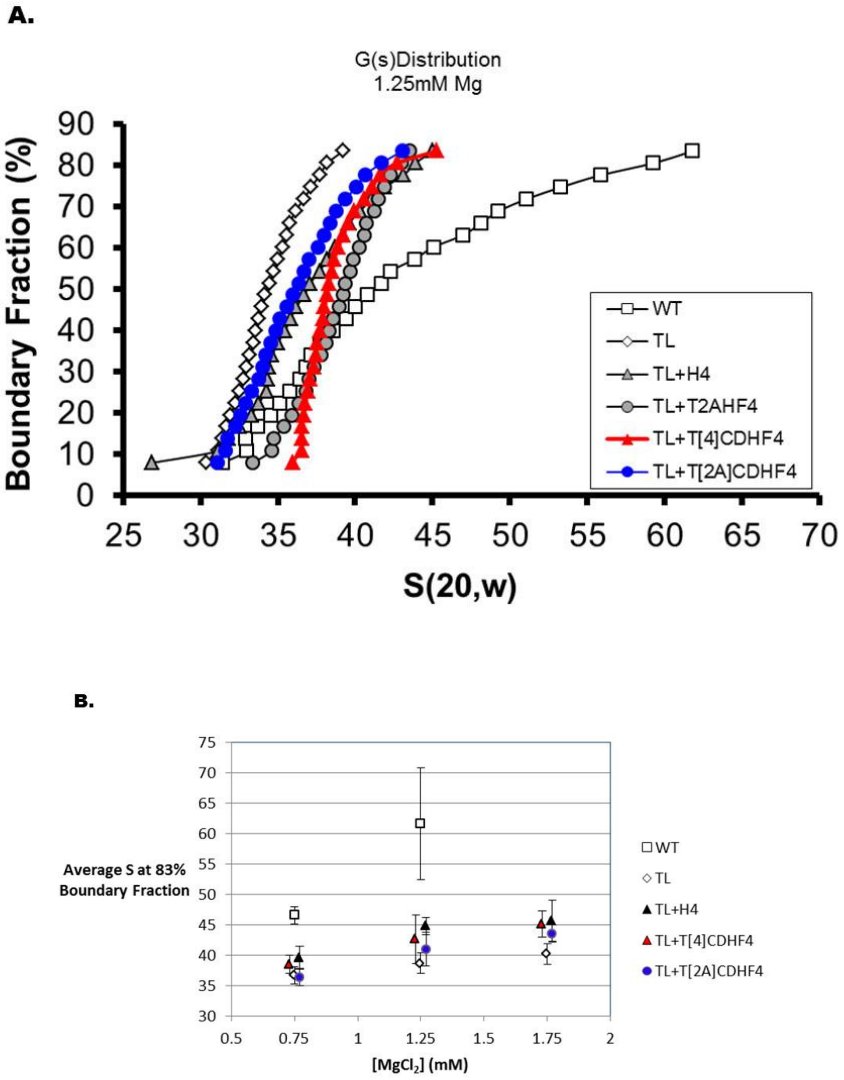

Figure 3.9 Folding of nucleosomal arrays containing a single "charge density" H4 tail mutant histone. Fully saturated nucleosomal arrays with a decreased positive charge density and WT length or increased positive charge density and a shorter tail length maximally compact the same extent. Decreasing the positive charge density facilitates moderate folding. A, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter 2.4 for corresponding G(s) Distributions in 1xTEN B, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays.

3.3.9 Consequence of replacing lysines with arginines in the H3 and H4 tail on the folding of nucleosomal arrays

To further explore the nature of the H4 tail's amino acid sequence and positive charge density, I tested H4 and H3 tail mutants that change the nature of the positively charged tails, but not the charge density or charge distributions. By using K→R tail mutants only the amino acid sequences changes and the positive charge density and distribution within the tail domain remain the same (Fig. 2.2). Using the TL+H4KR and TL+H3KR/H4KR arrays described in section 2.3.10 I performed folding experiments as described above. Whereas for oligomerization, the K→R mutants either mildly decreased the Mg50 (TL+H4KR) or increased the Mg50 (TL+H3KR/H4KR), during folding both arrays were able to substantially fold to a higher degree than that of their wild type counterparts, TL+H4 and TL+H3/H4, respectively (Fig. 3.10 A and B).

The TL+H4KR, reached a maximum of 52S in 1.25mM $MgCl_2$, with almost half the distribution above 40S, compared to a maximum of 45S and less than 20% of the distribution above 40S of the TL+H4 array (Fig. 3.10 A and B). Across the lower 50% of the distribution, the TL+H4KR array was considerably more compact than that of the WT array and was able to fold extensively at WT like levels (Fig. 3.10 C and D). The TL+H3KR/H4KR array reached virtually identical maximum S values as that of the TL+H3/H4 and WT array, each sedimenting between 59-62S at 1.25mM $MgCl_2$ (Fig. 3.10 C and D). The difference in the fraction of arrays sedimenting above 40S is striking between the three. WT and TL+H3/H4 arrays fold very similarly, each maintaining 32 and 29% of their populations between 40 and 55S at 1.25mM MgC_2, whereas the TL+H3KR/H4KR array is significantly more extensively folded with 56% of the distribution between 40 and 55S, see figure 3.10C. Similar to the TL+H4KR array, the TL+H3KR/H4KR array was significantly more compacted than the WT array and did not completely unfold into the extended beads-on-a-string conformation (Fig. 3.10 A and C)

This is analogous to what is observed with the folding of the TL+^{T}[2A]^{HF}4 and TL+^{T}[4]$_{CD}$^{HF}4 arrays (compare Fig. 3.6A and 3.9A with 3.110A and C). None of these four mutant arrays were able to unfold into the fully extended 29S conformation in $MgCl_2$, as do all other arrays tested in this chapter.

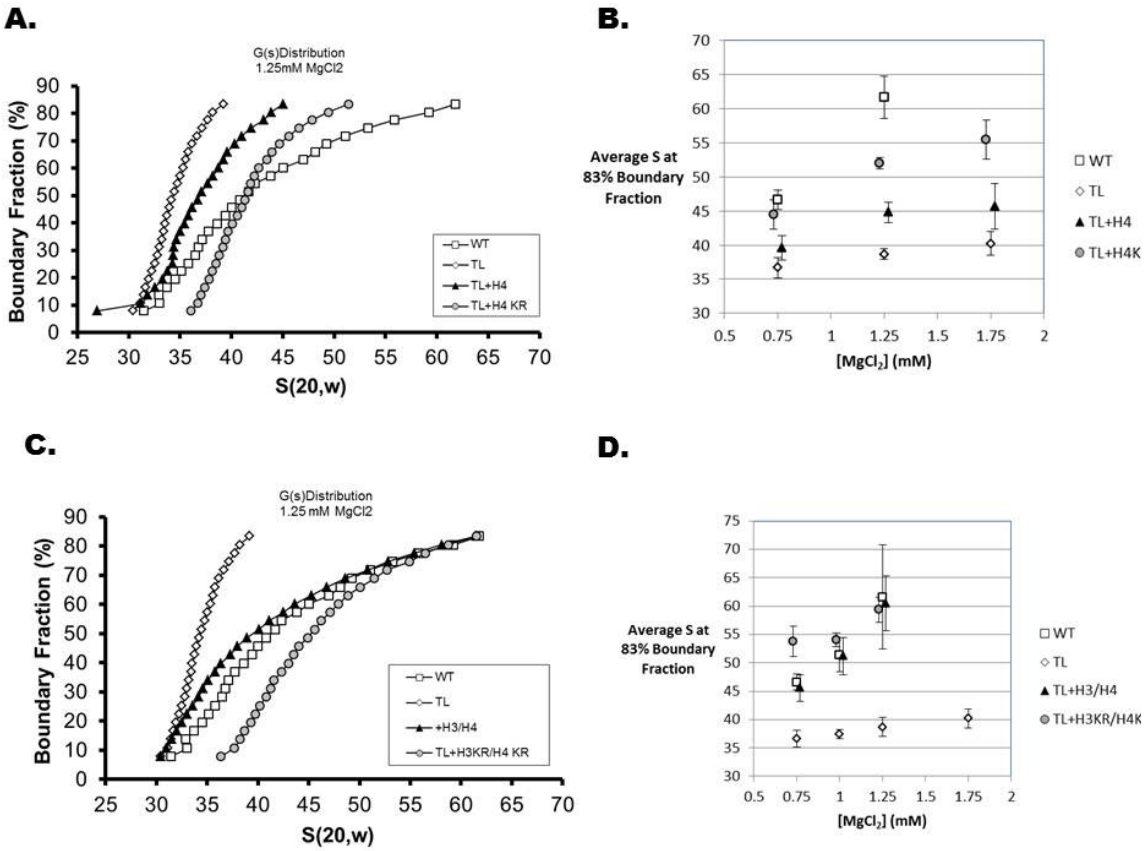

Figure 3.10 Folding of nucleosomal arrays containing H3 and H4 tails with increased arginine content. Fully saturated nucleosomal arrays (TLH4KR, A and B) or TL+H3/H4KR (C and D). Increasing the arginine content of the H4 tail induces similar Mg50s as the WT tail, increasing the arginine content of both the H3 and H4 tails disrupts oligomerization (B and D). A and C, Integral distribution of sedimentation coefficients corrected for diffusion and water at 20ºC ($S_{20,w}$) in 1.25 mM $MgCl_2$. Sedimentation velocity experiments were performed at 21,000-27,000 rpms, room temperature. See chapter two for corresponding G(s) Distributions in 1xTEN B and D, Maximum sedimentation coefficients of nucleosomal arrays. The sedimentation coefficient at 83% boundary fraction was averaged from a minimum of three separate experiments. Error bars represent the S.D. of three or more independent oligomerization assays

3.4 Discussion

3.4.1 Contributions of the core histone tails to nucleosomal array folding

The role of the core histone tails during nucleosomal array folding transitions intramolecularly has previously been studied using the 601_177-12 array (Dorigo *et al*, 2003). Richmond and colleagues reported that the H4 tail was the only tail domain to significantly contribute to array folding (Dorigo *et al*, 2003). However, an important difference between this study and what is presented here (besides the array template being used) is how the data is presented. Average midpoint sedimentation coefficients were used in by Dorigo *et al*, leaving out important details that can be garnered from examining the entire integral distribution (2003). For example in figure 3.4A, all four single tail containing arrays maintain similar midpoint sedimentation coefficients or mean S values (36-37S), however clear differences arise in the top half the distribution between the individual tetramer tails (TL+H4 and H3) and the individual dimer tails (TL+H2A and H2B). These differences are important and significant and indicate that both of the H3 and H4 tails are involved in folding.

There has been compelling evidence in past literature showing that the core histone tail domains were required for maximal nucleosomal array folding (Schwarz and Hansen 1994; Tse and Hansen, 1997; Dorigo *et al*, 2003). I show here that both the WT and TL 601_207-12 nucleosomal array were no different. WT arrays clearly were able to extensively fold, with more than 20% of the boundary sedimenting above 40S in 0.75mM $MgCl_2$ (Fig. 3.1A). Furthermore, figure 3.1A and B clearly indicate a requirement for the histone tail domains to induce extensively folded fibers reaching 55S sedimentation coefficients by 1.0mM $MgCl_2$, where the TL array is only able to reach a maximum of 40S. This is in very good agreement with previously published work (Tse and Hansen, 1997; Carruthers and Hansen, 2000; Dorigo *et al*, 2003).

In 2003, Richmond and colleagues also observed that only with deletion of the H4 tail was maximal array compaction inhibited, whereas deletion of the H3, H2A or H2B tail did not affect folding (Dorigo *et al*, 2003). This was in contrast to what was observed in figure 3.2 where deletion of either H4 or the H3 tail disrupts extensive folding. Folding of 601_207-12 arrays clearly indicate that both the H4 and H3 tails are required for folding at lower $MgCl_2$ concentrations (Fig. 3.2-3.5, Table1). Only at the increased $MgCl_2$ concentration of 1.75mM $MgCl_2$ was the H4 tail playing a more dominant role in array folding than the H3 tail (Fig. 3.5). The nucleosomal array model system used in this dissertation differs only from the 2003 research by an increase in 30 base pairs of DNA for each of the 12 repeats, so why would the H3 tail take on an increased role in array folding with longer linker DNA? The answer may lie within the structural difference between a folded 177-12 nucleosomal array versus the 207-12 array. The shorter separation between nucleosomes on the 177-12 array may limit the number of possible nucleosome-nucleosome conformations and/or interactions within a folded array. Conformations that favor interactions between the H4 tail and its interacting partner(s) likely render the H3 tail out of position to play a major role in the 177-12 array folding. Neighboring nucleosomes separated by longer linker DNA would presumably have more flexibility to sample a higher number of possible conformations. This would potentially allow the H3 tail more opportunities to come into contact with its interacting partner(s). The H4 tail, which has been shown to interact with multiple partners during condensation events, would be unaffected by a change in conformation.

Another interesting observation brings further insight into the individual roles of the histone tails in array folding. It is obvious that both tetramer tails are only able to maximally fold at WT levels of $MgCl_2$ together but not individually (compare Fig. 3.3 and 3.4). Therefore the TL+H3/H4 array meets the minimal “tail” requirement needed to extensively fold (Fig. 3.3; Table 1). $MgCl_2$ was unable to replace the H4 tail (H4TL) to

extensively fold at 1.75mM $MgCl_2$. However, 1.75mM $MgCl_2$ is enough to replace the H3 tail when both the H2A and H2B tails are present (H3TL), suggesting that the H4 tail is still the primary factor in array folding. The H4 tail within the 601_207-12 nucleosomal array plays a reduced role in folding compared to the 601_177-12 nucleosomal array (Fig. 3.5, Table 1). This is in good agreement with previously published data and adds a finer resolution to the individual role of the core histone tail domains (Table 1; Dorigo *et al*, 2003).

The role of the H3 tail in nucleosomal array folding has specific implications for how the chromatin field thinks about the role of histone tails. The predominant school of thought has centered on the H4 tail as the only tail required for nucleosomal array folding (Dorigo *et al*, 2003). The H3 tail has only played a supporting role at best and has generally thought to play important roles in post-translation modifications and epigenetic markers (Fingerman *et al*, 2005; Morillon *et al*, 2005; Berger, 2007; Dia *et al*, 2008). However, previous studies have indicated a role for the H3 tail in oligomerization (Gordon *et al*, 2005; Kan *et al*, 2007). It is interesting to note that the increased role of the H3 tail in nucleosomal array folding becomes more relevant in chapter 4, where the H3 centromeric variant CENP-A containing arrays fold into extensively folded fibers more readily than the WT H3 arrays, potentially due in part to the tail domain (Panchenko *et al*, 2011).

Table 1. Condensation properties of wild type and mutant nucleosomal arrays.

Nucleosomal Array[a]	S_{ave} [b]	Folding [c]	Mg50 [d]	Conclusion
		S (+/- SD)	mM	
WT	29.5S	61.6 (9.2)	1.89	• WT nucleosomal arrays are able to oligomerize and extensively fold. • TL nucleosomal arrays are only able to moderately fold.
TL	25S	38.7 (1.7)	NA	
Single Tail Deletion Arrays				
H4TL	28.0S	40.4 (1.8)	4.03	• Deletion of only the H4 tail leads to the largest loss in oligomerization (first reported by Gordon *et al*, 2005). • Deletion of either the H3 or H4 tails leads to largest deficient in folding (Dorigo *et al*, reported that only loss of H4 tail decreased array folding, 2003). • Deletion of H2A or H2B tails does not affect folding.
H3TL	28.0S	43.9 (2.9)	2.90	
H2ATL	27.5S	68.9 (2.5)	2.37	
H2BTL	28.2S	62.8 (5.0)	2.96	
Double Tail Deletion Arrays				
TL+H3/H4	27.5S	60.5 (4.9)	3.15	• The H3 and H4 tails together induce WT-levels of folding and induce oligomerization at lower [Mg^{2+}] than the H2A and H2B tails together (first reported: Dorigo *et al*, 2003 and Gordon *et al*, 2005).
TL+H2A/H2B	28.3S	46.5 (4.4)	7.2	
Single WT Tail Containing Arrays				
TL+H4	26.3S	44.8 (1.4)	8.47	• The H3 and H4 tails induce oligomerization at similar [Mg^{2+}], the H2B tail induces oligomerization at almost double the [Mg^{2+}] (Gordon *et al*, reported the H4 tail was the largest contributor, followed by the H3 tail, neither the H2A or H2B tails were unable to induce oligomerization individually, 2005). • No histone tail can individually induce WT levels of folding (first reported by Dorigo *et al*, 2003).
TL+H3	26.2S	43.5 (1.9)	7.27	
TL+H2A	26.4S	39.6 (1.0)	NA	
TL+H2B	25.9S	39.1 (0.7)	15.10	
Single "Tail-Swap" Containing Arrays				
TL+$^{T}3^{HF}4$	26.6S	48.1 (3.0)	7.17	• The H3 tail induces compaction similarly whether positioned from the H3 or H4 tail position. • The H2B tail induces compaction at lower [Mg^{2+}] when positioned in the H4 tail position. • The H2A tail is unable to induce oligomerization from either the H4 or the native H2A position, however the H2A tail induces folding to a greater degree from the H4 tail position.
TL+$^{T}2B^{HF}4$	26.7S	42.8 (1.1)	7.77	
TL+$^{T}2A^{HF}4$	26.0S	43.0 (2.0)	NA	
TL+$^{T}4^{HF}2B$	26.5S	39.2 (1.7)	13.33	
Single H4 Tail "Length" Arrays				
TL+$^{T}[4]_{2}^{HF}4$	26.0S	57.8 (2.7)	3.57	• Increasing the length of the H4 tail domain induces a higher degree of maximum folding and oligomerization (McBryant *et al*, 2009 reported similar results with oligomerization only).
TL+$^{T}[4]_{3}^{HF}4$	25.0S	NA	1.84	
Single H4 Tail "Charge-Density" Arrays				
TL+$^{T}[4]_{CD}^{HF}4$	26.1S	42.6 (3.9)	NA	• Positive charge density is inversely proportional to the Mg50 (McBryant *et al*, 2009). • Positive charge density or scrambling the H4 tail sequence does not change the maximum extent of folding.
TL+$^{T}2A_{CD}^{HF}4$	26.7S	41.0 (2.8)	10.44	
TL+$^{T}[4]_{R1}^{HF}4$	26.9S	42.8 (3.2)	13.1	
K—R Mutant Arrays				
TL+H3/H4 K—R	27.9S	59.4 (2.2)	5.63	• Increasing the arginine content of the H3 tail decreases oligomerization. • Increased arginine content induces a higher degree of moderate and extensive folding.
TL+H4 K—R	26.9S	52.0 (0.8)	7.72	

[a] WT, wild-type; TL, tailless; T, tail; HF, histone fold ; CD, charge density; R1, scrambled H4 tail

[b] The s_{ave} was obtained at 50% boundary fraction from the G(s) plots shown in chapter 2.

[c] The maximum sedimentation coefficient in the presence of 1.25 mM $MgCl_2$ obtained from the average of three or more experiments at 83% boundary fraction.

[d] $MgCl_2$ concentration at which 50% of sample is oligomerized, see chapter 2.

3.4.2 Multiple molecular determinants of the H4 tail domain drive nucleosomal array folding

Although it is clear that the H4 tail plays the leading role in nucleosomal array folding, it is unknown whether the mode of action of the H4 tail is the same for folding and oligomerization. Our lab has shown, and the data presented in chapter 2 further supports our published data that the molecular determinants of the H4 tail that induce array oligomerization are amino acid composition (including positive charge density), intrinsic disorder, domain length and the specific spatial position of the H4 tail within each nucleosome (McBryant *et al*, 2009). However, oligomerization and folding are separate condensation events and the role of the histone tails in oligomerization transitions is likely to bind to DNA and neutralize the negatively charged linker DNA (see chapter 2 discussion). More importantly, past folding experiments point to the primary sequence as the only molecular determinant driving folding, in particular the basic stretch of amino acids 14-25 (Dorigo *et al*, 2003). This stretch of amino acids makes crystal contacts with the H2A/H2B acidic patch in neighboring nucleosomes (Luger *et al*, 1997a). Two different crosslinking studies captured the H4 tail-H2A H2B acidic patch interaction in nucleosomal array folding events (Schalch *et al*, 2005; Kan *et al*, 2009). Finally, acetylation of H4K16 leads to a similar decrease in compaction as deleting the entire tail, indicating H4K16 participates in a specific interaction important to array folding (Shogren-Knaak *et al*, 2006). There is no lack of evidence that indicates that the primary sequence of the H4 tail is a molecular determinant in array folding events and therefore many of the mutants described in chapter 2 for oligomerization should predictably decrease an arrays ability to extensively fold due to lack of a conserved sequence.

The data presented here are contrary to past folding studies and indicate that the H4 tail's involvement in array folding is more complicated than a specific binding

interaction(s) between the H4 tail and its neighboring nucleosomes. It is clear that the H4 tail's primary sequence is not a necessity for nucleosomal array folding as replacing the H4 tail with any of the other three tails does not significantly impact the maximum extent of compaction, as measured at 83% boundary fraction (Fig. 3.6A, Table 1). Furthermore, scrambling the 27 amino acids of the H4 tail does not change the overall folding profile or maximum extent of folding compared to the wild type H4 sequence (Fig. 3.7a and b, Table 1). However, individually only the H3 tail and the randomized H4 tail are able to maintain analogous folding profiles as the H4 tail when extending from the H4 histone fold (figure 3.6a and 3.7a). Whereas the TL+^{T}2B^{HF}4 array requires more $MgCl_2$ to reach TL+H4 levels of folding and the shorter TL+^{T}2A^{HF}4 array forms a very homogenous and right shifted population of moderately folded fibers compared to the TL+H4 array. Of note, by 1.75mM $MgCl_2$ the TL+^{T}2B^{HF}4 array behaves analogous to that of the wild type TL+H4 array. Regardless, each of the three tail-swap mutant arrays and the scrambled tail domain have little primary sequence homology with the wild type H4 tail domain, yet each mutant is able to reach similar maximum extents of compaction as the TL+H4 array, indicating that the primary sequence of the H4 tail is not a contributing factor to nucleosomal array folding.

A second observation can be garnered from the "tail-swap" mutant arrays. First, the H3 tail behaves similarly when located at its native position within the nucleosome or from the H4 tail's position, compare figure 3.4 and 3.6. However the H2A and H2B tails do not overly contribute to folding when located in their native dimer positions (Figs. 1.2 and 3.4). Yet when either dimer tail is relocated to the H4 tail position they actively contribute to folding, reaching similar maximum states of compaction compared to TL+H4, indicating that the position of the H4 tail within the nucleosome is an important molecular determinant of the H4 tail's contribution to folding. This observation is further supported by the tail-swap mutant where the H4 tail is relocated to the H2B tail's

position. In this case, the TL+^{T}4^{HF}2B array folds almost identically to a completely TL array (Fig. 3.6C and D). This relocation effectively renders the "potent" H4 tail unable to participate in nucleosomal array folding. Interestingly, most folding models of the 30 nm chromatin fiber depict the linker DNA at the center of the folded chromatin (Schalch *et al*, 2005). The H2A and H2B tails would then be positioned on the outer edges of the folded fiber and unable to interact with the linker DNA (Fig. 1.2). The TL array is able to moderately fold and the addition of either or both of the dimer tails only minimally increases the state of compaction (Figs. 3.3, Table 1). This increase in compaction would most like come from the H2A and/or the H2B tails interacting with inter-nucleosomal DNA within the array. However both the H3 and the H4 tail are well positioned within the nucleosome to interact with DNA or protein within the center of the folded fiber. In hind sight it should not be unexpected that the H3 and H4 tails are able to independently compact arrays to similar degree, or that the H3 tail can replace the H4 tail in array folding.

Although oligomerization and folding are separate and independent transitions, it appears that the mode of action of the H4 tail in both transitions is similar. Length and charge density have also been shown to be important molecular determinants to nucleosomal array oligomerization, figure 2.9 and 10. Importantly, increasing the length or positive charge density increases the arrays ability to oligomerize, the reverse is also true. As with amino acid composition/intrinsic disorder and tail position, the length and positive charge density are also important determinants of the H4 tail in array folding. Doubling the length of the H4 tail (TL+^{T}[4]$_2$^{HF}4) generates arrays that are able to substantially compact more so than the wild type length array (TL+H4) and rival that of the WT array (with a full complement of tails) at each $MgCl_2$ concentration tested (Fig. 3.8, Table 1). However, increasing the length also disrupts the arrays ability to moderately fold, as judged by the decrease or left-shifted sedimentation coefficient in the

lower portion of the distribution compared to the TL+H4 array (Fig.3.9A). Observations from comparing arrays with tails of increasing lengths, TL+^{T}2A^{HF}4, TL+H4 and TL+^{T}[4]$_2$^{HF}4 (12, 27 and 54 amino acid residues in length respectively) show that at each $MgCl_2$ concentration tested, incremental increases in H4 tail domain length decreases the arrays ability to moderately fold (Fig. 4.1, Table 1).

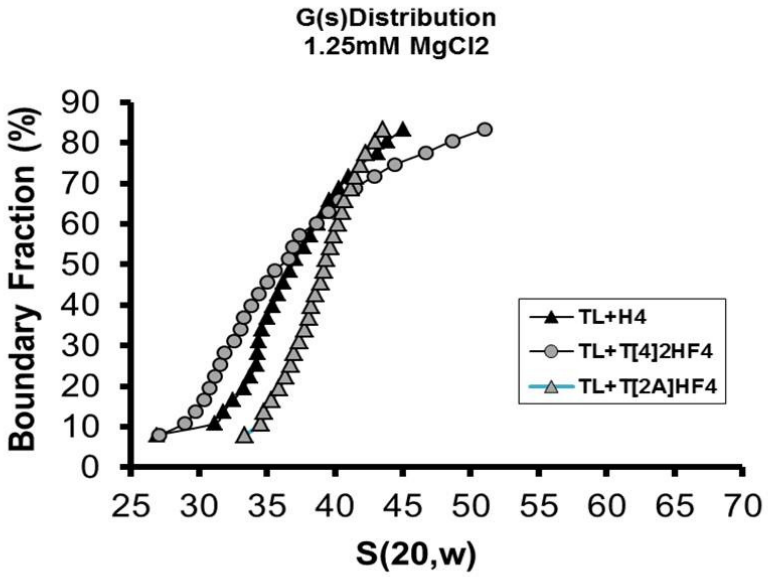

Figure 4.1 Influence of H4 tail length on moderately folded nucleosomal arrays. Increasing the length of the H4 tail disrupts moderate folding (as judged by the left-shifted distribution of TL+$^{T}[4]_{2}{}^{HF}4$ below 70% of the boundary). Decreasing the length facilitates moderate folding (as judged by the right-shifted distribution of TL+$^{T}[2A]^{HF}4$ below 70% of the boundary).

Similar to H4 tail length, the positive charge density of the H4 tail appears to adversely impact moderately folded nucleosomal arrays. Whereas the shorter tail domain containing array TL+^{T}2A^{HF}4 is able to generate a more homogenous population of moderately folded structures compared to TL+H4 arrays, increasing the number of positively charged amino acids from a ratio of 1:3 to 2:3 (TL+^{T}2A$_{CD}^{HF}$4) shifts the distribution of moderately folded arrays to the left and generates a more "wild type" distribution (Fig. 3.10). Conversely, when the charge distribution of the wild type H4 tail is decreased from 1:3 to 1:5 (TL+^{T}[4]$_{CD}^{HF}$4) the distribution is right shifted and gives a similarly homogenous distribution as the shorter TL+^{T}2A^{HF}4 array. However, both of the charge density mutants are able to reach similar maximum sedimentation coefficients as the wild type H4 tail (Fig. 3.9B).

The H4 tail has been shown to have multiple binding partners during nucleosomal array condensation events (Dorigo *et al*, 2004; Kan *et al*, 2007). It is attractive to hypothesize that changing the length and/or the charge density changes the nature or proportion of H4 tail interactions between its multiple partners. For instance, it is apparent that decreasing the charge density or the length of the H4 tail generates a highly homogenous population of arrays below 40S. Therefore, arrays containing a shorter tail or less positive charge may contain a higher number of H4 tail interactions that favor the 40S conformation. Hypothetically, the ^{T}2A^{HF}4 histone contains a tail domain that may be too short to participate in normal "wild type" interactions and instead is in a position to interact with a different target (possibly the DNA entry exit site of a neighboring nucleosome) that easily facilitates moderate folding. Increasing the positive charge of the shorter H4 tail ^{T}[2A]$_{CD}^{HF}$4 brings the distribution back in line with the wild type length/charge density H4 tail, compensating for the reduced length. Secondly, decreasing the positive charge of the full length H4 histone ^{T}[4]$_{CD}^{HF}$4 decreases the positive charge density from 33% to 19% as well as increasing the percentage of

hydrophobic residues in the tail domain from 48% to 56% (Fig. 2.2). An increase in hydrophobicity can potentially generate new interaction sites with hydrophobic patches on the nucleosome surface of neighboring nucleosomes forming a stable interaction between nucleosomes within the nucleosomal array.

3.4.3 Increased arginine content of the H3 and H4 tail facilitates a greater degree of nucleosomal array folding

To this point I have shown that the H3 and H4 tail are the most important contributors to nucleosomal array folding and that the mode of action the H4 tail uses to direct folding is through its position within the nucleosome, its amino acid composition, intrinsic disorder, length and positive charge density. To further exploit the histone tail's high positive charge density I changed the nature of the positive charge predominantly from lysine to arginine. Using H3 and H4 mutant histones in which four lysines where mutated to arginines, changed the nature of the positively charged tail domain, not the charge distribution or density. What I found was that these conserved changes have dramatic consequences for the folded nucleosomal array.

As with the TL+$^{T}[4]_{CD}{}^{HF}4$ and TL+$^{T}2A^{HF}4$ array, increasing the arginine content increases the extent of moderately folded arrays, however both the TL+H4KR and TL+H3/H4KR arrays are able to form extensively folded fibers to a higher degree than that of their wild type counterparts (Fig. 3.10). How can we reconcile the observation that decreasing the arginine content of one tail domain (TL+$^{T}[4]_{CD}{}^{HF}4$) and increasing the arginine content of another tail domain (TL+H4KR) both facilitate a highly homogenous population of moderately folded arrays? The answer I think lies in the potency of the H4KR mutant. In $MgCl_2$, the TL+$^{T}[4]_{CD}{}^{HF}4$ mutant array potentially forms an entropically favorable moderately folded conformation due to the increased hydrophobicity of the tail domain. Whereas increasing the arginine content of the H4 tail domain may lead to a more stable interaction between a subset of H4KR tail domains and DNA that does not

allow the array to fully unfold as judged by the ~35S data point at the bottom of the distribution, versus 29S of the TL+H4 array. What the KR mutants are effectively doing is shifting the equilibrium to the right. It is not surprising therefore that the TL+H3KR/H4KR array shifts the equilibrium even further to the right, as the H3 tail has been shown to interact with intra-array DNA in the presence of $MgCl_2$ (Kan *et al*, 2007).

Chapter 4

Replacement of Histone H3 with CENP-A Directs Global Nucleosomal Array Condensation and Loosening of Nucleosome Superhelical Termini

Tanya Panchenko, **Troy C. Sorensen,** Christopher L. Woodcock, Zhong-yuan Kan, Stacey Wood, Michael G. Resch, Karolin Luger, S. Walter Englander, Jeffrey C. Hansen and Ben E. Black. 2011. Replacement of Histone H3 with CENP-A Directs Global Nucleosomal Array Condensation and Loosening of Nucleosome Superhelical Termini. *Proceedings of the National Academy of Sciences*. **108**:16588-16593.

Chapter 4 was published in 2011 as a collaboration primarily between the Hansen lab and Ben Black lab at the University of Pennsylvania, Perelman School of Medicine. My contribution as a co-first author was to assemble and fold all the nucleosomal arrays tested. The article in its full length is printed below. I am responsible for the data and experiments presented in figures 4.1A-C, 4.2A,B, 4.3E and 4.4C. Appendix 1 contains additional data from mutants that I analyzed that were not in the final published version. Our major findings include that wild type H3 and CENP-A containing nucleosomal arrays assemble in the same stepwise manner during salt dialysis, that CENP-A containing nucleosomal arrays compact to a higher degree than the wild type containing nucleosomal arrays and that H3 R49 is a key component to the folding difference observed between the two arrays.

4.1 Introduction

The centromere is the control locus that directs the faithful inheritance of eukaryotic chromosomes at cell division (Cleveland *et al*, 2003; Allshire et al, 2008). The histone H3 variant, CENP-A, is a highly conserved constituent of all eukaryotic centromeres and is the most attractive candidate for carrying the epigenetic information that specifies the location of the centromere (Panchenko and Black, 2009). Recent findings have led to several fundamentally different proposals for how CENP-A marks centromere location: *i*. CENP-A confers structural and dynamic changes to octameric nucleosomes (Black *et al*, 2007; Sekulic *et al*, 2010), *ii*. CENP-A confers alterations of nucleosomal histone stoichiometry (Williams *et al*, 2009 ; Dalal *et al*, 2007), including the incorporation of non-histone proteins into nucleosome-like structures (Mizuguchi *et al*, 2007), *iii*. CENP-A directs the reversal of handedness of DNA wrapping from left to right (Furuyama and Henikoff, 2009). The nature and composition of CENP-A-containing nucleosomes remain controversial and areas of intense investigation. The available data regarding their structure, however, could be reconciled by distinct CENP-A-containing complexes existing over the course of a cell cycle-coupled maturation program of newly expressed CENP-A protein that propagates the epigenetic centromere mark (Black and Cleveland, 2011).

In the bulk chromatin fiber, nucleosome-nucleosome interactions are central to packaging eukaryotic DNA into the nucleus, to compacting chromosomes during mitosis, and to organizing functional sub-chromosomal domains (Szerlong and Hansen, 2011). Although much is known about how chromatin fibers condense *in vitro*, the extent to which the structured helical histone core of the nucleosome is physically impacted by contact with neighboring nucleosomes in a folded chromatin fiber is not yet known. Eukaryotic centromeres are composed of lengthy arrays of CENP-A-containing nucleosomes. An exception is the budding yeast point centromere which harbors a

single CENP-A-containing nucleosome (Meluh *et al*, 1998; Furuyama and Biggins, 2007). Despite the central role of the specialized CENP-A-containing nucleosomal array in specifying centromere location and directing chromosome inheritance, the internucleosomal interactions of nucleosomal arrays in which CENP-A replaces canonical histone H3 are completely unexplored.

To address the subunit structure and folding of CENP-A-containing nucleosomal arrays, we couple folding measurements using analytical ultracentrifugation (AUC) with mass spectrometry-based hydrogen/deuterium exchange (H/DX-MS). The AUC studies measure the bulk behavior of the arrays, and we find that CENP-A-containing arrays are somewhat more condensed upon folding than canonical arrays. H/DX-MS is an approach capable of measuring the dynamic behavior of the polypeptide backbone of each histone in the nucleosome core. Prior H/DX-MS experiments with sub-nucleosomal $(CENP\text{-}A/H4)_2$ heterotetramers (Black *et al*, 2004) and CENP-A-containing mononucleosomes (Black *et al*, 2007) found that the CENP-A/H4 interface is substantially rigidified by side-chain interactions that restrict transient unfolding of the contacting α-helices (Sekulic *et al*, 2010). In the present study, we demonstrate that the αN-helices of canonical H3-containing nucleosomes are substantially restricted (50 to 100-fold) at their superhelical termini upon nucleosome array folding, indicating an unexpected consequence of nucleosome-nucleosome interactions during chromatin folding. Importantly, both the initial rigidity of CENP-A at its own αN-helix and the rigidity imposed upon chromatin folding are reduced compared to its conventional counterpart containing H3, indicating looseness at the nucleosome superhelical termini.

4.2 Material and Methods

4.2.1 H/DX Reactions.

Nucleosome arrays (assembled and characterized by EM, AUC and MNase digestion as detailed in SI Methods) were incubated for 30 min at 4ºC or 23ºC with either

1xTEN or 1xTEN with 1.25 mM $MgCl_2$ prior to starting deuterium on-exchange reactions. Deuterium on-exchange was carried out by adding 5 µl of the array (containing ~3.8 µg of array) to 15 µl of deuterium on-exchange buffer (10 mM Tris, pD 7.0, 0.25 mM EDTA, 2.5 mM NaCl in D_2O; supplemented with 1.25 mM $MgCl_2$ where indicated) so that the final D_2O content was 75%. Reactions were quenched at the indicated time points by withdrawing 20 µl of the reaction volume, mixing in 30 µl ice cold quench buffer (2.5 M GdHCl, 0.8% formic acid, 10% glycerol), and rapidly freezing in liquid nitrogen prior to proteolysis and LC-MS steps (detailed in SI Methods).

4.2.2 H/DX Data Analysis.

MATLAB based MS data analysis tool, ExMS, was used for data processing (Zhang and Smith, 1993). Detailed information regarding the ExMS algorithm is described elsewhere (Zhang and Smith, 1993) and briefly in SI Methods. The level of H/DX occurring at each time point is expressed as either the number of deuterons or the percentage of exchange within each peptide. In each case, corrections for loss of deuterium label by individual peptides during H/DX-MS analysis (back exchange) were made through measurement of loss of deuterium from reference samples (fully deuterated control, FD) that had been deuterated under denaturing conditions as described elsewhere (Tsunaka *et al*, 2005). The loss of deuterium for the FD sample was ~ 10% for most peptides, and the measures centroid values for several of the αN helix-containing peptides are shown as part of Fig. S3. Calculation of deuterium loss correction and other data operations were performed using MATLAB.

4.3 Results

4.3.1 Stepwise Assembly of CENP-A-containing Polynucleosome Arrays.

Conventional nucleosomes and nucleosomal arrays can be assembled by adding the four core histones at the appropriate molar ratios to defined DNA templates consisting of nucleosome positioning sequences in order to saturate all of the nucleosome binding

sites. This is followed by salt dialysis assembly from 2 M to 2.5 mM NaCl (Simpson *et al*, 1985; Hansen *et al*, 1989 and Hansen *et al*, 1991). The stepwise assembly is characterized by the initial binding of the $(H3/H4)_2$ heterotetramer to DNA to form a tetrasome as the [NaCl] is lowered to 1 M, followed by binding of H2A/H2B heterodimers to the tetrasome to complete nucleosome formation as the [NaCl] is lowered to ≤0.6 M (Hansen *et al*, 1991). Whether CENP-A nucleosomes assemble by such a stepwise mechanism has not been tested.

To determine if CENP-A-containing nucleosomal arrays assemble via the same sequential pathway as canonical arrays, samples that contained $(CENP\text{-}A/H4)_2$ tetramers, (H2A/H2B) dimers, and a linear DNA template containing twelve tandem copies of the so-called 601 sequence (Lowary and Widom, 1998 and Dorigo *et al*, 2003) were mixed in 2 M NaCl and then successively dialyzed against buffer containing 1 M NaCl, 0.6 M NaCl and 2.5 mM NaCl. At each dialysis step, the reactions were assayed by AUC (Figs. 4.1A-C). AUC has been successfully employed in classic sedimentation velocity experiments to measure physical properties of nucleosomal arrays including array composition as well as global changes that accompany intra-array nucleosome-nucleosome interactions and inter-array oligomerization (Hansen, 2002). The CENP-A containing samples were compared with otherwise identical samples containing canonical $(H3/H4)_2$ in place of $(CENP\text{-}A/H4)_2$, as well as samples that contained only $(H3/H4)_2$ (Fig. 4.1A,B). As expected, the canonical H3-containing nucleosomal arrays assembled as previously reported, with the $(H3/H4)_2$ tetramers bound to each DNA repeat at 1 M NaCl to form a 16-20 S 12-mer tetrasomal array (we measured the free 12-mer DNA template sedimenting at ~13 S). This is followed by dimer binding at 0.6 M NaCl to generate the completely assembled 28-30 S beads-on-a string species also present in low salt (i.e. 2.5 mM NaCl) (Figs. 4.1B,C). Similarly, for the centromeric counterpart complexes, the $(CENP\text{-}A/H4)_2$ tetramer binds to DNA in 1 M NaCl and forms

a 17-20 S tetrasomal array (Figs. 4.1B). In reactions that contain (CENP-A/H4)$_2$ tetramers as well as H2A/H2B dimers the H2A/H2B binding is prevented by the presence of 1 M NaCl and we still observe a 17-20 S array (Fig. 4.11 B and C). Upon dialysis into 0.6 M NaCl, CENP-A-containing nucleosomal array assembly is completed as H2A/H2B dimers bind to the tetrasomes to form a stable 28-30 S nucleosomal array (Figs. 4.1B and C). We confirmed by electron microscopy that the 28-30 S CENP-A-containing arrays are saturated with 12 nucleosomes per input linear DNA fragment (Figs. 4.1D). These results indicate that CENP-A arrays form octameric nucleosomes via the same set of intermediate steps as do conventional nucleosomal arrays.

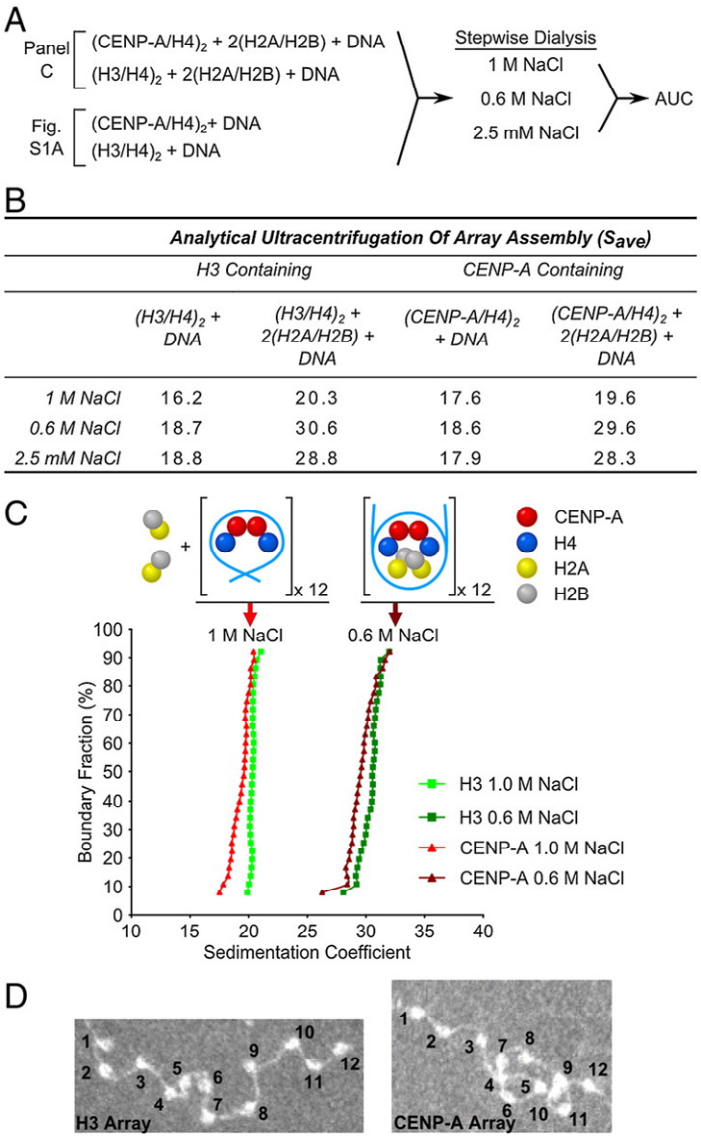

Figure 4.1. CENP-A Nucleosome Array Assembly. (A) Experimental scheme for defining the assembly pathway of CENP-A containing nucleosome arrays. (B) Average sedimentation coefficients (S_{ave} as defined at 0.5 boundary fraction) that were measured by AUC of all the different assembly species atvarying NaCl concentrations as outlined in (A). (CENP-A/H4)2 assembles onto the DNA as a tetramer forming a species sedimenting at ~19 S, with complete assembly of a ~29 S complex corresponding to a12-mer octameric nucleosome array that occurs at NaCl concentrations at or below 0.6 M. For corresponding AUC profiles of all 12 conditions shown in (B). (C) Representative AUC profiles of CENP-A or H3 containing nucleosome arrays analyzed at 1 M and 0.6 M NaCl. All sedimentation coefficients have been corrected for temperature and normalized for water. (D) Electron microscopy (EM) images of CENP-A and H3 containing nucleosome arrays showing that both form similar “beads-on-a-string” 12-mer arrays.

4.3.2 Protection from H/DX Upon Nucleosome Array Folding.

Despite no predicted gross change in their static structures (Sekulic *et al*, 2010 and Luger *et al*, 1997a), the histone fold domains of $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$ undergo >1000-fold slowing in H/DX rates upon incorporation into nucleosomes (Black *et al*, 2007). Further protection from H/DX upon nucleosome array folding, however, was previously untested. Nucleosome-nucleosome interactions are restricted in nucleosomal arrays in the absence of cations, most notably Mg^{2+} (Schwarz and Hansen, 1994). Addition of 1-2 mM Mg^{2+} leads to nucleosomal array folding mediated by internucleosomal contacts that condense the structure so significantly that the 29 S array now sediments at 40-55 S ((Schwarz and Hansen, 1994) and Fig. 4.2A and B). Importantly, such folding behavior is observed in 1.25 mM Mg^{2+}, both with our conventional arrays assembled with canonical recombinant human histones, and with the centromeric counterparts containing CENP-A in place of H3 (Fig. 4.2A and B). This indicates that CENP-A does not function by locally decondensing nucleosomal array structure. Indeed, CENP-A-containing nucleosome arrays show a small, but highly reproducible shift towards a state that is condensed relative to canonical arrays (Fig. 4.2A and B).

H/DX measures how polypeptide backbone amide protons are exchanged with deuterons from heavy water in solution. Folded regions (e.g. α-helices and β-sheets) only exchange upon transient unfolding events when amide protons lose main chain hydrogen bonding. Slow exchange can be achieved by many stabilizing interactions (Englander, 2006), including, in the case of DNA-binding proteins, assembly into higher-order complexes with DNA (Black *et al*, 2007; Kalodimos *et al*, 2004 Hansen *et al*, 2011). In order to test the extent to which the polypeptide backbone dynamics of the structured histone core of the nucleosomal subunits are altered by array folding, we monitored H/DX exchange behavior of folded and unfolded array fibers (time points at

10, 100, 1000, and 10,000 s at 23ºC) (Fig. 4.2C). To potentially detect changes on the most rapidly exchanging regions we also included an additional 10 s time point at 4ºC since such a reduction in temperature leads to a nearly 10-fold slowing in chemical exchange rates at the amide protons that we can measure (Englander, 2006). Throughout this time course, both H3- and CENP-A-containing arrays remain intact and extensively folded as determined by AUC (Fig. 4.2A and B). We monitored the H/DX behavior of overlapping peptides spanning the majority of the folded core of the nucleosome with ~90% coverage of the histone fold domains and ~60% coverage of the total polypeptide length of all histones for both H3- and CENP-A-containing arrays. There is a striking lack of folding-dependent protection at most locations throughout either type of nucleosome core. However, at peptides spanning much of the αN helix of either H3 or CENP-A, we observed additional protection from H/DX (Figs. 4.2D and E). In the canonical nucleosome structure (Luger *et al*, 1997a), the αN helix contacts the DNA at the superhelical termini, i.e. the DNA entry/exit site, of the nucleosome.

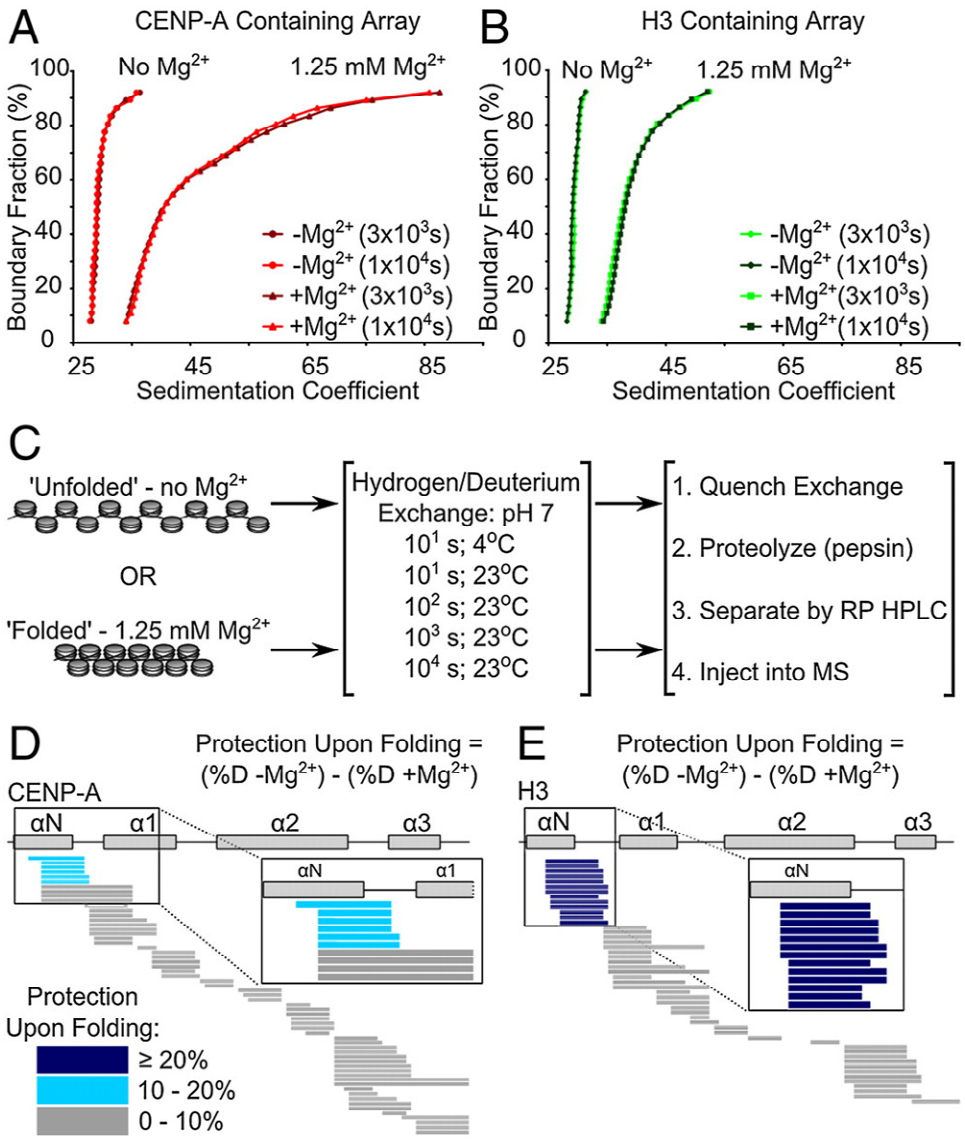

Figure 4.2. Protection from H/DX Upon Nucleosome Array Folding. (A) and (B) AUC profiles showing $MgCl_2$ dependent array folding and array stability over the course of the H/DX experiment. (s_{ave} is defined at 0.5 boundary fraction). (C) Experimental scheme for determining differences in protection from H/DX upon nucleosome array folding. Protection profiles of CENP-A containing nucleosome arrays at 100 s are represented in (D) and protection profiles of H3 containing nucleosome arrays at 100 s are represented in (E). Protection upon array folding is calculated for each peptide in each array type as the difference of percent deuterium content of the peptide from array in the 'unfolded' state minus the percent deuterium content of the same peptide in the 'folded' state. Peptides are represented by horizontal bars and color coded based on the difference in protection upon folding.

4.3.3 The αN Helix of CENP-A is Less Protected than that of H3 Upon Folding of Nucleosomal Arrays.

Although peptides spanning the αN helix of CENP-A are additionally protected upon array folding, the magnitude of this protection is less pronounced than for the same region in H3 arrays (Figs. 4.2D and E and 4.3A and C). The smaller magnitude of protection in this region of CENP-A, relative to H3, was observed in several time points from three independent experiments. H/DX data from several peptides that cover the same region, spanning a portion of the αN helix and following loop in each complex, can be used to compare the local effect of nucleosomal array folding. Representative peptides are shown in Figs. 4.3B and C. While nucleosomal array folding slows the H/DX of the αN helix of H3 by 50-100 times compared to unfolded arrays (Fig. 4.3B, compare $-Mg^{2+}$ to $+Mg^{2+}$), array folding only leads to a 5-10-fold slowing of H/DX in the corresponding region of CENP-A (Fig. 4.3C). We also noted that in addition to not being as restricted by nucleosomal folding at its αN helix, this region of CENP-A exchanges ~10 times faster than the corresponding region in H3 prior to polynucleosome folding (i.e. its beads-on-a string; compare $-Mg^{2+}$ in Fig. 4.3B and 4.3C). Indeed, the αN helix of CENP-A is nearly as flexible in folded chromatin as is the αN helix of H3 in completely unfolded arrays (compare $-Mg^{2+}$ in Fig. 4.3B with $+Mg^{2+}$ in Fig. 4.3C).

In the canonical H3-containing nucleosome, the αN helix lies at the DNA entry/exit site (Fig. 4.3D; (Luger *et al*, 1997a)). CENP-A is ~80% identical to H3 in this region, and reconstituted CENP-A-containing nucleosomes wrap ~150 bp of DNA in a left-handed manner (Sekulic *et al*, 2010). The finding that CENP-A nucleosomal arrays maintain local flexibility after chromatin folding builds on the earlier observation that topologically constrained DNA minicircles containing a single CENP-A nucleosome prefer a more 'open' conformation where the exiting DNA strands do not cross (Conde *et al*, 2007). In our reconstituted arrays on a repetitive and strongly positioning DNA sequence,

digestion with MNase clearly protects less DNA (Fig. 3E). CENP-A-containing arrays have a clear pause site at ~150 bp (Fig. 4.3E, 0.2 U MNase), consistent with a heavily populated steady-state species that is a fully wrapped octameric nucleosome. Further, CENP-A-containing nucleosomes dynamically and transiently release their DNA superhelical termini to allow the digestion of an additional one or two turns of DNA (i.e. 10-20 bp) at either terminus to yield a fragment of ~125 bp upon extensive digestion (Fig. 4.3E, 2 U MNase).

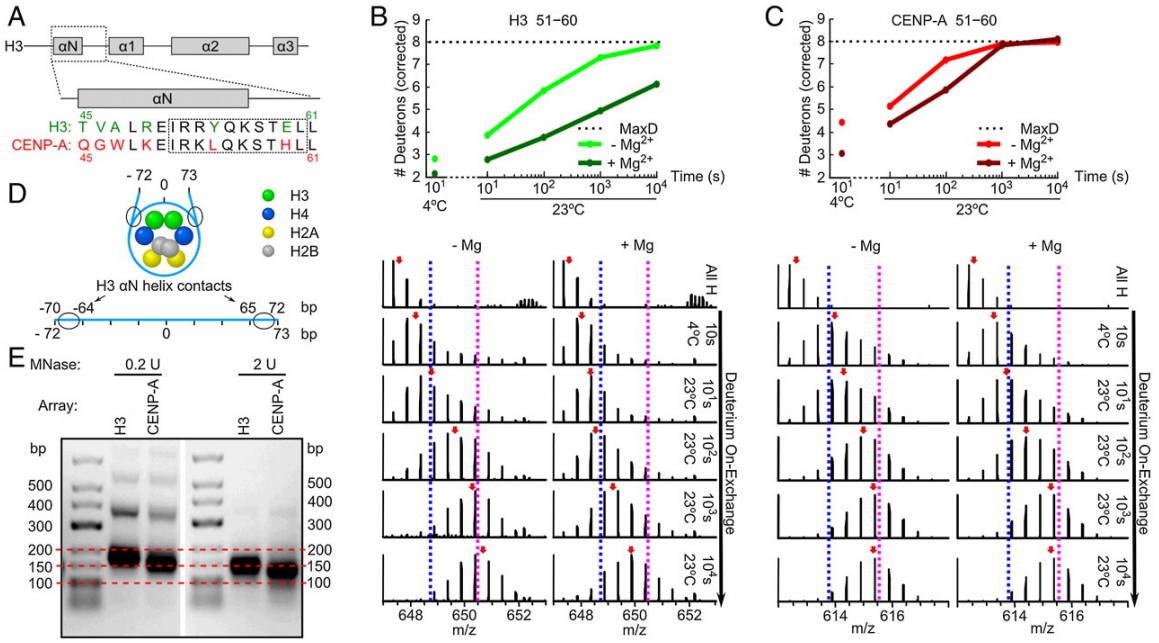

Figure 4.3. Increased Flexibility at the αN helix of CENP-A in Nucleosomal Arrays. (A) Alignment of H3 and CENP-A sequences in the αN and adjacent loop region. Boxed region highlights the sequence of the representative peptides in panels (B) and (C). (B) H/DX of a representative αN-H3 peptide from nucleosome arrays containing H3. Top panel shows normalized deuterium levels at each time point for arrays that are either "unfolded" (- Mg^{2+}) or "folded" (+ Mg^{2+}) and bottom panels show raw peptide data where the dotted blue and pink lines are drawn as guides to visualize differences in H/DX and red arrows indicate peptide centroid values. (C) H/DX of a representative αN-CENP-A peptide from nucleosome arrays containing CENP-A. Top and bottom panels are displayed as in (B). (D) Schematic representation of the locations where H3 αN helix contacts nucleosomal DNA. (E) MNase digestion of CENP-A or H3 containing nucleosome arrays.

4.3.4 Arg49 of H3 Contributes to Rigidity at the Superhelical Nucleosome Termini in Folded Arrays.

One likely contribution to the increased flexibility in folded CENP-A-containing nucleosomal arrays is the substitution of a lysine at the position corresponding to Arg49 in histone H3. Arg49 of H3 intercalates into the DNA one half turn from the superhelical terminus of the nucleosomal DNA (Fig. 4.4A; (Luger *et al*, 1997a)). Indeed, with mononucleosomes assembled onto topologically constrained minicircles, the R49K mutation leads to a preference for an open DNA entry/exit arrangement (as opposed to with entering/exiting strands crossed in a closed arrangement) to nearly the same extent as seen with CENPA mononucleosomes (Conde *et al*, 2007). The R49K mutation generates increased local flexibility, measured in 6 out of 6 peptides that span at least some portion of the αN helix, including adjacent peptides whose amino acid composition is identical to WT H3 (Figs. 4.4B). The increased flexibility in H3 R49K nucleosomes is more pronounced in folded than unfolded arrays. While this substitution does not account for the full extent of the flexibility observed in CENP-A-containing nucleosomal arrays (Fig. 4.3), the lysine in place of the DNA intercalating arginine is clearly a major contributor. Thus, the R49K mutation in H3 generates an intermediate level of flexibility that also corresponds with an intermediate level of sensitivity to digestion of superhelical DNA termini by MNase (Fig. 4.4C).

Beyond the αN helix, we also observed a region of H2A (a.a. 91-134) with 10 out of 11 peptides showing faster exchange rates in CENP-A containing arrays (~10-fold faster than in H3-containing nucleosomes; representative peptides, data not shown). This observation was missed in earlier efforts (Black *et al*, 2007), but was readily detected in the present study where many more nucleosome-derived peptides could be monitored

due to major technical improvements in peptide resolution. Unlike the αN helices of H3 and CENP-A, the difference in H2A at this location is independent of nucleosomal folding.

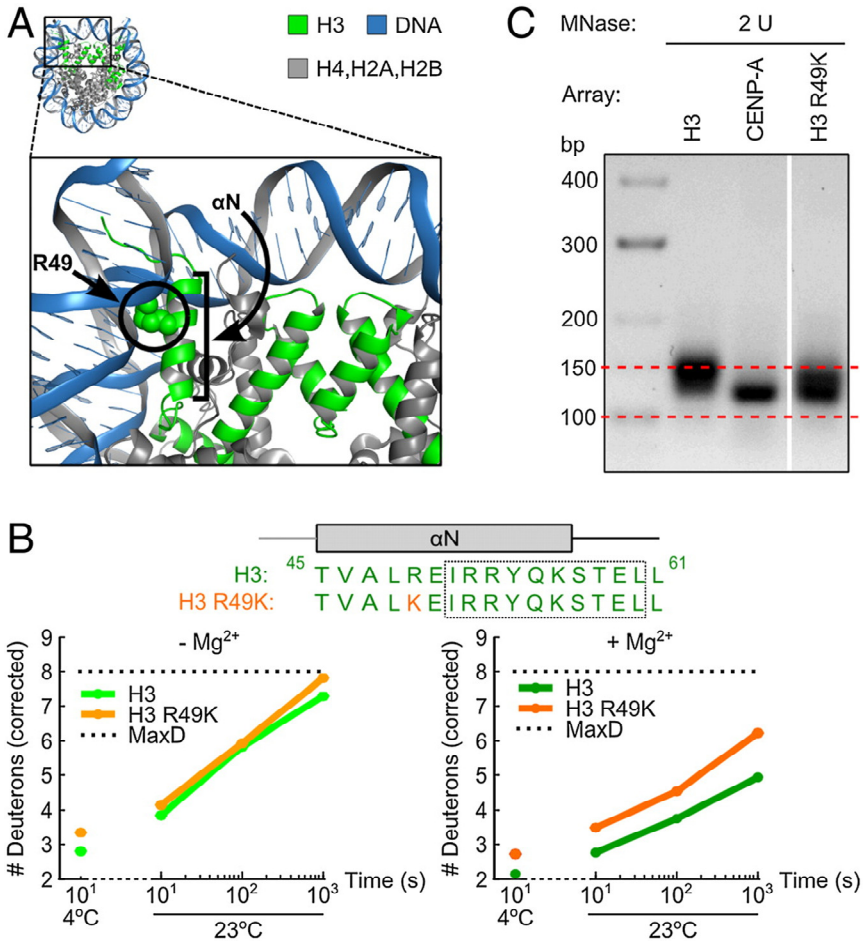

Figure 4.4 Arg49 of H3 contributes to a more rigidified folded fiber. (A) Crystal structure of the human nucleosome particle containing histone H3 (PDB 2CV5 (Tsunaka *et al*, 2005)) is shown to highlight the location of Arg49 in relation to the αN helix and the DNA entry/exit site. (B) Representative peptides from αN region of H3 containing arrays and mutant H3 R49K arrays. (C) MNase digestion of the H3 R49K nucleosome containing arrays.

4.4 Discussion

4.4.1 CENP-A in the Nucleosomal Array Context.

The measurement of >100 partially overlapping peptide 'probes' that span the majority of each histone subunit in canonical and CENP-A-containing nucleosomal arrays provide a high-resolution view of these structures. Previous high-resolution/site specific studies of the dynamic and structural impact of CENP-A focused on the sub-nucleosomal tetramer that it forms with histone H4 (Sekulic *et al*, 2010; Black *et al*, 2004; Cho and Harrison, 2011), mononucleosomes (Black *et al*, 2007 and Conde *et al*, 2007), and the ternary complex that CENP-A and H4 forms with the centromeric chromatin assembly protein HJURP (or the HJURP counterpart, Scm3, in yeast) (Cho and Harrison, 2011; Hu *et al*, 2011; Zhou *et al*, 2011). The CENP-A targeting domain (CATD), comprised of the L1 and α2-helix of CENP-A, contributes hydrophobic stitches that rigidify the interface between CENP-A and H4 (Sekulic *et al*, 2010 and Black *et al*, 2004). This rigidity is maintained after assembly into mononucleosomes (Black *et al*, 2007). L1, within the CATD, generates a surface on the face of the CENP-A nucleosome that is divergent in shape and electrostatic charge (Sekulic *et al*, 2010) from the corresponding surface on canonical H3-containing nucleosomes (Luger *et al*, 1997a). These features structurally and dynamically distinguish CENP-A mononucleosomes from conventional ones. The essential role of the CATD in centromere function (Black *et al*, 2007) strongly suggests that these features are key to distinguishing centromeric chromatin from the rest of the chromosome.

Beyond individual mononucleosomes, each centromere is made up of many CENP-A mononucleosome subunits. CENP-A-containing nucleosomes coalesce on the face of the chromosome to define the location of the mitotic kinetochore assembly, but they are not arranged in a linear context. Instead they are interspersed with conventional H3-containing nucleosomes (Blower *et al*, 2002). Whether or not there is a regular geometry

to centromeric chromatin organization (Blower *et al*, 2002; Zinkowski *et al*, 1991; Santaguida and Musacchio, 2009; Ribeiro *et al*, 2010), it is very likely that internucleosomal contacts between CENP-A-containing nucleosomes are fundamental in organizing centromeric chromatin. Using H/DX-MS, we have found that a major difference between the subunit structures of folded CENP-A- and H3-containing nucleosomal arrays is increased flexibility at the αN helix of CENP-A which contacts superhelical termini of each nucleosome.

After this work was complete, a crystal structure of an octameric CENP-A-containing nucleosome was reported (Tachiwana *et al*, 2011). In the crystal structure, the terminal 13 bp of DNA are not visible (Tachiwana *et al*, 2011), consistent with our observations of additional flexibility in the beads-on-a-string configuration (Fig. 4.3B and C; -Mg^{2+}). The H/DX studies of nucleosome array folding show the difference at this site between CENP-A- and H3-containing nucleosomes is much greater in folded than unfolded arrays (Fig. 4.3B and C). This finding greatly extends our understanding of CENP-A beyond mononucleosomes (Black *et al*, 2007a; Sekulic *et al*, 2007; Tachiwana *et al,* 2011) since it provides the first view of the dynamics and structure of the CENP-A-containing chromatin fiber.

4.4.2 Dynamics at the Superhelical Termini of the Nucleosomal Subunits of Folded Arrays.

For both $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$ heterotetramers, assembly into nucleosomes causes major H/DX protection along their polypeptide backbones (~1000-fold slower exchange throughout the respective histone fold domains; (Black *et al*, 2007a)). However, the internucleosomal contacts that accompany nucleosomal array folding apparently remain fluid and do not further restrict the backbone dynamics of the bulk of the octameric histone core of the nucleosome (Figs. 4.2D and E). The observation that the αN helix of H3-containing arrays is substantially affected upon folding (50-100-fold

slower H/DX; Figs. 4.3B and C) was not predictable from earlier crystal structures. In both nucleosome (Luger *et al*, 1997a) and tetranucleosome (Schalch *et al*, 2005) static structural models, the local structure of histone and DNA are not changed (Luger *et al*, 1997a; Schalch *et al*, 2005). However, it is interesting that rigidity is required for linker DNA and up to 10 bp of terminal nucleosome core DNA in order to make an idealized chromatin fiber model based on the tetranucleosome crystal structure (Schalch *et al*, 2005). It is also interesting that in the leading models for nucleosome higher-order packing, the nucleosome superhelical termini always face the interior of the fiber (Dorigo *et al*, 2004). Our results suggest that the fiber interior imposes the most rigidity on nucleosome cores.

Our studies also conclusively demonstrate that CENP-A-containing arrays can be readily assembled in a step-wise manner and form condensed fibers in a manner highly similar to canonical nucleosomal arrays (Figs. 4.1A-C and 4.2A and B). Prior to our present study, the data indicating whether or not CENP-A-containing complexes can assemble nucleosomes in a step-wise manner was not clear. Indeed, it has been reported that after assembly of the $(CENP\text{-}A/H4)_2$ heterotetramer a 147 bp DNA template (i.e. the number of base pairs that wraps the canonical nucleosome core particle) resulted in formation of complexes that showed native gel migration consistent with two stacked heterotetramers, not a single one as for canonical $(H3/H4)_2$ heterotetramers assembled onto the same template (Conde *et al*, 2007). However, the exact nature of the CENP-A 'tetrasomes' was not explored further. Nonetheless, the apparent deviation from the canonical tetrasome (i.e. $[H3/H4]_2$ + 147 bp DNA) behavior (Conde *et al*, 2007) was used to strongly question the relevance of subsequently reconstituted octamers to bona fide centromeric chromatin (Henikoff and Furuyama, 2010).

The histone H3 N-terminal tail is known to participate in array folding (Zheng *et al*, 2005). We note that the CENP-A N-terminal tail, which is completely divergent in sequence identity from bulk H3 but maintains the strongly basic charge, does not preclude the nucleosome-nucleosome interactions that lead to formation of extensively folded array structures (Fig. 4.2A and B). Indeed, we always observe that at the same Mg^{2+} concentration, CENP-A nucleosomal arrays are somewhat more condensed than canonical arrays (as judged by the right-shifted G(s) profiles). It is attractive to speculate that the enhanced propensity to fold is related to the loosened superhelical termini of CENP-A nucleosomes, perhaps working in conjunction with the specialized CENP-A N-terminal tail.

In budding yeast, one of a subset of fungal species where centromere location is determined not by epigenetic information but by a specific ~125 bp DNA sequence, the CENP-A counterpart, Cse4p, assembles into an octameric nucleosome that protects only 110-125 bp of DNA with no evidence of further wrapping (Kingston *et al*, 2011; Dechassa *et al*, 2011). We note that at the position corresponding to Arg49 in histone H3 (where substitution to the lysine corresponding to human CENP-A leads to increased flexibility; Figs. 4.4, data not shown), Cse4p has substituted a tyrosine that likely prevents any DNA wrapping by its αN helix. It seems likely that the Arg to Tyr substitution evolved to prevent full 145 bp DNA wrapping and to accommodate the constraints imposed by a system in which the centromere is defined by a particular DNA sequence.

In many eukaryotes, including animals, CENP-A-containing nucleosomes are thought to be arranged in several local clusters of adjacent nucleosomes per centromere interspersed by clusters of H3 containing nucleosomes (Blower *et al*, 2002). Although the exact numbers of nucleosomes that make up each cluster is not known, our 12-mer arrays reconstitute a single cluster. Our AUC experiments show that CENP-A containing

arrays are generally more condensed than canonical arrays upon folding, while our H/DX experiments demonstrate local flexibility at the contact point between the CENP-A αN-helix and the nucleosomal terminal DNA. We suggest that both characteristics (Fig. 4.5) are important for centromeric chromatin. The condensed nature of the array may reflect strong self-self-interactions that culminate in the coalescence of discontinuous CENP-A nucleosomes on each centromere into a discrete focus on the surface of the chromosome (Blower *et al*, 2002). In addition, the condensed nature of the array may be important to help maintain structural integrity of centromeric chromatin under mitotic spindle stretching forces. The increased flexibility at the DNA entry/exit site may accommodate inter-nucleosomal DNA paths unique to the centromere (Blower *et al*, 2002; Santaguida and Musacchio, 2009; Ribeiro *et al*, 2010) and/or nucleosomal (or nucleosome-proximal) binding proteins that require increased local access to the DNA and/or histone core. Alternatively, it is formally possible that the increased flexibility at the nucleosomal termini of the human CENP-A protein is a vestige of a common ancestral orthologue that may have lacked any stable DNA binding at the superhelical termini (i.e. similar to the behavior of Cse4 in S. cerevisiae (Kingston et al, 2011 and Dechassa *et al*, 2011)).

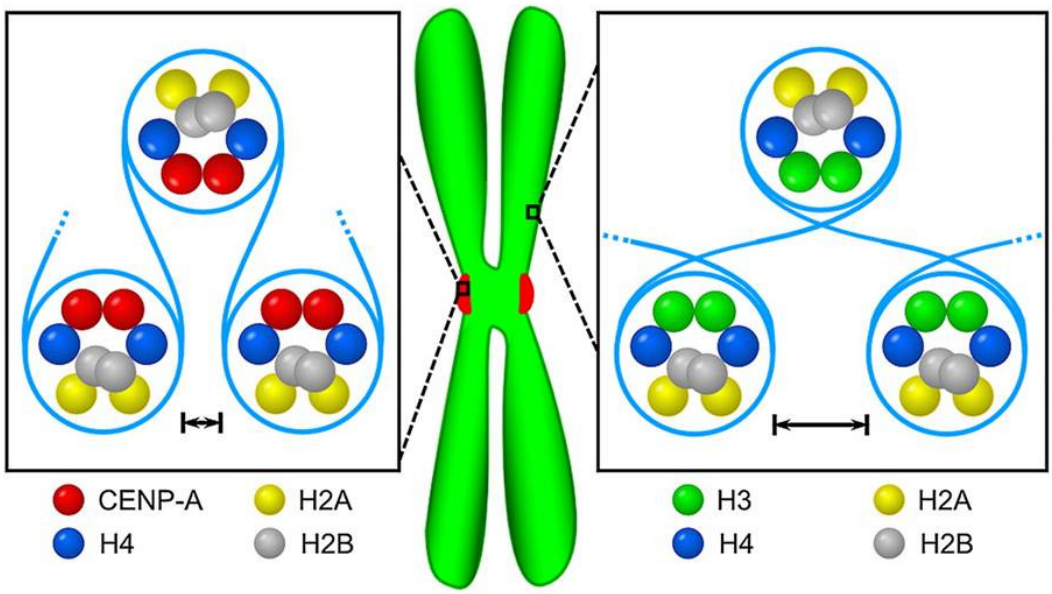

Figure 4.5. Summary of physical differences identified in this study between CENP-A- and H3containing nucleosomal arrays. AUC experiments showed that CENP-A-containing arrays aremgenerally more condensed than canonical arrays upon folding (indicated by closer spacing of adjacentmCENP-A-containing nucleosomes). At the same time the H/DX experiments measured local relative flexibility at the CENP-A αN-helix, indicating that the DNA at the entry/exit sites is less constrained (indicated by uncrossed internucleosomal DNA for CENP-A-containing arrays). See the text for a discussion of the potential implications of these findings.

4.4.3 Rapid Exchange on the C-terminus of Histone H2A in CENP-A-containing Nucleosomes.

The region of H2A (a.a. 91-134) that undergoes rapid exchange in CENP-A-containing nucleosomes is juxtaposed to the αN-helix of H3 and is exposed on the surface of the nucleosome in the conventional nucleosome structure (Luger *et al*, 1997a). Therefore, the increased exchange rates in H2A (a.a. 91-134) is possibly related to the flexibility of the αN-helix of CENP-A. This seems unlikely, however, since there is no correlation with the αN-helix behavior. The H2A H/DX in this region is not substantially slowed in H3 containing arrays upon array folding at the same time points where up to a 100-fold slowing is observed in the αN-helix of H3 (Fig. 4.3C). Another possibility is that the increased H2A H/DX in CENPA-containing nucleosomes reflects a different orientation of H2A that accompanies rotation of the H2A/H2B dimers. (CENP-A/H4)$_2$ heterotetramers prefer a compact state, either in solution or in crystals, which comes about by rotation at the CENP-A/CENP-A four-helix bundle (Sekulic *et al*, 2010). Upon nucleosome formation, the CENP-A/CENP-A interface may rotate to form a nucleosome of similar shape to the canonical nucleosome (Sekulic *et al*, 2010 and Kan *et al*, 2011), or the H2B/H4 four-helix bundles could rotate H2A/H2B dimers away from the central axis of the nucleosome to avoid steric clashes (Sekulic *et al*, 2010). Our H/DX data suggests that the possibility of whether one or both of these two states of CENP-A-containing nucleosomes are substantially populated at centromeres should be the subject of further careful analysis.

Chapter 5

Conclusion

This study provides novel insights into how the core histone N-terminal tail domains function to compact nucleosomal array model systems. Previous studies have indicated that the H4 tail alone is the primary "condensing" agent of the four histone tail domains (Dorigo *et al*, 2003; Gordon *et al*, 2005). However, in the fully saturated 601_207-12 nucleosomal array model system the role of the H3 tail is increased in both intra-array and inter-array condensation events. This is an important finding because little has been done to investigate the role of the H3 tail in nucleosomal array compaction (Kan *et al*, 2007; Dorigo *et al*, 2005). Instead, many studies have focused on the role of the H3 tail in (but not limited to) heterochromatin silencing (Fingerman *et al*, 2005; Dai *et al*, 2008) and transcription initiation (Morillon *et al*, 2005; Berger, 2007). Like the H3 tail, the H2B tail also has an increased role in nucleosomal array oligomerization.

Using a single tail containing nucleosomal array model system allowed me to investigate the mode of action of the H4 tail without the added contributions of the other tails. Perhaps the most significant contribution from this work is the observation that the position of the H4 tail within the nucleosome is a substantial molecular determinant. The H3 tail actually induced a higher level of array compaction when relocated to the H4 tail position then from its native position and the WT H4 tail. The position of the histone tails has previously been indicated as vital in protein-protein interactions in trans (Dai *et al*, 2008), but until recently the role of the H4 tail position has not been implicated in

chromatin condensation events (McBryant et al, 2009). Arrays containing only the H3 tail or only the H3 tailless histone behave similarly to arrays with the H4 counterparts. Due to this, similar studies should be performed to determine if the same the H3 tail has evolved to share the same molecular determinants as the H4 tail in array compaction. In particular, the position of the H3 tail contributes to array compaction.

In contrast to past studies, the evidence presented here indicates that the primary sequence of the H4 tail is not a primary molecular determinant of the H4 tail (Dorigo *et al*, 2003). Instead there are multiple molecular determinants of the H4 tail that induce nucleosomal array compaction. Amino acid composition, positive charge density and intrinsic disorder are all shared attributes of the core histone tail domains, and are also important molecular determinants of the H4 tail. Each of the core histone tail domains was able to reach similar maximum extents of compaction when located at the H4 tail position. I believe that each of the core histone tail domains has evolved to share these attributes because each tail is undoubtedly involved in similar DNA-protein interactions in higher order chromatin structures.

REFERENCES

Allshire, R. and Karpen, G. 2008. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet*. **9**:923-937.

Anderson, J., and Widom, J. 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.* **296**:979-987.

Arents, G., Burlingame, R., Wang, B., Love, W., Moudrianakis, E. 1991. The nucleosomal core histone octamer at 3.1 angstrom resolution: A tripartite protein assembly and a left-handed superhelix. *PNAS*. **88**:10148-10152.

Bednar, J., Horowitz, R., Grigoryev, S., Carruthers, L., Hansen, J., Koster, A., Woodcock, C.1998. Nucleosomes, linker DNA, and linker histones form a unique structural motif that directs the higher-order folding and compaction of chromatin. *PNAS*. **95**:14173-14178.

Berger, S. 2007. The complex language of chromatin regulation during transcription. *Nature*. **447**:407-212.

Black, B., Foltz, D., Chakravarthy, S., Luger, K., Woods, V., and Cleveland, D. 2004. Structural determinants for generating centromeric chromatin. *Nature*. **430**:578-582.

Black, B., Brock, M., Bedard, S., Woods, V., and Cleveland, D. 2007. An epigenetic mark generated by the incorporationof CENP-A into centromeric nucleosomes. *PNAS*. **104**:5008-5013.

Black, B., Jansen, L., Maddox, P., Folts, D., Desai, A., Shah, J., and Cleveland, D. 2007b. Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol. Cell*. **25**:309-322.

Black, B. and Cleveland, D. 2011. Epigenetic centromere propagation and the nature of CENP-A nucleosomes. *Cell*. **144**:471-479.

Blacketer, M., Feely, S., Shogren-Knaak, M. 2010. Nucleosome interactions and stability in an ordered nucleosome array model system. *J. Biol. Chem.* **285**:34597-607.

Blower, M., Sullivan, B. and Karpen, G. 2002. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell*. **2**:319-330.

Butler, P., Thomas, J. 1980. Changes in chromatin folding in solution. *J. Mol. Biol*. **140**:505-529.

Carruthers, L, Bednar, J., Woodcock, C., and Hansen, J. 1998. Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding. *Biochem*. **37**:14776-14787.

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

Cole, J., Lary, J., Moody, T., and Laue, T. 2008. Analytical Ultracentrifugation: Sedimentation velocity and sedimentation equilibrium. *Methods in Cell Biol*. **84**:143-179.

Cho, U. and Harrison, S. 2011. Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. *PNAS*. **108**:9367-9371.

Cleveland, D., Mao, Y., and Sullivan, K. 2003. Centromers and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**:407-421.

Conde e Silva, N., Black, B., Sivolob, A., Filipski, J., Cleveland, D., and Prunell, A. 2007. CENP-A containing nucleosomes: Easier disassembly versus exclusive centromeric localization. *J. Mol. Biol*. **370**:555-573.

Dai, J., Hyland, E., Yuan, D., Huang, H., Bader, J., Boeke, J. 2008. Probing nucleosome function: A highly versatile library of synthetic histone H3 and H4 mutants. *Cell*. **134**:1066-1078.

Dalal, Y., Wang, H., Lindsay, S. and Henikoff, S. 2007. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS Biol*. **5**:e218.

Davey, C., Sargent, D., Luger, K., Maeder, A., and Richmond, T. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. *J. Mol. Biol*. **319**:1097-1113.

Dechassa, M., Wyns, K., Li, M., Hall, M., Wang, M., and Luger, K. 2011 Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat. Commun*. **2**:313.

Demeler, B. and van Holde K. 2004. Sedimentation velocity analysis of highly heterogeneous systems. *Anal. Biochem*. **335**:279-288.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T. 2003. Chromatin Fiber Folding: Requirement for the histone H4 N-Terminal tail. *J. Mol. Biol*. **327**:85-86.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R., Richmond, T. 2004. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*. **306**:1571-1573.

Durrin, L., Mann, R., Kayne, P., and Grunstein, M. 1991. Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. C*ell*. **65**:1023-1031.

Eltsov, M., MacLellan, K., Maeshima., K., Frangakis., A., and Dubochet., J. 2008. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromoses *in situ. PNAS*. **105**:19732-19737.

Englander, S. 2006. Hydrogen exchange and mass spectrometry: A historical perspective. *J. Am. Soc. Mass. Spectrom*. **17**:1481-1489.

Fan, J., Gordon, F., Luger, K., Hansen, J., and Tremethick, D. 2002. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nat. Struct. Mol. Biol.* **9**:172-176.

Finch J, Klug A. 1976. Solenoidal model for superstructure in chromatin. *PNAS*. **73**:1897-1901.

Fingerman, I., Wu, C., Wilson, B., and Briggs, S. 2005. Global loss of Set1-mediated H3 Lys4 trimethylation is associated with silencing defects in S. cerevisiae. *J. Biol. Chem*. **280**:28761-28765.

Fletcher, T. and Hansen, J. 1995. Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *J. Biol. Chem.* **270**:25359-25362.

Furuyama, S. and Biggins, S. 2007. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *PNAS*. **104**:14706-14711.

Furuyama, T. and Henikoff, S. 2009. Centromeric nucleosomes induce positive DNA supercoils. *Cell.* **138**:104-113.

Garcia-Ramirez, M., Dong, F., Ausio, J. 1992. Role of the histone “tails’ in folding and oligonucleosomes depleted of histone H1. *J. Biol. Chem*. **267**:19587-19595.

Gordon, F., Luger, K., and Hansen, J., 2005. The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerizatin of nucleosomal arrays. *J. Biol. Chem.* **280**:33701-33706.

Grigoryev, S., Arya, G., Correll, S., Woodcock, C., and Schlick, T. 2009. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *PNAS*. **32**:13317-22.

Hansen, J., Ausio, J., Stanik, V. and van Holde, K. 1989. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochem*. **28**:9129-9136.

Hansen, J. van Holde, K. and Lohr, D. 1991. The mechanism of nucleosome assembly onto oligomers of the sea urchin 5 S DNA positioning sequence. *J. Biol. Chem*. **266**:4276-4282.

Hansen, J., and Lohr, D. 1993. Assembly and structural properties of subsaturated chromatin arrays. *J. Biol. Chem.* **268**:5840-5848.

Hansen, J., Kreider, I., Demeler, B., and Fletcher, T. 1997. Analytical ultracentrifugation and agarose gel electrophoresis as tools for studying chromatin folding in solution. *Methods: A companion to Methods in Enzymology*. **12**:62-72.

Hansen, J. 2002. Conformational Dynamics of the Chromatin Fiber in Solution: Determinants, Mechanisms, and Functions. *Annu. Rev. Biophys. Biomol. Struct*. **31**:361-392.

Hansen, J., Wexler, B., Rogers, D., Hite, K., Panchenko, T., Ajith, S., and Black, B. 2011. DNA binding restricts the intrinsic conformational flexibility of methyl CpG binding protein 2 (MeCP2). *J. Biol. Chem*. **286**:18938-18948.

Hewish, D., and Burgoyne, L. 1973. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun*. **52**:504-510.

Henikoff, S. and Furuyama, T. 2010. Epigenetic inheritance of centromeres. *Cold Spring Harb. Symp. Quant. Biol*. **75**:51-60.

Hizume, K., Nakai, T., Araki, S., Prieto, E., Yoshikawa, K., Takeyasu, K. 2009. 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*. doi:10.1016/j.ultramic.2009.03.014.

Hu, H., et al. 2011.Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev*. **25**:901-906.

Kalodimos, C., Biris, N., Bonvin, A., Levandoski, M., Guennueques, M., Boelens, R., and Kaptein, R. 2004. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*. **305**:386-389.

Kan, P., Lu, X., Hansen, J., and Hayes, J. 2007. The H3 tail domain participates in multiple interactions during folding and self-association of nucleosome arrays. *Mol Cell Biol*. **6**:2084-2091.

Kan, P., Caterino, T., and Hayes, J. 2009. The H4 tail domain participates in intra- and internucleosome interactions with protein and DNA during folding and oligomerization of nucleosome arrays. *Mol Cell Biol*. **2**:538-546.

Kan, Z., Mayne, L., and Englander, S. 2011. ExMS: A data analysis program for HX-MS experiments. *J. Am. Soc. Mass Spectrom*. in press.

Kingston, I., Yung, J., and Singleton, M. 2011. Biophysical characterization of the centromere-specific nucleosome from budding yeast. *J. Biol. Chem*. 286:4021-4026.

Kornberg, R. and Thomas, J. 1974. Chromatin structure: Oligomers of the histones. *Science*. **184**:865-868.

Kornberg, R. 1974. Chromatin structure: A repeating unit of histones and DNA. *Science*. **184**:868-871.

Korolev, N., Allahverdi, A., Yang, Y., Fan, Y., Lyubartsev, A., and Nordenskiold, L. 2010. Electrostatic origin of salt-induced nucleosome array compaction. *Biophysical Journal.* **99**:1896-1905.

Kruithof, M., Chien, F., Routh, A., Logie, C., Rhodes, D., and van Noort, J. 2009. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat. Struct. Mol. Biol.* **16**:534-540.

Lenfant, F., Mann, R., Thomsen, B., Ling. X., and Grunstein, M. 1996. All four core histone N-termini contain sequences required for the repression of basal transcription in yeast. *EMBO*. **15**:3974-3985.

Lowary, P. and Widom, J 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol*. **276**:19-42.

Lu, X., Simon, M., Chodaparambil, J., Hansen, J., Shokat, K. and Luger, K. 2008. The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat. Struct. Mol. Biol.* **15**:1122-1124.

Luger, K., Mader, A., Richmond, R., Sargent, D., and Richmond, T. 1997a. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature*. **389**:251-260.

Luger, K., Rechsteiner, T., Flaus, A., Waye, M., and Richmond, T. 1997b. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**:301-311.

Luger, K. and Richmond, T. 1998. DNA binding within the nucleosome core. *Curr. Op. in Struct. Biol*. **8**:33-40.

Maeshima, K., Hihara, S., and Eltsov, M. 2010. Chromatin structure: does the 30-nm fiber exist in vivo? *Current Opinion in Cell Biology*. **22**:291-297.

McBryant, S., Klonoski, J., Sorensen, T., Norskog, S., Williams, S., Resch, M., Toombs III, J., Hobdey, S., and Hansen, J. 2009. Determinants of histone H4 N-terminal domain function during nucleosomal array oligomerization. *J. Biol. Chem*. **284**:16716-16722.

Meluh, P., Yang, P., Glowczewski, L. Koshland, D. and Smith, M. 1998. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell*. **94**:607-613.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M. and Wu, C. 2007. Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell.* **129**:1153-1164.

Moazed, D. 2001. Enzymatic activities of Sir2 and chromatin silencing. *Curr. Opin. Cell Biol.* **13**:232-238.

Morillon, A., Karabetsou, N., Nair, A., and Mellor, J. 2005. Dynamic lysine methylation on histone H3 deines the regulatory phase of gene transcription. *Mol. Cell*. **18**:723-734.

Olins, D., and Olins, A. 1972. Physical studies of isolated eukaryotic nuclei. *J. Cell Biol*. **53**:715-736.

Panchenko, T. and Black, B. 2009. The epigenetic basis for centromere identity. *Prog. Mol Subcell. Biol*. **48**:1-32.

Panchenko, T., Sorensen, T., Woodcock, C., Kan, Z., Wood, S., Resch, M., Luger, K., Englander, W., Hansen, J., and Black, B. 2011. Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *PNAS.***108**:16588-16593.

Polach, K., Lowary, P. and Widom, J. 2000. Effects of core histone tail domains on the equilibrium constants for dynamic DNA site accessibility in nucleosomes. *J. Mol Biol.* **298**:211-223.

Raisner, R., Hartley, P., Meneghini, M., Bao, M., Liu, C., Schreiber, S., Rando, O., and Madhani, H. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*. 123:233-248.

Ribeiro, S., Vagnarelli, P., Dong, Y., Hori, T., McEwen, B., Fukagawa, T., Flors, C., and Earnshaw, W. 2010. A super-resolution map of the vertebrate kinetochore. *PNAS*. **107**:10484-10489.

Richmond, T., Finch, J., Rushton, B., Rhodes, D., and Klug, A. 1984. Structure of the nucleosome core particle at 7 angstrom resolution. *Nature*. **311**:532-537.

Robinson, P., Fairall, L., Huynh, V., Rhodes, D. 2006. EM measurements define the dimensions of the “30-nm” chromatin fiber: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci.* USA. **103**:6506-6511.

Robinson, P., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R., and Rhodes, D. 2008. 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol Biol.* **381**:816-825.

Routh, A., Sandin, S., and Rhodes, D. 2008 Nucleosome repeat length and linker histone stoichiometry determin chromatin fiber structure. *PNAS*. **105**:8872-8877.

Santaguida, S. and Musacchio, A. 2009. The life and miracles of kinetochores. *EMBO. J.* **28**:2511-2531.

Schalch, T., Duda, S., Sargent, D., and Richmond T. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fiber. *Nature*. **436**:138-141.

Schwarz, P. and Hansen, J. 1994. Formation and stability of higher order chromatin structures. *J. Biol. Chem*. **269**:16284-16289.

Schwarz, P., Felthauser, A., Fletcher, T., and Hansen, J. 1996. Reversible oligonucleosome self-association: Dependence on divalent cations and core histone tail domains. *Biochemistry*. **35**:4009-4015.

Sekulic, N., Basset, E., Rogers, D. and Black, B. 2010. The structure of (CENP-A-H4)(2) reveals physical features that mark centromers. *Nature*. **467**:347-351.

Shogren-Knaak, M., Ishii, H., Sun, J., Pazin, M., Davie, J., Peterson, C., 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*. **311**:844-847.

Simpson, T., Thoma, F., and Brubaker, J. 1985. Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: A model system for study of higher order structure. *Cell*. **42**:799-808.

Sinha, D., and Shogren-Knaak, M. 2010. The role of direct interactions between the histone H4 tail and the H2A core in long-range nucleosome contacts. *J. Biol. Chem*. **22**:16572-81**.**

Szerlong, H. and Hansen, J. 2011. Nucleosome distribution and linker DNA: Connecting nuclear function to dynamic chromatin structure. *Biochem. Cell. Biol.* **89**:24-34.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., Kimura, H., Kurumizaka, H. 2011. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature*. **476**:232-235.

Thoma, F., and Koller, T. 1977. Influence of histone h1 on chromatin structure. *Cell*. **12**:101-107.

Thoma, F., Koller, T., and Klug, A. 1979. Invovlement of histone H1 in the organization of the nucleosome and the salt-dependent superstrucutres of chromatin. *J. Cell Biol*. **83**:403-427.

Tremethick, D. 2007. Higher-order structures of chromatin: The elusive 30 nm fiber. *Cell*. **128**:651-654.

Tse, C. and Hansen, J. 1997. Hybrid Trypsinized Nucleosomal Arrays: Identification of multiple functional roles of the H2A/H2b and H3/H4 N-termini in chromatin fiber compaction. *Biochemistry*. **36**:11381-11388.

Tsunaka, Y., Kajimura, N., Tate, S. and Morikawa, K. 2005. Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res*. **33**:3424-3434.

van Holde, K. 1988. Chromatin. Springer-Verlag, New York.

van Holde, K., and Zlatanova, J. 2007. Chromatin fiber structure: Where is the problem now? *Cell Devel. Biol.* **18**:651-658.

Vasudevan, D., Chua, E., and Davey, C. 2010. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol*. **403**:1-10.

Wang, X., and Hayes, J. 2008. Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. *Mol. Cell. Biol*. **28**:227-236.

Williams, J., Hayashi, T., Yanagida, M. and Russell, P. 2009. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol. Cell.* **33**:287-298.

Woodcock, C., Frado, L., and Rattner, J. 1984. The higher-order structure of chromatin: Evidence for a helical ribbon arrangement. *J. Cell Biol*. **99**:42-52.

Zhang, Z. and Smith, D. 1993. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci*. **2**:522-531.

Zheng, C., Lu, X., Hansen, J. and Hayes, J. 2005. Salt-dependent intra- and internucleosomal interactions of the H3 tail domain in a model oligonucleosomal array. *J. Biol. Chem.* **280**:33552-33557.

Zhou, Z., Feng, H., Zhou, B., Ghirlando, R., Hu, K., Zwolak, A., Miller, L., Xiao, H., Tjandra, N., Wu, C., and Bai, Y. 2011. Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature*. **472**:234-237.

Zinkowski, R., Meyne, J. and Brinkley, B. 1991. The centromere-kinetochore complex: A repeat subunit model. *J. Cell Biol*. **113**: 1091-1110.

APPENDIX A

H3R49K Increases Nucleosomal Array Folding

Appendix A, details the additional experiments I performed that were not published in Panchenko *et al*, 2011. Figure A.I indicates how the H3R49K mutation was partially responsible for the increased flexibility of the nucleosomal DNA superhelical termini. Increased flexibility of the entry/exit DNA also allowed for increased compaction of the H3 R49K nucleosomal array. Integral distributions show that the H3 R49K nucleosomal array in 1.25 mM $MgCl_2$ was more folded than the WT H3 arrays, but unable to compact to the same degree as CENP-A arrays (Fig. A.1). The single arginine to lysine mutation increases the extent of compaction to a state intermediate to that of the WT H3 and CENP-A arrays (Fig. A.1). Likewise, the counter mutant CENP-A K49R, would theoretically decrease the flexibility of the superhelical termini and decrease the extent of compaction of the nucleosomal arrays (Fig. A.1). Nucleosomal arrays with the K49R mutation in CENP-A causes a decrease in compaction compared to the WT CENP-A array (H/DX experiments of CENP-A R49K arrays +/- Mg^{2+} produced no significant changes compared to WT CENP-A arrays (Fig. A.1; data not published, T. Panchecko). H3 R49K does not revert to CENP-A folding levels nor does the CENP-A K49R array revert to the WT H3 arrays folding levels, instead both mutant arrays fold to similar and intermediate states of the two WT arrays (Fig. A.1).

The R49K mutation in CENP-A does not fully account for the increased state of folding in CENP-A arrays compared to the WT H3 arrays. Although not tested in this

study, I believe that the tail domain of the CENP-A histone is contributing to the CENP-A containing arrays increased compaction over the WT H3 containing arrays, based off of two observations. First chapter 3.3.2 through 3.3.4 indicates that the H3 tail contributes to array folding. Secondly, chapter 3.3.9 indicates how increasing the arginine content of the H3 or H4 tail domain induces a more "folded" nucleosomal array compare to the WT counterparts. As noted in chapter 4, the H3 and CENP-A tails share no sequence homology, yet both are rich in lysine and arginine. Importantly, the CENP-A tail domain is highly enriched in arginines compared to the WT H3 tail, and in fact contains five more arginines than the WT H3 tail domain (Fig. 2.1).

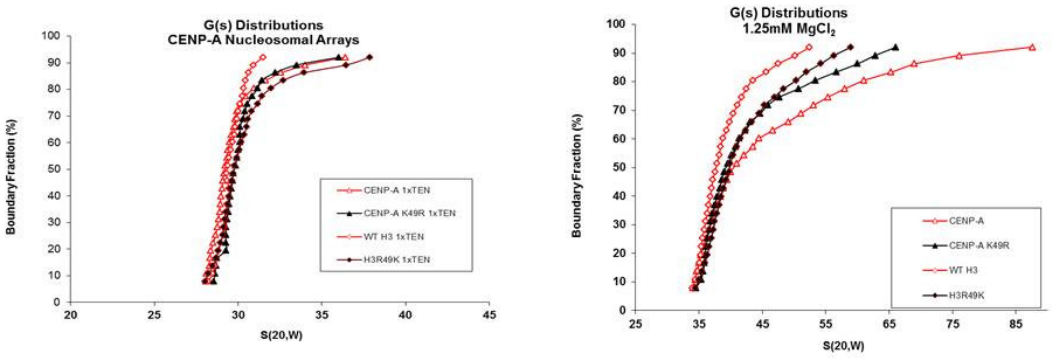

Figure A.1. Arg49 of H3 decreases nucleosomal array folding. Arrays containing H3 R49K or CENP-A K49R fold similarly, but at intermediate levels as their WT counterparts. Integral distribution of sedimentation coefficients of the H3 R49K and CENP-A K49R nucleosomal arrays in 1xTEN buffer (left) and with 1.25mM $MgCl_2$ (right).

APPENDIX B

Genetic Analysis of Core Histone Tail Mutants *In Vivo*

The role of the core histone tails in cell viability has been investigated in *Saccharomyces cerevisiae*. Simultaneous deletions of H3 and H4 tails are lethal, although cells are still viable after deletion of only the H3 or H4 tails (Kayne *et al*, 1988; Morgan *et al*, 1991; Ling *et al*, 1996). As was observed *in vitro* for nucleosomal array condensation (Chapter 2 and 3; McBryant *et al*, 2009), both the H3 and H4 tails can functionally replace each other (Morgan *et al*, 1991; Ling *et al*, 1996). A similar scenario has been established for the H2A and H2B tails (Wallis *et al*, 1983; Schuster *et al*, 1986). A more recent study involving 79 different deletions of either H3 or H4 tail domain residues has shown that the majority of these deletions are not lethal (Dai *et al*, 2008). Generally, large deletions of the H3 tail domain or deletions removing 4-5 lysines within the H4 tail domain are lethal (Dai *et al*, 2008). These studies document the feasibility of studying core histone tail mutants *in vivo* using *S. cerevisiae* as a model system. I was interested in investigating the role of the H4 tail in chromatin hierarchical structures *in vivo* using similar H4 mutants as those studied in chapter 2 and 3.

To begin to understand the role of the H4 tail in genomic chromatin organization, I cloned *S. cerevisiae* histone counterparts for many of the same histone tail mutants detailed in chapters 2 and 3. The specific budding yeast strains and plasmids used are detailed in Table B.1. H4 tail domain mutants were cloned into the YCp22 plasmid that contained the H3 and H4 ORF and maintained tryptophan selectivity. Newly cloned

mutant plasmids were sequenced before being transformed into the ROY1281 yeast strain using electroporation techniques (Fig. B.1). Newly transformed yeast strains were maintained on CAA-Trp medium to maintain the YCp22 plasmid.

Table B.1. Yeast strains and plasmids

Strain and Plasmid	Genotype
ROY1281	MATα lys2 trp1 his3 leu2 ura3 hht1-hht1::LEU2 hht2-hhf2::HIS3
pRO149	2 micron, URA3, HHT1-HHF1, HTA1-HTB1
YCp22	CEN4, ARS1, TRP1, HHT1-HHF1

FIGURE B.1. Yeast H4 Tail Mutant Sequences. A, sequence of the wild type yeast H4 tail followed by H4 tail deletion mutants. d1-20 H4 is the analogous H4TL mutant used in chapters 2 and 3. Underlined Isoleucine 21 is the sole difference in the H4 tail domain primary sequence (Xenopus H4 used in chapter 2 and 3 contains Valine 21). B, Sequence of H4 tail swap mutants. C, Sequences of H4 tail double length mutants. The proximal repeat contains K—R mutations, the Distal repeat is highlighted in red. Specific K—R mutations are underlined.

Initial plasmid shuffling of the transformed strains yielded extremely low shuffling efficiencies, even when shuffling in a wild type H4 plasmid (Fig. B.2). To document the lethality of the H4 tail mutants and low plasmid shuffling efficiency, each indicated yeast strain was grown to an OD_{600} of 1.0. Serial dilutions starting with an OD of 1.0 down to 0.0001 were performed on CAA-Trp plates containing FOA. Seven micro-liters of each dilution was spotted onto the FOA, CAA-Trp plates and stored in the dark at 30ºC (Fig. B.2).

Of the H4 mutant strains, only $^{T}[4]_{2}{}^{HF}4$, $^{T}2A^{HF}4$ and d14-27H4 strains are viable, only the $^{T}[4]_{2}{}^{HF}4$ is as healthy as the WT H4 strains in the assays test thus for. The H4 length mutant, $^{T}[4]_{2}{}^{HF}4$ (two head to tail repeats of the H4 tail) transformed, shuffled and grows comparable to wild type H4 strains (data not shown). However, the K(5,8,12,16,20)R counterpart, $^{T}[4]_{2}Pr_{(K—R)}{}^{HF}4$ (repeat proximal to the histone fold contains all 5 lysines mutated to arginines) has a dominant-negative phenotype in the meroploid strain and is lethal when the wild type H4 plasmid is shuffled out of the cell (Fig. B.2 A and C). Two suppressor mutants were generated during shuffling of this strain. Both suppressor mutant strains grow similar to wild type strains under non-stressful conditions, maintain the original $^{T}[4]_{2}Pr_{(K—R)}{}^{HF}4$ YCp22 plasmid and still have both the His and Leu markers intact that knocked out the endogenous H3-H4 ORFs (data not shown). No further analysis of the two suppressor mutants has been performed.

Even in the presence of the WT H4 that was expressed from a 2 micron plasmid the $^{T}[4]_{2}Pr_{(K—R)}{}^{HF}4$ mutant histone induced a very sick yeast strain. Because, H4 K16 has been shown to be an important sight for lysine acetylation, I cloned a third length mutant similar to original $^{T}[4]_{2}Pr_{(K—R)}{}^{HF}4$, except arginine 16 in the proximal repeat is reverted back to lysine 16 ($^{T}[4]_{2}Pr_{(K—R)K16}{}^{HF}4$). Importantly, this one amino acid change (R16K) alleviates some of temperature sensitive phenotypes seen in the original

$^{T}[4]_2Pr_{(K—R)}{}^{HF}4$ meroploid strain (Fig. B.2C). However, the $^{T}[4]_2Pr_{(K—R)K16}{}^{HF}4$ mutant is still lethal in the absence of the WT H4 histone, and the meroploid strain remains very slow growth phenotypes (Fig. B.2A and C). Western blots using anti-H4 and anti-acetylated H4 antibodies indicate that in the meroploid strains, each of the length mutants are expressed and acetylated (Fig. B.2). Importantly, $^{T}[4]_2Pr_{(K—R)}{}^{HF}4$ and $^{T}[4]_2Pr_{(K—R)K16}{}^{HF}4$ are acetylated in the distal repeat furthest from the nucleosome (Fig. B.2). In light of both the length and K→R mutant nucleosomal arrays discussed in chapter 3, it is possibly that the combination of a longer H4 tail and the lysine to arginine mutations is generating a highly compact genome, limiting DNA accessibility, and disrupting nuclear functions.

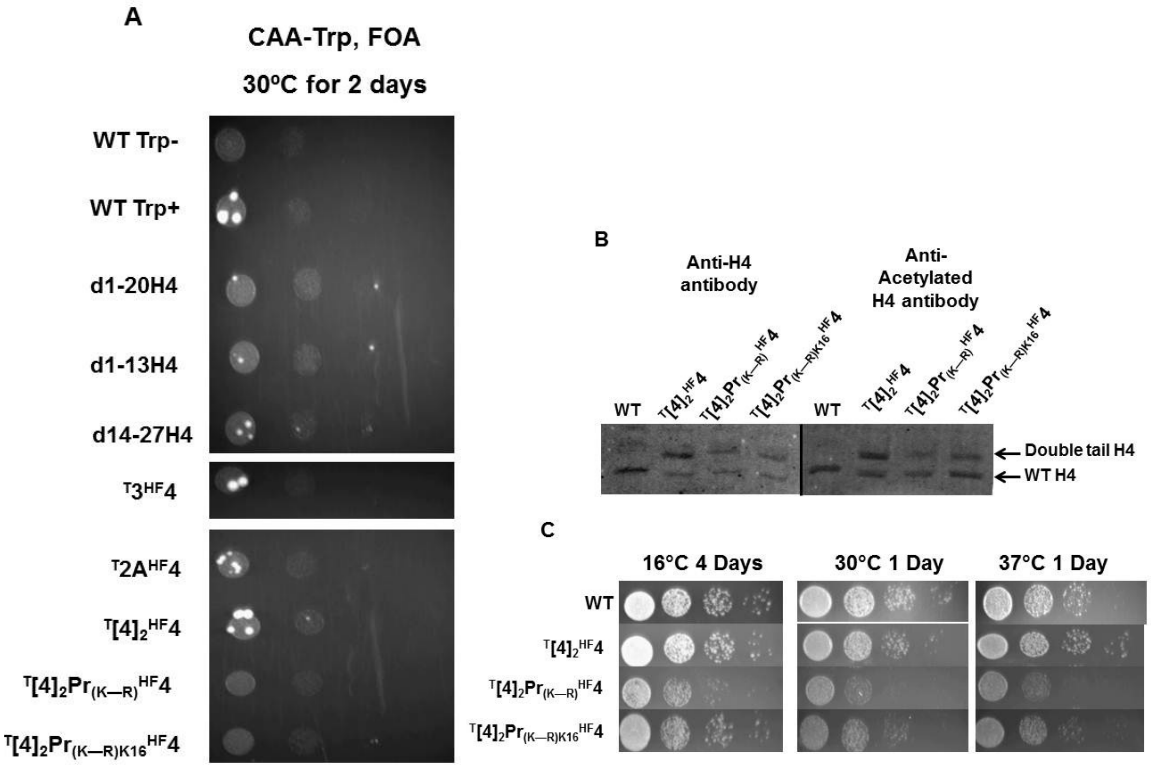

Figure B.2. Analysis of double H4 tail mutant meroploid strains A, Serial dilution, plasmid shuffle of meroploid strains on CAA –Trp media containing FOA. Only the WT, d14-27H4, $^{T}2A^{HF}4$ and $^{T}[4]_{2}{}^{HF}4$ mutants are viable (data not published). B, Western blot of meroploid strains containing double tail length mutants probed with anti-H4 and anti-acetylated H4 antibodies. C, Serial dilution spot tests of the three indicated H4 tail length mutant meroploid strains at 16℃ 4 days, 30 and 37℃ 1 day on CAA-Trp media.

To test whether the $^{T}[4]_{2}^{HF}4$ mutant strain affects chromatin structure *in vivo*, I performed MNase digestion on purified yeast nuclei from both a wild type strain and the viable length mutant strain, $^{T}[4]_{2}^{HF}4$. The MNase protocol was performed essentially the same as previously reported (Eriksson *et al*, 2005). Figure B.3, shows that at least for this initial experiment, nucleosomes are positioned at apparently the same repeat length (~160bp) in both the wild and the $^{T}[4]_{2}^{HF}4$ strain. Repeat experiments need to be performed with better phenol-chloroform extractions to help diminish the amount of blurring in each of the lanes (Fig. B.3).

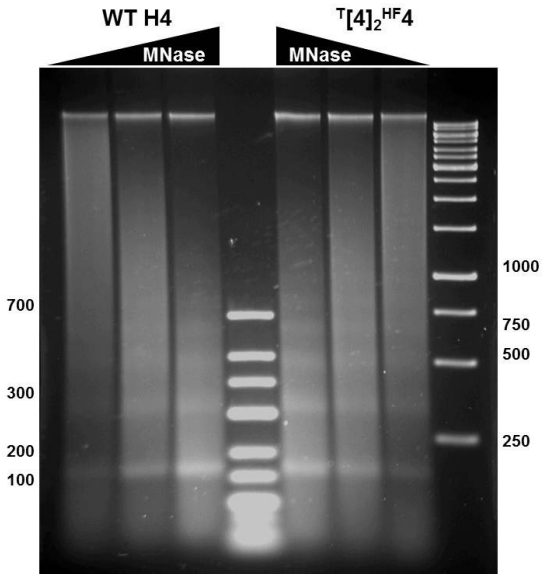

Figure B.3. Chromatin accessibility is not affected by the longer H4 tail domain. MNase digestion of purified nuclei. Shown is 2% agarose gel with increasing amounts of MNase, each lane was digested for two minutes.

REFERENCES

Dai, J., Hyland, E., Yuan, D., Huang, H., Bader, J., Boeke, J. 2008. Probing nucleosome function: A highly versatile library of synthetic histone H3 and H4 mutants. *Cell*. **134**:1066-1078.

Eriksson, P., Mendiratta, G., McLaughlin, N., Wolfsberg, T., Marino-Ramirez, L., Pompa, T., Jainerin, M., Landsman, D., Shen, C., and Clark. D. 2005. Global regulation by the yeast Spt10 protein is mediated through chromatin structure and the histone upstream activating sequence elements. *Mol. Cell. Biol.* **25**:9127-9137.

Kayne, P., Kim, U., Han, M., Mullen, J., Yoshizaki, F., and Grunstein, M. 1988. Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell*. **55**:27-39.

Ling, X., Harkness, T., Schultz, M., Fisher-Adams, G., and Grunstein, M. 1996. Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. *Genes & Development*, **10**(6):686-99.

McBryant, S., Klonoski, J., Sorensen, T., Norskog, S., Williams, S., Resch, M., Toombs III, J., Hobdey, S., and Hansen, J. 2009. Determinants of histone H4 N-terminal domain function during nucleosomal array oligomerization. *J. Biol. Chem*. **284**:16716-16722.

Morgan, B., Mittman, B., and Smith, M. 1991. The highly conserved N-terminal domains of histones H3 and H4 are required for normal cell cycle progression. *Mol. Cell Biol*. **11**:4111-4120.

Panchenko, T., Sorensen, T., Woodcock, C., Kan, Z., Wood, S., Resch, M., Luger, K., Englander, W., Hansen, J., and Black, B. 2011. Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *PNAS.***108**:16588-16593.

Schuster, T., Han, M., and Grunstein, M. 1986. Yeast histone H2A and H2B amino termini have interchangeable functions. *Cell*. **45**:445-451.

Wallis, J., Rykowski, M., and Grunstein, M. 1983. Yeast histone H2B containing large amino terminus deletions can function in vivo. *Cell*. **35**:711-719.