NOTE TO USERS

DISSERTATION

STRUCTURAL AND FUNCTIONAL EFFECTS OF HISTONE VARIANTS ON THE NUCLEOSOME CORE PARTICLE

Submitted by

Srinivas Chakravarthy

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO

Fall 2004

COLORADO STATE UNIVERSITY

November 8, 2004

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY SRINIVAS CHAKRAVARTHY ENTITLED "STRUCTURAL AND FUNCTIONAL EFFECTS OF HISTONE VARIANTS ON THE NUCLEOSOME CORE PARTICLE" BE ACCEPTED AS FULFILLING IN PART THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

(Oren P. Anderson)

(Jennifer Nyborg)

(Olve Peersen)

(Laurie A. Stargell)

(Karolin Luger)
Advisor

(Marvin Paule)
Department Head

ABSTRACT OF DISSERTATION

Structural and Functional Effects of Histone Variants on the Nucleosome Core Particle

DNA in eukaryotes is packed in association with a roughly equal mass of histone proteins to form a nucleo-protein complex known as chromatin. The fundamental unit of chromatin is the nucleosome core particle, which consists of two copies each of the four core histones H2A, H2B, H3 and H4 wrapped ~1.65 times by ~146 base pairs of DNA. In addition to compaction of DNA, the nucleosome core particle also plays an important role in regulating gene expression. It is the primary determinant of DNA accessibility in key cellular processes such as transcription, replication and repair. Several covalent modifications of the core histones are known to influence the structure and function of a nucleosome. However replacing one or more of the core histones with the corresponding histone variants can also vary the biochemical composition of the nucleosome. MacroH2A is a histone H2A variant that preferentially localizes to the inactive X-chromosome of adult female mammals. Unlike other core histones, macroH2A has a tripartite structural organization that consists of a histone domain and a non-histone domain connected by a linker domain. We have studied the structural and functional consequences of incorporation of the histone domain of macroH2A into nucleosome core particles. We show that the nucleosome structure remains unaffected by the sequence changes in most regions of the histone fold. The L1-loop of macroH2A is the only region that shows considerable structural differences compared to that of major type H2A. We show that this region is also responsible for the anomalies in the biochemical behavior of nucleosome core particles containing macroH2A (macro-NCPs). It has also been

established in studies inspired by our structural and biochemical insights that the L1-loop is sufficient for the *in vivo* targeting of macroH2A. We have determined the 1.6Å crystal structure of the non-histone domain of macroH2A. We find that it is an α/β fold that contains a 7 stranded β-sheet and 5 α-helices. It is also remarkably devoid of distinct features on the surface of the protein and the charge distribution is unusually neutral for a chromatin-associated protein. We have also investigated the stoichiometry of nucleosomes containing different histone variants. We show that different histone variants show different propensities to form hybrid nucleosomes (nucleosomes containing one variant and one non-variant histone). MacroH2A in fact 'prefers' to form hybrid nucleosomes *in vitro.* The stoichiometry of variant nucleosomes *in vivo* may depend on several factors such as local concentrations but this brings forth yet another potential level of structural and functional heterogeneity.

Srinivas Chakravarthy
Department of Biochemistry
And Molecular Biology,
Colorado State University,
Fort Collins, CO, 80523.
Fall 2004

ACKNOWLEDGEMENTS

The body of work presented in this thesis was made possible by the kind assistance from a large number of people in and outside the department. This department is uniquely close knit and thrives on collaborative effort between different labs and their respective areas of expertise. I have thoroughly enjoyed working in an environment that has been extremely conducive to my scientific growth.

My guide Dr. Karolin Luger has been largely instrumental in any success I may report in my scientific endeavors. She has been a constant source of encouragement and support in addition to being extremely resourceful and creative. Her exemplary work ethic and enthusiasm for science are infectious. For the best possible four years as a Ph.D student she has my heartfelt gratitude.

I would also like to thank Dr. Oren Anderson, Dr. Jennifer Nyborg, Dr. Olve Peersen, and Dr. Laurie Stargell for serving on my graduate committee. Their encouragement, ideas, and constructive criticism have contributed significantly to developing and improving this project. I would also like to thank them for being extremely patient with my mistakes.

I would specifically like to thank Dr. Oren Anderson for lending his crystallographic expertise and valuable time. Dr. Olve Peersen has been of inestimable help in solving crystallographic and computer related problems. Discussions with him regarding different aspects of crystallography and science in general are always stimulating and edifying. I thank Dr. Jennifer Nyborg for her support, and her dedication to research and teaching, which have been a source of inspiration. It has always been a pleasure to discuss the different modalities of my project with Dr. Laurie Stargell, who epitomizes

the combination of scientific acumen and patience. I thank her for being approachable and a gifted teacher. I would like to thank Dr. Robert Woody for his support, encouragement and readiness to lend a different perspective to my project.

I have been fortunate to work with a wonderful set of colleagues. None of the work presented here would be possible without their help and it has been a pleasure working with them. I especially thank Pam Dyer who has taken on the intimidating responsibility of managing the lab and I admire the effortlessness with which she executes her job.

Last but not the least I would like to thank my family and friends whose unflinching faith in me constantly pushes me to improve myself. My wife Maya's unconditional love and tolerance for my unpredictability make life eminently bearable and the joy it is. She remains the one "unrelenting" source of happiness in my life. My parents Venkataramana Rao and Madhavi Rao have taught me the value of hard work and sincerity. They will always represent "the" most important and stable aspect of my life and for their love, friendship and immense faith in me, I will be eternally grateful to them.

TABLE OF CONTENTS

CHAPTER 1

Literature Review

1.1 Introduction

DNA in eukaryotes is packed in association with an equal mass of proteins to form a nucleo-protein complex popularly known as chromatin (Van Holde, 1988). The fundamental structural and functional unit of chromatin is the nucleosome core particle (NCP). The proteins primarily responsible for the formation of this complex are called histones. An NCP consists of a protein core made of two copies each of four types of histones H2A, H2B, H3, and H4 wrapped ~1.65 times by 147 base pairs of DNA (Luger et al., 1997a; Luger et al., 1997b). Arrays of nucleosomes arranged in a beads on a string type of structure are the first level of compaction eukaryotic DNA undergoes (for review see Hansen, 2002). Several relatively less understood subsequent levels of compaction that involve the linker histone H1 and non-histone chromatin associated proteins, lead to the formation of the metaphase chromosome. However, chromatin is a dynamic physical entity and undergoes constant inter-conversion between different states of compaction (Horn and Peterson, 2002). In addition to packing a disproportionately large amount of DNA in eukaryotic nuclei, chromatin also regulates DNA-dependent cellular processes like transcription, replication, and DNA repair. Understanding the structural organization of chromatin is key to elucidating its role in the regulation of these DNA-dependent nuclear processes.

The past decade has produced an enormous amount of structural information on the nucleosome. Several high-resolution crystal structures of nucleosomes containing histones from different species have thrown light on the intricacies of the protein-protein

and protein-DNA interactions that stabilize nucleosomes (Davey and Richmond, 2002; Davey et al., 2002; Luger et al., 1997a; Richmond and Davey, 2003a; Suto et al., 2000; White et al., 2001). It is becoming increasingly evident that nucleosome structure is inordinately accommodating to variations in amino acid sequence, and biological context.

1.2 Nucleosome Structure

The first high-resolution structure of a nucleosome was made of recombinant histones from *Xenopus laevis* and a 146 bp palindromic DNA strand derived from the human α-satellite sequence (Luger et al., 1997a) (Fig. 1.1). This DNA is wrapped in ~1.65 turns of a flat, left-handed superhelix around the histone octamer made of the four core histones H2A, H2B, H3, and H4. Core histones have been remarkably well conserved in length and amino acid sequence through evolution (Sullivan et al., 2000). Histones H3 and H4 are most conserved and H2A and H2B are relatively less conserved (Sullivan et al., 2002a). All core histones are small (11 to 16kDa) basic proteins with an abundance of lysine and arginine residues. The core histones have a structural motif called the histone fold that mediates several protein-DNA and protein-protein interactions. The histone fold has a large α-helix flanked by two smaller helices and connected by two loop regions (commonly depicted as α1-L1-α2-L2-α3) (Arents et al., 1991; Arents and Moudrianakis, 1995; Luger et al., 1997a) (Fig. 1.2). These regions form crescent shaped heterodimers in the pairings H3-H4 and H2A-H2B. The protein octamer is divided into four histone-fold dimers defined by H3-H4 and H2A-H2B histone pairs (Fig 1.2). The two H3-H4 pairs interact through a 4-helix bundle formed from the two H3 histone folds involved thus defining the $(H3\text{-}H4)_2$ tetramer (Fig. 1.1).

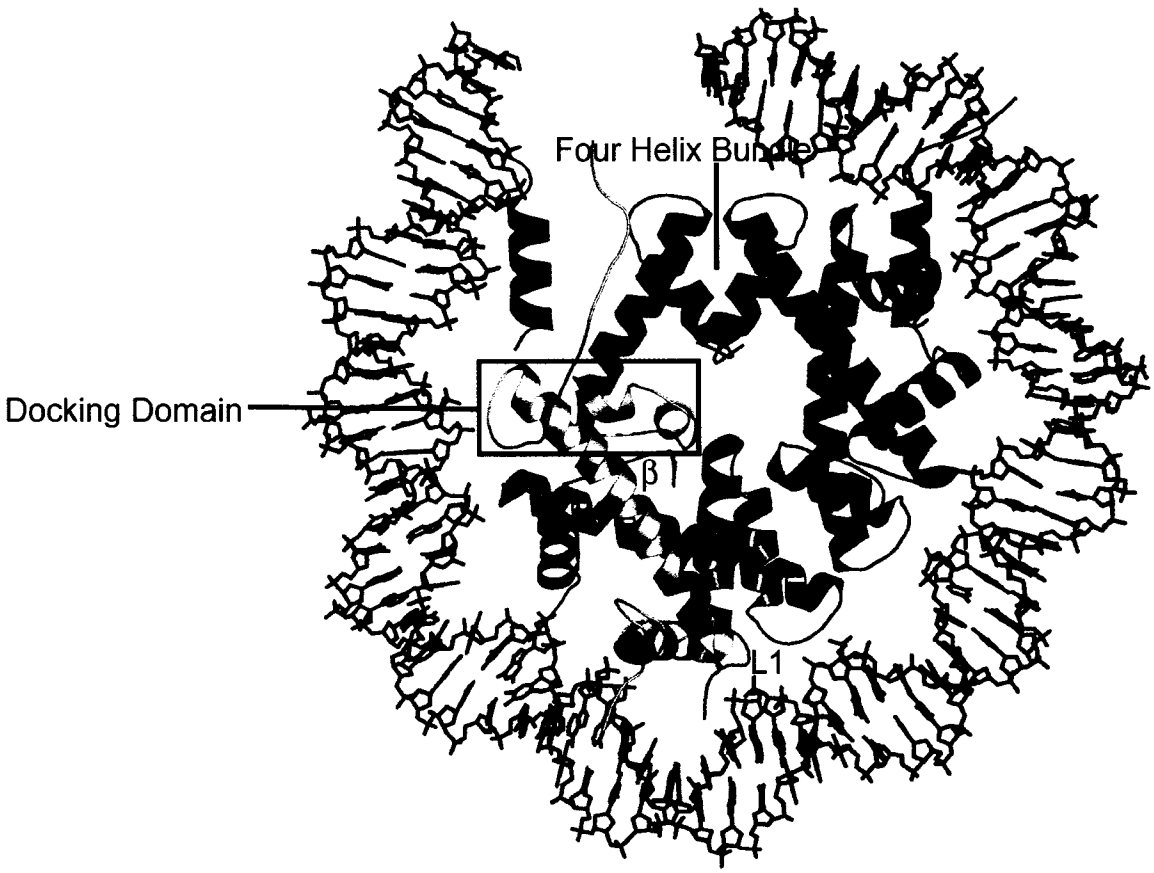


Figure 1.1 Overview of Nucleosome Structure. *Yellow*: H2A; *red*: H2B; *blue*: H3; *green*: H4; *teal*: DNA. Only 73 bp of DNA and associated proteins are shown. Important stabilizing features are also indicated (docking domain, the four helix bundle, the L1-loop of H2A and the β interaction between the docking domain and H4).

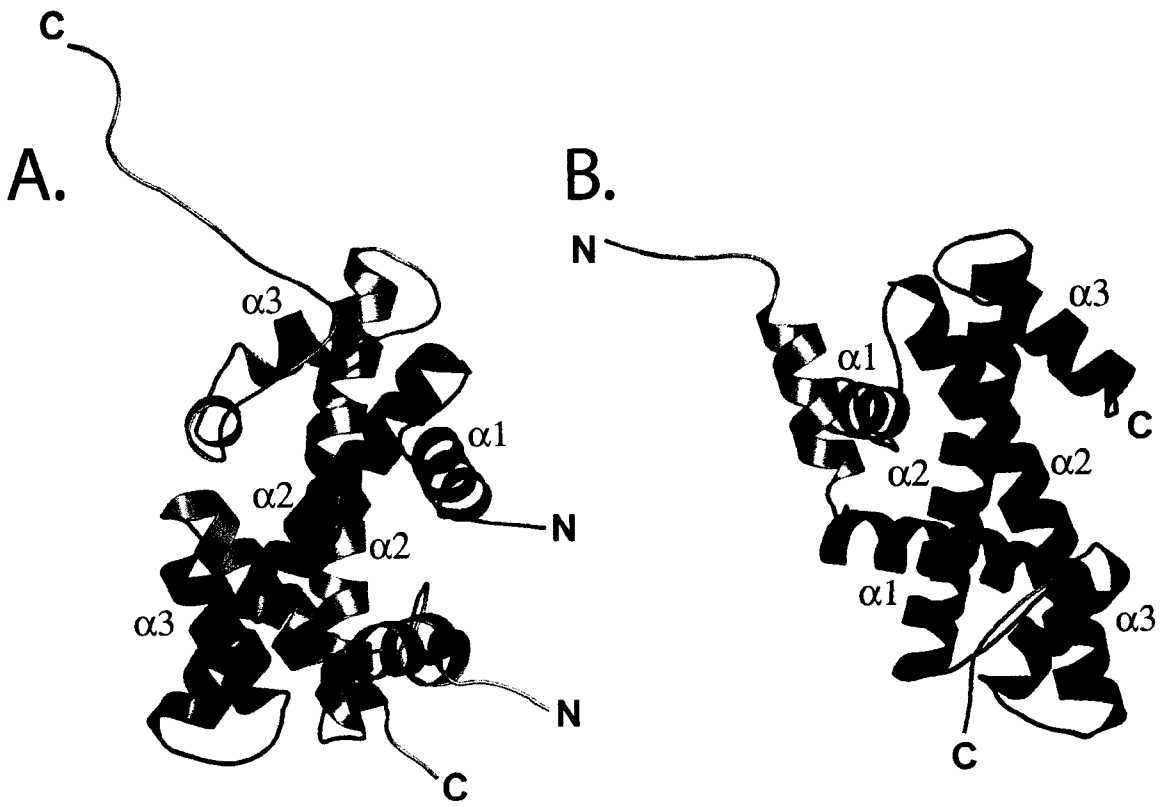


Figure 1.2. All core histones have a conserved characteristic fold called the histone fold. **A.** Histone H2A-H2B heterodimer. *Yellow*: H2A; *red*: H2B. All the components of the histone fold are indicated (α1, α2, and α3, L1 and L2). **B.** Histone H3 and H4 heterodimer. *Blue*: H3; *green*: H4.

Each H2A-H2B pair interacts with the tetramer through another 4-helix bundle between H2B and H4 histone folds. The two H2A-H2B dimers interact with each other only in the presence of DNA via a small interface between the L1-loops of the two H2A moieties (Luger et al., 1997a). Extensions of the histone domains (both helices and coil domains) also form integral parts of the protein core of the nucleosome.

The crescent shaped heterodimers formed by the central histone-folds bind ~30 bp of the DNA double helix. Binding is primarily to the DNA-phosphodiester backbones as they face the protein over 2.5 turns of the double helix. Several kinds of DNA-protein interactions take place whenever the DNA backbone is facing the histone core. These include: hydrogen bonds to phosphates made from main chain amide nitrogen atoms of amino acids; arginine side chains from the histone fold enter the minor groove 10 of the 14 times it faces the histone octamer and the other four times the arginine side chain is from the tail regions; non-polar contacts made with the deoxyribose groups; hydrogen bonds and salt links occurring between DNA-phosphate oxygen atoms and protein basic and hydroxyl chain groups.

The H3 αN helix extension of the histone fold and the preceding tail region in the minor grooves are responsible for binding the 13 bp of DNA at each terminus of the superhelix. A section of the H2A C-terminus runs antiparallel to the αN and further links it to the underlying H3-H4 histone fold domains. Preceding this region of the H2A C-terminus, the H2A tail (aa 92-108) folds into a docking domain with the H2A α3 helix (Fig. 1.1). This buried domain contains a short α-helix and makes β-sheet interactions with the short H4' C-terminal tail folded back over its α3 helix. These extensions are apparently

necessary to stabilize the helical ramp formed by the histone folds. The N-terminal tail of one of the two H4 moieties also plays a major role in stabilizing the crystal lattice of the nucleosome. One of the two very basic H4 N-termini interacts with what is known as an acidic patch contributed to by H2A (6 amino acids) and H2B (1 amino acid) of the adjacent nucleosome (Luger et al., 1997a). While the importance of this interaction *in vivo* is not known, it is essential for crystallization.

The parameters of the double helix show substantial variation over the length of the superhelix. The double helix has an average diameter of 41.8 Å but is not uniformly bent. The overall twist is 10.2 bp per turn in the local reference frame (independent of superhelical pitch). Dividing the DNA into segments delimited by the arginine side chains inserted into the minor groove, the twist value ranges from a minimum of 9.4 bp per turn to a maximum of 10.9 (Luger et al., 1997a; Richmond and Davey, 2003a). The overall overwinding of the superhelical DNA by 0.3 bp compared to linear free DNA creates an alignment of minor grooves between superhelical gyres that allows an essentially straight passage for the H3 and H2B tails through the channels formed. This alignment creates seven minor and six major DNA "supergrooves" in which 14-16 bp of DNA would be accessible for site-specific interaction with a single DNA binding ligand. Each supergroove brings together sequence elements that are 80 bp apart in linear DNA into close structural proximity. These supergrooves may be molecular interaction platforms that can be exploited for synergistic protein-protein interactions (Edayathumangalam et al., 2004).

A detailed picture of structured water at the protein-DNA interface was obtained from the 1.9 Å structure of an NCP that had a 147 bp (as opposed to 146 bp) fragment of DNA

derived from the human α satellite sequence (Davey et al., 2002). The overall structure of NCP146 and NCP147 are virtually indistinguishable. The only difference is that the two halves of NCP147 are related by a dyad axis that passes through the central base pair whereas in NCP146 it passes through one of the two central base pairs, therefore dividing the DNA into two unequal halves (one 73 bp and the other 72 bp long). This results in an overwinding and stretching of the 72 bp half in NCP 146. The absence of this distortion in NCP147 leads to a much improved electron density map for the DNA. This structure shows that water molecules not only contribute significantly to the stability of DNA-binding, but take part in adapting the histone surface to conformational variation in DNA. In other words, the bridging water may have a significant role in facilitating nucleosome mobility by providing an interaction pathway for changing phosphate group position.

High resolution structures of nucleosomes made of the 146 base pair α-satellite DNA and histones from other species have since been determined and it is found that the global features of the nucleosome structure are faithfully preserved through evolution in spite of sequence variations. The *Sce*-NCP (made of histones from *S.cereviciae*) has sequence differences distributed through out the histones, which are located both on the surface of the histone octamer and buried deep within the nucleosome structure (White et al., 2001). The most significant changes in the protein – protein interactions within the yeast nucleosome core particle are located in the H2A L1 loops where the two H2A – H2B dimers interact. This region is found to be destabilized compared to that of the *Xle*-NCP (White et al., 2001). It was also demonstrated that small changes in the surface of the nucleosome caused by minor amino acid sequence differences could alter inter-nucleosomal interactions. We have also determined the structure of a nucleosome core

particle containing major type core histones from *Drosophila melanogaster* (*Dme*-NCP) (unpublished data; see chapter 6).

In addition to nucleosomes from different species we have also performed structural studies on nucleosomes with core histones replaced by histone variants. The 2.6 Å crystal structure of an H2A.Z-NCP was determined in the lab (Suto et al., 2000). It is found again that globally the nucleosome structure can be superimposed on the Xla-NCP with an rmsd <1. There are however two regions in the H2A histone-fold that differ significantly in comparison with the non-variant NCP, the L1-loop and the docking domain. The docking domain mediates interactions between the H2A-H2B dimer and the $(H3\text{-}H4)_2$ tetramer and the L1 loops of the two H2A moieties form the only interface between the two H2A-H2B dimers in the nucleosome. The structural differences alluded to a potential weakening of the interaction between the H2A-H2B dimer and the H3-H4 tetramer, but the overall stability of the nucleosome may not be affected. FRET (fluorescence resonance energy transfer) was used to determine that the H2A.Z–NCPs are relatively more stable than major type nucleosomes under increasing salt concentrations (Park et al., 2004a).

Part of this thesis will focus on the structural studies done with nucleosome core particles in which the histone domain of macroH2A replaced major type H2A (macro-NCPs). We find that yet again the L1-loop of macroH2A is the site of major structural upheaval. Also noteworthy is the fact that macro-NCPs are made of the histone domain of human macroH2A, and the other core histones (H2B, H3, and H4) from mouse. This is the first structural study on a mammalian nucleosome. Comparison of nucleosomes from *M.musculus* (mammal), and *D.melanogaster* (fruit fly) with those from *X.laevis*

(amphibian) and *S.cerevisiae* (yeast) (Luger et al., 1997a; White et al., 2001) shows that the nucleosome structure is remarkably well preserved across evolution.

1.3 Chromatin and Transcription

Compaction of eukaryotic DNA into chromatin is a prerequisite for packing in excess of a meter of DNA into the extremely restricted volume of the eukaryotic nucleus. Such a high degree of compaction leads to a general repression of transcription. Although substantial progress has been made towards understanding *in vivo* chromatin structure, a considerable amount of detail encompassing the "higher order" structure remains by and large uncharacterized. Analytical ultracentrifugation analysis with model nucleosome arrays reconstituted with purified histones and a DNA template composed of 12 tandem repeats of a 208 bp nucleosome positioning sequence revealed that in low ionic strength conditions these arrays exist as extended "beads on a string" type of structure (for review see Hansen, 2002) (Fig. 1.3). The fiber assumes progressively tighter conformations with increasing bivalent cation concentration. Interactions within the fiber lead to formation of the 30nm fiber and inter-fiber interactions induce the formation of structures of higher compaction (Fig. 1.3). These folding steps need the N-terminal tails of the histones (Carruthers et al., 1998; Hansen, 2002).

In addition to the four core histones there is a fifth histone called the linker histone. The most common linker histone is called H1. The linker histone is highly rich in lysine and therefore very basic. It is slightly larger than core histones (>20 kDa). Linker histones are the easiest to dissociate from DNA with increasing salt concentrations. It has a central structured winged helix domain and highly charged tails at both the N and C-

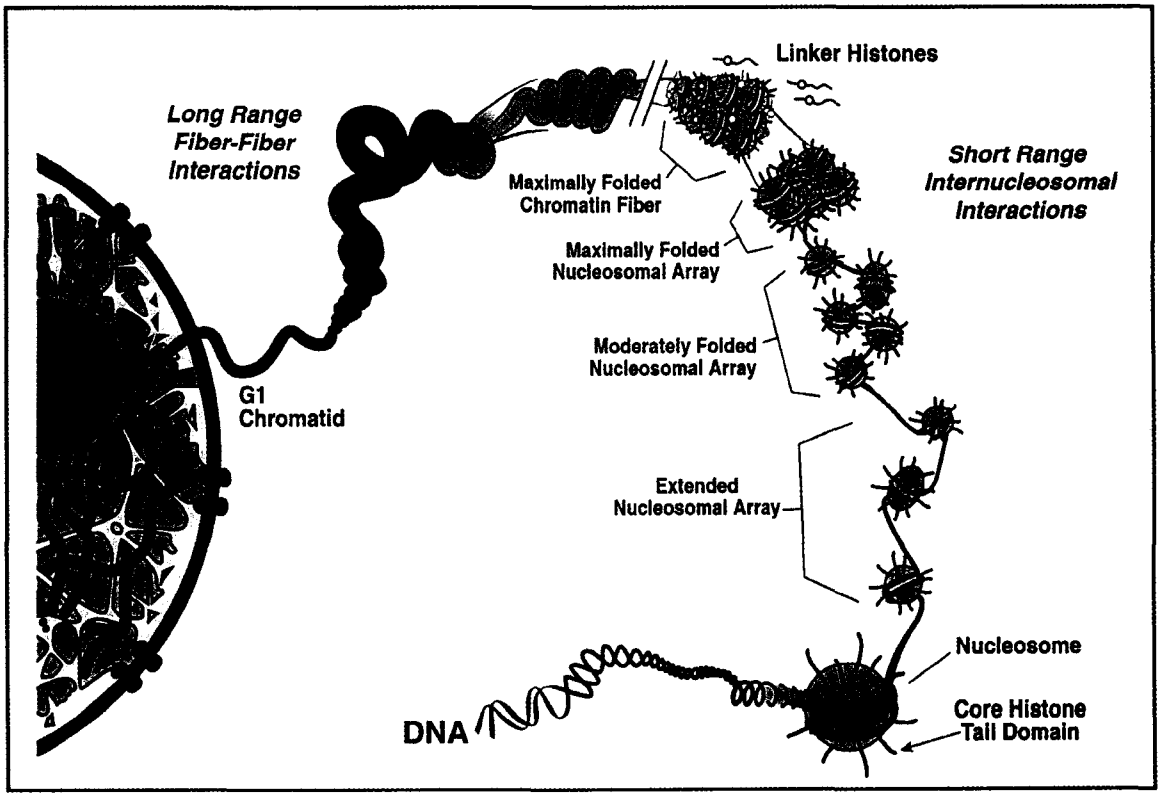


Figure 1.3 Schematic showing several known and predicted levels of compaction that are involved in formation of chromatin Hansen JC; (2002), Annu Rev Biophys Biomol Struct. 31:361-92.

termini. The central domain contains three α-helices attached to a three-stranded β-sheet (Ramakrishnan et al., 1993). The structured domain of the linker histone associates with the nucleosome core and in addition the tail regions interact with the linker DNA between nucleosomes. There is considerable diversity in the structure of linker histones in different lineages. This may lead to significant differences in the way chromatin is folded into higher order structures. The linker histone precipitates the folding process and leads to a more homogeneous mixture of highly compacted fibers (Carruthers et al., 1998). Chromatin fibers thus folded may form the infrastructure for higher degrees of chromatin compaction (Fig. 1.3). Considering the divalent ion concentrations needed to induce the 30 nm fiber state *in vitro* (1 to 2 mM) (for review see Hansen, 2002), it would be rational to expect chromatin to be highly compacted under physiological divalent ion concentrations (4-6 mM Ca^{2+}, and 2-4 mM Mg^{2+}) (Strick et al., 2001).

The transcriptional machinery may therefore have to unravel a highly compacted structure to enable gene expression. The recognition of chromatin structure as an influencing factor on nuclear events such as transcription is a relatively recent event. There exist wide variations in the degrees of chromatin structure dependent activity or repression. The transcriptional activity of a particular region of the genome may be dependent upon the position of genes within accessible (euchromatic) or an inaccessible (heterochromatic) chromatin environment. Multiple strategies have evolved to optimize the use of chromatin as a substrate for transcription (or any other DNA-directed process such as replication). Several factors contribute to the extremely fine-tuned regulation of gene-expression. Post-translational covalent modifications of the amino-terminal tails of histones for example alter the biochemical properties of the chromatin to provide signals

to which other factors can respond (histone code) (for review see Jenuwein and Allis, 2001). Posttranslational modifications of the core histones include phosphorylation, methylation, and ubiquitination in addition to the extensively studied acetylation. In addition nucleosomes can be remodeled by ATP-dependent enzymes to facilitate the interaction of transcription factors with the DNA template.

Alterations in the core histone acetylation status are found to have a role in determining transcriptional activity in chromatin (Wolffe, 1998). Being extremely lysine rich, the N-terminal tails of the core histones are suitable targets for acetylation. Chromatin immunoprecipitation studies have indicated that the histones hyper-acetylated at certain sites of particular histones in combination with a specific set of other modifications are prevalent in regions of the genome that show high transcriptional activity (Hebbes et al., 1994; Hebbes et al., 1992). Histone acetylation and a concomitant increase in the sensitivity to nucleases are now acknowledged markers of transcriptionally active chromatin (Hebbes et al., 1994; Struhl, 1998). Acetylation neutralizes the lysine residues in the amino-termini that may weaken the DNA – histone interactions and inter-nucleosomal interactions thus making the chromatin more susceptible to the machinations of the transcriptional apparatus. Conversely hypoacetylated chromatin is found to be transcriptionally inert (Taunton et al., 1996). Histone acetyltransferases (HATs) and histone deacetylases (HDACs) of different substrate specificity are now known (Roth et al., 2001). Among HATs for example, Gcn5 under some circumstances shows a preference for H3 over H4 but CBP/p300 on the other hand is indiscriminate and is equally efficient in acetylating all four histones (Schiltz et al., 1999). In addition to

directly loosening chromatin structure, acetylation also influences the binding affinity of chromatin for non-histone proteins. This may be facilitated either by exposure of nucleosomal DNA or by direct modulation of protein-protein interactions (Edmondson et al., 1996; Workman and Kingston, 1998).

Methylation is also emerging as an important factor in modulating chromatin structure. There are several proteins that have histone methyl transferase (HMTase) activity (SUV39H1 – human, and Clr4 – fission yeast for instance) (for review see Jenuwein and Allis, 2001). Lysines (especially 4, 9, and 27, from histone H3) can be mono-, di-, and tri- methylated adding another level of complexity. The methylation of lys 9 in H3 is known to be instrumental in the recruitment of heterochromatin protein 1 (HP1) (Lachner et al., 2001; Nakayama et al., 2001). Methylated lys 9 of histone H3 is in fact an accepted marker of transcriptionally silent chromatin (Nakayama et al., 2001). Also noteworthy is the intricate interplay between different kinds of modifications, various permutations and combinations of which engender structural alterations in chromatin. Methylation of lys 9 evidently inhibits phosphorylation at ser 10 and is inhibited itself by acetylation at lys 9 (Rea et al., 2000). Phosphorylation at ser 10 is known to facilitate acetylation of lys 14 of H3 (Cheung et al., 2000; Lo et al., 2000).

The mechanism by which phosphorylation regulates transcription is not well known. It is possible that the negatively charged phosphate groups neutralize the basic histone tails and reduce their affinity for DNA. Phosphorylation is also seen to stimulate HAT activity in some cases (Cheung et al., 2000; Lo et al., 2000). It is therefore likely that phosphorylation in combination with acetylation may destabilize the higher order compaction of chromatin. Phosphorylation of the C-terminal tail of the histone H2A

variant H2A.X is key for double-stranded DNA-repair (Rogakou et al., 1998). In this case the phosphorylation is important for the recruitment of repair proteins (Celeste et al., 2003).

Chromatin remodeling factors mobilize and reposition nucleosomes and have been known to expose DNA sequences to nuclease digestion and are also known to generate superhelical torsion in DNA (Flaus and Owen-Hughes, 2001; Kingston and Narlikar, 1999). The chromatin remodeling complexes exhibit a diverse range of compositions and functions. Examples of some of the known families of remodeling complexes are SNF2, ISWI, RAD54, INO80 and several others (Lusser and Kadonaga, 2003). The SWI/SNF complex, which is in the SNF2 family, is an extensively studied chromatin-remodeling complex. It is a ~2Mda multi-protein complex among the primary components of which is the swi2/snf2 ATPase subunit. An ATPase subunit is present in most of the complexes mentioned above and they have been found to be capable of remodeling activity on their own (Phelan et al., 1999). Recent studies have corroborated one of these subunits' (ISWI) involvement in ATP-dependent DNA translocation (Whitehouse et al., 2003). While many of these proteins do directly affect nucleosome positioning, it is seen that nucleosomes can also be thermally mobilized (Flaus and Owen-Hughes, 2003). It is therefore possible that these ATP-dependent translocases are merely accelerating a natural response to the turbulence chromatin experiences during transcription. The non-ATPase subunits like INI1, BAF155, BAF170 subunits of the SWI/SNF complex may either influence the motor activity of the ATPase (Phelan et al., 1999) or may be instrumental in recruitment of the complexes via interactions with sequence specific transcription factors [e.g. NURF301 of the NURF complex (Xiao et al., 2001)].

Two models are gaining popularity for explaining the mechanism of action of nucleosome remodeling factors. 1. Twist diffusion entails propagation of small local alterations from mean DNA twist around the nucleosome (reviewed by Flaus and Owen-Hughes, 2004; Lusser and Kadonaga, 2003). 2. Bulge diffusion involves the unraveling of DNA from the nucleosome and re-establishment of the same protein – DNA contacts at a distal position (Flaus and Owen-Hughes, 2004). This model is based on the assumption that excess DNA is available for this purpose, which is looped out in a 'bulge'. It is also becoming increasingly evident that remodeling of nucleosomes often entails removal of one of the H2A-H2B dimers (Flaus and Owen-Hughes, 2004). This might destabilize the nucleosome and may result in the removal of the whole nucleosome as suggested in case of the SWI/SNF complex in recent studies (Boeger et al., 2003; Boeger et al., 2004). Since there is evidence to support several models it is possible that particular remodeling machines use different mechanisms to regulate the expression of specific genes.

1.4 Histone Variants

It is possible to exert influence on nucleosome structure by altering the biochemical composition of the nucleosome protein core. This may be achieved by substitution of one or more of the core histones with the corresponding histone variants (for review see Brown, 2001; Malik and Henikoff, 2003). Although histone genes are present in varying number of copies depending on the species, some histones have non-allelic variants that differ from them in the protein sequence. These variations in sequence can at least in principle alter nucleosome structure and higher order structure organization. The bulk of histones in eukaryotic cells are S-phase specific – i.e. they are deposited during DNA

replication to pack the new DNA that is produced. The genes encoding these variants are found on the outside of the gene clusters of major histones and their synthesis is usually replication-independent and therefore not restricted to the S-phase.

Almost all histone variants are found in special regions of the genome that show unique higher order structure and levels of transcriptional activity. The differential regulation of transcriptional activity could be a direct effect of the histone variants on the local chromatin structure or the consequence of an altered set of covalent modifications on the histones or a combination of both. Most if not all known histone variants are sequence variations of the core histones H3 and H2A (Chakravarthy et al., 2004).

Both H3 and H4 are very well conserved evolutionarily but distinct variants of H3 have emerged for special roles in transcription and even chromosome segregation. H4, on the other hand has remained constant throughout eukaryotic evolution. H3 has an important role to play in stabilizing the nucleosome. The C-termini of the two H3 moieties in the nucleosome are involved in forming the four-helix bundle that results in the formation of the H3-H4 tetramer (Luger et al., 1997a). In addition H3 is also involved in interactions with H2A and has a few important contacts with the DNA (Luger et al., 1997a).

The known H3 variants are CenH3 and H3.3. CenH3 is the centromere specific H3 variant, which is exclusively found in the centromeric chromatin (Palmer et al., 1991). It is essential for centromere assembly (Black et al., 2004a; Howman et al., 2000; Vermaak et al., 2002). Mitotic division in eukaryotes is dependent on the presence of centromeres, which mediate chromosomal attachment to the spindle fibers and therefore facilitate their correct segregation. CenH3s are only 50% identical to the major type H3 in the histone

fold domain and its N-terminal tail while showing no sequence similarity with that of major type H3 varies in length from 20 to ~200 amino acids in different species (reviewed by Malik and Henikoff, 2003). Also remarkable is the fact that CenH3s are relatively a lot less conserved between species than canonical H3 is. This rapid evolution seems to be driven by a sequence specificity in DNA – binding by the N-terminal tail and the L1-loop (Malik and Henikoff, 2001; Malik and Henikoff, 2003). There are however some consistent structural features in CenH3s from different species like the length of the L1-loop, which is much longer than the L1-loop of canonical H3.

H3.3 is a replacement *H3* variant encoded by a gene that is devoid of the transcriptional, and post-transcriptional controls that regulate genes encoding major histones. It is the major H3 molecule that is available for deposition outside of S phase (Wu et al., 1982). These constitutively expressed histones could be important to replace nucleosomes that are disrupted during cellular processes such as transcription. H3.3 from different species while being highly similar to major type H3 consistently shows sequence differences at four sites: one in the N-terminal tail and three in the α2 helix of the histone fold domain (A31S, S87A, V89I, and M90G). Studies show that no single mutation could be completely responsible for replication independent incorporation. But any one change in H3 allows low degrees of replication independent deposition (Ahmad and Henikoff, 2002c). Also noteworthy is the fact that the N-terminal tail is necessary for replication dependent but not for replication independent deposition (Ahmad and Henikoff, 2002c). In *S.cerevisiae* the only H3 found is H3.3 (reviewed in Malik and Henikoff, 2003).

There are interesting similarities between the H2A-H2B and H3-H4 dimers. Like H3, the two H2A molecules interact with each other in the nucleosome, whereas H2B like H4

does not (Luger et al., 1997a). H2B, probably as a consequence does not have many variants but H2A has many. The first identified H2A variants were H2A.X and H2A.Z (Hatch and Bonner, 1988; Mannironi et al., 1989).

Histone H2A.X shares 95% amino acid identity with core H2A. It is present in all eukaryotes. The most conspicuous feature of H2A.X is a C-terminal extension that fits the consensus sequence: SQ(E/D)Φ, where Φ indicates a hydrophobic residue. This motif is crucial for chromatin compaction and repair by non-homologous end joining (NHEJ). The serine residue in this motif is phosphorylated in response to DNA double strand breaks (Rogakou et al., 1998). This event facilitates and is apparently followed by the recruitment of repair proteins such as Rad50, Rad51, and BRCA1 (Celeste et al., 2003; Rogakou et al., 2000; Rogakou et al., 1998). *S.cerevisiae* has only the H2A.X gene and *C.elegans* has only the canonical H2A (Thatcher and Gorovsky, 1994). Accordingly, one sees a high incidence of homologous recombination in the former and a relatively low level in the later (Malik and Henikoff, 2003). Meiotic prophase in mammalian males the X and Y chromatin condense to form a sex body or an XY body. Transcription of the X- and Y- linked genes is repressed in the XY body. This event is also known to precede meiotic recombination induced double stranded breaks. Recently H2A.X has been implicated in the condensation of the XY body during spermatogenesis (Fernandez-Capetillo et al., 2003).

H2A.Z is highly conserved throughout eukaryotic evolution. It is found in boundary elements that protect euchromatic (transcriptionally active) regions from the spreading of adjacent heterochromatin (Meneghini et al., 2003). A comparison of H2A.Z sequence with that of canonical H2A reveals two regions where it varies most. First, the docking

domain, a site of extensive interactions of H2A with the $(H3\text{-}H4)_2$ tetramer. This region has also been determined to be essential during development (Clarkson et al., 1999). Structural studies have suggested that this interface may be mildly weakened by H2A.Z (Suto et al., 2000). Looser packaging of H2A.Z nucleosome arrays is corroborated by analytical ultracentrifugation studies (Fan et al., 2002a). H2A.Z function is also found to be redundant with that of nucleosome remodeling factors. Deletion of gene encoding H2A.Z strongly increases the requirement for SWI/SNF and SAGA (Santisteban et al., 2000). Second, the L1-loop region is quite different between H2A and H2A.Z. This region is involved in formation of an interface that is the only site of contact between the two H2A-H2B dimers in the nucleosome. This interface may have a role in regulating the stoichiometry of the variant nucleosomes (one variant per nucleosome versus two). The relative affinity between L1-loops of different variants and the canonical H2A along with local concentrations of a given H2A subtype may determine the composition of nucleosomes in a particular region of the genome. Recent studies have also shown a H2A.Z specific chaperone complex called SWR1. At the core of this complex is a Swi2/Snf2-related adenosine triphosphatase, Swr1 which is required for the deposition of H2A.Z at specific loci *in vivo* (Krogan et al., 2003; Mizuguchi et al., 2004). It is not known whether this complex replaces both H2A moieties in the nucleosome or only one of them with H2A.Z. This is another potential factor that can regulate the stoichiometry of the nucleosome. The possibility of hybrid nucleosomes may give rise to yet another level of functional and structural heterogeneity not only with H2A.Z but also with other H2A variants.

H2A.Bbd is the latest H2A variant to be discovered (Chadwick and Willard, 2001b). It is also the least similar to canonical H2A among all variants. As the name suggests H2A.Bbd (Barr body deficient) is entirely excluded from the inactive X-chromosome in female mammals. It is seen in regions that have hyperacetylated H4, which is an acknowledged marker of transcriptionally active chromatin (Chadwick and Willard, 2001b). Recent studies in the lab have shown that H2A.Bbd makes the nucleosome conformationally more open. H2A.Bbd-NCPs protected only 118 bp of DNA as opposed to major H2A-NCPs, which protected 146 bp (Bao et al., 2004a). H2A.Bbd like H2A.Z also showed the propensity to form hybrid nucleosomes *in vitro* (Chakravarthy et al., 2004; see chapter 4).

Probably the most intriguing of the H2A variants reported to date is macroH2A. It is found in particularly high concentrations in the inactive X-chromosome (transcriptionally inactive) of adult female mammals.

The mechanism of the function of these histone variants remains unknown. In order to study the structural basis of their function, nucleosome structures with major type histone H2A replaced by its different variants have either been determined or are being worked on. The focus of this thesis is going to be macroH2A.

1.5 MacroH2A

MacroH2A is a 367 aa, (42kDa) protein whose N-terminal third (aa1-122) is ~64% identical to major H2A (Fig. 1.4). This region precedes a highly basic region (aa132-159) that is homologous to the C-terminus of the linker histone H1 (57% identical to sea urchin H1γ over 30 residues) (Pehrson and Fried, 1992). Amino acids 160-370 constitute

the non-histone domain. Although a putative leucine zipper-like structure has been attributed to the region between aa 181-208 it is now clear from the high resolution crystal structure of the non-histone domain that this region is, in fact, part of a 7-strand beta sheet (see chapter 3).

MacroH2A displays a very unique nuclear localization pattern. It is expressed in equal quantities in both males and females (Rasmussen et al., 1999). It is present in particularly large quantities in testes and is expressed in differentiating male and female embryonic stem cells. The non-sex chromosome specific portion of the macroH2A intracellular pool is characterized by diffuse, low-level nuclear staining. In mammalian females dosage compensation is achieved by inactivation of one of the two X-chromosomes present (Lyon, 1999). The inactive X-chromosome (Xi) resembles constitutive heterochromatin in that it replicates late in the S-phase, and is hypoacetylated on histone H4. It is often observed as a densely staining mass known as a Barr body (Lyon, 1999) at the periphery of human nuclei. MacroH2A is also highly abundant in the inactive X-chromosome (Xi) of mammalian females manifested as a Macro Chromatin Body (MCB) (Costanzi and Pehrson, 1998b). It is thought that the histone domain of macroH2A may alter the local chromatin structure at a nucleosomal level through sequence differences in the histone fold and the non-histone domain might interact with other protein or nucleic acid candidates that play a role in heterochromatinization and / or x-inactivation. One possibility is that this region of macroH2A interacts with XIST (X Inactivation Specific Transcript) a non-translated RNA molecule encoded by a gene on the X-chromosome. XIST is known to coat the Xi (Brown et al., 1992) and this event is an important essential marker of the initiation of inactivation (Brown and Willard, 1994).

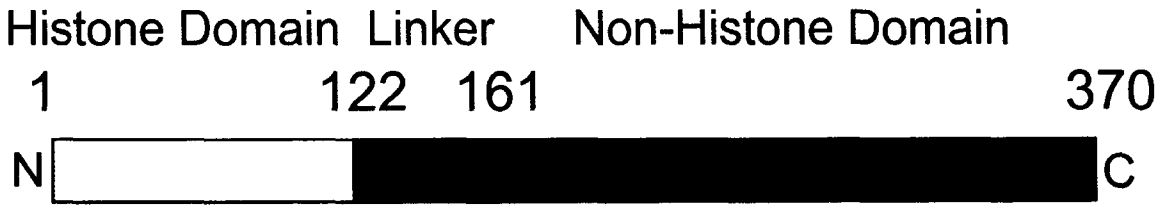


Figure 1.4 Schematic showing the structural organization of macroH2A. + indicates a highly lysine and arginine rich basic region that is homologous to the C-terminal tail of the linker histone H1. "Zip" indicates the region that was predicted to be the leucine zipper-like region.

On one hand it is observed that in fibroblasts with some mutations in XIST the mutant Xi remains late replicating and under-acetylated on histone H4 indicating that these characteristics are not dependent on Xist for maintenance (Csankovszki et al., 1999). On the other hand macro chromatin bodies are found only in fibroblasts with wild type Xist. These results suggested that macroH2A deposition to the Xi needs Xist RNA. It is also evident that the maintenance of the inactive state does not depend on the persistent association of macroH2A with Xist. Furthermore, it has also been reported that macroH2A exists in undifferentiated ES cells as prominent nocodazole (microtubule disrupting drug) sensitive focal accumulations centered on centrosomes (Rasmussen et al., 2000). This localization then shows a sequential shift from the centrosome to an inactive X chromosome during differentiation and X inactivation. The timing of macroH2A incorporation into the Xi is variable. It is a relatively early event in pre-implantation female mouse embryos (Costanzi et al., 2000). On the other hand it is a late event (7 days after initiation) in differentiating female mouse embryonic stem cells. The relative timing of MCB and Xist RNA colocalization suggests that accumulation of macroH2A1.2 on the Xi is unlikely to play a role in the primary events of X-inactivation (Mermoud et al., 1999). This makes it difficult to determine the stage (initiation, propagation or maintenance) of X-inactivation that macroH2A contributes to. It is now known that macroH2A is not 'essential' for X-inactivation because disruption of macroH2A localization to the Xi does not lead to reactivation of the Xi.

In addition to dosage compensation in female mammals, X-inactivation also occurs during spermatogenesis in male mammals resulting in the formation of the XY body

(Monesi et al., 1967; Richler et al., 1992). Gene dosage constraints cannot explain this phenomenon in males. Among the common characteristics in the X-inactivation processes in males and females are the expression of Xist RNA exclusively in the testes and the deposition of macroH2A1 at the inactive X chromosome.

MacroH2A is also found to localize to the centromeric heterochromatin in meiosis (Hoyer-Fender et al., 2000) but not during mitosis (Costanzi and Pehrson, 1998a). It is shown to co-localize with M31 (Turner et al., 2001). M31 is a mammalian member of the highly conserved HP1 (heterochromatin protein 1) protein family (Jones et al., 2000). Immunostaining analysis has revealed that M31 is a component of constitutive heterochromatin in both mitotic (Wreggett et al., 1994) and meiotic (Motzkus et al., 1999) cells. M31 was first seen to associate with the centromere of the x-chromosome in early pachytene and then with the Y chromosome in late pachytene stages of meiosis. This led to the suggestion that M31 functions in meiotic sex chromosome inactivation (MSCI) and that M31-mediated MSCI initiated on the X chromosome in early pachytene and later spread to the Y chromosome via the synapsed pseudoautosomal region (PAR) – the site of XY pairing during meiosis. MacroH2A1 and M31 in addition to their association with the centromeric heterochromatin and with the XY body also co-localize during late prophase to the PAR. Current ideas on meiosis suggest that chromosome condensation at the zygotene stage facilitates synapsis and recombination, whereas condensation at diplotene allows resolution of chiasmata and individualization of chromosomes prior to disjunction (Cobb et al., 1999; Hunter et al., 2001). It may be said in light of these observations that M31 might partly control the early condensation step, whereas macroH2A1 and M31 together control the later step. Studies have revealed that

HP1 is also enriched in the Barr body (in the female context) (Chadwick and Willard, 2003). MacroH2A co-localizes with HP1β in pericentromeric heterochromatin of autosomes and of the x-chromosome in mid to late pachytene spermatocytes, and to heterochromatin regions of somatic Sertoli cells. This suggests a possible function in heterochromatin formation in these sites for macroH2A. The spatial and temporal patterns of macroH2A and Hp1β staining suggest that macroH2A could have a role in the recruitment of Hp1β-containing protein complexes and the regulation of chromatin function (Hoyer-Fender et al., 2004).

MacroH2A is emerging as a "family" of variants. At least two different genes are known to encode this protein with significant variations in sequence but the same basic architecture and are called macroH2A1 and macroH2A2. MacroH2A1 has two splice variants macroH2A1.1 and macroH2A1.2, which are non-identical in a very small region of the non-histone region starting at aa 195 (Pehrson et al., 1997) (Fig. 1.5). MacroH2A2 is 68% identical (overall) to macroH2A1.2 (Fig. 1.5). The histone region is 84% identical to that of macroH2A1 and only 66% identical to major H2A. The basic region is the most varied and is only 25% identical to that of macroH2A1 (Chadwick and Willard, 2001a; Costanzi and Pehrson, 2001) (Fig. 1.5 linker domain). The basic domains of both macroH2A1 and macroH2A2 contain nearly 50% basic amino acids, but the location of each residue is not conserved. If the region simply serves as a patch of basic character, the site of each residue may not be crucial for function. The non-histone region is 64% identical to that of macroH2A1.2. The region in macroH2A2 corresponding to that of non-identity between macroH2A1.1 and macroH2A1.2 bears a closer resemblance to macroH2A1.2 (48% identical) (Fig. 1.5). MacroH2A1.2 and

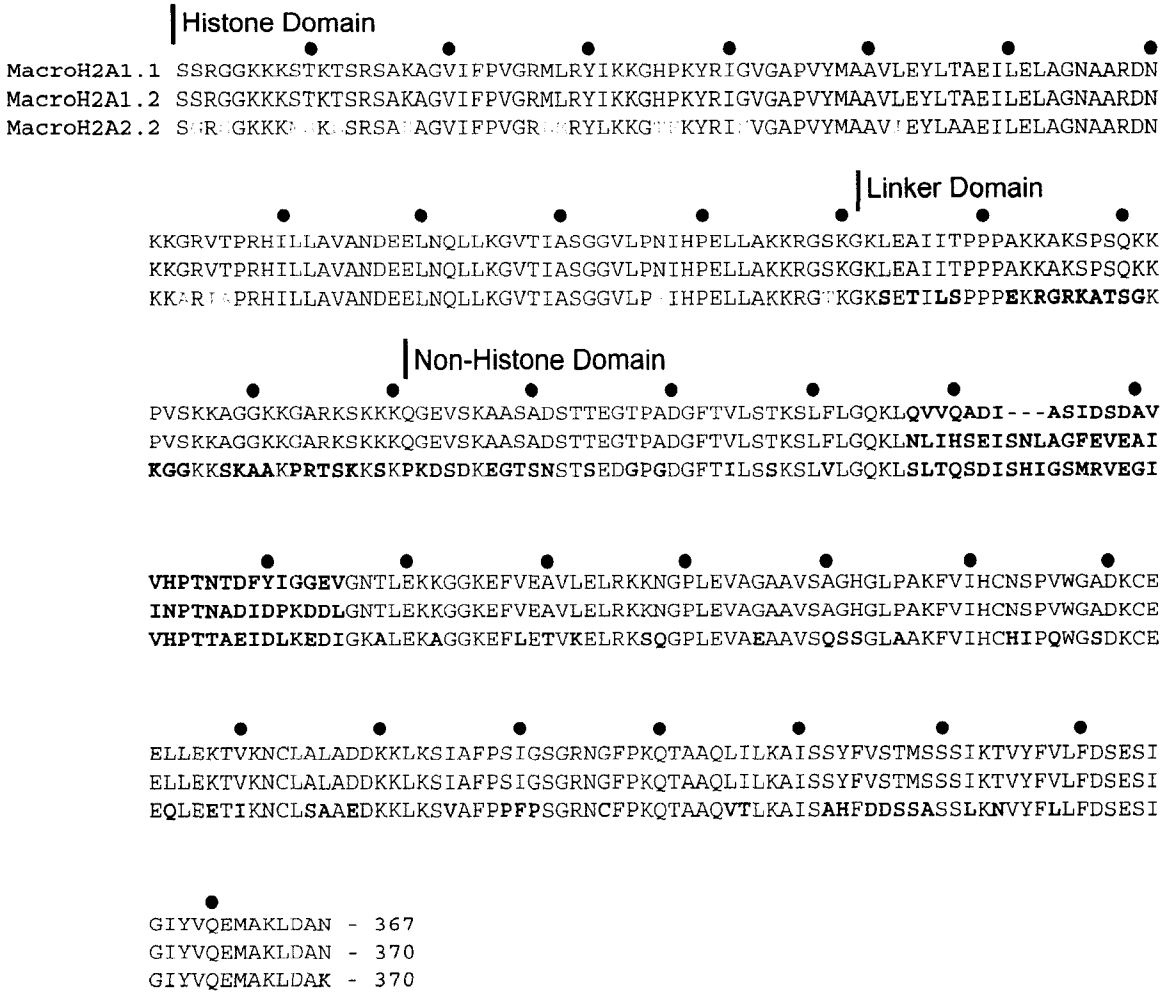


Figure 1.5. Sequence alignment of human macroH2A1.1 and macroH2A1.2 with human macroH2A2.2. Bullets indicate every tenth residue in macroH2A1.1. The beginning of each domain is indicated (histone, linker, and non-histone domains). Sequence differences in the histone domain are in orange, sequence differences in the linker domain are in blue, sequence differences in the non-histone domain are in green. The entire region of non-identity (due to alternative splicing between macroH2A1.1 and macroH2A1.2) is shown in red.

macroH2A2 display very similar (and sometimes overlapping) nuclear localization patterns (Chadwick and Willard, 2001a; Costanzi and Pehrson, 2001).Studies to understand the significance of the different regions of macroH2A have led to some interesting insights into the mode of deposition into the inactive X-chromosome (Xi). The histone domain alone (fused to GFP) was found to form the Xi associated macro chromatin body (MCB) with efficiencies comparable to full-length macroH2A (Chadwick et al., 2001). Deletion mutant analysis revealed that the N- and C-terminal tails of the histone domain are inessential for MCB formation. The targeting function therefore resides in the globular histone fold and is called Macro Chromatin Domain 1 (MCD1). Of the 19 amino acids that are different in macroH2A compared to major H2A no single amino acid could be attributed this function by point mutation studies. It is therefore very likely that a combination of these differences contributes to MCB formation. Surprisingly, the non-histone domain of macroH2A can be targeted to the Xi independently of the histone domain (therefore called MCD2) (Chadwick et al., 2001). It was found to localize to the Xi in association with the histone fold of major H2A, which on its own is not capable of forming an MCB. The basic region 120-175 influences the efficiency with which the MCB is formed (Chadwick et al., 2001). This was determined by domain swap experiments between macroH2A1.2 and macroH2A2. MacroH2A1.2 is more efficient than macroH2A2. The efficiencies of full-length macroH2A1.2 and macroH2A2 however are comparable to each other. Results from insertion and deletion mutation studies showed that the region between 182-371 is very crucial for the incorporation of the non-histone domain into Xi and MCB formation (Chadwick et al., 2001).

Recent experiments have shown that the proximity of the non-histone region of macroH2A to a promoter leads to repression of transcription (Perche et al., 2000a). MacroH2A may act as a 'roadblock' to the transcriptional machinery (Angelov et al., 2003a). Nucleosomes containing full-length macroH2A were refractory to NF-κB (transcription factor) binding and ATP dependent SWI/SNF mediated remodeling. Another noteworthy observation is that nucleosomes containing only the histone domain of macroH2A (aa 1-122) were equally refractory to SWI/SNF mediated remodeling but were amenable to NF-κB binding. This indicates that the different domains perform different aspects of macroH2A function, which in turn leads to a transcriptionally repressed chromatin structure.

On the basis of its high degree of evolutionary conservation, nuclear distribution, and unique structural organization, we postulate that macroH2A plays a significant role in establishing structurally and functionally unique chromatin domains (including the inactive X-chromosome) that are expected to be transcriptionally silent.

1.6 Specific Aims

The primary aim of the work presented in this thesis is to study the structural consequences of incorporating the histone H2A variant macroH2A into nucleosome core particles. MacroH2A has an unusual structural organization compared to canonical histones. It has an N-terminal histone domain that is 64% identical to major type H2A, connected to a C-terminal non-histone domain by a linker region. We propose that the histone domain alters the nucleosome structure by means of the sequence differences, and that the non-histone domain's function is extra-nucleosomal, i.e., it influences the higher

order chromatin structure either directly or indirectly by acting as a recruiting platform for other chromatin associated proteins. We tested this hypothesis by determining the 3 Å crystal structure of a nucleosome core particle in which the histone domain of macroH2A replaces major type H2A. We performed *in vitro* domain swap studies to determine which regions in the histone fold of macroH2A are responsible for macro-like behavior. We investigated the tendency of different H2A variants to form nucleosomes with non-canonical stoichiometry (hybrid nucleosomes) to test the hypothesis that the relative affinity between the L1-loops of different H2A variants regulates the histone composition of nucleosomes. As a first step towards understanding its function we determined the 1.6 Å *de novo* crystal structure of the non-histone domain of macroH2A.

We attempted structural studies on nucleosomes containing the centromeric H3 variant CenH3. We used the *D.melanogaster* and *H.sapiens* heterologs of CenH3 (cid and CENP-A respectively) for this purpose. We find significant differences in their biochemical behavior and crystal morphology. Optimization of the crystallization process is underway. As a control we also determined the 2.3 Å crystal structure of the nucleosome containing major-type core histones from *D.melanogaster*.

CHAPTER 2

Distinct regions in the histone-fold of macroH2A are responsible for its structure and function.

We determined the 3Å crystal structure of the nucleosome core particle containing the histone domain of macroH2A (macro-NCP). We find that the major structural changes are limited to the L1-loop of the macroH2A histone fold. We also find biochemical and *in vivo* evidence that indicates that the L1-loop is largely responsible for the structure and function of a macro-NCP. This data is being prepared for submission in combination with the contents of chapter 3.

2.1 Abstract

MacroH2A is an H2A variant that is enriched on the inactive X-chromosome in mammalian cells. It is atypical in its structural organization in that it has a C-terminal non-histone domain connected to the N-terminal histone domain by a flexible linker. The H2A-like domain is only 64% identical to non-variant H2A. The 3.0 Å structure of nucleosome core particles containing the histone domain of macroH2A (macro-NCP) revealed significant structural changes in the L1 loop of macroH2A, which forms the only interface between the two histone dimers within the NCP. Biochemical characterization showed that four amino acids in the L1 loop of macro-H2A cause dramatic changes in the stability of the histone octamer *in vitro*. The same four amino acids are sufficient for *in vivo* targeting of macro-H2A to the inactive X-chromosome. Our data provide structural and functional evidence that a four amino acid stretch exerts a pronounced effect on the *in vitro* and *in vivo* assembly pathway of a histone variant.

2.2 Introduction

All eukaryotic DNA is complexed with a roughly equal mass of protein to form chromatin (Van Holde, 1988). Its fundamental repeating structural and functional unit is the nucleosome core particle (NCP) that consists of an octamer that contains two copies of the four core histone proteins H2A, H2B, H3, and H4 around which 146 base pairs of DNA are wrapped in 1.65 superhelical turns (Luger et al., 1997a). The compaction of DNA into chromatin is an important regulator of DNA accessibility, and thus nucleosomes play a central role in transcription, replication, and repair. An important emerging mechanism to alter the fundamental biochemical composition and

characteristics of chromatin is the substitution of major-type core histones with histone variants (Malik and Henikoff, 2003).

Although the modes by which histone variants exert their function are yet to be understood, it has become apparent that they play vital and very specific roles in the formation and maintenance of structurally and functionally unique chromatin domains. This may be facilitated by structural alterations in the NCP and / or in chromatin higher order structures that are brought about by the amino acid sequence differences between the histone variants and their corresponding core counterparts (Suto et al., 2000); (Fan et al., 2002b). Many key questions regarding the mechanism by which histone variants exert their diverse and often essential functions are only now being addressed. For example, how do histone variants affect nucleosome structure? Is chromatin, which harbors histone variants, stabilized or destabilized? How are histone variants targeted to specific regions in the genome? Given the many emerging roles of different histone variants, it is likely that each will yield its own set of answers to these central questions.

MacroH2A1, with a molecular weight of 42 kDa, is almost three times the size of major, replication dependent H2A and is unique among known histone variants due to its unconventional tripartite structural organization (Pehrson and Fried, 1992). The N-terminal third of its amino acid sequence (aa 1-122) is only 64 % identical to major H2A. Considering the high general degree of evolutionary conservation among histones (Sullivan et al., 2002b), this signifies a relatively high degree of sequence divergence. A C-terminal non-histone domain (161-371) is linked to the histone-homology domain via a linker region (aa 123-160) that exhibits an amino acid composition similar to the C-terminal region of the linker histone H1. The C-terminal non-histone domain in itself is a

member of a functionally diverse group of proteins that exist in organisms ranging from bacteria and archaea to eukaryotes, and its function remains unknown. Overall, the amino acid sequence of macroH2A shows a remarkably high degree of evolutionary conservation between different species. The histone domain (1-122) is 100 % conserved in all species from which macroH2A has been identified; other regions of macro-H2A are only slightly less highly conserved (Pehrson and Fuji, 1998).

MacroH2A1 exists in two isoforms denoted macroH2A1.1 and macroH2A1.2, which are splice variants of the same gene (Pehrson et al., 1997); (Pehrson and Fried, 1992) and have distinct expression patterns with respect to tissue specificity and stage of differentiation. The two isoforms differ from each other only in the non-histone region (Pehrson et al., 1997). Thus, our results pertaining to the histone-like domain of macroH2A are applicable to both isoforms. MacroH2A2 has the same structural organization as macroH2A1 but is expressed from a different gene. The histone domain is highly homologous to that of macroH2A1 and most differences are in the linker and non-histone domain (Chadwick et al., 2001; Costanzi and Pehrson, 2001).

MacroH2A1 preferentially localizes at the inactive X-chromosome of mammalian female cells, where it is one of multiple redundant factors required for stable maintenance of transcriptional silencing (Costanzi and Pehrson, 1998b). Since macroH2A is expressed in equal levels in males and is also found in species that do not use X-inactivation for dosage compensation (Costanzi and Pehrson, 1998b; Pehrson et al., 1997; Pehrson and Fuji, 1998), it has been suggested that it may have a biological function above and beyond X-inactivation. Recent studies have found macroH2A – containing nucleosomes to be repressive towards transcription (Angelov et al., 2003a; Perche et al., 2000a).

A number of high-resolution structures of NCPs reconstituted with histones from different species and histone variants, and 146-147 bp of DNA derived from the human α-satellite sequence have been determined (Luger, 2003), revealing the intricacies of protein – protein and protein – DNA interactions within the NCP. To understand how the significant sequence differences between macroH2A and major H2A affect nucleosome structure, and to pinpoint the regions in macroH2A that are responsible for targeting to the inactive X-chromosome, we have combined x-ray crystallography with biochemical studies. We show that the L1 loop, which forms the only interface between the two H2A-H2B dimers within a single NCP, displays major structural differences between macroH2A NCP and majorH2A NCP. A four-amino acid region within the L1 loop of macroH2A causes dramatic changes in the stability of the histone octamer in response to decreasing ionic strength, and in the absence of the stabilizing influence of the DNA. Thus, our data implicate the L1 loop as a motif that is important in determining structural characteristics of nucleosomes containing H2A variants.

2.3 Materials and Methods

2.3.1. Expression and purification of histone proteins, and reconstitution of nucleosomes

The coding region for the histone domain of human macroH2A (amino acids 1-120) was sub-cloned into a pET3a vector. Expression plasmids for H2B and H3 from mouse (*Mus musculus*) were a kind gift from Dr. David Tremethick. All histones were over-expressed in BL21 (DE3)-plysS (Stratagene) and purified using previously published protocols (Luger et al., 1999b). The histone domain of macroH2A (macroH2A-HD), together with mouse H2B, H3 and H4 were refolded to a histone octamer (macro-octamer). This was

reconstituted onto a 146 bp palindromic DNA fragment derived from human α-satellite regions (α-sat DNA; (Luger et al., 1997a) using salt gradient dialysis (Dyer et al., 2004), resulting in macro-NCP. NCP reconstituted with histones from *Xenopus laevis* (*Xla*-NCP) were used as control. Milligram amounts of macro-NCP or *Xla*-NCP were heat-shifted and purified by preparative gel electrophoresis using published protocols (Dyer et al., 2004).

2.3.2. Crystallographic procedures

Macro-NCP was crystallized using salting in vapor diffusion at NCP concentrations ranging from 8-12 mg/ml with salt concentrations of 34 to 37.5mM KCl and 40-45mM $MnCl_2$. The crystals were soaked in 24% 2-methyl-2,4-pentanediol and 5% trehalose and frozen in liquid nitrogen as described previously (Luger et al., 1997a). Data were collected at Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline 8.2.2. Data from a single crystal were processed using Denzo and Scalepack (Otwinowski and Minor, 1997). Molecular replacement was performed using Protein Data Bank entry 1AOI as the search model. Molecular replacement and subsequent rounds of refinement were performed using CNS (Rice et al., 1998). The program O was used for model building (Jones et al., 1991). The veracity of the model was checked using SA-OMIT maps for critical regions during various stages of refinement, and a composite omit map at the end.

2.3.3. NAP-1-dependent NCP reconstitution

GST-tagged yeast NAP-1 (yNAP-1) was purified using published protocols (McBryant et al., 2003). Macro- or *Xla*-octamer; or $(H3\text{-}H4)_2$ tetramer and H2A-H2B (or macroH2A-H2B) dimer that had been stored in refolding buffer (2 M NaCl, 10 mM Tris-HCl pH 7.5,

1mM EDTA, 5 mM β-mercaptoethanol) was dialyzed against 100 mM NaCl TE buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA). The components were mixed with α-sat DNA in the presence of a two-fold molar excess of yNAP-1 (assuming a yNAP-1 monomer) over histone octamer (or dimer + tetramer). The reactions were incubated at room temperature for 3 hours, followed by a one-hour incubation at 37° C. NCPs were analyzed by native polyacrylamide gel electrophoresis (Luger et al., 1999b).

2.3.4. Domain swapping

We performed site directed mutagenesis using the quickchange mutagenesis kit (Stratagene) to change the L1 loop of *Xenopus laevis* majorH2A to that of macroH2A (amino acids 38-41 in majorH2A, which correspond to amino acids 35-38 in macroH2A, changed from NYAE to HPKY, resulting in the mutant mL1-H2A). Similarly, we mutated amino acids 83, 84 and 88 in major H2A, which correspond to amino acids 80, 81 and 85 in macroH2A (L-I, Q-L, and R-A respectively), to simulate relevant changes in the docking domain, resulting in mDD-H2A. Mutants were expressed, purified, and reconstituted into nucleosomes along with the other core histones from *Xenopus laevis* as described above, resulting in mL1-NCP and mDD-NCP, respectively.

2.3.5. Analytical Gel Filtration

The relative stability of non-variant and variant histone octamers in varying salt concentrations was studied using a Superdex-200 10/30 (Amersham) size-exclusion column. We injected sample (*Xla*-octamer, macro-octamer, or mL1-octamer) onto the column at strategic points in a decreasing salt gradient (2M NaCl to 0.5M NaCl), in a modification of a procedure described earlier (Endo et al., 1983). The injection points

were 2 M, 1.7 M, 1.4 M, 1.1 M, 0.8 M, and 0.5 M. With decreasing salt concentration the octamer dissociate into $(H3\text{-}H4)_2$ tetramer and H2A-H2B dimer.

2.4. Results

2.4.1. The nucleosome core particle containing the histone domain of macroH2A is structurally very similar to canonical nucleosomes

The amino acid sequence of the histone domain of macroH2A (amino acid 1 to 122) is only 64 % identical to major H2A from both *Xenopus laevis* and *Mus musculus* (Fig. 2.1A), but 100 % conserved between different species, indicating that the sequence of this variant has evolved to fulfill functions that are distinct from those of major-type H2A. There are two particular regions of sequence divergence that have the potential to cause structural and functional differences. First, the ladle-shaped H2A docking domain formed by H2A amino acids 83-108 is involved in an extensive interaction interface with one half of the $(H3\text{-}H4)_2$ tetramer, and guides the N-terminal helix of H3 (H3 αN) to interact with the penultimate 15 base pairs of the nucleosomal DNA. Second, the H2A L1 loops form the only interface between the two H2A-H2B dimers within a single NCP, perhaps facilitating the cooperative incorporation of the second H2A-H2B dimer in the NCP. This interface seemingly holds together the two gyres of the DNA superhelix, and may thus play a role in regulating the dynamic behavior of nucleosomes.

We investigated how the sequence differences between macroH2A and major H2A would affect nucleosome structure, by determining the crystal structure of macro-NCP.

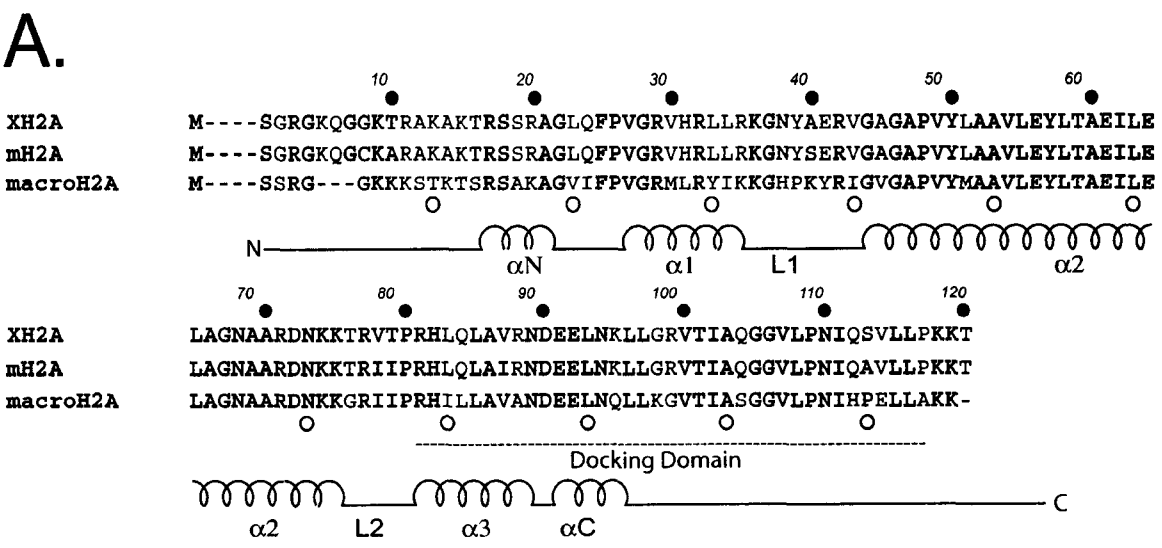


Figure 2.1A. Histone domain of macroH2A is 64% identical to the major type H2A. Sequence alignment of *Xenopus laevis* H2A, mouse H2A and the histone domain of human macroH2A. Filled circles indicate intervals of 10 amino acids in major H2A. Open circles indicate intervals of 10 amino acids in macroH2A. Differences between major H2A (mouse and *X.laevis,* or macroH2A and *Xla*-H2A, respectively) and macroH2A are shown in red. Differences between mouse and *Xla*-H2A are shown in blue. The secondary structure elements of the histone fold (α1, α2, and α3) and extensions (αN and αC) are indicated. The dashed line delineates the docking domain.

Histone octamers were refolded from mouse H2B, H3, H4, and the histone domain of human macroH2A (macroH2A$_{HD}$), which is 100% conserved between mouse and human (accession no. NP_036145 and NP_613075 respectively). The refolded histone octamer was reconstituted with a 146 bp palindromic human α-satellite DNA sequence (Luger et al., 1997a) using salt gradient dialysis, to yield macro-H2A$_{HD}$-containing nucleosome core particle (macro-NCP). Crystals diffracting to a resolution of ~ 3.0 Å were obtained by salting in vapor diffusion. Phases were obtained by molecular replacement, using *Xenopus laevis* NCP (*Xla*-NCP; PDB entry 1AOI) as a search model, and the structure was refined to a crystallographic R factor of 20.6 % (R_{free} = 26.0 %; Table 2.1). A representative region of the final electron density map, contoured at 1 sigma, is shown in Fig. 2.1B. The region of the macroH2A-NCP structure shown differs between macroH2A and major H2A in residues 83, 84, and 88, which correspond to residues 80, 81, and 85 in macroH2A (L-I, Q-L, and R-A respectively) and these differences in sequence were clearly visible even in the initial electron density map.

Despite the significant sequence differences between *Xla*-H2A and macroH2A, the structures of macro-NCP and *Xla*-NCP are super imposable with a root mean square deviation (rmsd) of < 1Å. Figures 2.1C and 2.1D show a superposition of the two structures in two different orientations. The path of the DNA is highly similar to that observed in major NCP (consistent with DNase footprinting experiments (Abbott et al., 2001)), however, it shows significant differences in the way in which macro-NCP responds to the crystallization induced twist defect of the short half of their DNA (see (Suto et al., 2003); (Muthurajan et al., 2003) for a general discussion of this

TABLE 2.1: DATA COLLECTION AND REFINEMENT STATISTICS

DATA COLLECTION	
Space group	$p2_12_12_1$
Unit cell dimensions (Å)	a = 105.5, b = 109.6, c = 176.0
Resolution range (Å)	50 – 2.95
Unique reflections	43,366
Completeness (%; overall / highest resolution shell)	99.5 / 100
R_{merge}[a] (%; overall / last shell)	9.5 / 42
REFINEMENT	
Number of amino acid residues in the final model[b]	759
Number of base pairs in the DNA	146
Number of water molecules	105
Total number of atoms in the final model	11,952
R-factor[c] / R_{free}	0.206 / 0.260
Resolution range (Å)	50 – 3.0
R.m.s. deviation from ideality	
Bonds (Å)	0.0068
Angles (°)	1.092
Average B-factors $(Å)^2$	
Protein	69.2
DNA	127.0
Solvent	67.6

[a] $R_{merge} = \Sigma \mid I_h - \langle I_h \rangle \mid / \Sigma I_h$, where I_h is the mean of the measurements for a single hkl.
[b] Residues included in each histone subunit: H3: 38 – 135, H3': 38:135, H4: 24:102, H4': 20:102, macroH2A: 14:119 (majorH2A aa-numbers), macroH2A': 14:119 (majorH2A aa-numbers), H2B: 30:122, H2B': 27:122. The remaining histone tails were too disordered to be included in the final model. [c] R-factor $= \Sigma \mid F_{obs} - F_{calc} \mid / \Sigma F_{obs}$

B.

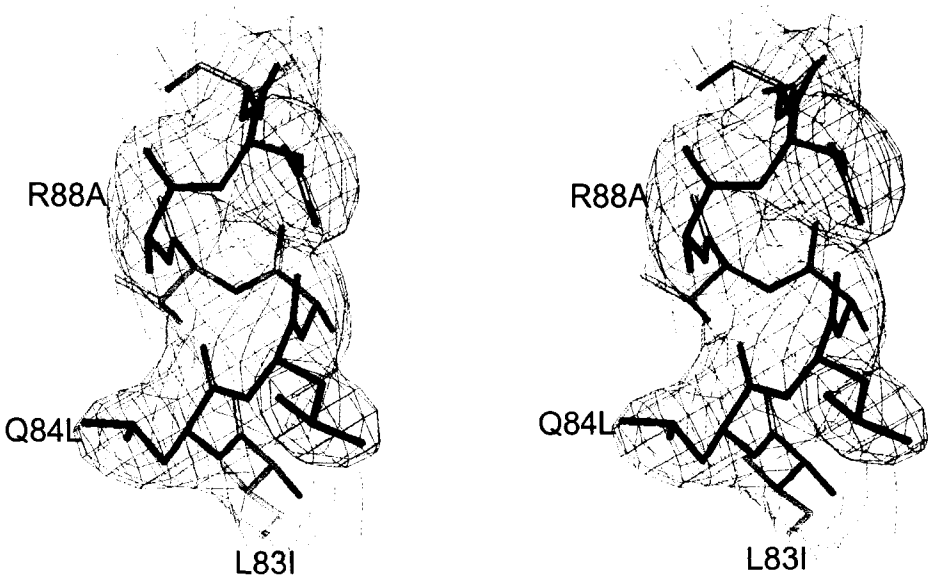


Figure 2.1 B. Differences in sequence are evident in the electron density map. Stereo view of a section of the |2Fo-Fc| electron density map, calculated at 3Å and contoured at 1σ, showing sequence differences between macro-H2A and Xla-H2A L83 to I80, Q84 to L81, and R88 to A85.

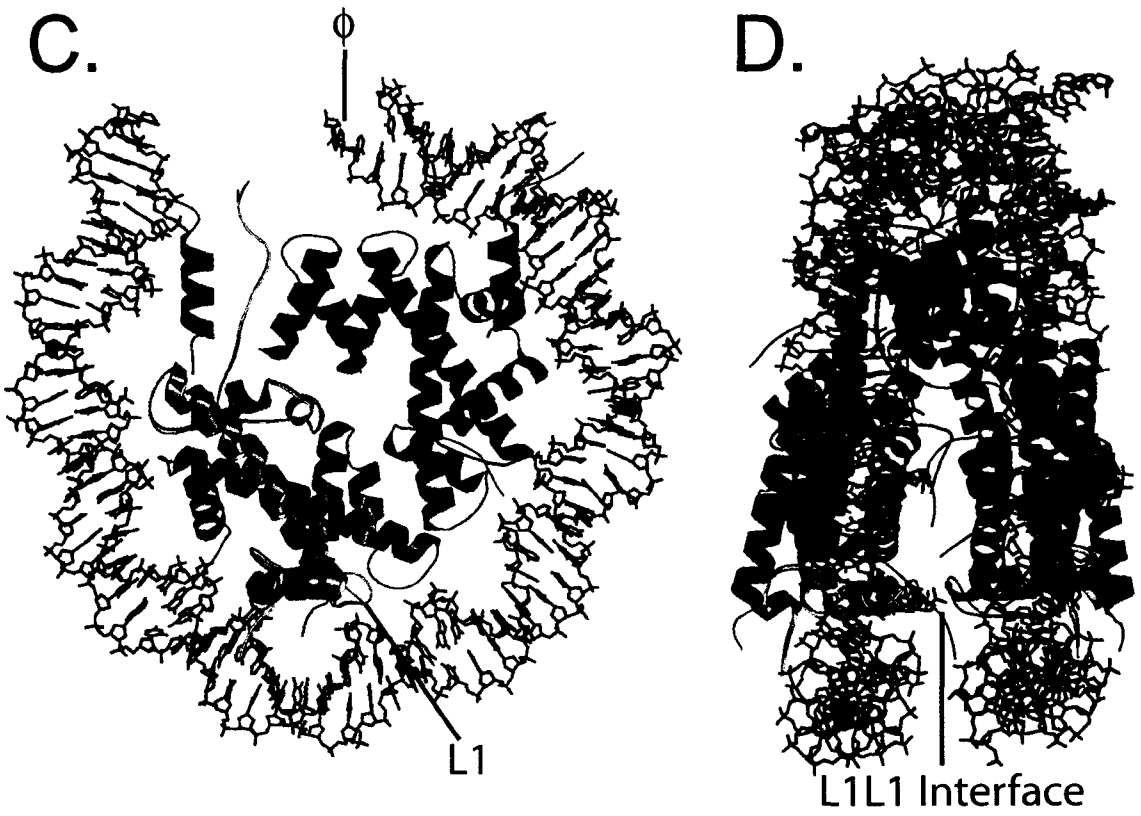


Figure 2.1 C & D. The overall structure of macro-NCP is similar to that of major NCP. C. Superposition of major-NCP and macro-NCP (only histone octamers are superimposed), viewed down the superhelical axis. Only 73 bp of the DNA and associated proteins are shown. The central base pair is indicated (ϕ). H3 is colored blue, H4 green, H2B red, H2A yellow, macroH2A gray and DNA turquoise. The L1 loop is indicated. **D.** Side view of the superimposed nucleosomes in (C), rotated by 90° around the y-axis with parts of the DNA removed for clarity. The L1 – L1 interface is indicated.

phenomenon). As a practical consequence, the refined structure contains only 145 base pairs of DNA. The histone domain of macro-H2A interacts with H2B in a manner very similar to *Xla*-H2A. Major interactions between the α2 helices of the two histones are maintained. In addition, there are no major differences in the protein-protein or protein-DNA interactions that stabilize macro-NCP, with the notable exception of the L1-L1 interface between the two macroH2A moieties (Fig 2.1C & D), where the path of the main chain in macro-NCP deviates significantly from that of major-NCP (see below).

2.4.2. Macro-NCP differs from non-variant NCP primarily in the L1-loop

We were particularly interested in the structural alterations caused by the sequence differences in the L1 loop and docking domain (Fig 2.1A). These are the two regions within the histone domain of H2A that play important structural roles in an NCP.

Indeed, the L1 loop is the only region where the path of the main chain deviates between macro-H2A and *Xla*-H2A (Fig. 2.2A and B). In *Xla*-NCP, the L1-L1 interface is stabilized by two intermolecular salt bridges between Glu41-Asn38' and Glu41'-Asn38 (Fig. 2.2C). These interactions are replaced in macro-NCP by two sets of hydrophobic interactions. Lys37 is sandwiched between Tyr38 and Pro36', and by interactions that are related via the axis of pseudo-symmetry of the NCP, Lys37' is sandwiched between Tyr38' and Pro36 (Fig. 2.2D). Together, these structural changes may render the L1-L1 interface in macro-NCP less flexible and more hydrophobic, and may alter the overall stability of the histone octamer.

In contrast, the interactions between the macroH2A$_{HD}$-H2B dimers and the (H3-H4)$_2$ tetramer remain unaffected by the numerous sequence variations in the macroH2A

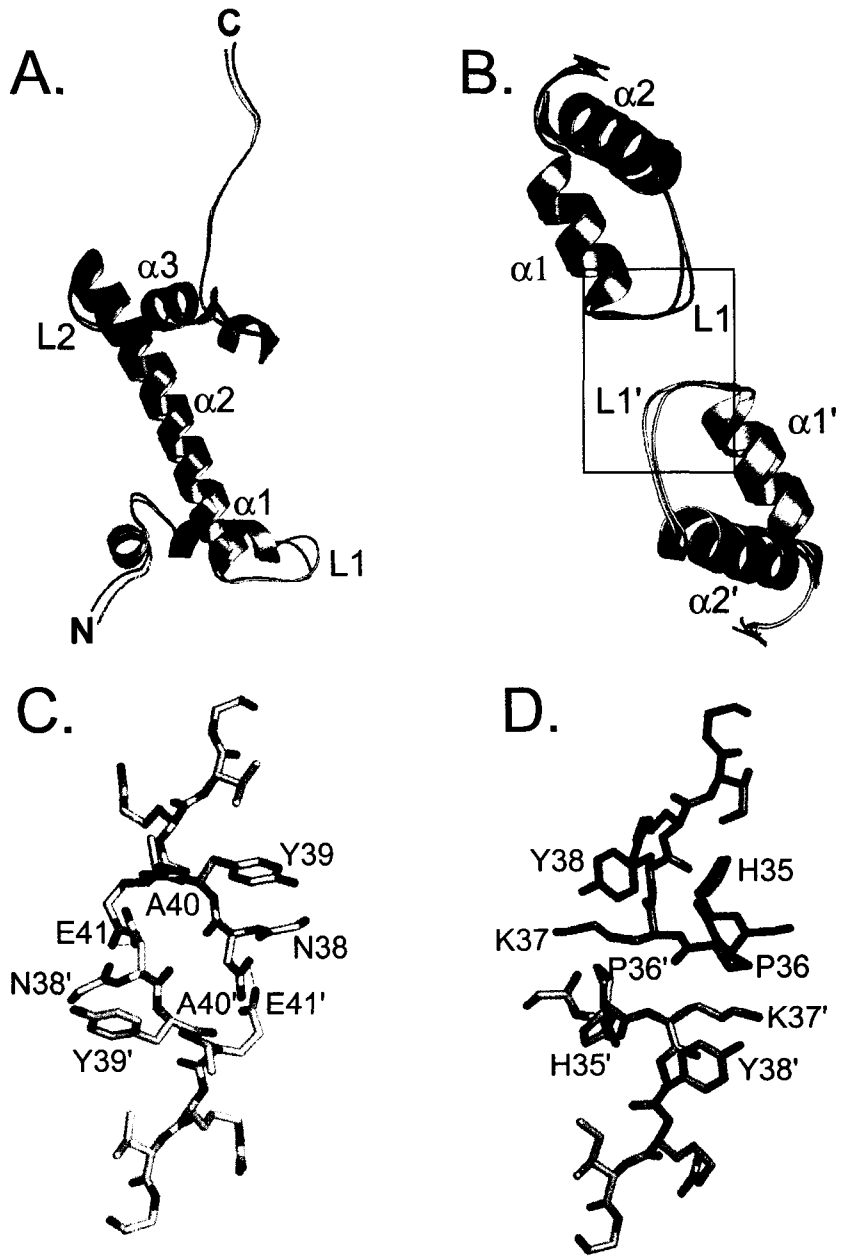


Figure 2.2 The interface formed by two L1 loops differs significantly between macro-NCP and *Xla*-NCP. A. Superposition of H2A and the histone domain of macroH2A in a view similar to that in 1c. MacroH2A is shown in gray, major H2A in yellow. **B.** Superposition of the L1 loops and the α-helices of macroH2A and macroH2A' (gray and off-white) and of H2A and H2A' (yellow and light yellow). Only minor changes in the path of the main chain of the L1 loop are observed.
C & D. Detailed view of the boxed area in (B) shows fundamental differences in the intermolecular interactions between two macroH2A chains (gray and off white) and two major H2A chains (yellow and light yellow) molecules respectively.

docking domain (Fig. 2.3). In *Xla*-H2A, residues Leu83, Gln84, and in particular Arg88 appear to be essential in stabilizing the conformation of the docking domain in absence of the $(H3\text{-}H4)_2$ tetramer (Fig. 2.3B). In macroH2A, these residues are changed to Ile80, Leu81 and Ala85 (Fig. 2.3C). Surprisingly, these substitutions do not result in structural changes in the histone main chain (Fig. 2.3A). The absence of the four hydrogen bonds that are being made by H2A R88 (Fig. 2.3B) is somewhat compensated by a small hydrophobic interface between the unique macroH2A residues I80, L81, and I99 (Fig. 2.3C). Interestingly, the acidic side-chains of H2A (E56, E61, E64, D90, E91, and E92) that form a pronounced acidic patch on the surface of the NCP are all conserved in macroH2A. This surface is essential for crystal contacts (Luger et al., 1997a), and represents the most distinct surface characteristic that is present in all major-type NCPs.

2.4.3. Macro-NCP and major-NCP behave differently during salt-dependent reconstitution

We have previously shown that NCP preparations reconstituted from *Xla* histones (and other major-type histones) and the DNA fragment used in this study are heterogeneous with respect to the position of the histone octamer on the DNA (Muthurajan et al., 2003). These positional isoforms are distinguished by their different electrophoretic mobility, and can usually be converted to a thermodynamically favorable position by incubation at 37° C for 20 – 60 min ('heat shifting', Fig. 2.4A, lanes 1 & 2). Macro-NCP also reconstitutes as two different species, but this heat-induced redistribution is reproducibly incomplete. A significant fraction of macro-NCPs remain in the upper band after heating at 37° C (Fig. 2.4A, lanes 3 & 4) and even higher temperatures (55° C; data not shown). The histone content of the bands representing shifted and unshifted nucleosomes was

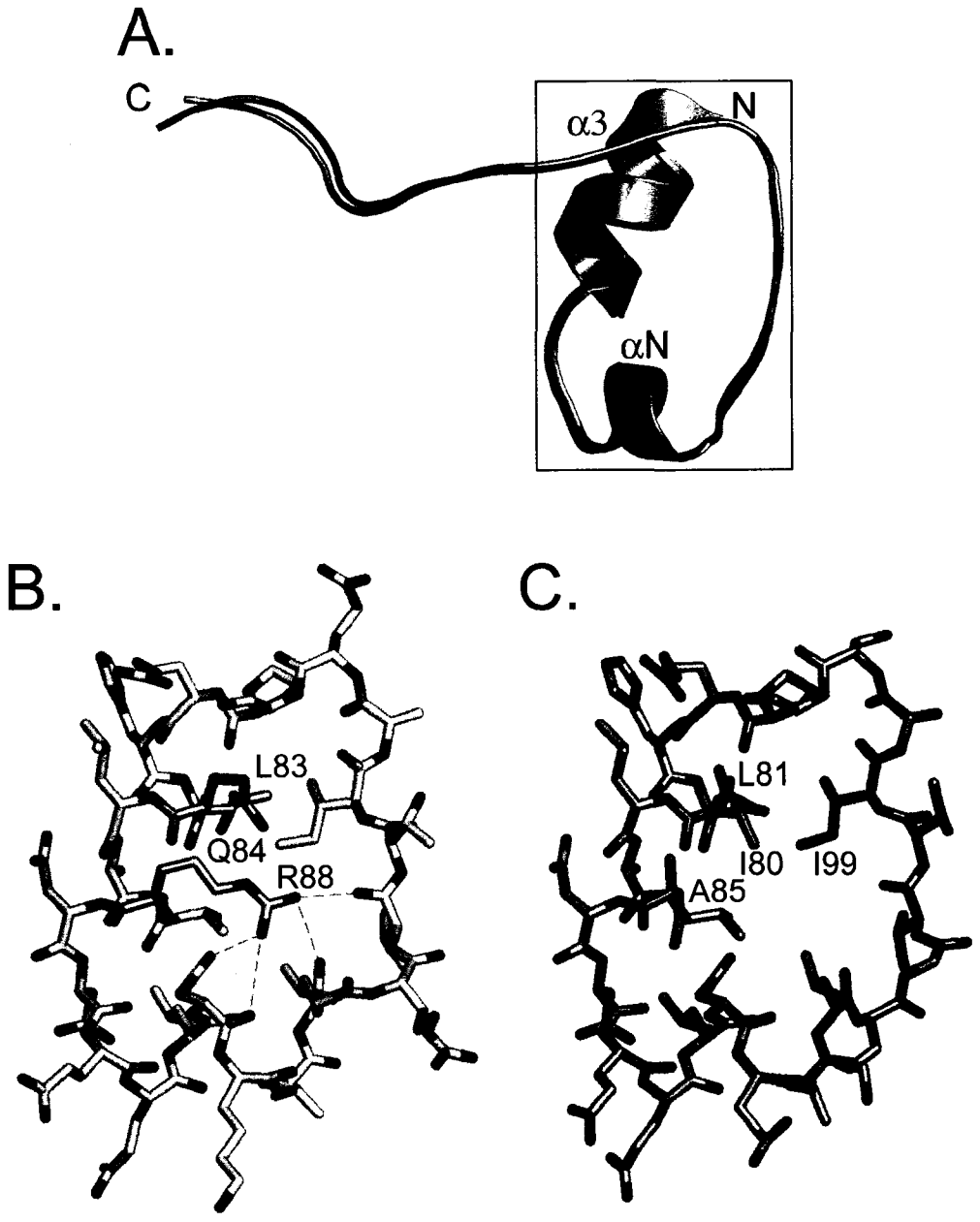


Figure 2.3 A. The H2A docking domain remains unchanged in spite of several sequence differences. A. Superposition of the docking domains (amino acid residues 80-119) of macroH2A (light gray) and *Xla*-H2A (light yellow) demonstrates that there are no differences in the path of the main chain between the two docking domains. **B, C.** Detailed view of the boxed area (amino acids 80-105 in majorH2A) in (A) highlights significant amino acid differences (L83 to I80, Q84 to L81, and in particular R88 to A85) that may alter the stability of this area. The sequence differences are shown in dark gray (macroH2A) and dark yellow (*Xla*-H2A). Also shown are some of the hydrogen bonds formed by R88 to stabilize this area in *Xla*-H2A.

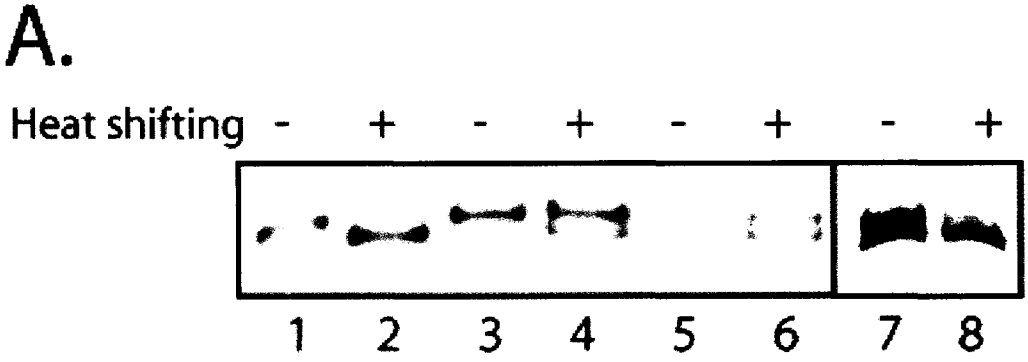


Figure 2.4 A . The L1 loop is responsible for the *in vitro* behavior of macro-NCP. Analysis of major- and macro NCP by 5% native PAGE, before and after a 2 hour incubation at 37 °C. Lanes 1 & 2, unshifted (-) and shifted (+) *Xla*-NCP; lanes 3 & 4, unshifted and shifted macro-NCP; lanes 5 & 6, unshifted and shifted mL1-NCP; Lanes 7 & 8, unshifted and shifted mDD-NCP. The gel was stained with commassie blue.

compared by SDS-PAGE. This analysis demonstrates that shifted macro-NCP contains a full complement of histones in stoichiometric amounts (Fig. 2.4B, compare lanes 5 and 6). However, the slower migrating species appears to be partially depleted for macroH2A$_{HD}$ and H2B, most likely due to loss of one macroH2A$_{HD}$-H2B dimer (Fig. 2.4B, compare lanes 3 and 4). We therefore hypothesize that the slower migrating band of macro-NCP consists of two species, one which can be fully shifted and thus represents a population with different translational positioning, and one which cannot be converted to the thermodynamically more stable form by heating due to a non-canonical histone composition.

The slower-migrating species may be specific to assembly of macro-NCP by salt gradient dialysis. This method relies on the dissociation of the histone octamer (which is stable at 2 M NaCl, but not under physiological conditions) into a (H3-H4)$_2$ tetramer and two H2A-H2B dimers at relatively high ionic strength, followed by the sequential deposition of (H3-H4)$_2$ tetramers and H2A-H2B dimers onto the DNA at consecutively lower salt concentrations (Luger et al., 1999b); (Dyer et al., 2004). The delicate balance between the stability of the histone octamer on the one hand, and the relative stability of the H2A-H2B dimer and (H3-H4)$_2$ tetramer interaction with DNA at critical ionic strength may be altered in macro-NCP. To address this possibility, we reconstituted macro-NCPs using two different approaches. First, instead of mixing refolded macro-octamer with DNA, we used independently refolded macroH2A$_{HD}$-H2B dimers, (H3-H4)$_2$ tetramer and 146 bp DNA for salt-gradient deposition. This approach has previously been shown to give identical results to salt-gradient dialysis assembly from refolded, purified histone octamer for *Xla*-NCP (Dyer et al., 2004). Intriguingly, macro-NCP obtained with this method

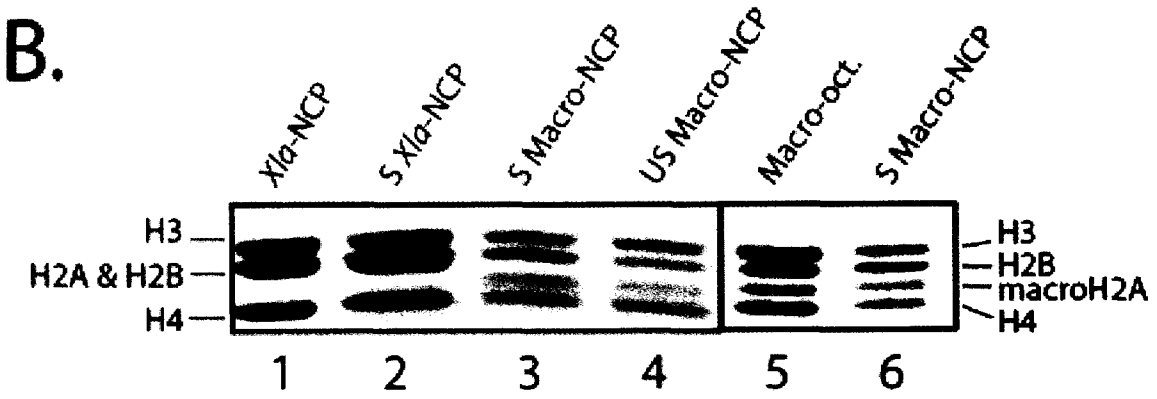


Figure 2.4 B . Histone content of individual nucleosome species, analyzed by SDS-PAGE. Lane 1. *Xla*-NCP control; lane 2: histone content of *Xla*-NCP (excised lane 2 in 4A), lane 3: histone content of shifted macro-NCP (lane 4 in 4A, faster band); lane 4: histone content of unshifted macro-NCP (lane 4 in 4A, slower band); lane 5: macro-octamer; lane 6: histone content of shifted macro-NCPs (lane 4 in 4A, faster band).

reproducibly exhibited complete heat-induced shifting, as observed for canonical *Xla*-NCP (Fig. 2.4C, lanes 1 & 2). Second, we reconstituted major and macro-NCP at low ionic strength, using yeast Nucleosome Assembly Protein 1 (yNAP-1, Fig. 2.4C, lanes 3 to 8). Again, macro-NCP obtained by this method was shifted completely upon incubation at 37 °C. Thus, macro-NCP did not assemble normally from refolded histone octamer using salt dialysis. MacroH2A did assemble into normal nucleosomes from individually purified macroH2A-H2B dimers and $(H3\text{-}H4)_2$ tetramers using salt gradient deposition or NAP-1. This suggests that the macro-H2A containing histone octamer may differ in its ability to dissociate into assembly-competent histone sub-complexes appropriately upon lowering the ionic strength, resulting in an altered *in vitro* assembly pathway.

To test this hypothesis, we monitored the ternary structure of the histone octamer (in the absence of DNA) by analytical gel filtration under varying ionic strengths. Histone octamer, $(H3\text{-}H4)_2$ tetramer and H2A-H2B dimer exhibit distinctly different elution profiles on a size-exclusion column (Dyer et al., 2004). Purified *Xla*- or macro-octamer, previously refolded at 2 M NaCl, was injected onto a column equilibrated with 1.7, 1.4, 1.1, 0.8, and 0.5 M NaCl, and the elution profiles were compared. *Xla*-octamer dissociates into H2A-H2B dimers and $(H3\text{-}H4)_2$ tetramer at 1.1 M NaCl, (Fig. 2.5, upper panel and Table 2.2). In striking contrast, the macro-octamer remains intact down to a concentration of 0.5 M NaCl (Fig. 2.5, and middle panel Table 2.2). Because histones (in particular the $(H3\text{-}H4)_2$ tetramer) bind to DNA at a concentration of ~ 1.0 M NaCl during salt-gradient reconstitution, this inability to dissociate properly may force the deposition of incompletely dissociated histone octamers on the DNA, resulting in the observed non

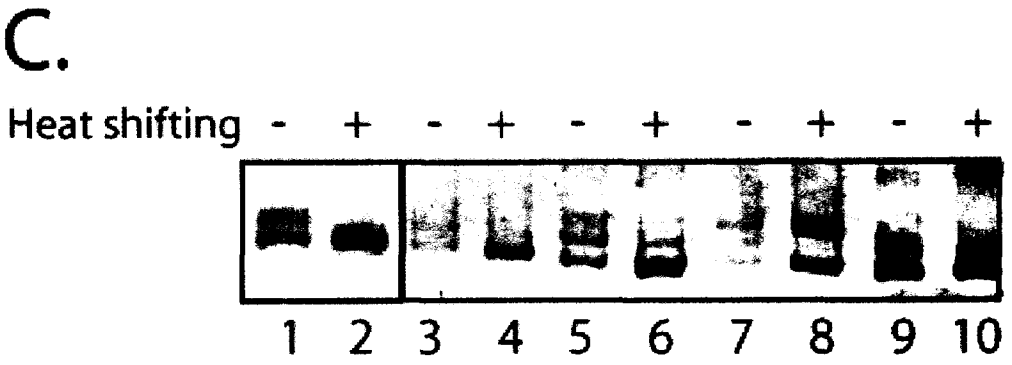


Figure 2.4 C. Comparison of different reconstitution methods. Lane 1 & 2. Unshifted and shifted macro-NCP reconstituted by salt-gradient from independently refolded $(H3\text{-}H4)_2$ tetramer, H2A-H2B dimer and 146 bp DNA. Lanes 3 & 4: unshifted and shifted macro-NCP, obtained by mixing $(H3\text{-}H4)_2$ tetramer, macro$H2A_{HD}$-H2B dimer, yNAP-1 and 146 bp DNA under physiological ionic strength. Lanes 5 & 6. Unshifted and shifted *Xla*-NCP, reconstituted with yNAP-1 using *Xla*-octamer. Lanes 7 & 8: unshifted and shifted yNAP-1 reconstituted macro-NCPs, obtained by mixing macro-octamer with yNAP-1 and 146 bp DNA. Lanes 9& 10: unshifted and shifted *Xla*-NCP reconstituted by salt gradient dialysis from *Xla*-octamer and 146 bp DNA.

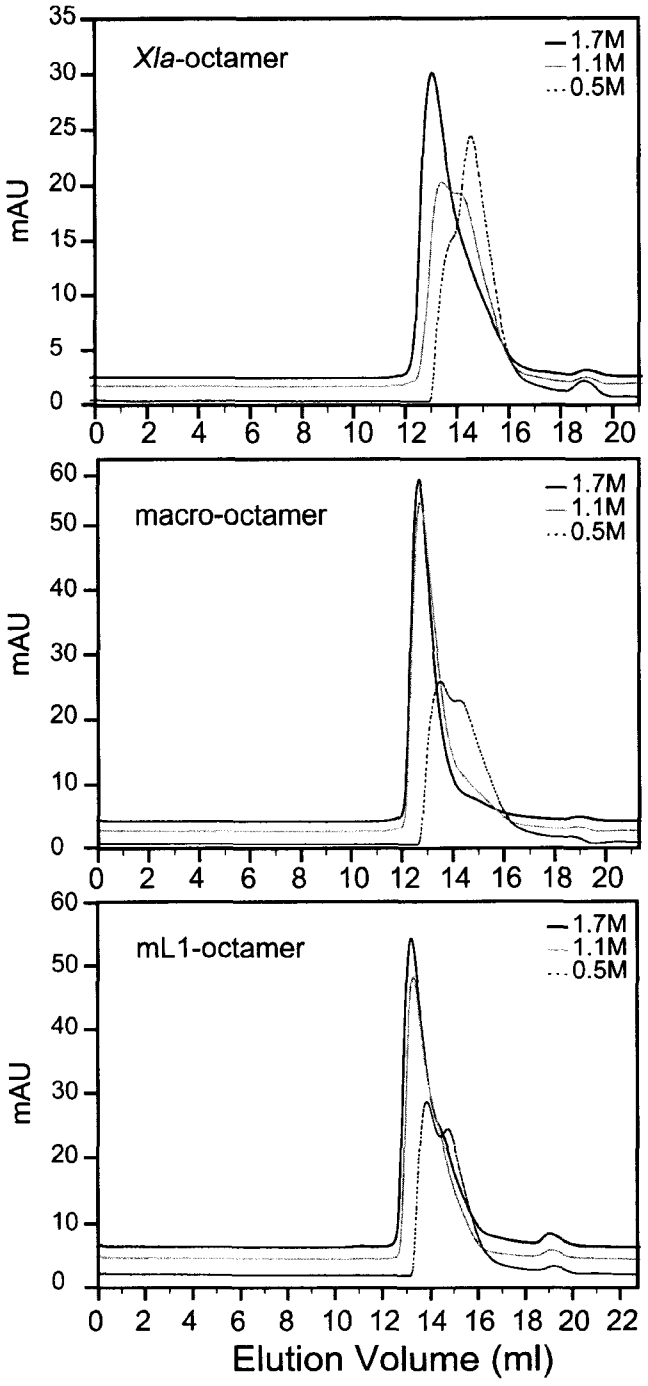


Figure 2.5. Gel filtration elution profiles of different octamers. Profile of *Xla*-octamer (upper panel), macro-octamer (middle panel), and mL1-octamer (bottom panel) at different concentrations of NaCl (1.7 M NaCl, black solid line; 1.1 M NaCl, gray solid line; and 0.5 M NaCl, dashed line). Data from these and additional runs at intermediate salt concentrations (not shown) are summarized in Table 2.2.

Table 2.2: The L1 loop of macroH2A is responsible for the different behavior of macro-octamer. Samples of refolded histone octamer (in 2 M NaCl) were injected onto a Superdex-200 column equilibrated with the indicated salt concentrations, and the elution of peaks (in ml) was tabulated.

	XLA-OCTAMER (ML)			MACRO-OCTAMER (ML)			ML1-OCTAMER (ML)		
	Oct.	Tet.	Dimer	Oct.	Tet.	Dimer	Oct.	Tet.	Dimer
1.7	13.2	-	-	12.8	-	-	13.3	-	-
1.4	13.2	-	-	13.0	-	-	13.3	-	-
1.1	-	13.7	14.4	13.0	-	-	13.4	-	-
0.8	-	13.8	14.6	13.1	-	-	13.4	-	-
0.5	-	13.8	14.85	-	13.7	14.5	-	13.8	15.00

canonical nucleosome species with aberrant electrophoretic behavior and non-canonical histone composition.

2.4.4. Four amino acids in the L1 loop are solely responsible for the properties of macro-octamer and macro-NCP.

In order to determine which region within the histone-like part of macroH2A is responsible for this unusual behavior, we generated mutants of *Xla*-H2A in which parts of the amino acid sequence were exchanged for the corresponding regions from macroH2A. We focused on regions in the macroH2A sequence that were shown to be structurally significantly different compared to major H2A, and that are implicated in interactions that contribute to stabilizing the nucleosome; the L1-loop and the docking domain (Fig. 2.1A).

We replaced the L1 loop of major H2A with that of macroH2A by site directed mutagenesis ($N_{38}YAE_{41}$ to $H_{38}PKY_{41}$) resulting in a mutant designated mL1-H2A ('macro L1 loop'). We used the same approach to replace the docking domain of major H2A with that of macroH2A (L83I, Q84L, and R88A) to generate a mutant designated mDD-H2A ('macro docking domain'). Histone octamers were refolded from these mutant histones as described above, and NCPs in which major H2A was replaced with either mL1-H2A or mDD-H2A (mL1-NCPs and mDD-NCPs, respectively) were reconstituted by salt-gradient dialysis. Both mutant NCPs were incubated at 37° C for 1 hr. While mL1-NCP showed incomplete heat induced shifting (Fig. 2.4A, lanes 5 & 6), and behaved essentially like macro-NCP, mDD-NCP behaved exactly like major NCP, i.e. heat-induced distribution to the shifted form was complete (Fig. 2.4B, lanes 7 & 8).

This suggests that the L1 loop of macroH2A is solely responsible for the changes in assembly that lead to the formation of the aberrant nucleosome species.

We next compared the ability of histone octamers reconstituted with this latter H2A / macroH2A chimera to dissociate into histone sub-complexes. Consistent with the results described above, mL1-octamer behaved exactly like macro-octamer in that it dissociated at 0.5 M NaCl (instead of 1.1 M NaCl), and the elution profiles at different salt concentrations were highly comparable (Table 2.2 and Fig. 2.5 lower panel). Together, these results confirm our hypothesis that the increased stability of the macro-octamer, caused by sequence differences in the four-amino acid stretch of the L1 loop (NYAE changed to HPKY) is the determining factor for the formation of an alternative species represented by the unshiftable band in native gels.

2.5. Discussion

The incorporation of histone variants into chromatin represents an important pathway to alter chromatin structure. Here we show that structural and biochemical differences between NCPs containing the histone-like domain of macroH2A and those reconstituted with major-type H2A can largely be reduced to a four amino acid region in the L1 loop that connects two α-helices of the histone fold domain of H2A or macro-H2A. The nature of this interface is completely altered in macro-NCP, in that the salt-bridges that are present in most major-type H2A sequences from higher eukaryotes are replaced by a conformationally more rigid and hydrophobic interface in macro-NCP. These changes are responsible for a less stringent requirement for high ionic strength to hold the histone octamer together in the absence of DNA. Finally, replacement of four amino acids within

the L1 loop of major H2A with the corresponding amino acids from macroH2A is sufficient to alter the localization of H2A, such that the chimeric protein localizes to the inactive X chromosome with the same efficiency as full length macroH2A, implicating the L1 loop as a motif important in regulating the sites of H2A deposition.

Our crystallographic studies of an NCP containing the histone domain of macroH2A showed that most of the many sequence differences between macro-H2A and major-type H2A (with the exception of the L1 loop) were accommodated without significant changes in nucleosome structure. For example, sequence differences between the docking domain of H2A and macro-H2A, which one would predict to have a significant effect on the structure and stability of this sub-domain, appear to have been compensated for by additional stabilizing alterations, allowing it to maintain its structure. Major surface features of the histone octamer, such as the acidic patch (Luger et al., 1997a), have been maintained in macro-NCP.

Since macro-H2A has been combined with mouse H2B, H3, and H4 to form nucleosomes, the structure presented here also represents the first structure of a mammalian nucleosome core particle. As observed before in a comparison of the structures of *Xenopus laevis* and *Saccharomyces cerevisiae* NCP (White et al., 2001), sequence differences between mouse and *Xenopus laevis* histones are scattered throughout the structure of the NCP and are not restricted to surface-exposed residues (data not shown). Together with the previously published data on yeast and mutant nucleosomes (White et al., 2001; Muthurajan et al., 2004), our data re-emphasize that the overall structure of the nucleosome is actually quite forgiving towards sequence changes,

and that the reason for the high degree of sequence conservation observed in histones can not be attributed only to the maintenance of the structure.

The H2A L1 loop is the only region within the entire nucleosome structure where the two H2A-H2B dimers interact and where the two gyres of the DNA superhelix are tethered together. The four amino acids primarily involved in forming this interface have a profound effect on the stability of the histone octamer at lower ionic strengths. This can be explained by the changed character of the interface (i.e. salt bridges versus hydrophobic interactions). Although the histone octamer is not a physiological entity in the absence of DNA, and preliminary experiments have shown that the overall stability of macro-NCP remains largely unchanged (Chakravarthy et al., 2004), we believe that changes in the character of this interface could have profound effects on the dynamic properties of macro-H2A containing chromatin. For example, the efficiency with which a gene is transcribed by RNA polymerase relies partly on the ability of the transcription machinery and associated factors to temporarily displace one or both H2A-H2B dimers (Kireeva et al., 2002; Belotserkovskaya et al., 2003). Furthermore, the ability of macro-NCP to be remodeled by ATP-dependent chromatin remodeling factors may be affected. Macro-NCP is remodeled in an inefficient manner by the ATP-dependent chromatin remodeling factor SWI/SNF, and the histone-like portion of macro-H2A is responsible for this behavior (Angelov et al., 2003b). Additional studies are required to address whether the inability of macro-NCP to be remodeled depends on the four amino acids in the L1 loop, or whether other regions of sequence divergence are responsible. Finally, our results indicate that relatively minor differences in amino acid sequence can lead to profound differences in *in vitro* assembly pathways.

In vivo targeting studies, which were guided by results obtained from biochemical and structural analysis, confirmed the central role of the L1 loop in H2A function (communication from Dr. Barbara Panning, UCSF). Strikingly, replacing only four amino acids in the L1 loop with the homologous residues from macroH2A are sufficient to bring about *in vivo* localization of H2A to the inactive X-chromosome at levels comparable to wild type macroH2A. This is reminiscent of the situation with H3.3, where single amino acid changes between major-H3 and H3.3 are sufficient to channel these histones to replication-dependent and replication-independent assembly pathways, respectively (Ahmad and Henikoff, 2002b). It is likely that specific chromatin assembly / exchange factors, like those found for histone H2A.Z (Krogan et al., 2003; Kobor et al., 2004; Mizuguchi et al., 2004) and H3.3 (Tagami et al., 2004), will be found for other histone variants, including macro-H2A, and it will be of interest to see whether the regions pinpointed in this study are responsible for targeting specificity.

It appears that sequence and structural differences in the L1 loop are a common theme in histone H2A variants. For example, structural studies have shown significant and completely different changes in the L1 loop of H2A.Z (Suto et al., 2000). The second member of the macroH2A family, macroH2A2 has an L1-loop that seems to be a convolution between the L1-loop of major H2A and that of macroH2A (major H2A: NYAE, macroH2A1: HPKY, and macroH2A2: TFKY). H2A.Bbd, a histone variant that is explicitly excluded from the inactive X-chromosome and is thought to be associated with transcriptionally active chromatin, also exhibits pronounced sequence differences in this area (Chadwick and Willard, 2001b). It is noteworthy that the L1-L1 interface in yeast is also strikingly different from that in higher eukaryotes, perhaps reflecting the

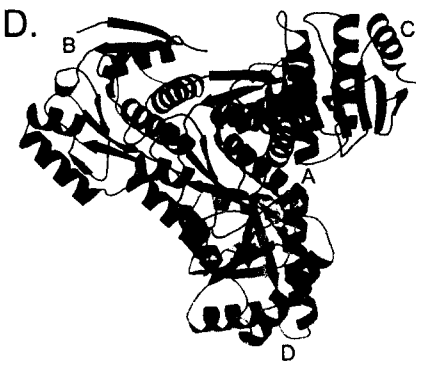
predominantly active status of most chromatin in yeast (White et al., 2001). Why has the L1 loop been chosen as a target of variation in the evolution of H2A variants? Our finding that the L1 loop contributes substantially to the *in vivo* targeting of H2A is certainly a strong indicator that this region may be recognized by specialized, yet-to-be identified assembly factors or chaperones. It is also possible that the L1-L1 interface facilitates the cooperative incorporation of the second histone H2A variant-H2B dimer to ensure that only nucleosomes with two identical H2A variant-H2B dimers are formed. Equally likely, the L1 interface may preclude the formation of such nucleosomes, and may instead favor the incorporation of a major-type H2A-H2B dimer to form nucleosomes with two types of H2A (Chakravarthy et al., 2004). Given the high degree of divergence of the L1 loop region in all H2A histone variants known to date, different histone variants may exhibit different preferences, adding an additional level of complexity to the already complex pathways of chromatin modification by histone variant incorporation.

2.6. Acknowledgements

We thank David Tremethick who kindly donated expression clones for mouse H2B and H3. We also thank Pam Dyer for invaluable help with reagents and GYSK Swamy for sharing his crystallographic expertise. Supported by research grants from the March of Dimes Birth Defects Organization and the NIH to K.L. The crystallographic coordinates and structure factors have been deposited in the protein data base (pdb ID 1U35).

CHAPTER 3

1.6 Å Crystal Structure of the Non-Histone Domain of MacroH2A



We determined the 1.6Å de novo crystal structure of the non-histone domain of macroH2A. We find that the fold of this domain is an α / β fold with a surface that is remarkable in its lack of distinguishing features except for a large hydrophobic patch which may be instrumental in recruiting other heterochromatin associated proteins. All the work presented here is a collaborative effort between GYSK Swamy and Srinivas Chakravarthy, and SC performed the structural analysis. The contents of this chapter in combination with those of chapter 2 are in preparation for publication.

3.1 Abstract

MacroH2A is a histone H2A variant that has an unusual structural organization. It has an N-terminal histone H2A-like domain (aa 1-122) that is 64% identical to major type H2A, a linker region (aa 123-160) and a C-terminal non-histone domain (aa 161-370). It is becoming increasingly plausible that the different domains are functionally independent. While it is now known that the histone domain affects the structure and function of the nucleosome core particle, the role of the linker and non-histone domains remains to be elucidated. To this end, we have determined the 1.6 Å x-ray structure of the non-histone domain spanning a region between aa 180 to aa 367 of macroH2A1.1. It is an α/β fold that is similar both in terms of sequence and structure to a functionally diverse group of proteins, which are functionally heterogeneous. The presence of this fold in a variety of contexts indicates that this fold may be capable of performing a wide range of functions by adapting to different circumambient conditions.

3.2 Introduction

Eukaryotic DNA complexed with histone and several non-histone proteins forms a complex called chromatin (Van Holde, 1988). The structural and functional unit of chromatin is the nucleosome core particle (NCP), which consists of a histone octamer made of two copies each of the four core histones H2A, H2B, H3, and H4 wrapped ~ 1.65 times by 147 bp of DNA (Luger et al., 1997a). One of the mechanisms to regulate accessibility of the nucleosomal DNA to the transcriptional machinery is altering the biochemical composition of the nucleosome by replacing one or more of the core histones with their corresponding histone variants (Malik and Henikoff, 2003). Through subtle differences in sequence, histone variants can bring about changes in the

nucleosome structure that can potentially influence DNA accessibility and as a consequence transcriptional activity of the local chromatin (Fan et al., 2002a; Suto et al., 2000).

Histone H2A has a large number of variants most of which have unique localization patterns. With the exception of H2A.X, which is implicated in DNA repair (Celeste et al., 2003), all histone variants are known to reside in regions of the genome that display unique chromatin structure and function. H2A.Z is hypothesized to be instrumental in preventing the spread of heterochromatin into actively transcribing regions of the genome (Meneghini et al., 2003). H2A.Bbd is excluded from the chromatin of the inactive X-chromosome and is seen in regions that have hyper-acetylated H4, which is one of the markers of transcriptionally active chromatin (Chadwick and Willard, 2001b). MacroH2A is seen in high concentrations at the inactive X-chromosome of adult female mammals (Costanzi and Pehrson, 1998b).

MacroH2A is an unusual H2A variant found in all vertebrates (Pehrson and Fuji, 1998). It is larger than all other known core histones (42kDa) and displays a non-canonical structural organization. In addition to an N-terminal histone domain (aa 1-122), macroH2A has a linker region (aa 123-160) that connects the histone domain to the non-histone domain (161-370) (Pehrson and Fried, 1992). The chromatin of an inactivated X-chromosome shares many of its characteristics with heterochromatin. It has also been seen that macroH2A is more repressive than major type H2A to transcription *in vitro* (Angelov et al., 2003a; Perche et al., 2000b). MacroH2A is found in comparable quantities in mammalian males and several non-mammalian species (Pehrson and Fuji, 1998). This indicates that this histone variant while possibly being important in the

process of X-inactivation also plays a significant role as a down-regulator of gene-expression in other regions of the genome.

MacroH2A is remarkably well conserved between various species. The histone domain is 100% conserved while the linker region is ~93% conserved and the non-histone domain is ~95% conserved between birds and mammals (Pehrson and Fuji, 1998). The non-histone domain of macroH2A has sequence homology with the A1pp domain (Martzen et al., 1999), which is found in organisms ranging from archaebacteria, and bacteria to higher eukaryotes in a variety of cellular contexts. This domain was discovered in a genome wide screen in yeast for proteins involved in tRNA splicing. The A1pp domain was shown to have ADP ribose 1-"phosphate (Appr-1"P) processing activity. Appr-1"P is a byproduct of the NAD^+ dependent reaction that removes the 2' phosphate group from the 3'5' –phosphodiester-2'-phosphomonoester linkage produced during tRNA splicing in yeast (Culver et al., 1993). Studies have shown that the histone domain and non-histone domain are capable of targeting themselves to the inactive X chromosome independently of each other (Chadwick et al., 2001). The histone and non-histone domain of macroH2A may therefore have independent functions, which in turn may influence the structure and function of the local chromatin.

As a first step towards determining the functional significance of the non-histone domain of macroH2A we have determined the 1.6Å crystal structure of the non-histone domain from aa 180-367. It is an α/β fold with a seven strand β-sheet and five α-helices. We find that the net charge of this domain is ~ 0, which is unusual for a chromatin associated protein. There is also an unusually large hydrophobic patch on the surface, which could either mediate self-association or could be part of an interface between macroH2A and

other proteins involved in heterochromatinization. We found considerable structural homology with the N-terminal regulatory domain of Leucine aminopeptidase (Strater and Lipscomb, 1995) and also with some NTPases that have the P-loop motif (Saraste et al., 1990). It is yet to be determined whether the macro domain serves a structural or an enzymatic role to affect the structure and function of chromatin.

3.3. Materials and Methods

3.3.1. Expression and Purification of the non-histone domain of macroH2A

The expression plasmid (pGEX 6P2) for macroH2A (161-367) was a kind gift from Dr. John Pehrson. We modified this construct by introducing a restriction site (NdeI) at amino acid 180 by site directed mutagenesis (Stratagene) and subcloning into pGEX4T2 in order to circumvent the problem of proteolysis and for optimum crystal growth. This construct was then used to transform the E.coli strain BL21-DE3 pLysS. The transformed cells were used to inoculate 6ml of 2xTY medium in the presence of the drugs ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml) and 5% glucose. After being allowed to grow until the medium was visibly turbid this primary culture was expanded to a volume of 100 ml in the presence of the same drugs as above and 5% glucose. This culture was allowed to grow for 1.5 to 2 hrs and then amplified to 3 liters. The culture was then allowed to grow at 37 °C till it reached an optical density (at 600 nm) of 0.4 to 0.6. We then induced expression with 0.4 mM Isopropyl β-D thiogalactopyranoside (IPTG) and brought the culture to a temperature of 25°C and grew it overnight. Bacteria were harvested by centrifugation and were resuspended in $1/10^{th}$ the volume of lysis buffer (150 mM NaCl, 5 mM EDTA, 20 mM Tris HCl -pH7.5, and 1

mM DTT, and 1 mM Benzamidine). This suspension was then flash frozen in liquid nitrogen and stored at -20°C.

The cell suspension was then thawed at 37°C and sonicated until the consistency was water like. The insoluble portion was spun down and the supernatant was loaded on $1/10^{th}$ the volume of pre swollen glutathione-agarose (Sigma) beads and rocked overnight at 4°C. Before loading the sample the beads were washed in water and three volumes of lysis buffer.

The beads were spun down (gently) and the supernatant pipetted out. They were then washed thoroughly in three volumes of PBS (80 mM disodium hydrogen orthophosphate, 20 mM sodium dihydrogen orthophosphate, and 100 mM sodium chloride) followed by three volumes of prescission protease buffer (50 mM Tris.HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT). Prescission protease buffer was added with prescission protease enzyme (Amersham Biosciences) (60-70 µl for every ml of swollen beads). The beads were rocked overnight at 4°C and spun down. The supernatant from this spin contains macroH2A without the GST tag. The beads were washed with three volumes of prescission protease buffer. All the washes and the elute were then analyzed on an 18% SDS-PAGE gel. The protein was then concentrated to 5-10 mg/ml and loaded on a superdex-75 10/30 (size-exclusion) column. The fractions were analyzed on an 18% SDS-PAGE gel and the right fractions were pooled and concentrated to ~25 mg/ml to be used in crystallization screens.

3.3.2. Crystallographic procedures

The non-histone domain of macroH2A (180-367) was crystallized at 28% PEG 2000, 0.2 M ammonium sulphate, 0.1 M sodium acetate pH 5.9. Heavy atom derivatives were obtained by co-crystallizing the non-histone domain of macroH2A with potassium dicyanoaurate [K Au $(CN)_2$]. Data were collected for both native and gold-derivatized crystals at Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline 8.3.1 to a resolution of 1.6 Å and 2.1 Å respectively. Phases were obtained by MAD (Multiple wavelength Anomalous Dispersion) in a two-wavelength experiment. Phases were then extended to 1.6 Å using the data from the native crystal. Denzo and Scalepack were used to index and scale the data (Otwinowski and Minor, 1997). Experimental phases were obtained using the program SOLVE / RESOLVE (Terwilliger, 2003). We then performed multiple rounds of refinement using CNS (Rice et al., 1998). We used the program O was used for model building (Jones et al., 1991). The validity of the model was checked using SA-OMIT maps at different stages of refinement and a composite omit map at the end of refinement.

3.4. Results

3.4.1. Overall Structure of the non-histone domain of macroH2A (aa 180-367)

The non-histone domain of macroH2A spans the region between amino acids 161-367 (Fig. 3.1A). We purified this region by batch affinity chromatography followed by size exclusion chromatography (superdex 75 10/30 column) (Fig 3.1B). Various crystallization screens were tried for this construct. We obtained crystals that showed sub-optimal diffraction properties and the reason was found to be proteolysis of the purified protein and a consequent heterogeneity of the sample used for crystallization. Proteolysis occurred precisely and consistently at amino acid 168 (Fig. 3.1 A), although

A.

Site of proteolysis
Construct 1 Construct 2

161-KQGEVSKAAS ADSTTEGTPA DGFTVLSTKS LFLGQKLQVV QADIASIDSD AVVHPTNTDF YIGGEVGNTL

EKKGGKEFVE AVLELRKKNG PLEVAGAAVS AGHGLPAKFV IHCNSPVWGA DKCEELLEKT VKNCLALADD

KKLKSIAFPS IGSGRNGFPK QTAAQLILKA ISSYFVSTMS SSIKTVYFVL FDSESIGIYV QEMAKLDAN-369

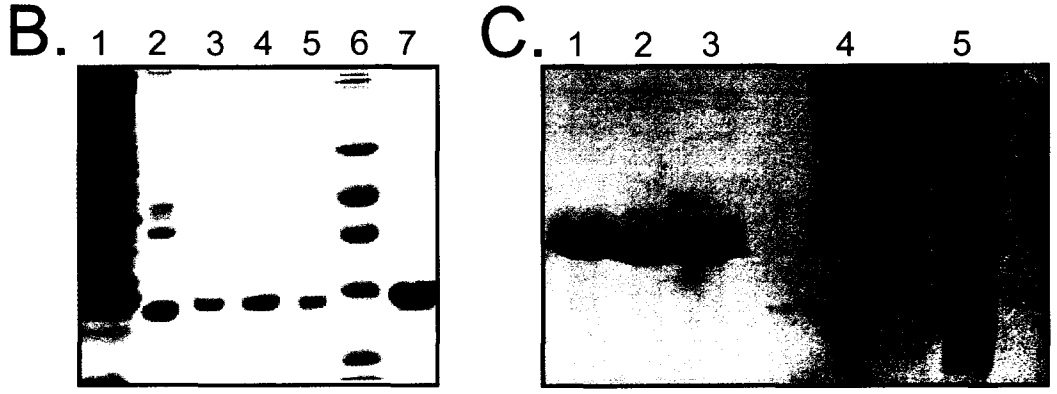
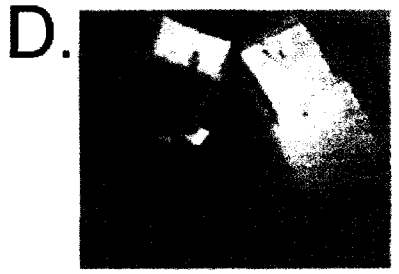


Figure 3.1. Purification of the non-histone domain of macroH2A. A. Amino acid sequence of the non-histone domain of macroH2A (aa 160 – 367). Also shown is the site of proteolysis seen after purification (see text). Sites of truncation for two new constructs are indicated (constructs 1 and 2 respectively and the sequences deleted are indicated in red and blue respectively). **B.** Lane 1: Crude cell-extract containing the non-histone domain of macroH2A (aa 160-367) loaded on the glutathione agarose beads. Lane 2: Glutathione agarose beads after digestion with prescission protease. Lane 3-5: Fractions from the superdex-75 size exclusion column. Lane 6: Molecular weight marker. Lane7: Purified and concentrated non-histone domain of macroH2A. **C.** Lane 1: Non-histone domain of macroH2A (aa 160-367 – original construct). Lane 2: Non-histone domain of macroH2A purified from construct 1 (3.1A: aa 168-367). Lane 3: Non-histone domain of macroH2A purified from construct 2 (3.1A: aa 180-367). Lane 4: Glutathione agarose beads after digestion with prescission protease. Lane 5: Crude cell extract which was loaded on the glutathione agarose beads.

no known protease recognition sequences were found (Fig. 3.1A). We therefore generated two new expression constructs starting at amino acid 168 (SADSTT...) and 180 (GFTVLS...) respectively (Fig 3.1A). The region between aa 180 – 367 was determined to be the most crucial region for *in vivo* targeting in recent studies (Chadwick et al., 2001). Both constructs were used to express and purify the corresponding protein using the same protocol as the original construct (Fig. 3.1C). We then used the newly generated truncated proteins in various crystallization screens (Hampton screen: 96 conditions; Wizard screen: 96 conditions; PEG screen: 24 conditions). The protein spanning the region between amino acids 180 and 367 yielded the best crystals (as characterized on the rotating anode X-ray generator – Rigaku) in the PEG screen (28% PEG 2000, 0.2M ammonium sulphate, 0.1M sodium acetate pH 5.9). We then co-crystallized this protein with potassium dicyanoaurate [K Au $(CN)_2$] to obtain the gold derivative of macroH2A (Fig. 3.1D). We collected data at the synchrotron (LBNL Berkeley, BL 8.3.1) to a resolution of 2.1 Å with the gold derivatized crystal and 1.6 Å with the native crystal (Fig. 3.1 E). Phases were obtained to a resolution of 2.1 Å using MAD (Multiple-wavelength Anomalous Dispersion) in a two wavelength experiment (Table 3.1). Initial electron density maps were calculated over ~45% of the asymmetric unit with a correlation coefficient of 0.7 (table 3.2). Phases were extended to 1.6 Å using the native data set. The structure was refined to a crystallographic R-factor of 0.235 (free R-factor = 0.260) (Table 3.2). A representative region of the final electron density map contoured at 2 σ is shown in Fig 3.2A. The non-histone domain of macroH2A is an α/β fold with a seven strand β-sheet and five α-helices (Fig. 3.2B& C).



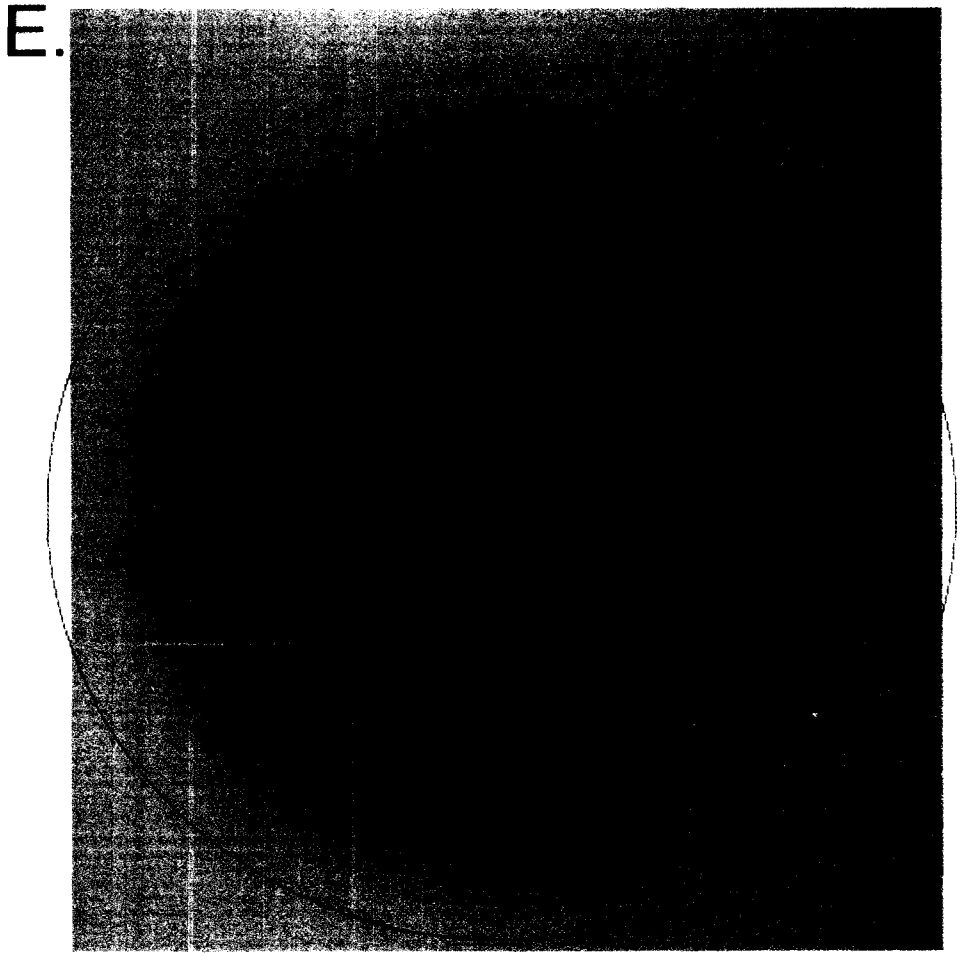


Figure 3.1 D & E. Crystallization and data collection. D. Gold derivatized [K Au $(CN)_2$] crystals of the non-histone domain of macroH2A (182-370). Size in μm indicated. See text for conditions. E. Diffraction pattern form native crystal obtained from BL 8.3.1 (ALS Berkeley). Highest resolution indicated.

Table 3.1: Data collection statistics for the MAD experiment

	Peak	High Energy Remote	Native
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Wavelength (λ) (Å)	1.074800	1.039990	1.074853
Resolution Range (Å)	50 – 2.1	50 – 2.1	50 – 1.6
Unique Reflections	42574	42686	93213
R_{merge} [a](% overall / last shell)	6.7 / 34.9	7.4 / 37.7	6.0 / 41.0

[a]$R_{merge} = \Sigma \mid I_h - \langle I_h \rangle \mid / \Sigma I_h$, where I_h is the mean of the measurements for a single hkl.

Table 3.2: Refinement statistics

Resolution Range	50 – 1.6
Cell Dimensions	a = 83.11, b = 89.791, c = 95.679 α = 90, β = 90, γ = 90
No. of Reflections (Working / Test)	88,260 / 2757
Map Correlation Coefficient[b]	0.7
No. of Protein Residues	750
No. of water molecules	540
No. of Molecules in Asymmetric Unit	4
R-factor[c] / R_{free}	0.235 / 0.260
B-Average (Protein) (Solvent)	30.94 39.36
RMSD from ideality Bonds (Å) Angles (°)	 0.0048 1.2613
Ramachandran Plot (% in allowed region)	93

[b]Correlation Coefficientt = $\int_v \rho_1(x)\,\rho_2(x)\,dx / [\int_v \rho_1(x)^2\,dx \int_v \rho_2(x)^2\,dx]^{1/2}$ (Read, 1986)
[c]R-factor = $\Sigma \mid F_{obs} - F_{calc} \mid / \Sigma F_{obs}$

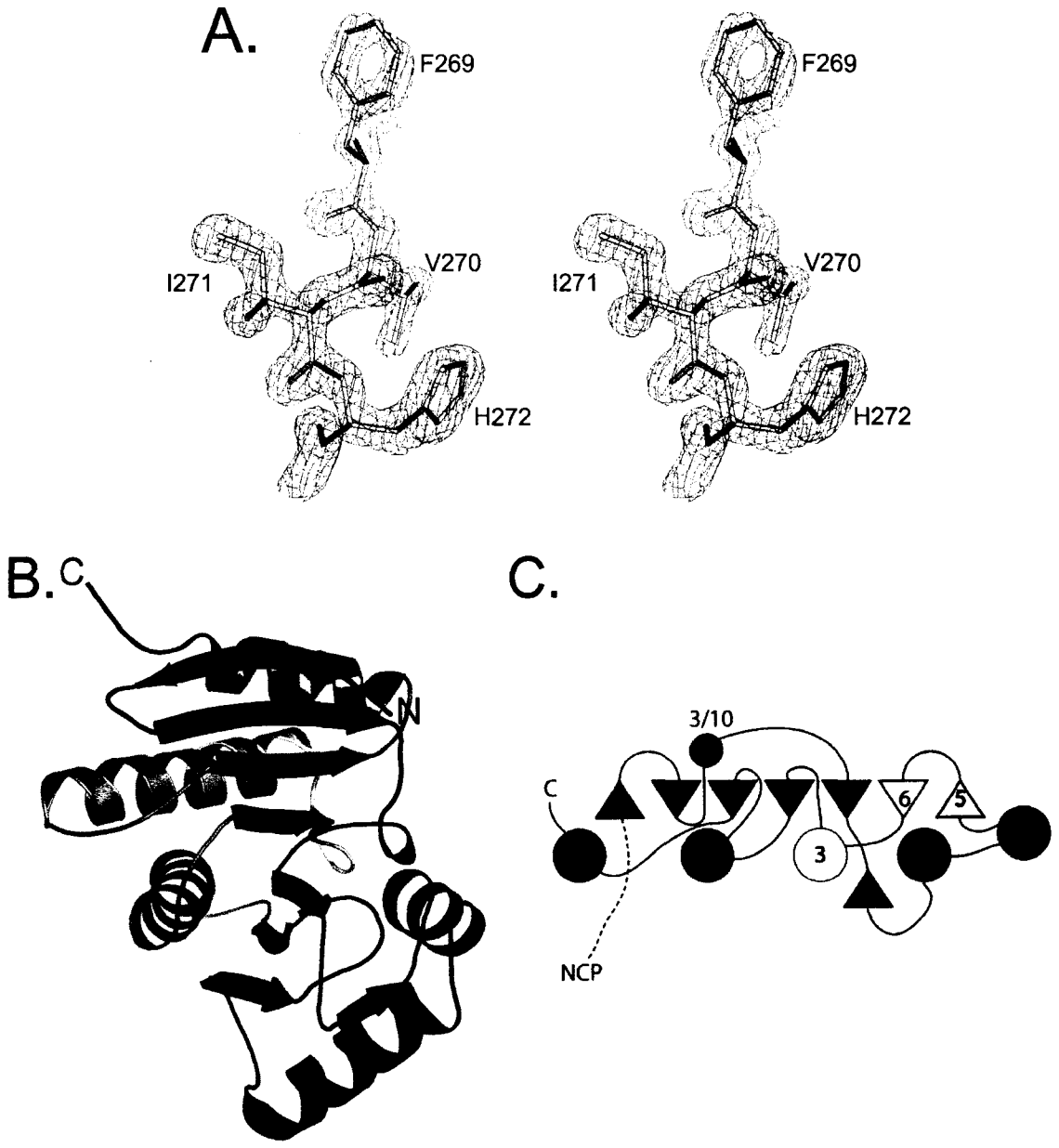


Figure 3.2 Overall structure of the non-histone domain of macroH2A. A. Stereo view of a section of the |2Fo-Fc| electron density map, calculated at 1.6Å and contoured at 2.0 σ, clearly showing a part of the sequence of the non-histone domain (F269 – H272). **B.** Overall structure of the non-histone domain (180 – 370). The N-terminus is in blue and the C-terminus in red with a gradient of the colors of the visible spectrum in between. **C.** A schematic representation of the fold in 1B. Beta-strands are depicted as arrowheads, and helices as circles.

and there are four molecules in the asymmetric unit (Fig. 3.2D). The β strands are arranged in the order 1287365. Strands 1 and 2 are anti-parallel, 2, 8,7, 3 and 6 are parallel, 6 and 5 are anti-parallel (Fig. 3.2B & C). The region between amino acids 180-208 which was predicted to be a leucine zipper like structure (Pehrson and Fried, 1992) in fact constitutes the first two strands of the 7 strand β-sheet (Fig.3.2E). A remarkable feature of this fold is that the β-sheet is completely protected only on one face by α-helices while the other face remains partially solvent exposed (Fig. 3.2 B & C).

3.4.2. Surface Features of the Non-Histone Domain of MacroH2A

The net surface charge of the non-histone domain of macroH2A is 0 (as calculated by the program DELPHI), which is unusual for a chromatin-associated protein (Fig 3.3 A). The surface is also devoid of any distinguishable contours and is unusually smooth. Due to these properties, it is less likely that macroH2A functions via a direct association with polynucleotide. It is still likely that macroH2A recruits some other protein(s), which in turn can bind to polynucleotides. A large hydrophobic patch on the surface that includes residues 183 to 186 (FTVL) and residues 356 to 360 (IGIYV) could be part of an interface between macroH2A and any other heterochromatin-associated protein (Fig 3.3). To see if this hydrophobic surface mediates self-association between multiple copies of the protein (dimerization or oligomerization) we performed analytical ultracentrifugation studies (sedimentation velocity). The results show that the non-histone domain does not self-associate and exists in solution as a monomer (data not shown). This leaves the hydrophobic patch available as a potential scaffold / platform for the recruitment of other heterochromatin-associated proteins.

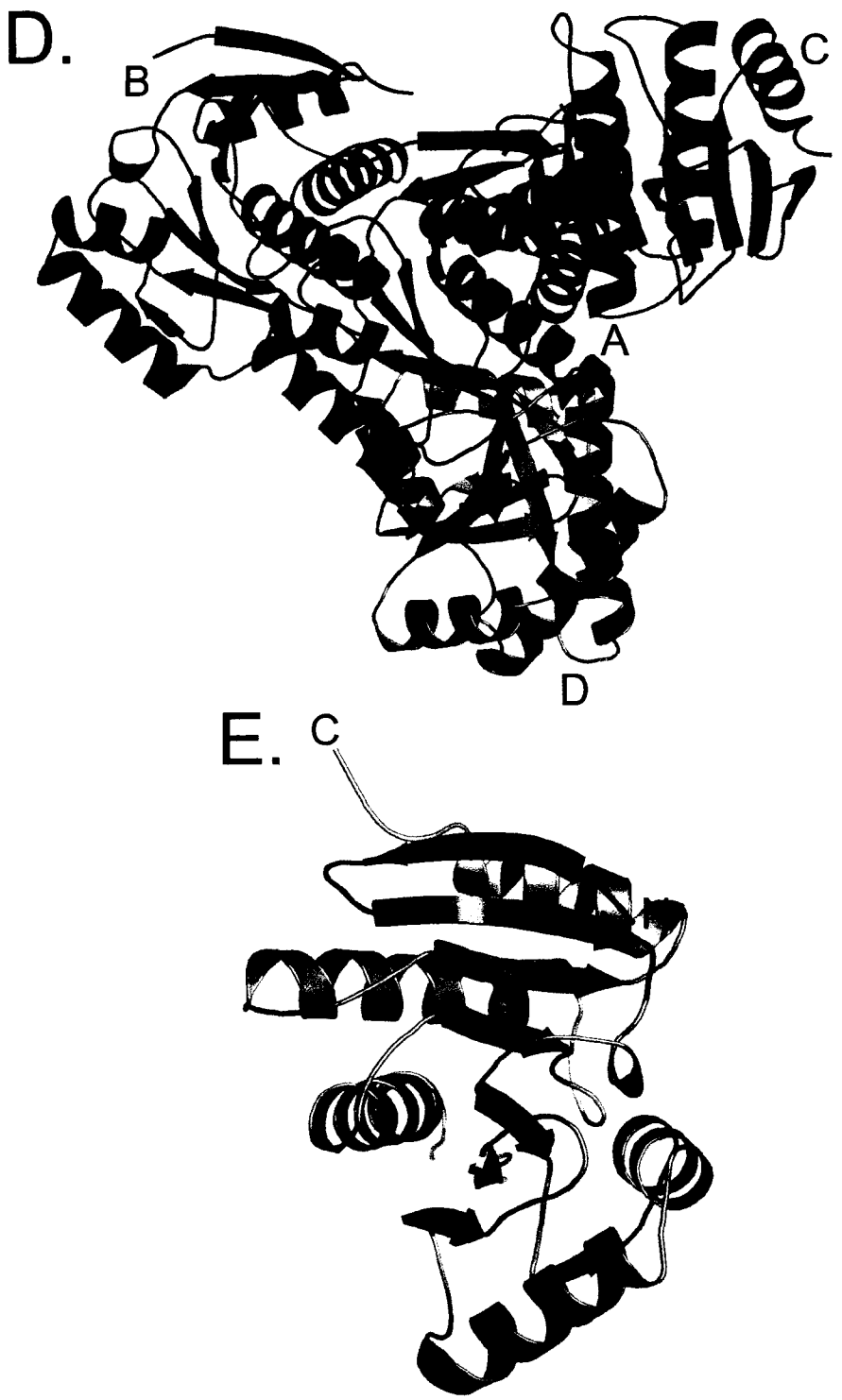


Figure 3.2 D. Arrangement of the 4 molecules of the macroH2A$_{NHD}$ in the asymmetric unit. No protein-protein interactions were seen that would suggest oligomerization. **E.** The region between aa 180 – 208 is not a leucine zipper as suggested in literature (Pehrson and Fried, 1992) but constitutes the first two strands of the 7-stranded beta-sheet (highlighted in red).

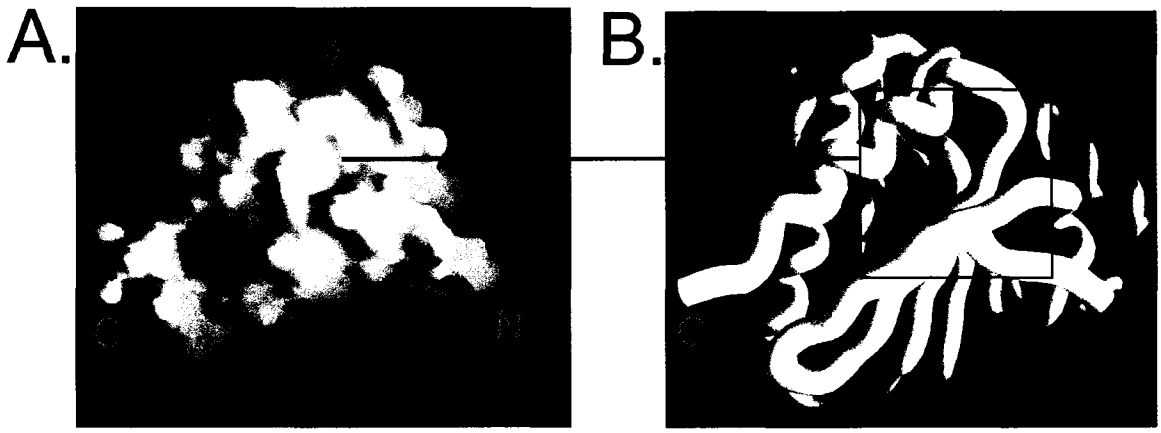


Fig.3.3. The surface of the macro domain (aa 182 – 369) has no distinct features. A. The surface charge representation of the non-histone domain of macroH2A. Basic regions are in blue, acidic regions in red and neutral regions in white. The boxed area encompasses a large hydrophobic region that includes residues F183 to L186 and I356 to V360. **B.** C-α backbone of the non-histone domain of macroH2A in the same orientation as A.

3.5. Similarities between the non-histone domain of macroH2A and other proteins.

The non-histone region of macroH2A is homologous to the A1pp domain (from the Appr-1"-p processing enzyme - YBR022Wp in yeast). This domain is found in various unrelated proteins. In addition to the C-terminus of macroH2A, the A1pp domain is found in the non-structural proteins of various ssRNA viruses, and in a variety of other contexts in eukaryotes. It is also found on its own in bacteria, archaebacteria, and eukaryotes. The presence of this domain in such an extensive range of organisms and cellular contexts alludes to an important and omnipresent function, but does not aid progress in our understanding of the specific function macroH2A performs.

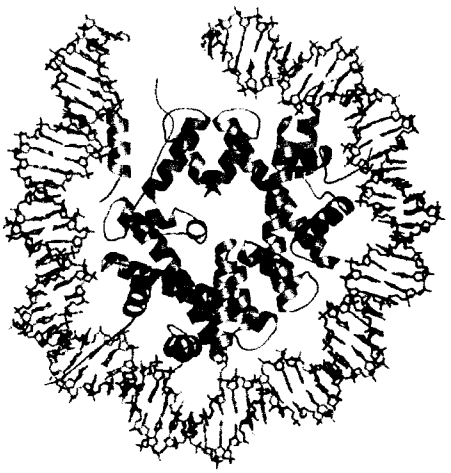
The non-histone domain of macroH2A also shows structural similarity with some proteins with which it doesn't share sequence homology as determined on the Dali server (Holm and Sander, 1993). It shows the strongest similarity with the N-terminal domain of leucine amino peptidase (Dali Z score 10.8). The function of this domain is not clear but its homologue in amino peptidase A (PepA) from *E.coli* is said to have the ability to bind DNA (Strater et al., 1999). While this may suggest a similar mode of function for macroH2A one must consider the sequence heterogeneity that can potentially influence the surface features of the protein. The net charge on the non-histone domain of macroH2A for example is 0, whereas that of the N-terminal domain (1-166) of PepA is 7.0 (as calculated in the program DELPHI). This in turn may affect the protein's ability to form a nucleoprotein complex. This domain is also similar to several NTP hydrolases. The highest degree of similarity among NTPases is with the hexamerization domain of N-ethylmaleimide-sensitive fusion protein (Dali Z score 5.2), which is an Mg^{2+}

dependent ATPase. Most of these NTPases are characterized by the presence of a well-studied and characterized structural motif called the P-loop. The sequence conservation in the nucleotide-binding pocket associated with the P-loop has been studied comprehensively (Saraste et al., 1990; Via et al., 2000). NTP-binding by the P-loop is contingent upon a set of specific sequence requirements, which are not obvious in the macro domain. While this does not entirely rule out nucleotide binding by macroH2A it remains to be substantiated by experimental evidence.

There is indirect evidence for macroH2A binding to XIST (X Inactivation Specific Transcript), an untranslated RNA that coats the inactive X-chromosome (Rasmussen et al., 2001). There is also a large region connecting the histone domain and the non-histone domain of macroH2A (123-179), which is particularly rich in the basic amino acids lysine and arginine (Pehrson and Fried, 1992). In the absence of direct evidence it is therefore not possible to exclude the possibility of this "linker" region binding to XIST RNA. Studies show that the macro domain acts as a direct road block to the recruitment of transcription factors (Angelov et al., 2003a). This may be achieved by either the direct DNA binding implied above or by recruitment of other factors necessary for chromatin silencing (both structural, e.g. HP1 and / or enzymatic e.g. HDACs). Recent studies have alluded to co-localization of macroH2A with HP1 (Chadwick and Willard, 2003; Hoyer-Fender et al., 2004; Turner et al., 2001). Work inspired by our structural studies has indicated that macroH2A might directly associate with HDAC1 through a large hydrophobic patch on the surface (Fig. 3.3) (communication from Dr. Saadi Khochbin; INSERM, Grenoble). In spite of the wide range of sequence and structural similarities, it is not possible to rule out an entirely novel *modus operandi* for macroH2A.

CHAPTER 4

Structural characterization of histone H2A variants



The contents of this chapter were published in the *69th Cold Spring Harbor Symposia for Quantitative Biology: Epigenetics.* We analyze the structural trends that have become clear in light of studies in the lab on histone H2A variants. Data from experiments that were done to determine the influence of these variants on nucleosome structure, and stability are also included. We are beginning to recognize patterns in the evolution of histone variants and their potential functional significance.

The author list for this paper is:

Srinivas Chakravarthy, Yunhe Bao, Vicki Roberts, David Tremethick, Karolin Luger.

All the experiments except those that involve H2A.Bbd were done by SC. YB contributed the data on H2A.Bbd. The surface charge diagrams were contributed by VR.

4.1. Introduction

Eukaryotic DNA associates with an equal amount of protein to form chromatin, the fundamental unit of which is the nucleosome core particle (NCP). An NCP consists of two copies each of the four core histones H2A, H2B, H3, and H4. This histone octamer binds 147 base pairs of DNA around its outer surface in 1.65 tight superhelical turns (Luger et al., 1997a; Richmond and Davey, 2003b); Fig. 4.1A). Linker histones and other non-histone proteins promote or stabilize the folding of nucleosomal arrays into superstructures of increasing complexity and largely unknown architecture (for review see Hansen, 2002). Covalent modification of the core histones and variations in the fundamental biochemical composition of nucleosomes distinguish transcriptionally active from inactive chromatin regions, by either changing the structure of the nucleosomes, by altering their ability to interact with other protein factors, or by modifying their propensity to fold into varying degrees of higher order structures (or by any combination of the above). Studying the mechanism for establishing distinct chromatin domains is essential to understanding differential regulation of gene expression and all other DNA-dependent processes. Much progress has been made in this direction in the past few years.

Substitution of one or more of the core histones with the corresponding histone variants has the potential to exert considerable influence on the structure and function of chromatin. Histone variants are distinct non-allelic forms of conventional, major-type histones that form the bulk of nucleosomes during replication and whose synthesis is tightly coupled to S-phase. Histone variants are characterized by a completely different

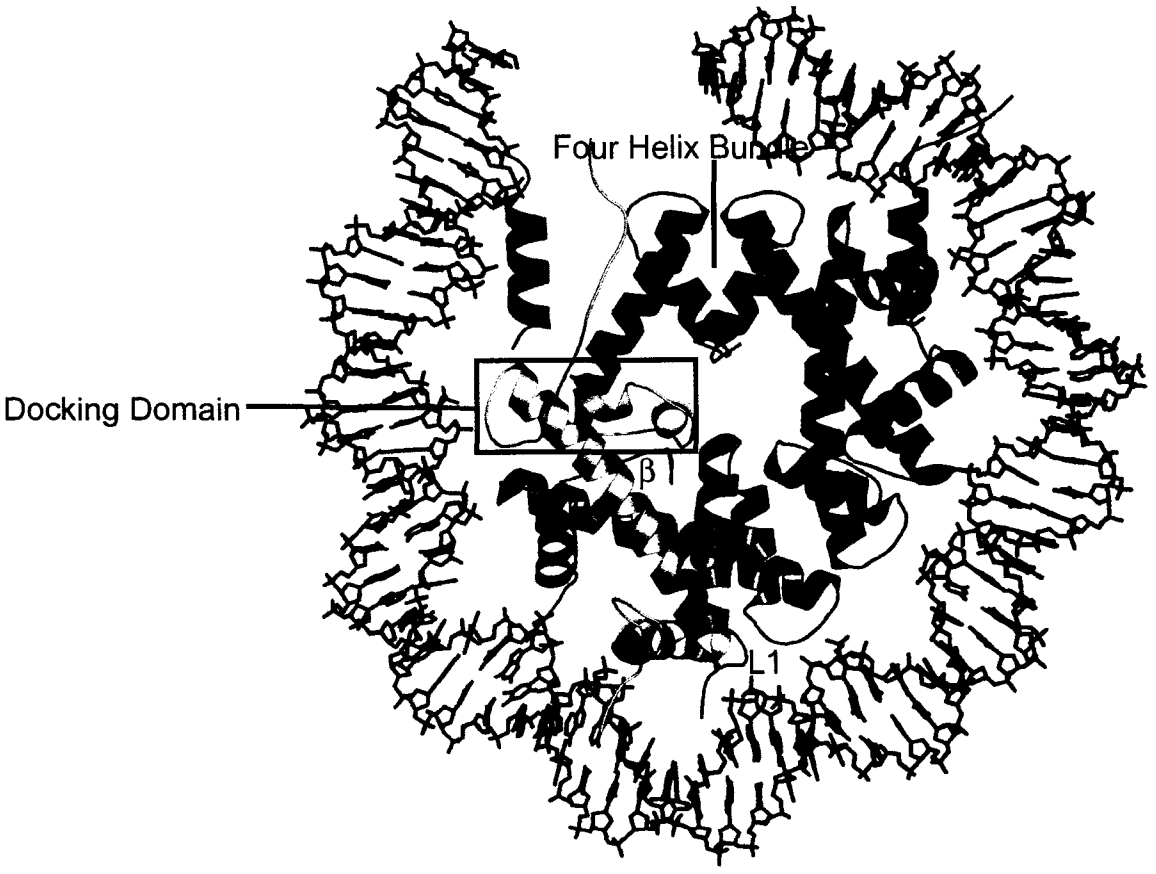


Figure 4.1. A. Overview of the nucleosome structure. Only 74 base pairs of DNA and associated proteins are shown. Yellow: H2A; red: H2B; blue: H3; green: H4. The four helix bundle formed by the H3 molecules and the H2A docking domain are indicated. Other structural features are also indicated.

expression pattern that is not restricted to S-phase. They are found in most eukaryotic organisms, and are expressed in all tissue types (unlike some H2B isoforms that are only found in specialized tissues such as testes). Compared to their major-type counterparts, histone variants exhibit moderate to significant degrees of sequence homology (Fig. 4.1B). H2A.X (82 %) and H3.3 (~96%) are the least divergent of all histone variants. H2A.Z (~ 60 %), macroH2A (~ 65 %), H2A.Bbd (40 %), and CenpA which has a 93 amino acid domain that is 62 % identical to H3 (Palmer et al., 1991; Sullivan et al., 1994) are increasingly divergent in their histone moiety from H2A and H3, respectively. As is the case with histones in general, the structured regions of the histones (encompassing histone folds and extensions) are more conserved than the histone tails. The structured region of H2A.X is 97% conserved to its major-type H2A counterpart, that of macroH2A 70%, H2A.Z 66%, and H2A.Bbd 48%, respectively. MacroH2A is unique in that it contains an additional non-histone like domain that is connected to the histone-homology domain by a flexible linker (Pehrson and Fried, 1992). All histone variants are highly conserved between different species. In many cases, they are even more conserved than their major-type paralogs (Sullivan et al., 2002b), indicating that they all have evolved to fulfill important functions that cannot be accomplished by major-type H2A and H3, as has been demonstrated for H2A.Z (van Daal and Elgin, 1992; Clarkson et al., 1999; Faast et al., 1999).

While the *modus operandi* of most of histone variants remains unknown they are all characterized by unique *in vivo* localization patterns, which in turn shed light on their putative function. H2A.X is distributed throughout the genome. It is implicated in double stranded DNA repair (Celeste et al., 2003; Rothkamm and Lobrich, 2003), and

B.

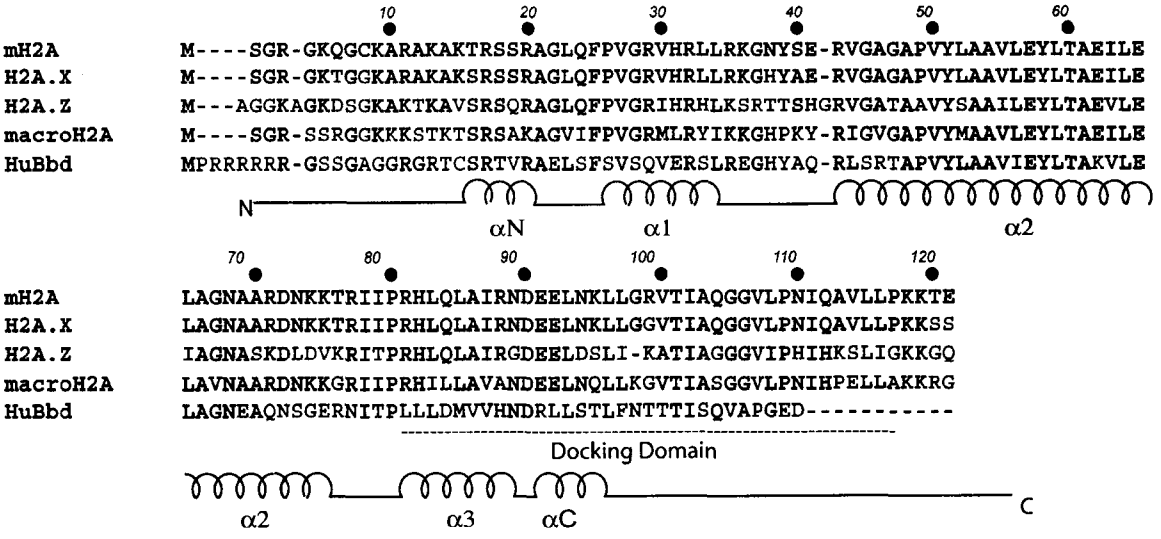


Figure 4.1 B. Sequence alignment of the histone domain of human H2A.X, mouse H2A.Z, human macroH2A, and mouse H2A.Bbd with major-type mouse H2A. Bullets: every tenth residue in major H2A; black: identical residues; blue: similar residues; red: different residues. Also indicated are the secondary structure elements of the histone fold (α1, α2, and α3) and the loops and extensions (L1, L2, αN, and αC).

is necessary for programming DNA breakage that occurs in developing lymphocytes (Bassing et al., 2003). The presence of H2A.Z within euchromatin plays a role in preventing silencing from spreading into regions of the chromosome that are normally transcriptionally active (Meneghini et al., 2003). Interestingly, H2A.Z can co-exist with Sir proteins at the telomere (Krogan et al., 2003). Most recently, H2A.Z has also been shown to play a role in chromosome segregation (Rangasamy et al., 2004). MacroH2A is found at the inactive X-chromosome of adult female mammals, which consists predominantly of heterochromatin and is transcriptionally inactive (Costanzi and Pehrson, 1998b), while H2A.Bbd co-localizes with hyper-acetylated histone H4, indicating that it might be associated with actively transcribed chromatin (Chadwick and Willard, 2001b). The histone H3 variant H3.3 is thought to be associated with active chromatin, and CENP-A is a major component of centromeric heterochromatin (Ahmad and Henikoff, 2002b; Vermaak and Wolffe, 1998).

Here we will summarize and review available structural information on nucleosomes and chromatin containing histone H2A variants, and will attempt to explain how structure relates to their varied function. We note that an exhaustive review of all biological and functional data would clearly exceed the scope of this manuscript. We will present a hypothesis why 'true' histone variants have only been identified for histone H2A and H3, and show data in support of our hypothesis that particular regions in the H2A amino acid sequence appeared to have been targets during the evolution of H2A histone variants.

4.2. Why are there no H4 and H2B histone variants?

All known true histone variants are replacements for either histone H3 or H2A (reviewed in (Malik and Henikoff, 2003); (Henikoff et al., 2004). From a structural vantage point, we hypothesize that this is the case because only H3 and H2A are engaged in homotypic interactions (Fig. 4.2). In contrast, neither H4 nor H2B interact with the other H4 or H2B molecule within the histone octamer. A four-helix bundle formed by residues from the two H3 chains holds together the $(H3\text{-}H4)_2$ tetramer, which is stable in the absence of DNA under physiological conditions (Fig. 4.2A, B). This interface, which is characterized by a combination of salt bridges, ionic interactions, and some hydrophobic contacts (Luger et al., 1997a) has been proposed many years ago to be conformationally flexible, especially in the absence of the (H2A-H2B) dimers (Hamiche et al., 1996; Chen et al., 1991; Protacio and Widom, 1996). In contrast, the interface formed between two H2A molecules is quite small and only exists in the context of a folded nucleosome (Fig. 4.2C, D). It is however the only point of contact between the two (H2A-H2B) dimers in a nucleosome, and thus may be in part responsible for highly cooperative incorporation of the two H2A-H2B dimers, as well as for tethering the two gyres of the DNA superhelix together. Intriguingly, major sequence differences between H3 and the centromeric H3 variant are found in this four-helix bundle region (Shelby et al., 1997; Black et al., 2004b). Similarly, variability among the many H2A variants themselves, and differences between variants and major-type H2A are found in the L1 loop (Fig. 4.1B). This suggests that sequence variability in these regions of self-interactions may serve to ensure that only nucleosomes with two identical H2A or H3 'flavors' are formed.

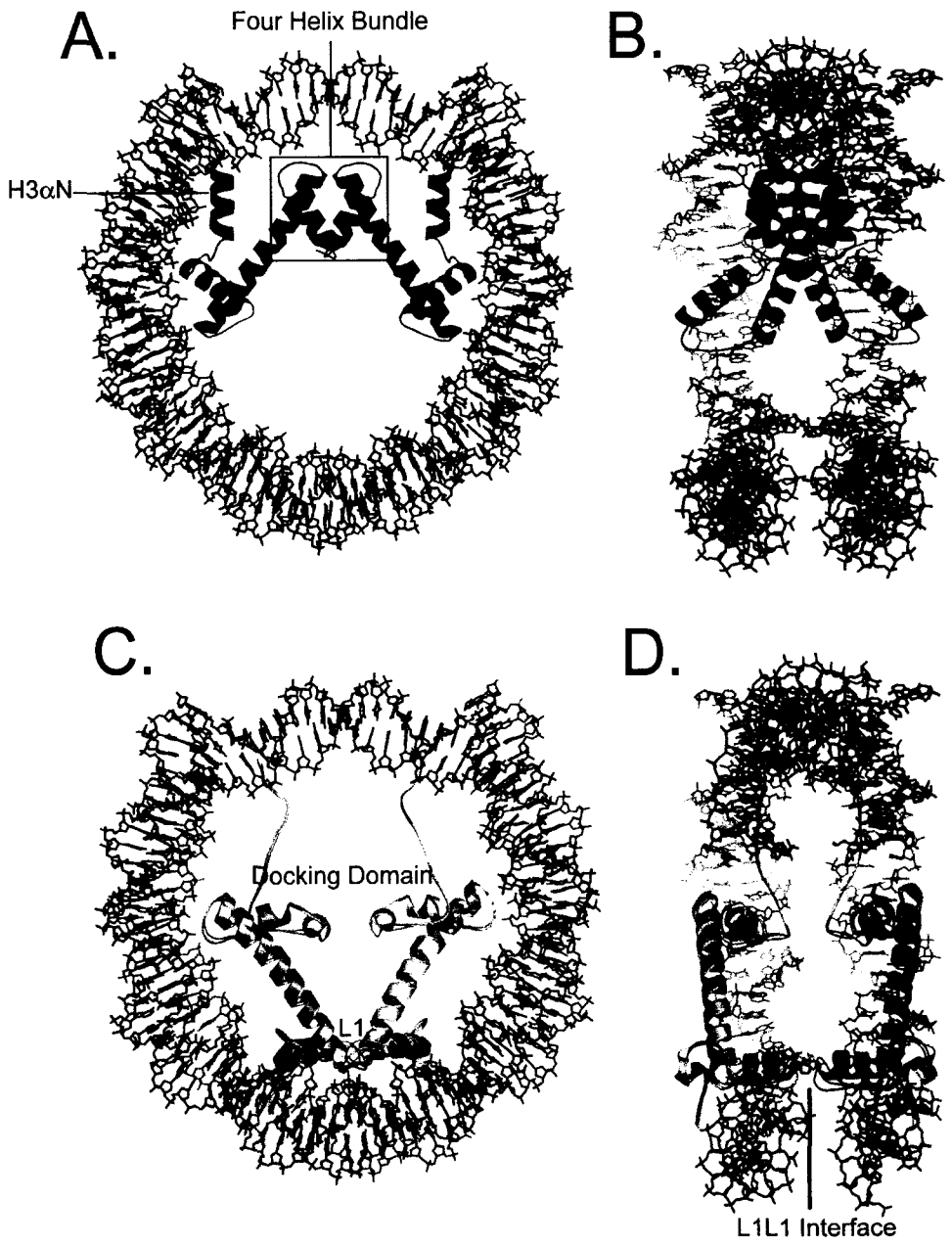


Figure 4.2. H3 and H2A fulfill special roles in the NCP. A. Structure of the NCP showing only the H3 chains (viewed down the superhelical axis). **B.** The same structure after rotation by 90° around the y-axis with parts of DNA omitted for clarity. **C.** Structure of the NCP viewed down the superhelical axis with only the H2A chains shown. The structures of NCPs containing major type H2A from Xenopus laevis (yellow: pdb entry 1AOI), H2A.Z (wheat: pdb entry 1F66) and macroH2A (gray: 1U35) have been superimposed. **D.** The structures shown in C shown in the same orientation as B.

Equally likely, these interfaces may preclude the formation of such nucleosomes, and may instead favor the incorporation of a major-type histone H3 with variant H3, or major-type H2A with variant H2A, respectively, as will be shown below.

A second common feature of H3 and H2A that is not shared by H4 and H2B is the fact that they both are involved in the organization of more than the requisite ~ 70 base pairs of DNA that are bound by a canonical histone fold dimer (Luger and Richmond, 1998a), due to quite extended regions in both H3 and H2A outside the histone fold. The N-terminal helix of H3 (αN) is positioned to interact with the penultimate turn of the DNA double helix before it exits the confines of the nucleosome (Fig. 4.1A and 4.2A). It is held in position by the H2A docking domain (Bao et al., 2004b). The C-terminal tail of H2A is poised to interact with linker DNA that extends beyond the 147 nucleosomal base pairs (Fig. 4.1A and 4.2C).

Thus, unlike H2B and H4, replacement of H3 and H2A with histone variants has the potential to affect DNA organization (or exit angle) at the penultimate 15 base pairs of nucleosomal DNA, with potential implications for higher order structure. As pointed out above, sequence variations that are specific to either H3 or H2A variants have the potential to increase the pool of theoretically possible nucleosomes, in being able to choose their interaction partners. Finally, in light of the current models for linker histone binding to the nucleosome (Crane Robinson, 1997; Hayes and Hansen, 2001), histone H1 and its variants and isoforms are more likely to interact with H3 and H2A than with H2B and H4. One intriguing (but as yet untested) model for how histone variants exert their function in chromatin may be by modulating the interaction between the nucleosome and linker histones or non-histone architectural chromatin proteins, such as HP1. For

example, it is conceivable that nucleosomes harboring certain H2A or H3 variants may be unable to interact with linker histones or other architectural proteins, or that they may prefer certain H1 isoforms over others, with profound effects on gene regulation and / or chromatin higher order structure.

4.3 Structural characteristics of nucleosomes and chromatin containing histone H2A variants

4.3.1 H2A.X

Amino acids (1-120) of H2AX are very similar in sequence to major H2A; indeed, with an only two amino acid difference in the structured domain it is safe to assume that the structural properties of a mono-nucleosome are likely to remain unaffected. The two sequence changes in the structured domain of H2A.X are in the L1 loop and in the docking domain, respectively, and are also found to distinguish macroH2A from major-type H2A (H2A N38 to H, and R99 to G, Fig. 4.1B).

The C-terminal domain of H2A.X is phosphorylated, and it is this phosphorylated form of this variant that is implicated in DNA repair (Celeste et al., 2003). One possible mode of H2A.X action could be via recruitment of repair proteins to the site of DNA-damage. It is indeed found that formation of Nbs1, 53bp1, and Brca1 foci at the damage sites is severely impaired in H2A.X -/- cells, as well as suppressing genomic instability (Celeste et al., 2003).

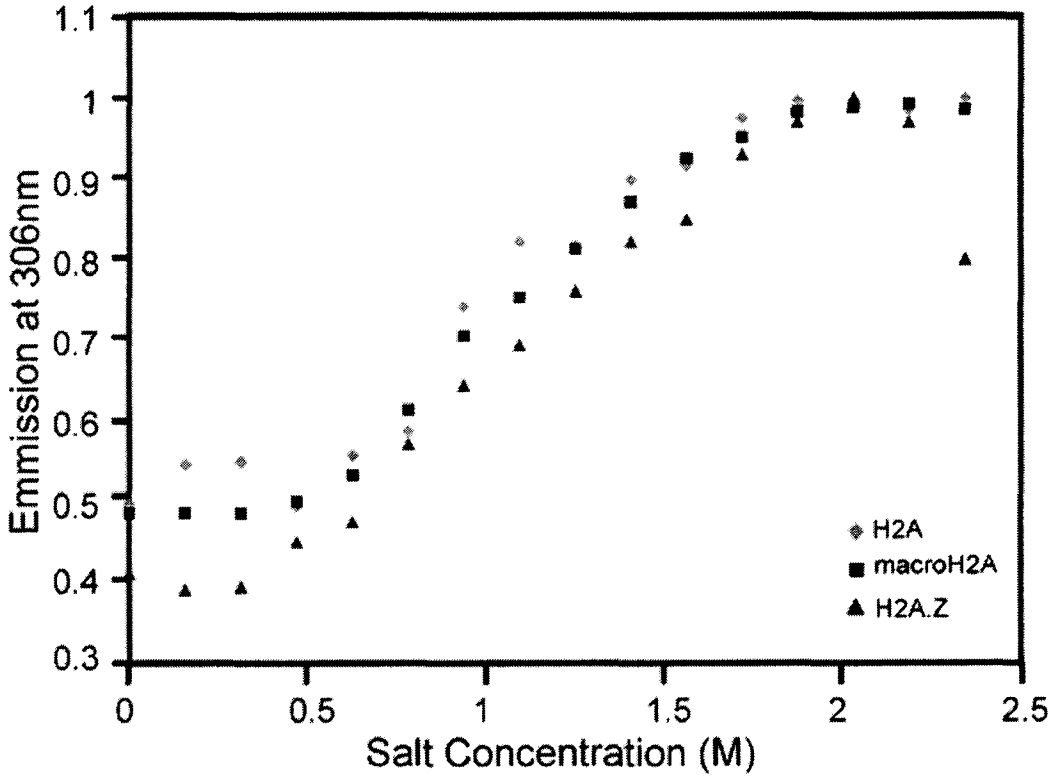
4.3.2 H2A.Z

H2A.Z, which is essential in *Drosophila*, mouse and *Tetrahymena* (van Daal and Elgin, 1992; Liu et al., 1996; Faast et al., 1999) is the histone variant whose structure and

function is perhaps best studied among all histone variants. While not being essential in budding yeast, H2A.Z (Htz1) functions to prevent the spread of heterochromatin into euchromatin (Meneghini et al., 2003) and plays a role in transcription with a function that is partially redundant with specific chromatin remodeling complexes.

The crystal structure of an NCP in which major H2A is replaced by H2A.Z reveals no major differences in the path of the DNA superhelix or in the nature of protein-DNA interactions (Suto et al., 2000). While tyrosine quenching revealed no major change in nucleosome stability *in vitro*, a more sensitive fluorescence resonance energy transfer (FRET) approach (Park et al., 2004b) revealed that the sequence changes of H2A.Z result in subtle stabilization of the (H2A-H2B) dimer interaction with the $(H3\text{-}H4)_2$ tetramer-DNA complex upon increasing ionic strength (Fig. 4.3).

Amino acid changes in the H2A.Z docking domain also contribute to an extended acidic patch, which is a prominent feature of the nucleosome surface (Suto et al., 2003). The docking domain is the region that is essential for *Drosophila* development (Clarkson et al., 1999). This region of H2A.Z (and the N-terminal tail) has been shown to interact with the pericentric heterochromatin-binding protein INCENP (Rangasamy et al., 2003), a protein critical for proper chromosome segregation. Thus, it appears that subtle sequence differences between major-type H2A and H2A.Z bear a direct relevance to its ability to interact with non-histone proteins. Restriction digest analysis of an *in vitro* chromatin model system (H2A or H2A.Z nucleosome arrays assembled on a DNA template that contains 12 repeats of a 208 base-pair nucleosome positioning sequence)



`Figure 4.3 The tyrosine quenching profile shows no pronounced differences in the overall stability of macro-NCP, H2A.Z-NCP, and major-NCP. Tyrosine quenching profiles in response to increased ionic strength for nucleosomes in which the histone domain of macroH2A, or H2A.Z, has replaced majorH2A. Values have been normalized using the formula $A_n = A / A_h$, where A_n is the normalized value, A is the observed signal, and A_h is the highest signal observed. *Diamonds*: major-NCP; *squares*: macro-NCP; *triangles*: H2A.Z-NCP.

in combination with sedimentation velocity studies established that H2A and H2A.Z nucleosomes assemble with comparable affinity (Fan et al., 2002b). Consistent with the increase in nucleosome core stability, H2A.Z arrays are more regularly spaced than H2A arrays. Sedimentation velocity analysis at elevated salt concentrations was used to determine that H2A.Z facilitates intramolecular folding of nucleosome arrays resulting in the formation of a folded state (55S). This folded state is thought to reflect the canonical 30nm fiber. On the other hand H2A.Z is found to hinder the oligomerization mediated by interactions between nucleosome arrays (Fan et al., 2002b). Therefore, H2A.Z arrays form higher-order chromatin structures that are distinctly different from H2A arrays. Consistent with this, H2A.Z is located at constitutive heterochromatin in mammalian cells and is required for faithful chromosome segregation (Rangasamy et al., 2003; Rangasamy et al., 2004). Taken together, H2A.Z may have a dual role in modulating the recruitment of non-histone proteins and in chromatin fiber folding. These features would enable H2A.Z to play a role in a number of different processes that include gene expression and mitosis.

4.3.3 MacroH2A

MacroH2A is perhaps the most unusual of all histone variants due to its large size (~370 amino acids and a molecular weight of 42 kDa) and tripartite structural organization. The N-terminal third (amino acid 1-122) is ~64 % identical to major H2A (Pehrson and Fried, 1992). This region is followed by a highly basic stretch (amino acid 132-160), which is homologous to the C-terminus of linker histone H1 (57 % identical to sea urchin H1γ over 30 residues; (Pehrson and Fried, 1992). Amino acids 161-370 form a tightly folded domain whose high-resolution structure we have determined recently (G.Y.S.K. Swamy

and S. C, unpublished data). The structure of an NCP in which H2A has been replaced by the histone domain of macro-H2A shows that the significant differences in sequence are accommodated with surprisingly minor structural changes (S. C. et al., unpublished data, Fig. 4.2 C, D). The most pronounced structural divergence is found in the L1 region of macroH2A. Changes in the electrophoretic behavior of mono-nucleosomes containing the H2A-like domain of macroH2A, and in the stability of the histone octamer at decreased ionic strength and in the absence of the stabilizing influence of the DNA were observed. A mutant of major-type H2A in which four amino acid in the L1 loop of major-type H2A (38 - NYAE – 41) were replaced with the corresponding amino acids from macroH2A (38 – HPKY – 41) was shown to confer all *in vitro* characteristics of macro-H2A containing octamer and NCP onto major-type H2A (S.C. et al., unpublished data).

Despite the pronounced effect on histone octamer stability, preliminary data show that the overall stability of the nucleosome in response to increased ionic strength is not affected (Fig. 4.3). However, given our results with H2A.Z-containing nucleosomes (Park et al., 2004b), it is possible and even likely that more subtle changes exist. It should be pointed out that since the first and foremost task of histone variants is to form nucleosomes, subtle changes are expected at most, but clearly these differences are fundamental to the function of histone variants.

MacroH2A- containing chromatin has been shown to be inhibitory to transcription *in vivo* (Perche et al., 2000a). Furthermore, macroH2A containing mono-nucleosomes are not remodeled by SWI/SNF, and are unable to bind transcription factors (Angelov et al., 2003b). More precisely, the non-histone domain or the flexible linker (or both) were

responsible for the inability of these variant nucleosomes to bind the transcription factor NF-kappaB. In contrast, the histone-like domain of macroH2A was shown to be responsible for the inability of macro-H2A-containing nucleosomes to be remodeled (Angelov et al., 2003b). It will be of interest to see whether the L1 loop alone can confer this inhibition of chromatin remodeling and transcription to nucleosomes. The key location of the L1 loop within the NCP structure (where it seemingly holds together the two gyres of the DNA double helix) certainly makes this region a prime candidate to affect the dynamic motions of the NCP that are associated with chromatin remodeling.

MacroH2A displays a very unique nuclear localization pattern. While it is expressed in equal quantities in both males and females (Rasmussen et al., 1999) in almost all the higher eukaryotes, it is enriched in the inactive X-chromosome (Xi) of adult female mammals manifested as a Macro Chromatin Body (MCB) (Costanzi and Pehrson, 1998b; Hoyer-Fender et al., 2000). Various GFP-tagged constructs have been used to determine the importance of the histone and the non-histone domain in the localization of macroH2A to the inactive X-chromosome (Chadwick et al., 2001). The histone domain alone (without N- and C-terminal tails) fused to GFP was found to form the Xi associated MCB with efficiencies comparable to full-length macroH2A. Of the 19 amino acids that are different in macroH2A compared to major H2A no single amino acid could be attributed this function by point mutation studies. On the basis of preliminary analysis of the crystal structure of macroH2A mono-nucleosomes and biochemical studies of L1 mutants of major H2A it seemed likely that the L1 loop is crucial for the localization on the Xi. This hypothesis has been tested experimentally (S. C. et al., unpublished data).

MacroH2A is gradually emerging as a "family" of variants. At least two different genes are known to encode this protein with significant variations in sequence but the same basic structural organization (H2A1 and macroH2A2, respectively). MacroH2A1 has two splice-variants (macroH2A1.1 and macroH2A1.2), which are non-identical in a very small region of the non-histone region starting at amino acid 195 (Pehrson et al., 1997). MacroH2A2 is overall 68% identical to macroH2A1.2. The histone region is 84% identical to that of macroH2A1 and only 66% identical to major H2A. The sequence of the macroH2A2 L1-loop (TFKY) seems to be a convolution between that of major H2A (NYAE), and macroH2A (HPKY). The basic region is the most varied and is only 25% identical to that of macroH2A1, whereas the non-histone region is 64% identical to that of macroH2A1.2. MacroH2A1.2 and macroH2A2 display very similar (and sometimes overlapping) nuclear localization patterns, at least at a global level, and the functional relevance of the sequence differences remains unclear (Chadwick et al., 2001; Costanzi and Pehrson, 2001).

4.3.4 H2A-Bbd

The structured region of H2A.Bbd is only 48% identical to that of major H2A, making H2A.Bbd the most divergent histone H2A variant known to date. So far, this histone variant has only been identified in humans and mice. Major hallmarks of the amino acid sequence of H2A.Bbd as compared to that of major H2A are (1) the presence of a continuous stretch of five arginines and the conspicuous absence of lysines in its N terminal tail, (2) the absence of a C terminal tail and the very last segment of the docking domain; (3) major sequence differences in the docking domain of H2A; (4) the presence of only one lysine in H2A.Bbd compared to fourteen in major H2A, resulting in a slightly

less basic protein (pI 10.7, compared to a pI of 11.2 for major H2A); and (5) the absence of the 'acidic patch' (Luger and Richmond, 1998b) on the docking domain (Fig. 4.1b).

We found that mono-nucleosomes containing H2A.Bbd had a more relaxed structure with less tightly bound DNA ends (Bao et al., 2004b). Only 118±2 bp of DNA are protected against digestion with micrococcal nuclease, in contrast to 146 bp in canonical nucleosomes. These results are consistent with the observed more rapid exchange of GFP-H2A.Bbd *in vivo* (Gautier et al., 2004). Intriguingly, we also found a lower repeat length in micrococcal nuclease digestion of nucleosomes reconstituted onto plasmids using a recombinant *in vitro* assembly system (Georges et al., 2002) (136 bp as opposed to ~160 bp for major nucleosome arrays), which suggests that the H2A.Bbd nucleosomes are deposited at a higher density. At this high density, H2A.Bbd represses transcription comparable to H2A. Intriguingly, domain swap experiments (in which the H2A docking domain was exchanged with that of Bbd) show that the H2A.Bbd docking domain is largely responsible for its behavior. The conservation of the histone-fold together with the nuclear localization pattern suggests that H2A-Bbd alters chromatin structure at the nucleosomal level, giving rise to transcriptionally active domains (Chadwick and Willard, 2001b).

4.4. Evolutionary Targets in the Histone Fold of H2A Variants

Sequence comparisons (Fig. 4.1B) together with analysis of the two available crystal structures of nucleosomes containing histone variants (Suto et al., 2000); Chakravarthy et al., unpublished data) and analysis of biochemical and biophysical data from ours and several other laboratories show that histone variants are true replacement histones in that they can form functional nucleosomes and chromatin. The majority of structural and

functional changes in histone H2A variants reside in the docking domain and in the L1 loops, with the latter being more structurally divergent than the former.

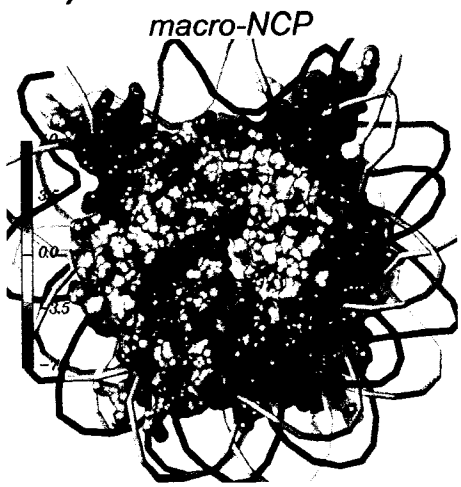
4.4.1. The Multi-Functional H2A Docking Domain

The docking domain is involved in interactions between the H2A-H2B dimer and the $(H3\text{-}H4)_2$ tetramer, and harbors three of the seven residues that form the 'acidic patch' on the surface of the nucleosome (Luger and Richmond, 1998b). Thus, sequence divergence in this region may affect the stability of the H2A variant-H2B dimer / $(H3\text{-}H4)_2$ tetramer interface, as has been observed for H2A.Z (Park et al., 2004b), which will have effects on chromatin remodeling and transcription. The inefficient organization of the penultimate ~ 15 – 20 base pairs of nucleosomal DNA can also be a consequence of relatively minor sequence changes in this domain, as described for H2A.Bbd (Bao et al., 2004b). Changes in amino acid sequence may also alter the surface of the nucleosome, with important implications for the ability of nucleosomes to interact with other factors or to form more compact higher order structures. It is interesting to note that while the acidic patch is decreased in H2A.Bbd (not shown), its size is actually increased in H2A.Z (Suto et al., 2000), and expanded towards the C-terminal end of the docking domain in macroH2A (Fig. 4.4).

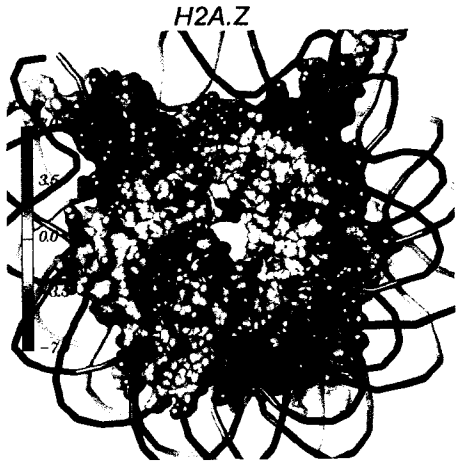
4.4.2. The H2A L1 loop may select for the second H2A-H2B dimer

The L1 loops of the two H2A moieties within the nucleosome are involved in the formation of the L1L1-interface, which is the only site of interaction between the two H2A-H2B dimers in the nucleosome. The L1L1-interface is responsible for the cooperative incorporation of the two H2A-H2B dimers. It may also stabilize the two gyres of the nucleosome core particle (Fig. 4.2D). Altering the biochemical nature of this

A)



B)



C)

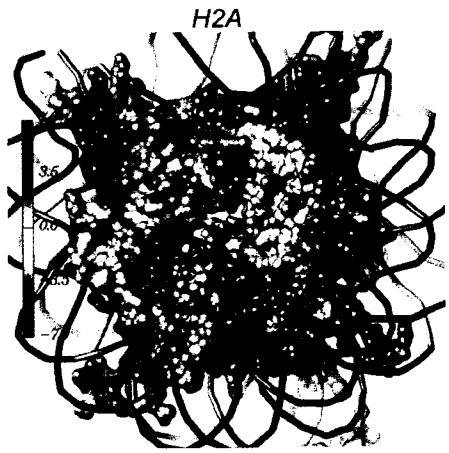


Figure 4.4: Subtle changes in the molecular surface of macro-NCP compared to H2A.Z-NCP and major-NCP. A. macro-NCP; **B.**, H2A.Z-NCP; **C.**, major-NCP. The electrostatic potential ranges from +7.0 (blue) to -7.0 (red) kcal/mol/e (color bar), shown is the value at the solvent-accessible surface (out 1.4 Å, the radius of a water molecule, from the molecular surface) mapped back on to the molecular surface. Molecular surfaces were calculated with the program MSMS (Sanner et al., 1996) using a 1.4 Å probe sphere. The acidic patch that provides an essential crystal contact with the N-terminal tail of histone H4 of a neighboring nucleosome core particle is indicated by a black arrow. Panels b and c are taken from (Suto et al., 2000), with permission.

interface should therefore result in an altered response to the transcriptional machinery and to chromatin remodeling factors, as well as determine the histone composition of the nucleosome.

There is no experimental evidence for the tacit assumption that one nucleosome contains, for example, two H2A.Z-H2B dimers. It is theoretically possible to have a nucleosome in which only one of the H2A moieties has been replaced by its corresponding variant, resulting in a nucleosome with one $(H3\text{-}H4)_2$ tetramer, one major H2A-H2B dimer and one variant H2A-H2B dimer. From a nucleosomal viewpoint, the primary determinant of the composition of a given nucleosome must be the compatibility of the L1-loop of major H2A with those of different H2A variants. To investigate this possibility *in vitro,* we performed salt-gradient reconstitutions with mixtures of $(H3\text{-}H4)_2$ tetramer, (H2A-H2B) dimers, and H2A.Z-H2B dimers (Fig. 4.5); or with H2A.Bbd-H2B dimers (Fig. 4.6). We used either nickel-affinity chromatography to isolate nucleosomes containing his-tagged H2A (Fig. 4.5), or gel elution of 'mixed nucleosome bands' (Fig. 4.6), followed by analysis of the histone content by SDS PAGE. Using these two approaches, we could show that while all histone H2A variants are capable of forming hybrid nucleosomes (Fig. 4.5 and 4.6), the propensity to do so clearly differs between macroH2A, H2A.Z, and H2A.Bbd. Intriguingly, hybrid nucleosomes reconstituted from a mixture of H2A-H2B dimers and H2A.Bbd-H2B dimers together with $(H3\text{-}H4)_2$ tetramer protect only ~130 bp of DNA against digestion with micrococcal nuclease, that is ~ 20 base pairs more than nucleosomes containing two H2A.Bbd chains, and ~ 15 base pairs less than canonical nucleosome (not shown). The ability of H2A.Bbd to form hybrid nucleosomes is

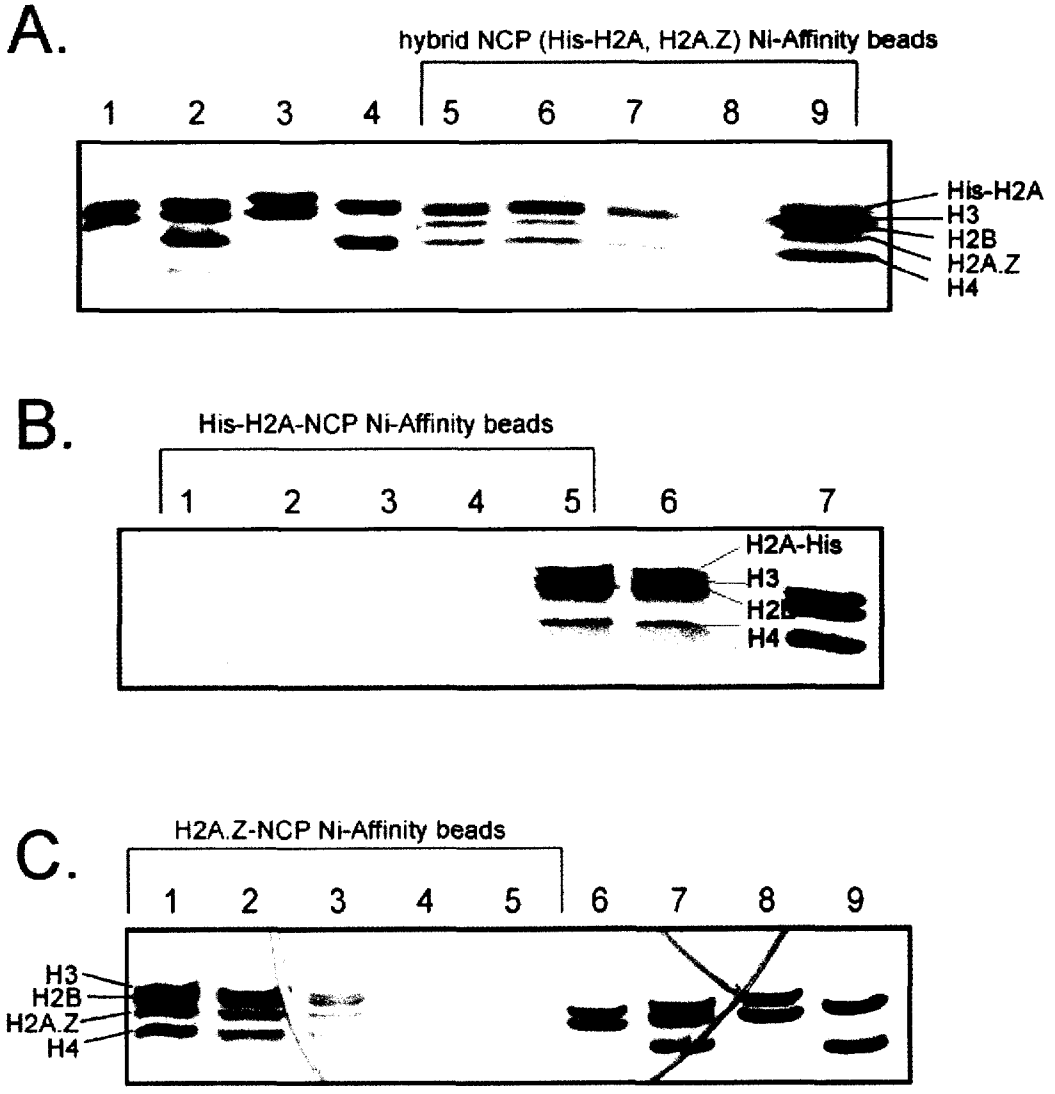
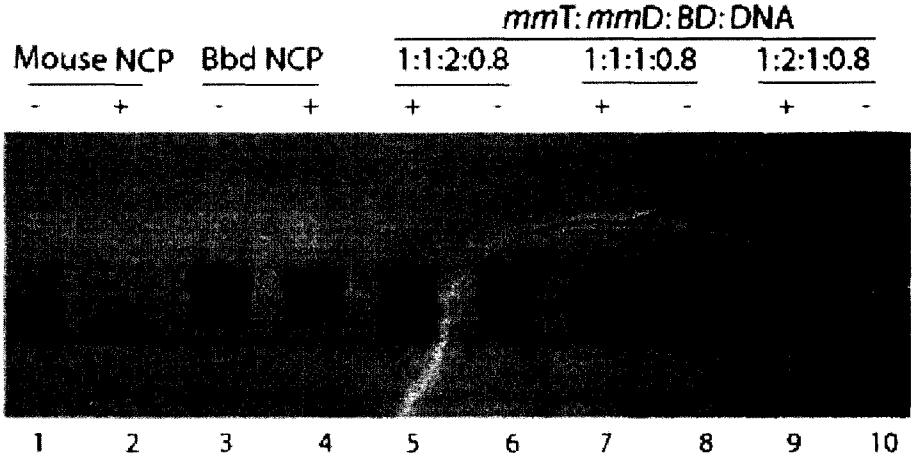
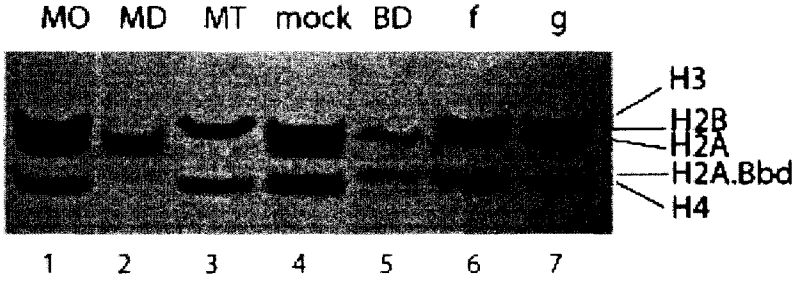


Figure 4.5 H2A.Z can form hybrid nucleosomes with major-type H2A. A. Lane 1: H2A.Z-H2B dimer. Lane 2: Histone octamer with untagged H2A. Lane 3: His-tagged H2A-H2B dimer. Lane 4: H3-H4 tetramer. Lane 5-8: flow-through and washes with 5 mM imidazole of NCP reconstituted from a mixture of H2A.Z, His-tagged H2A, H2B, H3, and H4 after a two-hour incubation with Ni-NTA beads. Lane 9: Elution with 1 M imidazole. **B.** Control with his-tagged H2A NCP. Lanes 1-4: flow through and washes with 5 mM imidazole after a two-hour incubation of His-tagged H2A NCP to Ni-NTA beads. Lane 5: Elution with 1 M imidazole. Lane 6: Onput. Lane 7: Histone octamer with untagged H2A. **C.** Control with untagged H2A.Z nucleosomes. Lanes 1-4: Flow through and washes with 5 mM imidazole after 2 hrs binding of H2A.Z NCPs to Ni-NTA beads. Lane 5: Elution with 1 M imidazole. Lane 6: H2A.Z-H2B dimer. Lane 7: Histone octamer with non His-tagged H2A. Lane 8: His-tagged H2A-H2B dimer (His-tagged H2A – upper band, H2B lower band). Lane 9: H3-H4 tetramer. All gels shown here are 18% SDS-PAGE gels stained with coomassie brilliant blue.

A.



B.



C.

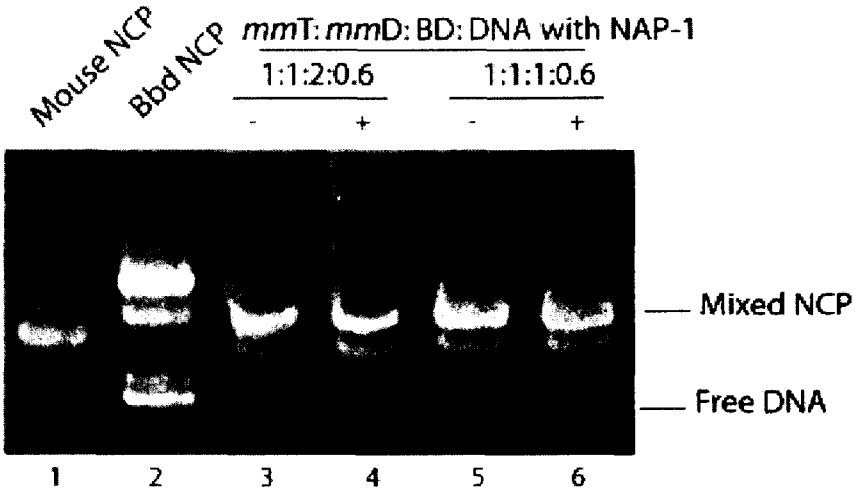


Figure 4.6 H2A.Bbd can form hybrid nucleosomes with major-type H2A. A. Salt gradient reconstituted mouse NCP (*mm*-NCP; lanes 1 and 2), Bbd-NCP (lanes 3 and 4) and hybrid NCPs (lanes 5-10), before (-) and after (+) a 1 hour incubation at 37°C, were analyzed by 5 % native gel and stained by coomassie brilliant blue. Ratios between mouse (H3-H4)2 tetramer (*mm*T), mouse (H2A-H2B) dimer (*mm*D), H2A.Bbd-H2B dimer (BD), and DNA for 'hybrid NCP' reconstitution are indicated. Bands that were subsequently excised from the gel for further analysis are labeled from a-g. **B.** Bands f, g (Fig. 4.6a) were analyzed by 18% SDS-PAGE, stained with coomassie brilliant blue. Mouse H2A and H2A.Bbd are indicated by stars. Mouse octamer (MO, lane 1), MD (lane 2), *mm*T (lane 3), *mm*T:*mm*D:BD mixtures of 1:1:1 (mock, lane 4) and BD (lane 5) are shown as controls. **C.** Hybrid NCPs are also obtained with yNAP-1 – dependent nucleosome assemblies. yNAP-1 – reconstituted NCP (reconstituted at indicated ratios) were compared with NCP reconstituted over a salt gradient on a 5% native gel. The gel was stained with ethidium bromide. Hybrid NCP is indicated by asterisks.

independent of the assembly pathway, since the same end result is obtained by yNAP-1 – dependent assembly under physiological ionic strength (Fig. 4.6C).

Our results clearly indicate the possibility that variants may be combined with major-type histone H2A in a single nucleosome, thus generating yet another level of structural and functional heterogeneity. The specific nature of the different L1L1-interfaces and their potential influence on the accessibility of nucleosomal DNA will become more obvious in light of future structural studies. How would such nucleosomes be assembled *in vivo*? Histone H2A-H2B dimers are in rapid exchange even in the absence of transcription and replication (Louters and Chalkley, 1985; Kimura and Cook, 2001). Recently discovered histone-variant specific assembly factors for H2A.Z (Kobor et al., 2004; Krogan et al., 2003; Mizuguchi et al., 2004) promote the replacement of one or both H2A-H2B dimer with a H2A.Z – H2B dimer. ATP-dependent chromatin remodeling factors are also capable of actively exchanging histone variant dimers into folded nucleosomes (Bruno et al., 2003). Since it is unlikely that both dimers are exchanged at the same time, the final 'equilibrium' makeup of a particular nucleosome will be determined by the L1 loop, and by the relative availability of major-type and variant histones.

4.5. Conclusions and Outlook

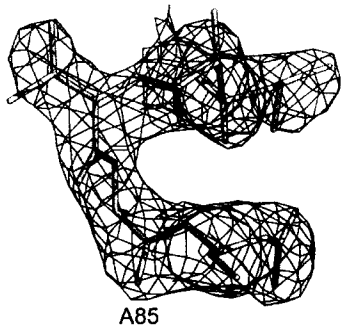
Histone variants are receiving an ever-increasing amount of attention, but much needs to be done before we reach a complete understanding of their role in the complex biology of chromatin. *In vitro,* structural and biophysical analysis of nucleosomes containing H2A.Bbd and CenpA (in addition to the already available structures for H2A.Z and macroH2A containing nucleosomes) will provide further insight into the ways in which nucleosome structure and dynamics is affected by histone variants. These studies will

also have to be expanded to take into account the possible existence of hybrid nucleosomes. Second, we will need to extend our investigations to include model nucleosomal arrays, as demonstrated for H2A.Z (Fan et al., 2002b). Third, the interaction of variant-containing chromatin to interact with various linker histones and non-histone proteins needs to be investigated. Fourth, more systematic analyses are needed to study the dynamic behavior of variant-containing nucleosomes and chromatin in the presence of various chromatin remodeling factors (see, for example, (Flaus et al., 2004). Finally, the transcriptional properties of variant-containing nucleosomes from several model promoters will have to be scrutinized carefully *in vitro*.

In vivo, the challenges are perhaps even greater. For example, knockout models are available for only few histone variants. Second, we need to investigate the distribution of histone variants at the single-nucleosome (and sub-nucleosome) level at a variety of chromatin loci. This represents a technical challenge that may not be easily overcome. Third, the pathways by which variant nucleosomes (or hybrid nucleosomes) are assembled needs to be further investigated. Finally, the ability of histone variants to be post-translationally modified must be scrutinized. Combined, these studies are likely to yield a picture of mind-boggling complexity that probably will approach or even exceed that of the histone modification network. Doubtlessly, histone variants add an entirely new dimension to the 'histone code'.

CHAPTER 5

Crystallographic Studies of MacroH2A Nucleosome Core Particles with Mixed Stoichiometry



In this chapter I present my work done in the attempt to test the hypothesis that macroH2A (and other H2A variants) are capable of forming mixed nucleosomes, i.e., nucleosomes in which only one of the major type H2A moieties is replaced by the histone domain of the variant. We have concluded from in vitro biochemical studies that macroH2A prefers to form mixed nucleosomes when present in equal molar quantities with major H2A. The crystallographic studies with mixed nucleosomes are underway and we hope to gain valuable structural insights from them.

5.1 Abstract

Histone variants are emerging as an important means to alter the biochemical composition of nucleosomes, thus establishing structurally and functionally unique chromatin domains. In all biochemical and structural studies done till date on variant nucleosomes it has been assumed that both the H2A moieties in the nucleosome are replaced by variant H2A. In the absence of *in vivo* evidence however it is logical to assume that in some cases only one of the H2A moieties may be replaced by variant H2A thus making hybrid nucleosomes. We have proved using biochemical experiments that hybrid nucleosomes with one of the major H2As replaced by macroH2A are not only feasible physical entities but *in vitro* they are in fact preferred over the homotypic nucleosomes. We have purified hybrid macro-NCPs and obtained well diffracting crystals of these hybrid-NCPs. The existence of hybrid nucleosomes in vitro opens up another avenue of structural and functional heterogeneity among nucleosomes *in vivo*.

5.2 Introduction

Histones are proteins that mediate the folding of eukaryotic DNA into a nucleoprotein complex known as chromatin. The fundamental unit of chromatin is the nucleosome core particle, which consists of a histone octamer made of two copies of the four core histones H2A, H2B, H3, and H4. Histone variants are non-allelic sequence variations of conventional histones. Most histone variants substitute for their major type counterparts in the nucleosomes of chromatin that is structurally and functionally unique. The two histones that have the most number of variants are H2A and H3. CenH3 (Centromeric histone H3) and H3.3 are the most common H3 variants. H2A.X, H2A.Z, H2A.Bbd, and

macroH2A are the known H2A variants. In order to understand the mode of function of these variants we have embarked on studying their structure in a nucleosomal context. The crystal structures of a nucleosome core particle in which both H2A moieties are replaced by H2A.Z or macroH2A have been determined (Suto et al., 2000) (also see Chapter 2). We find that the structure remains globally comparable to that of the non-variant nucleosome, but sequence differences in strategic regions of the histone fold do result in subtle structural alterations that may have potentially major functional implications. Among H2A variants the two regions that are consistently affected are the docking domain and the L1-loop (Chakravarthy et al., 2004). The docking domain is involved in interactions with the $(H3\text{-}H4)_2$ tetramer while the L1-loops of the two H2A moieties form the small yet vital interface between the two H2A-H2B dimers in the nucleosome. In all the structural and biochemical studies till now we have made the tacit assumption that both the H2A moieties are replaced by the variant H2A (homotypic nucleosomes). Recent studies have shown replication-independent deposition of histone variants that is contingent upon the local chromatin structure being 'open' (Ahmad and Henikoff, 2002a). Others have shown histone variant specific chaperone complexes that mediate the exchange of histone variants with major type histones (Mizuguchi et al., 2004). None of these studies show a predilection for either homotypic or hybrid nucleosomes (only one of the major H2A replaced by variant H2A), but it is tempting to assume that exchanging one of the major H2A-H2B dimers with a variant H2A-H2B dimer is easier than replacing both without compromising the structural integrity of the nucleosome.

The prevalence of either kind of nucleosome (homotypic or hybrid) *in vivo* may be regulated by the relative affinities between the L1-loops of major and variant H2As. We performed *in vitro* biochemical experiments that suggest that all H2A variants albeit to varying extents, exhibit the propensity to form hybrid nucleosomes. It is therefore reasonable to assume that depending on the relative local concentrations of histone H2A and its variants both homotypic and hybrid nucleosomes are feasible physical entities. To test this hypothesis we reconstituted different variant H2A nucleosomes in the presence of equal molar quantities of major type H2A. We find that when macroH2A and major type H2A are present in an equimolar ratio the nucleosomes that are reconstituted are predominantly hybrid. We purified and crystallized this nucleosome and obtained a 2.8 Å synchrotron data set. Molecular replacement using either the *Xle*-NCP or the macro-NCP as a search model yielded solutions with reasonable statistics but there seems to be a convolution between the two possible orientations of the nucleosome with respect to the two different H2A moieties and it is therefore difficult to distinguish the density for macroH2A from that for major type H2A. In order to determine the structure of the L1-L1 interface between macroH2A and major type H2A it is necessary to de-convolute the two orientations of the hybrid nucleosomes.

5.3 Materials and Methods

5.3.1 Expression and purification of histone proteins, and reconstitution of nucleosomes

The coding region for the histone domain of human macroH2A (amino acids 1-120) was sub-cloned into a pet3a vector. Expression plasmids for H2B and H2A from mouse (*Mus musculus*) were a kind gift from Dr. David Tremethick. The coding region for mouse

H2A was sub cloned into a pet15b vector, which has a his6-tag at the N-terminus of the protein. We also used the expression plasmids (pet3a) for histones H2B, H3 and H4 from *X.laevis*. All histones were over-expressed in BL21 (DE3)-plysS (Stratagene) and purified using previously published protocols (Luger et al., 1999b). The histone domain of macroH2A (macroH2A$_{HD}$), with Xenopus H2B, was refolded into a dimer as was His-tagged mouse H2A with mouse H2B and H3 with H4 was refolded into a heterotetramer using previously published protocols (Dyer et al., 2004). MacroH2A-H2B dimer, His-tagged H2A-H2B dimer, $(H3\text{-}H4)_2$ tetramer and a 146 bp palindromic fragment of DNA derived from human α-satellite regions (Luger et al., 1997a) were used in a 1:1:1:1 molar ratio for nucleosome reconstitution by salt gradient dialysis (Dyer et al., 2004), resulting in hybrid - NCPs. NCPs reconstituted with only macroH2A-H2B dimers or only His-tagged H2A-H2B dimers (tetramer: dimer: DNA = 1:2:1) were used as control. The nucleosomes thus obtained were analyzed on a 5% non-denaturing PAGE gel. Milligram amounts of all these NCPs were heat-shifted and purified by preparative gel electrophoresis using published protocols (Dyer et al., 2004).

5.3.2 Analysis of Nucleosome Contents Using Batch Ni-Affinity Chromatography

We used nickel-affinity chromatography to isolate nucleosomes containing his-tagged H2A. 100 µl of Ni-NTA beads (Sigma) were thoroughly washed in water and then in binding buffer (150 mM NaCl, 20mM Tris HCl pH 7.5, 5mM Imidazole). Nucleosomes (Hybrid-NCP or pure MacroH2A NCP or pure His-tagged H2A NCP) were loaded on to the beads and rocked for 2 hrs at 4°C. The beads were spun down and the flow through was collected followed by 3 washes with the binding buffer. The beads were then rocked for 15 mins in elution buffer (150mM NaCl, 20mM Tris HCl pH 7.5,

and 1M Imidazole). The flow through, the supernatant from three washes and the elute along with the on put were analyzed for their histone content on an 18% SDS –PAGE gel.

5.3.3 Crystallographic procedures

Hybrid-NCP was crystallized using salting in vapor diffusion at NCP concentrations ranging from 8-12 mg/ml with salt concentrations of 34 to 37.5mM KCl and 40-45mM $MnCl_2$. The homogeneity of the sample was checked on a 5% native PAGE gel and the contents of the hybrid nucleosomes were checked on an 18% SDS-PAGE gel. The crystals were soaked in 24% 2-methyl-2,4-pentanediol and 5% trehalose and frozen in liquid nitrogen as described previously (Luger et al., 1997a). Data were collected at Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline 5.0.2. Data from a single crystal were processed using Denzo and Scalepack (Otwinowski and Minor, 1997). Molecular replacement was performed using Protein Data Bank entry 1AOI as the search model. Molecular replacement and subsequent rounds of refinement were performed using CNS (Rice et al., 1998). The program O was used for model building (Jones et al., 1991).

5.4 Results

5.4.1 MacroH2A forms predominantly hybrid nucleosomes with equi-molar ratios of major H2A

Structural studies with H2A.Z suggest that the L1-loop of H2A.Z might be sterically hindered by that of major type H2A thus making a hybrid nucleosome (nucleosome with one H2A.Z and one major H2A) an unlikely physical entity (Suto et al., 2000). The L1

loops of the two H2A moieties are involved in formation of an L1L1 interface, which is the only site of interaction between the H2A-H2B dimers in the nucleosome (Luger et al., 1997a). It is also responsible for the cooperative incorporation of the H2A-H2B dimers during nucleosome assembly. Almost all H2A variants show sequence differences at the L1-loop (Chakravarthy et al., 2004). The propensity to form hybrid nucleosomes could therefore depend on the relative affinities between the L1-loops of different H2A variants and major H2A. In order to determine the readiness with which different H2A variants form hybrid nucleosomes *in vitro* we reconstituted nucleosomes by salt gradient dialysis (Dyer et al., 2004) using mixtures containing $(H3\text{-}H4)_2$ tetramer, His-tagged major H2A-H2B dimer, 146 bp α-satellite DNA and variant H2A-H2B (H2A.Z, H2A.Bbd, or macroH2A) dimer in a molar ratio of 1:1:1:1 (Fig 5.1A lanes 5 & 6). We then performed batch Ni-affinity chromatography to isolate nucleosomes that have the His-tagged H2A-H2B dimers (either homotypic nucleosomes containing only His-tagged H2A or hybrid nucleosomes that have one His-tagged H2A and one variant H2A) (Fig 5.1B). We see that in case of H2A.Z (contrary to the implications from the structural studies) and H2A.Bbd there is a mixture containing varying amounts of hybrid and homotypic nucleosomes (see chapter 4 figs. 4.5 and 4.6) but macroH2A forms only hybrid nucleosomes. For comparison, we performed this experiment with homotypic nucleosomes containing His-tagged major H2A (Fig 5.1A lanes 1 & 2) or homotypic nucleosomes containing macroH2A (Fig 5.1A lanes 3 & 4) (tetramer: dimer: DNA = 1:2:1 in both cases). We see that the His-H2A-NCPs bind with a high affinity to the Ni-NTA beads (Fig 5.1C) and the homotypic macroH2A-NCPs do not bind (Fig 5.1D). Nucleosomes show variable electrophoretic mobility on a 5% native

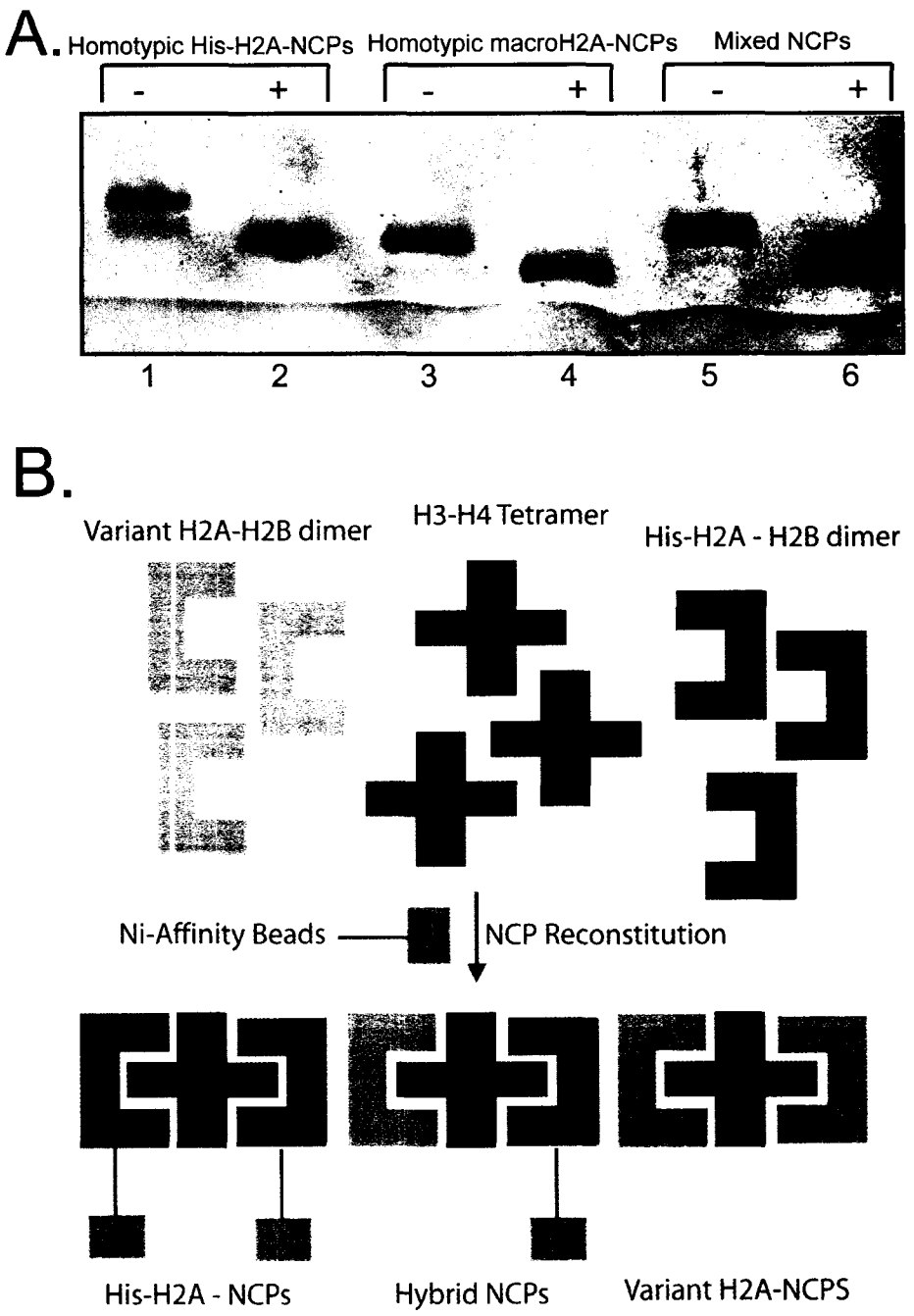


Figure 5.1A. MacroH2A can form hybrid nucleosomes with major-type H2A. Lanes 1 & 2. NCPs reconstituted by mixing (H3-H4)2 tetramer, H-s-H2A-H2B dimer, and DNA at a ratio of 1:2:1. Lanes 3 & 4. NCPs reconstituted by mixing $(H3\text{-}H4)_2$ tetramer, macroH2A-H2B dimer, and DNA at a ratio of 1:2:1. Lanes 5 & 6. NCPs reconstituted by mixing $(H3\text{-}H4)_2$ tetramer, macroH2A-H2B dimer, His-H2A-H2B dimer and DNA at a ratio of 1:1:1:1. – and + suggest before and after incubation at 37°C for 1hr respectively. 5% native PAGE gel stained with coomassie brilliant blue. **B.** Schematic depicting the Ni-affinity chromatography experiment. Possible NCPs and their predicted binding affinities are indicated. See text for experimental details.

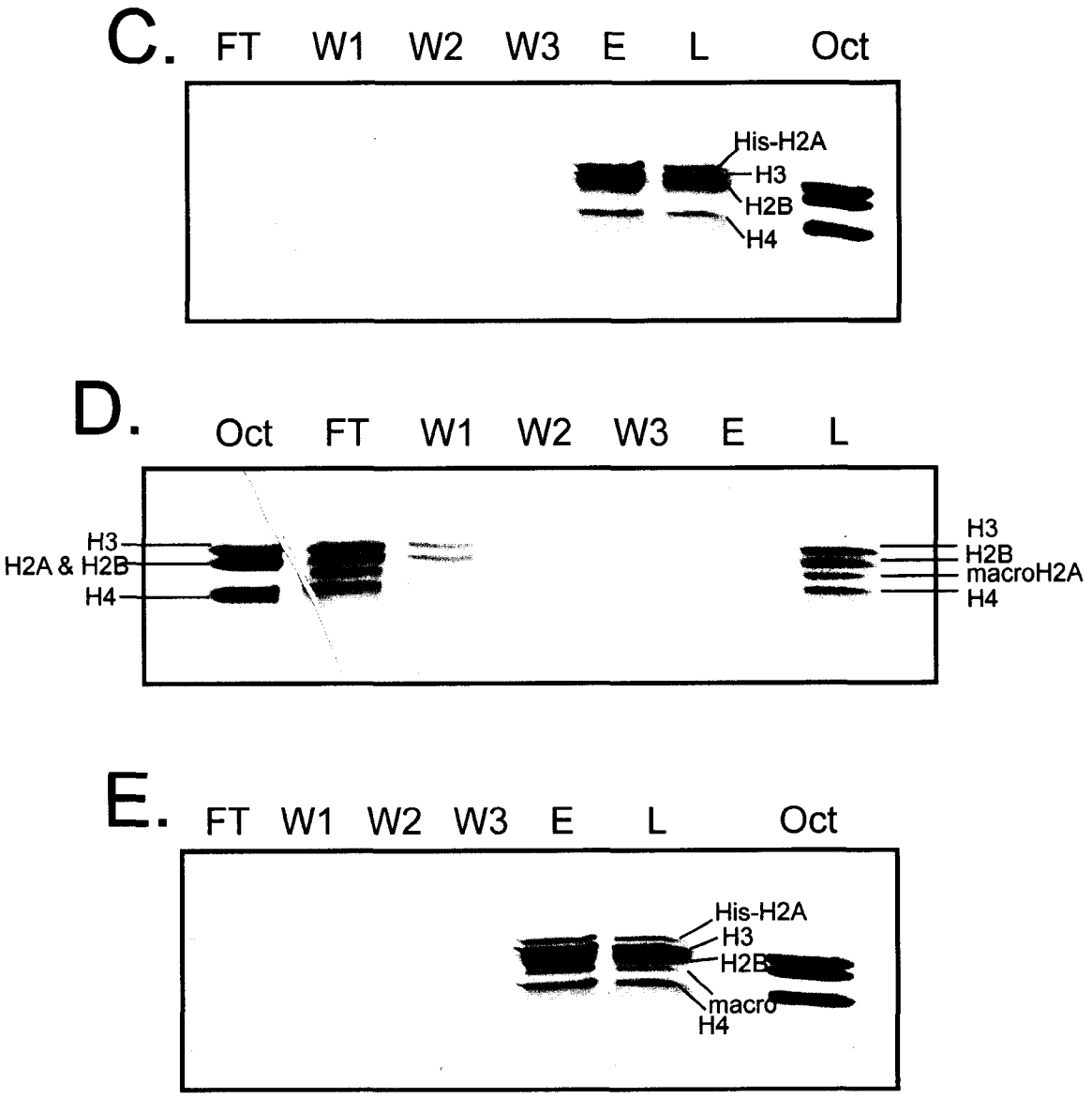


Figure 5.1. MacroH2A can form hybrid nucleosomes with major-type H2A. C. Control with his-tagged major H2A-NCP (lane 2 in A). **D.** Control with macroH2A-NCP (no his-tag) (lane 4 in A). **E.** Batch Ni-affinity chromatography with hybrid-NCPs (lane 6 in A). *FT*: flow through; *W1, W2, and W3*: washes; *E*: Elute; *L*: Load (Sample loaded on the beads); *Oct*: Octamer made of core histones from *X.laevis*. All the gels shown here are 18% SDS-PAGE gels stained with coomassie brilliant blue.

PAGE gel depending on the number of His-tagged H2A moieties they contain. Homotypic nucleosomes containing only variant H2As (no his-tag) migrate the fastest, and homotypic nucleosomes with only His- H2As migrate the slowest (compare lanes 2 and 4 in Fig 5.1A). Hybrid nucleosomes with one variant H2A and one His-tagged H2A show intermediate rates of migration (Fig 5.1A lane 6). When macroH2A-H2B dimers are present in an equi-molar ratio with His-tagged major H2A-H2B dimers in the reconstitution mixture we get predominantly nucleosomes that show intermediate electrophoretic mobility and bind to the Ni-NTA beads (Fig 5.1E). On examining the histone content of the nucleosomes bound to the Ni-NTA beads on an 18% SDS-PAGE gel, we find bands of equal intensity on the gel representing both macroH2A and His-tagged major H2A (Fig. 5.1E, elute). We therefore conclude that *in vitro,* macroH2A prefers to form hybrid nucleosomes.

5.4.2 Purification and crystallization of hybrid macro-NCP

Milligram amounts of hybrid nucleosomes were reconstituted using salt gradient dialysis and purified by preparative gel electrophoresis (Dyer et al., 2004). To make sure that the crystallization sample is homogeneous, we ran it on a 5% native gel, which showed only one band (Fig 5.2A). The histone content of this single band was analyzed on an 18% SDS-PAGE gel, which showed bands of equal intensity representing macroH2A and His-tagged major H2A (Fig 5.2B lane 2). We used a salting in vapor diffusion method to obtain diffracting crystals under conditions similar to those used for major type *Xle*-NCPs (Luger et al., 1997a) (Fig 5.2C). We collected a data set at the synchrotron (ALS-Berkeley – BL5.0.2) from a single crystal to a highest resolution of 2.9Å (Fig 5.3A). Hybrid-NCPs crystallize in the orthorhombic space group $P2_12_12_1$ and have unit cell

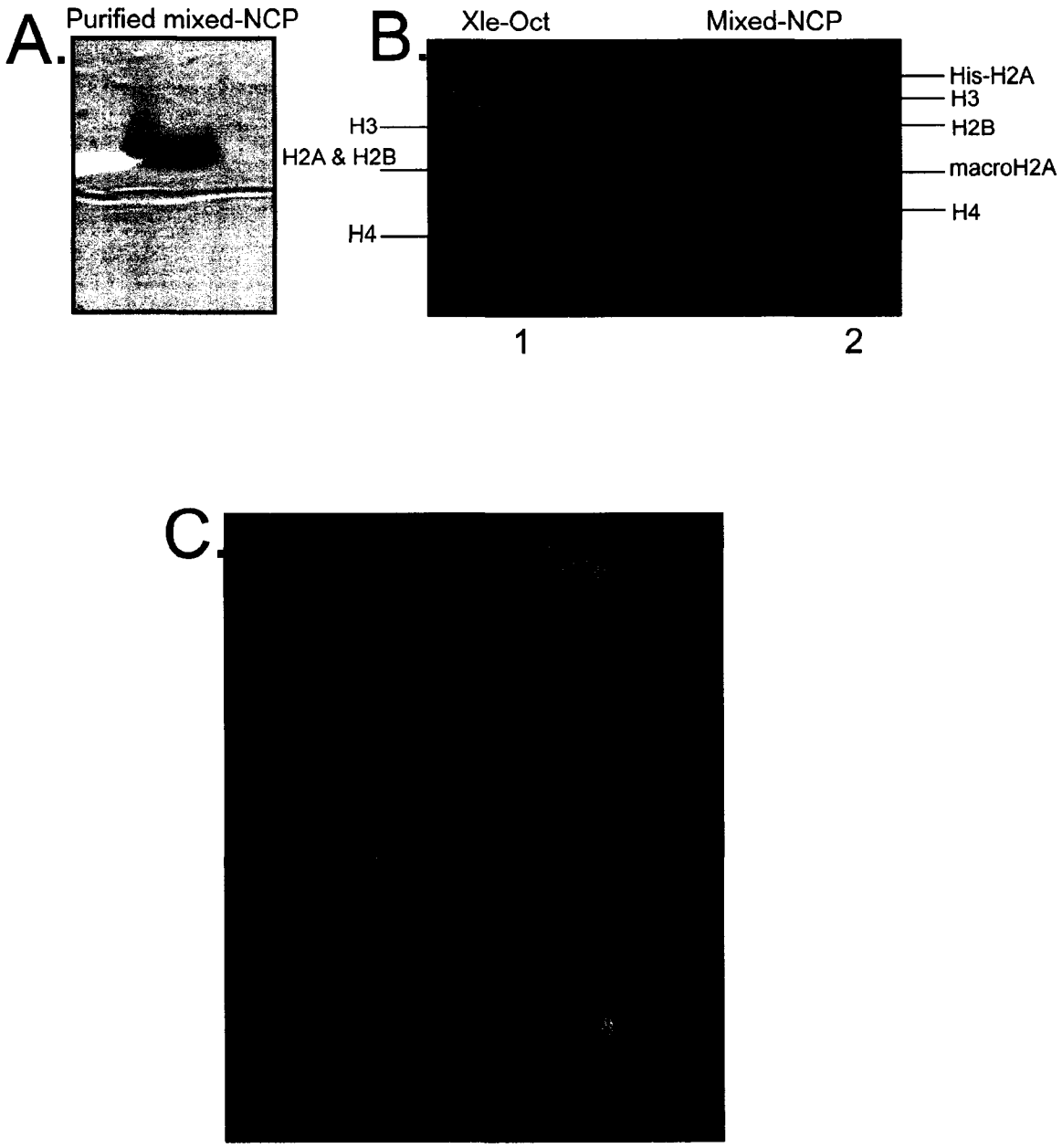


Figure 5.2 Purification and crystallization of hybrid-NCPs. A. Hybrid-NCPs purified by preparative gel-electrophoresis. 5% native PAGE gel stained with coomassie brilliant blue. **B.** Lane 1. Octamer made of histones from X.laevis. Lane 2. Purified hybrid-NCPs seen in A. 18% SDS-PAGE gel stained with coomassie brilliant blue. **C.** Hybrid-NCP crystals from sample seen in A. ~ 900 X 400 X 400 μm^3 in size.

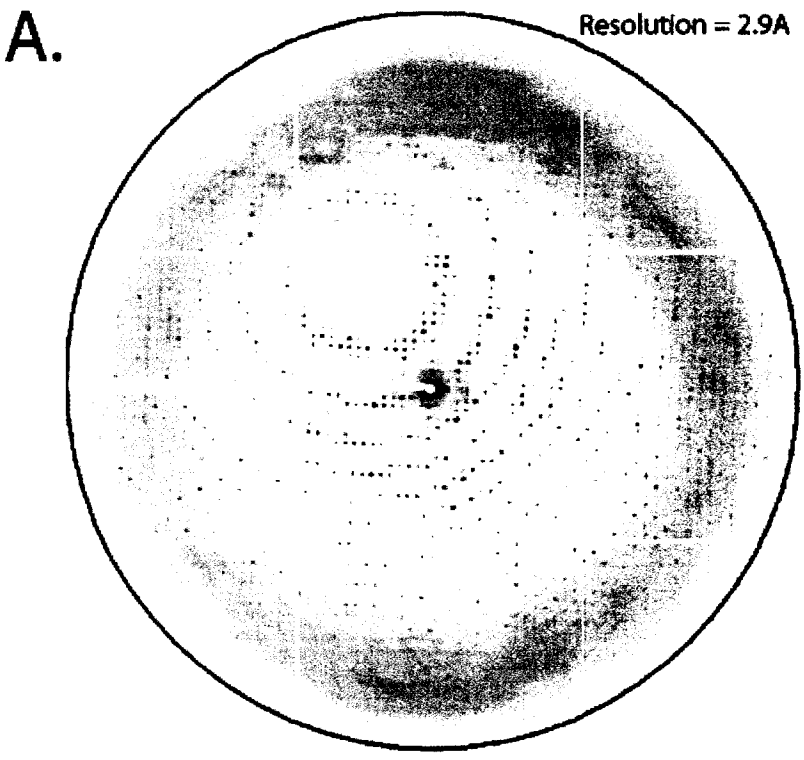


Figure 5.3 A. Diffraction pattern obtained from one of the crystals shown in 5.2C. Part of a data set collected at the synchrotron (ALS Berkeley, BL5.0.2).

dimensions that resemble those of macro-NCPs. The cell edge c both in macro-NCPs and macro hybrid NCPs decrease in length from ~182 Å to ~176 Å (table 5.1) (see also chapter 2, table 2.1). Molecular replacement with the *Xle*-NCP (pdb entry 1AOI) as a search model yielded a solution with a respectable R-factor (table 5.1), but it was difficult to distinguish the density for major-type H2A from that of macroH2A because of a convolution between the two possible orientations of the nucleosome with respect to the H2A moieties (Fig. 5.3B). Molecular replacement with the macro-NCP structure (pdb entry 1U35) as a search model in fact yielded a solution with a better R-factor (table 5.1). Representative regions of the electron density maps contoured at 1 – 2 σ from both solution are shown in Figure 5.4. The initial electron density maps from both cases fit the H2A moieties in the respective model (Fig 5.4 C – H). The electron density maps at the L1-loop region are particularly uninformative and thus make model building difficult (Fig. 5.4 C & E). We tried density modification and solvent flattening after molecular replacement to minimize model bias but the electron density remained as indistinct as before. In order to elucidate the nature of interaction between the L1-loops of major type H2A and macroH2A, we will try treating the two possible conformations as alternate conformations and we will refine the occupancies. This approach may elucidate the specific nature of the interaction between the L1-loops of major H2A and macroH2A.

5.5 Conclusions and outlook

We find that macroH2A prefers to form hybrid nucleosomes with equimolar quantities of major type H2A under conditions used for *in vitro* nucleosome assembly. However the variables that influence nucleosome assembly *in vivo* are several-fold more in number

Table 5.1: Crystallographic Statistics for Hybrid - NCPs

Space Group	$P2_12_12_1$
Unit Cell Dimensions	a= 106.1, b= 109.3, c= 176.3 α= 90, β= 90, γ= 90
Resolution (Å)	50 – 2.9
Unique Reflections	44,768
Completeness (%) (overall / last shell)	97.0 / 96.3
R_{merge} (%) (overall / last shell)	7 / 41
R-factor / R_{free}	
Mol. Rep. Model: 1AOI (*Xle*-NCP)	0.273 / 0.316
Mol Rep. Model: 1U35 (macro-NCP)	0.243 / 0.286

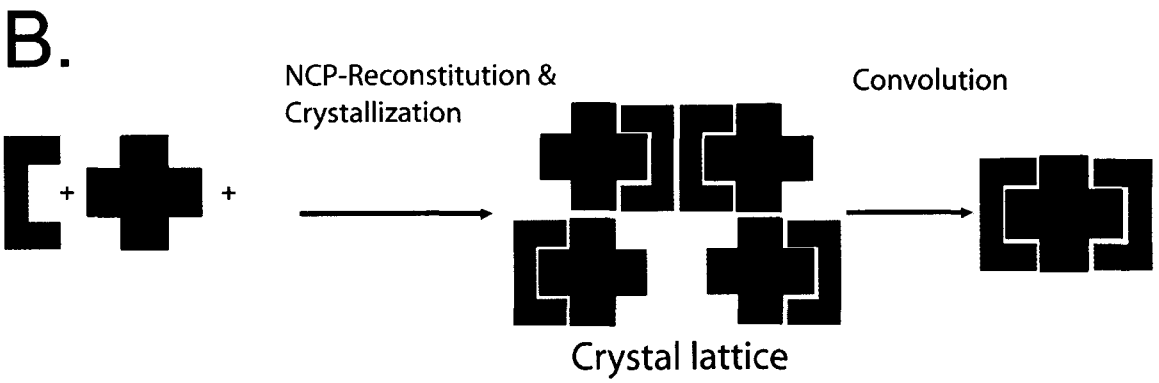
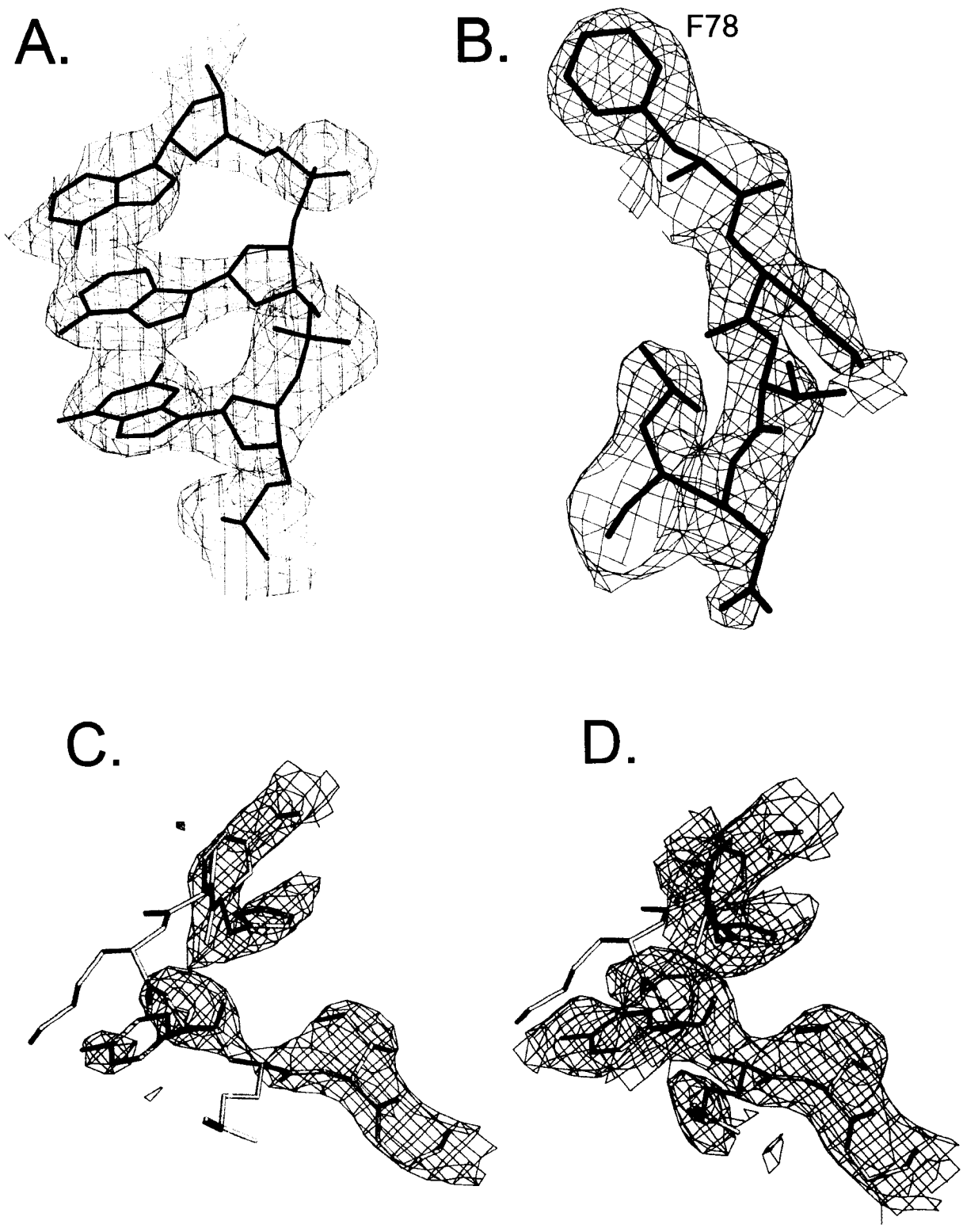


Figure 5.3 B. Two possible orientations of the hybrid nucleosomes with respect to the H2A moieties in the crystal lattice lead to a convolution between the two. *Red*: MacroH2A-H2B dimer; *yellow*: His-H2A-H2B dimer; *orange*: convolution; *blue*: H3-H4 tetramer.



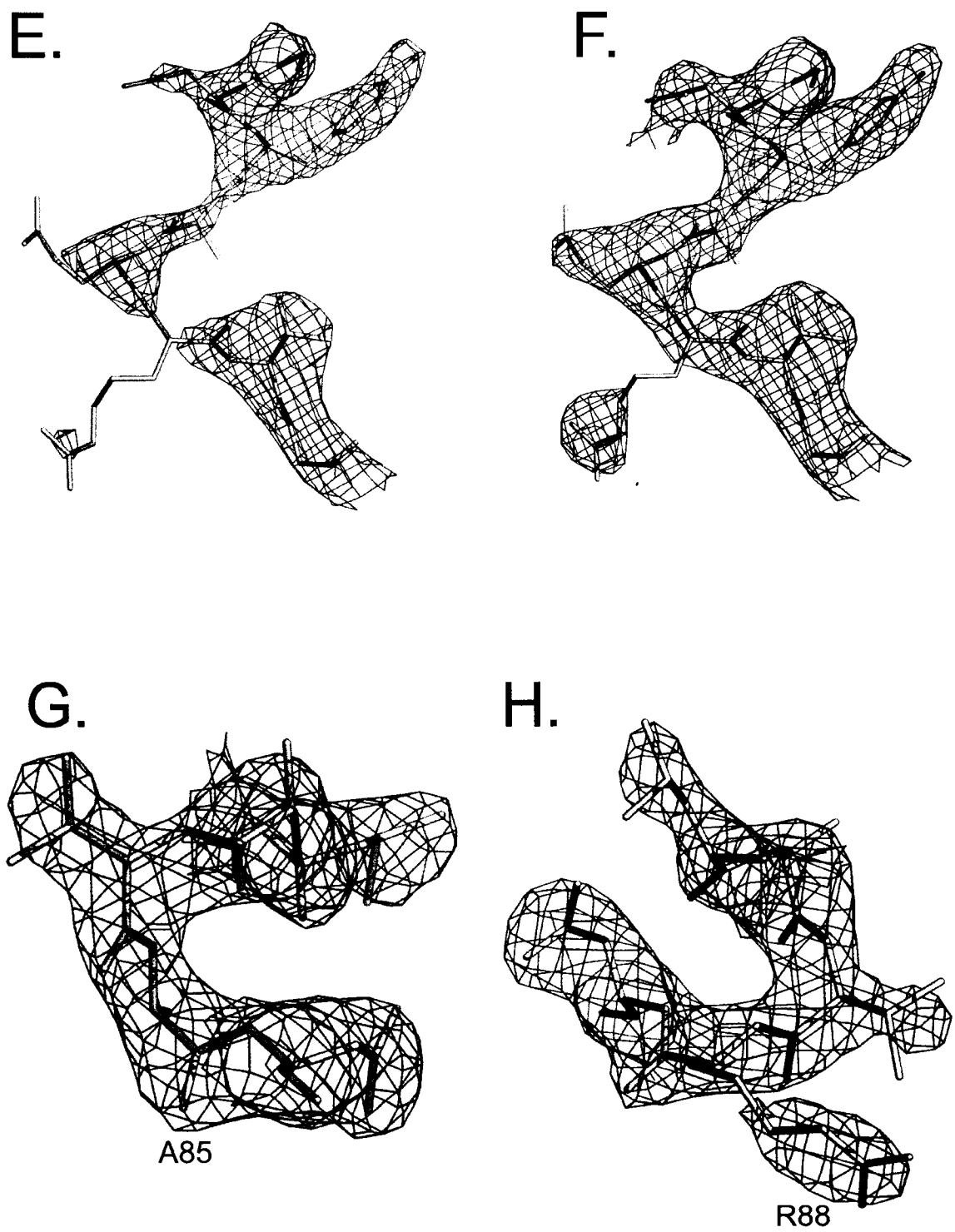
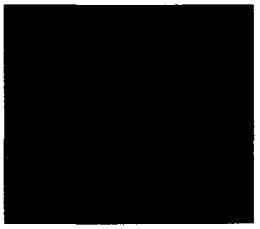


Figure 5.4 Representatives of the electron density form different regions of the hybrid-NCP. A. Initial 2Fo-Fc map contoured at 1.3 σ from the dyad region of the DNA (shown here is the base 71 – 73 on one of the strands. **B.** Initial 2Fo-Fc map contoured at 0.8 σ from one of the H3 chains in the hybrid nucleosome showing the region between aa 78 – 82. **C & D.** Initial 2Fo-Fc map contoured at 2.0 and 1.0 σ respectively from the L1-loop region of the H2A chain after molecular replacement with the structures of homotypic macro-NCPs (pdb id 1U35). Preference can be given to neither model on the basis of the electron density map. **E & F.** Initial 2Fo-Fc map contoured at 2.0 and 1.0 σ respectively from the L1 loop of the same H2A chain as in C and D of the hybrid – NCP after molecular replacement with the structure of homotypic H2A-NCPs from *X.laevis* (pdb id 1AOI). **G & H.** Initial 2Fo-Fc map contoured at 2.0 σ from the docking domain of the same H2A chain in the hybrid – NCP after molecular replacement with the structures of homotypic H2A-NCP from X.laevis (1AOI) and homotypic macro-NCP (1U35) respectively. Both seem to be accommodated equally well by the electron density map.

and complexity. The concentration of the variant and major type dimers in our reconstitution mixtures for example may not be an accurate reflection of the relative local concentrations of variant H2A and major H2A *in vivo*. Several chaperone proteins that are emerging for both major type histones and variant histones further complicate nucleosome assembly (Krogan et al., 2003; Mizuguchi et al., 2004). One must recognize that the existence of hybrid nucleosomes *in vitro* does open up a potential new level of heterogeneity in nucleosome structure and function. It is therefore important to obtain *in vivo* corroboration for the existence of hybrid nucleosomes. The three – dimensional structure of the hybrid nucleosome will throw light on the structural differences compared to the homotypic nucleosomes, which in turn may clarify the functional significance of the phenomenon. Most of the sequence differences between major H2A and macroH2A are buried and none of them have any affect on the crystal contacts of nucleosomes. This leads to a lack of discrimination between the two orientations of the hybrid-NCP with respect to the H2A moieties (Fig. 5.3B) that possibly results in a heterogeneous population of the two orientations in the crystal lattice. This complicates the molecular replacement solution and impairs our ability to distinguish between the major type H2A chain and the macroH2A chain. In addition to treating the two orientations as alternate conformations another possible means to address this problem is to try different molecular replacement programs. A recently developed program PHASER uses likelihood enhanced rotation and translation functions, which take advantage of fixed partial models (Storoni et al., 2004). This approach promises a proper accounting of partial solutions and a statistical weighting of multiple solutions. While these approaches may be helpful, it is possible that this problem may remain unsolvable.

CHAPTER 6

Crystallographic Studies with Nucleosome Core Particles from D.melanogaster and the Centromeric Histone H3 Variant CID



The centromeric histone H3 variant CenH3 has generated a lot of interest in recent years. We have attempted to determine the x-ray structure of a nucleosome in which cid (the *D.melanogaster* heterolog of CenH3) replaces major H3. While we could reconstitute cid-NCPs, crystallization is not optimal and we haven't been able to obtain diffracting crystals so far. Our recent work with human CENP-A – NCPs shows more promise. As a control we also determined the 2.3 Å crystal structure of a nucleosome with major histones from *D.melanogaster* (*Dme*-NCPs). We find that in spite of a few sequence differences, *Dme*-NCPs are virtually identical to *Xle*-NCPs (NCPs made of histones from *X.laevis*).

6.1 Abstract

Nucleosomes are the fundamental structural and functional units of chromatin, which is the nucleoprotein complex made up of equal amounts of DNA and histone proteins in eukaryotes. Two copies each of four kinds of histones (namely H2A, H2B, H3, and H4) constitute the protein core of each nucleosome. Besides DNA compaction the nucleosome also is the regulator of DNA accessibility during cellular processes such as transcription. Variants of the core histones, via sequence alterations in strategic areas of the histone fold, may influence the nucleosome structure thus regulating the transcriptional activity of a particular region of the genome. Here, we attempt the crystal structure determination of the nucleosome containing the centromeric H3 variant from *D.melanogaster*, cid (cid-NCP). For comparison, we have determined the 2.3Å crystal structure of a nucleosome core particle made up of all major type histones from *D. melanogaster* (*Dme*-NCP). We find that the *Dme*-NCP structure is by and large similar to the previously determined structure of the NCP made up of histones from *Xenopus laevis* (*Xle*-NCP). Cid-NCPs, on the other hand, exhibit anomalous biochemical behavior and show different crystal growth tendencies. Some of the crystallographic problems and possible solutions are discussed.

6.2 Introduction

Chromosomes replicated during S-phase of the cell cycle have to be separated and transported into separate daughter cytoplasmic domains formed during mitosis. Centromeres direct the formation of kinetochores the motor and microtubule-binding activities of which mediate this transport function. The centromere is a densely packed

heterochromatic domain capped by the trilaminar kinetochore. CenPA is a histone H3-variant originally identified as a centromere-specific auto-antigen that co-purified with nucleosomal core particles (Palmer et al., 1991). It is known as cid (**c**entromere **id**entifier) in *D.melanogaster* and cse4 in yeast (Henikoff et al., 2000; Stoler et al., 1995). Centromeric H3s from different species are more similar in the C-terminal histone fold which shows ~60% homology to the histone fold of major H3 (Malik and Henikoff, 2003). The N-terminal tail is more variable between different species both in length and sequence.

Studies have suggested a mechanism for specific molecular recognition of centromeric DNA at the nucleosomal level (Shelby et al., 1997). It was determined using deletion mutants and domain swapping experiments between CenPA and non-variant H3 that the highly basic N-terminal tail was dispensable for the function of centromere targeting. The histone fold domain was deemed essential for this function. More specifically the L1 loop (aa 75-86) and to a much larger extent the α-2 helix (aa 88-92 and 109-114) of the histone fold influence the targeting function of CenPA (Shelby et al., 1997). Site directed point mutagenesis revealed that Trp-86 is conserved between species and is crucial for targeting. The N-terminal helix also plays a relatively minor role. No independent component is capable of centromere targeting on its own.

Deletion mutant analysis showed that the identification of residues 27-55 as an essential region in the cse4 (Chen et al., 2000). The region between residues 28-60 fused with the histone fold domain of CenPA is fully functional and is thus designated the essential N-terminal domain. Further analysis led to the possibility that this region mediates interactions with other centromeric proteins (Chen et al., 2000).

Earlier studies addressing the timing and mode of incorporation of cid into the centromeric chromatin suggest that centromeric heterochromatin replicates at the same time as pericentric heterochromatin (Sullivan and Karpen, 2001). More recent studies suggest that replication of centromeric heterochromatin is isolated from, and earlier than the replication of pericentric heterochromatin (Ahmad and Henikoff, 2002a). Another intriguing hypothesis that comes out of this study arises from the assumption that cid and H3 nucleosomes are interspersed in this region. These alternating patches of H3 and cid nucleosomes replicate out of phase and this is facilitated by the presence of multiple origins of replication. This unique domain organization also displays a combination of histone modifications that is distinct from that of both euchromatin and the flanking heterochromatin (Sullivan and Karpen, 2004)

Recent studies show that CENP-A and histone H4 form hetero-tetramers that are more compact and conformationally more rigid than the tetramers made of major type H3 and H4 (Black et al., 2004a). They also show that substitution of the domain responsible for the added rigidity into major H3 is sufficient to target it to centromeres. In order to further elucidate the role of this unique variant we think it is imperative that the high-resolution structure of the nucleosome core particle containing CENP-A be determined. The histone fold of H3 is involved in numerous protein – protein and protein – DNA interactions in the nucleosome. Sequence differences in the regions that mediate these interactions will presumably have a tangible effect on the structure of the nucleosome. We have reconstituted nucleosomes with cid and other core histones from *D.melanogaster* (cid-NCP). As a control we also reconstituted a nucleosome containing all major type histones from *D.melanogaster* (*Dme*-NCP) and determined its 2.3 Å

crystal structure. We find that the *Dme*-NCPs in spite of sequence differences in the histone folds of H2A and H2B (H3 and H4 are highly conserved) behave similarly to *Xle*-NCPs (NCPs made of core histone form *X.laevis* (Luger et al., 1997a)) both biochemically and structurally. The cid-NCPs on the other hand show anomalous electrophoretic mobility on a 5% native PAGE gel and show different crystal size and morphology. We have also reconstituted nucleosomes containing human CENP-A and other core histones from *M.musculus* (CENPA-NCPs). While these nucleosomes show comparable electrophoretic mobility with *Xle*-NCPs and *Dme*-NCPs, they show substantial differences in crystal morphology. We have obtained crystals with both cid-NCPs and CENPA-NCPs of sub-optimal size and diffraction properties.

6.3 Materials and Methods

6.3.1 Expression, purification of cid and major Histones from *Drosophila melanogaster*, and reconstitution of cid-NCPs and *Dme*-NCPs

The expression plasmids (pET 3a) for major histones from *Drosophila melanogaster* were a kind gift from Dr. Carl Wu (NIH). The expression plasmid for an N-terminal truncated cid was a kind gift from Dr. Gary Karpen (UC, Berkeley). All histones were over-expressed in BL21 (DE3)-plysS (Stratagene) and purified using previously published protocols (Luger et al., 1999b). The major histones were refolded to a histone octamer (Dyer et al., 2004). Octamer refolding was precluded in presence of cid. Therefore we refolded (cid-H4)$_2$ tetramer and H2A-H2B dimer (Dyer et al., 2004). The *Dme*-octamer was reconstituted onto a 146 bp palindromic DNA fragment derived from human α-satellite regions - α-sat DNA; (Luger et al., 1997a) using salt gradient dialysis

(Dyer et al., 2004), resulting in *Dme*-NCP. Cid-NCPs were reconstituting using salt gradient dialysis by mixing varying molar ratios of $(cid\text{-}H4)_2$ tetramer, H2A-H2B dimer, and the above mentioned DNA fragment (Dyer et al., 2004). The optimal ratio of tetramer: dimer: DNA was determined to be 1:2:0.8. *Dme*-NCPs and cid-NCPs were heat-shifted and purified by preparative gel electrophoresis using published protocols (Dyer et al., 2004).

6.3.2 Purification of hCENPA-H3 tetramer

A bacterial codon-biased synthetic human CENP-A gene was constructed using polymerase chain reaction (PCR) from overlapping oligonucleotides. This was co-expressed with synthetic histone H4 (Luger et al., 1997a) from a bi-cistronic vector derived from pET-12 (Novagen) and BL21-DE3pLysS. The soluble CENP-A – H4 complex was purified to homogeneity using a series of chromatography columns (Black et al., 2004a). The tetramer thus purified was a kind gift from Dr. Don Cleveland (University of California, San Diego). H2A-H2B dimer from mouse histones was purified using published protocols (Dyer et al., 2004). Human CENP-A – NCPs were reconstituted using salt gradient dialysis by mixing $(hCENP\text{-}A - H4)_2$ tetramer, H2A-H2B dimer, and a 146 bp DNA fragment derived from the human α-satellite sequence in a molar ratio of 1:2:0.65. These NCPs were then heat shifted and purified by preparative gel electrophoresis using published protocols (Dyer et al., 2004).

6.3.3 Crystallographic Procedures

NCPs were crystallized using salting in vapor diffusion at NCP concentrations ranging from ~8-12 mg/ml with salt concentrations of 34 to 37.5mM KCl and 40-45mM $MnCl_2$. The crystals were soaked in 24% 2-methyl-2,4-pentanediol (MPD) and 5% trehalose and

frozen in liquid nitrogen as described previously (Luger et al., 1997a). For the *Dme*-NCP crystals data were collected at Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline 5.0.2. Data from a single crystal were processed using Denzo and Scalepack (Otwinowski and Minor, 1997). Molecular replacement was performed using Protein Data Bank entry 1AOI as the search model. Molecular replacement and subsequent rounds of refinement were performed using CNS (Rice et al., 1998). The program O was used for model building (Jones et al., 1991).

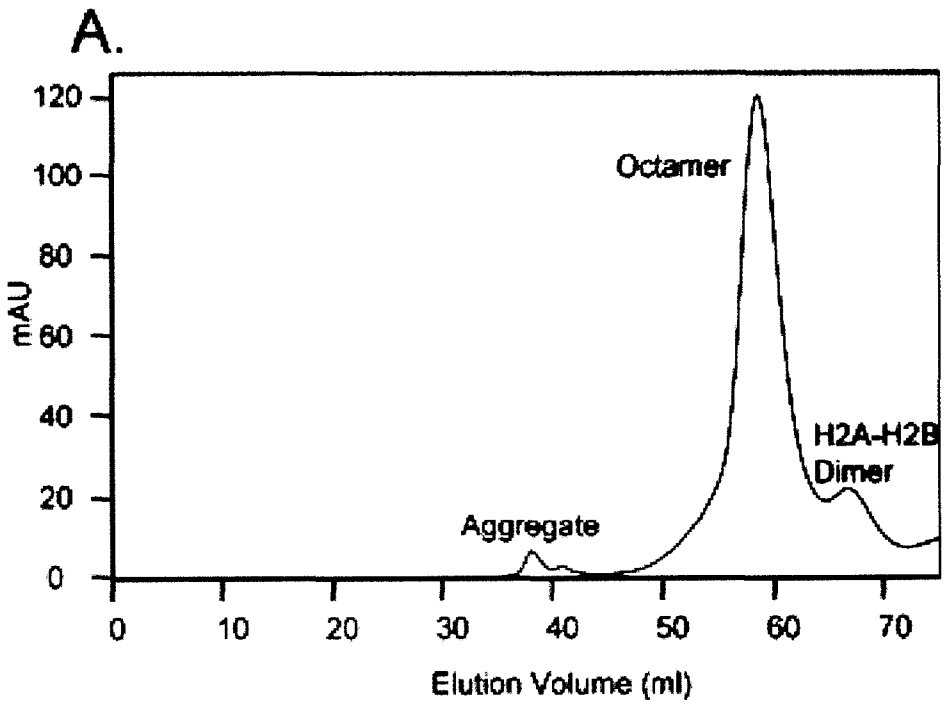
6.4 Results

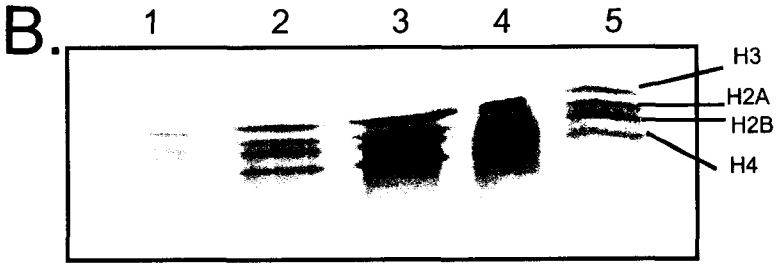
6.4.1 Overall Structure of *Dme*-NCP

Nucleosome structures from *Xenopus laevis* and *Saccharomyces cerevisiae* have been determined (Luger et al., 1997a; White et al., 2001). Although there were some interesting differences between these nucleosomes, their global structures are on the whole similar. In order to see if the general features of nucleosome structure are consistently maintained through out the evolutionary spectrum and as a control in crystallographic studies of cid-NCPs (cid is the *D.melanogaster* heterolog of CENP-A) we embarked on the structure determination of the nucleosome made of core histones from *D.melanogaster*. Between *X.laevis* and *D.melanogaster* H2A and H2B are more variable (84.5% and 82.8% identity respectively) than H3 and H4 (98.5% and 99% respectively) (Sullivan et al., 2002a). We purified histones from *D.melanogaster* using published protocols (Luger et al., 1999b) and refolded the histones into an octamer using pre-optimized procedures (Dyer et al., 2004) (Fig 6.1A & B). Nucleosomes were reconstituted with this octamer and a 146 base pair DNA fragment derived from the

human α-satellite sequence using salt gradient dialysis (Dyer et al., 2004)) (Fig 6.1C). These NCPs were then used in crystallization trials by a salting in vapor diffusion method. Well diffracting crystals characterized on the home source were used for data collection at the synchrotron (ALS Berkeley, BL5.0.2.). Data were obtained to a resolution of 2.3 Å from a single crystal. Phases were determined using the *Xla*-NCP structure (pdb entry 1AOI) as a search model for molecular replacement. We refined the structure to an R factor of 0.237 (R_{free} = 0.258) (table 6.1). A representative region of the electron density contoured to 1.3 σ is shown in Fig. 6.2 A & B.

We see that the structure of *Dme*-NCP is virtually identical to that of *Xle*-NCP and is super-imposable with a root mean square deviation (RMSD) of <1Å. The few sequence differences that do exist between H2A and H2B of the two species are either concentrated on the tail or are not significant enough to alter either the path of the DNA superhelix or the protein – DNA interactions. The only sequence differences that showed potential structural significance at the outset were the residues in the α2 of H2A L55M, T59A, and I62V (Fig. 6.2 C & D). These residues are known to be involved in interactions with H2B that stabilize the H2A-H2B dimer (Luger et al., 1997a). These changes lead to virtually no changes in the structure of the *Dme*-NCP. It is therefore reasonable to assume that whatever functional heterogeneity may exist between nucleosomes from *D.melanogaster* and *X.laevis* is probably not attributable to differences in nucleosome structure. Non-nucleosomal factors (non-histone chromatin associated proteins like heterochromatin protein 1 for example) that influence the chromatin structure and function may also play a vital role in defining the role of chromatin in different contexts.





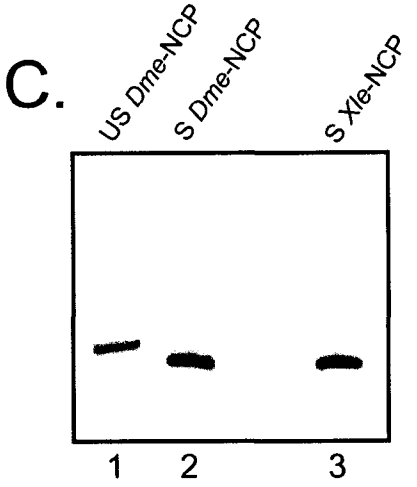


Figure 6.1 Reconstitution of Dme-NCPs. A. Chromatogram showing the elution profile of an octamer made up of core histones from *D.melanogaster* from a superdex 200 size-exclusion column. **B.** Fractions from the superdex-200 size-exclusion column showing octamer made of core histones from *D.melanogaster* (Dme-octamer). 18% SDS-PAGE gel stained with coomassie brilliant blue. **B.** Lane 1&2: Nucleosomes reconstituted by mixing Dme-octamer and a 146 bp fragment of DNA derived from the human α-satellite DNA sequence by salt gradient dialysis (*Dme*-NCPs). Lane 3: *Xle*-NCPs – NCPs containing core histones from *X.laevis*. US and S: before and after incubation at 37°C respectively. 5% non-denaturing PAGE gel stained with coomassie brilliant blue.

Table 6.1: Data Collection and refinement statistics for *Dme*-NCPs

DATA COLLECTION	
Space Group	$P2_12_12_1$
Unit Cell Dimensions (Å)	a = 106.14, b = 109.58, c = 182.04
Resolution (Å)	50 – 2.3
Unique Reflections	93,949
Completeness (%) (overall / last shell)	98.2 / 85.0
R_{merge}[a] (%) (overall / last shell)	4.5 / 30.2
REFINEMENT	
No. of Amino Acid Residues	760
No. of Base Pairs of DNA	146
No. of Water Molecules	156
Total No. of Atoms in the Model	12012
R-factor / R_{free}	0.237 / 0.258
RMSD from Ideality Bonds (Å) Angles (°)	 0.006 0.960
Average B-factors (Å)^2 Protein DNA Solvent	 44.3 96.8 50.9

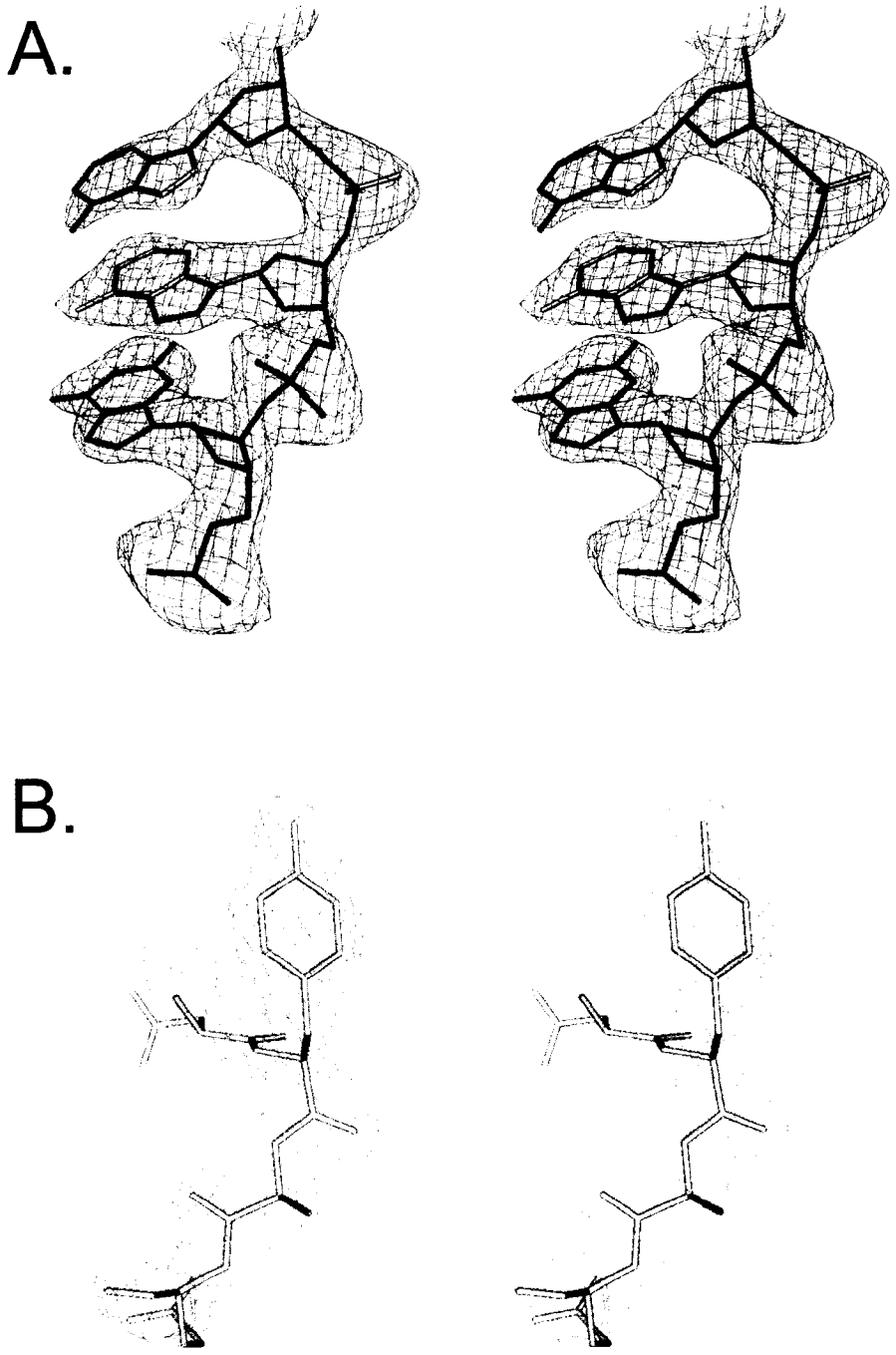


Figure 6.2. A. Stereo image of the 2Fo-Fc electron density map contoured at 1.3σ from the DNA in *Dme*-NCP. Shown here is the region between residue 71-73 (near the dyad) of one of the strands. **B.** Stereo image of the $2F_O$-F_C electron density map contoured at 2.0 σ from the L1-loop region of one of the H2A moieties in the *Dme*-NCP.

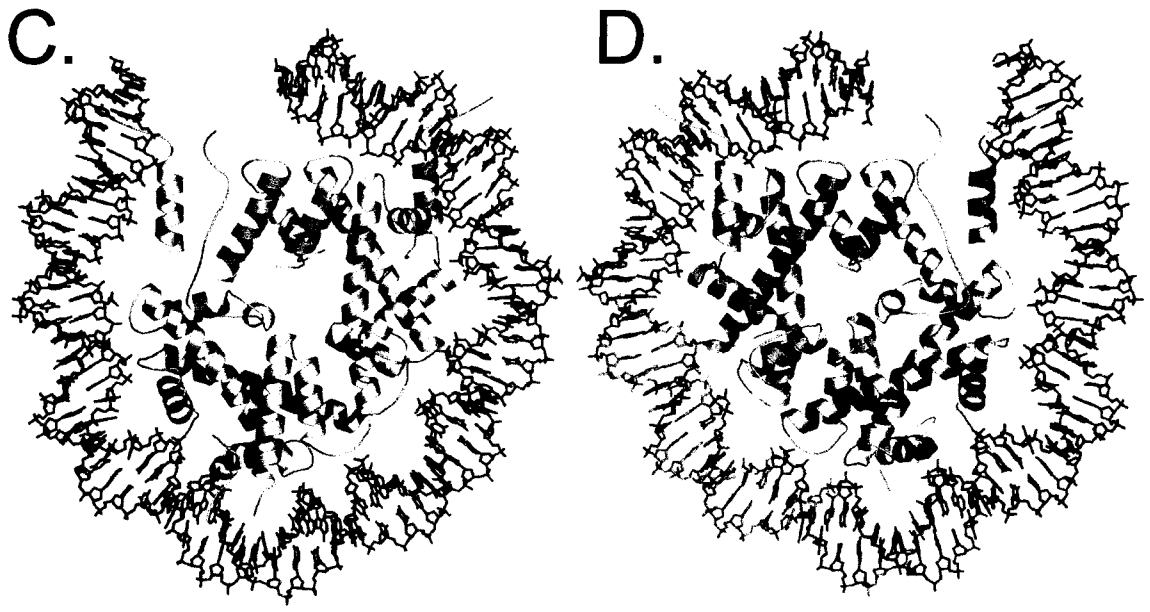


Figure 6.2 Sequence differences between histones from D.melanogaster and X.laevis are mostly concentrated in H2A and H2B. C. The solvent - exposed surface view of one half of the *Dme*-NCP is shown with sequence differences compared to *Xenopus laevis* major-NCP shown in yellow (H2A), and red (H2B). **D.** The same half of the nucleosome viewed from the interior surface between the two gyres of the DNA supercoil.

6.4.2 Purification and Crystallization of cid-NCPs

Cid was mixed with *D.melanogaster* major histones in order to keep the system homogeneous because centromeric histones are strikingly divergent evolutionarily and are known to be species specific (Malik and Henikoff, 2003) (Fig. 6.3A). Cid purified using a pre-established protocol (Luger et al., 1999a) was used in refolding (cid -$H4)_2$ tetramer which was mixed with H2A-H2B dimer from *D.melanogaster* and a 146 bp α-satellite DNA strand to reconstitute cid-NCPs (Dyer et al., 2004). Unlike major-type histones cid cannot refold into a histone octamer along with equi-molar quantities of the other core histones (Fig 6.3 B). Cid can only be refolded into a (cid-$H4)_2$ tetramer (Fig 6.3B and 6.3C lanes 1-4). Reconstitution of the cid-NCPs occurred with sub-optimal efficiency, which was manifested as a reduced intensity in staining with coomassie brilliant blue on a 5% native polyacrylamide gel (Fig 6.3D lanes 1 & 2). In order to optimize the process we then reconstituted cid-NCPs by mixing constant amounts of (cid-$H4)_2$ tetramer, and H2A-H2B dimer, with varying amounts of DNA. The efficiency improved with lower amounts of DNA (Fig 6.3D lanes 3 & 4) (tetramer: dimer: DNA = 1:2:0.8). The electrophoretic mobility of cid-NCPs was slower when compared to *Xle*-NCPs and *Dme*-NCPs (Fig. 6.3 D compare lanes 1-6 with 7). Preparative gel electrophoresis was used to purify cid-NCPs at a concentration of ~5mg/ml (Dyer et al., 2004). Crystals were set up using a salting in vapor diffusion method with varying concentration of MnCl2 and KCl. Crystals appeared almost through out the screen in approximately six weeks. Micro - seeding was tried to optimize crystallization. The morphology of these crystals was different compared to those of *Xle*-NCP and they also failed to grow to a size sufficient for X-ray diffraction experiments (Fig. 6.3D).

A.

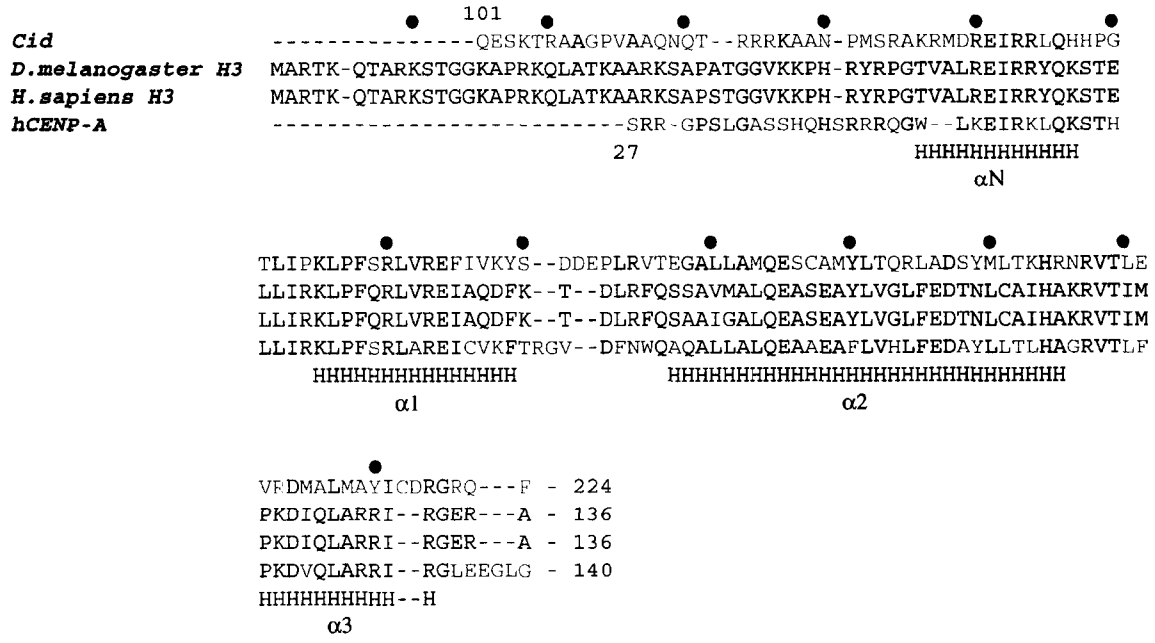
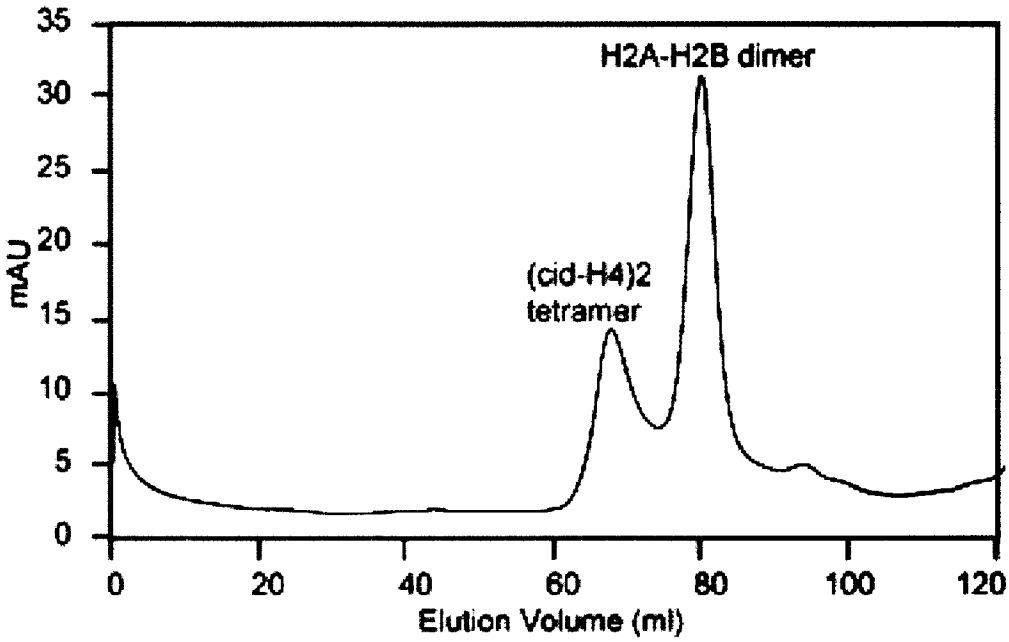
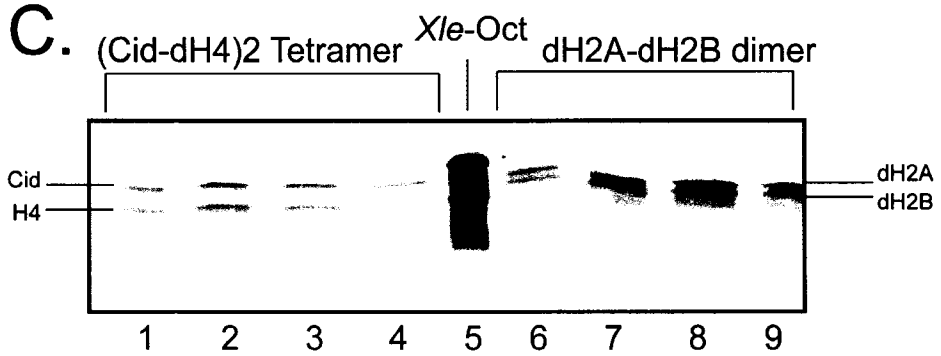


Figure 6.3. A. Sequence alignment between the N-terminal truncated constructs of cid and hCENP-A used in the structural studies and the major type H3 from *H.sapiens* and *D.melanogaster*. Start resides of the two constructs are shown. Filled circles represent every tenth residue in major type H3. Sequence differences present in all four proteins are in blue. Sequence differences found only in cid and CENP-A are in red. The secondary structure elements αN, α1, α2, and α3 are indicated.

B.



C.



D.

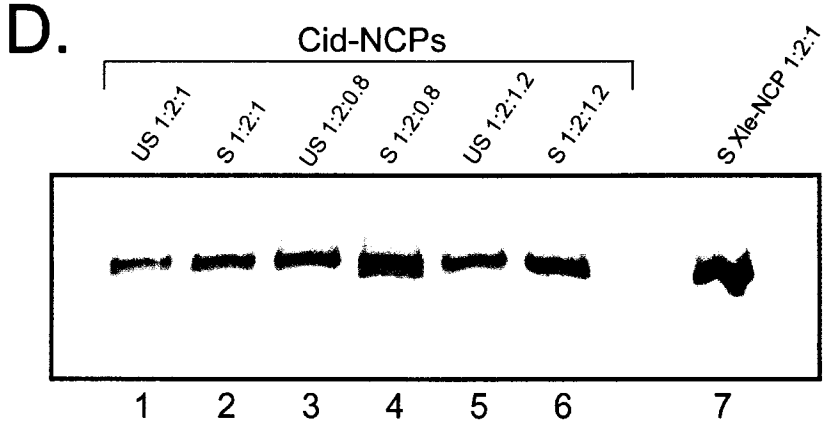


Figure 6.3. Reconstitution of Cid-NCPs. B. Chromatogram showing the elution profiles of cid-H4 tetramer and H2A-H2B dimer from a superdex-200 size-exclusion column. **C.** Lanes 1-4: Fractions from superdex 200 size exclusion column showing $(cid\text{-}H4)_2$ tetramer. Lane 5: Purified octamer containing core histones from *X.laevis*. Lanes 6-9: Fractions from superdex 200 size exclusion column containing H2A-H2B dimer. 18% SDS-PAGE gel stained with coomassie brilliant blue. **D.** Cid-NCPs reconstituted using different ratios of cid-H4 tetramer, H2A-H2B dimer and a 146 base pair α-satellite DNA sequence. Lane 7: *Xle*-NCP. US: Before incubation at 37°C for 1hr; S: After incubation at 37°C for 1hr. 5% non-denaturing gel stained with coomassie brilliant blue.

6.4.3 Reconstitution and Purification of hCENP-A – NCPs

An N-terminal truncated human CENP-A (Fig. 6.3A) was co-purified in association with histone H4 using a bi-cistronic vector derived from pET-12 (Novagen) (Black et al., 2004a). This was purified as soluble (hCENP-A – $H4)_2$ tetramer and along with H2A-H2B dimer and the 146 base pair α-satellite DNA in an optimized molar ratio of 1:2:0.65 was used in nucleosome reconstitution (Fig. 6.4A lanes 1 & 2). Human CENP-A – NCPs unlike cid-NCPs showed electrophoretic mobility comparable to that of non-variant NCPs (Fig 6.4A compare lanes 1-2 with 3). We purified these nucleosomes using preparative gel electrophoresis (Dyer et al., 2004) (Fig. 6.4B). We then set up a crystallization screen with the purified hCENP-A – NCPs that yielded a single crystal in salt concentrations of 36mM KCl, and 42.5mM $MnCl_2$ (Fig. 6.4C). Although this crystal is larger than the ones we obtained with cid-NCPs it still is sub-optimal for the purpose of X-ray diffraction experiments. We have set up sitting drops with larger sample quantities and crystallization considerably slowed down by mineral oil, which may yield better crystals.

6.5 Outlook and Conclusions

We have reconstituted nucleosomes with major histone H3 either substituted by cid or hCENP-A. We find that cid-NCPs do not resemble non-variant nucleosomes in their *in vitro* biochemical behavior but hCENP-A – NCPs do. We attempted crystallization of these variant nucleosomes. We find that nucleosomes with either cid or hCENP-A crystallize less readily than non-variant NCPs. Optimization attempts have not yielded

D.

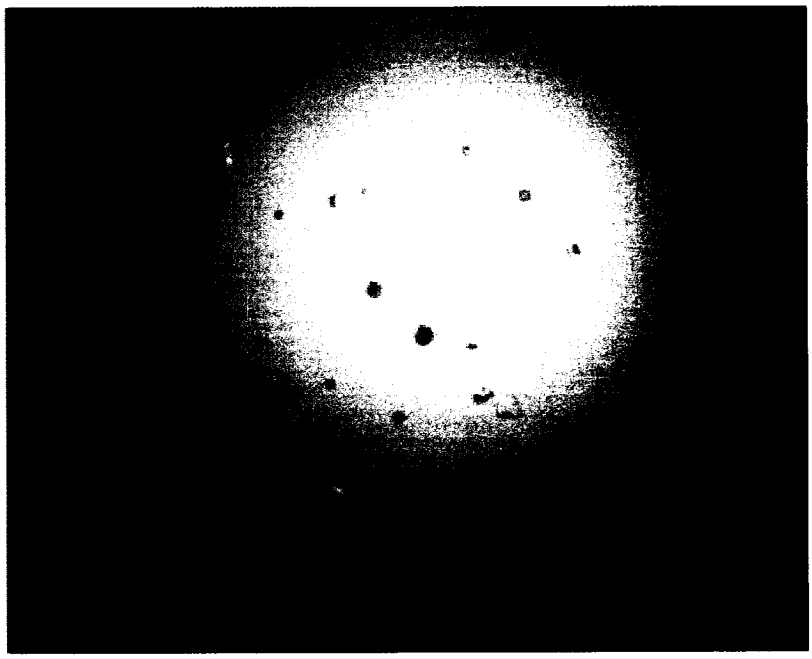


Figure 6.3 D. Crystals of cid-NCPs do not grow to optimal size. Hanging drop crystals grown at salt concentration of 34 to 37.5mM KCl and 40-45mM $MnCl_2$ using the cid-NCPs shown in B lane 4 after purification by preparative gel electrophoresis. ~ 10 μm in their largest dimension.

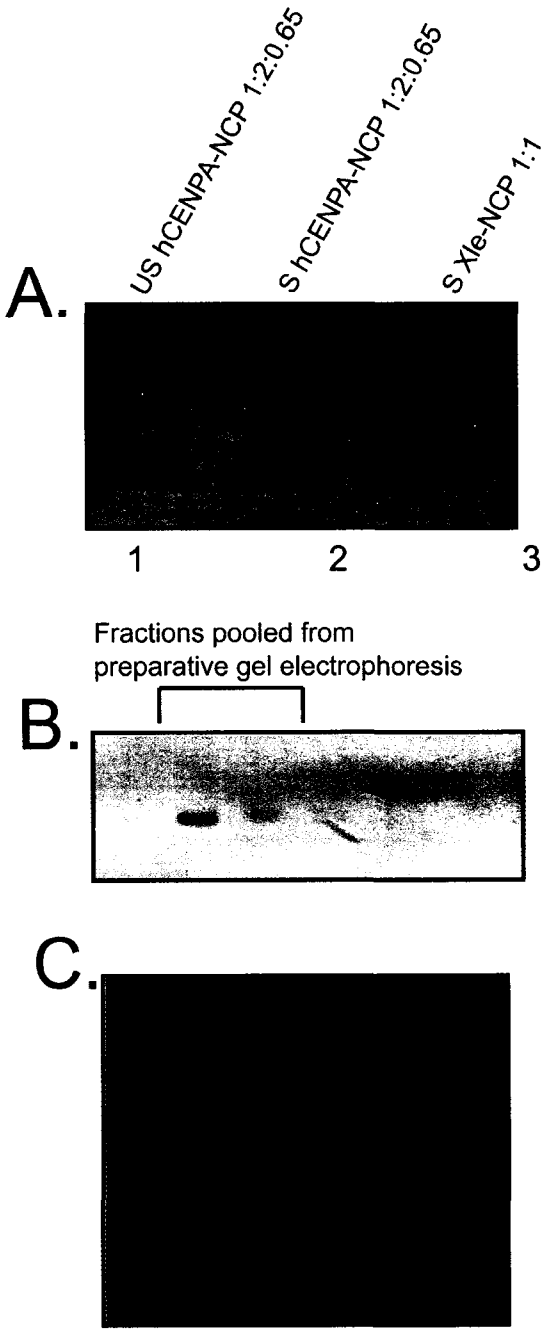


Figure 6.4 Purification and crystallization of hCENP-A – NCPs. A. Lanes 1 & 2: hCENP-A – NCPs reconstituted by mixing hCENP-A – H4 tetramer, H2A-H2B dimer, and 146 bp α-satellite DNA sequence in a molar ratio of 1:2:0.65. Lane 3: *Xle*-NCPs. *US*: Before incubation at 37°C for 1hr; *S*: after incubation at 37°C for 1hr. **B.** Fraction from preparative gel electrophoresis of hCENP-A – NCPs in A lane 2. Both A & B are 5% native PAGE gels stained with comassie brilliant blue. **C.** Crystal obtained from hanging drops set up with sample obtained from preparative gel-electrophoresis shown in B at salt concentrations of 36mM KCl and 42.5mM $MnCl_2$. ~ 400x75x40 μm^3

large enough crystals with cid-NCPs but the most recent trial with hCENP-A – NCPs have yielded by far the largest crystals ever obtained with any heterolog of CenH3. This unusual recalcitrance on the part of CenH3-NCPs in crystallization trials indicates that the structure of these nucleosomes may be more drastically altered than in the case of other variant nucleosomes (Suto et al., 2000). Recent studies in our collaborator Dr. Don Cleveland's lab have suggested that the regions in the H3 histone fold involved in the formation of the four-helix bundle that stabilizes the $(H3\text{-}H4)_2$ hetero-tetramer is involved in targeting hCENP-A to the centromere (Black et al., 2004a). This work also indicated that the hCENP-A – H4 tetramer might be more compact than the canonical H3-H4 tetramer. It may be realistic to assume that these structural changes are drastic enough to disrupt the crystal contacts that usually facilitate crystallization of nucleosomes. Having said that, one cannot exclude the possibility that the failure to obtain homogeneous sample for crystallization hinders crystal growth.

While we are still trying to optimize crystallization conditions, we are also exploring other alternatives to circumvent this problem. We will try reconstituting cse4-NCPs (cse4 is the yeast heterolog of CenH3). We will also try to study the structural influence of the amino acids in CenH3 sequence that are considered crucial for *in vivo* targeting (Black et al., 2004a; Chen et al., 2000; Shelby et al., 1997) by domain swap studies.

CHAPTER 7

Summary and Future Directions

The primary conclusion that can be drawn from the work presented in this thesis is that nucleosome structure on a global scale is unchanged by sequence differences that are characteristic of either histones from different species or histone variants. It is becoming increasingly clear that subtle structural alterations in strategic locations can engender significant functional differences in nucleosomes. We solved the 3Å crystal structure of a nucleosome core particle in which the histone domain of macroH2A replaced both H2A moieties. We see that while the global structure remains unchanged the L1-loop in the histone fold of macroH2A is considerably different, which leads to an altered L1L1-interface between the two macroH2A-H2B dimers. This small yet important interface has been implicated in the cooperative incorporation of the dimers during nucleosome assembly and stabilizing the two superhelical gyres of the nucleosome. A subtle change in this small interface is therefore potentially sufficient to alter the susceptibility of the nucleosome to processes such as transcription, replication, and DNA repair. MacroH2A for example, is known to be generally repressive to the process of transcription and is preferentially localized to the inactive X-chromosome of female mammals. The L1L1-interface of a macro-NCP is predominantly hydrophobic in nature as opposed to that of a non-variant NCP, which is stabilized by salt bridges. The macro L1-loop also has a proline residue that might render it conformationally less flexible and therefore limit the number of possible conformations the nucleosome can adapt thus making it less amenable to nucleosome remodeling that occurs during transcription. We find that the L1-loop in addition to affecting the in vitro assembly pathway of macro-NCPs, also

influences the in vitro stability of the macro-octamer. Studies guided by our biochemical and structural results revealed that the L1-loop is also sufficient albeit not exclusively responsible for the *in vivo* targeting of macroH2A (Communication from Dr. Barbara Panning; UCSF). These results in addition to observations made in studies with other histone H2A variants show that the L1-loop and the docking domain of H2A seem to be the chosen targets of evolution in histone variants.

We also determined the 1.6Å crystal structure of the non-histone domain of macroH2A (aa 180 – 370). We find that it is a α/β fold that has a seven stranded β-sheet and five α-helices, which protect only one face of the β-sheet and leave the other face partially open. This is unusual among proteins with β-sheets but it may suggest that the non-histone domain acts as a recruitment platform for some other protein(s) that are associated with heterochromatin. This domain is also characterized by a predominantly non-descript surface charge distribution except for an unusually large hydrophobic patch. A recent study revealed that this domain acts as a direct roadblock to transcription factor binding. This implies that there is a direct physical association between the non-histone domain and the promoter DNA but nothing we see in the structure alludes to an ability to bind polynucleotide directly. It may on the other hand act as an indirect roadblock by recruiting other protein that can bind chromatin or enzymes that can render the local chromatin impermeable to transcription factors. Preliminary results from studies in Dr. Saadi Khochbin's lab, guided by our structural studies suggest that this region may be involved in binding HDAC1 (histone deacetylase 1), an enzyme that has been implicated in the formation of heterochromatin. There have also been reports that the non-histone domain of macroH2A binds to SPOPI, a protein for which the function is not known.

Further studies therefore need to be carried out to determine the precise nature of the non-histone domain of macroH2A.

MacroH2A also has a linker region that connects the histone domain with the non-histone domain. This region is extremely basic and is homologous to the C-terminal tail of the linker histone H1. It is desirable to investigate the role of the linker region in higher order structure formation. We propose to do so by incorporating different constructs of macroH2A, the histone domain (aa 1-122), the histone domain with the linker region (aa 1-160), full-length macroH2A (1-370), into nucleosome arrays. We will then perform analytical ultracentrifugation to analyze the ability of the nucleosome arrays to fold into fibers of higher compaction. In order to understand the role of the non-histone domain we have to investigate the binding affinity with several protein and polynucleotide candidates implicated in recent studies (e.g. SPOP1, HDAC1, HP1, and Xist). Crystallographic studies need to be performed to understand the structural basis of these putative interactions. While the studies reported in this thesis bring to light significant structural details about the unusual histone variant macroH2A, which in turn have guided functional studies in our collaborators' labs, much more clearly needs to be learnt before we fully understand its functional significance in a physiological context.

REFERENCES

Abbott, D. W., Ivanova, V. S., Wang, X., Bonner, W. M., and Ausio, J. (2001). Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. J Biol Chem *276*, 41945-41949.

Ahmad, K., and Henikoff, S. (2002a). Histone H3 variants specify modes of chromatin assembly. Proc Natl Acad Sci U S A *99 Suppl 4*, 16477-16484.

Ahmad, K., and Henikoff, S. (2002b). The histone variant H3.3 marks active chromatin by replication- independent nucleosome assembly. Mol Cell *9*, 1191-1200.

Ahmad, K., and Henikoff, S. (2002c). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

Angelov, D., Molla, A., Perche, P. Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003a). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Mol Cell *11*, 1033-1041.

Angelov, D., Molla, A., Perche, P. Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003b). The Histone Variant MacroH2A Interferes with Transcription Factor Binding and SWI/SNF Nucleosome Remodeling. Mol Cell *11*, 1033-1041.

Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., and Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A *88*, 10148-10152.

Arents, G., and Moudrianakis, E. N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci U S A *92*, 11170-11174.

Bao, Y., Konesky, K., Park, Y. J., Rosu, S., Dyer, P. N., Rangasamy, D., Tremethick, D. J., Laybourn, P. J., and Luger, K. (2004a). Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. Embo J.

Bao, Y., Konesky, K., Park, Y. J., Rosu, S., Dyer, P. N., Rangasamy, D., Tremethick, D. J., Laybourn, P. J., and Luger, K. (2004b). Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. Embo J *In press*.

Bassing, C. H., Suh, H., Ferguson, D. O., Chua, K. F., Manis, J., Eckersdorff, M., Gleason, M., Bronson, R., Lee, C., and Alt, F. W. (2003). Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. Cell *114*, 359-370.

Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090-1093.

Black, B. E., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L., Jr., and Cleveland, D. W. (2004a). Structural determinants for generating centromeric chromatin. Nature *430*, 578-582.

Black, E. B., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L. J., and Cleveland, D. W. (2004b). Structural Determinants for Generating Centromeric Chromatin. Nature *In press*.

Boeger, H., Griesenbeck, J., Strattan, J. S., and Kornberg, R. D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. Mol Cell *11*, 1587-1598.

Boeger, H., Griesenbeck, J., Strattan, J. S., and Kornberg, R. D. (2004). Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol Cell *14*, 667-673.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., and Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527-542.

Brown, C. J., and Willard, H. F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154-156.

Brown, D. T. (2001). Histone variants: are they functionally heterogeneous? Genome Biol *2*.

Bruno, M., Flaus, A., Stockdale, C., Rencurel, C., Ferreira, H., and Owen-Hughes, T. (2003). Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. Mol Cell *12*, 1599-1606.

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

Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., Bonner, R. F., Bonner, W. M., and Nussenzweig, A. (2003). Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. Nat Cell Biol *5*, 675-679.

Chadwick, B. P., Valley, C. M., and Willard, H. F. (2001). Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. Nucleic Acids Res *29*, 2699-2705.

Chadwick, B. P., and Willard, H. F. (2001a). Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum Mol Genet *10*, 1101-1113.

Chadwick, B. P., and Willard, H. F. (2001b). A Novel Chromatin Protein, Distantly Related to Histone H2A, Is Largely Excluded from the Inactive X Chromosome. J Cell Biol *152*, 375-384.

Chadwick, B. P., and Willard, H. F. (2003). Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Hum Mol Genet *12*, 2167-2178.

Chakravarthy, S., Bao, Y., Roberts, V. A., Tremethick, D. J., and Luger, K. (2004). Structural Characterisation of histone H2A variants. Cold Spring Harb Symp Quant Biol *69*, in press.

Chen, T. A., Smith, M. M., Le, S. Y., Sternglanz, R., and Allfrey, V. G. (1991). Nucleosome fractionation by mercury affinity chromatography. Contrasting distribution of transcriptionally active DNA sequences and acetylated histones in nucleosome fractions of wild-type yeast cells and cells expressing a histone H3 gene altered to encode a cysteine 110 residue. J Biol Chem *266*, 6489-6498.

Chen, Y., Baker, R. E., Keith, K. C., Harris, K., Stoler, S., and Fitzgerald-Hayes, M. (2000). The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. Mol Cell Biol *20*, 7037-7048.

Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M., and Allis, C. D. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell *5*, 905-915.

Clarkson, M. J., Wells, J. R., Gibson, F., Saint, R., and Tremethick, D. J. (1999). Regions of variant histone His2AvD required for Drosophila development. Nature *399*, 694-697.

Cobb, J., Miyaike, M., Kikuchi, A., and Handel, M. A. (1999). Meiotic events at the centromeric heterochromatin: histone H3 phosphorylation, topoisomerase II alpha localization and chromosome condensation. Chromosoma *108*, 412-425.

Costanzi, C., and Pehrson, J. R. (1998a). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Costanzi, C., and Pehrson, J. R. (1998b). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*.

Costanzi, C., and Pehrson, J. R. (2001). MacroH2A2, a new member of the MacroH2A core histone family. J Biol Chem *21*, 21.

Costanzi, C., Stein, P., Worrad, D. M., Schultz, R. M., and Pehrson, J. R. (2000). Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. Development *127*, 2283-2289.

Crane Robinson, C. (1997). Where is the globular domain of linker histone located on the nucleosome? Trends Biochem Sci *22*, 75-77.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat Genet *22*, 323-324.

Culver, G. M., McCraith, S. M., Zillmann, M., Kierzek, R., Michaud, N., LaReau, R. D., Turner, D. H., and Phizicky, E. M. (1993). An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate. Science *261*, 206-208.

Davey, C. A., and Richmond, T. J. (2002). DNA-dependent divalent cation binding in the nucleosome core particle. Proc Natl Acad Sci U S A *99*, 11169-11174.

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.

Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol *375*, 23-44.

Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. (2004). Molecular recognition of the nucleosomal "supergroove". Proc Natl Acad Sci U S A *101*, 6864-6869.

Edmondson, D. G., Smith, M. M., and Roth, S. Y. (1996). Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4. Genes Dev *10*, 1247-1259.

Endo, S., Saito, Y., and Wada, A. (1983). Denaturant-gradient chromatography for the study of protein denaturation: principle and procedure. Anal Biochem *131*, 108-120.

Faast, R., Thonglairoam, V., Schulz, T., Wells, J. R. E., Rathjen, P. D., Tremethick, D. J., and Lyons, I. (1999). H2AZ is essential for early mouse development. (manuscript in preparation).

Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C., and Tremethick, D. J. (2002a). The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nat Struct Biol *19*, 19.

Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C., and Tremethick, D. J. (2002b). The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nat Struct Biol *19*, 172-176.

Fernandez-Capetillo, O., Mahadevaiah, S. K., Celeste, A., Romanienko, P. J., Camerini-Otero, R. D., Bonner, W. M., Manova, K., Burgoyne, P., and Nussenzweig, A. (2003). H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. Dev Cell *4*, 497-508.

Flaus, A., and Owen-Hughes, T. (2001). Mechanisms for ATP-dependent chromatin remodelling. Curr Opin Genet Dev *11*, 148-154.

Flaus, A., and Owen-Hughes, T. (2003). Dynamic properties of nucleosomes during thermal and ATP-driven mobilization. Mol Cell Biol *23*, 7767-7779.

Flaus, A., and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Curr Opin Genet Dev *14*, 165-173.

Flaus, A., Rencurel, C., Ferreira, H., Wiechens, N., and Owen-Hughes, T. (2004). Sin mutations alter inherent nucleosome mobility. Embo J *23*, 343-353.

Gautier, T., Abbott, D. W., Molla, A., Verdel, A., Ausio, J., and Dimitrov, S. (2004). Histone variant H2ABbd confers lower stability to the nucleosome. EMBO Rep *5*, 715-720.

Georges, S. A., Kraus, W. L., Luger, K., Nyborg, J. K., and Laybourn, P. J. (2002). p300-Mediated Tax Transactivation from Recombinant Chromatin: Histone Tail Deletion Mimics Coactivator Function. Mol Cell Biol *22*, 127-137.

Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M. F., Revet, B., and Prunell, A. (1996). Interaction of the histone (H3-H4)2 tetramer of the nucleosome with positively supercoiled DNA minicircles: Potential flipping of the protein from a left- to a right-handed superhelical form. Proc Natl Acad Sci U S A *93*, 7588-7593.

Hansen, J. C. (2002). CONFORMATIONAL DYNAMICS OF THE CHROMATIN FIBER IN SOLUTION: Determinants, Mechanisms, and Functions. Annu Rev Biophys Biomol Struct *31*, 361-392.

Hatch, C. L., and Bonner, W. M. (1988). Sequence of cDNAs for mammalian H2A.Z, an evolutionarily diverged but highly conserved basal histone H2A isoprotein species. Nucleic Acids Res *16*, 1113-1124.

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

Hebbes, T. R., Clayton, A. L., Thorne, A. W., and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. Embo J *13*, 1823-1830.

Hebbes, T. R., Thorne, A. W., Clayton, A. L., and Crane-Robinson, C. (1992). Histone acetylation and globin gene switching. Nucleic Acids Res *20*, 1017-1022.

Henikoff, S., Ahmad, K., Platero, J. S., and van Steensel, B. (2000). Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci U S A *97*, 716-721.

Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet *20*, 320-326.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol *233*, 123-138.

Horn, P. J., and Peterson, C. L. (2002). Molecular biology. Chromatin higher order folding--wrapping up transcription. Science *297*, 1824-1827.

Howman, E. V., Fowler, K. J., Newson, A. J., Redward, S., MacDonald, A. C., Kalitsis, P., and Choo, K. H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci U S A *97*, 1148-1153.

Hoyer-Fender, S., Costanzi, C., and Pehrson, J. R. (2000). Histone macroH2A1.2 is concentrated in the XY-body by the early pachytene stage of spermatogenesis. Exp Cell Res *258*, 254-260.

Hoyer-Fender, S., Czirr, E., Radde, R., Turner, J. M., Mahadevaiah, S. K., Pehrson, J. R., and Burgoyne, P. S. (2004). Localisation of histone macroH2A1.2 to the XY-body is not a response to the presence of asynapsed chromosome axes. J Cell Sci *117*, 189-198.

Hunter, N., Borner, G. V., Lichten, M., and Kleckner, N. (2001). Gamma-H2AX illuminates meiosis. Nat Genet *27*, 236-238.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jones, D. O., Cowell, I. G., and Singh, P. B. (2000). Mammalian chromodomain proteins: their role in genome organisation and expression. Bioessays *22*, 124-137.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjelgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst *A47*, 110-119.

Kimura, H., and Cook, P. R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J Cell Biol *153*, 1341-1353.

Kingston, R. E., and Narlikar, G. J. (1999). ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev *13*, 2339-2352.

Kireeva, M. L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V. M. (2002). Nucleosome Remodeling Induced by RNA Polymerase II. Loss of the H2A/H2B Dimer during Transcription. Mol Cell *9*, 541-552.

Kobor, M. S., Venkatasubrahmanyam, S., Meneghini, M. D., Gin, J. W., Jennings, J. L., Link, A. J., Madhani, H. D., and Rine, J. (2004). A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin. PLoS Biol *2*, E131.

Krogan, N. J., Keogh, M. C., Datta, N., Sawa, C., Ryan, O. W., Ding, H., Haw, R. A., Pootoolal, J., Tong, A., Canadien, V., *et al.* (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. Mol Cell *12*, 1565-1576.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Liu, X., Li, B., and Gorovsky, M. A. (1996). Essential and nonessential histone H2A variants in Tetrahymena thermophila. Mol Cell Biol *16*, 4305-4311.

Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R., and Berger, S. L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell *5*, 917-926.

Louters, L., and Chalkley, R. (1985). Exchange of histones H1, H2A, and H2B in vivo. Biochemistry *24*, 3080-3085.

Luger, K. (2003). Structure and dynamic behavior of nucleosomes. Curr Opin Genet Dev *13*, 127-135.

Luger, K., Maeder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997a). X-ray structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251-259.

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

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

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

Luger, K., and Richmond, T. J. (1998a). DNA binding within the nucleosome core. Current Opinion in Structural Biology *8*, 33-40.

Luger, K., and Richmond, T. J. (1998b). The histone tails of the nucleosome. Curr Opin Genet Dev *8*, 140-146.

Lusser, A., and Kadonaga, J. T. (2003). Chromatin remodeling by ATP-dependent molecular machines. Bioessays *25*, 1192-1200.

Lyon, M. F. (1999). X-chromosome inactivation. Curr Biol *9*, R235-237.

Malik, H. S., and Henikoff, S. (2001). Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics *157*, 1293-1298.

Malik, H. S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat Struct Biol *10*, 882-891.

Mannironi, C., Bonner, W. M., and Hatch, C. L. (1989). H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. Nucleic Acids Res *17*, 9113-9126.

Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J., and Phizicky, E. M. (1999). A biochemical genomics approach for identifying genes by the activity of their products. Science *286*, 1153-1155.

McBryant, S. J., Park, Y. J., Abernathy, S. M., Laybourn, P. J., Nyborg, J. K., and Luger, K. (2003). Preferential binding of the histone (H3-H4)2 tetramer by NAP1 is mediated by the amino-terminal histone tails. J Biol Chem *278*, 44574-44583.

Meneghini, M. D., Wu, M., and Madhani, H. D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell *112*, 725-736.

Mermoud, J. E., Costanzi, C., Pehrson, J. R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J Cell Biol *147*, 1399-1408.

Mizuguchi, G., Shen, X., Landry, J., Wu, W. H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343-348.

Monesi, V., Grippa, M., and Zito-Bignami, R. (1967). The stage of chromosome duplication in the cell cycle as revealed by x-ray breakage and 3H-thymidine labeling. Chromosoma *21*, 369-386.

Motzkus, D., Singh, P. B., and Hoyer-Fender, S. (1999). M31, a murine homolog of Drosophila HP1, is concentrated in the XY body during spermatogenesis. Cytogenet Cell Genet *86*, 83-88.

Muthurajan, U. M., Bao, Y., Forsberg, L. J., Edayathumangalam, R. S., Dyer, P. N., White, C. L., and Luger, K. (2004). Crystal structures of histone Sin mutant nucleosomes reveal altered protein-DNA interactions. Embo J *23*, 260-271.

Muthurajan, U. M., Park, Y. J., Edayathumangalam, R. S., Suto, R. K., Chakravarthy, S., Dyer, P. N., and Luger, K. (2003). Structure and dynamics of nucleosomal DNA. Biopolymers *68*, 547-556.

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode, Vol 276, Macromolecular Crystallography, part A (New York, Academic Press).

Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H., and Margolis, R. L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A *88*, 3734-3738.

Park, Y. J., Dyer, P. N., Tremethick, D. J., and Luger, K. (2004a). 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.

Park, Y. J., Dyer, P. N., Tremethick, D. J., and Luger, K. (2004b). 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.

Pehrson, J. R., Costanzi, C., and Dharia, C. (1997). Developmental and tissue expression patterns of histone macroH2A1 subtypes. Cell Biochem *65*, 107-113.

Pehrson, J. R., and Fried, V. A. (1992). MacroH2A, a core histone containing a large nonhistone region. Science *257*, 1398-1400.

Pehrson, J. R., and Fuji, R. N. (1998). Evolutionary conervation of histone macroH2A subtypes and domains. Nucleic Acids Research *26*, 2837-2842.

Perche, P., Vourc'h, C., Konecny, L., Souchier, C., Robert-Nicoud, M., Dimitrov, S., and Khochbin, S. (2000a). Higher concentrations of histone macroH2A in the barr body are correlated with higher nucleosome density [In Process Citation]. Curr Biol *10*, 1531-1534.

Perche, P. Y., Vourc'h, C., Konecny, L., Souchier, C., Robert-Nicoud, M., Dimitrov, S., and Khochbin, S. (2000b). Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density. Curr Biol *10*, 1531-1534.

Phelan, M. L., Sif, S., Narlikar, G. J., and Kingston, R. E. (1999). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol Cell *3*, 247-253.

Protacio, R. U., and Widom, J. (1996). Nucleosome transcription studied in a real-time synchronous system: test of the lexosome model and direct measurement of effects due to histone octamer. J Mol Biol *256*, 458-472.

Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219-223.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D. J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. Embo J *22*, 1599-1607.

Rangasamy, D., Greaves, I., and Tremethick, D. J. (2004). RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. Nat Struct Mol Biol *11*, 650-655.

Rasmussen, T. P., Huang, T., Mastrangelo, M. A., Loring, J., Panning, B., and Jaenisch, R. (1999). Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. Nucleic Acids Res *27*, 3685-3689.

Rasmussen, T. P., Mastrangelo, M. A., Eden, A., Pehrson, J. R., and Jaenisch, R. (2000). Dynamic relocalization of histone MacroH2A1 from centrosomes to inactive X chromosomes during X inactivation. J Cell Biol *150*, 1189-1198.

Rasmussen, T. P., Wutz, A. P., Pehrson, J. R., and Jaenisch, R. R. (2001). Expression of Xist RNA is sufficient to initiate macrochromatin body formation. Chromosoma *110*, 411-420.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases [see comments]. Nature *406*, 593-599.

Read, R. Y. (1986). Improved Fourier Coefficients for maps using phases from partial structures with errors. Acta Cryst *A42*, 140-149.

Rice, L. M., Shamoo, Y., and Brunger, A. T. (1998). Phase improvement by moluti-start simulated annealing refinement and structure-factor averaging. J Appl Cryst *31*, 798-805.

Richler, C., Soreq, H., and Wahrman, J. (1992). X inactivation in mammalian testis is correlated with inactive X-specific transcription. Nat Genet *2*, 192-195.

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

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

Rogakou, E. P., Nieves-Neira, W., Boon, C., Pommier, Y., and Bonner, W. M. (2000). Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. J Biol Chem *275*, 9390-9395.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858-5868.

Roth, S. Y., Denu, J. M., and Allis, C. D. (2001). Histone acetyltransferases. Annu Rev Biochem *70*, 81-120.

Rothkamm, K., and Lobrich, M. (2003). Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses. Proc Natl Acad Sci U S A *100*, 5057-5062.

Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996). Reduced surface: an efficient way to compute molecular surfaces. Biopolymers *38*, 305-320.

Santisteban, M. S., Kalashnikova, T., and Smith, M. M. (2000). Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes. Cell *103*, 411-422.

Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci *15*, 430-434.

Schiltz, R. L., Mizzen, C. A., Vassilev, A., Cook, R. G., Allis, C. D., and Nakatani, Y. (1999). Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. J Biol Chem *274*, 1189-1192.

Shelby, R. D., Vafa, O., and Sullivan, K. F. (1997). Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. J Cell Biol *136*, 501-513.

Stoler, S., Keith, K. C., Curnick, K. E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev *9*, 573-586.

Storoni, L. C., McCoy, A. J., and Read, R. J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr *60*, 432-438.

Strater, N., and Lipscomb, W. N. (1995). Two-metal ion mechanism of bovine lens leucine aminopeptidase: active site solvent structure and binding mode of L-leucinal, a gem-diolate transition state analogue, by X-ray crystallography. Biochemistry *34*, 14792-14800.

Strater, N., Sherratt, D. J., and Colloms, S. D. (1999). X-ray structure of aminopeptidase A from Escherichia coli and a model for the nucleoprotein complex in Xer site-specific recombination. Embo J *18*, 4513-4522.

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.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes Dev *12*, 599-606.

Sullivan, B., and Karpen, G. (2001). Centromere identity in Drosophila is not determined in vivo by replication timing. J Cell Biol *154*, 683-690.

Sullivan, B. A., and Karpen, G. H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol.

Sullivan, K. F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol *127*, 581-592.

Sullivan, S., Sink, D. W., Trout, K. L., Makalowska, I., Taylor, P. M., Baxevanis, A. D., and Landsman, D. (2002a). The Histone Database. Nucleic Acids Res *30*, 341-342.

Sullivan, S., Sink, D. W., Trout, K. L., Makalowska, I., Taylor, P. M., Baxevanis, A. D., and Landsman, D. (2002b). The Histone Database. Nucleic Acids Res *30*, 341-342.

Sullivan, S. A., Aravind, L., Makalowska, I., Baxevanis, A. D., and Landsman, D. (2000). The histone database: a comprehensive WWW resource for histones and histone fold-containing proteins. Nucleic Acids Res *28*, 320-322.

Suto, R. K., Clarkson, M. J., Tremethick, D. J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol *7*, 1121-1124.

Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. (2003). Crystal Structures of Nucleosome Core Particles in Complex with Minor Groove DNA-binding Ligands. J Mol Biol *326*, 371-380.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.

Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p [see comments]. Science *272*, 408-411.

Terwilliger, T. C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. Methods Enzymol *374*, 22-37.

Thatcher, T. H., and Gorovsky, M. A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res *22*, 174-179.

Turner, J. M., Burgoyne, P. S., and Singh, P. B. (2001). M31 and macroH2A1.2 colocalise at the pseudoautosomal region during mouse meiosis. J Cell Sci *114*, 3367-3375.

van Daal, A., and Elgin, S. C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. Mol Biol Cell *3*, 593-602.

Van Holde, K. E. (1988). Chromatin (New York, Springer-Verlag).

Vermaak, D., Hayden, H. S., and Henikoff, S. (2002). Centromere targeting element within the histone fold domain of Cid. Mol Cell Biol *22*, 7553-7561.

Vermaak, D., and Wolffe, A. P. (1998). Chromatin and chromosomal controls in development. Dev Genet *22*, 1-6.

Via, A., Ferre, F., Brannetti, B., Valencia, A., and Helmer-Citterich, M. (2000). Three-dimensional view of the surface motif associated with the P-loop structure: cis and trans cases of convergent evolution. J Mol Biol *303*, 455-465.

White, C. L., Suto, R. K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. Embo J *20*, 5207-5218.

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D., and Owen-Hughes, T. (2003). Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Mol Cell Biol *23*, 1935-1945.

Wolffe, A. P. (1998). Packaging principle: how DNA methylation and histone acetylation control the transcriptional activity of chromatin. J Exp Zool *282*, 239-244.

Workman, J. L., and Kingston, R. E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu Rev Biochem *67*, 545-579.

Wreggett, K. A., Hill, F., James, P. S., Hutchings, A., Butcher, G. W., and Singh, P. B. (1994). A mammalian homologue of Drosophila heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. Cytogenet Cell Genet *66*, 99-103.

Wu, R. S., Tsai, S., and Bonner, W. M. (1982). Patterns of histone variant synthesis can distinguish G0 from G1 cells. Cell *31*, 367-374.

Xiao, H., Sandaltzopoulos, R., Wang, H. M., Hamiche, A., Ranallo, R., Lee, K. M., Fu, D., and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. Mol Cell *8*, 531-543.