Investigation of parameters of chromatin higher order structure and the making of multiply-acetylated histone H3 via nonsense suppression

by

Isaac A Young

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:
Michael Shogren-Knaak, Major Professor
Richard Honzatko
Scott Nelson
Eric Underbakke
Yanhai Yin

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2021

DEDICATION

In loving memory, Susan Jo Young. March 22, 2021

TABLE OF CONTENTS

ABSTRACT

Chromatin, is the primary compactor of, as well as a regulator of all functions related to the eukaryotic genome. Chromatin generally is observed to exist in three states in vitro: the 30nm fiber, the 10 nm extended fiber, and the self-associated array particle (oligomers). Historically, the tightly packed 30nm fibers were the most important system in vivo with oligomers having less physiological relevance but in fact strong evidence was never found for the 30nm fiber's existence in vivo. Technology now exists which is helping to prove that the 30nm fiber is niche behavior for chromatin and in fact in vitro oligomers may help to probe chromatin's role in regulating genomic functions. Utilizing dynamic light scattering (DLS), differential magnesium induced sedimentation assays, and inductively coupled plasma mass spectrometry, we probed oligomer's growth kinetics, size, magnesium binding stoichiometry, and threshold magnesium concentration requirement for oligomerization ($[Mg^{2+}]_{50}$) and looked for changes in their behavior in response to several factors: pH, nucleosomal concentration, magnesium concentration, and array length.

First, we found that DLS was a valuable technique for rapidly and accurately detecting size of arrays, intra-array compacted arrays, and array oligomers as well as the growth kinetics of actively growing oligomers. Second, we found that out of all the factors we tested that: 1. Only pH had a strong correlation with $[Mg^{2+}]_{50}$, 2. Only array concentration had a strong correlation with the growth kinetics of actively oligomerizing particles. 4. Only magnesium concentration had a correlation with magnesium binding stoichiometry and that array may bind an excessive amount of magnesium relative to their charge. Finally, using DLS we have captured preliminary evidence of an intermediately sized array oligomer which is in good agreement with the size of topologically associated domains in vivo.

A common post-translational modification (PTM) of proteins is lysine acetylation. This is an especially ubiquitous PTM in the histones of chromatin, and is important for helping to regulate both structural and mechanistic aspects of chromatin. A number of strategies exist for generating acetylated nucleosomes for the in-vitro study of chromatin though they all have various advantages and disadvantages. An especially attractive approach is to genetically encode acetyl-lysine residues using nonsense suppression. This strategy has been successfully applied to single sites of histone acetylation. However, because histone acetylation in nature can often occur at multiple sites simultaneously, we optimized procedures for recombinantly expressing and purifying histone H3 proteins that incorporates up to four sites of lysine acetylation on the histone tail. Histone octamers containing four sites of lysine acetylation were assembled into mono-nucleosomes and enzymatic assays confirmed that this acetylation largely blocks further acetylation by the yeast SAGA acetyltransferase complex.

In the future, it is our hope to combine the use of multiply acetylated histones to probe for what affects their inclusion might have on magnesium binding stoichiometry, size, kinetic growth, and magnesium threshold requirements for self-associated nucleosomal array particles. Any findings might help to differentiate between roles for how these lysine neutralizing PTM's behave as interactors with distal and proximal nucleosomes vs how their loss may change the electrostatics in the system.

CHAPTER 1. GENERAL INTRODUCTION

The genetic material of living organisms is more than just DNA

Life as we understand it, requires a set of instructions or genetic material to exist; it is the basis for determining the structure, and nature each life form on earth. To make life possible, the genetic material must be:

1. copied with high fidelity,
2. passed from one generation to the next without change,
3. utilized or read by the life form which contains it,
4. resistant to damage and repaired whenever damage occurs,
5. complex enough to account for the diversity of life, and finally
6. small enough to fit entirely within the smallest unit of life: the cell.

DNA which is composed of two strands of a polymer of phospho-deoxyribosyl Nucleobases which join together in an antiparallel double helix is well suited to be the genetic material of life and fulfills nearly all of the above listed requirements.

This is because among known biomolecules, only nucleic acids are capable of forming a double-stranded helix that is structurally largely independent of the underlying polymeric sequence of four different bases: adenine, thymine, cytosine, and guanine. Proper helix formation relies heavily on proper hydrogen bonding of paired bases between strands: A to T, and C to G. Thanks to this base pairing, each strand is informationally redundant. The infinite sequence complexity, structural consistency, and informational redundancy are key features which make DNA well suited to be repaired, copied and read/utilized by various cellular machinery such as repair complexes, the replisome, or RNA polymerases.

DNA alone is almost the perfect genetic material and it is in fact commonly presented as the carrier of genetic information in the cell, but it has some drawbacks which make it insufficient to be considered the genetic material, especially in eucaryotic organisms which are more complex than prokaryotic.

Firstly, to contain the full complexity of a multicellular organism such as a Human being, the total DNA sequence is large. (~$6.4x10^9$ bases for Humans) The smallest chromosome (of which there are 46) in Humans is just under $5x10^7$ bases in length (1). With a helical pitch of 3.4 angstroms per base, this equates to DNA of 16,000 microns in length which must fit inside a nucleus. The diameter of which averages about 10 microns but as variable and depends on cell type. One source found a range between 5-20 microns in diameter but the nucleus to cell size ratio is more constant within a single cell type (2). By length, the degrees of freedom by which the long DNA may wrap and coil about greatly increases the level of potential for damaging forces similar to how a fishing line can easily become hopelessly tangled in knots.

Secondly, by volume, the entire human genome takes just 0.1-5.9% of the nucleus space however each phosphate in DNA contributes a negative charge and in total, a considerable amount of work would be required to contain the total negative charge in a confined space like the nucleus. Indeed, even a simpler organism like E. coli with $4.6x10^6$ bases encounters similar charge density issues relative to cell size.

Finally, unicellular organisms must utilize different genes depending on their environment and additionally, for multicellular organisms, one cell differentiates its self from others in part thanks to differential utilization of the genome. DNA alone, though it may have certain sequences which are used as a method of recruitment that offer some level of differential utilization, does not have the ability to block access to cellular machinery that is needed for

differential utilization of the genome. In a sense, what makes DNA good for uniform access by cellular machinery also functions as a drawback since there is no way to recruit or block said machinery at any particular portion of the genome.

Many scientists, including notably Walter Flemming, through research dating back to the 1880's and through the 1960's, discovered that nature overcomes these obstacles evolutionarily by utilizing a suite of positively charged proteins which form a stable complex with the DNA which not only solves the charge compaction issue, but also provides a new platform from which to regulate the organization, and utilization of the genome. Thus, the more accurately defined genetic material is inclusive of both DNA and all of the proteins that are in complex with it. While these nucleoprotein complexes exist in all realms of life, the broad focus of this dissertation is around the nucleoprotein complex found specifically in eucaryotes, named "chromatin" by Flemming.

The primary function of eukaryotic chromatin is to compact the genome but, nature has evolved strategies to both contend with and exploit the regulatory opportunities chromatin provides simply by existing. Thus a second important function of chromatin is that it helps regulate the utilization of the DNA via transcription, (3) replication, (4) and repair (5). Structurally, chromatin is a polymer of protein/DNA complexes with monomeric units called nucleosomes which consist primarily of DNA and positively charged proteins called histones along with a few other proteins like cohesin that bind the DNA (6-10).

Histone proteins

Histone proteins are highly conserved in the eukaryotic kingdom in both sequence and structure. Beyond the eukaryotic nucleosome, the "Histone fold" structural moiety is seen in

other proteins and stretches beyond eukaryotes to encompass the proteins which carry out analogous functions in bacteria and archaea (11,12). Eukaryotes do however utilize many variants of each of the above histones which will act as a substitute and may imbue a nucleosome with unique properties relative to a canonical nucleosome (13).

Histone proteins are basic in nature, containing a relatively high abundance of lysine and arginine residues. They are thus readily able to interact with negatively charged DNA. Through this interaction, chromatin as a whole has ~3/5^{ths} the negative charge of DNA alone. (Table 1) Due to the reduced charge density, chromatin is much more readily compacted when compared with DNA alone.

Table 1. Net charge for the *Xenopus laevis* nucleosome based on 177bp of DNA backbone based on the protein sequences used in recombinant *E. coli* expression systems. (14) Amino acid counts indicated are on a per octamer basis as there are two copies of each histone

Histone type	Arginine	Lysine	Aspartate	Glutamate
Xenopus H2A	26	26	4	14
Xenopus H2B	16	40	6	12
Xenopus H3	36	26	8	14
Xenopus H4	28	22	6	8
Total Charges	+106	+114	-24	-48
177 bp DNA (354 phosphates) Net charge: 106 + 114 – 24 – 48 – 354 = **-206**				

Histone proteins are most highly produced during the S phase of the cell cycle when DNA replication takes place due to the requirement to assemble new nucleosomes as new DNA is made (15). Individual histone proteins will easily aggregate however mechanisms are in place to safely traffic newly expressed histones into the nucleus for deposition to form new nucleosomes on replicated DNA. Proteins called Histone chaperones (such as Nap1) are able to bind and transport histones into the nucleus and act as mediators for proper nucleosome assembly in-vivo (16).

The N and C terminal tails have been especially well studied. The tail domains are flexible, interacting with the local environment where they are critical to all nucleosome functions. They undergo numerous post-translational modifications (PTMs) which are the primary protein-based markers which differentiate one location in chromatin from another. The field of epigenetics is interested in understanding what factors govern cell-cell differentiation when the underlying genome is identical and PTM's are potentially important for understanding epigenetics since they help to differentiate different loci on the genome.

It is possible to purify these histone proteins individually as well as to purify entire nucleosomes from natural sources and this is required to investigate common and novel PTM combinations of which there is infinite potential complexity. In this dissertation however, all chromatin samples were prepared using *Xenopus laevis* histones recombinantly expressed in *E. coli.* This is advantageous for obtaining a homogeneous sample and it avoids the potential noise in the data introduced from whatever unknown affects different PTMs and their combinations may have.

Chromatin lower order structure: Nucleosomes and nucleosomal arrays

Chromatin is a polymer of nucleoprotein complexes called nucleosomes (mononucleosomes to specify singular), which were discovered using Mnase digestion by Kornberg and others in 1975 (17). Individual nucleosomes in chromatin are linked like beads on a string which may be purified from the cell and examined most commonly by EM (18). Through its formation, chromatin achieves a primary level of compaction by shortening the general contour length of the DNA polymer by about 7-fold. (19,20) One nucleosome is composed of a roughly equal mass protein and nucleic acid. It consists of a variable length linker region of DNA connected to 147 bp of DNA which wraps 1.7 times around an octameric

globular protein core of 4 different typical histones: two copies each of H2A, H2B, H3, and H4. (Figure 1)

Figure 1. Cartoon image of first high resolution nucleosome crystal structure (21) in two views. Left: the DNA wrapping around can be seen with Histone h3 in blue, Histone H4 in green, Histone H2B in yellow, and Histone H2A in red. Right: the DNA wrapping around makes the nucleosome's general shape that of a disk and the DNA makes just under two wraps around the Histone core.

The first crystal structure (7 Å resolution) of the mononucleosome was published in 1984 by Tim Richmond based on his work in the Klug lab (22). Since then, the Richmond lab has continued to refine the structure of the mononucleosome. Utilizing alpha-satellite palindromic DNA sequences from the Bunick lab, they have obtained high resolution crystal structure at 1.9 Å (21,23,24). The nucleosome appears in crystal structures as a slightly wedged disc. Protruding from the core are 8 of the N-terminal histone domains as well as 2 C-terminal domains which are most commonly nicknamed histone "tails". These tails are unstructured and thus, not visible in x-ray crystal structures. The tails are of special significance due to their exposure to the surface of the nucleosome where they are known to interact with neighboring nucleosomes, DNA, as well as various other proteins in the nucleus.

Nucleosomes have been well characterized structurally using many methods beyond crystallography. Electron microscopy, (18) cryo-EM, (25) fluorescence microscopy, (26), small angle X-ray scattering,(27) single molecule fluorescence and atomic force microscopy, (28,29) sedimentation, (30) NMR, (31) visible light scattering, (32) and others have been used to gain structural insights about nucleosomes. Most of these works indicate the non-static nature of the nucleosome with many different characterized movements of the DNA and protein components within.

From the beginning, an avenue of interest revolved around understanding DNA sequences which dictate nucleosome formation (33). The Nucleosome does not form randomly on any sequence of DNA with equal probability. Work in-vivo confirms that in a statistical manner characterizing millions of yeast genomes, many nucleosomes are well positioned and play a critical role in genetics (34). Biophysical and structural work in-vitro has often demanded that the position of the octamer on the DNA be homogeneous within the sample. The Widom-601 sequence was developed using SELEX methods to fill this exact role and has been widely used (35,36). This work has shown that Nucleosomes most easily deposit on AT rich sequences thanks to the increased flexibility which adenine-thymine base pairs imbue due to reduced steric resistance against bending when compared to guanidine-cytosine base pairs.

The core of nucleosomes, the octamer, is composed of two different purifiable histone sub-complexes: 2 identical dimers of H2A/H2B and one tetramer of H3/H4. Examining the aforementioned crystal structures, the structure of the octamer may be described as follows. Following the nucleosome DNA from the first base pair, and moving towards the dyad (center of the binding sequence), the structural characteristics of the octamer follow a 2-fold symmetrical helix. It begins with an H2A/H2B Dimer where the H2B interacts in a hydrophobic 4-helix

bundle interaction with H4 and finally H3 located near the dyad axis. The 2-fold symmetrical nature plays out when the inverse is repeated on the other half of the nucleosome where the other H3 histone interacts tightly in a 4-helix bundle right at the dyad axis, followed by H4, which interacts with the H2B of the other H2A/H2B dimer.

The histone octamer, may be reconstituted in-vitro as a stable complex only under very high 1.5-2M salt conditions. After the refolding of histone octamers in high salt, DNA may be added and the octamer will deposit and nucleosomes will form if ideal conditions are provided in the lab which very generally are a buffered pH, and a gradual decrease in salt concentration over a long period of time.

Notably, it has been found that the H3/H4 tetramers deposit around 1 molar salt into an intermediate structure called the tetrasome (37). By 0.8 molar salt, one H2A/H2B dimer deposits forming the hexasome intermediate structure and finally at around 0.6 molar salt, the final dimer deposits to form a full nucleosome. These intermediate structures may be detected in *in vitro* assemblies as having slightly different electrophoretic mobility under ideal native PAGE conditions. It has also been found to be important in and of themselves and are another example of the diversity within chromatin structure (38,39).

This order of deposition and the structural knowledge add evidence for the current model of the dynamics of the nucleosome where the DNA at the entry and exit locations pulls away from the nucleosome, bringing an H2A/H2B dimer along with it in a dynamic fluctuation process called nucleosome “breathing” (40).

Nucleosomes are most stable in low salt conditions but this stability is increased if multiple nucleosomes are refolded on a single long template of DNA (41) into what is called a

nucleosomal array or polynucleosome which follows the same general theories of assembly as mononucleosomes. The nucleosomal array is a polymer of nucleosomes greater than 1 assembled in vitro with the first published works appearing in the 1980's. A typical length of study is 12 nucleosomes; however, a wide variety of arrays have been prepared. As proof of concept, Simpson *et al.* showed that reconstitution was possible on arrays ranging in size from 6 to 25 in length, (42) and later works have shown successful reconstitutions of arrays as long as 80 nucleosomes (43).

Nucleosomal arrays are the most commonly studied model system of chromatin because they allow for the study of interactions between nucleosomes either proximally or distally, and these interactions are what enable the highly dynamic higher order structure or secondary compaction of chromatin beyond the primary 7-fold compaction provided by the DNA wrapping around histones.

Chromatin higher order structure in vitro

As is the case with all polymers, there are multiple levels of structural detail to be aware of and purely from an in vitro perspective, nucleosomal arrays have some interesting polymeric properties. The nucleosome's wedge-shaped disc structure as well as 1.7 turns of DNA give chromatin a shape that is naturally adapted to favor compaction. For example, the open face of the histone octamer contains a small acidic patch that coordinates with neighboring histone tails to allow for more degrees of freedom in interaction with lower energy costs due to charge-charge repulsion (44). The tails are intrinsically disordered positively charged protein domains which "mediate multivalent interactions with DNA and nucleosomes and other chromatin proteins." (45)

Nuclesosomal arrays have long been observed to react to the addition of divalent cations such as magnesium and in literary works as early as 1979, there are mentions of observed "aggregation of oligonucleosomal chains…". (46,47) Depending on divalent cation concentration, the extended 10nm array (figure 2A) will either intra-array associate to form 30 nanometer fibers (Figure 2B) or intra-array associate to form the so-called "aggregations" which are now given formal names like chromatin condensates, oligomers, or self-associated particles (Figure 2C).

Ever since the first observations of 30nm chromatin fibers were observed in EM in 1976 (18), EM studies as well as others have delved deep into learning structural determinants of 30nm fiber formation using short arrays of nucleosomes from 3 to 12 in length. Ultimately these have successfully shown how the 30nm fiber may compact: as a series of tetra-nucleosomal subunits which stack in a left-handed helical fashion (Figure 2B right). (48,49)

Chromatin higher order structure in vivo

Understanding the underlying structure of the nucleosome and nucleosomal array is only interesting insofar as it can help to probe more deeply into the overarching question of how chromatin in the nucleus leads to regulation of the genome. Chromatin is thought to, at least in part, regulate DNA related processes via variation in its secondary compaction. The defined levels of chromatin secondary compaction are that of Euchromatin (open, loose and more accessible to proteins) and Heterochromatin (closed, tight and less to proteins). Broadly speaking, by forming heterochromatin or euchromatin, the underlying DNA becomes

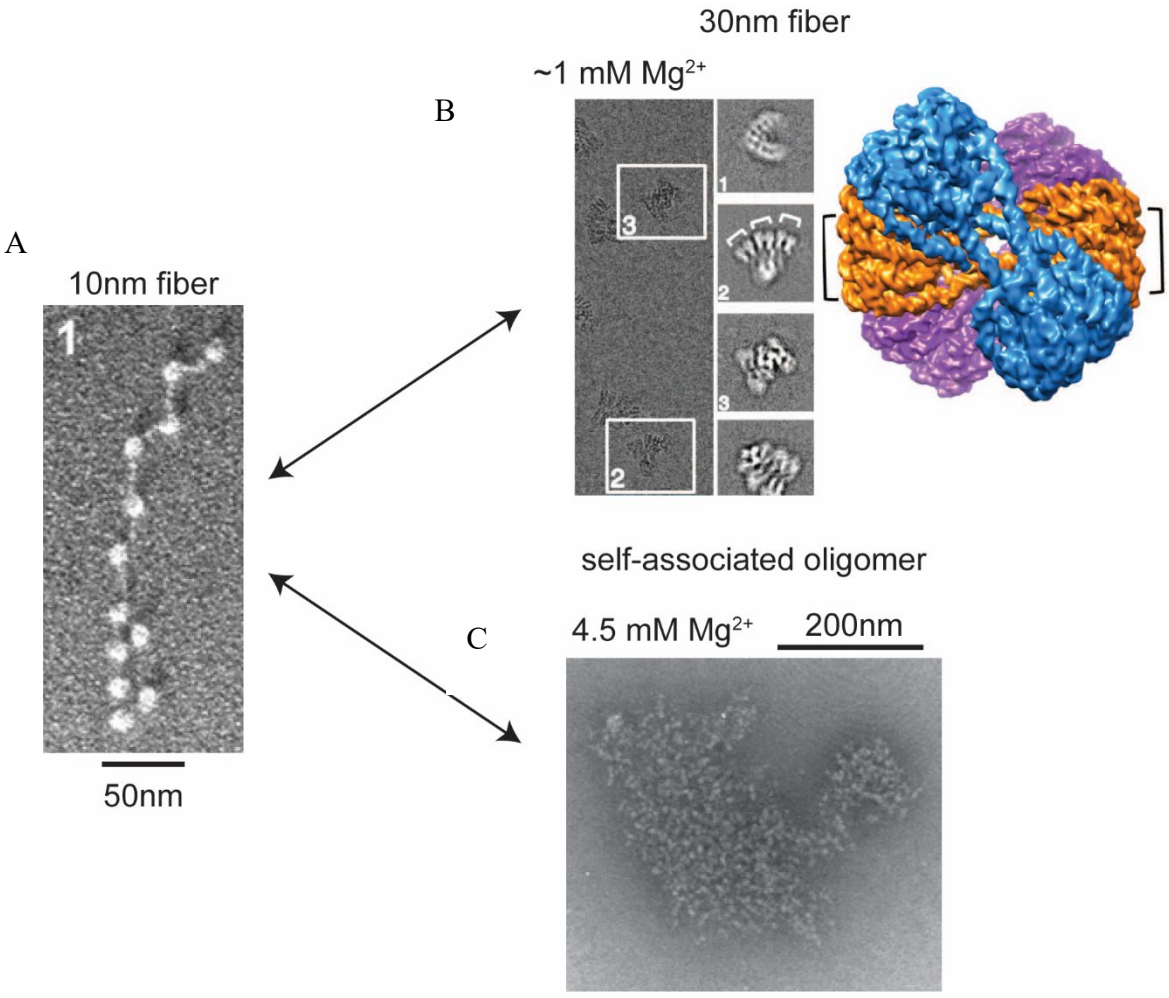

Figure 2. Visualizing the self-associated states of *in vitro* nucleosomal arrays. A) Adapted from (50), Electron micrograph with platinum shadowing portraying a 601-207-12 nucleosomal array in a 10nm fiber partially extended state in a low magnesium environment. **B)** Adapted from (48), Left: cryo-EM micrograph of 601-187-12 nucleosomal arrays induced to form 30nm fibers by addition of ~1mM Mg^{2+} Right: 3D cryo-EM map reconstructed from micrographs such as those portrayed on left. **C)** Adapted from (27), negative stain electron micrograph of self-associated array oligomer particle which were caused to form via addition of 4.5 mM Mg^{2+} to a sample of 601-207-12 nucleosomal array such as in A.

transcriptionally controlled. Many factors contribute to how compact or loose an area of

chromatin becomes including:

1. post-translational modifications of the histones (PTM's), (51-53)
2. variation in protein complexes including variable histones, (13,43,54)
3. chromatin remodeling, (55) and

4. DNA methylation (56,57).

Furthermore, additional proteins, so-called linker histones, exist that can further reduce chromatin negative charge and their binding results in especially dense regions of chromatin called heterochromatin. (43) All of these modifications do indeed cause 30nm fiber formation in vitro. These 30nm structures, by their nature of being easily observable using EM, SAX, and others, are stable in structure and so the prevailing viewpoint has been that chromatin in the nucleus has a stable and static structure of euchromatic and heterochromatic regions which were remodeled slowly by enzymes.

Thus, the dogma of chromatin has been one which portrayed aesthetically pleasing order in the compaction of chromatin *in vivo*. (58) (Figure 3A) In brief: Nucleosomes form a polymer like beads on a string. The nucleosomes come together to arrange themselves into regularly ordered 30nm fibers and indeed, chromatin model systems were seen to model this structure as well as early SAX studies indicated presence of something in the size of 30nm in multiple cell types (59-62). Overwhelming evidence however points a lack of 30nm fiber in the cell and arguments carefully illustrate that the early SAX studies likely had ribosomes of 30nm which were in the nucleus and were detected in those studies (63).

A more modern view of chromatin has rapidly taken the stage in the last 10 years which portrays chromatin as a dynamic complicated interdigitation of nucleosomes in a semi-solid or fluid like state which fills a given space as dictated by the local cellular regulatory modifications like linker histone binding, or cohesion binding, and formation of domains (Figure 3B) (45).

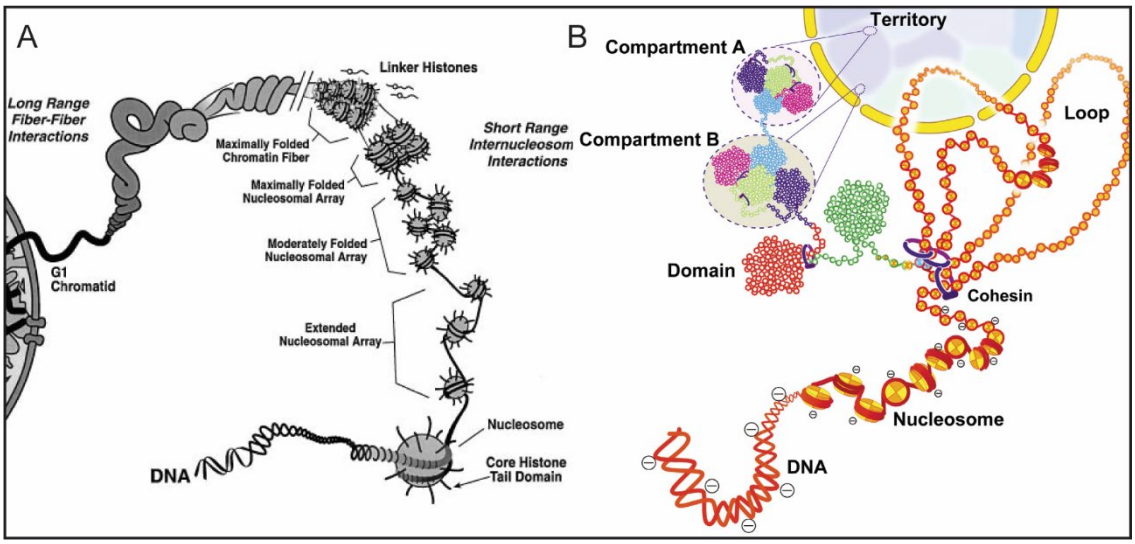

Figure -3. Comparison 2002 vs 2020 model of chromatin higher order structure. A) Figure by Jeff Hansen (58) portrayal of chromatin higher order structure as it was understood in 2002. Individual nucleosomes were thought to exist in states of intermediate folding ranging from extended 10nm arrays which would fold into 30nm arrays which would then further fold and eventually coil into chromatids which fill the nucleus. **B)** Figure by Kazuhiro Maeshima et al. (45) Modern simplified portrayal of chromatin higher order. Nucleosomal arrays generally associate in domains of disordered gel-like domains which join together into compartments in transcriptionally active euchromatin (compartment B) and transcriptionally inactive heterochromatin (compartment A). All compartments of a single chromosome then generally occupy a distinct chromosome territory within the nucleus.

Using in vitro structural insight to learn about in vivo properties

While initially it was thought that self-associated particle formation was less important than the 30nm fiber, the focus has now inverted. Several different methods have found that chromatin self-associated particles form particles with radius averaging about 250-500nm. Both the structural sizes and rigidity of particles in vitro, and the structural rigidity of chromatin are seen to recapitulate what is seen in vivo (26,27,64).

H3 Histone acetylation: a mechanism of Chromatin regulation

One mechanism for regulating chromatin is through PTMs of histone proteins. Histone modifications occur at many different sites within histone, especially on the histone tails, and occur with high frequency throughout the genome. Histone acetylation was one of the first

discovered histone PTMS (65). It is well documented that acetylated domains of chromatin are more commonly associated with highly transcribed areas found in the loosely packed euchromatin (66,67). The acetylation of chromatin occurs through catalysis by many lysine acetyl-transferase enzyme complexes. Acetyl-CoA and lysine are the substrates which form products sCoA and Acetyl-lysine. One found in yeast is of particular interest in the Shogren-Knaal lab: The Spt-Ada-Gcn5-acetyltransferase complex (SAGA). This complex is found to be involved with activation of transcription in genes which are inducible. It predominantly targets four loci on the histone H3 n-terminal tail: Lysine residues 9, 14, 18, and 23 (68) (Figure 4). Histones targeted by saga are not uniformly acetylated at all 4 sites. There is, in fact, a preferential order of acetylation. Lysine 14 is most heavily acetylated, followed by 9, then 18, and least preferred is residue 23. Every combination of acetylation possible may have specific unique effects in-vivo and it is of interest in our lab to investigate biophysical differences generated by these various combinations.

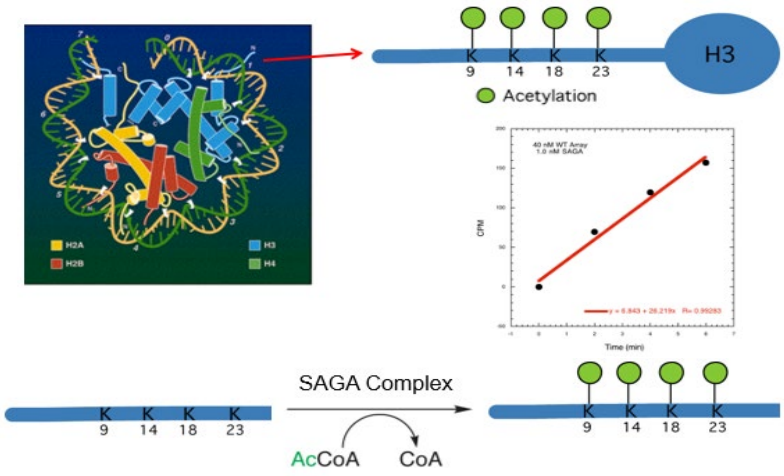

Figure 4. Major targets of acetylation by the SAGA complex. The N-terminal tail of Histone H3 protrudes from the core near the DNA entry/exit point and is acetylated primarily at sites 9, 14, 18 and 23. Lower: Yeast SAGA complex is routinely assayed for activity using the N-terminal tail peptide and radioactive acetyl-CoA. Inset graph: Kinetics of SAGA under saturating levels of substrate. Activity is measured by scintillation counting.

Acetylation of lysine in a histone is a common PTM and causes two chemical changes to the histone: One, the histone loses a positive charge under all physiological pH levels. And two, the histone has a different size and shape. The change in ionic charge of the histone has obvious implications when the highly negatively charged polyphosphate backbone of DNA is considered; the histones bind strongly to DNA thanks in part to the attraction between positively charged histones and negative charges of DNA. Indeed, it has been shown that acetylated nucleosomes are less stable than unacetylated nucleosomes (66,69,70). The change in size and shape of the histone also allows the acetylated lysine to serve as a binding site for chromatin modifying enzymes, transcription factors, as well as proteins involved in DNA replication (71), and DNA repair (72). In this way, the acetylation mark may recruit DNA machinery to a specific genetic locus in the organism. Specifically, any protein with one or more bromodomains may be recruited to acetylated nucleosomes in-vivo.

A major method by which the function of acetylated histones can been determined is through in-vitro biochemical and biophysical studies with systems containing acetylated lysine. Early studies focused on using acetylated peptides of the H3 tail. While such peptides can be quite useful, in many cases it has been revealed that chromatin-associated proteins interact with domains outside of isolated histone tails. For example, in the Shogren-Knaak lab, acetylated histone h3 tails in nucleosomal arrays (a common in-vitro chromatin model system) have been studied to ascertain information about SAGA preference for acetylating one tail in a nucleosome when the other tail is already acetylated (73).

In-vitro Techniques to Obtain Histones with Acetylation

Because of the utility of acetylated histones for understanding their function, a number of strategies have been explored, each with their strengths and limitations.

Purification from natural and unnatural sources

Histones from various eukaryotic tissues are full of unique PTM patterns that include recognizable epigenetic acetylation patterns. It is in large part thanks to PTM patterns that the cells of multicellular organisms are able to differentiate. It can be useful to study reconstituted nucleosomes using histone octamers purified from specific tissues. Novel properties might be elucidated by nucleosomes isolated from different tissue types. Immortal cell lines such as helacells and cancer cell lines will also have histones that may be purified and used for study (70).

The primary disadvantage of this technique is that the histone samples will always be highly heterogeneous. The histone code implies the idea that the differential expression is highly location specific within chromatin and technologies for studying a single length of chromatin from a single cell give limited information.

Chemical acetylation of histones from recombinant sources

E-coli was developed as a tool to express proteins from a plethora of different organisms. Expressing in e-coli has the advantage of generating proteins that are largely PTM free. This makes the resulting proteins an excellent substrate for subsequent experiments to be performed that study any enzyme that catalyzes the generation of PTMs on proteins. In order to study Histones that have only acetylation present, histones must first be produced in e-coli. Subsequently, it has been shown that mixing the histones with acetyl phosphate at 50 deg C will produce histones with lots of acetylation (70). Again, this technique has the drawback that the samples produced are always heterogeneous.

Using alternative residues as acetyl-lysine analogs (cysteine derivatives and glutamine)

Another advantage of using e-coli is that the DNA given to the e-coli may be easily modified using standard E-coli cloning and mutagenesis. Acetylated lysine is quite similar in terms of its properties to glutamine. (Figure 5) Indeed, in the literature, there are examples where this glutamine mutagenic strategy does not appear to do good job of recapitulating the activity of acetylated lysine (74).

Incorporating cysteine residues to replace choice lysine residues allows subsequent chemical steps to generate acetyl lysine analogs (75-79). This technique is useful because the cysteine residue is the most nucleophilic of all 20 amino acid residues. In this way, organic chemists can easily use unique cysteine side-chains on a protein to perform organic reactions. Cysteine acetyl-lysine analogs have disadvantages of not being exactly the same, chemically, as acetyl-lysine and are costlier and more time consuming than other methods available, including native chemical ligation. Cysteine analogs of specifically methylated lysines on H3 have been successful however (75,80).

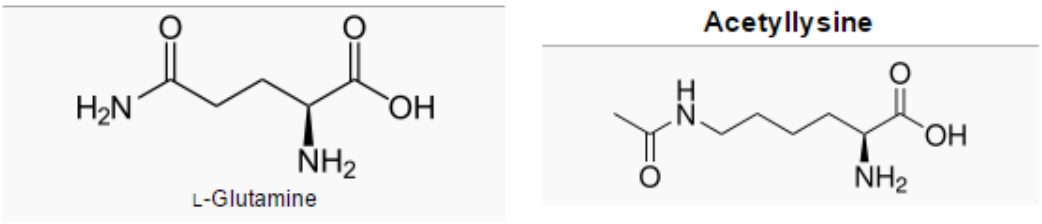

Figure 5. Comparison of Acetyllysine and a canonical amino acid analog, L-Glutamine. They both contain an acetyl group on the gama linked chain of the amino acid. Acetyl-lysine is however 2 carbons longer, and the primary amide of l-glutamine will be slightly more reactive than the secondary amide of acetyl-lysine.

In-vitro enzyme mediated acetylation

PCAF and CBP/P300 are histone acetyltransferases which are known to generate specific acetylated histone products (72). The primary disadvantage of this approach is that very few PTMs have a known enzyme that generates only one acetylated histone product. Additionally, products will be heterogeneous, as it is difficult to drive an enzymatic reaction to completion.

Native chemical ligation and expressed protein ligation

In native chemical ligation, an N-terminal tail of a histone is ligated to the globular domain of the histone protein (81,82). The C-terminal globular domain is mutated to cysteine at the N-terminal residue and the N-terminal tail is synthesized using solid phase peptide synthesis standard Fmoc chemistry. The C terminal residue is then thioesterified. The resulting thioesterified N-terminal tail must be purified via reversed-phase high pressure liquid chromatography (HPLC) purification which proves costly in terms of yield and is a time intensive process. (Figure 6) Additionally, a cysteine mutation in introduced by the process which may add complications if further chemistry is required to modify the histone product.

Enzyme mediated ligation of peptides

Recent work introduced an enzyme-mediated approach to ligate a synthesized N-terminal tail histone peptide to the globular domain of histone H3 (83). In this process, the F40 sortase requires a recognition motif on the C-terminal of the peptide in order to ligate to a di-glycine N-terminal globular histone h3 protein. This strategy requires the production of peptide, similar to native chemical ligation but uses less peptide. This is an improvement in that the cost is reduced, since less peptide needs to be purchased, and the peptide product needs to be purified less extensively due to the selectivity of the enzyme. However, this approach is still potentially

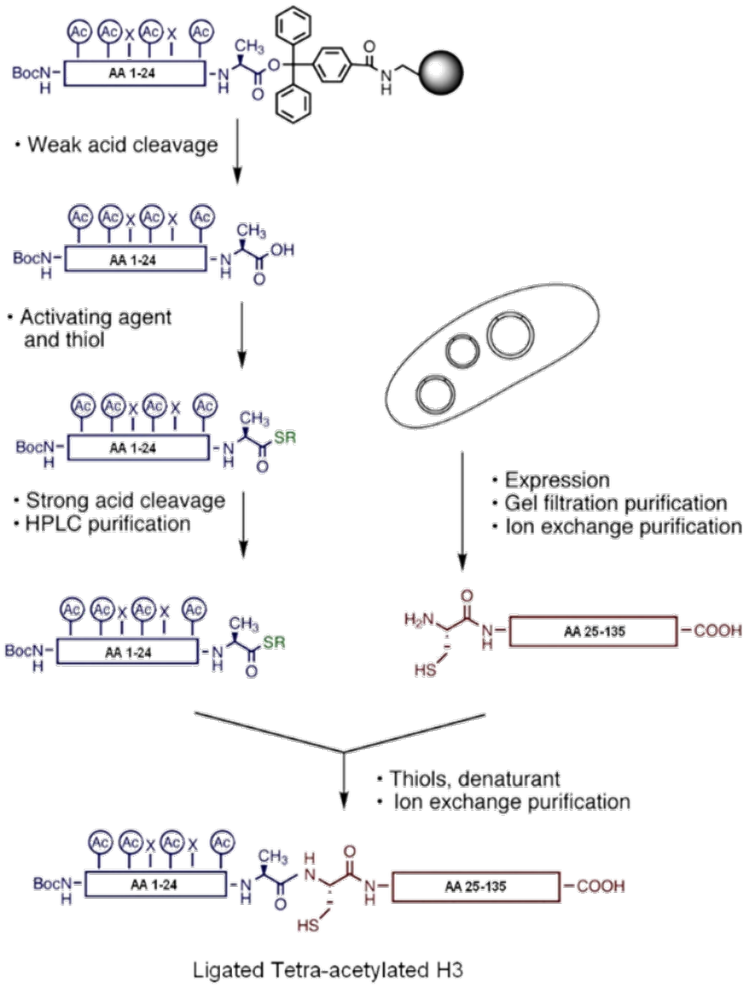

Figure 6. Diagram of Acetylated histone produced via native Chemical ligation. Weak acid cleavage removes acetylated N-terminal tail peptide produced via solid phase peptide synthesis without deprotecting the side chains. Thioesterification reaction followed by strong acid cleavage of protecting groups activates the c terminal of the peptide for ligation to a globular domain histone protein recombinantly expressed in e-coli.

difficult to scale-up, and the introduced ligase recognition sequence incorporated into the histone could alter the properties of the histone.

Nonsense suppression: a technique to generate acetylated histone H3

To address some of the shortcomings of the available approaches for generating multiply-acetylated H3 histones, this chapter 3 explores their generation and application via nonsense suppression mutagenesis. This technique, developed initially by the Schultz lab (84), utilizes a

unique tRNACUA-Synthetase pair in E. coli to specifically incorporate a non-coding amino acid at an introduced nonsense codon site in a gene of interest (Figure 7). Recently, this technique was re-engineered to generate histone H3 acetylated at K56 (85). It was not, however, shown to be useful to generate histones with more than a single lysine acetylation. Investigating whether more acetyl-lysine residues could be incorporated using this strategy is worthwhile, as it would expand the number and combination of acetyl-lysine sites that could be investigated, and potentially reduce the time and cost of generating such species. However, uncertainty remained as to whether multiple nonsense suppression events could be supported. At the time of this research, there was no clear precedent for multiple nonsense suppression events, but recently it was shown that two unnatural amino acids could be incorporated to generate an intra-protein FRET pair(86).

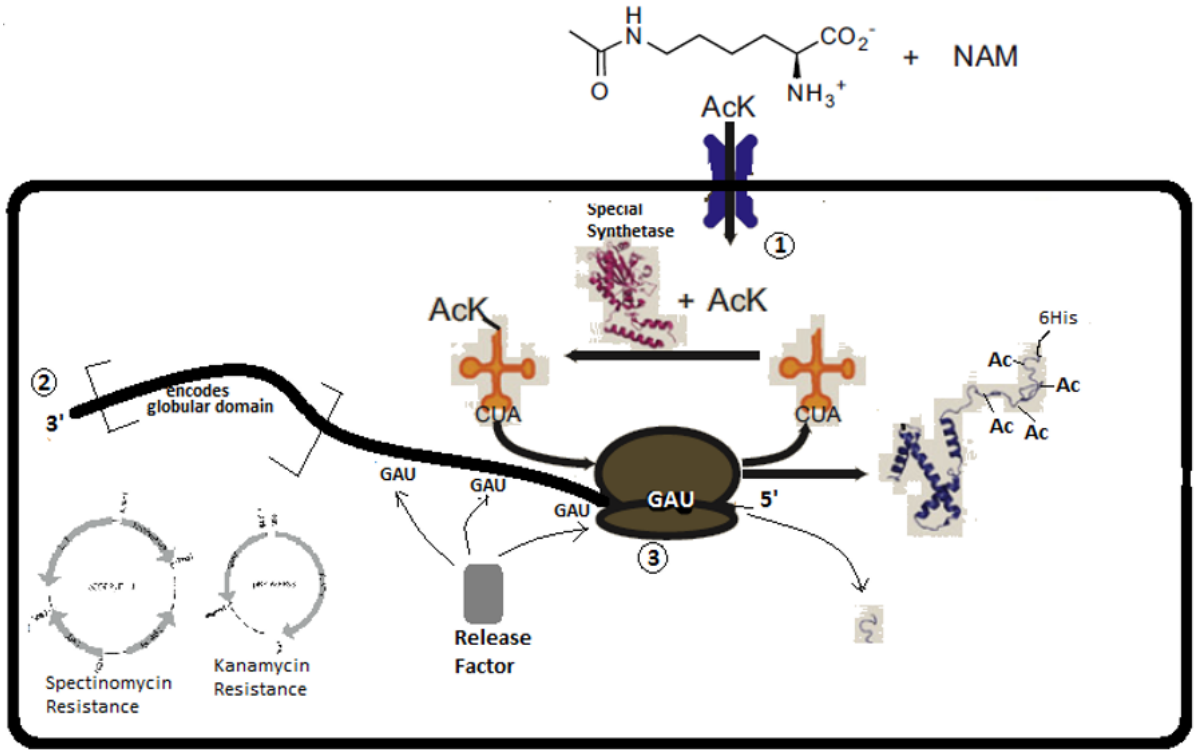

Figure 7 General mechanism of nonsense suppression to produce acetylated histones. (84,85) Rounded large black box represents an E-Coli¬ cell. Inside the cell, two plasmids are inserted

which confer the cell with resistance to Spectinomycin and Kanamycin. The cell produces a special t-RNACUA and a special acetyl-lysine synthetase. (1) The t-RNACUA is recognized by the Synthetase which has been engineered to accept and install acetyl-lysine (ACK: which is made available in the media) on the t-RNACUA. (2) mRNA containing amber stop codons (portrayed as 3'-GAU-5') transcribed from the plasmid under an inducible LAC promoter by the cell. (3) A fully assembled ribosomal complex with mRNA attached which is actively translating the mRNA into protein containing both full length and truncated versions of acetylated histone. The truncated histone (lower right) is an undesirable side product produced due to Release factor which competes for binding to the amber stop codon and prematurely terminates translation.

Content of this dissertation.

The next 2 chapters and appendix of this dissertation represent manuscripts. Chapter 2 was published in 2016 in the journal of protein expression and purification. Chapter 3 is not yet published and we have not yet selected a Journal for submission to publish. The Appendix represents a manuscript where contributions by Isaac Young did not include production of figures or writing but rather involved in early work to develop the strategy for production of reagents. Thus, this manuscript was included as an appendix.

Bibliography

1. Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J. T., Threadgold, G., Torrance, J., Wood, J., Clarke, L., Koren, S., Boitano, M., Li, H., Chin, C.-S., Phillippy, A. M., Durbin, R., Wilson, R. K., Flicek, P., and Church, D. M. (2016) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *bioRxiv*, 072116

2. Moore, M., Sebastian, J., and Kolios, M. (2019) Determination of cell nucleus-to-cytoplasmic ratio using imaging flow cytometry and a combined ultrasound and photoacoustic technique: a comparison study. *Journal of Biomedical Optics* **24**, 106502

3. Reinke, H., and Horz, W. (2003) Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol Cell* **11**, 1599-1607

4. Lipford, J. R., and Bell, S. P. (2001) Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. *Mol Cell* **7**, 21-30.

5. Wellinger, R. E., and Thoma, F. (1997) Nucleosome structure and positioning modulate nucleotide excision repair in the non-transcribed strand of an active gene. *Embo J* **16**, 5046-5056.

6. Olins, D. E., and Olins, A. L. (2003) Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**, 809-814

7. Lukács, D. (1981) Walter Flemming, discoverer of chromatin and mitotic cell division. *Orvosi Hetilap* **122**, 349-350

8. Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014) Chromatin as dynamic 10-nm fibers. *Chromosoma* **123**, 225-237

9. Michaelis, C., Ciosk, R., and Nasmyth, K. (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35-45.

10. Gruber, S., Haering, C. H., and Nasmyth, K. (2003) Chromosomal Cohesin Forms a Ring. *Cell* **112**, 765-777

11. Sanders, T. J., Marshall, C. J., and Santangelo, T. J. (2019) The Role of Archaeal Chromatin in Transcription. *Journal of Molecular Biology* **431**, 4103-4115

12. Dillon, S. C., and Dorman, C. J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology* **8**, 185-195

13. Kamakaka, R. T., and Biggins, S. (2005) Histone variants: deviants? *Genes & Development* **19**, 295-316

14. Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999) Expression and Purification of Recombinant Histones and Nucleosome Reconstitution. in *Chromatin Protocols* (Becker, P. B. ed.), Humana Press, Totowa, NJ. pp 1-16

15. Marzluff, W. F., and Duronio, R. J. (2002) Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Current Opinion in Cell Biology* **14**, 692-699

16. Burgess, R. J., and Zhang, Z. (2013) Histone chaperones in nucleosome assembly and human disease. *Nature structural & molecular biology* **20**, 14-22

17. Kornberg, R. D. (1977) Structure of chromatin. *Annu Rev Biochem* **46**, 931-954

18. Woodcock, C. L. F., Safer, J. P., and Stanchfield, J. E. (1976) Structural repeating units in chromatin. *Experimental Cell Research* **97**, 101-110

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

20. Németh, A., and Längst, G. (2004) Chromatin higher order structure: Opening up chromatin for transcription. *Briefings in Functional Genomics* **2**, 334-343

21. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal Structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

22. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984) Structure of the nucleosome core particle at 7 [angst] resolution. *Nature* **311**, 532-537

23. Vence, T. (2015) Crystal Unclear. in *TheScientist* (Aberlin, M. B. ed., LabX Media Group, Midland, Ontario, Canada

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

25. Migl, D., Kschonsak, M., Arthur, C. P., Khin, Y., Harrison, S. C., Ciferri, C., and Dimitrova, Y. N. (2020) Cryoelectron Microscopy Structure of a Yeast Centromeric Nucleosome at 2.7 Å Resolution. *Structure* **28**, 363-370.e363

26. Strickfaden, H., Tolsma, T. O., Sharma, A., Underhill, D. A., Hansen, J. C., and Hendzel, M. J. (2020) Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell* **183**, 1772-1784.e1713

27. Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., DeLuca, J., Ishikawa, T., and Hansen, J. C. (2016) Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *The EMBO Journal* **35**, 1115-1132

28. Shlyakhtenko, L. S., Lushnikov, A. Y., and Lyubchenko, Y. L. (2009) Dynamics of Nucleosomes Revealed by Time-Lapse Atomic Force Microscopy. *Biochemistry* **48**, 7842-7848

29. Park, Y. J., Dyer, P. N., Tremethick, D. J., and Luger, K. (2004) A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. *J Biol Chem* **279**, 24274-24282

30. Azzaz, A. M., Vitalini, M. W., Thomas, A. S., Price, J. P., Blacketer, M. J., Cryderman, D. E., Zirbel, L. N., Woodcock, C. L., Elcock, A. H., Wallrath, L. L., and Shogren-Knaak, M. A. (2014) Human heterochromatin protein 1alpha promotes nucleosome associations that drive chromatin condensation. *J Biol Chem* **289**, 6850-6861

31. Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J., and Musselman, C. A. (2018) The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *eLife* **7**, e31481

32. Blacketer, M. J., Gannon, M. K., Young, I. A., and Shogren-Knaak, M. A. (2017) Solid-phase synthesis of highly repetitive chromatin assembly templates. *Analytical Biochemistry* **531**, 12-15

33. Lacy, E., and Axel, R. (1975) Analysis of DNA of isolated chromatin subunits. *Proc Natl Acad Sci U S A* **72**, 3978-3982

34. Brogaard, K., Xi, L., Wang, J. P., and Widom, J. (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496-501

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

36. Huynh, V. A., Robinson, P. J., and Rhodes, D. (2005) A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone. *J Mol Biol* **345**, 957-968

37. Hansen, J. C., van Holde, K. E., 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

38. Levendosky, R. F., Sabantsev, A., Deindl, S., and Bowman, G. D. (2016) The Chd1 chromatin remodeler shifts hexasomes unidirectionally. *eLife* **5**, e21356

39. Rhee, Ho S., Bataille, Alain R., Zhang, L., and Pugh, B. F. (2014) Subnucleosomal Structures and Nucleosome Asymmetry across a Genome. *Cell* **159**, 1377-1388

40. Chen, Y., Tokuda, J. M., Topping, T., Sutton, J. L., Meisburger, S. P., Pabit, S. A., Gloss, L. M., and Pollack, L. (2014) Revealing transient structures of nucleosomes as DNA unwinds. *Nucleic Acids Res* **42**, 8767-8776

41. Blacketer, M. J., Feely, S. J., and Shogren-Knaak, M. A. (2010) Nucleosome interactions and stability in an ordered nucleosome array model system. *J Biol Chem* **285**, 34597-34607

42. Simpson, R. T., Thoma, F., and Brubaker, J. M. (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

43. Routh, A., Sandin, S., and Rhodes, D. (2008) Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proceedings of the National Academy of Sciences* **105**, 8872-8877

44. Sinha, D., and Shogren-Knaak, M. A. (2010) Role of direct interactions between the histone H4 Tail and the H2A core in long range nucleosome contacts. *J Biol Chem* **285**, 16572-16581

45. Maeshima, K., Tamura, S., Hansen, J. C., and Itoh, Y. (2020) Fluid-like chromatin: Toward understanding the real chromatin organization present in the cell. *Current Opinion in Cell Biology* **64**, 77-89

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

47. Straetling, W. H. (1979) Role of histone H1 in the conformation of oligonucleosomes as a function of ionic strength. *Biochemistry* **18**, 596-603

48. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R. M., Zhu, P., and Li, G. (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380

49. Bednar, J., Horowitz, R. A., Dubochet, J., and Woodcock, C. L. (1995) Chromatin conformation and salt-induced compaction: three-dimensional structural information from cryoelectron microscopy. *J Cell Biol* **131**, 1365-1376.

50. Grigoryev, S. A., and Woodcock, C. L. (2012) Chromatin organization - the 30 nm fiber. *Exp Cell Res* **318**, 1448-1455

51. Garcia, B. A., Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2007) Pervasive combinatorial modification of histone H3 in human cells. *Nature Methods* **4**, 487-489

52. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A., and Garcia, B. A. (2009) High Throughput Characterization of Combinatorial Histone Codes. *Molecular & Cellular Proteomics* **8**, 2266-2284

53. Hansen, J. C., Tse, C., and Wolffe, A. P. (1998) Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry* **37**, 17637-17641

54. Robinson, P. J. J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R. G., and Rhodes, D. (2008) 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. *Journal of Molecular Biology* **381**, 816-825

55. Lorch, Y., Zhang, M., and Kornberg, R. D. (1999) Histone octamer transfer by a chromatin-remodeling complex. *Cell* **96**, 389-392

56. Qureshi, I. A., and Mehler, M. F. (2010) Emerging Role of Epigenetics in Stroke: Part 1: DNA Methylation and Chromatin Modifications. *Archives of Neurology* **67**, 1316-1322

57. Honda, S., Lewis, Z. A., Shimada, K., Fischle, W., Sack, R., and Selker, E. U. (2012) Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. *Nat Struct Mol Biol* **19**, 471-477, S471

58. Hansen, J. C. (2002) Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct* **31**, 361-392

59. Scheffer, M. P., Eltsov, M., and Frangakis, A. S. (2011) Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences* **108**, 16992-16997

60. Paulson, J. R., and Langmore, J. P. (1983) Low angle x-ray diffraction studies of HeLa metaphase chromosomes: effects of histone phosphorylation and chromosome isolation procedure. *Journal of Cell Biology* **96**, 1132-1137

61. Kizilyaprak, C., Spehner, D., Devys, D., and Schultz, P. (2010) In Vivo Chromatin Organization of Mouse Rod Photoreceptors Correlates with Histone Modifications. *PLOS ONE* **5**, e11039

62. Scheffer, M. P., Eltsov, M., Bednar, J., and Frangakis, A. S. (2012) Nucleosomes stacked with aligned dyad axes are found in native compact chromatin in vitro. *J Struct Biol* **178**, 207-214

63. Joti, Y., Hikima, T., Nishino, Y., Kamada, F., Hihara, S., Takata, H., Ishikawa, T., and Maeshima, K. (2012) Chromosomes without a 30-nm chromatin fiber. *Nucleus* **3**, 404-410

64. Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S., and Rosen, M. K. (2019) Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e421

65. Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964) Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**, 786-794

66. Tse, C., Sera, T., Wolffe, A. P., and Hansen, J. C. (1998) Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III. *Molecular and Cellular Biology* **18**, 4629-4638

67. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **325**, 834-840

68. Grant, P. A., Eberharter, A., John, S., Cook, R. G., Turner, B. M., and Workman, J. L. (1999) Expanded Lysine Acetylation Specificity of Gcn5 in Native Complexes. *Journal of Biological Chemistry* **274**, 5895-5900

69. Gansen, A., Tóth, K., Schwarz, N., and Langowski, J. (2009) Structural Variability of Nucleosomes Detected by Single-Pair Förster Resonance Energy Transfer: Histone Acetylation, Sequence Variation, and Salt Effects. *The Journal of Physical Chemistry B* **113**, 2604-2613

70. Tóth, K., Brun, N., and Langowski, J. (2006) Chromatin Compaction at the Mononucleosome Level. *Biochemistry* **45**, 1591-1598

71. Wang, X., Li, J., Schowalter, R. M., Jiao, J., Buck, C. B., and You, J. (2012) Bromodomain Protein Brd4 Plays a Key Role in Merkel Cell Polyomavirus DNA Replication. *PLoS Pathog* **8**, e1003021

72. Das, C., Lucia, M. S., Hansen, K. C., and Tyler, J. K. (2009) CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature* **459**, 113-117

73. Li, S., and Shogren-Knaak, M. A. (2009) The Gcn5 Bromodomain of the SAGA Complex Facilitates Cooperative and Cross-tail Acetylation of Nucleosomes. *Journal of Biological Chemistry* **284**, 9411-9417

74. VanDemark, A. P., Kasten, M. M., Ferris, E., Heroux, A., Hill, C. P., and Cairns, B. R. (2007) Autoregulation of the rsc4 tandem bromodomain by gcn5 acetylation. *Mol Cell* **27**, 817-828

75. Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. (2007) The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128**, 1003-1012

76. Guo, J., Wang, J., Lee, J. S., and Schultz, P. G. (2008) Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angewandte Chemie International Edition* **47**, 6399-6401

77. Huang, R., Holbert, M. A., Tarrant, M. K., Curtet, S., Colquhoun, D. R., Dancy, B. M., Dancy, B. C., Hwang, Y., Tang, Y., Meeth, K., Marmorstein, R., Cole, R. N., Khochbin, S., and Cole, P. A. (2010) Site-Specific Introduction of an Acetyl-Lysine Mimic into Peptides and Proteins by Cysteine Alkylation. *Journal of the American Chemical Society* **132**, 9986-9987

78. Li, F., Allahverdi, A., Yang, R., Lua, G. B. J., Zhang, X., Cao, Y., Korolev, N., Nordenskiöld, L., and Liu, C.-F. (2011) A Direct Method for Site-Specific Protein Acetylation. *Angewandte Chemie International Edition* **50**, 9611-9614

79. Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J., and Davis, B. G. (2012) Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angewandte Chemie* **124**, 1871-1875

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

81. Shogren-Knaak, M. A., and Peterson, C. L. (2003) Creating Designer Histones by Native Chemical Ligation. in *Methods in Enzymology*, Academic Press. pp 62-76

82. Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **311**, 844-847

83. Ringel, A. E., Cieniewicz, A. M., Taverna, S. D., and Wolberger, C. (2015) Nucleosome competition reveals processive acetylation by the SAGA HAT module. *Proceedings of the National Academy of Sciences* **112**, E5461-E5470

84. Ellman, J. A., Mendel, D., and Schultz, P. G. (1992) Site-specific incorporation of novel backbone structures into proteins. *Science* **255**, 197+

85. Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J. W. (2009) A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Molecular Cell* **36**, 153-163

86. Lee, T. C., Kang, M., Kim, C. H., Schultz, P. G., Chapman, E., and Deniz, A. A. (2016) Dual Unnatural Amino Acid Incorporation and Click-Chemistry Labeling to Enable Single-Molecule FRET Studies of p97 Folding. *Chembiochem : a European journal of chemical biology* **17**, 981-984

CHAPTER 2. EXPRESSION AND PURIFICATION OF HISTONE H3 PROTEINS CONTAINING MULTIPLE SITES OF LYSINE ACETYLATION USING NONSENSE SUPPRESSION

Isaac A Young[1,4], Chitvan Mittal[2,4], Michael Shogren-Knaak[3,4]

[1]Primary author, all research except [2], [2]Contributed figure and method for histone acetyl-transferase assay, [3]Principle investigator and grantsmanship, [4]Iowa State University Affiliation

Modified from a manuscript published in *Protein Expression and Purification*

Abstract

Lysine acetylation is a common post-translational modification, which is especially prevalent in histone proteins in chromatin. A number of strategies exist for generating histone proteins containing lysine acetylation, but an especially attractive approach is to genetically encode acetyl-lysine residues using nonsense suppression. This strategy has been successfully applied to single sites of histone acetylation. However, because histone acetylation can often occur at multiple sites simultaneously, we were interested in determining whether this approach could be extended. Here we show that we can express histone H3 proteins that incorporate up to four sites of lysine acetylation on the histone tail. Because the amount of expressed multi-acetylated histone is reduced relative to the wild type, a purification strategy involving affinity purification and ion exchange chromatography was optimized. This expression and purification strategy ultimately generates H3 histone uniformly acetylated at the desired position at levels and purity sufficient to assemble histone octamers. Histone octamers containing four sites of lysine acetylation were assembled into mononucleosomes and enzymatic assays confirmed that this acetylation largely blocks further acetylation by the yeast SAGA acetyltransferase complex.

Introduction

Lysine acetylation is a pervasive post-translational protein modification, where approximately 3,600 sites of lysine acetylation in over 1,700 different proteins have been identified through proteomic studies of human cells (1). This modification helps to regulate proteins with a wide range of functions, including those involved in metabolism (2), cell structure (3), and cell signaling (4).

Histone proteins that comprise chromatin constitute an especially large class of acetylation targets. Histone proteins H2A, H2B, H3 and H4, are highly basic proteins that contain a high density of lysine residues. Two copies of each of these proteins can form a histone octamer, around which DNA can wrap to form a nucleosome (5), the fundamental structural unit of chromatin. In nucleosomes, the bulk of the histone sequence is contained within the wraps of the DNA in the globular histone domain. However, less structured regions of each histone extend past the DNA to form histone "tails." At least thirty different sites of lysine acetylation have been identified in histones, with multiple sites identified in each (6). Modifications occur both in the globular and tail regions of the histone, although a significantly greater number of sites are present in the histone tails.

One strategy, to better understand how histone acetylation affects chromatin structure and function, is to reconstitute chromatin in vitro using acetylated histones. In this strategy, the main challenge is to obtain histones that contain lysine acetylation at the desired sites. Direct purification of uniformly acetylated histones is generally impractical, because the heterogeneity of acetylation sites coupled with their similar physical properties makes isolation difficult. Enzymatic acetylation of histones has had some success. However, ensuring that only desired lysines are acetylated, and are acetylated fully, is often not easy (7). As an alternative, we have

previously adapted the strategy of native chemical ligation to incorporate a range of histone modifications in the H3 and H4 tails (8,9). In this strategy, a histone tail peptide containing desired modifications is generated by standard solid-phase peptide synthesis strategies, and then ligated to a recombinantly expressed histone fragment corresponding to the remaining sequence. This strategy yields uniformly modified, full-length histone in reasonable yields. However, because peptide synthesis is relatively costly, and the steps involved are laborious, we were interested in exploring alternative strategies for incorporating different patterns of lysine acetylation into the histone tails.

A promising strategy that has recently been applied to generating acetylated histone is nonsense suppression expression. In nonsense suppression expression (10), the idea is to genetically encode for a non-standard amino acid that becomes incorporated during translation. In its most common form, the non-standard amino acid is introduced into a specific site in a protein by first mutating its sequence to an amber stop codon. By including a modified tRNA that contains an amber anticodon and a site for attaching a desired nonstandard amino acid, as well as a tRNA synthetase that can charge the tRNA with the desired amino acid, the nonstandard amino acid becomes incorporated into the protein during translation. Chin and coworkers showed that a modified pyrrolysyl- tRNA synthetase derived from M. barkeri can efficiently charge a tRNA with acetyl-lysine, and that this residue can be incorporated into histone H3 with good yields (11). With this strategy acetyl-lysine was incorporated at multiple sites within the H3 histone. However, in all cases, each histone contained only a single acetyl-lysine.

In vivo, acetylation of multiple sites within the same histone is common (12), where the pattern of this acetylation can change the binding of chromatin-associated proteins or directly

change chromatin structure and stability. For examples, dual bromodomains found in a number of nuclear proteins recognize and bind to specific pairs of lysine acetylation, thereby targeting these proteins to specific chromatin regions (13). Similarly, because lysine acetylation reduces the total change on histone tails, multiple acetylations can work synergistically to reduce the folding of chromatin into 30 nm fibers and reduce its stability (14,15). Because of our interest in how different combinations of lysine acetylation can modulate the structure and function of chromatin, we were interested in exploring to what extent this strategy could be extended to incorporating multiple sites of acetylation within a single H3 histone.

Materials and methods

Plasmid production

Xenopus laevis H3 histone in a pET3c expression plasmid (16) was mutated to incorporate 1-4 amber codons by sequential application of Quikchange mutagenesis (Strategene). The mutated H3 histone ORFs were then cloned into a pCDF-PylT-H3K14amb plasmid to replace the single amber codon containing H3 histone (11).

Expression of acetylated histone H3

Expression methods were adapted from those previously published (11). BL21 DE3 cells were co-transformed with the pAcKRS-3 plasmid containing the tRNA synthetase and the pCDF-PylT plasmid containing both the H3 histone with the desired number of amber codons and the amber suppressor tRNA. Transformed cells were grown overnight at 37°C in 50 mL standard Luria Broth media under selection of 50 μg/ml kanamycin and 50 μg/ml spectinomycin. 250 mL of prewarmed non-standard 2xYT broth (2xYT-KS: 0.5% w/v sodium chloride, 1% w/v yeast extract, 2% w/v tryptone), containing 50 μg/ml kanamycin and 50 μg/ml spectinomycin, was inoculated with overnight culture to a final OD600 of 0.08 at 37°C. The dilute 2xYT-KS

culture was grown to 0.7 OD600. N-ε-Acetyl-L-lysine (≥98% pure, Novabiochem) was then added to a final concentration of 10, 20, 40, or 80 mM. Nicotinamide (NAM) was added to a final concentration of 20 mM. After 30 minutes, protein expression was induced with the addition of 0.5 mM IPTG. After 4 hours, 250 mL cultures of cells were pelleted and resuspended in 30-ml of Wash Buffer (WB: pH 7.5, 50 mM Tris-HCl, 100 mM NaCl, 1 mM Na-EDTA, 1 mM benzamidine, 1 mM DTT, 20 mM NAM) then frozen at -80°C overnight.

Purification of acetylated histone H3

Inclusion bodies from 250 ml of culture were isolated as previously described (17). The insoluble final pellet containing histone was macerated with 0.25 mL DMSO and then thoroughly resuspended in 15 mL unfolding buffer (UB: 6M guanidine hydrochloride, 20 mM Tris hydrochloride, 5 mM β mercaptoethanol, pH 8.0) and stirred at room temp for 1 hour. The mixture was clarified via centrifugation (12,000 RCF, 10'). The supernatant was added to 2.5 mL of pre-equilibrated and drained Ni^{2+}-NTA Bead Resin (Qiagen) in a 30-ml disposable column and stirred at room temperature for 1 hour. The column was drained and the resin washed two times with 30-ml Guanidine Wash Buffer (GWB; 6M guanidine hydrochloride, 100 mM monosodium phosphate, 5 mM β-mercaptoethanol, pH 6.3). The resin was then washed with 100 mL 1X TEV Cleavage buffer (50 mM Tris hydrochloride, 50 mM imidazole, 0.5 mM EDTA, 5 mM β mercaptoethanol, pH 8.0). The resin was resuspended in 23 mL 1X TEV cleavage buffer and 10,000 units of TURBOTEV protease (Nacalai USA) were added. Digestion was carried out with nutation for 16 hours at room temperature. The column was then drained and resuspended in 15 mL UB and stirred at RT for 1 hour. UB eluent containing histones was dialyzed into SAU buffer (7M urea, 20 mM sodium acetate, 1 mM DTT, 1 mM sodium EDTA, pH 5.2). The histone sample was then purified via cation exchange on a salt gradient from 0 mM to 600 mM NaCl

over 225 mL using a HiTrap SP HP 5 mL column (GE Life Sciences). Histones eluted in a highly pure state and were concentrated and desalted using Sep-Pak C-8 reverse phase purification columns (Waters) before lyophylization. Purified acetylated histone H3 was checked via MALDI-TOF analysis to verify the correct number of acetylations were present (John Leszyk at University of Massachusetts Medical School). Histone was quantified using densitometry against standard curves of H3-tetra-alanine containing histone.

Assembly of mononucleosomes

Wild type histones and the tetra-alanine H3 histone were expressed, purified, and quantified according to literature protocols (17). 177 bp DNA templates containing the 601 strong positioning sequence (18) and BglI non-palindromic sticky ends were prepared as previously described (19). Octamers were assembled from the appropriate recombinant histones and purified via size exclusion chromatography as published previously (17). Mononucleosomes were reconstituted from the DNA template and histone octamer via rapid dilution methods (20), and then were characterized via 4% native PAGE Gel with staining for the DNA. The assembled nucleosomes were dialyzed into native buffer (2.5 mM NaCl, 10 mM Tris, 0.25 mM EDTA, pH 7.4) and concentrated 7-10 fold by volume using 30 kDa MWCO concentrators (EMD Milipore). Mononucleosomes were immobilized onto beads largely as previously described (19). Briefly, 4.66 pmol of the biotinylated adapter were bound to 200 □g of paramagnetic streptavidin beads (NEB) at RT for 20 minutes in 1X binding buffer (10 mM Tris-HCl, pH 7m.4, 100 mM NaCl) to get a final reaction volume of 100 □l. The supernatant was removed and beads were washed three times with 50 □l of 1X ligation buffer (50 mM Tris, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT; pH 7.5). 6.99 pmol of WT, TA or TAc nucleosomes were added to the beads in 1X ligation buffer, in a final reaction volume of 100 □l. 800 U of T4 DNA ligase (NEB) were used to ligate

the nucleosomes to the adapter. The reactions were carried out at RT for 5-6 hours. Excess unbound nucleosomes were then washed off. The beads were resuspended in 31 □l of 1X ligation buffer to get a final nucleosome concentration of 150 nM. The integrity of the assembled substrates was analyzed by digesting the beads with PstI at 37°C, for 5 hours. The liberated nucleosomes were analyzed on 4% native PAGE.

Histone acetylation assay

To compare the extent of acetylation of WT, TA and TAc nucleosomes, substrates were subjected to standard acetylation assays by ySAGA (Spt-Ada-Gcn5-acetyltransferase complex from Saccharomyces cerevisiae), under initial rate, sub-saturating nucleosome concentrations, as previously described (19). Briefly, 1.5 nM ySAGA was used to acetylate 10 nM of each kind of nucleosome in a 50 □l reaction volume of 1X HAT buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% glycerol, 1 mM DTT, 1 mM PMSF, 10 mM sodium butyrate, 4.0 μM acetyl CoA with a specific activity of 5.78 Ci/mmol (Moravek Biochemicals)). The acetylation reaction was carried out at 30°C for 10 min. The beads were washed 4 times at RT with 50 □l 1X wash buffer 1 (WB1- 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1% triton X-100) and for additional 5 times at 37°C with 15 □l 1X WB2 (50 mM NH2OH , pH 7.5, 0.1% triton X-100), with 25 minutes incubation between each wash. The beads were then resuspended in 15 □l of 1X WB1 and added directly to 6 ml of scintillation cocktail. The samples were counted in Tri carb 4910 TR Liquid Scintillation Analyzer (Perkin Elmer). To account for non-H3 nucleosomal acetylation, the counts obtained from the WT and TAc nucleosomes were subtracted from those obtained from TA nucleosomes. The resultant extent of acetylation obtained from TAc substrate was compared to that of WT, which was normalized to 100%.

Results

We have been interested in how histone H3 tail acetylation is established by the SAGA family of transcriptional co-activitors (9,19,21), and how this acetylation affects chromatin structure and function, such as subsequent histone acetylation (21). The SAGA complex from budding yeast has been shown to target four major residues on the H3 tail, with lysine 14 being the most prominent acetylation site, followed by lysine 9, lysine 18, and then lysine 23 (Figure 8) (22). Nonsense suppression has been successfully employed to generate H3 histones containing lysine 14 acetylation (11), and we sought to determine if such a strategy could be applied to generating histones containing as many as all four primary acetylations. Histone proteins can be recombinantly expressed in E. coli (16), and H3 K14 acetylated histone was previously generated from an overexpression plasmid containing an amber codon at amino acid residue 14 (Figure 8) (11). Building off this sequence, we introduced additional putative acetyl-lysine residues by using site directed mutagenesis to convert lysine codons to amber stop codons. Plasmids were constructed to contain up to four amber codons at the major sites of SAGA acetylation, where histones with intermediate levels of acetylation were made to match the known preference of lysine acetylation - i.e. H3 histone containing one acetylation contained H3 K14 acetylation; K14 and K9 acetylation for two acetylations; and K14, K9, and K18 acetylation for three acetylation sites.

To test the efficiency of acetyl-lysine incorporation, we first over-expressed singly acetylated H3 K14 acetylated histone according to literature protocols (11). In this protocol, H3 protein expression is induced by IPTG addition in E. coli BL21 DE3 cells. However, for suppression of the amber stop codons, the cells also must contain a plasmid that constitutively expresses a tRNA containing an amber anticodon and a tRNA synthetase that can charge the

tRNA with acetyl-lysine. Acetyl-lysine is added to the media, as well as NAD, which inhibits acetyl-lysine deacetylation. As expected, we observed visible amounts of H3 histone in whole cell extracts (data not shown). In the inclusion bodies, where overexpressed histones localize, the full-length histone was especially prevalent (Figure 9A). However, when the same conditions

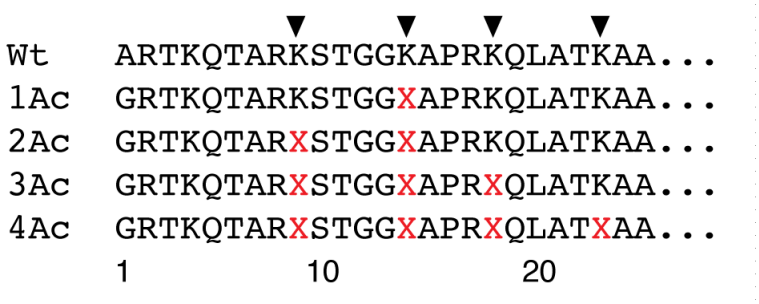

Figure 8 Sequences of acetylated histone H3 proteins. Top row shows the sequence of the N-terminal tail of histone H3. Triangles point to the major sites of SAGA-mediated lysine acetylation. Subsequent rows show the acetylated histones generated, where 'X' denotes a site of amber codon replacement in the histone expression vector and acetyl-lysine residues in the expressed protein. The C-terminal portion of the histone sequence is not shown. The N-terminal portion of the histone reflects the sequence present after TEV cleavage of the 6-His affinity tag: HHHHHHSQDPENLYFQG, with TEV cleavage between the last two residues, leaving the first H3 residue glycine.

were used for overexpression of the tetra-acetylated H3 histone, no H3 histone was observed in the inclusion body (Figure 9A). Because full length expression of the tetra-acetylated histone requires efficient suppression of all four non-sense codon, and suppression involves a number of components, lack of expression could be due to a number of factors. During optimization, we found that increasing the concentration of acetyl-lysine in the cell media dramatically increased the amount of full-length histone expressed (Figure 9B). Further, expression of histones with fewer than four acetyl-lysines was also feasible, with greater expression occurring when fewer amber codons are suppressed (Figure 9C).

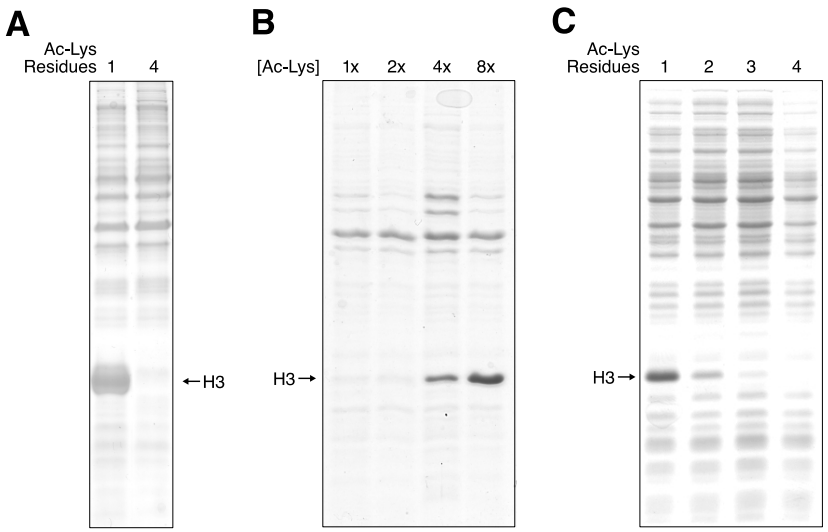

Figure 9. Expression of acetylated histone H3 proteins. Proteins were resolved on an 18% SDS-PAGE gel and stained with Coomassie Blue. **A)** Comparison of amounts of full-length mono- and tetra-acetylated histone protein in inclusion bodies under published nonsense suppression conditions. **B)** Comparison of amounts of full-length tetra-acetylated histone protein in inclusion bodies with increasing concentrations of acetyl-lysine in the media. 1X represents previously utilized acetyl-lysine concentrations of 10 mM. **C)** Comparison of expression of full-length histone containing one to four acetyl-lysines in whole cell extracts using 40 mM acetyl-lysine concentrations.

With improved tetra-acetylated H3 histone expression, we set out to purify the histone. Standard purification protocols for histones overexpressed in E. coli involve isolating inclusion bodies and then further purifying denatured histone by size exclusion and cation exchange chromatography. With these steps, we were able to increase the purity of the H3 histone, but not enough to obtain pure protein (data not shown). In the literature protocol for the singly acetylated histone (11), a TEV-cleavable, six-His affinity tag was fused to the N-terminus to facilitate purification, and a similar tag was investigated for the multiply acetylated histones. For the tagged, tetra-acetylated histone, induction and inclusion body purification proceeded to enrich the amount of full-length H3 histone (Figure 10A, lanes 1-3). Histone could then be unfolded and captured on nickel-NTA beads under denaturing conditions to further enrich purity (Figure 10A, lane 4). After solvent exchange into a non-denaturing buffer compatible with TEV

protease, H3 histone was liberated from the nickel beads (Figure 10A, lane 5). We found that to get complete cleavage required optimization of the both the binding and cleavage steps. For histone binding, we utilized just enough nickel-NTA beads to get complete histone binding, because additional beads decreased cleavage efficiency. Because commercially available TEV protease is also His-tagged, we believed that the nickel-NTA beads could be sequestering away TEV protease. Consistent with this idea, we found that addition of up to 50 mM imidazole improved the cleavage efficiency of the immobilized histone (higher concentrations of imidazole resulted in elution of uncleaved histone). The amount of TEV protease was also optimized to minimize the amount of protease required to give complete cleavage of the tetra-acetylated histone. Despite full cleavage, the histone could only be eluted from the beads under denaturing conditions, where a significant amount of contaminating proteins coelute. A cation exchange chromatography step under denaturing conditions resulted in histone protein sufficiently pure for subsequent applications. (Figure 10A, lane 6). Because we were working with relatively small amounts of protein, we found that reversed phase C8 Sep-Pak purification was an efficient way to concentrate and desalt the purified tetra-acetylated histone after cation exchange chromatography.

While the purification step generated full-length H3 histone, additional characterization was necessary to confirm the acetylation state of the histone. Full-length H3 protein lacking tetra-acetylation could result from amber suppression by non-acetyl-lysine codons, or by enzymatic deacetylation of the acetyl-lysine residues. To rule out these possibilities, the mass of the purified histone was determined by MALDI mass spectrometry (Figure 10B), and this analysis confirmed that the isolated protein was tetra-acetylated (15,425 Da expected, 15,428 Da observed, 42 Da per acetyl group). The ultimate yield of the tetra-acetylated H3 histone was 0.06

mg/g of cells. While this is significantly less than the 0.82 – 2.45 mg/g of cells that is typically obtainable for unacetylated H3, it is not significantly worse than the 0.17 mg/g of cells that was obtained from original single-site nonsense suppression studies (11). A similar purification protocol was applied to each of the other acetylated histones, and yielded 2.45 mg/g of cells, 0.65 mg/g of cells, and 0.24 mg/g of cells for H3 histones containing one, two and three acetylated lysines, respectively.

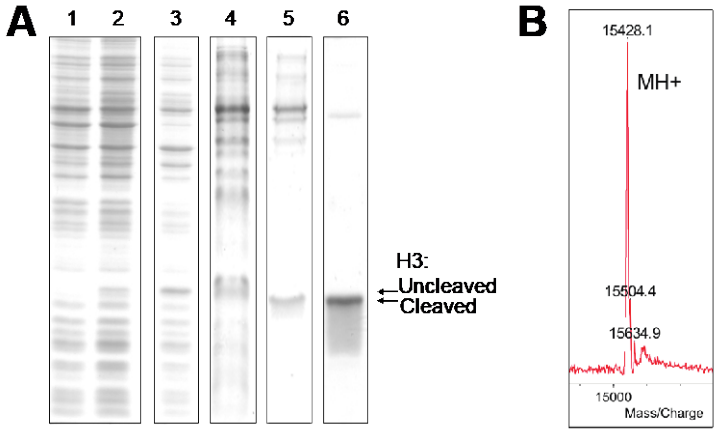

Figure 10 A) Analysis of tetra-acetylated H3 histone purity at various steps of the modified purification protocol. Proteins were resolved on an 18% SDS-PAGE gel and stained with Coomassie Blue. Pre- and post-induction of whole cell extracts are shown in lanes 1 and 2, respectively. Lane 3 shows protein composition following inclusion body purification. Lanes 4 and 5 show proteins bound to nickel-NTA-bead and then released by TEV protease, respectively. Lane 6 shows the protein composition following cation exchange chromatography. **B)** MALDI-TOF mass spectrometry analysis of purified tetra-acetylated H3 protein.

The purity and amount of tetra-acetylated H3 was sufficient to incorporate into a histone octamer and then a mononucleosome. To generate histone octamer, the tetra-acetylated H3 was combined with recombinant H2A, H2B, and H4 histones under denaturing conditions and dialyzed into a high salt (2M NaCl) solution (16). Histone octamer was resolved from incomplete assembly products by size exclusion chromatography, and eluted identically to wild-type octamers. Denaturing protein gel electrophoresis of the wild type, tetra-acetylated H3, and

tetra-alanine H3 octamers look similar (Figure 11A), with some changes in electrophoretic mobility of acetylated and alanine-containing H3 histone relative to the wild-type H3. Equal amounts of octamers were then incorporated into mononucleosomes by deposition onto a 177 base pair 601 double stranded DNA template by rapid dilution (20), and all three mononucleosomes appear largely similar in their extent of assembly and their electrophoretic mobility (Figure 11B). Each mononucleosome was immobilized onto a paramagnetic bead, and

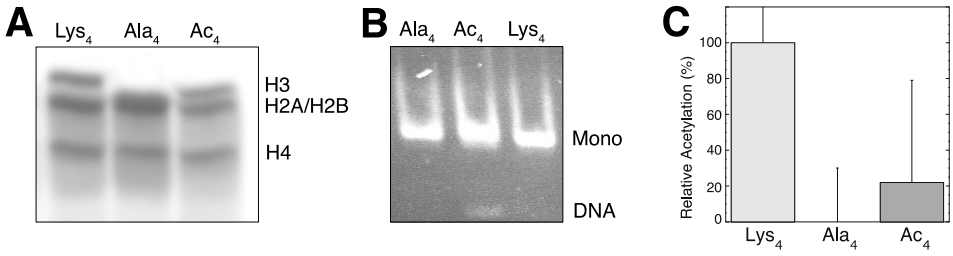

Figure 11. Utilization of tetra-acetylated histone H3 protein. A) Denaturing protein gel analysis of histone octamers containing unacetylated, tetra-alanine, or tetra-acetylated H3 histones. Histones were resolved on an 18% SDS PAGE gel with Coomassie Blue staining. **B)** Native gel analysis of mononucleosomes containing tetra-alanine, tetra-acetylated, or unacetylated H3 histone. Species were resolved on a 4% native PAGE gel with ethidium bromide staining. **C)** Comparative extent of SAGA-mediated mononucleosome acetylation relative to mononucleosomes in which the major H3 tail acetylation sites were mutated to alanine. To determine the extent of acetylation, the amount of radioactive acetyl incorporation was subtracted from that of the tetra-alanine mononucleosome and then normalized to the wild-type mononucleosome. Data represents four independent trials.

then used as a substrate for a SAGA-mediated acetyltransferase assay (Figure 11C). As expected, the nucleosome containing acetyl-lysine at the major SAGA acetylation sites showed a significant reduction of new acetylation relative to a nucleosome that was not pre-acetylated. The reduction in the amount of acetylation was similar to that observed for a nucleosome in which the four major H3 tail acetylation sites were mutated to alanine, indicating that the tetra-acetylated histone prevents SAGA-mediated histone acetylation in a manner similar to loss of the lysine residue.

Discussion

We have shown that nonsense suppression can be used to incorporate multiple acetyl-lysine residues into the tail of H3 histone protein. One major hurdle we encountered was expressing full-length protein. This difficulty presumably arises because full-length expression requires efficient nonsense suppression at every amber codon. Consistent with this idea, we observed that the amount of full-length protein decreases with increasing numbers of amber mutations to suppress (Figure 9C). Thus, conditions to optimize nonsense suppression were necessary. We found that increasing the amount of acetyl-lysine present in the media increases full-length protein expression (Figure 9B), suggesting that a key difficulty in nonsense suppression in this system is charging the nonsense suppression tRNA with acetyl-lysine. It is likely that acetyl-lysine is limiting in the cell and by increasing its concentration in the media, intracellular concentrations increase to drive the action of the synthetase. For our purposes, a four-fold increase in acetyl-lysine concentrations was sufficient for generating amounts of tetra-acetylated histone we needed. However, for the increasing acetyl-lysine concentrations used (Figure 9B), we did not observe saturation of histone expression, suggesting that even greater levels of expression should be possible. In this case, the only trade-off may ultimately be the cost of acetyl-lysine. Further, because nonsense suppression is a complicated process, other steps in the reaction might also benefit from optimization. This could include improving the expression levels and properties of the tRNA and synthetase, as well as optimizing the length and timing of histone induction relative to acetyl-lysine addition.

The other major hurdle encountered was in the purification of the multiply acetylated histones. Inclusion body preparations, nickel-NTA beads binding, and cation exchange chromatography were highly efficient, while proteolytic cleavage of the histone from the nickel-

NTA required significant optimization in the bead, imidazole, and TEV protease amounts. Even under these optimized conditions, a significant amount of TEV protease was required. The inefficiency of TEV cleavage could be due in part to sequestration of the protease to the nickel-NTA beads, and alternative forms of TEV-protease that do not contain a His-tag might improve cleavage efficiency. Another potential contributor could be the inability of H3 histone to form a well-behaved globular structure under native conditions. In such a case, it is likely that affinity captured histones form aggregates on the surface of the nickel-NTA beads when solvent is exchanged for proteolysis, making the cleavage site less accessible. A cleavage strategy that can be performed under denaturing conditions would likely avoid this issue.

In our studies, we focus on incorporating multiple acetyl-lysine residues into the H3 histone at known sites of SAGA-mediated acetylation. However, we feel our strategies should be applicable to other sites within the H3 histone, to other acetylation sites in other histones, and even to acetylation of other proteins. Within the H3 histone, Chin and coworkers have previously shown that many different sites within the H3 histone can be individually acetylated, suggesting that, in general, nonsense suppression is not highly sequence dependent in the H3 histone. This idea is further supported by the fact that we observe a similar degree in decrease in histone expression with every amber codon added. Thus, we expect that other combination of multiple lysine acetylations in H3 histone should be possible. With respect to other histones, multiple acetylations may also be possible. For example, while the H4 histone had proven recalcitrant to single site acetyl-lysine nonsense suppression, recent studies have shown that with codon optimization, H4 K16 acetylated histones can be generated at level comparable to those found for single sites acetylation H3 histones in the original Chin study. Our improvements in H3 expression and purification may be directly applicable to H4 histones, as well as others. Finally,

for non-histone proteins, especially those that can be purified more effectively than histone, increasing acetyl-lysine concentrations might prove sufficient to allow for improved incorporation of single or multiple acetyl-lysine residues.

We have shown that we can extend a nonsense suppression strategy for incorporating acetyl-lysine into H3 histones to generate histones containing up to four acetyl-lysine residues. Key to achieving this is improved histone expression via an increase in the amount of acetyl-lysine added to the cell media, and a combined affinity capture and ion exchange chromatography purification strategy. The tetra-acetylated H3 histone is generated in sufficient yield and purity to be incorporated into histone octamers and nucleosomes, and we expect that the insight gained from our study could aid in utilizing nonsense suppression to incorporate acetyl-lysine residues into other histones and non-histone proteins.

Acknowledgements and Funding: This work was supported by American Cancer Society Research Scholar Grant 1206501 to M. A. S.-K.

Bibliography

1. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **325**, 834-840

2. Xiong, Y., and Guan, K.-L. (2012) Mechanistic insights into the regulation of metabolic enzymes by acetylation. *The Journal of Cell Biology* **198**, 155-164

3. Perdiz, D., Mackeh, R., Poüs, C., and Baillet, A. (2011) The ins and outs of tubulin acetylation: More than just a post-translational modification? *Cellular Signalling* **23**, 763-771

4. Mowen, K. A., and David, M. (2014) Unconventional post-translational modifications in immunological signaling. *Nat Immunol* **15**, 512-520

5. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal Structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

6. Zhang, L., Eugeni, E. E., Parthun, M. R., and Freitas, M. A. (2003) Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* **112**, 77-86

7. Robinson, P. J. J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R. G., and Rhodes, D. (2008) 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. *Journal of Molecular Biology* **381**, 816-825

8. Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **311**, 844-847

9. Li, S., and Shogren-Knaak, M. A. (2008) Cross-talk between histone H3 tails produces cooperative nucleosome acetylation. *Proceedings of the National Academy of Sciences* **105**, 18243-18248

10. Young, T. S., and Schultz, P. G. (2010) Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry* **285**, 11039-11044

11. Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J. W. (2009) A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Molecular Cell* **36**, 153-163

12. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A., and Garcia, B. A. (2009) High Throughput Characterization of Combinatorial Histone Codes. *Molecular & Cellular Proteomics* **8**, 2266-2284

13. Ladurner, A. G., Inouye, C., Jain, R., and Tjian, R. (2003) Bromodomains Mediate an Acetyl-Histone Encoded Antisilencing Function at Heterochromatin Boundaries. *Molecular Cell* **11**, 365-376

14. Tse, C., Sera, T., Wolffe, A. P., and Hansen, J. C. (1998) Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III. *Molecular and Cellular Biology* **18**, 4629-4638

15. Gansen, A., Tóth, K., Schwarz, N., and Langowski, J. (2009) Structural Variability of Nucleosomes Detected by Single-Pair Förster Resonance Energy Transfer: Histone Acetylation, Sequence Variation, and Salt Effects. *The Journal of Physical Chemistry B* **113**, 2604-2613

16. Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999) Expression and Purification of Recombinant Histones and Nucleosome Reconstitution. in *Chromatin Protocols* (Becker, P. B. ed.), Humana Press, Totowa, NJ. pp 1-16

17. Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999) Preparation of nucleosome core particle from recombinant histones. in *Methods in Enzymology*, Academic Press. pp 3-19

18. Lowary, P. T., and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning1. *Journal of Molecular Biology* **276**, 19-42

19. Mittal, C., Blacketer, M. J., and Shogren-Knaak, M. A. (2014) Nucleosome acetylation sequencing to study the establishment of chromatin acetylation. *Analytical Biochemistry* **457**, 51-58

20. Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M., and Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *Journal of Molecular Biology* **267**, 807-817

21. Li, S., and Shogren-Knaak, M. A. (2009) The Gcn5 Bromodomain of the SAGA Complex Facilitates Cooperative and Cross-tail Acetylation of Nucleosomes. *Journal of Biological Chemistry* **284**, 9411-9417

22. Grant, P. A., Eberharter, A., John, S., Cook, R. G., Turner, B. M., and Workman, J. L. (1999) Expanded Lysine Acetylation Specificity of Gcn5 in Native Complexes. *Journal of Biological Chemistry* **274**, 5895-5900

CHAPTER 3. KINETICS AND PROPERTIES OF NUCLEOSOMAL ARRAY SELF-ASSOCIATION IN VARIABLE PH, CONCENTRATION, AND LENGTH

Isaac A Young[1,2], Xun Wu[2,3], Wenyu Huang[2,3], Michael Shogren-Knaak[2,4]

[1]Primary Author and contributed all research except [3], [2]Iowa State University affiliation [3]Ames Laboratory affiliation and contributed ICP-MS data, [4]Principle investigator

Modified from a manuscript to be submitted to *Journal of Biological Chemistry*

Abstract

Nuclear chromatin adopts a random polymer conformation that self-associates into a hierarch of structures, including topological associated domains (TADs) and chromatin compartments. Recent work with nucleosomal arrays has illuminated a number of aspects of chromatin self-association, including their physical phase. In this paper, we use nucleosome arrays to probe the structure and kinetics of chromatin self-association, and the factors that contribute to them. We find that array self-association is independent of array length and moderately sensitive to array concentration. We see that arrays are stable over a relatively wide range of pH values from pH 5-11. Over the entire range, reduction of pH increases the sensitivity of the arrays to magnesium-induced self-association. Analysis by inductively coupled plasma mass spectrometry analysis shows that arrays sequester enough magnesium ions to neutralize their negative charge plus additional magnesium. The excess magnesium ions suggest the array particles also sequesters a near equivalent volume of solvent - a situation that could provide greater access to the interior of the array particles or packed array particles. We also characterize self-associated array by dynamic light scattering, showing that this technique is a sensitive method for measuring array particles size and growth rate. We observe that there are at least two kinetic phases in array particle formation with a distinct structural intermediate. This behavior is

consistent with a nucleation model of array particle growth and the size of the species involved correspond to the observed sized of TADs and chromatin compartments.

Introduction

Chromatin is the native form of DNA found in the nucleus of all eucaryotic cells, and it serves to both compact and organize the genome (1,2). What the structure and dynamics of chromatin are has long been, and continues to be, a central question in biology. The fundamental structural unit of chromatin is the nucleosome, composed of an octameric core complex of positively charged histones (H2A, H2B, H3, and H4) around which 147bp of DNA is wrapped (3). Protruding from the core are positively charged histone tails, which are unstructured N- and C-terminal peptides that can interact with the negatively charged DNA (4) and portions of the core histone (5,6). Individual nucleosomes are joined by variable lengths of linker DNA resulting in $10^5 – 10^6$ nucleosomes per chromosome in humans (7). The structure and frequency of nucleosomes simultaneously compact the DNA contour length by ~5-8-fold and act as a counterion to DNA(3,8), reducing the energetic cost of keeping the DNA in the nucleus.

Beyond the level of the nucleosomes, chromatin adopts a hierarchy of structures. It was long believed that the first level of this hierarchy was the formation of a 30 nm fiber, followed by other well-ordered structures. While at low concentrations of chromatin, 30 nm fibers are observed *in vitro* (9-12), recent studies have indicated that the predominant form of chromatin *in vivo* is a molten globule (13,14), in which chromatin adopts a random polymer conformation. Despite not being as uniformly organized as originally thought, chromatin does organize into a hierarchy of structures, from topologically associated domains (TADs), chromosome compartments, and chromosomal territories (15).

The use of chromatin model systems consisting of multiple tandem nucleosomes, nucleosomal arrays (16,17), has proven invaluable in probing the structure and properties of chromatin, traditionally with regards to characterizing 30 nm fiber formation. These chromatin model systems have also been shown to undergo reversible self-association (18). This structural transition can be induced with concentrations of divalent magnesium cations present in the nucleus during interphase, and which rise during entry into mitosis (19,20). Recent studies have shown that self-associated nucleosomal arrays chromatin model systems closely reflects the state of nuclear chromatin, and its physical properties (21,22).

However, despite the utility of nucleosomal arrays for understanding in vivo chromatin structure, the factors, structure, and dynamics of chromatin remains poorly understood. To address this, we were interested in defining some of the parameters of nucleosomal array self-association, including: the contribution of array length, array concentration, and pH; magnesium binding stoichiometry; and self-association kinetics.

Materials and methods

Preparation of array template and carrier DNA

174bp Carrier DNA was prepared by PCR amplification and purification as described(6). Plasmid containing 601-177-12 template was purified from E-coli using qiagen gigaprep kits and fragment excised using EcoRV as previously described. (Richmond JMB 2003) 601-177-24, 36, and 48 templates were cut out of E-coli plasmid as previously described using PstI/BamHI dual digestion (23). All excised template DNA fragments were purified using preparative agarose gel electrophoresis (16). The band containing template was extracted using electro-elution. Ethidium bromide was removed, and DNA concentrated via three rounds of butanol extraction

concentration (24) followed by ethanol precipitation. DNA was resuspended in water and insoluble agarose was removed via 2 cycles of freeze/thaw/centrifugation.

Preparation of octamer

Recombinant Xenopus laevis histones were expressed and purified from E-coli and histone octamers were refolded and purified similarly to literature methods (25). 40&50 micromolar histone octamer in refolding buffer (2M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM Na-EDTA, and 1 mM freshly added DTT) was then flash frozen in luquid nitrogen without inclusion of any protease inhibitors as 30 or 300 microgram aliquots and stored at -80°C as done previously by others(21). AUC-SV assays, ScaI digestions, and self-association assays were performed to verify that arrays prepared using frozen octamer were similar to those prepared using non-frozen octamer.

Reconstitution of 601-177-12 arrays

As needed, aliquots of octamer were thawed on ice and used to assemble arrays following generally standard stepwise salt dialysis methods. (6,17,23,26,27) Specifically our method used here is described for clarity. Octamer, 601-177-12 template, and Carrier DNA were combined in a octamer:nucleosome positioning site:carrier ratio of 1.2:1:0.3 in 2M NaCl. Our step-wise salt dialysis buffer system entails a constant 10 mM Tris pH 8.0 and 0.25 mM EDTA and variable NaCl concentration. Dialysis was carried out for 3 hours at four degrees per step going from 1.4M, 1.2M, 0.8M, and 0.6M NaCl. Finally, the remaining NaCl was removed via dialysis in LDB: 2.5 mM NaCl, 10 mM HEPES pH 8.5, 0.1 mM TCEP. Array samples were purified via addition of 4 mM magnesium and sedimentation to remove carrier mononucleosomes and carrier free DNA which remain in the supernatant. Arrays were resuspended in LDB to ~200 ng/µL

[DNA]. Array samples were kept at 4 deg C for no longer than 50 days prior to use or being discarded.

The reconstitution of 601-177 arrays of variable nucleosome repeat length proceeded similarly to above. During pilot assays, we found that optimal ratios of octamer relative to positioning sites for 24, 36, and 48mer were 1.15, 1.05, and 0.95 respectively for ideal saturation. As array length increased, we found it increasingly difficult to remove excess carrier nucleosomes using the magnesium centrifugation sedimentation technique. However, we also found that longer arrays required a smaller octamer ratio to achieve saturation as indicated by ScaI digestion possibly due to differences in quantitation.

Quantitation of array

Arrays were quantified using uv-vis absorbance spectroscopy measured at wavelengths of 260nm, 280nm, and 320nm. 320nm absorbance was subtracted from 260nm and Beer–Lambert law was applied using the typical DNA mass extinction coefficient (260nm) of 20 $L*(g*cm)^{-1}$ to find the concentration of array. Absorbance at 280nm was baselined against 320nm as well and a 260/280 ratio of 1.8 was observed indicating no significant protein contamination. When conversion to molarity of array was used in calculations, the molecular weight of array was assumed to be 1.3 megadaltons which is the weight of the DNA alone (613.5 $g*mol^{-1}*bp^{-1}$).

High concentration array sample treatment

For concentration variable experiments, two different batches of array were used. 252&264 micrograms of 629&670 ng/μL array was concentrated using 0.5 ml capacity 3,000 MWCO centrifuge concentrators. A final concentration of 1.251&2.757 mg/ml was achieved at a recovery rate of 86% both times. Prior to quantitation, in the 2.757 mg/ml array solution, a small

amount of whitish precipitate was observed and removed via centrifugation. ScaI digestion assays were performed to verify stability of arrays at higher concentrations. Samples were diluted to target concentrations for self-association assays.

Variable pH array sample treatment

For pH variable experiments, an iso-ionic strength variable pH buffer system was formulated based on (28). An aqueous 100x "MTA" stock composed of 420 mM MES (2-(N-morpholino)ethanesulfonic acid), 420 mM Tricine (N-(2-Hydroxy-1,1-bis(hydroxymethyl)ethyl)glycine), 840 mM Aminomethyl propanol (2-Amino-2-methylpropan-1-ol), and 250 mM NaCl was prepared. 1 mL of 0.1M TCEP, and 10 ml of 100x stock was mixed with 950 ml of water then either 1M NaOH or 1M HCl was added to achieve the desired target pH. Volumes of acid or base required to achieve target pH varied from ~1 mL to get to pH values close to the starting pH of 8.7 to ~13 ml to get to the furthest pH values of 5 or 11. The solution was finally brought to a total volume of 1 Liter. Array samples were dialyzed at room temperature for at least 3 hours but no longer than 12 hours. Dialyzed array samples could then be stored at 4 deg C until needed.

Validation of array stability under variable pH conditions

ScaI digestions were performed to verify saturation using ScaI-HF similarly to previously described with modifications(23). A 10x ScaI digestion buffer was prepared (500 mM NaCl, 100 mM Tris pH 7.4, 25 mM $MgCl_2$, and 0.5% Triton-X100) and stored at room temperature and used as needed. 40 ng of array was mixed in 1x ScaI digestion buffer with 20 U of ScaI-HF restriction endonuclease from NEB in a total volume of 20 ul. After 2 hours digestion at 37 deg C, 6x glycerol reaction stop dye (50% glycerol, ~0.1% bromophenol blue, 50 mM EDTA) was added.

Native PAGE method

Analysis of ScaI digests was performed based on adaptation of previously published methods by Gaykalova et al. which we found to offer especially fine detail in our lab (29). 10ng of array ScaI digest was run on native gels containing 4.5% acrylamide:bisacrylamide (40:1), 5% glycerol, 20 mM Na-HEPES pH8.0, 0.1 mM EDTA. Gel running buffer was 20 mM HEPES pH 8.0 and 0.1 mM EDTA. Gel parameters were 10 or 15 well combs in plates by biorad with dimensions 1.5mm thick, 6 cm tall by 8.3 cm wide. 10 volts/cm or 60 volts DC current were applied for 60 minutes and stained with sybr gold (Invitrogen).

Analytical ultracentrifugation sedimentation velocity (AUC-SV) experiments

AUC-SV was performed using Beckman XL-A analytical ultracentrifuge as previously described (3,30). Ultrascan III software was used to perform boundary analysis using the van Holde & Weischet technique. AUC-SV runs were performed as a quality control step and to examine their behavior in different pH buffers.

Array self-association assays

Self-Association assays were performed similarly to previously published with situation specific modifications (31). In standard assays, 40 µL of array in LDB at 20 ng/µL template DNA were mixed with an equal volume of freshly prepared 2x $MgCl_2$ in array buffer. After incubation for 15 minutes at room temperature, array oligomers were sedimented via centrifugation for 10 minutes at 14,000 x g. A differential sedimentation plot was generated by comparing the amount of array remaining in the supernatant with the supernatants treated identically in 0 mM $MgCl_2$ (determined via uv-vis spectroscopy by comparing relative A_{260}-A_{320} measurements). From the graph of fraction in supernatant vrs $[Mg^{2+}]$, an $[Mg^{2+}]_{50}$ was determined by tracing a line between the nearest points of magnesium concentration where % in

the supernatant was above and below 50% and recording the $[Mg^{2+}]_{50}$ as the point where the line crossed the 50% mark.

Length variable self-association assays were performed as above except 35 µL of array was used instead of 40.

Concentration variable self-association assays were performed as above with the following changes. Concentrated array samples were diluted with LDB to concentrations between 20 ng/µL and 2.757 mg/ml. To save material, sedimentations were performed in a minimum volume of 10 µL where 7 or 8 µL of array was mixed with 3 or 2 µL of $MgCl_2$ in LDB. After centrifugation, 4 µL of supernatant was carefully extracted and systematically diluted for measurement using uv-vis spectroscopy with the goal of seeing an A_{260} of 1.2 if 100% of the array remained in the supernatant.

Variable pH self-association assays: After dialysis, arrays were quantified to verify that no array was lost during dialysis. They were then diluted with pH specific last dialysis MTA buffer (pHLDMTA). 2x $MgCl_2$ buffers were prepared using the same buffer and assays were performed as above. Over multiple iterations, it became clear that there was a trend in $[Mg^{2+}]_{50}$ measurements with different pH and it became necessary to strategically adjust magnesium concentrations used in the assay accordingly depending on pH. Low pH array samples (pH range 5-7) were assayed at magnesium concentrations between 0.2 and 1.5 mM $MgCl_2$. High pH array samples (pH range 9-11) were assayed at magnesium concentrations between 1.7 mM and 4.7 mM $MgCl_2$. Arrays with pH between 7 and 9 were assayed at magnesium concentrations between 0.7 and 3.0 mM $MgCl_2$.

Inductively coupled plasma mass spectrometry (ICP-MS) quantification of magnesium in array samples

Array self-association was induced via addition of $MgCl_2$ buffer in a final composition of 5.5 or 3.3 micrograms of array (based on template DNA) at 10 ng/µL in LDB or pHLDMTA and 15 minutes of incubation at room temp. Oligomers were sedimented via centrifugation at 14,000xg for 10 minutes. Supernatant was carefully removed via pipetting and a secondary short centrifugation to remove the last 1-2 µL. Array samples were resuspended in 1 mL water, then organic matter was destroyed via addition of 3 ml of 2% Nitric Acid. Mock controls were prepared similarly where addition of array was omitted. When array was measured in the absence of any magnesium via dialysis in 0.1 mM EDTA, no magnesium ICP-MS signal was detected. Any outliers were discarded where over 10-fold magnesium was detected. These occurred seemingly randomly and are likely the result of failure to fully remove the magnesium chloride buffer. Samples were always kept in 13 mL conical bottom plastic centrifuge tubes as it was found that magnesium contamination was a problem when using glass vials. Experiments were conducted using a Thermo X-Series 2 ICP-MS machine.

Data in figure 14B and 14C was produced on two separate days and were normalized to repeated positive controls of 4 mM $MgCl_2$ pH 8.5 array measurements because stoichiometric signal was inconsistent on a day-by-day basis but were relatively consistent within one day. This was due to systematic error where during the first day, resuspended array was first quantified and then measured on ICP-MS. Figure 14A data, figure 14B data points of 2.7 mM, 4 mM, and 6.7 mM $MgCl_2$ and the single pH 10.5 data point of figure 14C was generated on the first day. On the second day, a known quantity of array was sedimented, but the resuspended array was not quantified. The remaining data in figure 14B and 14C were collected during the second day.

Typical recovery of array after magnesium sedimentation is usually fairly consistent, however this is an unknown variable and is the reason why we normalized the data.

The concentration of the master magnesium chloride stock used for all magnesium related experiments was quantified via ICP-MS using 1000 ppm Mg^{2+} in 2% nitric acid as a standard solution purchased from BDH company. It was found to be 1.23 molar (n=4), range of 1.19-1.30 molar, median of 1.21 molar, and standard deviation of 0.04 molar.

Dynamic light scattering measurement of array samples

Sample preparation: Water and buffers used were filtered through 0.02 micron filters to remove dust. Nucleosome array samples of ~200 ng/µL template DNA were dialyzed 3h at 4 deg C in LDB containing 0.1 mM EDTA followed by 3h of dialysis at 4 deg C in LDB without EDTA. As needed, arrays were then dialyzed into pH variable MTA buffers. They were then filtered through 0.45 micron PES syringe filters (4 mm diameter) from Tisch Scientific.

DLS Data collection: All work was done using a wyatt DynaPro NanoStar with low-volume quartz SN JC-343 cuvette. Array samples were diluted to a range of concentrations between 1.3 and 10 ng/µL template DNA and 40 µL loaded into the freshly washed (3x1ml filtered H2O) quartz cuvette. All measurements were conducted using 20-140x, 5-second autocorrelation curve acquisitions at 25 deg C, with auto-attenuation enabled, and only DLS data was collected. Longer measurements were performed (2h, 3h, and overnight), but the data was noisy and discarded in the results reported.

Size measurements of array samples with $MgCl_2$: After DLS data collection was performed on the array sample without magnesium, 1, 2, 3, or 4 µL of a 55 mM stock of $MgCl_2$ in LDB was added to the 40 µL of solution and gently mixed with a pipette in the cuvette so as to

not introduce any bubbles. The door to the NanoStar was closed gently and DLS data collection was initiated immediately. The time from first addition of $MgCl_2$ buffer was variable between 20 and 40 seconds. Size progression was monitored for as long as 2 hours and became noisier over time as the nanostar reached its limit of detection using the active auto-attenuation function. All experiments in Figure 16B were carried out at 10 ng/μL array by template. All experiments in Figure 16C were carried out at 2 mM $MgCl_2$.

Data analysis: Using Wyatt DYNAMICS 7 software (Wyatt Technologies), outlier autocorrelation curves with bad baselines were automatically excluded. Solution Parameters necessary for calculation of hydrodynamic radius were set in the software: solvent viscosity = 1.000cP at 20 deg C, solvent refractive index at 589nm & 20C of 1.333. Regularization and cumulant fit analysis methods were then automatically applied to extract hydrodynamic radius information for each acquisition. Analysis methods and equations are published in the DYNAMICS 7 user's guide available from Wyatt Technologies.

Regularization fit data analysis method in figure 12A: 80-second measurements with 5-second acquisitions were carried out and a regularization fit derived hydrodynamic radius histogram was extracted. For concentrations of 2.7 mM $MgCl_2$ and above, array samples were allowed to incubate for 10 minutes after adding magnesium prior to the 80-second measurement.

Cumulant fitting data analysis method in figures 12B-D: The hydrodynamic radius of each 5-second acquisition from a selected 80-second measurement period was averaged and standard deviations determined.

Rate of growth kinetics data analysis: The cumulant fit derived hydrodynamic radius from each of the first 140 acquisitions was plotted against time with the first acquisition was

plotted at 30 seconds as an approximate interval of time that elapsed between magnesium addition and start of growth.

Results

Exploring factors which affect self-association

As a model for chromatin, standard 601-177-12 nucleosomal arrays were used. These arrays consist of histone octamers deposited onto DNA sequences consisting of multiple head-to-tail repeats of a nucleosome positioning sequence and separated by linker DNA sequence. The 601nucleosome positioning sequence was chosen because it yields a homogeneous population of array which have been extensively characterized (11,17,32). The first parameter investigated for its effect on self-association was the number of nucleosomes in the array systems.

Nucleosomal arrays with twelve nucleosomes demonstrate reversible self-association with the addition of anywhere from 1-10 mM of divalent magnesium ion (18,27), where this range of magnesium concentrations is comparable to the influx of magnesium ion observed in the nucleus during mitosis (19,20). One way of characterizing this array self-association *in vitro* has been by differential sedimentation analysis. In this technique, the amount of magnesium ion is titrated and at each magnesium ion concentration the amount of array that has undergone sufficient self-association as to be isolated by centrifugation is determined. Previous studies have shown that the saturation of the arrays affects how sensitive the arrays are to divalent magnesium (33), and our own group has shown that 4-nucleosome arrays undergo self-association comparable to 12-nucleosome arrays. (31) In cells, the number of nucleosomes present in a stretch of DNA can be much greater than twelve, so we investigated how length affects array self-association. Arrays with 12, 24, 36, and 48 nucleosomes, but with the same mass concentration, were titrated with magnesium ion and characterized by differential sedimentation

(Figure 12A). Each array demonstrated comparable behavior, with most arrays remaining in solution at lower magnesium concentrations, and the majority of the arrays being sufficiently self-associated to be sedimented at higher magnesium ion concentrations. As has been previously observed, the transition between these states is not linear, but is cooperative, and a proxy for this transition, the threshold concentration of magnesium ion required for differential sedimentation ($[Mg^{2+}]_{50}$) can be measured. Comparison of these $[Mg^{2+}]_{50}$'s show that increases in array length from 12 to 48 nucleosomes do not change (Figure 12B), indicating that the presence of additional nucleosomes does not seem to affect the self-association process.

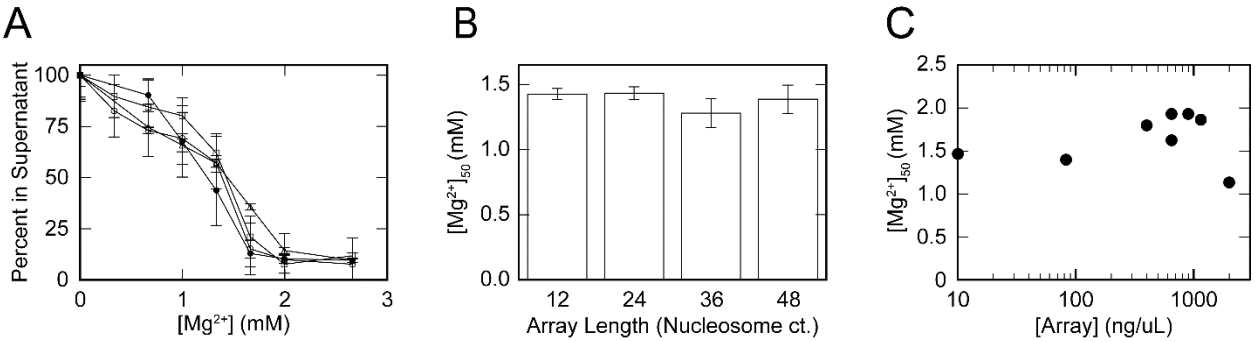

Figure 12. Effect of nucleosomal array length and concentration on self-association of nucleosome arrays. A) Mg^{2+} induced self-association assays of arrays at different nucleosome repeat lengths: Triangles= 12, Squares=24, Filled Circles=36, and Empty Circles=48. Each data point represents average of 3 repetitions. Assays performed with [array]=10 ng/uL, pH 8.5. **B)** Effect of nucleosomal array length on the threshold magnesium-ion concentration for array self-association. In this and part C, $[Mg^{2+}]_{50}$ represents the threshold $[Mg^{2+}]$ at which 50% of array has sedimented in the self-association assays seen in A. Error bars represent standard deviation of 3 replicates. **C)** Effect of nucleosomal array concentration on the $[Mg^{2+}]_{50}$ for array self-association assays performed similarly to A. Measurements performed at variable concentrations of array between 10 and 2000 ng/uL (n=9), pH 8.5.

Another factor that could affect array self-association is array concentration. The concentration of nucleosomes in the cell nucleus is considerably higher than concentrations typically used for *in vitro* studies of nucleosomes, and this concentration could change the propensity for arrays to self-associate. To address this possibility, we performed differential

sedimentation analysis over a range of concentrations of 12-nucleosome arrays (Figure 12C). These data suggest that there is a modest increase in magnesium ion required for increasing concentrations of array, suggesting that higher array concentrations discourage association and require more magnesium to overcome that. However, at the highest concentration of array the requirement is less, so this relationship seems complicated and requires more experiments to substantiate more fully.

Effect of pH on array stability and self-association

Another parameter of array self-association we were interested in investigating was the effect of pH. Although chromatin in the nucleus does not experience significant fluctuation of pH under normal circumstances, use of arrays in material applications very well could. Under these conditions their sensitivity to magnesium might change. In fact, such changes might even prove useful in allowing the arrays to sense their environment and transition from one structural state to another. At an even more fundamental level, changes in pH are likely to affect the stability of the arrays, and so we probed that first.

The stabilities of 12-nucleosome arrays were determined using ScaI digestion analysis at four pH values encompassing proton concentrations varying over eight orders of magnitude (Figure 13A). In this assay, arrays were prepared as normal and then dialyzed into a specific pH buffer. Arrays were then digested at ScaI sites between each nucleosome and resolved by native gel electrophoresis analysis. Arrays at the various pHs were also analyzed in the same manner after self-association, centrifugation, and resuspension of the self-associated array. It should be noted that for the pH studies, a three-component isoionic buffer was used (24). While array self-association has been shown to be more than an order of magnitude less sensitive to monovalent cations than divalent cation (18), we wanted to make sure the ionic strength of the buffer did not

change so that it was not a contributing factor to any pH differences we observed. The digestion analysis showed that at pH 7.5 and pH 10, the individual nucleosomes of the array remained intact. However, under the more basic pH of 12, the mononucleosomes appeared to show significant disassembly to hexasome and appeared not to resuspend at all after self-association. Similar behavior was seen under the more extreme acidic conditions, where the amount of observable mononucleosome and hexasome at pH 4.0 before self-association was additionally reduced.

Because ScaI digestion analysis involves returning the arrays to a more neutral pH and separation of the products by gel electrophoresis, we want to make sure that these factors were not influencing the array stability. To more directly characterize the arrays at various pHs, their sedimentation coefficient distribution was determined by sedimentation velocity analytical ultracentrifuguation (SV-AUC) (Figure 13B). In this analysis, arrays at pH 8.5 and 10.0 were observed to have a relatively uniform sedimentation coefficient distribution at an S value typically observed for these arrays (6). In contrast, at pH 4.0 the arrays had a broader and higher sedimentation value and at pH 12 they had a significantly lower sedimentation value. These differences from our standard arrays indicate that mass, shape, or both are significantly perturbed at the extreme pHs, and corroborate the ScaI digestion analysis results (Figure 13A). Thus, a more detailed pH analysis was performed at pH values within the range where array stability was not affected (Figure 13C). This analysis showed that arrays are stable over a wide range of pHs from pH 5.1 to pH 10.5.

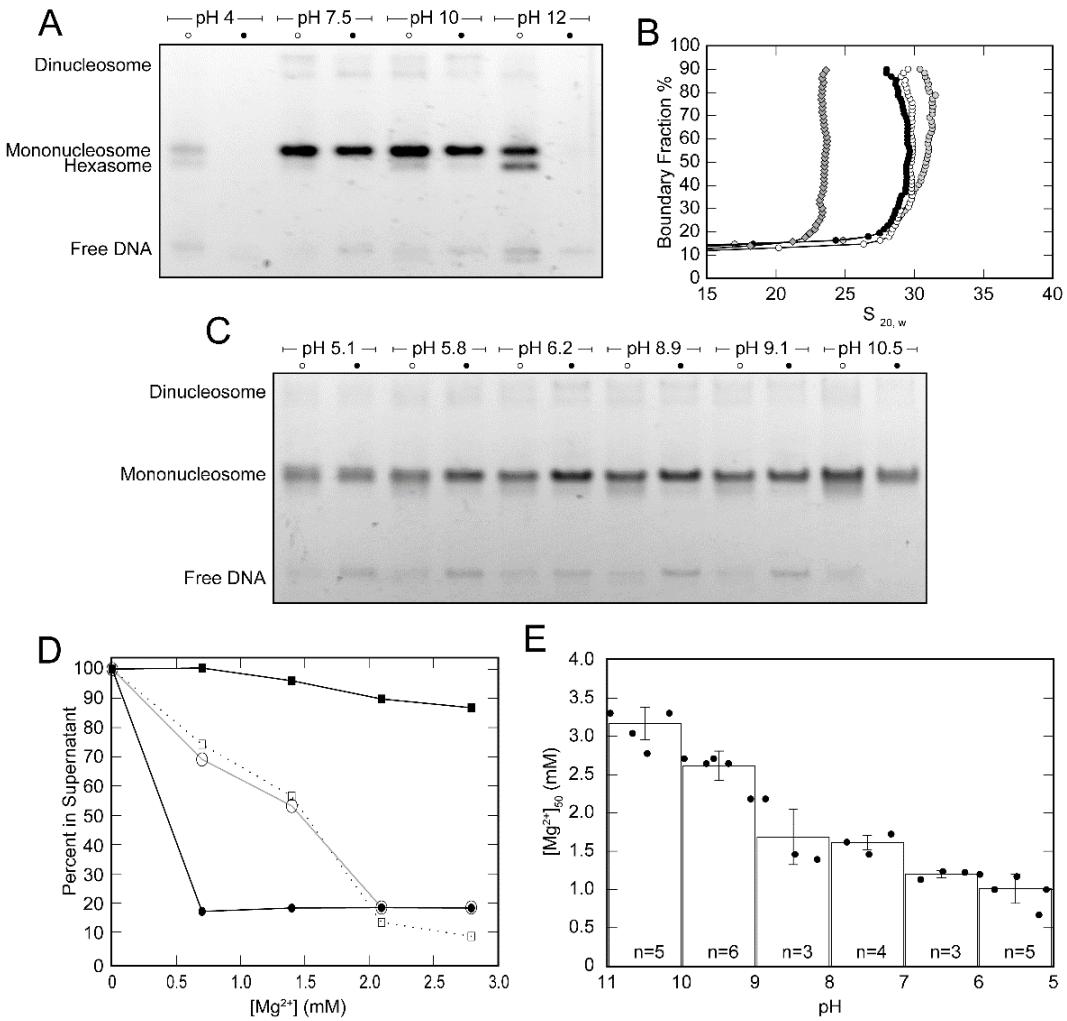

Figure 13. Effect of pH on stability and self-association of nucleosome arrays. A) Characterization of pH-dependent nucleosomal array stability by ScaI digestion. Arrays were either directly digested after being dialyzed into various iso-ionic strength buffers (open circles), or after dialysis, pelleting with 4 mM Mg^{2+}, and then resuspension (filled circles). Digestion products were resolved by native-PAGE analysis. **B)** Integrated sedimentation coefficient distribution of nucleosomal arrays at various pHs. Open circles are for the standard pH 8.5 buffer. The other arrays were characterized in iso-ionic strength buffer at pH 4.0 (gray circles), pH 10.0 (black circles), and pH 12.0 (gray diamonds) **C)** Extended characterization of pH-dependent nucleosome array stability by ScaI digestions. Conditions are as described in part A. **D)** Extent of array self-association as a function of magnesium ion concentration at selected pH values. Representative data of select pH arrays. |Open Squares: traditional buffer pH=8.5 | Open Circles: iso-ionic buffer pH=8.4 | | Filled Squares: pH=10.2 | Filled Circles: pH=5.7 **E)** Effect of pH on the threshold magnesium-ion concentration for array self-association. $[Mg^{2+}]_{50}$ represents the $[Mg^{2+}]$ at which 50% of array has sedimented in self-association assays such as those represented in D. Black dots (n=26) represent individual measurements from separate self-association assays performed at variable pH. The bars represent average $[Mg^{2+}]_{50}$ measurements with standard deviation error in steps of 1 pH.

The effect of pH on magnesium-induced 12-nucleosome array self-association was assessed by differential sedimentation (Figure 13D). Near the standard assay condition of pH 8.5, the use of the isoionic buffer did not change the extent of magnesium-ion induced array association from our standard buffer conditions. However, at pH 10.2 arrays were less sensitive to magnesium ion, while at pH 5.7 they were more sensitive. To more fully characterize this behavior, differential sedimentation was performed over more pH values from just above pH 5 to just below pH 11 (Figure 13E). In comparing the $[Mg^{2+}]_{50}$'s from these experiments, a clear trend is observed that reducing the pH also reduces the amount of magnesium ion required to trigger array self-association, where there is just over a 3-fold difference over this range. Thus, nucleosomal array self-association is sensitive to changes in pH.

Magnesium ion association

In our studies of the self-association of nucleosomal arrays, divalent magnesium ion drives this transition. However, relatively little is known about the amount of magnesium associated with these arrays. To address this question, we used inductively coupled plasma mass spectrometry (ICP-MS), a technique that can sensitively quantify atomic ions in solution. For the initial experiment (Figure 14A), 12-nucleosome arrays were self-associated at a concentration of 4.0 mM magnesium, isolated by differential sedimentation, and then analyzed by ICP-MS. This analysis showed an average stoichiometry of magnesium ions to array of roughly 3,400:1. In terms of relative charge, this represents an approximate charge ratio of 2.7:1 or 6,800 positive charges from the divalent magnesium ion to 2,486 net negative charges from the array DNA and all theoretically charged protein side chains. Overall, this indicates that the amount of magnesium ion present is in significant excess relative to what is required for simple charge neutralization. This observed excess was not due to the individual components of the self-

association, as neither mock experiments with just magnesium nor with just array showed comparable amounts of associated magnesium ion. Magnesium buffers at all pHs were also tested and had similar minimal interference in the experiment. (Data not shown) Thus, an excess of associated magnesium ion requires both magnesium and arrays and is presumably part of the self-associated array in some fashion.

To better understand factors that contribute to the amount of associated magnesium, we analyzed self-associated 12-nucleosome array samples generated from different concentrations of magnesium ion (Figure 14B). For the same amount of array, these data indicate a clear positive correlation between the concentration of magnesium used to generate the self-associated array and the amount of magnesium associated with the self-associated array, suggesting a near linear relationship. Because pH can change the $[Mg^{2+}]_{50}$ for array self-association (Figure 13E), we also investigated the effect of pH on associated magnesium (Figure 14C). In this experiment a consistent magnesium ion concentration of 4 mM was used for two reasons: first, because we did not want variations in magnesium at the various pHs to affect the amount of associated magnesium, as was seen in part 'B'; second, because this concentration is sufficiently past the $[Mg^{2+}]_{50}$ for array self-association for all of the tested pHs that differences in the extent that the arrays were self-associated would not complicate the analysis. These experiments showed that the amount of magnesium ion associated with the self-associated array was the same regardless of the pH. Thus, together these results indicate that the higher the magnesium concentration used to drive array self-association, the more magnesium ion that is ultimately associated with the array, but this association appears to be independent of the pH at which the self-associated array is generated.

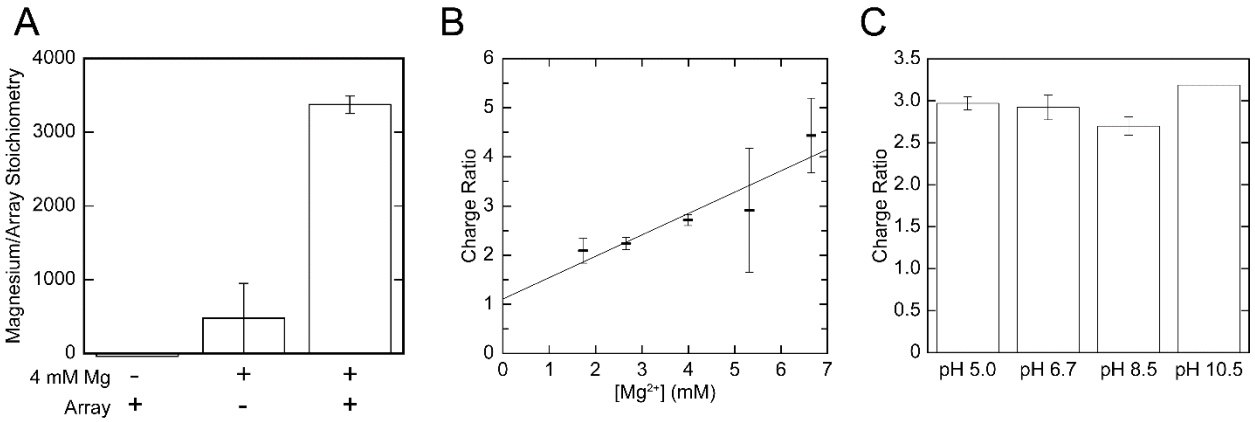

Figure 14. Amount of Mg^{2+} associated with oligomerized nucleosomal arrays under various conditions. A) Relative amount of magnesium ion associated with nucleosomal arrays following magnesium-ion induced oligomerization and sedimentation. Measurements of magnesium amount were determined by ICP-MS. All data in this and subsequent parts represents three replicates and 12-nucleosome repeat length. **B)** Effect of varying amounts of magnesium ions in solution on the amount of magnesium ions associated with oligomerized nucleosomal array. In this and part C, the amounts of associated magnesium were converted to the charge ratio: (#positive charges in Mg^{2+} bound to array) ÷ 2486: the theoretical absolute value net charge of an array. Solid line represents a linear fit with slope of 0.44 and y intercept of 1.1. **C)** Effect of pH on the amount of Mg^{2+} associated with nucleosomal arrays which were self-associated under 4 mM magnesium. pH 10.5 represents a single measurement.

Measuring kinetics and size of nucleosomal arrays and self-associated particles

While differential sedimentation provides an effective way to identify arrays that exceed a certain threshold of self-association, the method does not directly provide the size of these particles nor offer an easy way to probe the kinetics of the process. Dynamic light scattering (DLS) is a technique that reports on the effective size of particles in real time (34) and we were interested to see if it could be applied to array self-association. To test this, we used DLS to characterize the distribution of nucleosomal array particle sizes under several different conditions (Figure 15). In the absence of divalent magnesium ion (Figure 15A top), we observed a single array species that had a hydrodynamic radius centered at 28 (n=22) nm. This value is in very close agreement to sedimentation-based methods of these same kinds of arrays (6), where the median sedimentation coefficient suggests a hydrodynamic radius of ~27-29 nm. The

distribution of the array in DLS is not uniform, but this too is expected, as the population of arrays adopt different degrees of compaction as was observed in EM-based studies (35). When the magnesium concentration of the arrays is changed to 1.3 mM (Figure 15A middle), two species are apparent. The major species is one with a radius centered around a hydrodynamic radius of 20 nm, with a smaller population centered around 100 nm. The 20 nm particle is smaller than the average array without magnesium, and is consistent with folding of the array into a more compact “30 nm fiber” (11). Moreover, from sedimentation velocity experiments, comparable magnesium ion concentrations result in a roughly 2/3-fold change in sedimentation coefficient, similar to the change in radius observed from the top panel to the middle panel. At 1.3 mM of divalent magnesium ion, differential sedimentation shows that while most of the array is not self-associated (Figure 12A), the fraction that does sediment may correspond to the minor population centered around 100 nm. When the magnesium ion concentration is raised to 5.3 mM (Figure 15A bottom), we expect all of the arrays to become self-associated, and indeed that appears to be the case. The major species has a radius between 400-500 nm, with a small amount of array particles at 100 nms, potentially matching the minor species observed in part B. Recent fluorescent microscopy and EM studies of similar self-associated nucleosomal arrays show particles that range in radius from 150 nm to 500 nm observed in fluorescence microscopy, 400nm in negative staining TEM, and weight-averaged size of ~400 nm using SV-AUC when array is 100% self-associated (7) This is in good agreement with the major species observed. Altogether, these experiments validate that DLS can be used to assess nucleosomal array particle size.

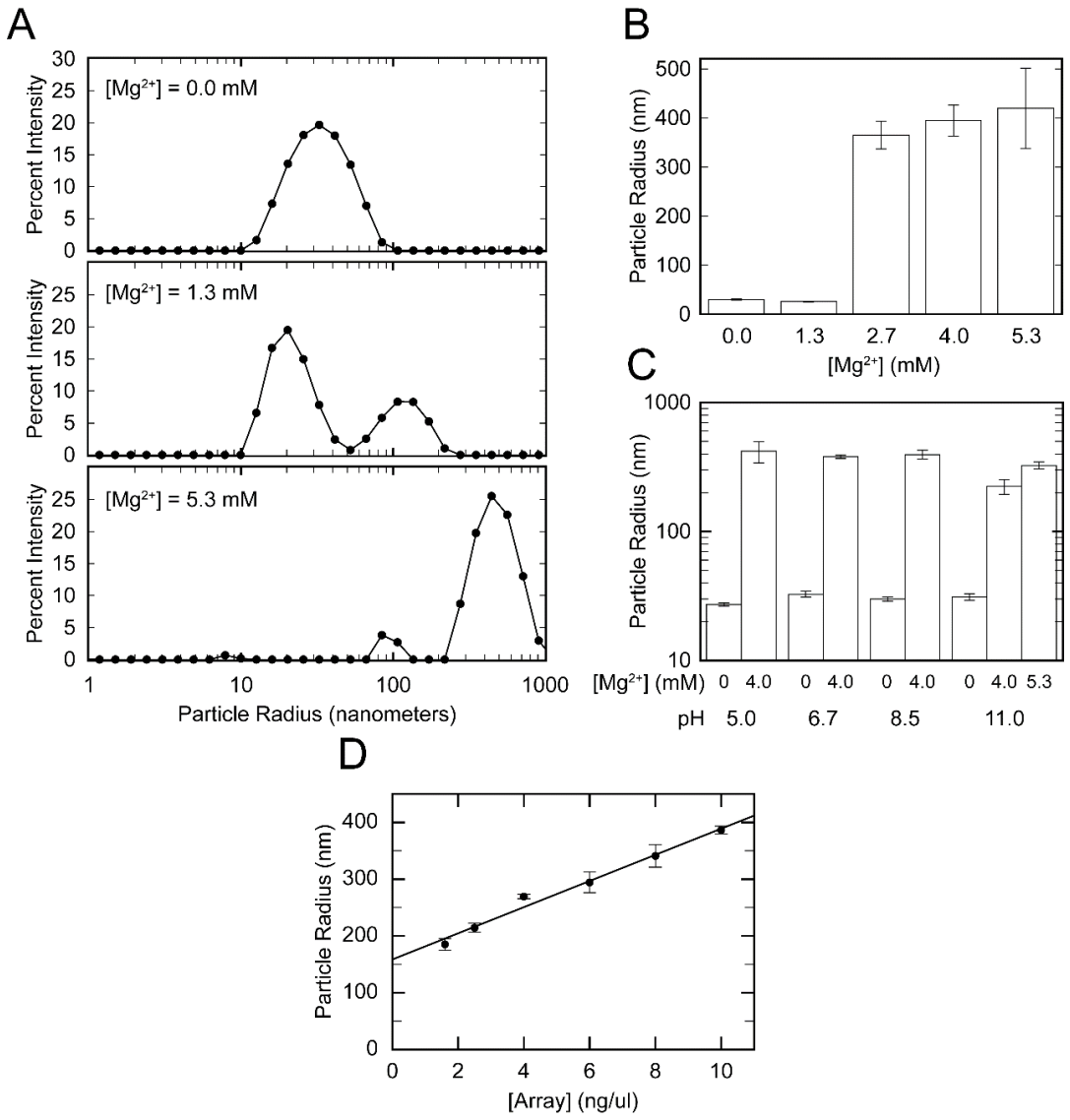

Figure 15 Determination of nucleosomal array sizes using dynamic light scattering. A) Comparison of the nucleosomal array particle size distributions under conditions that correspond to distinct structural states. Shown are representative regularization fits of autocorrelation data. No particles were observed above 1000 nm though the technique has the capacity to detect larger particles. Under 0 mM MgCl2, from 22 repetitions, center peak of array size was 28 nm in radius. Under 1.3 mM MgCl2, from 3 repetitions, the center peak of array size was 20 nm in radius with a second species observed consistently of size 100nm. Under higher concentrations of Mg after 10 minutes incubation, the center peak varied from 320 to 450 nm from 9 repetitions. The breadth of the peaks shown describe the polydispersity of the measurement. **B and C)** Effects of magnesium ion concentration at 10 ng/uL array template (parts B and C) and pH (part C) on average nucleosomal array size. Data is based on three replicants of cumulant fitting. **D)** Average self-associated array particle size as measured after 10 minutes of incubation. Dashed lines represent a linear fit to the data. Under constant 2.7 mM [Mg^{2+}], pH 8.5. The array DNA concentrations reflect effective nucleosome concentrations between 14 and 90 nanomolar.

We then used DLS to characterize how various parameters affect nucleosomal array particle sizes. When we looked at the effect of different magnesium ion concentrations (Figure 15B), we observed that beyond 2.5 mM, a concentration that differential sedimentation indicated was sufficient to induce full array self-association (Figure 12A), the size of the nucleosomal particles was roughly the same. This result suggests that beyond a threshold concentration of magnesium, more magnesium generates the same sort of self-associated array species. We also probed how pH affects array particle sizes (Figure 15C). In the absence of magnesium ion, the arrays were all small (20-30 nm) and relatively uniform in size. At 4.0 mM magnesium ion, all of the array particles were large (200-500 nm), although the arrays at pH 11.0 were smaller. However, from our pH titration studies (Figure 13E), the $[Mg^{2+}]_{50}$ at pH 11.0 is around 3.3 mM, and the fact that these arrays were not significantly past the $[Mg^{2+}]_{50}$ could be responsible for this smaller size. Indeed, with more magnesium ion, the array particle size at pH 11 increases to nearly that of the other arrays. These pH results suggest that despite the fact that at different pHs the $[Mg^{2+}]_{50}$ varies (Figure 13E), there is similarity in amount of associated magnesium ion independent of pH (Figure 15C). These results suggest that regardless of the pH, the arrays form a similar-sized particle once sufficiently past the threshold magnesium ion concentration.

Because particle sizes can be monitored by DLS in real time, we used it to study the kinetics of nucleosomal array particle growth. When our standard concentration of array (10 ng/μL template) was subjected to our standard magnesium ion concentration for full self-association (2.7 mM), we observed a time-dependent growth in particle radius size (Figure 16A). The kinetics of this process are complicated, but were simplified to two phases. The first phase consisted of rapid growth from roughly 30 nm to 100 nm, and this transition occurred faster than we could measure. To approximate the minimum rate of growth in this phase, we fit the initial

and first post-magnesium addition time point for a rate of 2.3 nm/s. Subsequent time points appeared roughly linear and the rate for this second phase provided a rate of 0.4 nm/s. This second phase is considerably slower than the initial phase, being 6-fold less. Although the time interval for the second phase was short enough to approximate the growth rate as linear, the growth rate past these timepoints did not appear so, ultimately slowing down. However, these data were highly noisy and irreducible, potentially due to the system signal attenuation necessary to measure such large particles and to fluctuations in relatively few particle numbers.

The ability to robustly measure the slower phase of nucleosomal array particle growth allowed us to probe how different factors affected particle growth rate. With variation in magnesium ion concentration, the particles, as expected did not grow below the threshold magnesium ion concentration. (Figure 16B) Above this threshold, at best there was a very modest increase in growth rate from 2.7 mM to 5.3 mM magnesium ion concentration (1.3x increase). However, these values were not statistically significant ($p=0.22$ where a p-value>0.05 suggests no significant difference). In contrast to the relative uniformity of array particle growth rates at magnesium ion concentrations past the threshold concentration, there was a clear dependence on growth rate on array concentration (Figure 16C), and was well fit by a linear curve.

Similarly, the sizes of these particles as measured after 10 minutes of incubation showed a clear dependence on array concentration. Arrays at 14 nanomolar reached a size of 200nm and 90 nanomolar array reached a size of about 400 nanometers. (Figure 15D) It has recently been shown that self-associated array particles reach a similar size of 250-500nm radius at all concentrations from 2.5 nanomolar all the way up to 100 nanomolar although at higher concentrations the particles appear to start to stick together in a non-uniform manner. (22) We

expect this indicates that as array concentration is reduced that it may take longer for arrays to reach an equilibrium size. Additionally, at larger concentrations, the slow phase growth that we observe may be accelerated by individual spherical particles sticking together.

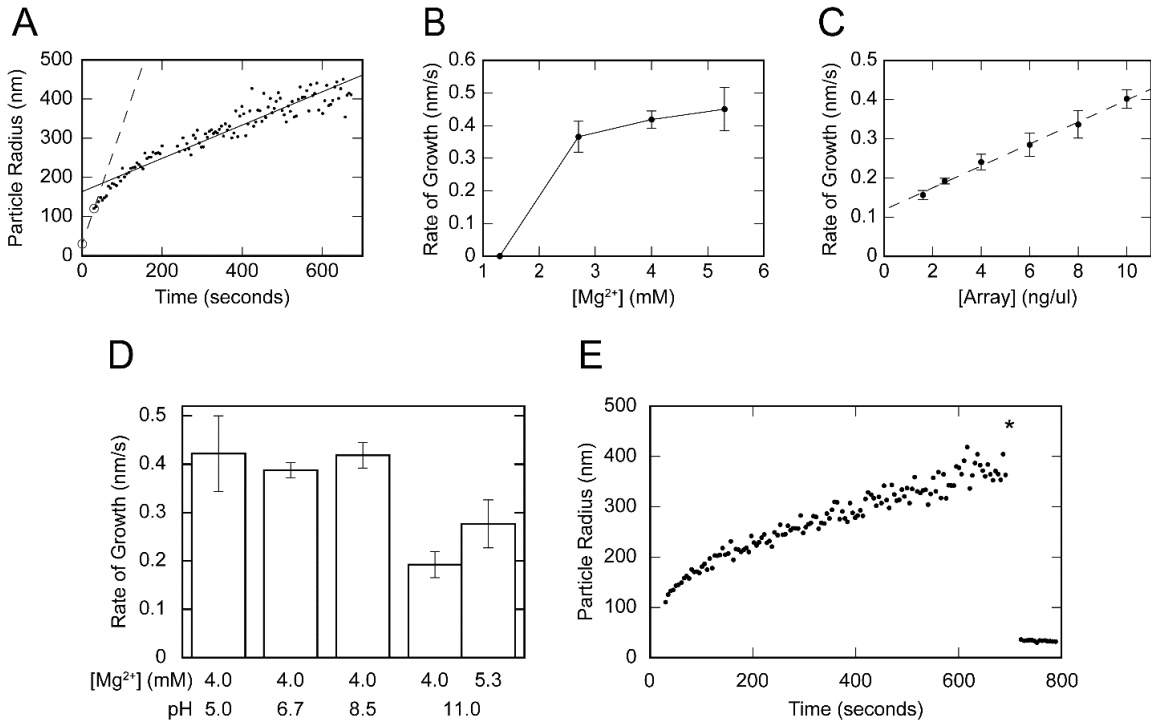

Figure 16. Kinetics of nucleosome array self-association and dissociation. A) Representative time course of nucleosomal array particle size growth. Raw particle size data is shown as individual dots, where magnesium ion concentration is brought to 2.7 mM at time zero. The dashed line was used to estimate the minimum first phase rate of change from the array particle size right before time zero to the first measurement at about 30s. The solid line is a linear fit of the data from 30s to 700s and was used to calculate the second phase rate. **B)** Effect of magnesium ion concentration on the nucleosomal array particle growth rate. Experiment performed at 10 ng/uL, pH 8.5. For this and parts C and D, each data point represents three measurements of the second phases rates. **C)** Effect of array concentration on the nucleosomal array self-associaiton particle growth rate. Dashed lines represent a linear fit to the data. Under constant 2.7 mM [Mg^{2+}], pH 8.5. The horizontal axis in array DNA concentrations reflect effective nucleosome concentrations between 14 and 90 nanomolar. **D)** Effect of pH and magnesium ion concentration on 10 ng/uL nucleosomal array particle growth rate. **E)** Representative timecourse of nucleosomal array particle dissociation. Array size was measured as described in part A for the initial timepoints. EDTA was added to a final concentration of 4 mM at 700 seconds (represented by the asterisk).

Investigation of the effect of pH on the growth rate indicated that at most pHs, the array particles grew at a similar rate (Figure 16D). The one exception was growth of array particles at pH 11.0. However, as with particle size at this pH (Figure 15C), the rate of growth appeared to approach those observed at the lower pH values when the magnesium ion concentration was increased past the $[Mg^{2+}]_{50}$.

As chromatin and nucleosomal array self-association is a reversible process, we sought to characterize the kinetics of array dissociation by DLS. Starting from arrays that had grown to between 300-400 nm in size, a concentration of EDTA was added that was just slightly in excess (1.5-fold) of the magnesium ion concentration (Figure 16E). Dissociation to individual nucleosomal array with hydrodynamic radii identical to the pre-magnesium addition was very rapid, and no intermediates in going to this state were observed in the time it took between adding the EDTA and measuring the array particle sizes. Because EDTA chelation of magnesium is energetically favorable and could be driving the particle dissociation, we wondered if reduction of magnesium ion concentration by rapid dilution might allow us to observe dissociation intermediates. However, rapid dilution of concentrated magnesium-induced self-associated arrays into a volume that reduced the total magnesium ion concentration to $1/10^{th}$ the $[Mg^{2+}]_{50}$ also generated individual array particle in the timeframe of measuring. Thus, these studies indicate that dissociation of nucleosomal arrays from their self-associated state is rapid.

Discussion

Structure of self-associated arrays

Nucleosomal arrays can self-associate to form either a solid- or liquid-like phase both in vitro and in vivo (7,21,22). Which phase is adopted in vitro appears to depend on the assay component (22), and our reaction conditions largely matched the conditions in which arrays

associate in a solid-phase. Additionally, prior unpublished fluorescence microscopy results from our lab showed that these arrays self-associated to form solid particles, and our DLS results show sizes in good agreement with the solid particles characterized by others (7). Array particles can adopt more irregular chain-like structures (22), and these especially large species may be present as species that were difficult to analyze by DLS in later assay timepoints.

Formation of the self-associated array particles is induced by divalent magnesium ions, and to what extent these ions are part of the particle and how has not been clear. Our studies indicate that isolated array particles have large number of magnesium ions per array, with roughly 5,500 magnesium ions per array at the highest magnesium concentration tested. This number of magnesium ions is a significant excess in positive charge over what is required to neutralize the net negative charge of the array (4.4 positive charges to 1 negative charge at the highest magnesium concentration tested). Presumably other counterions, such as the chloride ions also present, counteract this excess. The amount of magnesium ion associated with the array increases in roughly a linear fashion as a function of the magnesium concentration used to generate the self-associated species. This linear behavior suggests that the associated magnesium is likely to represent two distinct populations of magnesium cations: The constant component of the curve, seen in the the y-intercept, represents an amount of magnesium that is associated with the array regardless of how much was used to drive self-association. This value is close to the 1:1 charge ratio required to neutralize the negative charge of the array, indicating that this amount of magnesium ions is required for array self-association and is more directly held by the array. The variable part of the curve, seen in the slope, represents the magnesium ions whose concentration increases in proportion to the magnesium ion concentration used to drive self-association. Because this population of magnesium seems to simply reflect the amount present in solution, it

is likely not required for self-association but is part of the solution trapped either in or around the particle. From the magnitude of the slope, the fraction of solution to total volume can be determined and is found to be roughly 40%.

While our data cannot readily determine how solution is held by the array particles, one way that such behavior is possible is if the self-associated array is a hydrogel-like species, where the particle is a network of arrays that is solvated by aqueous solution. Hydrogels include substance such as agarose gels that are only roughly 1% agarose by mass, and proteins associated with nucleic acids have also been shown to form hydrogels (36). Such a hydrogel species could include sufficient amounts of water to effectively carry solution magnesium ions with it, where higher concentrations of magnesium ion used to induce array association would trap a proportionally larger amount of magnesium ion in the array particle. Functionally, a hydrogel would be a useful structure for array particles, as they would still allow some access to the interior of the structure, especially for small molecules. Such a structure could explain the rapid dissociation kinetics we observe, where EDTA could enter the interior of the array particle via channels in the hydrogel and then chelate the magnesium ions necessary for self-association.

Factors in array self-association

A number of factors can influence array self-association. Our previous studies indicated that arrays composed of as few as four nucleosomes undergo differential sedimentation in a manner similar to arrays with twelve nucleosomes. (31) Our results with longer arrays, show that even more nucleosomes do not have a significant impact on array association. This result suggests that any entropic benefit derived from linking nucleosomes together does not contribute beyond four nucleosomes, and four nucleosomes may serve as a core domain for cross-array contacts. Our array concentration differential sedimentation results (Figure 12C) show some

increase in the amount of magnesium ion required for self-association, though at the highest concentration of arrays, this trend is reversed. These trends are relatively modest and require more extensive study. We also observe that array concentration, unlike magnesium concentration, appears to also affect self-associated array particle size, with larger arrays generated by with higher array concentrations. However, our kinetics data indicates that these differences likely reflect difference in growth rate, and that the ultimate particle sizes achieved may not differ, something observed in other studies with longer incubation times (22).

Our studies also show that pH affects array self-association, where reductions in the pH increases array sensitivity to magnesium ions. These results agree with previous studies with isolated chromatin fractions from Hela cells, which found changes in pH could perturb oligomerization (37). This agreement suggests that our trends could persist with heterogeneity in array length, nucleosome saturation, and histone modification. The change in the $[Mg^{2+}]_{50}$ required for self-association with pH could be due to several different factors. Under more basic conditions, magnesium ions can form magnesium hydroxides, which may not be as effective in promoting array self-association. In fact, magnesium hydroxides are reported to precipitate out of solution at high pHs (38). However, ICP-MS analysis of sedimented samples without arrays at pH 10.5, 9.1, 6.1, and 5.1 showed consistently insignificant amounts of magnesium present, suggesting that magnesium was present in solution in some form over the range of pHs we tested.

Another possible factor that could contribute to the observed pH trends is that protonation of arrays at lower pHs could reduce the overall negative charge of the arrays, promoting array self-association in a manner similar to neutralization by divalent magnesium cation. There are numerous ionizable groups in our arrays that could play such a role, and it seems likely that the

side-chain groups of the histone proteins are major contributors. Specifically, there are an abundance of residues with pKa's that are within one unit of the pH's tested: per recombinant *Xenopus laevis* nucleosome there are 114 lysines (pKa 10.5), 30 tyrosines (pKa 9.5-10), 18 histidines (pKa 7.0), 48 glutamic acids (pKa 4.4), 24 aspartic acids (pKa 4.0), and 1 cysteine (pKa 8.3)(39,40). It is likely that the actual protonation state of these residues is more complicated than just their pKa's, as the high charge density of the proteins and DNA in the nucleosome is likely to perturb the extent of their protonation.

While the phosphate backbone of the DNA has many oxygens that could be protonated (354 per nucleosome), the pKa of 0-2 for the phosphodiester is likely to be too far outside of the pH range to contribute significantly (38). It should be noted that DNA bases also have ionizable groups, and a number of these have pKa values in individual nucleosides that are near the extremes of the pH's probed: 9.2 for guanosine, 9.2 for thymidine, and 4.2 for cytidine (38). Nonetheless, the sites of protonation or deprotonation for each of these bases is involved in hydrogen bonding in the double stranded DNA, which significantly alters their pKa's (41). Thus, we suspect changes to their protonation state does not significantly influence the pH dependences of array self-association. However, outside of the tested range, ionization of these groups, as well as those from the histones, are likely to be responsible for the decreased array stability that we observe.

Array self-association kinetics and intermediates

The formation of self-associated nucleosomal arrays appears to have at least several distinct phases. On a timescale faster than our initial DLS measurements, individual array above the $[Mg^{2+}]_{50}$ for self-association rapidly form particles that have a hydrodynamic radius roughly 100 nm in size. Following that, there is a significantly slower phase in which the particles grow

to size of 400 nm and above in hydrodynamic radii. From probing the slower phase, it appears that above the $[Mg^{2+}]_{50}$, additional magnesium does not accelerate the process, suggesting that arrays are saturated in associated magnesium necessary for slow phase growth. However, the rate of growth does increase linearly with array concentration, indicating that having more array present, either individually or in particle form speeds up growth.

Although, mechanistic models for chromatin systems are potentially very complex, our results are consistent with a nucleation-based model of growth. In the early phase, a relatively small number of arrays come together to form a seed particle. Once formed, these particles can be added to from individual arrays or from the coalescing of, or sticking to, other particles. In such a model, once enough magnesium is present to reduce the zeta potential of the particles, they can associate by the frequency of their encounters, where increased array concentrations lead to a higher encounter frequency and therefore faster growth rates. Interestingly, the smaller array particle intermediate can be seen not only in the early phases of array self-association, but also under conditions in which significant differential sedimentation is not seen. At magnesium ions concentrations below the $[Mg^{2+}]_{50}$, but above zero, DLS analysis shows that most arrays undergo compaction to a state consistent with the 30 nm fiber, but some are the same size as the 100 nm seed particle, while no larger particles are observed. It is not clear if the most compacted state is an intermediate or an off-pathway product that antagonizes formation of the seed particle, but previous work in our lab has suggested that formation of this compacted state has a negative impact on forming particles large enough for differential sedimentation (6).

The separation of array self-association into distinct intermediates is reminiscent of the hierarchy of observed chromosome domains. Chromatin is observed to form topologically associated domains (TADs) in vivo with sizes between with radii between 80-100 nm (15).

Association of these TADS are believed to then form chromosomal compartments with radii between 500-1000 nm. The hierarchical nature of our self-associated array and the close correspondence of the sizing of our array particles to TAD and chromosomal compartments is highly suggestive and merits future study.

In conclusion, our studies have shown how pH, array concentration, array length, and magnesium concentration affect the kinetics, formation, and structure of array self-associated states. In addition to advancing our knowledge of the structure and function of chromatin in living systems, a better understanding of array self-association could be useful in other contexts, such as DNA-based materials. The complementary nature of double stranded DNA has led to its ubiquitous use in a wide range of fabricated materials, nano-devices, and sensors, and the unique structural transitions that chromatin imparts upon the underlying DNA may augment and expand the range of uses that are possible.

Bibliography

1. Prieto, Eloise I., and Maeshima, K. (2019) Dynamic chromatin organization in the cell. *Essays in Biochemistry* **63**, 133-145

2. Maeshima, K., Ide, S., and Babokhov, M. (2019) Dynamic chromatin organization without the 30-nm fiber. *Current Opinion in Cell Biology* **58**, 95-104

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

4. Hayes, J. J., and Hansen, J. C. (2001) Nucleosomes and the chromatin fiber. *Curr Opin Genet Dev* **11**, 124-129

5. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal Structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

6. Sinha, D., and Shogren-Knaak, M. A. (2010) Role of direct interactions between the histone H4 Tail and the H2A core in long range nucleosome contacts. *J Biol Chem* **285**, 16572-16581

7. Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., DeLuca, J., Ishikawa, T., and Hansen, J. C. (2016) Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *The EMBO Journal* **35**, 1115-1132

8. Németh, A., and Längst, G. (2004) Chromatin higher order structure: Opening up chromatin for transcription. *Briefings in Functional Genomics* **2**, 334-343

9. Woodcock, C. L. F., Safer, J. P., and Stanchfield, J. E. (1976) Structural repeating units in chromatin. *Experimental Cell Research* **97**, 101-110

10. Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138-141

11. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R. M., Zhu, P., and Li, G. (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380

12. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., and Woodcock, C. L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A* **95**, 14173-14178.

13. Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A. S., Imamoto, N., Ishikawa, T., and Maeshima, K. (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J* **31**, 1644-1653

14. Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025

15. Szabo, Q., Bantignies, F., and Cavalli, G. (2019) Principles of genome folding into topologically associating domains. *Science Advances* **5**, eaaw1668

16. Simpson, R. T., Thoma, F., and Brubaker, J. M. (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

17. Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T. J. (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol* **327**, 85-96

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

19. Strick, R., Strissel, P. L., Gavrilov, K., and Levi-Setti, R. (2001) Cation-chromatin binding as shown by ion microscopy is essential for the structural integrity of chromosomes. *J Cell Biol* **155**, 899-910

20. Maeshima, K., Matsuda, T., Shindo, Y., Imamura, H., Tamura, S., Imai, R., Kawakami, S., Nagashima, R., Soga, T., Noji, H., Oka, K., and Nagai, T. (2018) A Transient Rise in Free Mg2+ Ions Released from ATP-Mg Hydrolysis Contributes to Mitotic Chromosome Condensation. *Current Biology* **28**, 444-451.e446

21. Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S., and Rosen, M. K. (2019) Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e421

22. Strickfaden, H., Tolsma, T. O., Sharma, A., Underhill, D. A., Hansen, J. C., and Hendzel, M. J. (2020) Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell* **183**, 1772-1784.e1713

23. Blacketer, M. J., Gannon, M. K., Young, I. A., and Shogren-Knaak, M. A. (2017) Solid-phase synthesis of highly repetitive chromatin assembly templates. *Analytical Biochemistry* **531**, 12-15

24. Stafford, D. W., and Bieber, D. (1975) Concentration of DNA solutions by extraction with 2-butanol. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **378**, 18-21

25. Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., and Luger, K. (2003) Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. in *Methods in Enzymology*, Academic Press. pp 23-44

26. Carruthers, L. M., Tse, C., Walker, K. P., 3rd, and Hansen, J. C. (1999) Assembly of defined nucleosomal and chromatin arrays from pure components. *Methods Enzymol* **304**, 19-35

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

28. Ellis, K. J., and Morrison, J. F. (1982) [23] Buffers of constant ionic strength for studying pH-dependent processes. in *Methods in Enzymology* (Purich, D. L. ed.), Academic Press. pp 405-426

29. Gaykalova, D. A., Kulaeva, O. I., Bondarenko, V. A., and Studitsky, V. M. (2009) Preparation and Analysis of Uniquely Positioned Mononucleosomes. in *Chromatin Protocols: Second Edition* (Chellappan, S. P. ed.), Humana Press, Totowa, NJ. pp 109-123

30. Muthurajan, U., Mattiroli, F., Bergeron, S., Zhou, K., Gu, Y., Chakravarthy, S., Dyer, P., Irving, T., and Luger, K. (2016) Chapter One - In Vitro Chromatin Assembly: Strategies and Quality Control. in *Methods in Enzymology* (Marmorstein, R. ed.), Academic Press. pp 3-41

31. Blacketer, M. J., Feely, S. J., and Shogren-Knaak, M. A. (2010) Nucleosome interactions and stability in an ordered nucleosome array model system. *J Biol Chem* **285**, 34597-34607

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

33. Schwarz, P. M., and Hansen, J. C. (1994) Formation and stability of higher order chromatin structures. Contributions of the histone octamer. *J Biol Chem* **269**, 16284-16289

34. Kamgar-Parsi, K., Hong, L., Naito, A., Brooks, C. L., III, and Ramamoorthy, A. (2017) Growth-incompetent monomers of human calcitonin lead to a noncanonical direct relationship between peptide concentration and aggregation lag time. *Journal of Biological Chemistry* **292**, 14963-14976

35. Azzaz, A. M., Vitalini, M. W., Thomas, A. S., Price, J. P., Blacketer, M. J., Cryderman, D. E., Zirbel, L. N., Woodcock, C. L., Elcock, A. H., Wallrath, L. L., and Shogren-Knaak, M. A. (2014) Human heterochromatin protein 1alpha promotes nucleosome associations that drive chromatin condensation. *J Biol Chem* **289**, 6850-6861

36. Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., and McKnight, S. L. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767

37. Guo, X. W., and Cole, R. D. (1989) Chromatin aggregation changes substantially as pH varies within the physiological range. *Journal of Biological Chemistry* **264**, 11653-11657

38. Thaplyal, P., and Bevilacqua, P. C. (2014) Chapter Nine - Experimental Approaches for Measuring pKa's in RNA and DNA. in *Methods in Enzymology* (Burke-Aguero, D. H. ed.), Academic Press. pp 189-219

39. Hass, M. A. S., and Mulder, F. A. A. (2015) Contemporary NMR Studies of Protein Electrostatics. *Annual Review of Biophysics* **44**, 53-75

40. Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999) Expression and Purification of Recombinant Histones and Nucleosome Reconstitution. in *Chromatin Protocols* (Becker, P. B. ed.), Humana Press, Totowa, NJ. pp 1-16

41. MOODY, E. M., LECOMTE, J. T. J., and BEVILACQUA, P. C. (2005) Linkage between proton binding and folding in RNA: A thermodynamic framework and its experimental application for investigating pKa shifting. *RNA* **11**, 157-172

CHAPTER 4. GENERAL CONCLUSION

Our findings in relation to factors of pH, length, magnesium concentration and nucleosome concentration have shown exciting possibilities for learning more about kinetic and structural properties of chromatin. Our explorations have given us a plethora of new physical parameters to report on self-association beyond differential sedimentation including kinetics of formation and dissociation, magnesium binding stoichiometry, as well as a new faster way to measure the size of arrays in all states. In the future, we hope to uncover even more factors which impact these properties of chromatin as well as delve deeper into the relationships between factors and properties we have seen including the following factor/property relationships:

1. how factors impact formation of the 100nm intermediate self-associated particle,
2. how the factor of saturation impacts all properties,
3. how the factor of including PTMs, especially acetylation, impact all properties,
4. how factors impact hydration state of the self-associated particle state,
5. how the factor of array length impacts self-associated particle kinetics, and
6. extend our knowledge of how the factor of concentration and array length impact all properties.

Furthermore, we anticipate the use of DLS will prove invaluable as another means to perform quality control on new samples of prepared array, and maybe tell us information about saturation, and perhaps the level of intactness of histone tails which are vulnerable to cleavage by protease enzymes.

Our work with nonsense suppression mutagenesis has provided us with an effective means to readily access H3 histones with a wide range of acetylation states. These reagents offer us the means to probe to an unprecedented degree how acetylated lysine residues individually and in combinations modulate nucleosome structure and function. Future work in the Shogren-Knaak lab will initially focus on two questions – 1) How acetylation of specific H3 tail lysine residues influences the subsequent acetylation of the histone lysines within and between nucleomes, and 2) How specific H3 tail acetylations promote the progression of nucleosome disassembly.

Investigating SAGA's postulated acetylation spreading mechanism

In vivo studies of nucleosome acetylation, such as those using Chromatin IP, have revealed that a basal level of histone acetylation is present ubiquitously throughout the genome, and that when further acetylation is introduced, it tends to rise significantly within a nucleosome, and to spread outward to neighboring nucleosomes (1). These results have led to models in which pre-existing nucleosome acetylation can promote subsequent acetylation, and that this can lead to a processive spreading of acetylation between nucleosomes. This model is further bolstered by the fact that histone acetyl-transferase enzymes, such as the SAGA complex, contains domains that bind specificially to acetylated histone tails (2). Nonetheless, mechanistic details are largely absent, and can be studies with our ability to readily install specific patterns of lysine acetylation in the H3 histone tail.

Within a nucleosome, recent results from our group and others have indicated that nucleosome acetylation is processive, and that pre-existing acetylation can potentially facilitate this processivity. Pre-existing acetylation on only one of the two N-terminal H3 histone tails within a nucleosome results in increased levels of acetylation on the unacetylated tail (3).

Trimethylation of lysine residue 4 of histone H3 has recently been proven to increase processivity of the SAGA complex (4). Recent preliminary evidence suggests the pre-steady state kinetics of SAGA contain a burst-phase of multiple acetylations. Work is ongoing to determine the number of acetylations measured. (Mittal and Shogren-Knaak, Unpublished results). Building on these results, we seek to better understand how specific individual and combinations of lysine acetylations and their recognition by the bromodomain contributes to this processivity. To that end, we will first continue studying pre-steady state kinetics of SAGA on nucleosome substrates to fully characterize the burst-phase. Then secondly, incorporate the use of nucleosomes containing pre-existing acetylation in measuring the burst-phase of SAGA and compare those results to SAGA mutants where one or more of the bromodomains have been deleted. We expect these experiments may tell us something about the factors which dictate the parameters of the burst-phase pre-steady state kinetics of nucleosomes.

Processivity between nucleosomes has not been previously demonstrated, but access to acetylated histone and tools developed from our lab will provide a means to probe to what extent this processivity occurs. Chitvan Mittal has developed a technique in our lab which we call nucleosome acetylation sequencing (NAS) to measure levels of catalytically installed acetylation of individual nucleosomes within nucleosomal arrays(5). NAS follows a strategy of first ligating individual nucleosomes sequentially onto a solid substrate, catalytically acetylating using a tritium labeled acetyl group, and finally cleaving individual nucleosomes from the bead to read the levels of radioactivity in a scintillation counter. Scintillation counts may then be correlated to an average number of acetylations per nucleosome.

NAS and nonsense suppression give us unique tools to investigate a possible answer to the aforementioned question of SAGA's ability to navigate and acetylate chromatin in a selective

manner. In our future investigations, we want to examine a nucleosomal array by NAS which contains non-radioactive acetylation (generated via nonsense suppression) on 1 out of many positioned nucleosomes in the array. Higher levels of acetylation at those nucleosomes which are neighboring the pre-acetylated nucleosome when compared to those nucleosomes at the outskirts of the array would be sufficient to indicate that SAGA is, at least in part, preferentially acetylating nucleosomes that have acetylated neighboring nucleosomes.

Characterizing how Specifically Acetylated Lysines Disrupt Nucleosome Stability

Studies have shown that nonspecific bulk acetylation of histones destabilizes nucleosomes (6-9). However, it is not clear which sites or combination of sites contribute most significantly. Utilizing multiply acetylated histones generated via nonsense suppression, we can study effects of acetylation at a number of different sites. Single molecule techniques suggest the first step of nucleosome dissociation is a peeling away of the DNA at the entry exit domain with an H2A/H2B dimer adhering to the DNA and not the rest of the histone octamer (10-13). Histone H3 tail DNA is located near the entry/exit sites of the DNA that wraps around the histone core and potentially interacts with DNA that becomes unpeeled from the H2A/H2B dimers. Thus, by systematically varying the number and position of lysine acetylation, we should be able to test whether all H3 lysines are equally involved in stabilizing the nucleosomes, potentially by providing general electrostatic interactions with the nucleosome entry-exit DNA, or if neutralizing lysing residues in a specific order could facilitate the peeling away of the nucleosomal DNA from the histones.

We will detect the potential small changes in structure brought on by acetylation of H3 using single molecule total internal reflection fluorescence microscopy (smtirfm). In this technique, fluorescently labeled nucleosomes will be assembled with internal FRET pairs. Using

SMTIRFM, live changes in FRET of single nucleosomes can be observed in real time (14). A graph is generated which graphs nucleosome dissociation events as an exponential decay curve. The decay constant is the inverse of the average lifetime of molecules and corresponds. And the decay constant is analogous to the kinetic off rate. Nucleosomes are stable and off rates are slow in low salt buffers. Using smTIRFM, salt buffer may be added at time 0. The salt destabilizes the nucleosome and make the off-rate easily measured.

Single molecule techniques make it possible to see changes with very little noise because the user selects only molecules which are free of artifacts a histogram can be generated which shows changes in populations of high FRET and low FRET efficiency species. The effect of acetylation at lysine 56 on histone H3 was shown to reduce counts of individual nucleosomes with high FRET between dye pairs at the entry-exit region but not at dye pairs located on an internal position within the nucleosome (15).

Minor works in solid-phase ligation strategies during graduate career

Explanation for below work.

During my time as a graduate student, I began my work with nonsense suppression (see chapter 2). As I was finishing up this work, I began the work of cloning and pilot experiments for the work of solid-phase ligation of DNA shown below. My foundational work was handed off to my colleagues who completed the work while I moved on to investigate the work shown in chapter 3. As my work did not ultimately yield a data figure used in solid phase ligation, we felt it was not prudent to include this work as an official chapter in my dissertation. However, as my work was instrumental in establishing the methods employed and I am an author on this published work, we still felt it was prudent to include within my dissertation.

Solid-phase synthesis of highly repetitive chromatin assembly templates

Melissa J. Blacketer[1], Margaret K. Gannon[1], Isaac A. Young[2], Michael A. Shogren-Knaak[3]

[1]Equal contributions, [2]Contributed to conceptual design, and early experimental methods development, [3]Principle investigator

Modified from a manuscript published in *Analytical Biochemistry*

Abstract

DNA templates for assembling chromatin model systems typically consist of numerous repeats of nucleosome positioning sequences, making their synthesis challenging. Here we describe a solid-phase strategy for generating such templates using sequential enzymatic ligation of DNA monomers. Using single nucleosome site monomers, we can either generate a twelve-nucleosome site target, or systematically access intermediate-sized templates. Using twelve nucleosome positioning site monomers, longer templates can be generated. Our synthesized templates assemble into well-defined chromatin model systems, demonstrating the utility of our solid-phase approach. Moreover, our strategy should be more widely applicable to generating other DNAs containing highly repetitive DNA sequences.

Introduction

Chromatin is composed of a linear string of nucleosomes (16), DNA wrapped around a spool of histone proteins (17). Chromatin model systems can be used to better understand the structure of chromatin and its use as a substrate (18,19), and are often assembled from DNA templates that contain head-to-tail repeats of nucleosome positioning sequences (20). These highly repetitive templates are typically generated by random insertion of individual DNA repeat

units into a vector (21,22). However, with this approach, the size of the resulting template is not readily controlled, nor is it easy to generate templates with sequence differences at specific nucleosome positioning sites. Recently, our group developed a bead-based method for step-wise assembly of trinucleosome model substrates from individual nucleosomes (23). Here we describe our effort to adapt and extend this approach to make longer DNA templates for chromatin model system assembly.

Materials and methods

Preparation of DNA fragments

Plasmids containing the 601-172-1 and 601-177-12 ligation fragments were made from 601-177-1 (24) and 601-177-12 (22) templates, respectively, using PCR and cloning. Fragments were excised by BglI digestion (NEB, for this an all restriction enzymes) and gel purified. Primer and vector details are included in the supplementary materials. Biotinylated, double strand DNA adapter and cap fragments with non-palindromic sticky ends were made by annealing synthesized oligonucleotides (IDT) (23). Oligo details are included in the supplementary materials.

DNA Solid-Phase Ligation

The 601-172-12 ligated DNA template was created as follows: 0.95 pmoles of biotinylated nucleotide adapter, ADT-BglI', was immobilized onto either 50 μg or 500 μg of hydrophilic streptavidin magnetic beads (NEB) (23). Next, a two-hour room temperature ligation on the beads was performed: a 2.25:1 ratio of non-palindromic 601-172-1 fragment to ADT-BglI' was mixed to a total volume of 10 μl in 1X Mighty Mix Ligation Kit solution (Takara Bio USA). The beads were then washed with 50 μl of 1X NEB Ligation Buffer to remove any

excess, unligated fragment. Ligation and washes were continued, alternately attaching non-palindromic 601-172'-1 and 601-172-1 fragments. After twelve rounds, ligation of CAP-BglI was performed using the above conditions. The ligated and "capped" fragments were photocleaved from the beads directly into 1 x NEB 3.1 buffer for 10 min with exposure to 312 nm UV light. 20 units each of PstI and BamHI were added and the DNA digested in in a total volume of 30 μl for 2 hours at 37 °C. The released fragments containing various numbers of 172 bp repeats were gel purified and cloned into p601X (25) to create plasmids p601X-172-8, p601X-172-10, and p601X-172-12. DNA templates NPD-172-8, NPD-172-10, and NPD-172-12, respectively, could be excised with PstI/BamHI digestion and gel purified.

601-177-24, 36 and 48 non-palindromic DNA templates NPD-177-24, NPD-177-36 and NPD-177-48 were created as above using non-palindromic 601-177-12 and 601-177'-12 fragments with the following difference: 0.388 pmoles of biotinylated nucleotide adapter was used with 120 μg of beads. Ligations were performed with 1,200 U of T4 ligase (NEB) in a final concentration of 1x Ligation Buffer for 4 hours with 1.5:1 molar ratio of fragment to ADT-BglI. For NPD-177-24 and NPD-177-48, ligation with 1CAP-BglI at a 1.5:1 molar ratio was performed after two and four cycles of ligation, respectively. For NPD-177-36, CAP-BglI' was ligated after three cycles. After photocleavage and digestion, fragments were cloned to create plasmids p601X-24, p601X-36 and p601-X48 from which templates NPD-177-24, NPD-177-36 and NPD-177-48 respectively could be excised with PstI/BamHI digestion and gel purification.

Assembly and Analysis of Nucleosomal Arrays

Recombinant *Xenopus laevis* histones were expressed, purified, and then assembled into histone octamers according to standard protocols (26). Histone octamers were deposited onto various DNA template, via step-wise salt deposition using previously described methods (20,24).

Nucleosomal arrays were digested with restriction enzymes to determine correct assembly, saturation, and stability. The general protocol was as follows: 400 ng of array was digested with 20 U of ScaI-HF (or BglI for 172 array) in a total of 20 µl of digestion buffer (50 mM NaCl, 10 mM Tris pH 7.4, 2.5 mM $MgCl_2$, 0.05% Triton-X) for 2 hours at 37 °C. The nucleosomes and free DNA generated by restriction digestion was characterize by 4% native PAGE analysis in 0.5X TBE, according to standard protocols (20). The absolute molecular weight of nucleosomal arrays was determined by size exclusion chromatography with multi-angle light scattering as previously described (24), with the following changes: In the protein conjugate analysis used to determine array saturation, the array was decomposed into free DNA monomers and mononucleosome components, using the following parameters determined from free DNA monomer and mononucleosome model systems: DNA ε_{260} of 20 ml•mg^{-1}•cm^{-1} and *dn/dc* of 0.1269 ml•g^{-1}. Mononucleosome ε_{260} of 17.74 ml•mg^{-1}•cm^{-1} and *dn/dc* of 0.1269 ml•g^{-1}.

Results and discussion

To generate chromatin assembly templates, we utilized a solid-phase synthesis strategy (Figure 1A). In this process, a cleavable DNA adapter is attached to a magnetic bead and then undergoes multiple cycles of enzyme-mediated DNA fragment ligation. After capping and cleaving the full-length DNA, the product is cloned, amplified, and purified from *E. coli*, to give a DNA template suitable for chromatin-model system assembly. In this approach, the DNA ligation fragments contain different non-palindromic overhangs, to enforce head-to-tail ligation, while preventing ligation in the opposite orientation or fragment self-ligation. As few as two

different fragment types can be used, limiting the number of different fragments required. However, additional fragments can be used to change the sequence at specific sites in the template. With solid-phase synthesis, an excess of fragment can be used in each ligation step to drive the reaction to completion, and unincorporated reagents can be readily washed away from the beads.

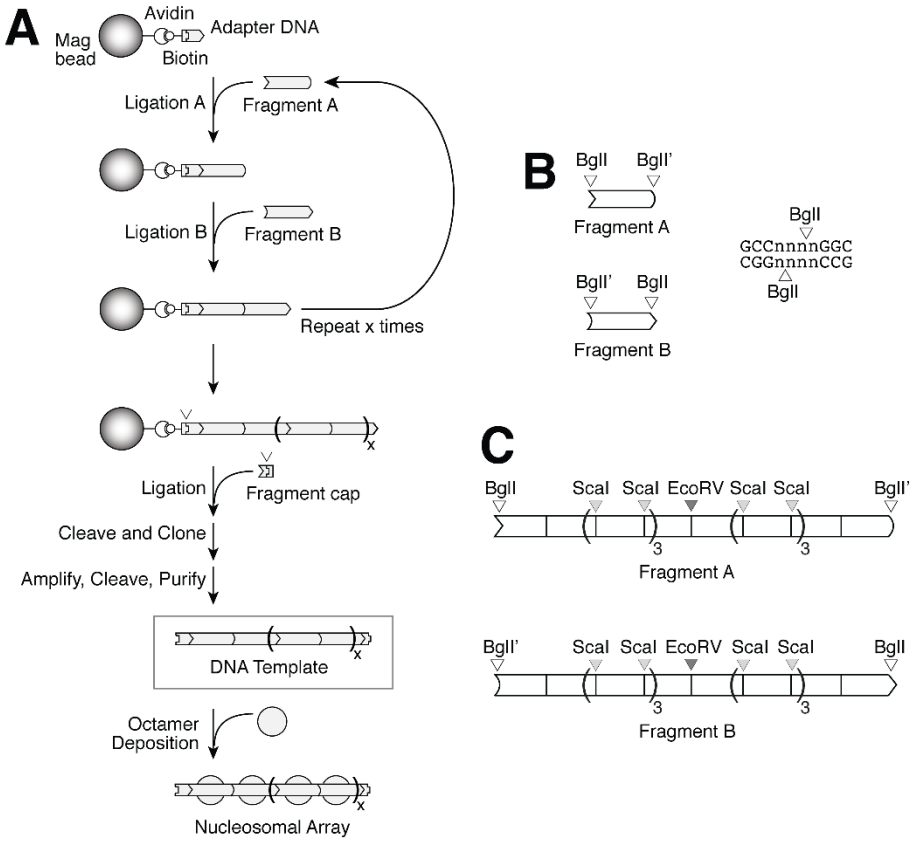

Figure 1. A) Solid-phase DNA fragment ligation strategy to synthesize DNA templates for chromatin model system assembly. Shown are ligation fragments containing: **B)** a single nucleosome binding site, or **C)** twelve binding sites.

Initial ligations were performed with a 172-base pair DNA template containing a single nucleosome positioning sequence, the standard 147-base pair 601 strong nucleosome positioning sequence (Figure 1B, left) (27). Different non-palindromic ends were generated by BglI digestion, as BglI recognizes sequences distal from its site of non-palindromic cleavage (Figure

1B, right). Under optimized conditions, we found that twelve rounds of ligation predominantly generated a single ligation product (Figure 2A, left lane), and DNA sequencing confirmed the desired twelve head-to-tail 172 fragment repeats. We found that at higher levels of bead substitution (Figure 2A, right lane), a greater amount of intermediate sized fragments was generated, presumably due to electrostatic repulsion near the bead reducing the efficiency of DNA extension. Sequencing of the expected 8-mer, 10-mer, and 12-mer product confirmed their identity, and showed that these conditions are useful for readily generating a range of different template sizes in one ligation reaction. The 12-mer template was assembled into chromatin model system and analyzed further. BglI digestion of the chromatin model system generated from our ligated 172-framents at sites between the nucleosome positioning sequences revealed that, as expected, relatively few of the positioning sites were free at a 0.9:1.0 ratio of histone octamers to nucleosome positioning sites (Figure 2B), and the digestion pattern was comparable to that for a chromatin model system assembled from an equivalent amount of the standard 12-mer 177 base-pair 601 repeat template (Figure 2B). Further confirmation that the chromatin model system generated from the ligated fragments was similar to the standard assembly template was demonstrated by size-exclusion chromatography with multi-angle light scattering (Figure 2C). The two species elute with very similar elution times, and have absolute molecular weights consistent with comparable saturations of the nucleosome positioning sites.

Ligations were also performed with fragments containing twelve 177-base pair repeats of the 601 nucleosome positioning sequence (Figure 1C). These fragments, like those containing a single nucleosome positioning site, have non-palindromic ends generated by BglI digestion. Additionally, because of the way they were generated, they contain a number of additional restriction enzyme cleavage sites between the individual positioning sequence. Using these

longer fragments, multiple rounds of ligation were performed (Figure 2D). Fragments representing 24, 36, and 48 nucleosome positioning sequences were generated and cloned. Like

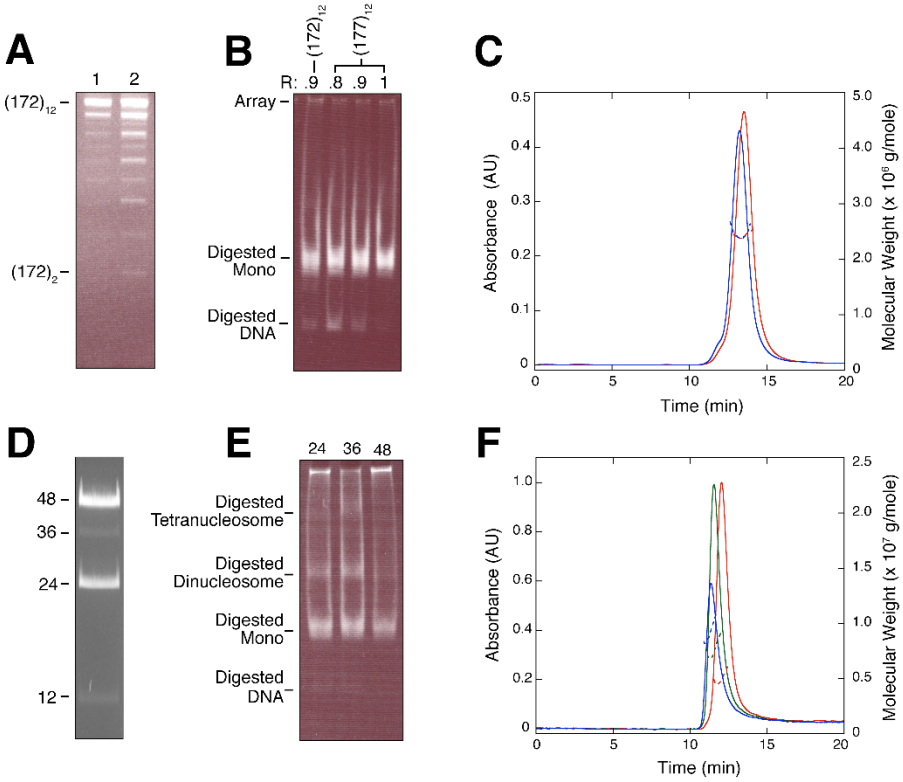

Figure 2. A) Gel analysis of twelve rounds of 172 bp fragment ligation. Synthesis was either optimized for full-length ligation product (lane 1), or production of all sizes of ligation intermediates (lane 2). **B)** Native gel analysis of digested nucleosomal arrays assembled either from the 12-mer DNA templates shown in 'A', or a standard 601-177-12 DNA template. R refers to the molar ratio of histone octamers to nucleosome binding sites. **C)** Size exclusion chromatography elution profiles of R=0.9 arrays from part 'B'. The red and blue traces correspond to arrays as-sembled onto 601-172-12 and 601-177-12 templates. Their respective absolute masses were measure to be 2.48 +/- 0.3 MDa and 2.49 +/- 0.6 MDa, correspond to arrays saturations of 11.1 and 10.9 nucleosomes/array. **D)** Gel anal-ysis of four rounds of 12-site fragment ligations optimized for generating dimeric and tetrameric products. **E)** Native gel analysis of digested nucleosomal arrays assembled from templates containing 24, 36, or 48 nucleosome binding sites. **F)** Size exclusion chromatography elution profiles of nucleosomal arrays assemble from templates containing 24 (red), 36 (blue), and 48 (green) nucleosome binding sites. The corresponding absolute molecular weights associ-ated with each elution peak are 5.16 +/- 0.20 MDa, 7.60 +/- .01 MDa, and 10.8 +/- 2.0 MDa, and correspond to 23.5, 34.0, and 51.0 nucleosomes/array.

the 12-mer array template described above, these templates could be assembled into chromatin model systems. Digestion with ScaI, which cuts between many, but not all, of the nucleosome

positioning sites, revealed relatively little free DNA (Figure 2E), indicating a high saturation of the DNA templates with histone octamer binding. Additionally, size exclusion chromatography with multi-angle light scattering analysis corroborate the high saturation of the model system (Figure 2F).

Altogether, these results show our solid-phase DNA fragment ligation strategy is an effective method for creating unique DNA templates for chromatin model systems, including those in which the number, identity, and placement of specific nucleosome positioning sites need to be carefully controlled. The ability to generate such templates will be useful for better understanding chromatin structure and function, and we expect that our technique will be generally useful for generating other DNA constructs containing repetitive sequences.

Acknowledgements

This work was supported by American Cancer Society Research Scholar Grant 1206501 to M. A. S.-K.

Bibliography

1. Bulger, M. (2005) Hyperacetylated Chromatin Domains: Lessons from Heterochromatin. *Journal of Biological Chemistry* **280**, 21689-21692

2. Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., and Workman, J. L. (2002) Function and Selectivity of Bromodomains in Anchoring Chromatin-Modifying Complexes to Promoter Nucleosomes. *Cell* **111**, 369-379

3. Li, S., and Shogren-Knaak, M. A. (2008) Cross-talk between histone H3 tails produces cooperative nucleosome acetylation. *Proceedings of the National Academy of Sciences* **105**, 18243-18248

4. Ringel, A. E., Cieniewicz, A. M., Taverna, S. D., and Wolberger, C. (2015) Nucleosome competition reveals processive acetylation by the SAGA HAT module. *Proceedings of the National Academy of Sciences* **112**, E5461-E5470

5. Mittal, C., Blacketer, M. J., and Shogren-Knaak, M. A. (2014) Nucleosome acetylation sequencing to study the establishment of chromatin acetylation. *Analytical Biochemistry* **457**, 51-58

6. Tse, C., Sera, T., Wolffe, A. P., and Hansen, J. C. (1998) Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III. *Molecular and Cellular Biology* **18**, 4629-4638

7. Widom, J. (1999) Equilibrium and Dynamic Nucleosome Stability. in *Chromatin Protocols* (Becker, P. B. ed.), Humana Press, Totowa, NJ. pp 61-77

8. Tóth, K., Brun, N., and Langowski, J. (2006) Chromatin Compaction at the Mononucleosome Level. *Biochemistry* **45**, 1591-1598

9. Gansen, A., Tóth, K., Schwarz, N., and Langowski, J. (2009) Structural Variability of Nucleosomes Detected by Single-Pair Förster Resonance Energy Transfer: Histone Acetylation, Sequence Variation, and Salt Effects. *The Journal of Physical Chemistry B* **113**, 2604-2613

10. Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* **12**, 46-53

11. Koopmans, W. J. A., Buning, R., Schmidt, T., and van Noort, J. (2009) spFRET Using Alternating Excitation and FCS Reveals Progressive DNA Unwrapping in Nucleosomes. *Biophysical Journal* **97**, 195-204

12. Andrews, A. J., and Luger, K. (2011) Nucleosome Structure(s) and Stability: Variations on a Theme. *Annual Review of Biophysics* **40**, 99-117

13. Bohm, V., Hieb, A. R., Andrews, A. J., Gansen, A., Rocker, A., Toth, K., Luger, K., and Langowski, J. (2011) Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic Acids Res* **39**, 3093-3102

14. Luo, Y., North, J. A., and Poirier, M. G. (2014) Single molecule fluorescence methodologies for investigating transcription factor binding kinetics to nucleosomes and DNA. *Methods* **70**, 108-118

15. Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J. W. (2009) A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Molecular Cell* **36**, 153-163

16. van Holde, K. E. (1988) *Chromatin*, Springer-Verlag, New York

17. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal Structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

18. Poirier, M. G., Bussiek, M., Langowski, J., and Widom, J. (2008) Spontaneous access to DNA target sites in folded chromatin fibers. *J Mol Biol* **379**, 772-786

19. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R. M., Zhu, P., and Li, G. (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380

20. Carruthers, L. M., Tse, C., Walker, K. P., 3rd, and Hansen, J. C. (1999) Assembly of defined nucleosomal and chromatin arrays from pure components. *Methods Enzymol* **304**, 19-35

21. Simpson, R. T., Thoma, F., and Brubaker, J. M. (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

22. Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T. J. (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol* **327**, 85-96

23. Mittal, C., Blacketer, M. J., and Shogren-Knaak, M. A. (2014) Nucleosome acetylation sequencing to study the establishment of chromatin acetylation. *Anal Biochem* **457**, 51-58

24. Azzaz, A. M., Vitalini, M. W., Thomas, A. S., Price, J. P., Blacketer, M. J., Cryderman, D. E., Zirbel, L. N., Woodcock, C. L., Elcock, A. H., Wallrath, L. L., and Shogren-Knaak, M. A. (2014) Human heterochromatin protein 1alpha promotes nucleosome associations that drive chromatin condensation. *J Biol Chem* **289**, 6850-6861

25. Blacketer, M. J., Feely, S. J., and Shogren-Knaak, M. A. (2010) Nucleosome interactions and stability in an ordered nucleosome array model system. *J Biol Chem* **285**, 34597-34607

26. Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999) Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**, 1-16

27. Anderson, J. D., and Widom, J. (2000) Sequence and Position-dependence of the Equilibrium Accessibility of Nucleosomal DNA Target Sites. *J Mol Biol* **296**, 979-987

Supplementary

Non-palindromic 601-172-1 precursor: PCR fragments (BglI sites bold, PstI sites underlined)

- Template: 601-177-1 fragment (1)

- 172BglIF Primer: 5’-GCATGCTGCA**GCCAGAATGGC**TACATGCACAGGATGTATATATC-3’

• 172Bgl1IR Primer: 5’-GCATGCTGCA**GCCACCTTGGC**GGCCGCCCTGGAGAATCCCG-3’

• 172’BglIF Primer: 5’-GCATGCTGCA**GCCAAGGTGGC**TACATGCACAGGATGTATATATC-3’

• 172’Bgl1IR Primer: 5’-GCATGCTGCA**GCCATTCTGGC**GGCCGCCCTGGAGAATCCCG-3’

Non-palindromic 601-172-1 plasmids

• PstI digest PCR fragment and pRS315 plasmid, purify, and then ligate

---

Non-palindromic 601-177-12 precursor: 601-177-2 plasmid

• ScaI digest 601-177-12 plasmid to release all but two 177-1 sequences, purify and religate

Non-palindromic 601-177-12 precursor: 601-177-3 fragment

• Partial ScaI digestion of 601-177-12 plasmid, and isolation of 601-177-3 fragment

Non-palindromic 601-177-12 precursor: 601-177-5 fragment

• ScaI digest 601-177-2 plasmid and ligate 601-177-3 fragment
• Digest product with EcoRV

Non-palindromic 601-177-12 precursor: Non-palindromic 601-177-2 plasmid (BglI sites bold, PstI sites underlined)

• Template: 601-177-2 plasmid

- 177BglIF Primer: 5'-GCATGCTGCA**GCCAGAATGGC**TCTTACATGCACAGGATGTATATATC-3'
- 177BglIR Primer: 5'-GCATGCTGCA**GCCACCTTGGC**GCGGCCGCCCTGGAGAATCCCG-3'

- 177BglF' Primer: 5'-GCATGCTGCA**GCCAAGGTGGC**TCTTACATGCACAGGATGTATATATC-3'
- 177BglIR'Primer: 5'-GCATGCTGCA**GCCATTCTGGC**GCGGCCGCCCTGGAGAATCCCG-3'

- PstI digest PCR product and p601X (2), purify, and then ligate

Non-palindromic 601-177-12 plasmid:

- ScaI digest non-palindromic 601-177-2 plasmid and ligate with 601-177-5 fragment.

  Screen for constructs with two 601-177-5 fragment insertions

---

Adapter fragment (Pcb refers to photocleavable biotin, Pho refers to phosphorylated. PstI sites underlined. BglI compatible overhang in bold):

- ADT-BglIF': 5'-Pcb-GGCCGCCTGCAGATATC**GAA**-3'
- ADT-BglIR': 5'-Pho-GATATCTGCAGGCGGCC-3'.

Cap fragments (BamHI sites underlined. BglI and BglI' compatible overhangs in bold):

- CAP-BglIF: 5'-GATATCGGATCCAATTAT-3'
- CAP-BglIR: 5'-ATAATTGGATCCGATATC**TTC**-3'

- CAP-BglIF': 5'-GATATCGGATCCAATTAT-3'

• CAP-BglIR': 5'-ATAATTGGATCCGATATC**CCT**-3'

1. Blacketer, M. J., Feely, S. J., and Shogren-Knaak, M. A. (2010) Nucleosome interactions and stability in an ordered nucleosome array model system. *J Biol Chem* **285**, 34597-34607

2. Azzaz, A. M., Vitalini, M. W., Thomas, A. S., Price, J. P., Blacketer, M. J., Cryderman, D. E., Zirbel, L. N., Woodcock, C. L., Elcock, A. H., Wallrath, L. L., and Shogren-Knaak, M. A. (2014) Human heterochromatin protein 1alpha promotes nucleosome associations that drive chromatin condensation. *J Biol Chem* **289**, 6850-6861