THE UNIVERSITY *of* EDINBURGH

A characterisation of CenH3 nucleosomes

A Thesis Submitted for the Degree of
Doctor of Philosophy
By

Matthew Miell

To

The University of Edinburgh
September 2012

Declaration

I declare that the work presented herein has been composed by myself and represents my own work, unless otherwise stated. I also declare that this work has not been submitted for any other degree or professional qualification.

Matthew Miell

September 2012

Acknowledgements

I wish to extend my sincere thanks to my principal supervisor, Robin. Thank you for providing me with the opportunity and freedom to pursue my research interests, for your committed support, and for our valuable discussions. I am deeply grateful for your guidance and insight, which has helped me to develop both scientifically and personally over the past four years.

I would also like to thank my second supervisors, Tom and Jim, whose continuous support and expertise has been invaluable.

I am grateful to Ken and Chris, who as members of my thesis panel have often offered stimulating discussions and beneficial suggestions. Also to SULSA and the BBSRC for funding.

A huge thank you to the numerous people who have generously mentored and supported me in various techniques with patience. To Andy Bowman and others in the Owen-Hughes lab for teaching me to make nucleosomes. To Andy Downes for introducing me to AFM. To Harald and Nick for all their help with bioinformatics. To Juri and Helene, for help with mass spectrometry. Finally, to all the staff of the Edinburgh Protein Production Facility and the Biophysical Characterisation Facility, for their invaluable advice on so many aspects of protein biochemistry.

Special thanks extend to all the members of the Allshire lab, past and present, for their support, advice, friendship, and for generally making life in the lab so enjoyable and productive. Particular thanks go to Ali, whose valued suggestions, expertise and support have been outstanding. Also to

George, without whom the lab could not function as smoothly as her cheery presence ensures.

On a personal note, I wish to thank all those who have supported me personally to accomplish this work. I am particularly grateful to the tea-break gang, Sandra, Pauline, Lisi and Anne. The friendship and daily doses of caffeine they inspire is highly valued. I am especially grateful to Sandra, whose friendship and support has been consistent and indispensable since we first met back at our interviews. This work would not have been possible without the support and encouragement that my parents, Dorothy and David have provided throughout every aspect of my life. I am truly grateful for their influence in everything I have achieved. Lastly, I would like to express my heartfelt thanks for the unwavering support, encouragement and fun that my partner Hannah has ensured is present in every day.

Abstract

As a centromere-specific protein complex in direct contact with the DNA, CenH3-containing nucleosomes are generally thought to act as the distinguishing epigenetic mark of active centromere location. Confusingly, seemingly disparate models have been proposed for the structure of CenH3 nucleosomes. The most widely supported model is an octameric structure that, like histone H3 nucleosomes, contains two subunits of each histone. Another more contentious, yet persistent model is the hemisome model proposed for fly and human CenH3 nucleosomes. In this case it is suggested that CenH3 nucleosomes contain only single subunit of each histone.

One reason for this lack of consensus is that seemingly contradicting models are often proposed, even with material from the same organism, with little overlap in experimental approaches. For example, the proposed hemisome model for fly and human CenH3 nucleosomes is predominantly based on atomic force microscopy (AFM) imaging where the height of nucleosomes on a surface is measured. These AFM measurements are the main data used by protagonists for the hemisome model. However, data supporting an octameric model for human, and other, CenH3 nucleosomes is largely based on biochemical analysis of nucleosomes prepared in vitro, with little cross-over in the methodology used to generate data to support either model.

In order to reach a consensus the same analyses needs to be applied to CenH3 nucleosomes assembled *in vitro* or extracted from cells. Here, recombinant *Schizosaccharomyces pombe* CENP-A^{Cnp1} and H3 histones expressed and purified from *E. coli* have been assembled into nucleosomes.

To our knowledge this is the first time that recombinant *S. pombe* nucleosomes have been produced, allowing the stoichiometry and composition of these nucleosomes to be examined in detail by a variety of biochemical and biophysical assays. The application of AFM has enabled the height of these recombinant nucleosomes to be measured and tests the ability of AFM to infer stoichiometry using defined material. The intriguing conclusion is that octameric CenH3 nucleosomes uniquely behave as tetrameric "hemisomes" as defined by AFM.

In recent years the contribution of DNA sequence to directing H3 nucleosome location has received a great deal of interest. Since CENP-A^{Cnp1} nucleosomes wrap DNA differently to H3 nucleosomes their preference for sequences that produce a stable nucleosome is expected to be altered. The development of protocols to assemble recombinant CENP-A^{Cnp1} nucleosomes *in vitro* has also been used here to assess the contribution of primary DNA sequence to CENP-A^{Cnp1} nucleosome positioning. CENP-A^{Cnp1} and H3 nucleosomes were reconstituted on genomic DNA at low density and the resulting nucleosomal DNA from CENP-A^{Cnp1} and H3 particles compared by Illumina sequencing. The stability of CENP-A^{Cnp1} and H3 nucleosomes on specific 'H3' and 'CENP-A^{Cnp1}' sequences was cross-checked. Comparing these data with *in vivo* CENP-A^{Cnp1} nucleosome positions has allowed the contribution of primary DNA sequence to CENP-A^{Cnp1} nucleosome positioning to be explored.

Table of Contents

List of Figures

List of Tables

Abbreviations

°C	degrees centigrade
AFM	Atomic force microscopy
ATP	Adenosine triphosphate
Da	Daltons
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	diaminoethanetetra-acetic acid
EGTA	ethylene glycol tetra-acetic acid
FRET	Förster (Fluorescence) resonance energy transfer
g	gram
g	x gravity
GFP	Green fluorescent protein
h	hour
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
IP	immunoprecipitation
kb	kilobase
kDa	kilodaltons
M	molar
mg	milligram
min	minute
ml	millilitre
mM	millimolar
nM	nanomolar

OD	optical density
PAGE	polyacrylamide gel electrophoresis
PALM	photo-activated localisation microscopy
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pmol	picomole
PMSF	Phenylmethyanesulphonyfluoride
SDS	Sodium dodecyl sulfate
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
UV	ultraviolet
μg	microgram
V	volts

Chapter 1

Introduction

Chromatin organisation and function

DNA effectively stores around 400 million times more binary digits of information per mm^3 than a typical Blu-ray DVD (Church *et al*., 2012). Yet even with such an efficient storage medium, compacting a genome of billions of DNA base pairs into the nucleus of a cell remains an engineering marvel. Especially considering that genomes are dynamic structures, continually requiring small regions to be located, locally decompacted, processed and packaged away again within minutes, and then for the whole genome to be duplicated and segregated reliably over each cell cycle.

To compact the genome in an organised manner, all eukaryotes have adopted a common mechanism by which ~ 147 bp sections of DNA are tightly wrapped ~ 1.65 times around a cylindrical core of highly conserved histone proteins (H2A, H2B, H3 and H4), forming a complex called a nucleosome (Luger *et al*., 1997). These nucleosomes typically form every ~ 200 bp along the genome, leaving ~ 20 – 60 bp between nucleosomes as a relatively flexible linker that in some higher eukaryotes is bound by "linker" histone proteins (H1 and isoforms). Arrays of nucleosomes are further compacted by supercoiling into higher order structures that are mediated both by inter-nucleosomal interactions and the binding of a large number of non-histone proteins (Robinson and Rhodes, 2006), although some evidence

is emerging that disputes the widespread compaction of nucleosomes into a "30nm fibre" *in vivo* (Fussner *et al*., 2012; Maeshima *et al*., 2010; Nishino *et al*., 2012). This combination of nucleosomes, linker histones and other non-histone proteins is collectively known as chromatin.

Historically chromatin has been broadly classified into two distinct groups, euchromatin and heterochromatin, according to the level to which regions of the genome stained with histological dyes (Elgin, 1996). Euchromatin accounts for the majority of the genome, it is largely transcriptionally active, and stains poorly during interphase because it is less compact than heterochromatin. As more is understood about how chromatin influences cellular processes, certain post-translational modifications of histones have become recognised as markers for chromatin features. Canonical heterochromatin, typically required at subtelomeres and flanking centromeres of higher eukaryotes, is di- or tri-methylated on lysine 9 of histone H3 which directly leads to the recruitment of a protein associated with transcriptional silencing called Heterochromatin Protein 1 (HP1) and associated factors (Bannister *et al*., 2001; Fischer *et al*., 2009; Rea *et al*., 2000; Sullivan and Karpen, 2004). Facultative heterochromatin is formed at otherwise euchromatic regions in a developmentally linked manner, the primary examples of which are the transcriptional inactivation of a single X chromosome in somatic cells of female mammals and repression of the *Hox* genes. Facultative heterochromatin is enriched in H3 lysine 9 methylation but does not necessarily interact with HP1 proteins (Peters *et al*., 2002). Instead, repression of *Hox* genes is dependent on another group of chromodomain proteins, the repressive Polycomb Group (PcG) proteins that bind to histone

H3 methylated at lysine 27 (Cao *et al*., 2002). Recently it has been proposed that chromatin can be classified more specifically according to current understanding of correlations between histone post-translational modifications, interacting non-histone proteins and different chromatin states, with up to 5 distinct flavours of chromatin proposed (Filion *et al*., 2010).

Whilst histone post-translational modifications are clearly an important regulatory mechanism for altering chromatin structure and accessibility this is also achieved through the incorporation of specific variant histones and other non-histone proteins. For example, the H2A variant histone macroH2A is enriched on the inactive X chromosome (Xi) and thought to aid silencing (Chadwick and Willard, 2002), whilst the histone H3 variant H3.3 is found often found at sites of active transcription, where is it proposed to form inherently unstable nucleosomes (Jin and Felsenfeld, 2007). In all eukaryotes that have been studied, the location of active centromeres appears to be defined by the incorporation of a centromere-specific histone H3 variant known generically as CenH3 (Allshire and Karpen, 2008). How CenH3 is specifically recruited and incorporated at centromeres is unclear but appears to be predominantly regulated epigenetically in all studied eukaryotes, with the notable exception of budding yeast. The incorporation of CenH3 nucleosomes are essential to define a distinctive region of chromatin that leads to kinetochore assembly and microtubule attachment (Sullivan and Karpen, 2004), yet functional and heritable "neocentromeres" are able to form and recruit kinetochore proteins on previously euchromatic regions of DNA without a change in the underlying DNA sequence (Ishii *et al*., 2008;

Ketel *et al*., 2009; Saffery *et al*., 2001; Voullaire *et al*., 1993). Similarly, in chromosomes that contain two centromeres as a result of chromosomal fusion one will be inactivated, no longer able to recruit CenH3 or kinetochore components, again without a change in the underlying DNA sequence (Agudo *et al*., 2000; Earnshaw and Migeon, 1985; Sullivan and Willard, 1998). Thus centromeres demonstrate an epigenetic plasticity, dependent on the specific incorporation of CenH3 nucleosomes rather than a particular DNA sequence. This plasticity might have evolved as a consequence of the crucial requirement for chromosomes to contain a single active centromere, as defects in centromere function lead to chromosome breakage, loss and gain that typically result in cell death or diseases such as cancer (Weaver *et al*., 2007). Thus, defining how centromeres are established and maintained to allow normal function and accurate chromosome segregation is critical to our understanding of these processes. The purpose of this thesis is twofold, firstly to critically assess the two predominant models for CenH3 nucleosome structure and secondly to characterise the contribution of DNA sequence to the location of CenH3 nucleosomes.

Nucleosome structure and function

Histones

Histones are a group of highly conserved proteins that bind DNA and function to compact, organise and regulate access to the DNA. There are two major families of histone proteins, the "core" histones, H2A, H2B, H3 and H4

and the "linker" histone H1 (including the avian H5 isoform). The core histones are extremely conserved across eukaryotes, particularly histones H3 and H4, and two copies of each core histone form the octameric particle around which DNA is wrapped to make nucleosomes (Luger *et al*., 1997). Linker histones bind the DNA between nucleosomes to aid nucleosome spacing and compaction of nucleosome arrays and have been proposed to play a limited role in transcriptional regulation but are less well conserved than the core histones (Robinson and Rhodes, 2006; Schäfer *et al*., 2008; Wong *et al*., 2007a).

Histones are small proteins, typically between 11 – 15 kDa and are highly basic, allowing them to bind DNA non-specifically. The core histones have three major domains: 1) a histone-fold domain consisting of three alpha-helices connected by two loops, 2) an unstructured lysine-rich amino-terminal tail and 3) a short, relatively unstructured carboxyl-terminal domain (Luger *et al*., 1997) (Figure 1-1). To date, the vast majority but by no means all post-translational modifications on histones occur in the N-terminal tails (Figure 1-2), which extend out from the nucleosome and so are easily accessible to a variety of enzymatic modifiers. The long unstructured histone tails appear to "hug" adjacent nucleosomes when unmodified (in *in vitro* studies) (Luger *et al*., 1997) whilst acetylation neutralises the basic charge of tails, releasing them and allowing them to interact with various ligands (Luger and Richmond, 1998a). Methylation of histones also directly facilitates the binding of proteins with specific domains such as PHD fingers or chromodomains (Zhang and Reinberg, 2001). Further, evidence that suggests cross-talk between various modifications and their binding partners has led

to the proposal of a "histone code" (Strahl and Allis, 2000) which predicts that the presence of a pre-existing histone modification influences subsequent histone modifications.

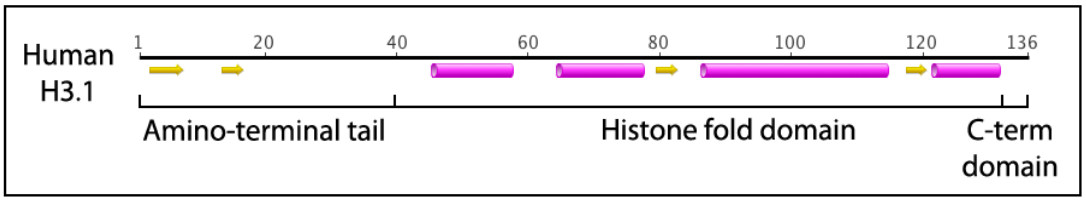


Figure 1-1 – Secondary structure of human histone H3.1

Each of the core histones has a similar secondary structure, with a long unstructured amino-terminal tail that extends from the nucleosome core, a histone fold domain of alpha helices separated by short loops, and a short unstructured carboxyl-terminal domain.

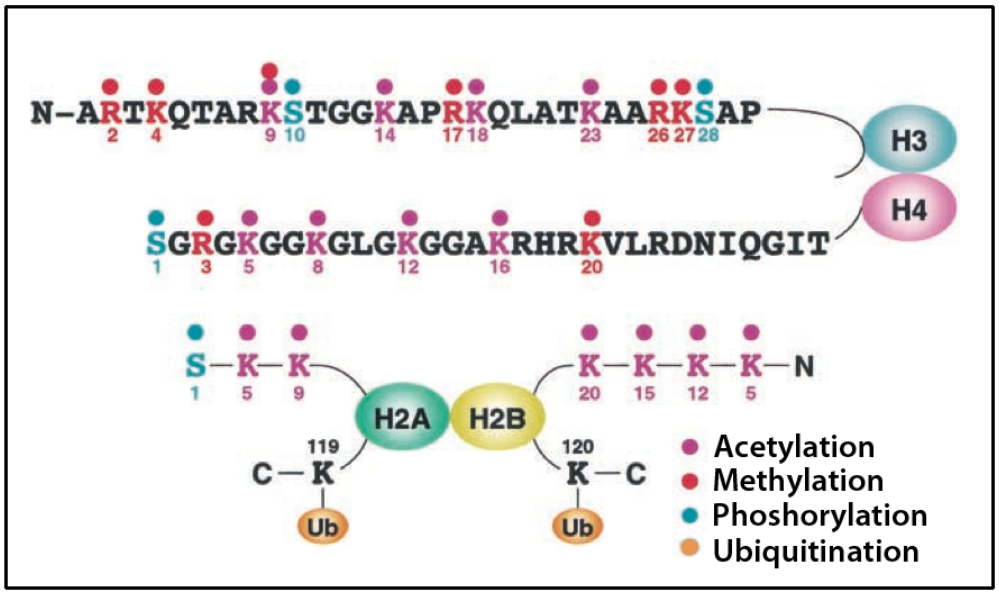


Figure 1-2 – Sites of post-translational modifications on the histone tails

Illustration of the core histone tails highlighting the sites of acetylation (purple), methylation (red), phoshorylation (green) and ubiquitination (orange). Figure adapted from Zhang *et al* (2001).

Core histones are expressed and incorporated into nucleosomes during S phase in a replication-dependent manner, so that DNA left exposed after replication is bound by newly synthesised histones and wrapped into nucleosomes (Kamakaka and Biggins, 2005). A number of variant histones exist for each of the canonical core histone types, except H4 (Table 1-1). Typically these variant histones differ by no more than a few amino acids (Figure 1-3) yet confer distinct nucleosomal architectures and specialised functions. For example, phosphorylation of H2A.X (known as gamma-H2A.X once phosphorylated) is specific to sites of DNA damage and is thought to help recruit factors required for the DNA damage response (Paull *et al*., 2000; Rogakou *et al*., 1999). In keeping with their transient requirement at such sites, variant histones are typically synthesised and incorporated in a replication-independent manner, to be used where and when required (Kamakaka and Biggins, 2005).

Variant	Species	Chromatin effect	Function
$H1^0$	Mouse	Chromatin condensation	Transcription repression
H5	Chicken	Chromatin condensation	Transcription repression
spH1	Sea urchin	Chromatin condensation	Chromatin packaging
H1t	Mouse	Open chromatin	Histone exchange, recombination?
MacroH2A1/2	Vertebrate	Condensed chromatin	X-chromosome inactivation
H2ABbd	Vertebrate	Open chromatin	Transcription activation
H2A.X	Ubiquitous	Condensed chromatin	DNA repair/recombination/transcription repression
H2A.Z	Ubiquitous	Open/closed chromatin	Transcription activation/repression, chromosome segregation
spH2B	Sea urchin	Chromatin condensation	Chromatin packaging
CenH3	Ubiquitous		Kinetochore formation/function
H3.3	Ubiquitous	Open chromatin	Transcription

Table 1-1 – Histone variants and their functions

The species distribution and likely functions of major histone variants are shown. Figure adapted from Kamakaka *et al* (2005).

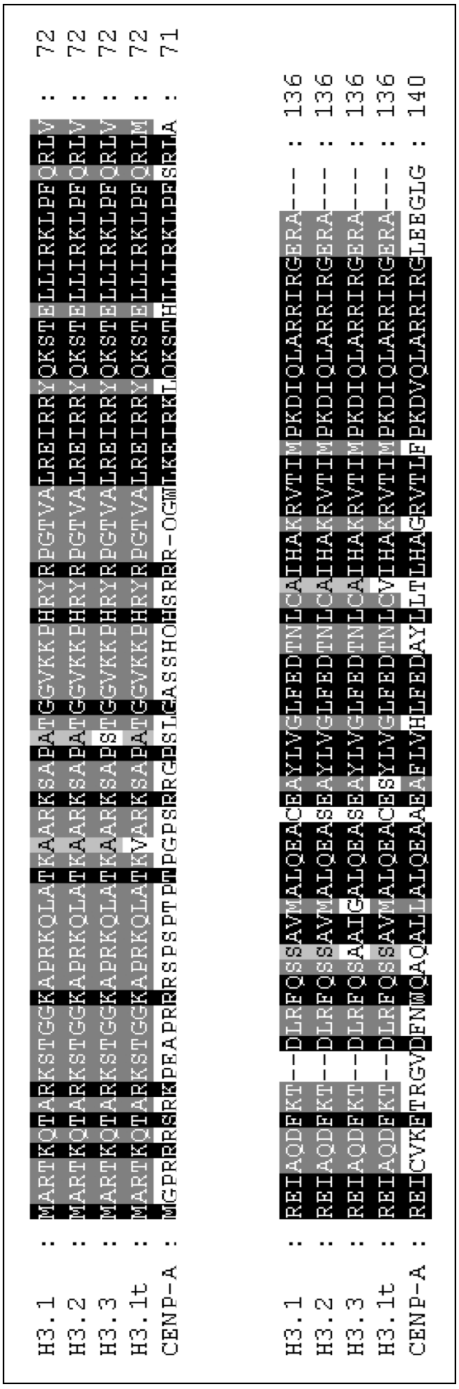


Figure 1-3 – Protein sequence alignment of human H3 variants

ClustalW was used to align the protein sequences of the human H3 variants. CENP-A is by far the most divergent among all the H3 variants.

Nucleosomes

Nucleosomes are formed by wrapping ~147 bp of DNA ~ 1.65 times around a cylindrical octameric core of histone proteins, containing two copies of H2A, H2B, H3 and H4, with the long amino-terminal tails of each histone extending out from the nucleosome (Figure 1-4). Within the core histone octamer two H3 / H4 histone pairs form a heterotetramer through a 4-helix bundle between the two H3 histones, whilst H2A / H2B heterodimers interact with the H3 / H4 tetramer through the formation of 4-helix bundles between H2B and H4 (Luger *et al*., 1997). Negatively charged DNA is wrapped around the nucleosome and bound primarily by electrostatic interactions between positively charged amino groups and the phosphodiester backbones as they face the histones, but also through a range of hydrogen bonds, salt bridges and nonpolar interactions. In the absence of DNA the *in vitro* stable state of core histones is a heterotetramer of H3 and H4 and heterodimers of H2A and H2B. Correspondingly, nucleosome assembly occurs first by a H3 / H4 tetramer binding and assembly onto ~ 60 bp of DNA followed by two dimers of H2A / H2B which contact ~ 30 bp each, with short 4 bp linkers between each histone pair to give an octameric nucleosome.

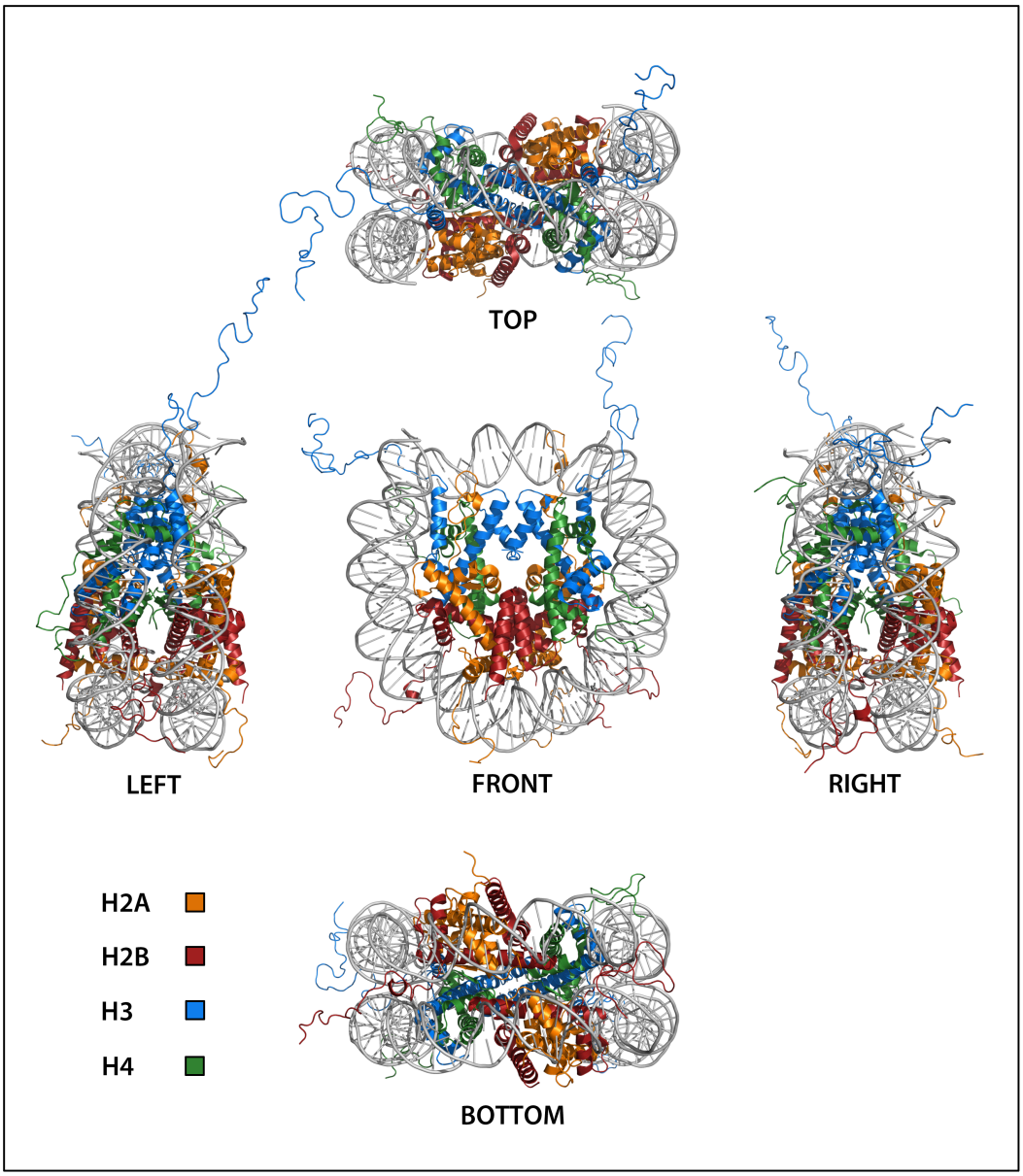


Figure 1-4 – Nucleosome structure

Illustration of the typical structure of a canonical H3 nucleosome from different angles. In a typical nucleosome roughly 147 bp of DNA (grey) is wrapped ~ 1.65 times around an octameric core of histone proteins, containing two copies of H2A (orange), H2B (red), H3 (blue) and H4 (green). Unstructured histone tails extend outwards from the nucleosome core further than is shown here but are typically not apparent in crystal structures. Image created using the crystal structure of the *Xenopus lavis* H3 nucleosome (Protein Data Bank ID 1KX5) published by Davey *et al* (2002).

Nucleosome sequence preference

DNA wrapping around the nucleosome requires extreme bending of DNA, particularly at ± 1.5 and ± 4 to 5 superhelical turns from the dyad axis where the DNA is pushed out by the interaction with the histones. This bending is achieved in part by local increases in DNA twist, particularly at the 14 sites where the minor groove faces the histone core (Figure 1-5). At these sites arginine side chains inserted into the minor groove stretch the DNA by narrowing the minor groove, increasing the twist from 10.5 bp per turn (for free DNA) to as little as 9.4 bp per turn. In concert with this, insertion of the H3 and H2B N-terminal tails through the minor groove facing away from the histone core locally decrease twist to a maximum of 10.9 bp per turn. The energy required to bend DNA, particularly into the minor groove, therefore favours A / T base pair dinucleotide sequences at these locations as they naturally adopt a relatively narrow minor groove and so allow for stronger salt links to form across the phosphate groove (Wu *et al*., 2010). Correspondingly, C / G dinucleotide sequences are favoured at sites where the minor groove faces away from the histone core due the required decrease in twist (Luger and Richmond, 1998b; Luger *et al*., 1997). The ability of a given DNA sequence to position a nucleosome therefore depends on the sum of the energetic costs from adapting the DNA conformation at each of these 14 contact sites around the nucleosome where the minor groove faces inwards (Luger and Richmond, 1998b). Thus, whilst nucleosomes are generally capable of forming on the majority of DNA sequences they do exhibit a so-called "sequence preference", whereby

nucleosomes will show preference for positioning on one sequence over another (Segal *et al*., 2006).

By comparing the positions of *S. cerevisiae* nucleosomes extracted from cells with those of chicken and fly nucleosomes assembled *in vitro* onto a range of DNA sequences from *S. cerevisiae* the sequence preference of canonical H3 nucleosomes (containing H2A/H2B/H3/H4) has been modelled and can be used to accurately predict *in vitro* nucleosome positioning on DNA sequence from a range of organisms (Kaplan *et al*., 2009; Segal *et al*., 2006). However, the extent to which the primary DNA sequence actually determines nucleosome organisation *in vivo* is less clear and contentious (Gkikopoulos *et al*., 2011; Kaplan *et al*., 2009; 2010; Zhang *et al*., 2010). The current consensus is that the probability that some part of a nucleosome will occupy a given sequence *in vivo* (the "occupancy" score) correlates positively with the overall content of C / G dinucleotides and negatively with the content of A / T dinucleotides, but the probability that nucleosomes occupying a given sequence will be precisely aligned with respect to their dyad axes (the "positioning" score) is best predicted by the periodicity of A / T and C / G dinucleotides (Kaplan *et al*., 2009; Locke *et al*., 2010). *In vivo,* nucleosome occupancy can be predicted from the DNA sequence with a high degree of accuracy across a number of genomes, but the exact positioning of these nucleosomes can only be predicted for approximately 20 % of well positioned nucleosomes (Locke *et al*., 2010; Zhang *et al*., 2009). One explanation for this difference is that nucleosome positioning is likely to be determined by other *in vivo* biological activities that move and process nucleosomes (such as the many chromatin remodelling

factors or transcription). Yet nucleosome occupancy arguably has more biologically relevant implications because it represents the extent to which a given DNA sequence (such as a transcription factor recognition site) will be occluded within a population and thus prevent binding of proteins such as transcription factors (Mai *et al*., 2000). A comparison of nucleosome occupancy at the promoters of respiration genes in aerobic and anaerobic yeast negatively correlated nucleosome occupancy with transcription (Field *et al*., 2009). In contrast, no observable function has yet been associated with nucleosome positioning, despite the characterisation of nucleosome positioning sequences at promoters (Ioshikhes *et al*., 2006). Accordingly, transient transfection into the chicken 6C2 cell line of the synthetic strong nucleosome positioning sequence "601" found the sequence quickly lost the ability to position nucleosomes (Gracey *et al*., 2010).

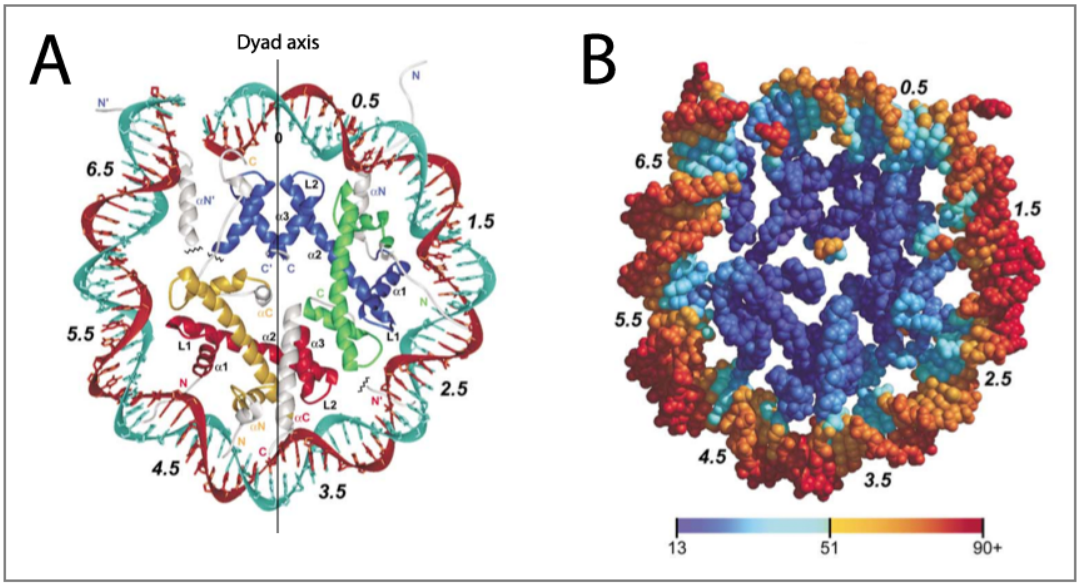


Figure 1-5 – Structure of DNA wrapping around a nucleosome

View down the superhelix axis to illustrate the wrapping of DNA around the histone core particle. For clarity, only one half of the nucleosome is shown. A) H3 (blue), H4 (green), H2A (yellow) and H2B (red), with histone fold extensions and segments of the tails coloured grey. The dyad axis runs through the central base pair, splitting the nucleosome in a two-fold symmetry with ~ 73 bp of DNA on either side. The superhelix locations (SHL) where DNA interacts directly with histone fold domains (and H3 alpha-N helix) are numbered from the dyad axis. B) The crystallographic B-factor (a measure of thermal motion) is mapped to the van der Waals surface of atoms and coloured in a scale from 13 to 90 + $Å^2$. Increased DNA twist at sites where the minor groove faces towards the histone core shows a relative reduction in the B-factor, and a corresponding increase as it faces outwards and twist is relatively reduced. Image adapted from Davey *et al* (2002).

Organisation of centromeric chromatin

During mitotic cell division the replicated genome must be faithfully segregated such that a complete copy of the genome is transferred to both daughter cells. Movement of the chromosomes during mitosis is accomplished through the attachment of each duplicated chromosome to dynamic microtubules originating from opposite poles of the spindle. In most eukaryotes the microtubules attach to a unique chromosomal locus called the centromere. A large multi-protein complex called the kinetochore is assembled on this site that captures and binds microtubules. This attachment is tightly regulated by the spindle assembly checkpoints, to correct errors in microtubule attachment that would otherwise result in chromosomal loss and / or gain during anaphase causing aneuploidy which can lead to cell death or contribute to oncogenesis (Weaver *et al*., 2007).

Centromeres are highly variable in both size and DNA sequence between species (Allshire and Karpen, 2008) (Figure 1-6), with the most simple the "point" centromere found in the budding yeast *Saccharomyces cerevisiae*, which is defined by a specific 125 bp DNA sequence that directs centromere activity. In contrast, higher eukaryotes have "regional" centromeres of various sizes depending on the species, which do not require a specific DNA sequence. Instead, complex centromeres appear to be epigenetically defined by the incorporation and propagation of the histone H3-variant CenH3 and the assembly of CenH3 nucleosomes at the chosen site (Allshire and Karpen, 2008; Earnshaw and Migeon, 1985). This begs the question, how are these sites for CenH3 selected? Although regional centromeres do not require a specific DNA sequence to attract CenH3, centromeric DNA from different

organisms do share common features. Regional centromeres are typically AT-rich relative to the rest of the genome. For example, *S. pombe* centromeric central domain regions are ~ 72 % A/T compared with ~ 65 % A/T across the whole genome, whilst *S. cerevisiae* centromeres are ~ 86 % A/T rich, compared with ~ 61 % across the whole genome. They are also gene-free, though in *D. melanogaster* and humans centromeres contain interspersed blocks of H3K4me2 (Sullivan and Karpen, 2004) and are transcribed to produce non-coding RNAs in a number of organisms (Carone *et al*., 2009; Choi *et al*., 2011; Wong *et al*., 2007b). Regional centromeres often contain repetitive elements (Sullivan *et al*., 2001). For example, in the fruit fly *D. melanogaster* the only centromere characterised at the DNA sequence level forms over a 420 kb region of repetitive sequences and transposable elements (Sun *et al*., 1997). Similarly, human centromeres form over 0.1 - 4 Mbp of repetitive α-satellite DNA that is interspersed with 17 bp binding sites for CENP-B, a centromere protein that promotes the incorporation of CENP-A nucleosomes (Allshire and Karpen, 2008; Okada *et al*., 2007; Schueler *et al*., 2001). The fission yeast *Schizosaccharomyces pombe* has regional centromeres similar to higher eukaryotes, but with just three chromosomes and defined centromeric DNA sequences it represents an experimentally useful model organism for the study of centromeres. Fission yeast centromeres span 40 - 100 Kb and are organised with heterochromatic inverted "outer" repeats surrounding a ~ 10 – 12 Kb central domain region that incorporates CENP-A^{Cnp1} nucleosomes and consists of a non-repetitive central core region flanked by inverted "inner most repeats" (Pidoux and Allshire, 2004).

Rearrangements of *Drosophila* and human chromosomes that result in the fusion of two chromosomes so that it contains two distinct arrays of centromere repeats only recruit centromere proteins to one centromere, indicating that one centromere has been inactivated (Agudo *et al*., 2000; Earnshaw and Migeon, 1985; Sullivan and Willard, 1998). The inactivation of a previously functional centromere, which occurs without change to the DNA sequence, demonstrates the flexible epigenetic nature of regional centromeres. Moreover, rearrangements that cause the complete deletion of a normal centromere can result in the formation of "neocentromeres" at conserved sites along the chromosome arrays in otherwise euchromatic regions. Importantly, neocentromeres recruit all known kinetochore proteins and thus form active and heritable centromeres without changes to the DNA, again demonstrating the epigenetic regulation of regional centromere location (Ishii *et al*., 2008; Ketel *et al*., 2009; Saffery *et al*., 2001; Voullaire *et al*., 1993).

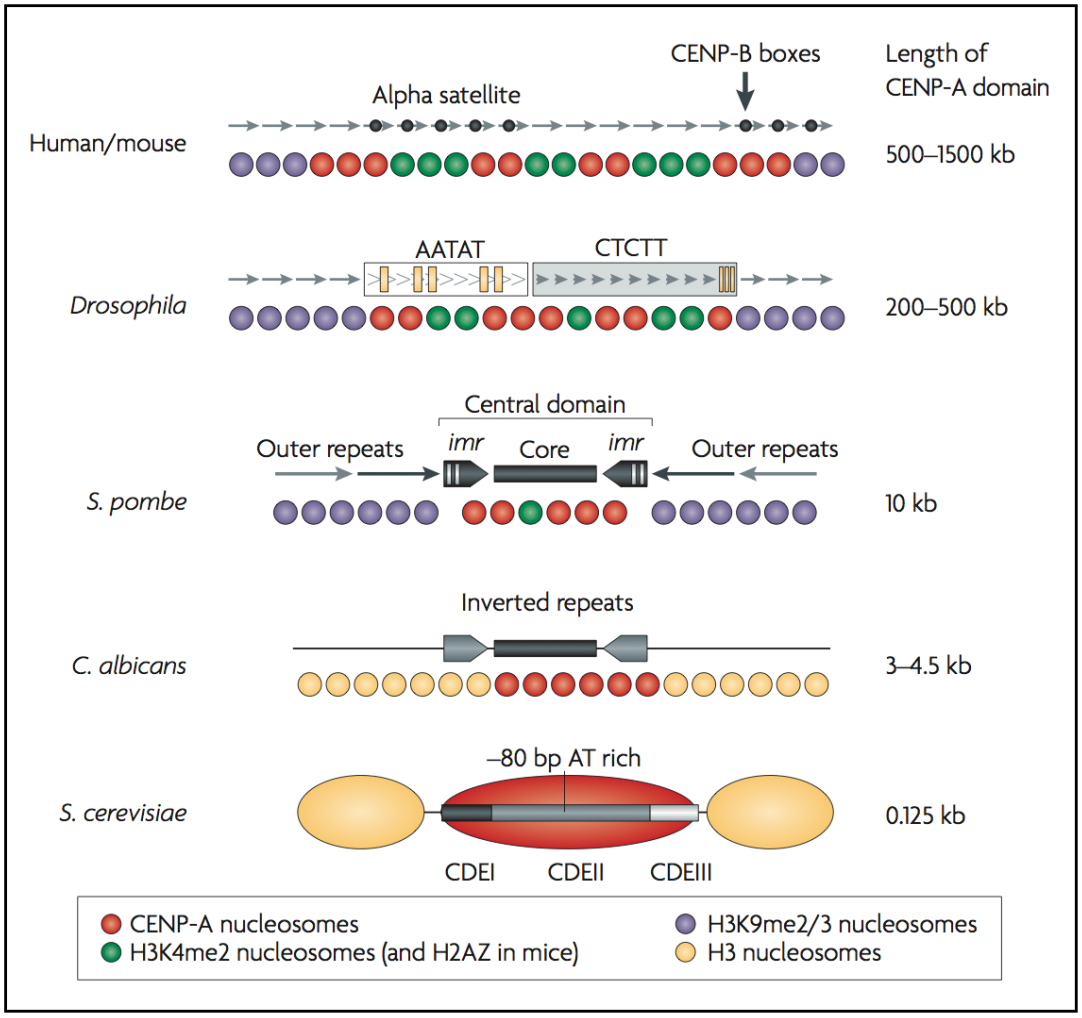


Figure 1-6 – Organisation of centromeres

Schematic illustrates the organisation of centromeric DNA in humans, mice, *Drosophila melanogaster*, *Schizosaccharomyces pombe*, *Candida albicans* and *Saccharomyces cerevisiae*. Figure adapted from Allshire and Karpen (2008).

Post-translational histone modifications

Nucleosomes play an essential role in compacting and regulating access to the underlying DNA, but the dynamic nature of the genome requires that nucleosomes themselves must be regulated to allow for changes in the chromatin state. How nucleosomes interact with one another, with a large range of non-histone chromatin proteins and with other interacting proteins, alters the accessibility of the underlying DNA and is regulated by the post-translational modification of histones (Ferreira *et al*., 2007; Li *et al*., 1993). These modifications are most often on the accessible amino-terminal histone tails that extend from the nucleosome cores and confer distinct structural properties to the nucleosome that can generate synergistic or antagonistic interactions with a host of other proteins. For example, *in vitro* nucleosomes carrying the H4K20me3 modification require less Mg^{2+} to condense than unmodified nucleosomes (Lu *et al*., 2008), consistent with a role for H4K20me3 in contributing to a locally condensed chromatin state. Furthermore, modifications may act in combination to define a particular chromatin state. H3S10 phosphorylation in budding yeast and *Tetrahymena* facilitates the acetylation of H3K14 via GCN5 (Lo *et al*., 2000), whilst H3S10 phosphorylation and H3K9 methylation are inhibitory towards one another *in vitro* (Rea *et al*., 2000).

Euchromatic regions typically contain a number of modifications associated with active transcription, including H3K4/K36 methylation, and H3K9/14 and H4K5 acetylation, not necessarily on the same nucleosome (Jenuwein and Allis, 2001). Similarly, transcriptionally silent heterochromatin regions are methylated at H3K9/K27 and H4K20 (Martin and Zhang, 2005;

Rosenfeld *et al*., 2009), and in budding yeast are hypoacetylated at all tested lysines (Suka *et al*., 2001). In *S. pombe*, silencing at peri-centromeric and mating type heterochromatin is dependent on the recruitment of Swi6 and Chp2 (the fission yeast homologues of HP1) (Allshire *et al*., 1995), which associate with methylated H3K9 via chromodomains (Bannister *et al*., 2001).

After DNA replication, histone modifications must be re-established on the newly deposited histones to ensure faithful propagation of the chromatin state. Targeting and spreading of chromatin-modifying complexes may occur both *de novo* after each cell cycle or by recognition of existing modifications on "old" histones. In cases where the recognised modification is the same as the deposited modification, a positive feedback loop is established that is able to epigenetically maintain a particular chromatin state through multiple cell cycles (Bonasio *et al*., 2010). Propagation of the heterochromatic H3K9 methylation modification is a well characterised example of a positive feedback loop, whereby the modification is bound by the heterochromatin protein HP1 (responsible for the stochastic silencing of nearby genes observed in *S. pombe* and *D. melanogaster*) which in turn recruits the histone methyltransferases SUV39H1/2, spreading the methyl-H3K9 mark to surrounding nucleosomes. However, the methyl-H3K9 mark alone is not always sufficient to establish this feedback loop. Notably, although the inactive X chromosome has elevated methyl-H3K9 marks it lacks HP1 (Heard *et al*., 2001). It is possible that distinct methyl-H3K9 binding proteins, or Xist RNA and the factors that it recruits exclude HP1 (Maison and Almouzni-Pettinotti, 2004). During the initial stages of X chromosome inactivation methylated H3K27 spreads along the chromosome. The

contribution of methyl-H3K27 to silencing at inactive X chromosomes is unknown, but methyl-H3K27 is also active in Polycomb silencing, where it recruits the the repressive PRC1 complex through an interaction with the chromodomain-containing Polycomb protein (Martin and Zhang, 2005). It is well established that histone post-translational modifications can have effects both in isolation and combinatorially. Thus there are a range of possible epigenetic chromatin "read-outs" and this has lead to the proposal that post-translational modifications may provide a layer of information about the chromatin state in a "histone code" (Jenuwein and Allis, 2001; Strahl and Allis, 2000).

Histone variants

Conserved variants for each of the core histones have been identified with the notable exception of H4, providing increased diversity in nucleosome types (Kamakaka and Biggins, 2005). Histone variants closely resemble their canonical counterparts, for example, canonical H3.1 and variant H3.3 are ~ 96 % identical in mammals (Figure 1-3). These differences confer altered structural and biochemical properties to nucleosomes that contain them (Kamakaka and Biggins, 2005). Unlike canonical histones, histone variants may be expressed outside S phase, and are incorporated into chromatin in a replication-independent manner (Loyola and Almouzni-Pettinotti, 2007).

Nucleosomes composed of the canonical H2A, H2B, H3 and H4 histones are the most widespread across the genome. In contrast, nucleosomes containing histone variants have specialised functions resulting in their incorporation at more specific regions. H3.3, for example, is incorporated

specifically by the HIRA histone chaperone and accumulates at actively transcribed genes (Ahmad and Henikoff, 2002; Ray-Gallet *et al*., 2011; Tagami *et al*., 2004). In keeping with a role promoting transcription, H3.3 is proposed to destabilise nucleosomes (Jin and Felsenfeld, 2007), although the structural basis for this is unclear (Tachiwana *et al*., 2011b) and instability may instead result from being relatively more acetylated than H3.1 (Hake *et al*., 2006; Johnson *et al*., 2004; McKittrick *et al*., 2004; Waterborg, 1990). Intriguingly, whilst H3.1 can hold modifications indicative of both transcriptionally active and repressed chromatin (Loyola *et al*., 2006), human H3.2 is relatively enriched in repressive H3K27me3 marks compared with H3.1 (Hake *et al*., 2006), indicating a potential tendency for specific variant histones to gain particular modifications. Nucleosomes containing H3.1, H3.2 or H3.3 have no discernible structural differences (Tachiwana *et al*., 2011b) so specificity for these modifications would appear to stem from variant-specific chaperone mediated interactions. For example, H3.1 interacts with CAF, whilst H3.3 specifically interacts with HIRA and CENP-A with HJURP (Ray-Gallet *et al*., 2011). However, it also remains possible that the variant-specific enrichment of modifications is simply a consequence of the relative enrichment of these variants at actively transcribed or repressed regions.

Interestingly, the canonical H2A and H3 from budding yeast resemble the variants H2A.X and H3.3 of higher eukaryotes respectively (Kamakaka and Biggins, 2005). However, it is not known how closely the characteristics of yeast canonical H3 nucleosomes match with those containing H3.3 from higher eukaryotes. Certainly, the replication-independent incorporation of H3.3 is not observed for canonical yeast H3 (Loyola and Almouzni-Pettinotti,

2007). Whilst H2A.X and H3.3 have yet to be observed in the same nucleosome, nucleosomes composed of their apparent counterparts from budding yeast histones can be assembled and have been crystallised (White *et al*., 2001). It has been shown that *Drosophila* H2A.Z and H3.3 can be incorporated into the same nucleosome *in vitro* (Thakar *et al*., 2009), raising the possibility that different variant histones can combine within single nucleosomes to impart a spectrum of properties and functions.

Histone chaperones

Newly synthesised histones are highly basic, therefore, in order to prevent promiscuous interactions and aggregation they are bound and escorted by chaperone proteins (Woodland and Adamson, 1977). These chaperone proteins dissemble from histones during their deposition onto DNA. In addition, they are required whenever histones are removed and thus are essential to all processes involving nucleosome assembly and disassembly such as during transcription, replication and DNA repair (De Koning *et al*., 2007). Many chaperones appear capable of multiple roles, for example, the H3 - H4 chaperone Asf1 can directly aid the assembly of nucleosomes *in vitro* (Donham *et al*., 2011), but in *X. laevis* egg extracts it appears to act instead as an escort chaperone that delivers H3 - H4 to the HIRA chaperone for deposition (Ray-Gallet *et al*., 2007). Such differences highlight the difficulties in assigning specific functions to particular histone chaperones. Nevertheless, chaperones may be broadly grouped into three functional categories, those that act alone (such as Asf1), those that act in multi-chaperone complexes (such as the CAF-1 complex) and those that act

in complex with enzymatic activity (such as Arp-4 in the INO80 nucleosome remodelling complex) (De Koning *et al*., 2007). In this respect, RbAp48 provides a notable exception as, in different complexes it is capable of fitting into all three functional categories, further highlighting the multiple roles that chaperones can play in histone dynamics. Whilst little is known of how chaperones switch between their multiple roles there are hints that post-translational modifications of the chaperones themselves plays some part. For example, the FACT subunit SSRP1 is selectively phosphorylated and NAP2 dephosphorylated in response to DNA damage, promoting their interaction with histone amino-terminal tails *in vitro* (Dirksen *et al*., 2006). Different chaperones may also engage the same histone complex in different ways and thereby define different downstream interaction networks. For example, both Vps75 and Asf1 can bind the H3 - H4 complex, with both leading to interactions with the histone acetyltransferase Rtt109. Yet Vps75 and Asf1 present the H3 - H4 complex differently such that Vps75 - Rtt109 acetylates H3K9 and H3K27 whilst Asf1 - Rtt109 acetlyates H3K56 (D'Arcy and Luger, 2011).

Typically chaperones show preferences (though not exclusivity) for interaction with either H3 - H4 or H2A - H2B complexes. Moreover, some chaperones are highly selective for particular histone variants. A good example of this is the specificity for CAF-1 in the replication-dependent deposition of H3.1 and the transcription-coupled replication-independent deposition of H3.3 by HIRA (Tagami *et al*., 2004). How these specificities for particular histones are structurally defined is only beginning to be determined (Su *et al*., 2012). It also remains to be determined how well such

specificity is conserved between homologous chaperones in different model organisms (De Koning *et al*., 2007).

The centromere specific histone variant CenH3

Conservation

Centromere protein A, or CENP-A, was originally identified as a centromere-specific protein that was targeted by human autoimmune antibodies (Earnshaw and Rothfield, 1985). Since then, centromere-specific H3 variants have been found to be essential for centromere function in all eukaryotes but confusion arose as this protein was named differently in each species. For example, in *S. cerevisiae* Cse4 was identified in genetic screens (Stoler *et al*., 1995), whilst Cnp1 in *S. pombe* and CID in *D. melanogaster* were identified by homology to CENP-A and Cse4 (Henikoff *et al*., 2000; Takahashi *et al*., 2000). To clarify these names the consensus is to refer to specific versions in superscript, such as *S. pombe* CENP-A^{Cnp1}. More recently the term CenH3 is also used when referring to all CENP-A homologues to avoid confusion with the human CENP-A.

The histone H3-variant CenH3 is exclusively found at active centromeres and often described as the most attractive example of a *bona fide* epigenetic mark that defines a specific chromatin state and domain, in this case centromere location. CenH3 is conserved in all eukaryotes and appears to epigenetically regulate the position of centromeres in all species, with the exception of budding yeast where centromeres are defined by DNA sequence

(Camahort *et al*., 2009; Conde e Silva *et al*., 2007) and perhaps also in *C. elegans*, where CENP-A is assembled along the entire length of the chromosome in a pattern that appears epigenetically linked with transcription (Gassmann *et al*., 2012). Used generically, CenH3 refers here to CENP-A in mammals, CID in flies, Cse4 in budding yeast and Cnp1 in fission yeast (Talbert *et al*., 2012). CenH3 is the most divergent of the H3 variants. Inter-species conservation is limited to the histone fold domain, with essentially no sequence conservation in the amino-terminal tail that varies in length between 20 and 200 amino acids between different species (Figure 1-7). The rapid divergence of CenH3 has been proposed as an adaptation to maintaining the ability to bind rapidly evolving centromeric sequences (Henikoff *et al*., 2001), which is consistent with predictions that the A/T-rich sequences that are present at centromere in many species are unfavourable for canonical H3 nucleosome formation (Segal *et al*., 2006). Despite this divergence CenH3 function appears to be well conserved, with CENP-A^{Cse4} from budding yeast being able to functionally replace CENP-A in RNAi-depleted human cells (Wieland *et al*., 2004). Interestingly, using chimeric versions of human H3 / CENP- A it was found that just the 6 carboxyl-terminal amino acids of CENP-A are capable of assembling functional kinetochores *in vitro*, apparently through the recruitment of CENP-C (Guse *et al*., 2011). Oddly, these residues are not conserved and it is currently unclear whether these amino acids are recognised directly or else impart some greater structural distinction to the nucleosome.

The epigenetic nature of CenH3 is best observed when loss of a centromere is accompanied by activation of a new "neocentromere" at a

previously non-centromeric location on the same chromosome. In humans, neocentromeres do not share homology with centromeric DNA sequences but are fully functional and maintained at the same location, even through meiosis, without changes to the DNA sequence (Amor *et al*., 2004; Warburton *et al*., 1997). If primary DNA sequence is not implicated, how can centromere location be propagated? It is thought that pre-existing CenH3 can guide the deposition of newly synthesised CenH3 and ensure the recruitment of kinetochore components to the same location in subsequent cell cycles. It is likely that a feedback loop operates in which the kinetochore associated proteins direct the incorporation of new CenH3 at that site to allow self-propagation (Okada *et al*., 2009). In human cells, CENP-A binds directly to CENP-C and CENP-N, which in turn recruit CENP-T, -M, -N, -U and -H to form the constitutive CENP-A nucleosome-associated complex (NAC) (Figure 1-8). The NAC directly mediates the interaction between the CENP-A nucleosome and more distal kinetochore sub-complexes (CAD). Both the NAC and CAD complexes are stably localised to centromeres throughout the cell cycle and are required for proper mitotic progression. Collectively the NAC and CAD complexes are termed the constitutive centromere-associated network (CCAN) (Carroll *et al*., 2009; Foltz *et al*., 2006; Guse *et al*., 2011).

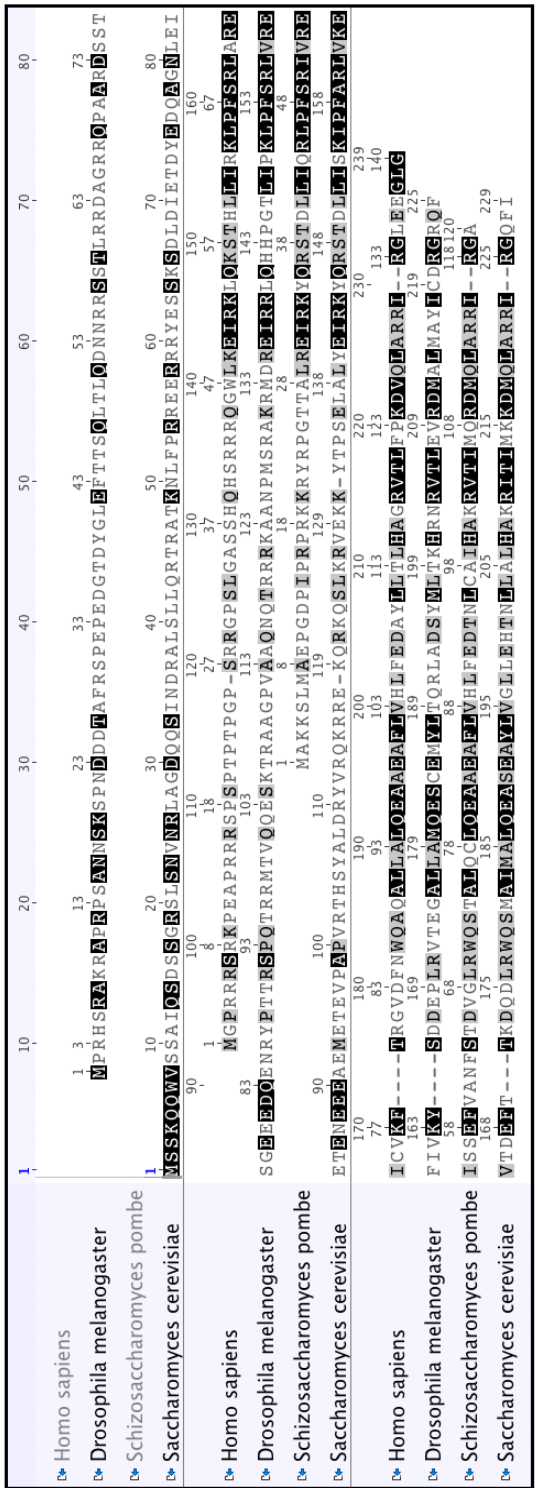


Figure 1-7 – CenH3 protein sequence alignment

Alignment of CenH3 protein sequences from humans (CENP-A), *Drosophila* (CID), *S. pombe* (Cnp1) and *S. cerevisiae* (Cse4) was made with the software package Geneious using the in-built alignment function. Amino acids are shaded according to the degree of similarity to the consensus.

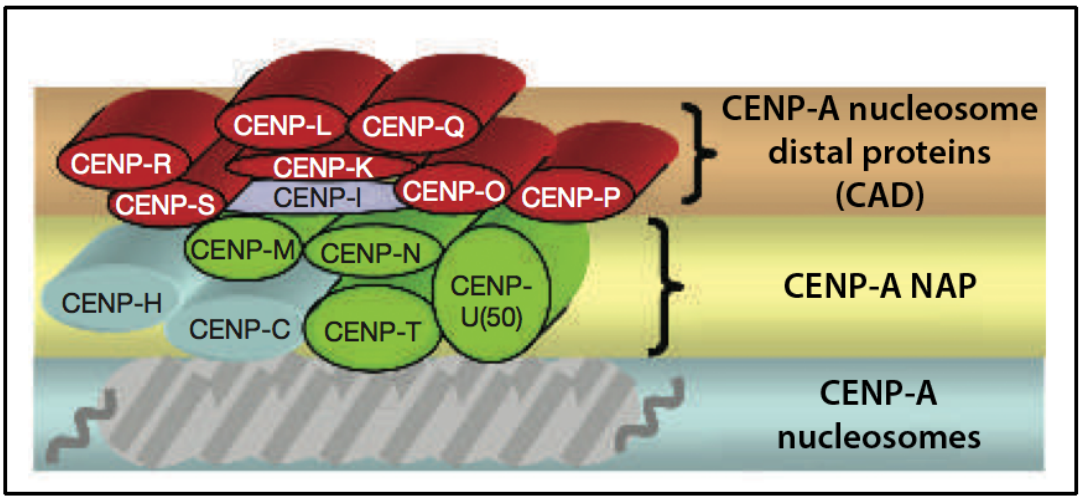


Figure 1-8 – Constitutive centromere-associated network (CCAN) factors in humans

Human CENP-A nucleosomes are associated with a large set of proteins that remain present at centromeres throughout the cell cycle. The CENP-A nucleosome associated complex (NAC) directly mediates the interaction between CENP-A nucleosomes and CENP-I and the other CENP-A distal component (CAD) complexes. Collectively these complexes are termed the constitutive centromere-associated network (CCAN). Figure adapted from Foltz *et al* (2006).

CenH3 deposition into chromatin

Initially, all newly synthesised histones were thought to be incorporated into the genome as it is synthesised during S phase (Wu and Bonner, 1981). In *S. pombe*, CENP-A^{Cnp1} is expressed early in S phase (Takayama *et al*., 2008) and centromeric DNA also replicates in very early S phase (Kim *et al*., 2003). In humans, CENP-A is up-regulated during S phase but reaches a maximum during G2 (Shelby *et al*., 1997) and thus centromeres were thought to replicate late in S phase. However, careful analysis later showed that human centromeres replicate asymmetrically and before CENP-A incorporation

(Shelby *et al*., 2000). The use of informative pulse-chase SNAP-tag assays identified that CENP-A assembly in human cells actually requires progression through mitosis, occurring at telophase and throughout G1 (Jansen *et al*., 2007). In *S. pombe*, whilst cell-cycle arrested cells can deposit additionally expressed CENP-A^{Cnp1} during S phase (Dunleavy *et al*., 2007; Takayama *et al*., 2008) careful quantification of centromere-associated CENP-A^{Cnp1} using PALM suggests that the majority of CENP-A^{Cnp1} is loaded during G2 (Lando *et al*., 2012). Thus the incorporation of newly synthesised CenH3 is replication-independent.

The human Mis18 complex of Mis18∝, Mis18β and Mis18BP1 (also known as hsKLN2) (Maddox *et al*., 2007) dissociates from centromeres during G1 and does not return until late anaphase / telophase, just as new CENP-A is incorporated (Jansen *et al*., 2007). In *S. pombe* both Mis18 and Mis16 (a homologue of human RbAp48) exhibit dynamic localisations but only dissociate for a brief period from mitotic prophase to mid/late anaphase, which is similar to Scm3 dynamics (Pidoux *et al*., 2009; Williams *et al*., 2009). The return of Mis18/Scm3 in *S. pombe* is coincident with the time when CENP-A^{Cnp1} can be incorporated (Dunleavy *et al*., 2007; Takayama *et al*., 2008). The Mis18 complex, together with interactive partners RbAp46/48 are required for CENP-A / CENP-A^{Cnp1} incorporation (Fujita *et al*., 2007; Hayashi *et al*., 2004). In human cells, siRNA knockdown assays showed that the Mis18 complex is required for HJURP targeting to chromatin and subsequent CENP-A incorporation (Barnhart *et al*., 2011). It was proposed that the Mis18 complex, together with RbAp46/48 in higher eukaryotes, prime centromeres for the incorporation of new CenH3 (Fujita *et al*., 2007; Jansen *et al*., 2007;

Maddox *et al*., 2007). Human RbAp46 is known to bind the histone acetyltransferase Hat1 (Verreault *et al*., 1996; 1998), and global inhibition of histone deactylases (HDACs) with trichostatin A partially rescues the loss of CENP-A associated with RNAi knockdown of Mis18∝, although this might be an indirect effect (Ma *et al*., 2008). Thus acetylation has been implicated in "licensing" the centromeres for new CENP-A incorporation. It also appears that, once incorporated, the deposition of new CenH3 may be "approved" by a GTPase switch driven by MgcRacGAP (Lagana *et al*., 2010). Using a SNAP-tag pulse-labelling method in human cells has observed that the new pool of CENP-A is specifically absent from centromeres following loss of MgcRacGAP function. It is proposed that such a surveillance mechanism may help to prevent ectopic incorporation of CENP-A, although the nature of this mark of approval has yet to be determined.

CenH3 chaperones

As CenH3 is incorporated only at centromeres specific chaperones are required to ensure that it is delivered to centromeres and to prevent its ectopic incorporation. In several cell types this system can be overwhelmed when CenH3 is over-expressed, resulting in its promiscuous incorporation (Heun *et al*., 2006; Hewawasam *et al*., 2010; Ranjitkar *et al*., 2010; Tomonaga *et al*., 2003; Van Hooser *et al*., 2001). Notably CENP-A is over-expressed in tumour cells and may contribute to their genomic instability by inappropriate incorporation (Tomonaga *et al*., 2003). In fission yeast, newly synthesised but over-expressed CENP-A^{Cnp1} is assembled at centromeres both in S phase and G2 (Lando *et al*., 2012; Pidoux *et al*., 2009; Takahashi *et al*.,

2005). In a search for mutants that disrupt central domain (CENP-A^{Cnp1}) chromatin the *sim1* and *sim3* mutants were identified (Pidoux *et al*., 2003). *Sim3*$^{+}$ encodes a protein that shares homology with human NASP and *X. laevis* N1/N2 (Dunleavy *et al*., 2007). The NASP / N1/N2 chaperones are involved in storing H3/H4 in oocytes and NASP also associates with H3.3 (Tagami *et al*., 2004). Since Sim3 is not specifically localised at centromeres it seems likely that it acts as an escort chaperone rather than a chromatin assembly factor. It has also been shown to associate with H3, so it is possible that it acts as an exchange factor, depositing CENP-A^{Cnp1} and collecting released H3 to ensure it is taken away (Dunleavy *et al*., 2007). The *sim1*$^{+}$ gene (later renamed *scm3*$^{+}$) encodes the *S. pombe* ortholog of *S. cerevisiae* Scm3 (Pidoux *et al*., 2009), which was first identified as a high-copy suppressor of a CENP-A^{Cse4} mutant (Chen *et al*., 2000). Scm3 directly associates with CenH3 in fission and budding yeasts and is required for CenH3 incorporation (Camahort *et al*., 2007; Mizuguchi *et al*., 2007; Pidoux *et al*., 2009; Stoler *et al*., 2007; Williams *et al*., 2009). Importantly, fission yeast Scm3 is localised to centromeres independently of CENP-A^{Cnp1} (Pidoux *et al*., 2009; Williams *et al*., 2009). Because fission yeast Scm3 is released from centromeres in prophase and reassociates in mid- to late anaphase it was proposed to be a CENP-A^{Cnp1} chaperone (Pidoux *et al*., 2009; Williams *et al*., 2009). In contrast, budding yeast Scm3 is stably associated with centromere and CENP-A^{Cse4} throughout the cell cycle, although it is also possible that it is released briefly (Mizuguchi *et al*., 2007; Pidoux *et al*., 2009). Uniquely, budding yeast Scm3 has been proposed to replace H2A / H2B as a component of an unconventional hexameric CENP-A^{Cse4} nucleosome (Mizuguchi *et al*., 2007; 2011), although

this remains contentious and the nature of CENP-A^{Cse4} nucleosome is still a matter of debate (Camahort *et al*., 2009; Shivaraju *et al*., 2011). Careful protein alignments have shown that Scm3 shares homology with a domain in the essential human CENP-A-specific chaperone HJURP (Sanchez-Pulido *et al*., 2009). Furthermore, structural analyses have shown that both Scm3 and HJURP can bind to prenucleosomal CenH3 in a trimeric complex with H4 (Cho and Harrison, 2011; Dechassa *et al*., 2011; Zhou *et al*., 2011), although the exact conformation of binding, and the residues involved are unclear.

Organisation of CenH3 at centromeres

Extended chromatin fibres show the centromeres of humans, *Drosophila*, mice and rice consist of long interspersed blocks of H3 and CenH3 nucleosomes, yet during mitosis centromeric chromatin appears to fold such that CenH3 is presented on the surface of the chromosome and H3 resides towards the centre (Figure 1-9) (Allshire and Karpen, 2008; Blower *et al*., 2002; Ribeiro *et al*., 2010). In humans, CENP-A only localises to a portion of the available centromeric alpha-satellite repeat arrays, which can span 100 – 5000 kb (Warburton *et al*., 1997). Thus, calculating the number of CenH3 nucleosomes at these regional centromeres has long been an outstanding question that has proven difficult to answer. The point centromeres of budding yeast, defined by a single 125 bp DNA sequence and binding just a single microtubule, contain a single CENP-A^{Cse4} nucleosome (Furuyama and Biggins, 2007). However, it remains debatable if this "one CenH3 nucleosome, one microtubule" rule applies to the much larger regional centromeres found in many eukaryotes. The fluorescence intensity

of CENP-A^{Cse4} nucleosomes from the 16 clustered centromeres of budding yeast has been used as a reference point to estimate the number of CenH3 nucleosomes at centromeres in other species (Joglekar *et al*., 2008). Comparison of the ratio of fluorescence per nucleosome of CENP-A^{Cse4}-GFP to the level of fluorescence in *S. pombe* centromeres suggested that just 2 – 3 CENP-A^{Cnp1} nucleosomes were present at each fission yeast centromere. It had previously been shown that each of the 3 kinetochores are contacted by 2 – 4 microtubules in *S. pombe* (Ding *et al*., 1993). Thus this estimated number of CENP-A^{Cnp1} nucleosomes is consistent with the idea that each CenH3 nucleosome specifies one microtubule attachment site. However, more recently the estimated number of CENP-A^{Cnp1} nucleosomes has been reassessed by different approaches (Coffman *et al*., 2011; Lando *et al*., 2012; Lawrimore *et al*., 2011). The original study proved incorrect and the authors revised their estimate to 15 copies per centromere (Lawrimore *et al*., 2011). In contrast a second study has proposed that there are in total ~680 and ~ 122 CenH3-GFP nucleosomes at the *S. pombe* and *S. cerevisiae* centromere clusters, respectively (~226 and ~7 CenH3 nucleosomes per centromere, respectively) (Coffman *et al*., 2011). In a very different approach, photo-activated localisation microscopy (PALM) was used to directly "count" the number of mEos2-CENP-A^{Cnp1} signals at the centromere cluster by iterative photobleaching. These analyses suggest that there are 10 – 20 CENP-A^{Cnp1} nucleosomes per centromere (Lando *et al*., 2012). It therefore appears that the number of CenH3 molecules at fission yeast centromeres exceeds the number of microtubules at each centromere. Thus CenH3 nucleosomes do not necessarily equate with microtubule attachment sites in either yeast. This is

consistent with observations that fission yeast centromeres remain functional despite a 2.7-fold reduction in CENP-A^{Cnp1} at centromeres (Strålfors *et al*., 2011). In human cells, normal kinetochore assembly also appears to proceed even with only 10 % of the normal level of CENP-A (Liu *et al*., 2006). But if not all CenH3 nucleosomes create microtubule attachment sites, what distinguishes those CenH3 nucleosomes that do? One possibility is the differential post-translational modification of CenH3 nucleosomes, or alternatively they may simply be selected stochastically.

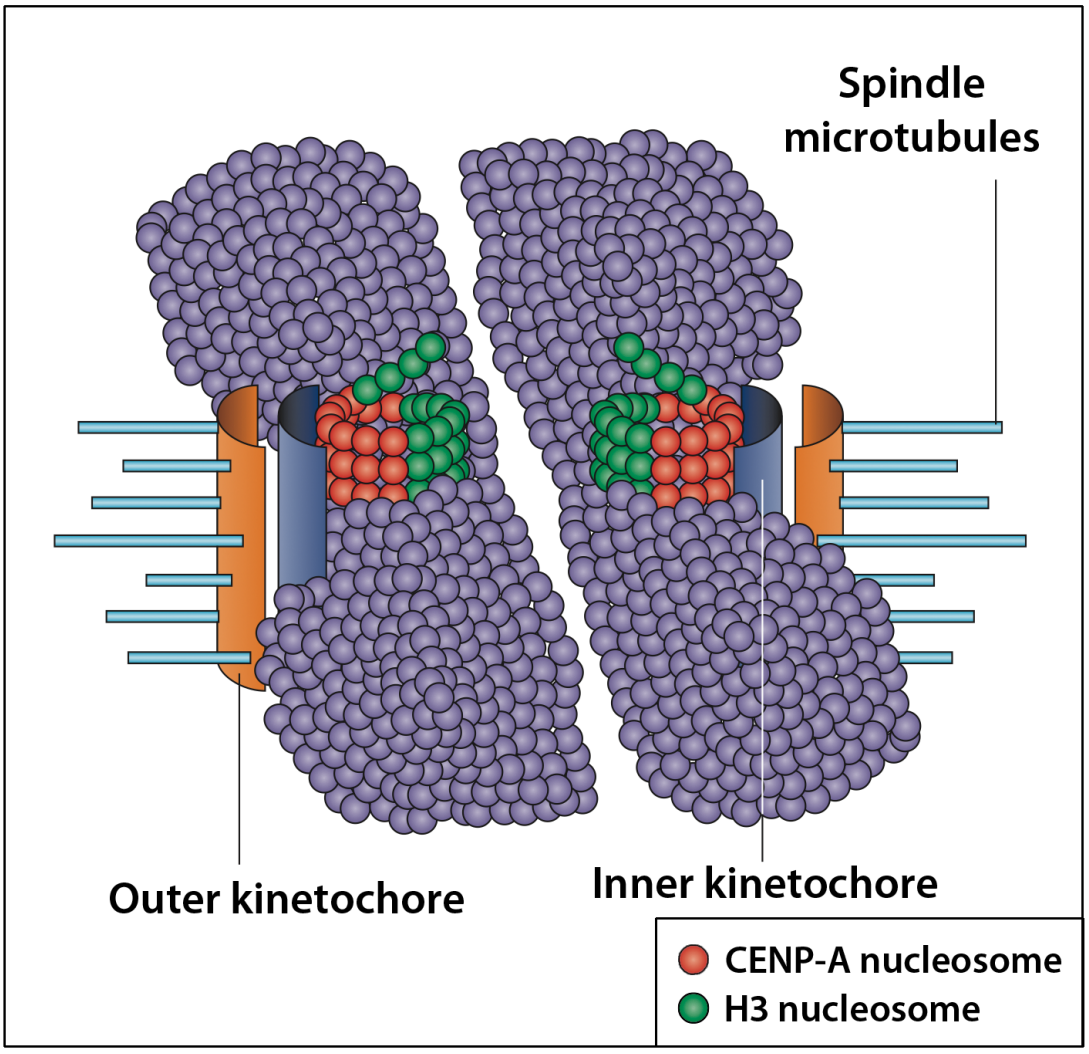


Figure 1-9 – Model of the three-dimensional structure of mitotic centromeres

Centromeric chromatin from Mammalian and *Drosophila* cells during mitosis appear organised such that H3 nucleosomes are located internally to CenH3 nucleosomes, which are presented on the chromosome surface to enable kinetochore assembly and the attachment of spindle microtubules. Figure adapted from Allshire and Karpen (2008).

Post-translational modifications of CenH3

The amino-terminal tail of CenH3 is highly divergent between different species and lacks many of the residues typically modified in H3 and other H3 variants, including K4, K9, K36, S10 and T3 (Figure 1-7) (Kouzarides, 2007). Relatively little is known about post-translational modifications of CenH3. It has recently been shown that budding yeast CENP-A^{Cse4} can be methylated on arginine 37 (R37) and this has been linked with the recruitment of specific kinetochore components (Samel *et al*., 2012). The R37 residue is not conserved in CenH3 proteins and cells with R37 mutated to alanine (R37A) are viable, demonstrating that R37 modification is not essential. However, when the R37A mutation is combined with mutations in kinetochore components it exhibits synthetic lethality (Samel *et al*., 2012). Thus R37 is implicit for kinetochore integrity.

In human cells, CENP-A was found to interact with PARP-1, which poly(ADP-ribosyl)ates CENP-A in response to DNA damage (Saxena *et al*., 2002). The strong negative charge of the ADP-ribose polymers is thought to prevent poly(ADP-ribosyl)ated proteins binding DNA and so may function to decompact chromatin surrounding the site of DNA damage (de Murcia *et al*., 1986). The poly(ADP-ribosyl)ation site has yet to be identified in CENP-A, however, like histone H3 it contains a consensus motif for poly(ADP-ribosyl)ation within the histone fold domain, suggesting that they both respond to DNA damage-induced poly(ADP-ribosyl)ation in the same manner (Saxena *et al*., 2002).

Human CENP-A and H3 do share similar phosphorlyation sites in their amino-terminal tails, serine 7 of CENP-A (CENP-A S7) and serine 10 of H3

(H3 S10). Interestingly, phosphorylation of H3 and CENP-A at these similar sites are temporally separated and appear to have different functions. H3 S10 phosphorylation is apparent from late G2 phase / early prophase and throughout mitosis whilst CENP-A S7 phosphorylation is initiated by Aurora-B and Aurora-C and is observed between prophase and mid-anaphase (Slattery *et al*., 2008; Zeitlin *et al*., 2001a). One consequence of H3 S10 phosphorylation is the removal of repressive HP1 protein from H3K9 methylated nucleosomes at pericentromeric regions (Hirota *et al*., 2005) however, the role of CENP-A S7 phosphorylation is less clear. The expression of dominant negative CENP-A S7 mutants does not have mitotic defects but instead a delay in cytokinesis with accompanying defects in the localisation of Aurora B, INCENP, and PP1γ1 (Zeitlin *et al*., 2001b).

Thus, known post-translational modifications of CenH3 are currently limited to human and budding yeast systems, and appear to be neither essential nor conserved.

Structural features of CenH3 nucleosomes

Prenucleosomal CenH3 is specifically bound by the HJURP / Scm3 chaperone through recognition of the CENP-A targeting domain (CATD) region, which in human CENP-A is made of 22 discontinuous amino acids between loop 1 and $\propto$2-helix that, when substituted into H3, are capable of targeting H3 to the centromere (Black *et al*., 2004; 2007b). Whilst it remains unclear precisely how HJURP / Scm3 binds CENP-A and H4 (Cho and Harrison, 2011; Dechassa *et al*., 2011; Zhou *et al*., 2011) the CATD region as a whole does appear to define a more rigid and compact CENP-A / H4

tetramer than is observed for H3 / H4 tetramers (Black *et al*., 2004; 2007b; Sekulic *et al*., 2010). However, the recent crystal structure for a CENP-A nucleosome, and analytical ultracentrifugation assays both suggest the hydrodynamic radius for CENP-A nucleosomes is essentially identical to that of H3 nucleosomes, thus a compact CENP-A / H4 tetramer structure appears to be specific to pre-nucleosomal CENP-A. (Panchenko *et al*., 2011; Tachiwana *et al*., 2011a). Deuterium exchange experiments comparing human H3 and CENP-A nucleosomes *in vitro* suggested that the CENP-A / H4 tetramer forms a more rigid structure within the nucleosome, although this is open to interpretation as the reduced exchange observed in CENP-A / H4 tetramers might also be caused by increased intra- and/or intermolecular hydrogen bonding of amide protons (Black *et al*., 2007a). Thus the conclusions regarding rigidity remains to be confirmed by other approaches.

Models of CenH3 nucleosome composition

There has been significant debate about what are the distinguishing features of CenH3 nucleosomes relative to H3 nucleosomes. Differences in opinion stem principally from seemingly conflicting data from a range of model organisms, consequently a number of different models for the composition of CenH3 nucleosomes and their stoichiometry have been proposed (Figure 1-10).

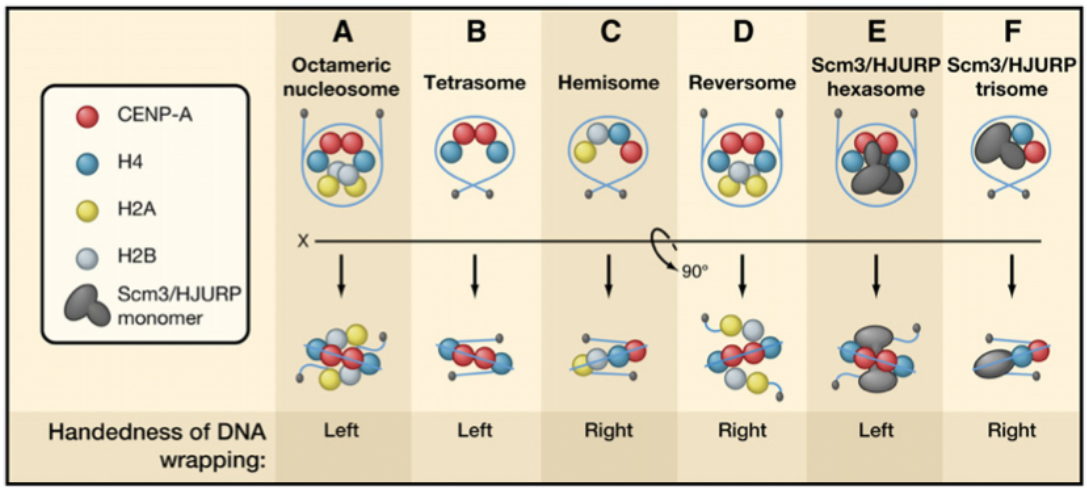


Figure 1-10 – CenH3 nucleosome models

Representation of current models for CenH3 nucleosome stoichiometry, components and DNA wrapping direction. Handedness of DNA wrapping corresponds to the induction of positive supercoiling for right handed wrapping and negative supercoiling for left handed wrapping. Adapted from Black *et al* (2011).

Octameric model

The simplest model for a CenH3 particle is that of a conventional octameric nucleosome (Figure 1-10 A), containing two copies of H2A, H2B, CenH3 and H4 and wrapping DNA in the typical left-handed direction (Camahort *et al.*, 2009; Conde e Silva *et al.*, 2007; Sekulic *et al.*, 2010; Tachiwana *et al.*, 2011a; Zhang *et al.*, 2012). Evidence for an octameric particle comes from a number of organisms and also from *in vitro* reconstitutions. Immunoprecipitations of CenH3 from *Drosophila* and human cell extracts contain CenH3 homodimers with stoichiometric amounts of H4, H2A and H2B (Foltz *et al.*, 2006; Shelby *et al.*, 1997; Zhang *et al.*, 2012). *In vitro*

assembled CENP-A and CENP-A^{Cse4} nucleosomes assemble octameric particles containing H2A, H2B and H4; which does not rule out other models, but demonstrates that CenH3 readily forms octameric nucleosomes (Camahort *et al*., 2009; Tachiwana *et al*., 2011a). These *in vitro* assembled nucleosomes appear to wrap DNA in the canonical left-handed direction but wrap DNA less tightly at the entry and exit sites (Tachiwana *et al*., 2011a) (Figure 1-11), which may account for the lack of a defined nucleosomal repeat length at centromeres (Polizzi and Clarke, 1991; Takahashi *et al*., 1992). In such an octameric model CenH3 nucleosomes would be distinguished from H3 nucleosomes by the specific recruitment of other centromere proteins by CenH3 (Guse *et al*., 2011) and, potentially, by an altered nucleosome / chromatin structure imparted by this distinct histone and the altered wrapping of DNA (Panchenko *et al*., 2011).

Hemisome and reversome models

Perhaps the most persistent alternative model for CenH3 nucleosome stoichiometry is that of the "hemisome" or half nucleosome (Figure 1-10 B), in which it is proposed that CenH3 particles contain a single copy of H2A, H2B, CenH3 and H4 (Dalal *et al*., 2007). Evidence for the hemisome model comes predominantly from height measurements of immunoprecipitated CENP-A^{CID} and CENP-A by atomic force microscopy (AFM) and remains to be independently confirmed (Bui *et al*., 2012; Dalal *et al*., 2007; Dimitriadis *et al*., 2010). In the initial study, cross-linked chromatin extracted from cells was analysed by SDS-PAGE and Western blotting to determine the size of the CENP-A^{CID} nucleosome and it was argued that an absence of cross-links

that are diagnostic of an octameric nucleosome and a propensity to wrap ~ 30 bp less DNA supported the adoption of an unusual hemisome structure (Dalal *et al*., 2007). However, subsequent commentaries have pointed out that the reduced cross-linking observed for CENP-A^{CID} chromatin may be expected because *Drosophila* CENP-A^{CID} lacks many of the cross-linkable lysine residues that are present in H3-containing nucleosomes (Black and Bassett, 2008). Moreover, the reduced wrapping of DNA is consistent with an octameric nucleosome structure, particularly as the octameric crystal structure of a human CENP-A nucleosome was subsequently shown to wrap a similarly reduced length of DNA (Tachiwana *et al*., 2012). Plasmid-based supercoiling assays have suggested that DNA wraps in a unique right-handed direction around *Drosophila* CENP-A^{CID} (*in vitro*) and *S. cerevisiae* CENP-A^{Cse4} (*in vitro* and *in vivo*) relative to H3 nucleosomes and so it has been argued that this wrapping is only compatible with a hemisome model (Furuyama and Henikoff, 2009). However, others have also argued right handed wrapping could occur via reconstitution intermediates, or alternatively, a theoretical "reversome" model in which an octameric histone core is flipped to accommodate a right-handed DNA wrapping (Figure 1-10 D) (Lavelle *et al*., 2009). To date, the reversome model remains purely hypothetical and it is argued that the high energy barrier associated with its formation would necessitate some currently unidentified cofactor (Bancaud *et al*., 2007). One explanation that remains untested is that the positive supercoiling observed in CENP-A^{CID} and CENP-A^{Cse4} nucleosome reconstitutions could result from the formation of non-nucleosomal reconstitution artefacts.

Recently, measurement of H3 and CENP-A nucleosome height using AFM in chromatin extracted from human cells at different stages of the cell cycle showed that the height of CENP-A nucleosomes increases for brief periods after new CenH3 is assembled into chromatin, but that following DNA replication falls back to lower heights that are indicative of hemisomes (Bui *et al*., 2012). Such cell cycle-specific changes in nucleosome stoichiometry could help to explain why an octameric stoichiometry is preferred over hemisomes when CenH3 nucleosomes are assembled *in vitro* (Camahort *et al*., 2009). However, in the same AFM measurement experiments that report cell cycle changes in CENP-A nucleosome height similar cell cycle-dependent changes were observed for the height of H3 nucleosomes (Bui *et al*., 2012). This suggests that the changes in apparent nucleosomal height observed during the cell cycle are common to all nucleosomes and are not CenH3 specific. These changes could instead reflect alterations in nucleosomal stability rather than real differences in height or stoichiometry, perhaps caused by cell cycle-specific post-translational modifications that alter how nucleosomes react under AFM imaging conditions.

Hexasome and trisome models

The hexasome model for CenH3 nucleosome particles is currently limited to budding yeast where it has been proposed that CENP-A^{Cse4} nucleosomes contain two copies of CENP-A^{Cse4}, H4 and the non-histone protein Scm3 (Figure 1-10 E) (Furuyama and Henikoff, 2009; Mizuguchi *et al*., 2011). Evidence for this model comes from chromatin immunoprecipitation (ChIP) assays where reduced levels of H2A, H2B and Htz1 (an H2A variant) were

detected at *S. cerevisiae* centromeres *in vivo*, relative to H4 (Mizuguchi *et al*., 2007). Additional support comes from *in vitro* reconstitutions in which Scm3 was able to replace H2A and H2B in CENP-A^{Cse4} nucleosomes (Mizuguchi *et al*., 2011). However, in ChIPs from strains depleted of CENP-A^{Cse4} or Scm3 H2A, H2B and Htz1 levels failed to recover (Mizuguchi *et al*., 2007). This raised the possibility of a more general problem with ChIPs from this chromosomal location. Moreover, other independent analysis of *in vitro* reconstitutions directly contradict this hexameric model and indicate that Scm3 is not a nucleosomal component (Camahort *et al*., 2009). The related trisome model (Figure 1-10 F) contains just one copy of CENP-A^{Cse4}, H4 and Scm3 and, like the reversome model, was proposed as an alternative explanation for the apparent right-handed wrapping of DNA around CENP-A$^{CID/Cse4}$ nucleosomes (Black and Cleveland, 2011; Furuyama and Henikoff, 2009).

Tetrasome model

Finally, the tetrasome model (Figure 1-10 B) suggests that CenH3 nucleosomes are formed by a tetramer of CenH3 / H4 similar to that found in the octameric nucleosome but lacking both H2A and H2B (Black and Cleveland, 2011). This model arose as an alternative to the hexasome model after Scm3 was found not to be a nucleosomal component (Camahort *et al*., 2009), but otherwise it is based on the original observation that H2A and H2B are depleted at budding yeast centromeres (Mizuguchi *et al*., 2007). Unlike the hexasome model, the tetrasome model also extends to fission yeast, with ChIPs that again show a relative reduction in H2A and H2B levels at

centromeres (Pidoux *et al*., 2009; Williams *et al*., 2009). However, here too, in mutants defective in Scm3 function (increasing the level of H3 at the centromeres) H2A and H2B levels do not recover to that observed elsewhere in the nucleus (Williams *et al*., 2009). This again suggests that some feature of centromeres leads to difficulty with respect to the interpretation of ChIP analyses and extrapolation of nucleosome stoichiometry at these locations. It is also possible that the reduced levels of H2A / H2B may reflect a less stable association of H2A - H2B dimers with CenH3 nucleosomes compared to H3 nucleosomes (Henikoff, 2008).

Thus it appears that the only consensus with respect to features of CenH3 nucleosome models are the presence of CenH3 and H4 histones and a relative reduction in the amount of DNA wrapped around these unusual nucleosomes. Typical H3 nucleosomes wrap ~ 147 bp whereas generally only ~ 120 bp has been observed to wrap around CenH3 nucleosomes (Dalal *et al*., 2007; Panchenko *et al*., 2011; Tachiwana *et al*., 2011a). This relative unwrapping of DNA at the entry / exit sites of nucleosomes has been attributed to the smaller, more flexible ∝N helix of CENP-A and can largely be reproduced with *in vitro* assembled human H3 nucleosomes containing a single amino acid substitution, H3 R49K, in the ∝N helix of H3 and CENP-A (Figure 1-11) (Panchenko *et al*., 2011; Tachiwana *et al*., 2011a). This residue is not conserved in CenH3 from different species, with both *D. melanogaster* and *S. pombe* CenH3 containing an arginine at this site. It has been suggested that the lack of conservation at this site would alter the length of DNA wrapped by these nucleosomes in different species and so might represent adaptations

to the species-specific nucleosomal repeat lengths imposed by the DNA sequences (Panchenko *et al*., 2011). Interestingly, recombinant mouse nucleosomes containing the H2A variant H2A.Bbd also wrap a similarly reduced length of DNA (~ 118 bp). This is thought to be a consequence of the alterations H2A.Bbd imposes on the interaction of the ∝N helix of H3 with DNA at the entry / exit sites (Bao *et al*., 2004). H2A.Bbd nucleosomes appear to associate with actively transcribed regions, although so far the evidence for this is indirect and mainly stems from the observation that H2A.Bbd colocalises with acetylated H4 (Chadwick and Willard, 2001). In contrast, in *in vitro* transcription assays H2A.Bbd nucleosomes assembled onto plasmids containing a natural promoter found H2A.Bbd repressed transcription relative to canonical nucleosomes (Bao *et al*., 2004). In CENP-A arrays the reduced wrapping of DNA at the nucleosome entry / exit sites appears to result in a more condensed chromatin fibre and it has been proposed that this may physically distinguish centromeric chromatin from general chromatin (Panchenko *et al*., 2011). However, *in vivo*, CENP-A nucleosome arrays are interspersed with H3 nucleosomes and it is not known how CENP-A nucleosomes would alter chromatin in this context (Blower *et al*., 2002; Ribeiro *et al*., 2010; Sullivan and Karpen, 2004).

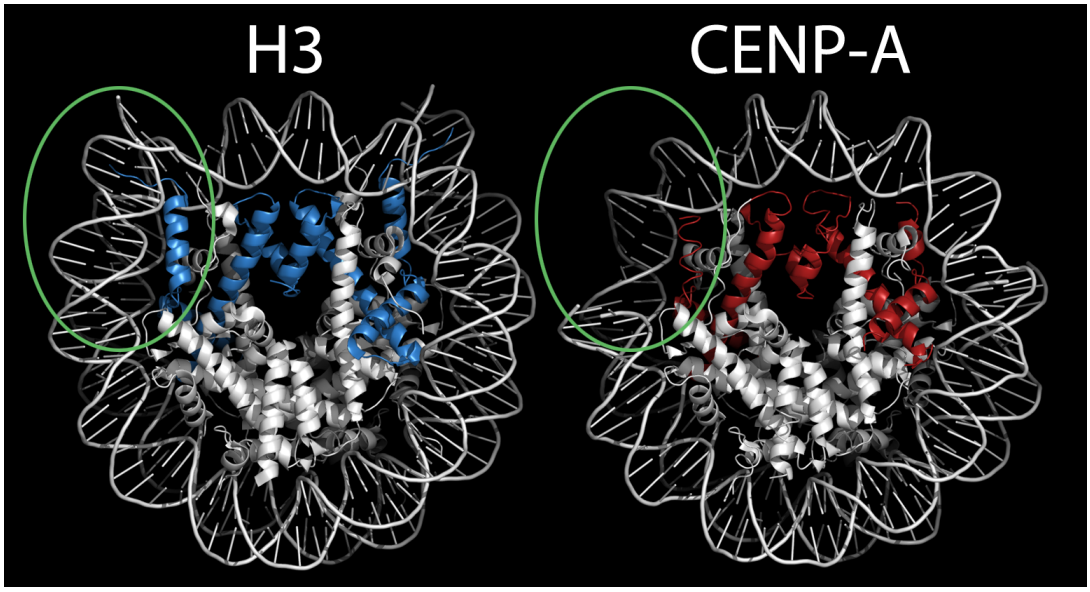


Figure 1-11 – Difference in DNA wrapping between human H3 and CENP-A nucleosomes

The difference in stable wrapping of DNA is apparent from the crystal structures of H3.1 [3AFA (Tachiwana *et al*., 2011b)] and CENP-A [3AN2 (Tachiwana *et al*., 2011a)] nucleosomes. The H3 nucleosome (left) wraps the entire 147 bp DNA, but in the CENP-A nucleosome (right) only 121 bp are stably bound, the terminal 13 bp at both DNA entry / exit sites (one of which is circled in green, the other is in the symmetrical location on the back face of the nucleosome) are not visible in this crystal structure. The transient wrapping of these terminal base pairs in the CENP-A nucleosome is attributed to the smaller, more flexible ∝N helix of CENP-A (Panchenko *et al*., 2011; Tachiwana *et al*., 2011a). For clarity, H3 is coloured blue and CENP-A coloured red, all other histones are in grey.

Aim and scope of this study

Centromere-specific CenH3 nucleosomes are currently the best candidate for the epigenetic mark that defines active regional centromeres in eukaryotes. However, how this mark is recognised and regulated is unclear. There are a number of seemingly contradictory models for CenH3 nucleosome composition and stoichiometry, but one consensus between the current models for CenH3 nucleosomes is the altered wrapping of DNA relative to H3 nucleosomes. This difference is predicted to impart a distinct nucleosome stability to CenH3 nucleosomes that may alter their sequence preference. In this context the purpose of analyses presented here was twofold: first, to critically assess the hemisome and octamer models of CenH3 nucleosome structure; second, to characterise the contribution of DNA sequence to the location of CenH3 nucleosomes.

Chapter 2

Materials and Methods

Cloning *S. pombe* histones

Untagged *S. pombe* genes for the histone proteins H2A, H2B, H3, H4 and CENP-A^{Cnp1} were codon optimised for expression in *Escherichia coli* by the commercial company Geneart (now part of Life Techologies). Geneart assembled fragments from synthetic oligonucleotides and PCR products, cloned the individual histones into pGA4 (ampR) using the Kpn1 and HindIII restriction sites and purified the plasmid from transformed bacteria using the Pure Yield Plasmid Midiprep kit (Promega). Sequencing verified the final constructs to be 100 % correct within the restriction sites used and ~ 10 μg of each was lyophilised for shipping.

The histone sequences were subcloned into bacterial pET3a plasmids (Novagen) using the Nde1 and BamH1 restriction sites so that expression was regulated by the T7 promoter, inducible by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma). Correct insertion of histone sequence was confirmed both by PCR spanning the insertion sites and Sanger sequencing across the whole inset. Each histone expression plasmid was transformed into DH5∝ and used for both glycerol stocks and plasmid purification with the Qiagen midiprep kit, used according to the instructions supplied by the manufacturer.

Histone expression

Histone pET3a expression plasmid was transformed into BL21 Star (Life Technologies) by heat shock and selected on Ampicillin plates. Single colonies were grown in ~ 100 ml 2xTY starter cultures containing 50 mg/l Carbenicillin (Sigma) at 37 C. As the OD600 reached 0.1 the starter culture was diluted into larger cultures, typically 6 x 1l of 2xTY, again with Carbenicillin selection. Cultures were induced with a final concentration of 1 mM IPTG as the OD600 reached 0.2 − 0.4 and a small sample of the pre-induced culture was stored frozen for use as the uninduced sample. Post-induction growth was continued at 37 C for ~ 3 h, except for expression of H4, for which expression time was reduced to 1 h to lessen the toxicity of H4 over-expression. A post-induction sample was taken and stored as for the uninduced sample, with the OD600 recorded. Cultures were then pelleted at 10 C, washed twice with cold wash buffer (50 mM Tris pH 7.5, 2 mM EDTA, 100 mM NaCl, 1 mM βME, protease inhibitor cocktail (Roche)) then resuspended in ~ 30 ml of wash buffer and frozen at - 80 C for storage.

An amount of uninduced and induced culture equivalent to 800 ul of a culture at 0.2 OD600 was lysed with Bugbuster HT (Merck Millipore) and pelleted at 20,000 x g for 5 mins to separate into soluble and insoluble fractions. Both the soluble and insoluble fractions were made up to < 20 ul with NuPAGE loading buffer containing 5 mM βME and run on a 4 − 12 % NuPAGE SDS-PAGE gel in MES running buffer (Life Technologies). Gels were stained with Instant Blue colloidal coomassie (Expedeon) to check both the level of histone expression and whether expression was in the soluble or

insoluble fraction. Invariably, expression of all histones was observed in the insoluble fraction.

Inclusion body preparation

Cells were thawed in a 37 C water bath then sonicated on ice with a large probe sonicator at full power for 3 cycles of 60 sec (10 seconds on / 10 seconds off), with a 2 minute incubation on ice between cycles to cool. 1250 U of Benzonase (Novagen) per litre of culture was added and the sample incubated at RT for 1 h or on ice for 2 h until the sample lost all viscosity. Sample was centrifuged in 30 ml screw cap polypropylene tubes at 4 C, 20,000 x g for 10 minutes in a JA 25.50 rotor (Beckman Coulter) with the supernatant discarded. The pellet was twice resuspended in wash buffer containing 1 % (v/v) Triton X-100 (Thermo Scientific) and pelleted as before, with the supernatant discarded each time. The pellet was washed twice more in wash buffer without Triton X-100 until the pellet was both white and free of detergent. The final pellet resuspended in 15 ml of unfolding buffer (7 M Guanidinium HCl, 20 mM NaAc pH 5.2, 1 mM βME) and left to unfold at RT with gentle rotation for 30 minutes. Undissolved matter was removed by centrifugation at 50,000 x g, RT with the supernatant removed to a fresh tube for dialysis to SAUDE buffer (7 M Urea, 20 mM NaAc pH 5.2, 5 mM βME, 1 mM EDTA) overnight at 4 C using dialysis tubing with a 6 − 8 kDa molecular weight cut off.

Cation exchange chromatography

Unfolded inclusion body preparations were clarified by centrifugation at 50,000 x g, 15 minutes, 4 C and filtered through a 0.2 μm filter before loading onto a 5 ml HiTrap SP HP cation exchange column at 5 ml/minute. The column was subjected to a continuous gradient of SAUDE buffer containing between 100 − 600 mM NaCl over 20 column volumes and peak fractions, including the lysate, flow-through, wash and elution fractions were sampled on a 4 − 12 % NuPAGE SDS-PAGE gel (Life Technologies) stained with Instant Blue (Expedeon). Samples containing the semi-purified histone were pooled and extensively dialysed over 24 h (with three changes of dialysis buffer) to water containing 5 mM βME at 4 C. After refolding, samples were clarified by centrifugation at 50,000 x g, 10 minutes, 4 C, were lyophilised and resuspended in ~ 10 ml of H_2O to reduce sample volume. Protein concentration was calculated by UV spectroscopy (Nanodrop - Thermo Scientific) and samples split into 200 nmol aliquots for lyophilisation and long term storage at - 20 C. The extinction coefficients used for each histone were 4470 $cm^{-1}M^{-1}$ for H2A, 7450 $cm^{-1}M^{-1}$ for H2B, 4470 $cm^{-1}M^{-1}$ for H3, 8605 $cm^{-1}M^{-1}$ for CENP-A^{Cnp1} and 5960 $cm^{-1}M^{-1}$ for H4.

Refolding histone octamer

For a given amount of H3 or CENP-A^{Cnp1} histone octamer 1.5 x the amount of H2A and H2B were required relative to the amount of H3 / CENP-A^{Cnp1} and H4. The required number of lyophilised histone aliquots (H2A, H2B, H3 and H4 to make H3 octamer; H2A, H2B, CENP-A^{Cnp1} and H4

for CENP-A^{Cnp1} octamer) were resuspended in 1 ml of unfolding buffer (7 M Guanidinium HCl, 20 mM NaAc pH 5.2, 10 mM βME) for each 200 nmol aliquot and left to unfold with gentle rotation at RT for 1 h. Undissolved material was removed by centrifugation at 50,000 x g, 10 minutes, 4 C and the concentration of each histone was quantified by UV spectroscopy (Nanodrop - Thermo Scientific). Histones were mixed to give a 50 % excess of H2A and H2B relative to the amount of H3 and H4 (for H3 octamer) or CENP-A^{Cnp1} and H4 (for CENP-A^{Cnp1} octamer) and dialysed at 4 C over 12 h against at least 3 changes of refolding buffer (2 M NaCl, 10 mM Tris-Cl pH 7.5, 1 mM EDTA, 5 mM βME, 1 M 3-(1-Pyridinio)-1-propanesulfonate (Sigma)). Any precipitate was removed by centrifugation at 12,000 x g for 15 minutes, 4 C and the supernatant kept at 4 C for further purification.

Purification of refolded octamer

Refolded octamer was concentrated to < 500 μl using Ultrafree 15 ml 5 kDa spin concentrators (Merck Millipore) and immediately loaded onto a HiPrep Superdex S200 16/300 column (Amersham Biosciences) equilibrated with filtered and degassed refolding buffer lacking both 3-(1-Pyridinio)-1-propanesulfonate and βME, kept at 6 C. 1 μl of sample was kept back as an input sample, stored at 4 C. The column was subjected to a flow rate of 0.5 ml / minute and maximum back pressure across the column of 1.5 MPa using the AKTA Purifier (GE Healthcare). Typically, the major peaks eluted at the following volumes: ~ 9 – 11 ml for aggregates, ~ 12.5 ml for octamer (the back of this peak often contain H3 / H4 or CENP-A^{Cnp1} / H4 tetramer as a visible shoulder to the octamer peak), and ~ 15 ml for H2A / H2B dimer. 5 μl

of peak fractions and the input fraction were sampled on 4 − 12 % NuPAGE SDS-PAGE gels (Life Technologies) run with MES buffer and stained with Instant Blue (Expedeon) to check the purity and fractions found to contain refolded octamer were pooled, concentrated to approximately 15 − 30 μM. The concentration was calculated by UV spectroscopy (Nanodrop - Thermo Scientific) and the refolded octamer stored at 4 C for ~ 2 − 3 months, or until the proteins began to degrade. To monitor for signs of degradation 2 μl of the octamer stocks were run on SDS-PAGE gels, any degradation was evident in the appearance of smeared protein bands by staining with Instant Blue (Expedeon).

PCR and purification of DNA for nucleosome reconstitutions

The pUC19 plasmid containing the 601.3 sequence (Anderson and Widom, 2001) was a gift from Tom Owen-Hughes (University of Dundee). DNA fragments for nucleosome reconstitutions were produced by PCR, often with a Cy3-labelled 5′ primer to aid visualisation of the resulting nucleosomes. These PCRs were typically 1 ml reactions and used both the high fidelity polymerase Platinum Pfx (Life Technologies) and a relatively high concentration of primers to improve yield. Conditions for a typical PCR reaction are shown in Table 2-1. The 1 ml PCR product was mixed with 300 ul of 1 M sucrose, run on a horizontal 5 % native PAGE gel in 0.2 x TBE buffer at 140 V for ~ 1 h and post-stained with ethidium bromide to visualise the DNA under UV. Ethidium bromide staining was not necessary if a Cy3-labelled primer was used as the large quantity of DNA could be visualised directly by eye. The specific product band was cut from the native PAGE gel

and extracted by electro-elution using a 6 – 8 kDa dialysis bag containing the gel slice and ~ 0.5 – 1 ml 0.2 x TBE run at 140 V in 0.2 x TBE for ~ 15 minutes. The eluted DNA was quantified by UV spectroscopy (Nanodrop - Thermo Scientific) and concentrated to ~ 1 mg / ml using Ultrafree 0.5 ml 5 kDa spin concentrators (Merck Millipore) if required. A typical final yield was ~ 1 μmol of purified DNA per 1 ml PCR.

	Volume (ul)
H2O	713
10 x Pfx Buffer	200
Fwd primer (100 uM)	15
Rev primer (100 uM)	15
dNTPs (10 mM each)	30
MgSO4 (50 mM)	20
Template DNA (~ 100 ng / ul)	2
Platinum Pfx polymerase	5
TOTAL	**1000**

Table 2-1 - PCR conditions for the large scale production of DNA

A typical PCR reaction for the production of DNA used in nucleosome reconstitutions. These large scale PCR reactions use a higher than normal concentration of primers to increase yield. Following the PCR reaction, the product was typically purified from a native PAGE gel by electro-elution.

Radiolabelling of DNA

For the purification of P^{32} end-labelled DNA, the PCR of the required fragment proceeded as described above for unlabelled DNA except that primers included the Ava1 restriction digest site (CCCGAG) at both the 5′ and 3′ ends. Purified PCR product was digested with Ava1 (New England Biolabs) in NEB buffer 4, purified from an agarose gel using a gel extraction kit (Qiagen) and the concentration quantified by UV spectroscopy (Nanodrop - Thermo Scientific). Digested DNA was end-labelled with P^{32} dCTP using Klenow exo- (New England Biolabs) in NEB buffer 2 and unincorporated label was removed using an illustra Microspin G50 spin column (GE Healthcare). The labelled was DNA stored at - 20 C.

Nucleosome reconstitution

Nucleosomes were prepared from existing stocks of histone octamer and purified DNA, with the DNA adjusted to 2 M NaCl before mixing with histone octamer (already in 2 M NaCl) at roughly equal molar ratios. Reconstitution reactions were placed within 0.5 ml mini dialysis devices (Thermo Scientific) which floated within a beaker of reconstitution buffer containing 2 M NaCl (2 M NaCl, 20 mM Hepes pH 7, 2 mM EDTA). A peristaltic pump slowly diluted the reconstitution buffer 5 - fold with reconstitution buffer containing 10 mM NaCl over 12 h, proceeded by a further 3 buffer changes to the low salt reconstitution buffer over a further 24 h period. The exact ratio of DNA : Octamer in reconstitutions varied according to the accuracy of the concentration calculations but for each new

batch of octamer or DNA a small titration reconstitution was set up over the DNA : Octamer range of 0.7 : 1 to 1.3 : 1 to establish a ratio that would provide a slight excess of DNA after reconstitution. The ratio of free DNA to nucleosomal DNA was assessed by mixing ~ 1 pmol of reconstituted nucleosome sample with ~ 5 μl of a 20 % sucrose (loading buffer), running on a 5 % native PAGE gel in 0.2 x TBE at 140 V for 45 mins and post-staining with Sybr Green 1 (Life Technologies). Input DNA and a 100 bp DNA ladder were run alongside nucleosome reconstitutions in order to identify the free DNA band. Assembled nucleosomes were stored at 4 C for no more than 2 weeks to avoid degradation.

Nucleosome reconstitution onto radiolabelled DNA

Reconstitutions proceeded as described for unlabelled nucleosomes except that labelled DNA was first mixed with an unlabelled version of the same DNA at a 1 : 10 ratio (respectively) to minimise inaccuracies in the ratio of DNA : octamer that may have arisen from the estimated quantification of labelled DNA concentration.

Nucleosome reconstitution onto genomic DNA

Genomic DNA was prepared from the wildtype 972 strain of *S. pombe* using a genomic DNA isolation kit (Qiagen) according to the manufacturer's instructions. Reconstitutions were performed as described for typical reconstitutions except that 20 μg (185 pmol) of octamer was mixed with either 200, 400 or 800 μg of genomic DNA for the 10, 20 and 40 % occupancy

samples respectively. These ratios of octamer to DNA give molar ratios of DNA : octamer of ~ 10 : 1, 20 : 1 and 40 : 1 (respectively) when assuming the typical nucleosome repeat length of ~ 165 bp observed *in vivo* for *S. pombe*.

Nucleosome fixation with $BS(PEG)_5$

Nucleosomes were dialysed to a fixation buffer of 20 mM Hepes pH 7, 2 mM EDTA and the primary amine cross-linker $BS(PEG)_5$ (Thermo Scientific) was then added at the required molar excess (1000 – 5000 x for full fixation of the histone octamer). Samples were left to fix for 2 h at 37 C with gentle shaking before fixation was quenched by adding Tris pH 7 to a final concentration of 200 mM. To check the extent of fixation, ~ 15 pmol of nucleosome was digested with 0.5 μl of Benzonase (Novagen) at RT for 10 minutes then boiled in SDS-PAGE loading buffer (Life Technologies), run on a 4-12 % NuPAGE SDS-PAGE gel in MES buffer (Life Technologies) alongside an unfixed control sample and the gel stained with a silver staining kit (Life Technologies). Fixed nucleosome samples could be stored at 4 C for a number of weeks but were typically used immediately after fixation.

Micrococcal nuclease digestion of nucleosomes

Nucleosomes were diluted ~ 5-fold in Micrococcal nuclease (MNase) digest buffer (20 mM Tris pH 8.0, 1 mM $CaCl_2$, 0.01 % Igepal) so that the final concentration of total DNA was 16.3 μg / ml and incubated in a 37 C water bath for 10 minutes. MNase (Sigma) was added at 0.6 mUnits per μg of total DNA and the sample shaken gently in a 37 C water bath for 1 minute before

the digest was stopped by making the sample up to 66 mM EDTA, 15 μg Proteinase K per μg of total DNA and 1 % (w/v) SDS. Samples were then incubated at 65 C for 2 h, extracted twice with phenol chloroform and ethanol precipitated. To remove excess salts the samples were briefly dialysed to TE before running on a 2 % agarose gel and gel extraction of the mono-nucleosome band (at ~ 150 bp). Samples were run alongside an undigested input control that was similarly treated but without the addition of MNase and a DNA-only sample that was treated as the other digested samples but using just 15 mU of MNase per 100 μg of DNA. The typical efficiency of this protocol was low, with typical yields of < 1 μg mono-nucleosomal DNA from 100 μg of total input DNA. Efficiencies were found to increase to ~ 7 % of input if PCR purification columns (Qiagen) were used in place of phenol extraction and ethanol precipitation.

Mass spectrometry of heavy / light labelled nucleosomes

Mixed heavy / light -labelled CENP-A^{Cnp1} (27 μg) or H3 (40 μg) nucleosomes in 50 mM HEPES buffer (pH 7.8, 200 mM $KClO_4$, 10 mM EDTA at 0.1 μg / μl) were incubated with a 1000-fold molar access of $BS(PEG)_5$ (Thermo Scientific) at 4°C o/n. Cross-linking was quenched by addition of saturated ammonium bicarbonate (ABC) to a final concentration of 50 mM for 30 min. Nucleosome were separated in SDS-PAGE (Life Techologies Nu-PAGE Novex 4 - 12 % Bis-Tris gel in MES buffer). The band containing specific cross-linking product (octamer) was cut out and digested "in gel" with trypsin, the resulting peptides were analysed by SCX using Stage-Tips as described previously (Chen *et al*., 2010).

Peptide fractions were loaded onto an analytical column, packed with C18 material (ReproSil-Pur C18-AQ 3 μm) using a flow rate of 0.7 μl / minute. Peptides were eluted using a linear gradient from 4 % acetonitrile in 0.5 % acetic acid to 24 % acetonitrile in 0.5 % acetic acid over 81 minutes at 0.25 μl / minute into an Orbitrap Velos (Thermo Scientific). MS spectra were acquired in the Orbitrap at a resolution of 100,000. The five most intense peptide precursor ions with charge stages z = 3 or higher were fragmented per cycle. MS/MS spectra were acquired in the Orbitrap mass analyser at resolution 7500 using CID at 35 %.

Raw data files were processed into peaks using MaxQuant (Cox and Mann, 2008) at default parameters, except *"Top MS/MS peaks per 100 Da"* was set to 200. To identify cross-linked peptides, searches were conducted using in-house software "Xi". Parameters used: MS accuracy 6 ppm; MS/MS accuracy 20 ppm, trypsin as enzyme; fixed modifications being carbamidometylation on cysteine; variable modifications being, oxidation on methionine, heavy labels: Lys8 Arg10 and user defined database: *S. pombe* H3, H4, H2A, H2B and CENP-A^{Cnp1} from UniProt (www.uniprot.org). Cross-links identified in Xi were manually validated and mapper onto the crystal structures for human H3 or CENP-A nucleosomes (Protein Data Bank IDs: 1KX5 and 3AN2 respectively).

Oxidative cross-linking of cysteines within H3 and CENP-A^{Cnp1} nucleosomes

The sulphydryl groups within H3 and CENP-A^{Cnp1} nucleosomes were catalytically oxidised with copper phenanthroline as described in Gould *et al* (1980). Briefly, nucleosomes were first dialysed into TEP buffer [50 mM triethanolamine, 0.5 mM EGTA, 0.1 mM phenylmethlsulphonyl fluoride (PMSF)] at pH 8 then incubated with 0.1 volume of 25 mM 1-10 o-phenanthroline (Sigma) and 12.5 mM copper sulphate at 4 C for at 24-48 h. Treatment was terminated with the addition of EDTA to a final concentration of 10 mM and dialysis to TE at 4 C. The presence of disulphide bonds were analysed by running samples on SDS-PAGE and visualised by silver staining the gel. To reduce disulphide bonds, samples were incubated at RT with 0.5 M 2-mercaptoethanol for 30 minutes prior to analysis by SDS-PAGE.

Dynamic light scattering

Dynamic light scattering was performed using the Zeitasizer Auto Plate Sampler (Malvern) which was temperature controlled to be at 25 C. Nucleosome samples were prepared in a final volume of 50 μl DLS buffer (50 mM Hepes pH 7.8, 50 mM KCl, 5 mM EDTA) at a concentration of > 0.2 mg / ml (DNA concentration calculated by UV spectroscopy) and loaded into the instrument in a 384-well plate format. Samples automatically loaded into the flow cell were equilibrated for 2 minutes before averaging the results of 15 x 10 second readings for each of 3 independent reconstitutions. Buffer was used as a control but typically showed no counts above background.

Atomic force microscopy (AFM) imaging of nucleosomes

Mica (V1 grade - SPI supplies) was freshly cleaved using a clean scalpel and functionalised with 3-aminopropyltriethoxy silane (APTES - Sigma) as described in Lyubchenko *et al* (2009) then coated with glutaraldehyde as described in Wang *et al* (2002). Samples were pipetted onto the functionalised surface at a titration of nucleosome concentrations centred around 1 nM. Deposited samples were left for 5 minutes to adhere at RT then rinsed twice by allowing molecular biology grade water (Sigma) to flow over the surface held at a 45 degree angle. A stream of argon was used to gently dry the surfaces and they could then be either imaged immediately or stored under a clean argon atmosphere for several months. AFM imaging was performed in tapping (non-contact) mode in air using either a Veeco Explorer or a Veeco Nanoman VS with Dimension 3100 controller (Bruker). Images were collected over an area of between $1 - 5\ \mu$m at a scan rate of ~ 1.2 Hz and the DLC-10 tips used (Bruker) had a nominal resonance of 160 kHz, stiffness of 5 N / m and a tip radius of 1 nm.

Automated processing of AFM images

Images were flattened using the Nanoscope software (Veeco) and exported for analysis in ImageJ (NIH). Within ImageJ, the image scale was set and a Gaussian blur applied using a radius of 1 nm to reduce noise. A duplicate image was created with the contrast enhanced to allow for the clear selection of multiple large areas of background (surface). The modal height of these background regions was determined by redirecting measurements to

the original image and the average modal background height from multiple areas was subtracted from the entire original image in order to set the average background height to 0 nm. The duplicate image was used to select 10 points along a number of naked DNA fragments and the maximum heights measured by redirecting to the original image. The average maximum DNA height was recorded for each image to enable normalisation between images (as DNA height should remain constant). The duplicate image was closed and a fresh duplicate made from the original image to act as the particle filter mask. The mask image was thresholded at a lower limit of 0.2 nm above the average DNA height to isolate nucleosomes from the DNA of the array. The image was converted to an 8-bit image and local thresholding used to define individual nucleosomes. The local thresholding filter used was chosen separately for each image to be appropriate for separating particles but was typically the midgrey filter, with a radius of ~ 5 nm. The maximum heights of particles were calculated for all particles with an area between 78.5 nm^2 and 2000 nm^2 and a circularity > 0.5 by redirecting measurements to the original image. Nucleosome heights from multiple images were normalised by addition or subtraction such that the DNA height in each image was adjusted to 0.5 nm (which was the average height, ± ~ 0.2 nm, of DNA across all images).

Manual selection of nucleosomes

Images were pre-processed as described for the automated analysis above. Using a high contrast duplicate of the image used to measure nucleosomes resulting from the automated analysis manually verified

nucleosomes were selected using the multi-point tool in ImageJ. All nucleosome-like particles within the image were selected so long as they were associated with DNA in a clearly defined beads-on-a-string pattern and were not overlapping other fragments of DNA or nucleosomes. Measurements of maximum height were redirected to the original image as for the automated analysis and heights were again normalised to the height of DNA within each image, as described above. In this manner overlapping DNA fragments, aggregated nucleosomes and particles not associated with DNA were easily excluded from analysis.

High throughput sequencing

Nucleosomes reconstituted onto genomic DNA (as described above) were heat-shifted by incubation at 37 C for 30-60 mins. Samples containing roughly 4 μg of DNA were diluted to 250 μl with MNase digest buffer (above) and digested with MNase to obtain predominantly mononucleosome-length DNA fragments. Mononucleosome-length DNA from three experimental replicas was purified by extraction from agarose gels using Qiagen gel extraction kits and was sent to the commercial company BGI (http://www.genomics.cn/en/index) for paired-end library processing and Illumina sequencing. The number of paired reads obtained for the CENP-A^{Cnp1} 10 %, 20 % and 40 % occupancy samples were 23,046,085, 24,207,376 and 19,632,340, respectively. For the H3 samples at 10 %, 20 % and 40 % the number of reads obtained were 20,561,388, 25,008,305 and 22,466,119, respectively. Reads that were between 100 and 147 bp in length were mapped using the novoalign software (http://www.novocraft.com),

with mappings for repeat regions divided by the repeat number. At the 10 % occupancy level, the number of mapped reads for the CENP-A^{Cnp1} and H3 samples were 17,283,993 and 16,461,663, respectively.

Chapter 3

Components and stoichiometry of CENP-A^{Cnp1} containing nucleosomes *in vitro*

Introduction

Despite the recent publication of a crystal structure (Tachiwana *et al*., 2011a) the long debate over the stoichiometry within CenH3 nucleosomes remains open. The crystal structure, like much of the biochemical data collected to date for CenH3 nucleosomes suggests an octameric stoichiometry similar to that of the typical H3 nucleosome - with two copies of each histone. The major competing model is that of a hemisome, containing just one copy of each histone (Dalal *et al*., 2007), which has persisted largely due to *ex vivo* nature of the data. One of the common criticisms of data supporting the octameric model regards the reliance on *in vitro* nucleosomes and whether these mimic true *in vivo* states. Yet the nature of the CenH3 nucleosome does not lend itself to direct *in vivo* or *ex vivo* analysis, being both at low abundance within a cell and relatively difficult to extract intact (Zhang *et al*., 2012). Data supporting the hemisome model comes predominantly from the *ex vivo* comparison of nucleosome dimensions by atomic force microscopy (AFM), which consistently finds both fly (CENP-A^{CID}) and human CENP-A nucleosomes appear smaller in height than their H3 counterparts (Bui *et al*., 2012; Dalal *et al*., 2007; Dimitriadis *et al*.,

2010). The advantage of AFM is its sensitivity, the ability to measure the dimensions of individual nucleosomes from pico-molar quantities, yet the reliance of the hemisome model on this technique raises concerns about inferring a difference in stoichiometry from a difference in height. What has been lacking is the same analyses applied to CenH3 nucleosomes assembled *in vitro*. In this chapter recombinant CENP-A^{Cnp1} nucleosomes from *S. pombe* were produced *in vitro* and their stoichiometries tested by various biochemical techniques, with the aim of comparing with an AFM analysis of these same nucleosomes in the next chapter. In this way these *in vitro* nucleosomes are used as a negative control to test whether CENP-A nucleosomes with defined octameric stoichiometries still appear as hemisomes by AFM. To our knowledge this is the first time recombinant *S. pombe* nucleosomes have been produced *in vitro*.

In vitro studies of nucleosomes generally take two forms, those using recombinant histones and those using histones isolated from chromatin. However, when using histones isolated from chromatin the various histone variants and post-translational modifications of the population reduce the homogeneity of the sample and can vary significantly between batches depending on the type and developmental state of the tissue from which they were isolated (Luger *et al*., 1999). Recombinant nucleosomes produced *in vitro* lack post-translational modifications, but as no modifications have been identified for CENP-A^{Cnp1} and the two modifications identified in other organisms (R37me2a in budding yeast and S7p in humans) are not conserved residues in fission yeast (Samel *et al*., 2012; Zeitlin *et al*., 2001a), recombinant

technology represents the only reliable method for obtaining the large quantities of purified histones required for many biochemical assays.

For this work, recombinant H3 and CENP-A^{Cnp1} *S. pombe* nucleosomes were produced following an established and commonly-used protocol (Luger *et al*., 1999). This protocol was originally used for the production of recombinant H3 nucleosomes from frog histones but has since been used for the production of nucleosomes from a wide range of species, including budding yeast (Camahort *et al*., 2009) and humans (Tachiwana *et al*., 2011a). Briefly, the proceedure entails the recombinant expression and purification of individual histones (which may either be tagged or untagged), followed by mixing of unfolded histones at an equal ratio and refolding by dialysis to a buffer containing 2 M salt. The high salt in this buffer is essential for maintaining the integrity of the histone complexes that form during refolding. Octameric histone complexes are then purified from any excess H3:H4 tetramer, H2A:H2B dimer, aggregates or free histone using a gel filtration step. The resulting histone octamer (or stoichiometric amounts of tetramer and dimer) is then mixed with the desired DNA and slowly dialysed to a low salt buffer (< 250 mM) to obtain a mixture of reconstituted nucleosomes and excess free DNA.

Results

Assembling recombinant *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes *in vitro*

Individual *S. pombe* histone genes were codon optimised for expression in *E. coli* by GeneArt (now part of Life technologies) and cloned into pET3a

expression vectors. Expression of each untagged histone was trialled in a range of cell types (CD41, CD43, Rosetta, BL21 and pLysS versions of each) before BL21 cells were chosen since they gave the strongest expression in all cases. Notably, two observations first noted by Luger *et al*, 1999 with *X. laevis* histones were also found to be true for *S. pombe* histones. Firstly that histone H4 expression levels were consistently lower than for the other histones (often also the case here for H2B) and secondly that starter cultures grown over an OD_{600} of 0.6 at any point prior to induction significantly reduce their ability to give good expression, even if subsequently diluted.

Histone expression was almost exclusively in the insoluble fraction when cells were lysed with the Bugbuster HT (Merck Millipore) detergent mix, yet they continued to bind a significant amount of nucleic acid even in denaturing and high salt conditions (typically ~ 1.4 mg of DNA per 1 l expression culture, based on UV spectroscopy). Instead, prior digestion with Benzonase nuclease (Novagen) was required for efficient removal of contaminating nucleic acids.

Purification of unfolded histones was achieved using cation exchange chromatography over a 5 ml HP SP column, with continuous salt gradients for each histone within the range of 100 mM − 1 M NaCl (Figure 3-1). The identities of the purified histones were confirmed by MALDI-TOF mass spectrometry (not shown) and each was dialysed to water then lyophilised in aliquots for storage at − 20 C.

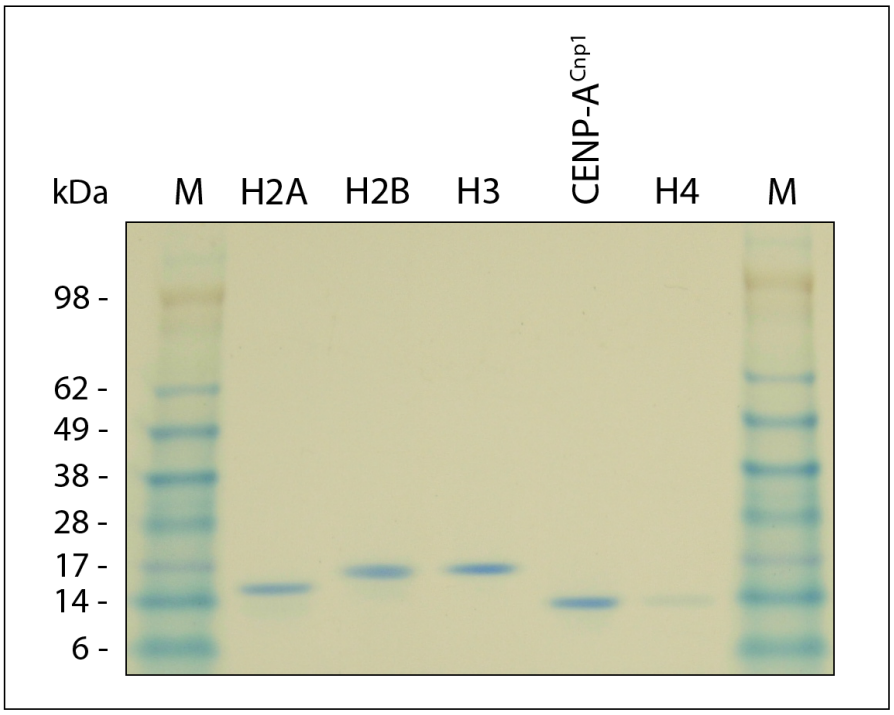


Figure 3-1 – Histones purified by cation exchange

SDS-PAGE gel of individual histones purified by cation exchange and stained with colloidal Coomassie (Instant Blue from Expedeon). As previously observed, histone H4 was often expressed at much lower quantities than other histones due to the toxic nature of H4 over-expression (Luger *et al.*, 1999).

The classical protocol for preparation of recombinant histone octamer calls for unfolded histones to be mixed at an equal ratio before refolding in 2 M salt (Luger *et al.*, 1999). Initial attempts to refold the *S. pombe* CENP-A^{Cnp1} histone complex in this way reproducibly resulted in precipitation of the majority of CENP-A:H4 during refolding, while H3 histone complexes refolded with very little or no visible precipitation. Attempts to improve the refolding of CENP-A^{Cnp1} histone complexes included the use of mild non-ionic detergents, reducing agents, a range of total protein concentrations,

altering the speed of refolding through controlling the rate of dialysis, and buffer pH. These varieties of conditions had little effect on the final level of CENP-A^{Cnp1} : H4 precipitation (not shown). Instead, soluble refolding of CENP-A^{Cnp1} histone complexes was only observed with either the use of a non-detergent sulfobetaine (NDSB - a zwitterionic detergent) at 0.6 M or when the ratio of input histones included a ~ 50 % excess of H2A : H2B relative to CENP-A^{Cnp1} : H4 (Figure 3-2). However, this experiment simply records the soluble proteins resulting from the refolding reaction and does not imply that the histones are correctly folded or formed into canonical histone complexes. The salt level into which the histones were dialysed also appear to make a difference, with an increase in the soluble levels of all histones when the salt was reduced from 2 M to 0.5 M. High salt is used in these refolding reactions to help neutralise the charge of the histone proteins such that they are able to interact in the absence of DNA - interaction of the H3/CENP-A^{Cnp1} : H4 complexes with DNA are required to maintain complex integrity below ~ 1.2 M salt. As 0.5 M salt appears to increase the amount of soluble protein it would seem 2 M salt may either be disrupting the interaction between histones or else encouraging interaction to such a degree that histones begin to aggregate into insoluble complexes. Attempts to first refold histones at 0.5 M salt then dialyse them to 2 M salt resulted in precipitation of all histones upon dialysis to 2 M salt (not shown), perhaps suggesting 2 M salt was inducing aggregation of the histones. It may still be possible to titrate the salt to a level that would allow formation of histone complexes without aggregation but for these and future reconstitutions 0.6 M NDSB and a 50 % excess of H2A : H2B relative to H3/CENP^{Cnp1} : H4 were

used instead, which appeared to aid CENP-A^{Cnp1} : H4 refolding or else to stabilise the correctly folded complex.

Once refolded, clarified histones were run on a gel filtration column to purify histone complexes from any soluble histone aggregates. The gel filtration traces showed soluble aggregates of H3 : H4 but almost no soluble aggregates in the CENP-A^{Cnp1} sample (Figure 3-3). In addition, little CENP-A^{Cnp1} : H4 is observed as soluble tetramer (or dimer) compared to the level of H3 : H4 tetramer formed. Consequently, it would seem the *S. pombe* CENP-A^{Cnp1} : H4 complex has a strong tendency to form insoluble aggregates under the conditions trialled here - a conclusion that is in agreement with observations of recombinant CENP-A^{Cse4} : H4 complexes from *S. cerevisiae* (Camahort *et al*., 2009) and the apparent tendency for CENP-A to aggregate (Palmer *et al*., 1991).

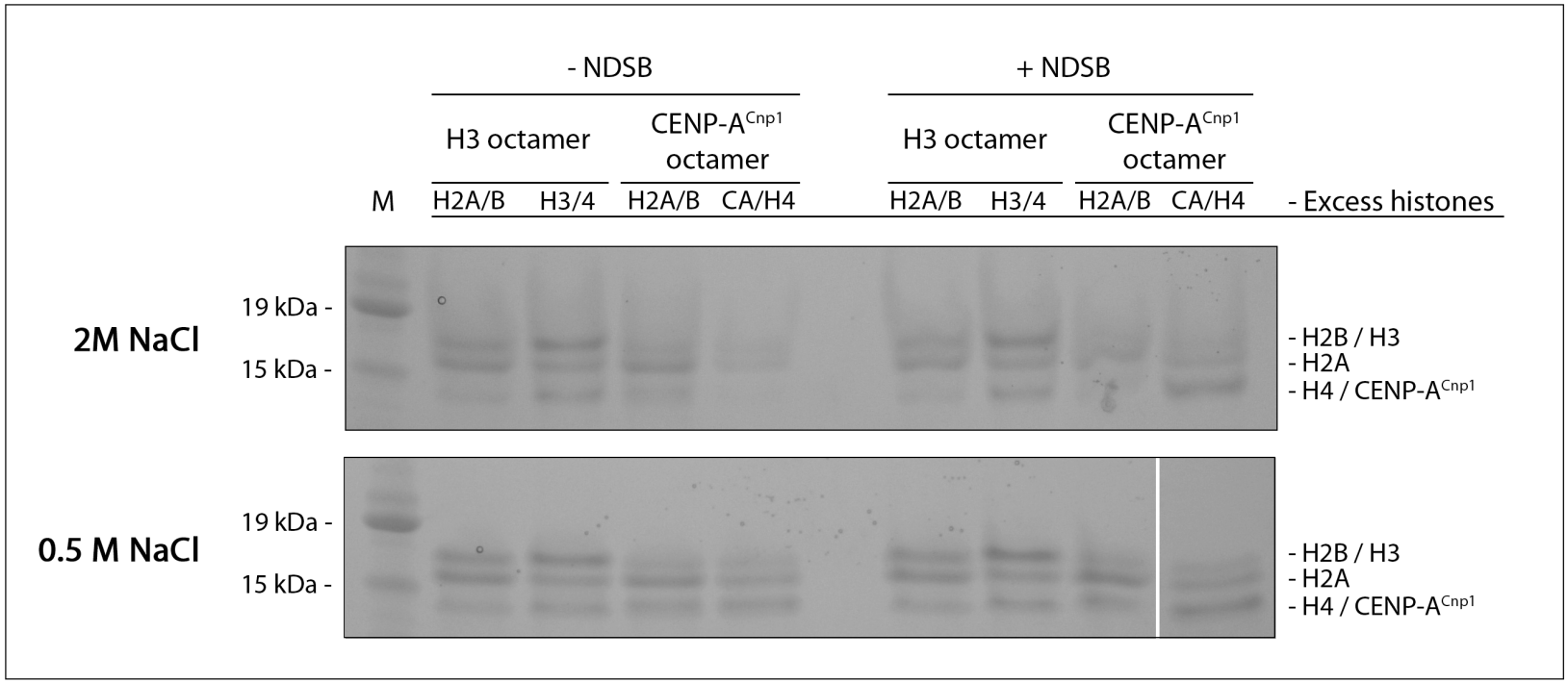

- NDSB
+ NDSB
H3 octamer
CENP-A^Cnp1 octamer
H3 octamer
CENP-A^Cnp1 octamer
M
H2A/B
H3/4
H2A/B
CA/H4
H2A/B
H3/4
H2A/B
CA/H4
- Excess histones
2M NaCl
19 kDa -
15 kDa -
- H2B / H3
- H2A
- H4 / CENP-A^Cnp1
0.5 M NaCl
19 kDa -
15 kDa -
- H2B / H3
- H2A
- H4 / CENP-A^Cnp1

Figure 3-2 – Comparing conditions for the refolding of histone complexes

Lyophilised histones were unfolded, quantified by UV spectroscopy and mixed with a 50 % molar excess of either H2A:H2B or H3/CENP-A^{Cnp1}:H4. Samples were then refolded in either 2 M or 0.5 M salt, with and without 0.6 M of the non-detergent sulfo-betaine 3-(1-Pyridinio)-1-propanesulfonate (Sigma). Clarified samples were run on the SDS-PAGE gels shown here and stained with colloidal Coomassie (Expedeon) to compare the relative amounts of each histone found in the soluble fraction after refolding in each of these conditions.

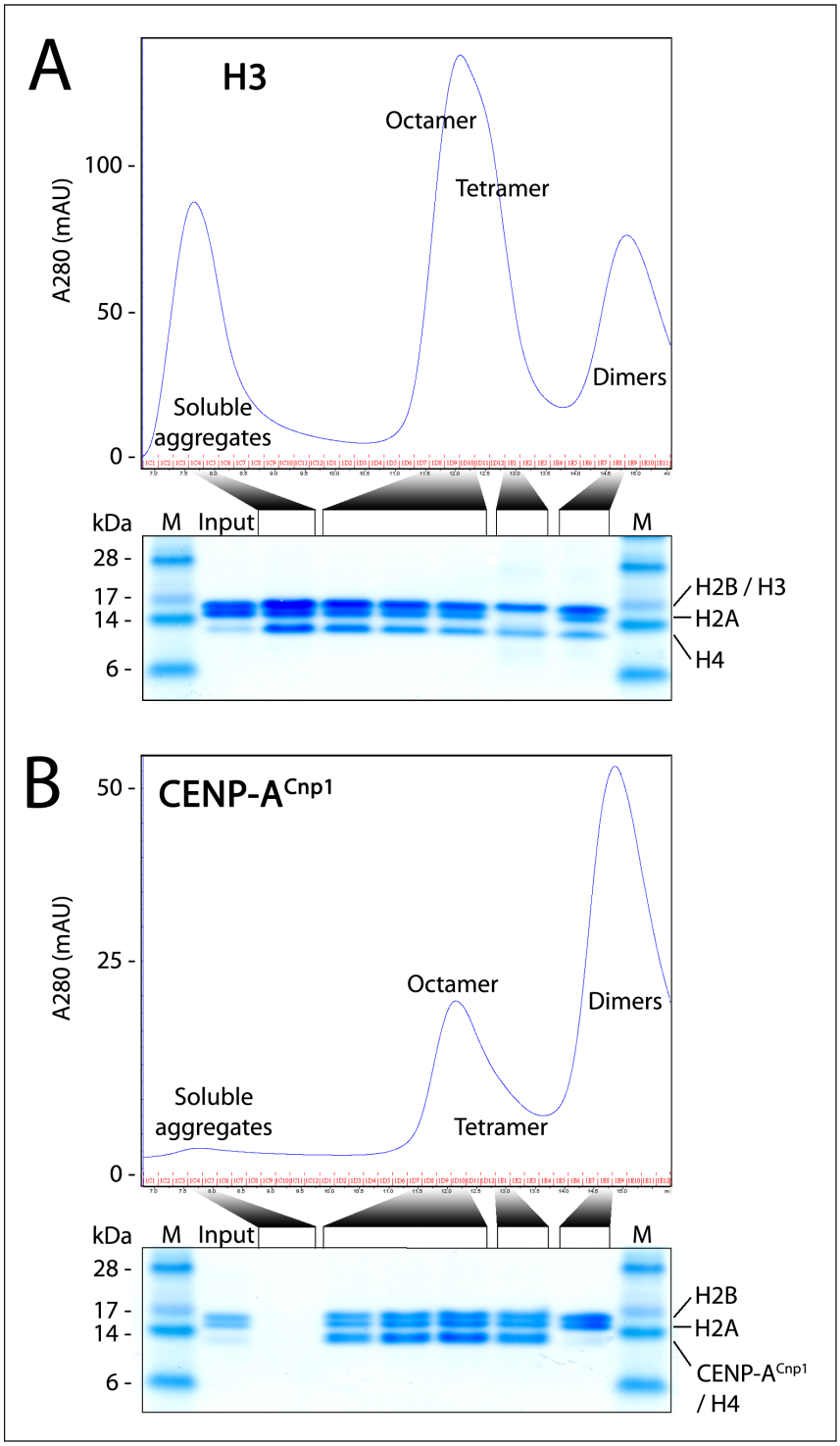


Figure 3-3 – Elution profiles of H3 and CENP-A^{Cnp1} octamer refolding

Unfolded input histones (a) H2A, H2B, H3 and H4 or (b) H2A, H2B, CENP-A^{Cnp1} and H4 were mixed for a 50 % molar excess of H2A and H2B and refolded by overnight dialysis to 2 M salt. The resulting histone complexes

were separated over Superdex S200 10/300 gel filtration columns and labelled as soluble aggregates, octamer (two copies of each histone), tetramer [$(H3 : H4)_2$ or $(CENP\text{-}A^{Cnp1} : H4)_2$] or dimers (H2A : H2B) according to the protein content of each fraction and the migration patterns defined by Bowman *et al* (Bowman *et al*., 2010) for histone complexes on this column. Peak fractions were chosen for SDS-PAGE analysis and octamer fractions containing an equal ratio of histones were pooled for reconstitution into nucleosomes. Input refers to the sample loaded onto the gel filtration column rather than the input into the refolding reactions.

Importantly, despite the requirement for an excess of H2A : H2B in the refolding reactions, gel filtration traces of H3 and CENP-A^{Cnp1} histone complexes after refolding showed similar elution profiles (Figure 3-3), which are in excellent agreement with published purifications of recombinant *Xenopus* histone complexes using a similar gel filtration column (Bowman *et al*., 2010). Furthermore, analysis of the ratio of CENP-A^{Cnp1} and H4 (tetramer) to H2A and H2B (dimer) from the fractions sampled in Figure 3-3b show the supposedly octameric complexes contained each of the input histones at an equal ratio, again suggesting these complexes are either tetrameric or octameric (with one or two copies of each histone, respectively) (Figure 3-4). The apparent (0.2 fold) excess of H2A and H2B in these quantitations is thought to reflect the ~ 0.4 fold decrease in Coomassie staining observed for H4 relative to H3 in the tetramer fraction of the H3 octamer, rather than a real difference in octamer stoichiometry (Figure 3-4).

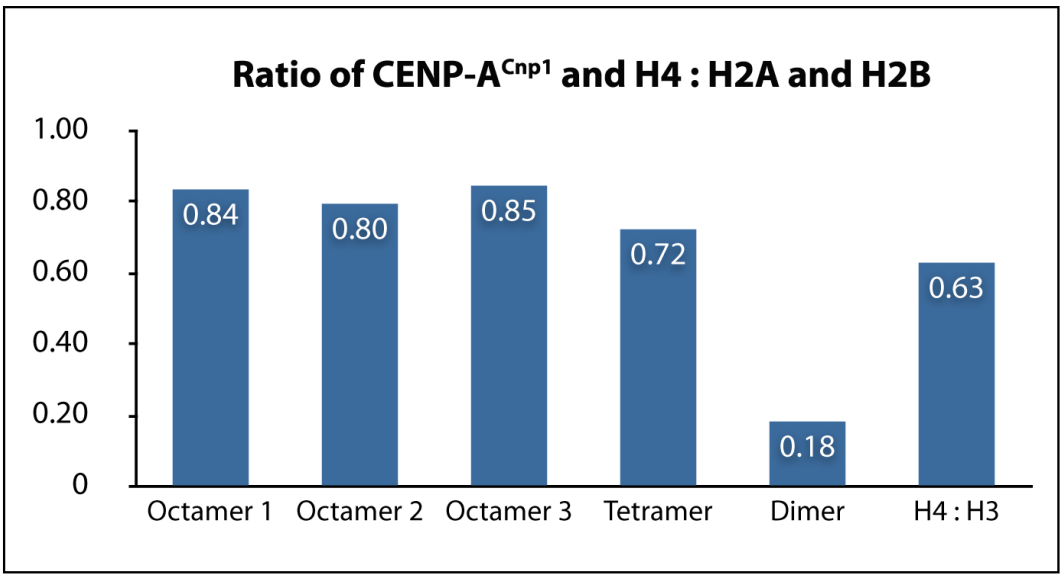


Figure 3-4 Quantifying histones in octamer fractions

The ratios of (CENP-A^{Cnp1} / H4) : (H2A / H2B) were calculated by densitometric analysis for each of the Coomassie stained fractions in Figure 3-3 B. The ~ 0.2 fold decrease in (CENP-A^{Cnp1} / H4) : (H2A / H2B) observed in the supposedly octameric fractions is thought to be a consequence of the relatively poor Coomassie staining of H4. Quantification of the ~ 0.4 fold reduction in H4 staining relative to H3 staining (shown in the final lane) was calculated from the tetramer fraction of Figure 3-3a.

To assemble the H3 and CENP-A^{Cnp1} histone complexes into nucleosomes the synthetic "601.3" DNA (Anderson and Widom, 2001) was chosen because its incorporation into nucleosomes has been well characterised previously and is known to readily reconstitute stable nucleosomes (Vasudevan *et al.*, 2010). 601 DNAs of various defined lengths (typically 147 bp) were produced by PCR from a pUC19 plasmid containing the 601.3 sequence (from Tom

Owen-Hughes, Dundee) and purified by electro-elution from a native PAGE gel. Reconstitution of nucleosomes proceeded by mixing DNA resuspended in 2 M salt at a roughly 1:1 molar ratio with histone octamer and thorough dialysis from 2 M to 10 mM salt over a period of 24 h. Completed reconstitutions were run on native PAGE gels to assess the extent of DNA incorporation into nucleosomes and to purify nucleosome particles from unincorporated free DNA and non-nucleosomal reconstitution intermediates by electro-elution (Figure 3-5). Whilst attempts were made to identify the protein components of these nucleosomes by MALDI mass spectrometry insufficient numbers of peptides were obtained for accurate identification and so this remains to be repeated. From quantification of the ratio of DNA present in the free DNA and nucleosome bands the efficiency of DNA incorporation into nucleosomes in a given reconstitution was typically between 50 – 80 % for both H3 and CENP-A^{Cnp1} nucleosomes, largely dependent on the exact ratio of DNA to protein in the reconstitution. Notably, both H3 and CENP-A^{Cnp1} *in vitro* nucleosomes exhibit similar migration patterns to those of nucleosomes reconstituted from *X. laevis* histones (a kind gift from Professor Tom Owen-Hughes) and also with previously published migration patterns of purified *ex vivo* nucleosomes (Camerini-Otero and Felsenfeld, 1977), strongly suggesting that these reconstitutions led to the formation of true nucleosome-like particles.

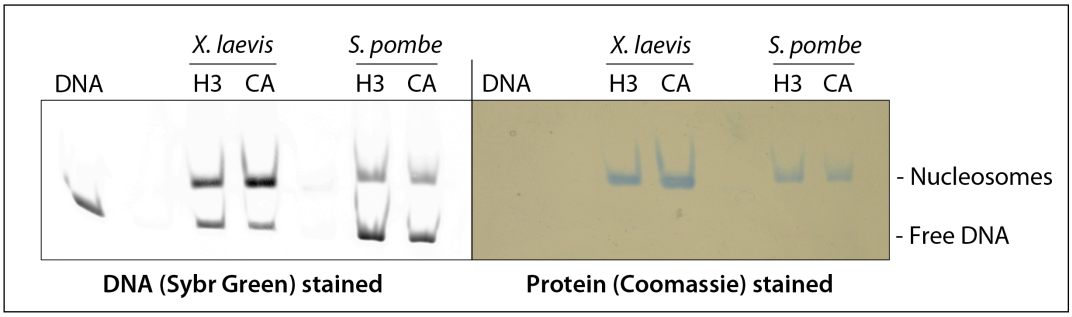


Figure 3-5 – Native PAGE of reconstituted nucleosomes from *S. pombe* and *X. laevis*

Reconstituted *S. pombe* H3 and CENP-A^{Cnp1} (CA) nucleosomes run on a 5 % native PAGE gel are shown alongside free DNA and similarly reconstituted *X. laevis* nucleosomes for comparison. The native PAGE gel was post stained with Sybr Green 1 (for DNA) and Coomassie (for protein). Reconstitutions typically result in mixture of nucleosomes (upper, protein-containing bands) and free DNA (lower bands). *X. laevis* nucleosomes were a kind gift from Professor Tom Owen-Hughes.

Assembled nucleosomes have typical nucleosomal diameters

The mobility of H3 and CENP-A^{Cnp1} nucleosomes through a native PAGE gel suggested that both had similar hydrodynamic radii (Figure 3-5), but to confirm the monomeric aggregation state of the samples dynamic light scattering was used. Dynamic light scattering measures the fixed angle scattering of laser light through a particle solution to infer particle diameters and molecular weights by modelling their rates of Brownian motion. Measurements of both H3 and CENP-A^{Cnp1} nucleosome samples were classified as monodisperse (not shown) which, when combined with their comparative mobility through native PAGE gels, suggests these *in vitro*

prepared CENP-A^{Cnp1} nucleosomes have a similar hydrodynamic radius to that of H3 nucleosomes.

The estimated diameters and molecular weights reported by dynamic light scattering measurements correspond to the Brownian motion of particles modelled as spheres and so, because nucleosomes are closer to wide cylinders than spheres this modelling introduces some degree of error. Nevertheless, the estimated diameters were 10.1 nm (± 1.6 nm) for H3 nucleosomes and 11.7 nm (± 1.9 nm) for CENP-A^{Cnp1}, with no particles above 1 nm in diameter present in buffer only samples (not shown), in excellent agreement with the 10 nm diameters observed in the crystal structures of both H3 and CENP-A nucleosomes (Luger *et al*., 1997; Tachiwana *et al*., 2011a) (Figure 3-6). Whilst rather more prone to modelling error, average molecular weights of 224 kDa for CENP-A^{Cnp1} and 166 kDa for H3 nucleosomes were also similar to their expected values of ~ 200 kDa. Whilst not a significant difference, the slight increase in apparent diameter of the CENP-A^{Cnp1} nucleosome sample is reminiscent of the increased diameter observed by small angle X-ray scattering (SAXS) for human CENP-A nucleosomes (Tachiwana *et al*., 2011a). A likely explanation for this difference (in both the dynamic light scattering and SAXS data) is the propensity of CENP-A nucleosomes to wrap ~ 13 bp less DNA (~ 120 bp DNA wrapped by CENP-A, compared to ~ 147 bp for H3) at the DNA entry / exit sites of the CENP-A^{Cnp1} nucleosome, which would increase the effective hydrodynamic radius of the CENP-A nucleosome were these unwrapped DNA arms to extend out from the nucleosome core as proposed (Tachiwana *et al*., 2011a).

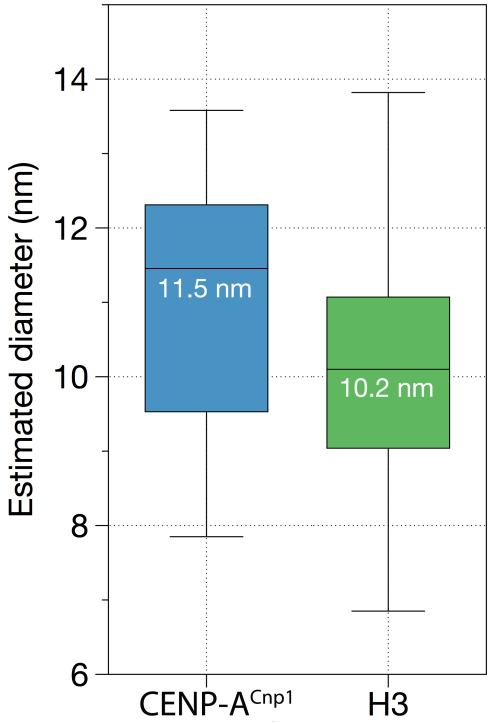


Figure 3-6 - Estimated diameters of CENP-A^{Cnp1} and H3 nucleosomes by dynamic light scattering

Dynamic light scattering was used to estimate the diameter of CENP-A^{Cnp1} and H3 *in vitro* nucleosomes in solution. The central lines through each box represent the median, the outer edges mark the first and third quartiles and the whiskers mark the range of three independent reconstitutions for each nucleosome type. The mean diameters (10.1 nm and 11.7 nm for H3 and CENP-A^{Cnp1} nucleosomes, respectively) are similar to the 10 nm diameter observed in the crystal structures of both *Xenopus* H3 and Human CENP-A recombinant nucleosomes (Tachiwana *et al.*, 2011a) (Luger *et al.*, 1997).

CENP-A^{Cnp1} nucleosomes wrap less DNA than H3 nucleosomes

The slight increase in the apparent diameter of CENP-A^{Cnp1} nucleosomes relative to H3 nucleosomes revealed by dynamic light scattering (Figure 3-6)

suggests that CENP-A^{Cnp1} nucleosomes may wrap slightly less DNA than H3 nucleosomes. This is in keeping with the observations in a range of species that CenH3 nucleosomes show a relaxed wrapping of ~ 7 – 30 bp of DNA at the entry / exit sites relative to H3 nucleosomes (Conde e Silva *et al*., 2007; Dalal *et al*., 2007; Tachiwana *et al*., 2011a). To determine how much DNA these *in vitro S. pombe* H3 and CENP-A^{Cnp1} nucleosomes wrap, nucleosomes were assembled onto long (> 5 kb) fragments of genomic DNA at low density, digested with micrococcal nuclease (MNase) to predominantly mononucleosome fragments (Figure 3-7) and the DNA purified by proteinase K digestion and phenol / chloroform extraction for loading onto an agarose gel. To accurately size the nucleosome-protected DNA the mononucleosome sized DNA band was extracted from the gel and analysed on a Bioanalayser (Agilent). The results of four separate reconstitutions show CENP-A^{Cnp1} nucleosomes protected an average of ~ 10 bp less DNA than H3 nucleosomes (Figure 3-8), mimicking the DNA wrapping of CENP-A nucleosomes *in vivo*. The exact length of DNA protected varied between repeats, which may support the transient unwrapping or "breathing" of DNA observed in human CENP-A nucleosomes (Conde e Silva *et al*., 2007) but could also represent slight differences in the extent of MNase digestion between samples.

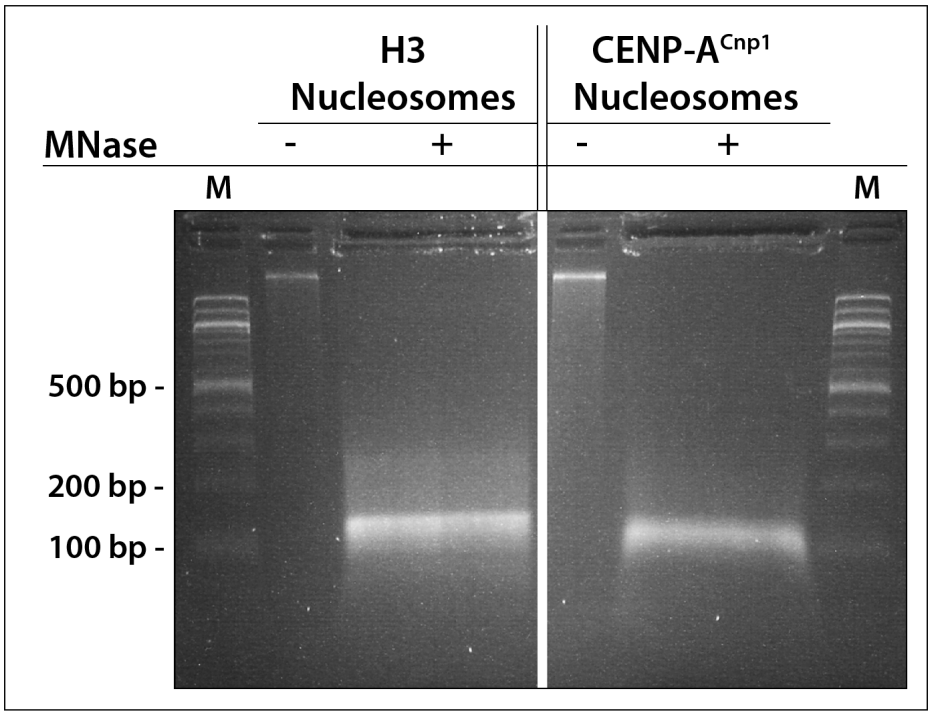


Figure 3-7 – MNase digest of nucleosome arrays

To determine the amount of DNA protected from micrococcal nuclease (MNase) digestion by nucleosome types, H3 and CENP-A^{Cnp1} nucleosomes were reconstituted onto long (> 5 kb) fragments of genomic DNA at low density and digested with MNase to predominantly mononucleosomes. The DNA was purified by proteinase K digestion and phenol / chloroform extraction before loading onto an agarose gel. The mononucleosome band (between 100 – 200 bp) was extracted and analysed on a Bioanalyser for accurate sizing (Figure 3-8).

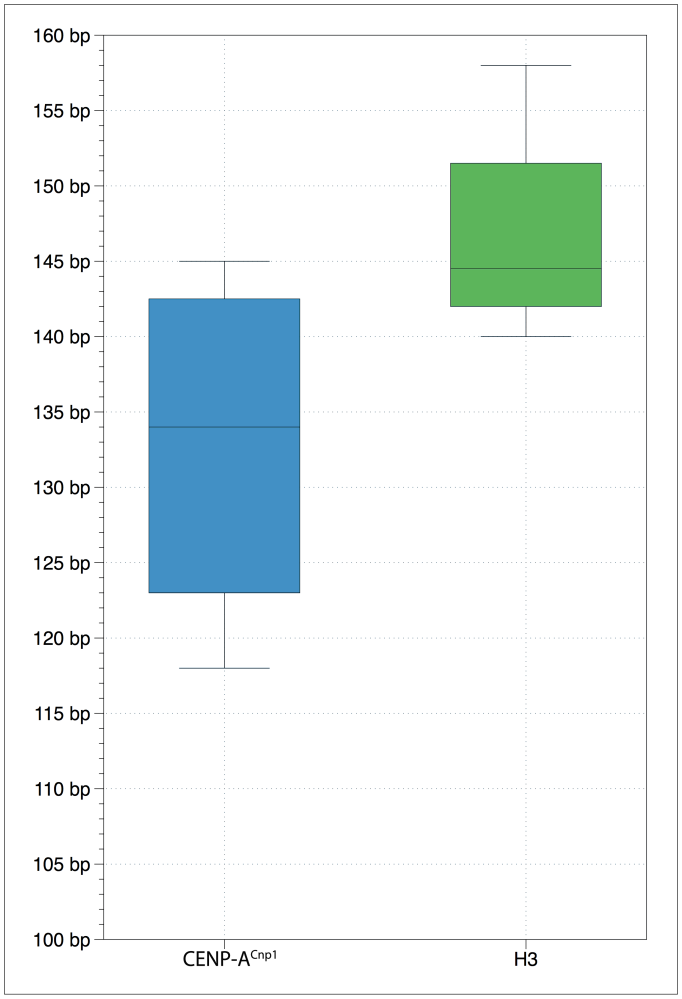


Figure 3-8 - CENP-A^{Cnp1} nucleosomes protect shorter lengths of DNA from MNase than H3 nucleosomes

Nucleosomes assembled onto long (> 5 kb) fragments of genomic DNA were treated with micrococcal nuclease to digest unprotected DNA. The resulting DNA was run on an agarose gel (Figure 3-7) to isolate mononucleosomal fragments, which were extracted and run on the high resolution Bioanalyzer (Agilent) to accurately determine fragment sizes. The distributions of the observed fragment lengths are shown. The central line in each box represents the median fragment length, with the outer edges marking the first and third quartile ranges. The whiskers indicate the minimum and maximum values.

When reconstituted onto 147 bp of DNA, *in vitro* CENP-A^{Cnp1} nucleosomes showed a slight reduction in native PAGE mobility relative to H3 nucleosomes (Figure 3-9), which is consistent with the idea that CENP-A^{Cnp1} (but not H3) nucleosomes may have small lengths of unbound DNA extending from the core particle. If this extended region of DNA were the cause of the reduced mobility of CENP-A^{Cnp1} nucleosomes in a native PAGE gel then one prediction would be that removing this "flapping" DNA would correspondingly increase nucleosome mobility. To test this H3 and CENP-A^{Cnp1} nucleosomes were reconstituted onto either 250, 147 or 122 bp of 601 DNA and their mobilities compared on a native PAGE gel. When H3 and CENP-A^{Cnp1} nucleosomes were wrapped with just 122 bp of DNA, essentially removing the DNA predicted to be poorly bound by CENP-A^{Cnp1} nucleosomes, CENP-A^{Cnp1} mobility was now equivalent to that of H3 nucleosomes (Figure 3-9). When wrapped with 250 bp of DNA the overall mobility was reduced but again there was also no obvious difference between H3 and CENP-A^{Cnp1} nucleosomes, presumably because the supposed 10 bp of extra unbound DNA in CENP-A^{Cnp1} nucleosomes contributes relatively little to increasing mobility when both H3 and CENP-A^{Cnp1} nucleosomes already have ~ 100 bp of unwrapped DNA. Taken together with the increased sensitivity of CENP-A^{Cnp1} nucleosomes to MNase, these data suggest *in vitro* CENP-A^{Cnp1} nucleosomes wrap ~ 10 bp less DNA than H3 nucleosomes (~ 134 bp for CENP-A^{Cnp1} and ~ 144 bp for H3) and is consistent with both the *in vitro* results of Tachiwana *et al* (2011) and *in vivo* data of Dalal *et al* (2007), validating these *in vitro S. pombe*

nucleosomes for use in further experiments as mimics of their *in vivo* counterparts.

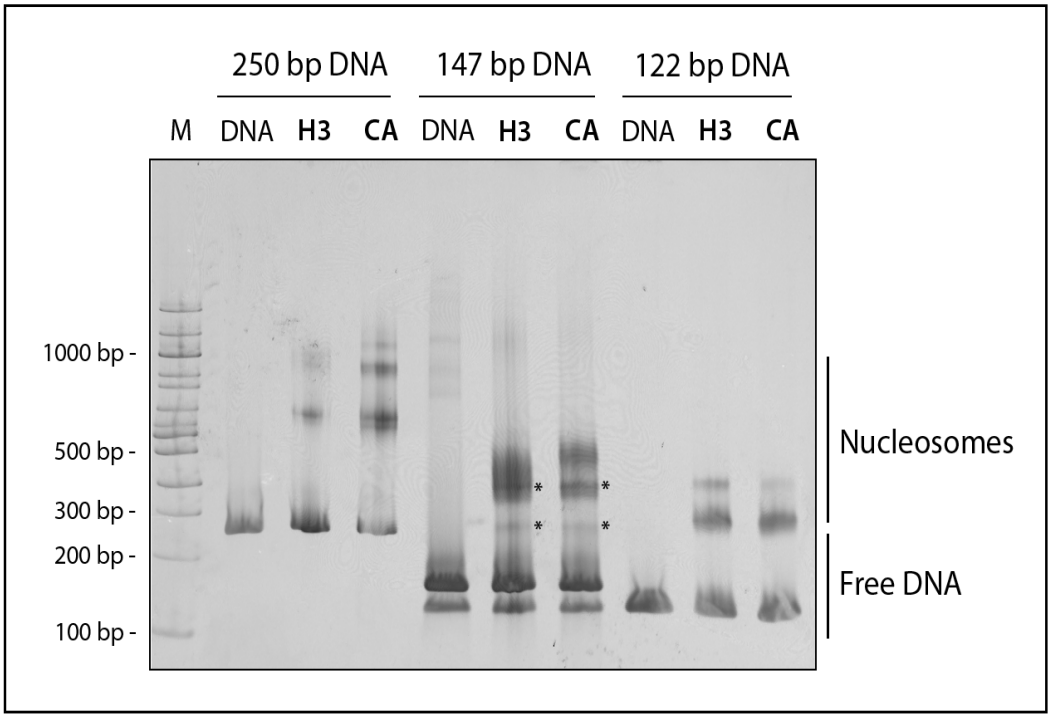


Figure 3-9 - Relative mobility of H3 and CENP-A^{Cnp1} nucleosomes differs with wrapped DNA length

H3 and CENP-A^{Cnp1} (CA) nucleosomes were reconstituted onto either 250 bp, 147 bp, or 122 bp 601 DNA and run on a native PAGE gel stained for DNA with Sybr Green 1. Relative to H3, CENP-A^{Cnp1} nucleosomes show slightly retarded mobility when wrapped by 147 bp of DNA but this is not the case for nucleosomes wrapped by either 250 bp or 122 bp. Mobility of both H3 and CENP-A^{Cnp1} nucleosomes is reduced on 250 bp DNA (relative to mobility on 147 bp DNA) and increased on 122 bp DNA. The 147 bp DNA was contaminated with a smaller (122bp) DNA species that also formed nucleosomes (the positions of which are marked with an asterisk).

Cross-linking of CENP-A^{Cnp1} and H3 nucleosomes indicates they form octameric complexes

In the hemisome model of CENP-A nucleosomes, each nucleosome contains a single copy of histones H2A, H2B, CENP-A and H4, rather than the two copies of each histone found in the octameric CENP-A model and H3 nucleosome. If the hemisome model is correct for *S. pombe* CENP-A^{Cnp1} nucleosomes then the CENP-A^{Cnp1} nucleosome would have a hemisomal molecular weight of ~ 53 kDa (not including the molecular weight of DNA). If the CENP-A^{Cnp1} nucleosome were octameric it would have a molecular weight of 106 kDa (without DNA), similar to the ~108 kDa of an octameric H3 nucleosome. Dynamic light scattering had estimated the molecular weight of CENP-A^{Cnp1} nucleosomes (including DNA) to be ~ 224 kDa but this was an estimate. To test the molecular weight of CENP-A^{Cnp1} and H3 nucleosomes fixation titrations were made using the primary amine cross-linker BS(PEG)$_5$ to increasingly cross-link the histones within the nucleosome core. Samples were then run on SDS-PAGE gels and the molecular weight of the cross-linked histone complexes could be directly determined. BS(PEG)$_5$ was chosen as the fixative since the ~ 22 Å PEG spacer between NHS esters is known to cross-link lysine residues over a relatively large distance, increasing the likelihood of capturing the entire histone core complex. The results clearly show the step increases in molecular weight as histones are increasingly fixed together (Figure 3-10) and starting from a 400-fold molar excess of fixative the molecular weight of the maximally fixed complex for both CENP-A^{Cnp1} and H3 nucleosomes corresponds to the octameric species

(~100 kDa), strongly suggesting both these CENP-A^{Cnp1} and H3 *in vitro* nucleosomes are octameric.

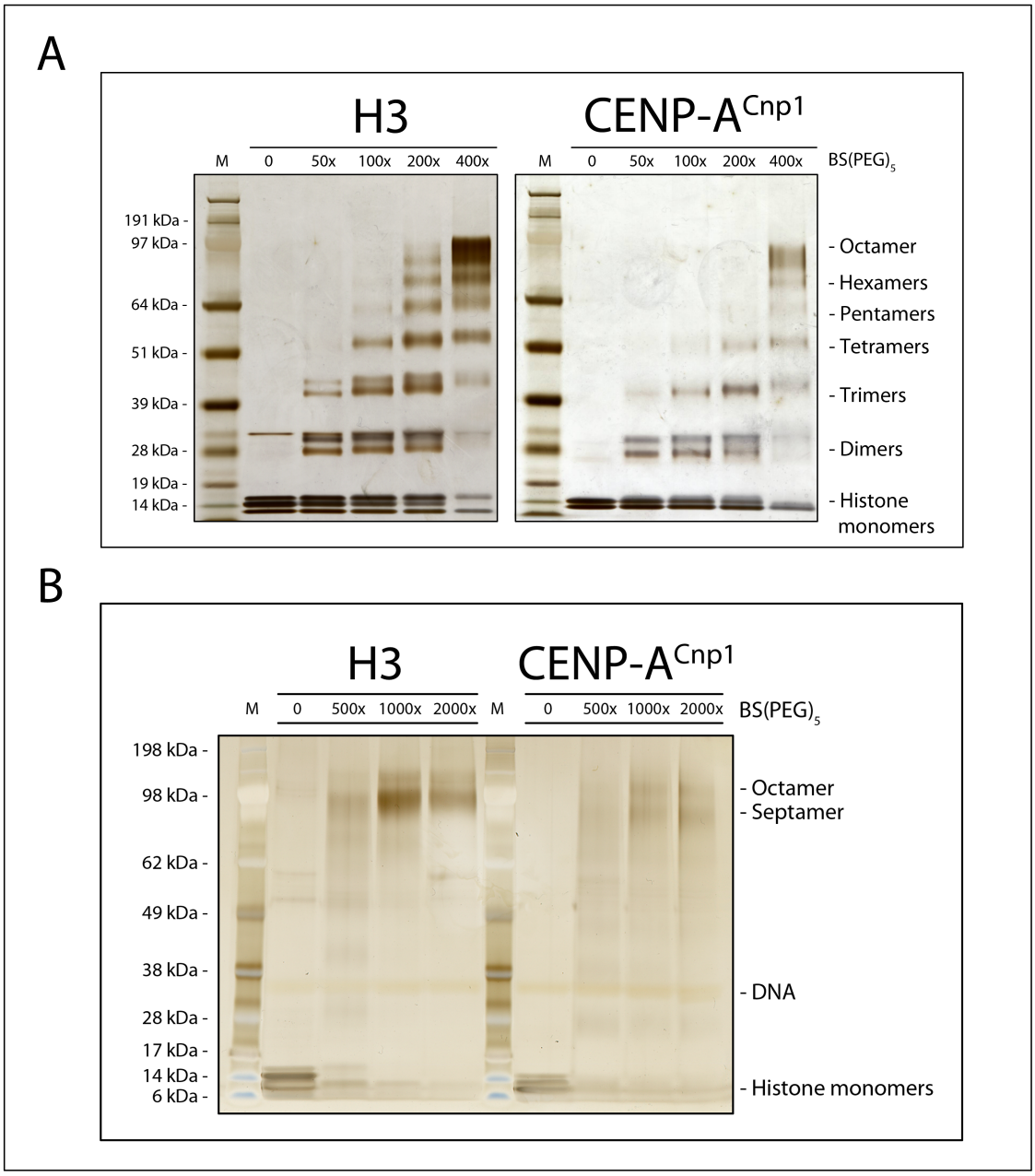


Figure 3-10 – Nucleosome cross-linking titration

H3 and CENP-A^{Cnp1} nucleosomes were fixed using a titration of the primary amine cross-linker BS(PEG)$_5$. Samples were then run on a SDS-PAGE gel and silver stained to determine their size. Both H3 and CENP-A^{Cnp1}

nucleosomes cross-link up to octameric complexes, with each of the intermediate complexes also visible (a). The concentrations of $BS(PEG)_5$ written above each lane refer to the molar excess of $BS(PEG)_5$ relative to nucleosome concentration. Increasing the cross-linker concentration a further 4-fold continued to show cross-linking at octameric sizes (b).

At the highest concentrations of cross-linker two bands are visible (Figure 3-10 B) and the upper band, particularly in the H3 sample, appears to be above octameric size (although it is difficult to accurately distinguish very large sizes on these gels). It was considered that for CENP-A^{Cnp1} nucleosomes at least, this may represent cross-linking of two or three hemisome particles so that they would appear above octameric and above. However, this was considered to be unlikely for three reasons, 1) because the low (nM) concentration of nucleosomes in these samples makes such interactions uncommon, 2) the identification of intermediate complexes between the hemisome and octamer sizes would not be expected from an interaction between two already cross-linked hemisomes (Figure 3-10 A) and 3) because a similar (even stronger) pattern was observed for H3 nucleosomes. Instead it seemed most likely that these bands represent the cross-linking of aggregated nucleosomes carried over from the reconstitution. To test this possibility, maximally cross-linked samples were run on native PAGE gels (Figure 3-11). These gels clearly show the presence of typical reconstitution aggregates above nucleosome sizes, particularly so for the H3 reconstitution, that would account for the large (~ 130 kDa) band on the cross-linked SDS-PAGE gel. Meanwhile, the monomeric nature of these maximally-fixed

nucleosomes demonstrate that octameric CENP-A^{Cnp1} nucleosomes are the result of true octameric complexes, not the over-fixation of two hemisomal nucleosomes (Figure 3-11).

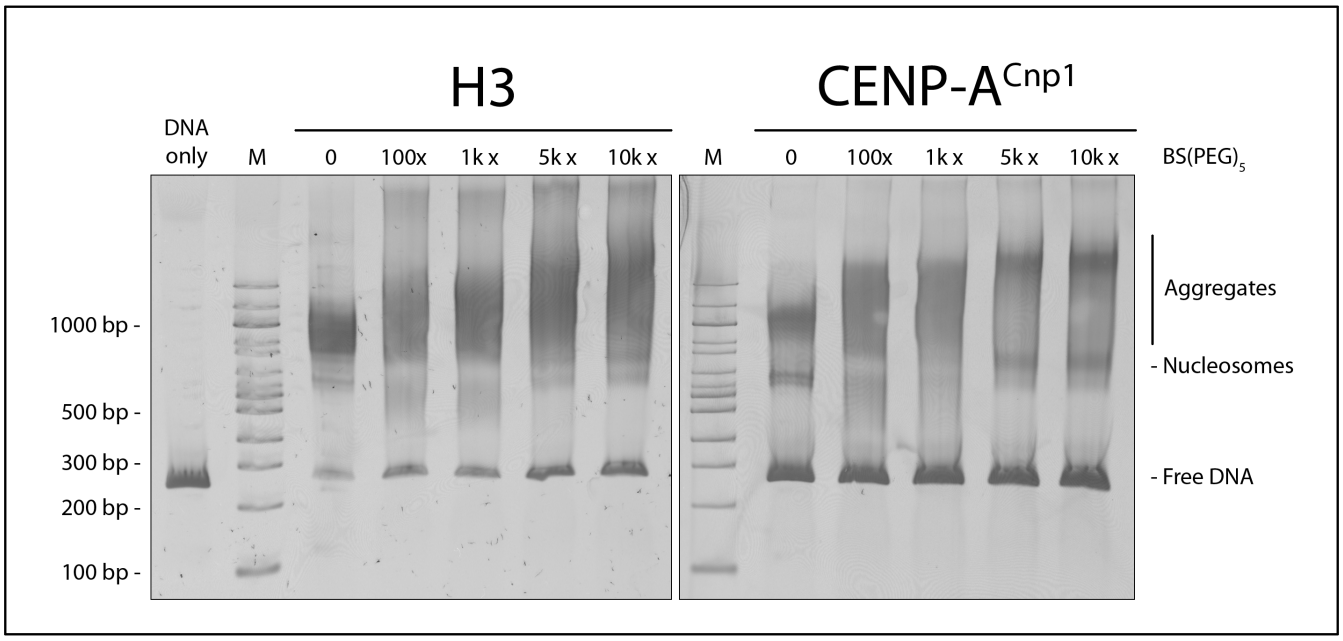


Figure 3-11 - Native PAGE mobility of cross-linked nucleosomes

H3 and CENP-A^{Cnp1} nucleosomes were cross-linked with a large molar excess of BS(PEG)$_5$ and run on a native PAGE gel, post-stained with Sybr Green 1 DNA stain. In both reconstitutions the predominant complexes are reconstitution-induced aggregates (~1000 bp) that become increasingly aggregated upon fixation. The nucleosome band (~ 600 bp), however, remains monomeric.

Mass spectrometry of hybrid heavy/light labelled nucleosomes suggest two copies of each histone

A second fixation experiment coupled nucleosome fixation and mass spectrometry to test whether these CENP-A^{Cnp1} nucleosomes contained one or two copies of each histone. Each histone was prepared using either "heavy" (N15, C13) labelled Arginine and Lysine or "light" unlabelled amino acids. Nucleosomes were prepared using an equal ratio of heavy and light labelled versions of each histone so that, if nucleosomes were octameric and contained two copies of each histone, a proportion of nucleosomes should contain both one heavy, and one light copy of a given histone. After nucleosome fixation at a 1,000-fold excess of $BS(PEG)_5$ to achieve maximal fixation samples were run on an SDS-PAGE gel (Figure 3-12) and the octameric band digested "in-gel" with trypsin and analysed by mass spectrometry. Uniquely, this isotope labelling allowed for the distinction to be made between inter-histone cross-links (within the same histone) and intra-histone cross-links (between two copies of the same histone), where cross-links between a heavy and a light version of the same histone would strongly suggest there were two copies of that histone within a nucleosome, a result that would be incompatible with a hemisome model.

The results of this experiment are summarised in Figure 3-13, with the full list of peptides available in Appendix 1. For both the H3 and CENP-A^{Cnp1} recombinant nucleosome samples heavy : light cross-links were identified between H4 histones, suggesting both H3 and CENP-A^{Cnp1} nucleosomes contained two copies of H4 (Figure 3-13). Cross-links between a heavy and light version of the other histones were lacking from CENP-A^{Cnp1} nucleosome

samples but were observed between H3 : H3 and H2A : H2A for the H3 nucleosome sample. That H3 cross-links are apparent when CENP-A^{Cnp1} cross-links are not is surprising as the crystal structures of these nucleosomes predict both H3 and CENP-A^{Cnp1} to contain lysine residues in essentially the same structural regions within their four-helix bundles. One cross-link was observed between a heavy and a light copy of CENP-A^{Cnp1} (not shown) but both residues were in the amino-terminal tails that extend from the nucleosome, increasing the likelihood that it was the result of an inter-nucleosomal cross-link, so this was excluded. Importantly, the absence of cross-links does not rule out their existence, it may therefore be useful to test a different cross-linking agent (see next section), or may simply require repetition of the experiment to collect CENP-A^{Cnp1} : CENP-A^{Cnp1} cross-links. That both H3 and CENP-A^{Cnp1} nucleosomes appear to contain two copies of H4 argues against a hemisome stoichiometry for these CENP-A^{Cnp1} nucleosomes.

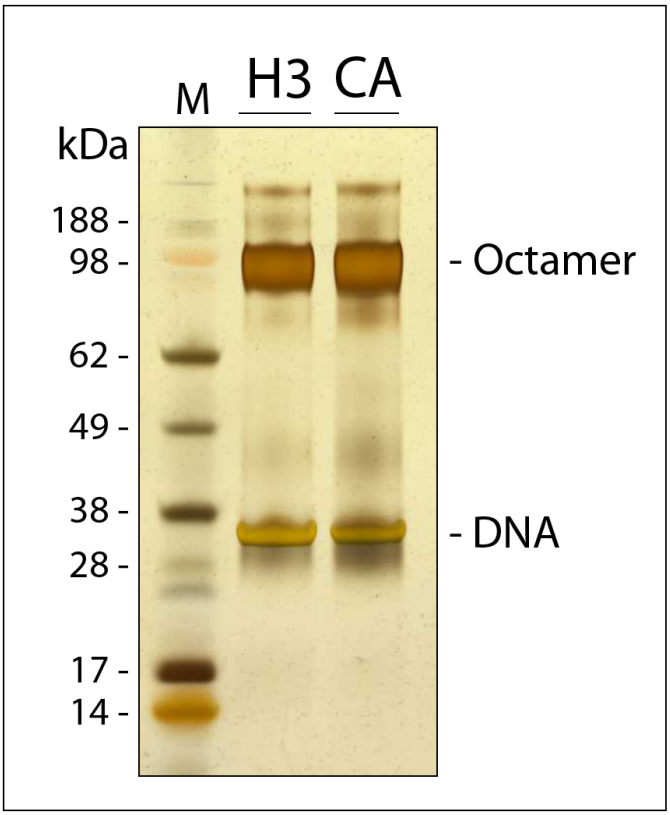


Figure 3-12 - Cross-linking of heavy : light nucleosomes

Two versions of each histone were prepared, one heavy labelled with N15 and C13 Arginine and Lysine and the other unlabelled. Nucleosomes made from an equal ratio of heavy and unlabelled H2A, H2B, H3 and H4 for the H3 nucleosome sample, and H2A, H2B, CENP-A^{Cnp1} and H4 for the CENP-A^{Cnp1} (CA) nucleosome sample, assembled onto 147 bp 601 DNA. The nucleosomes were cross-linked with a 1,000 -fold molar excess of the primary amine cross-linker BS(PEG)$_5$ and the fixed complexes loaded onto an SDS-PAGE gel to show the level of cross-linking, which corresponds to the molecular weight of an H3 or CENP-A^{Cnp1} octamer (respectively). Octamer bands were then cut from the gel and analysed by mass spectrometry (see text for details).

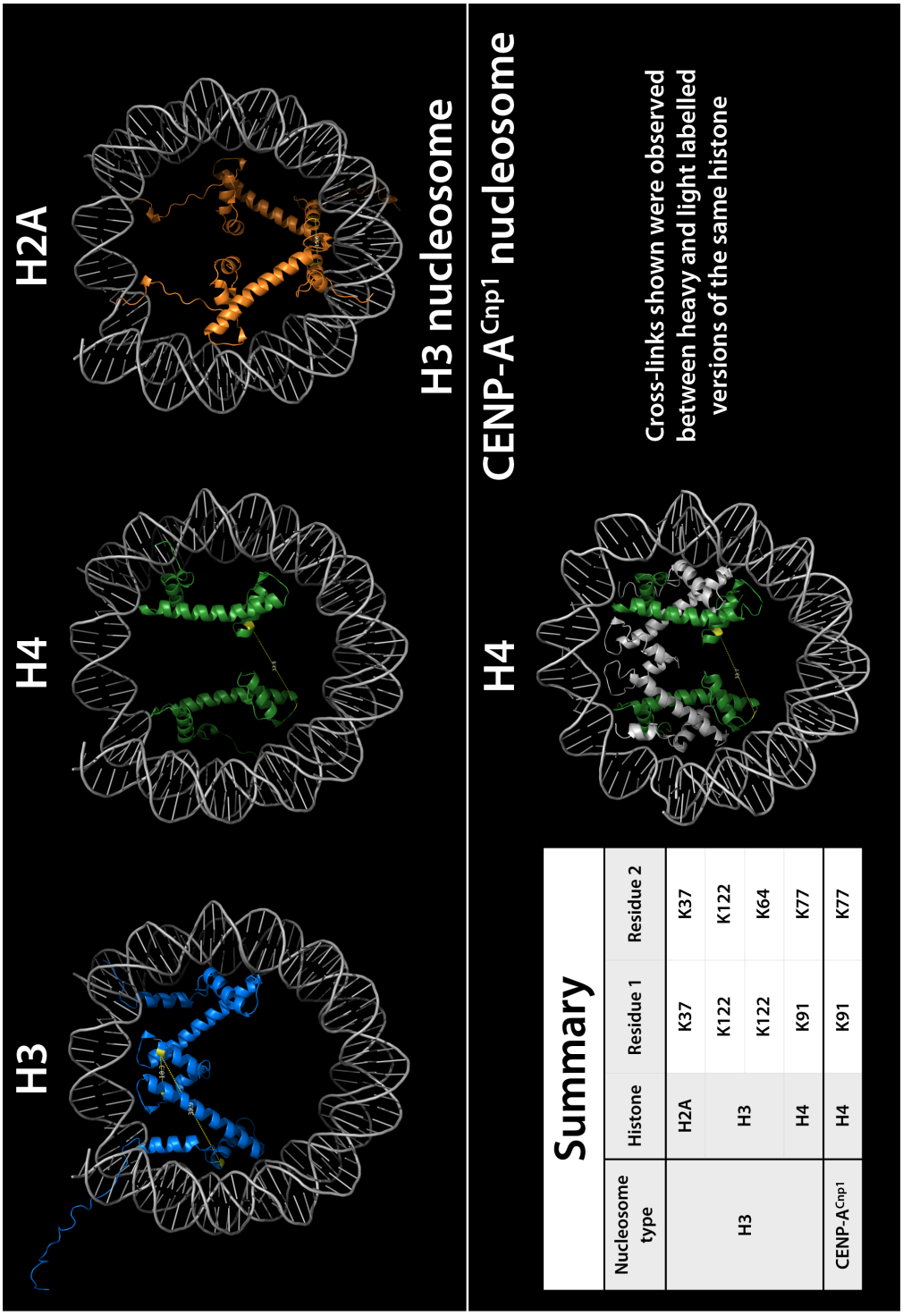


Figure 3-13 - Map of cross-links between two copies of the same histone

H3 and CENP-A^{Cnp1} nucleosomes were prepared from an equal mixture of heavy and light-labelled versions of each histone and cross-linked with BS(PEG)$_5$. Shown here are the cross-links identified by mass spectrometry

as being between one heavy labelled histone (labelled with N15 and C13 Arginine and Lysine) and one unlabelled copy of the same histone type, suggesting two copies of that histone are contained within the nucleosome. Such cross-links were observed for H3, H4 and H2A in H3 nucleosomes (upper panel) and for H4 in CENP-A^{Cnp1} nucleosomes (lower panel). In the lower panel the CENP-A^{Cnp1} histones are included (shaded grey) for orientation only. Cross-links were mapped onto the crystal structure of the H3 nucleosome (3AFA) (Tachiwana *et al.*, 2010) and human CENP-A nucleosome crystal structure (3AN2) (Tachiwana *et al.*, 2011a) respectively for H3 and CENP-A^{Cnp1} nucleosome cross-links. A full list of the peptides identified is available in Appendix 1.

CENP-A^{Cnp1} nucleosomes contain two adjacent CENP-A^{Cnp1} proteins

An octameric CENP-A^{Cnp1} nucleosome should contain two copies of CENP-A^{Cnp1} whilst a hemisomal nucleosome should only contain a single CENP-A^{Cnp1}. As no lysine : lysine peptide cross-links were observed within the histone-fold domains of two CENP-A^{Cnp1} proteins in the above cross-linking / mass spectrometry experiment an alternative cross-linking method was used. Nucleosomes were incubated in an oxidising environment with the aim of forming a disulphide bond between the two cysteine residues expected to be located with the four-helix bundles of both H3 and CENP-A^{Cnp1} octameric nucleosomes (C111 and C96, respectively) (Figure 3-14). This method has previously been successful used for analysis of both chicken (Camerini-Otero and Felsenfeld, 1977; Gould *et al.*, 1980) and budding yeast (Bowman *et al.*, 2011) H3 nucleosomes and in both H3 and CENP-A^{CID} nucleosomes from flies (Zhang *et al.*, 2012).

Following the method outlined in Gould *et al*, 1980, *in vitro* prepared H3 and CENP-A^{Cnp1} nucleosomes at a concentration of 200 nM were oxidised over 24 h and run on non-reducing SDS-PAGE gels, visualised by silver staining. A disulphide bond was apparent in the CENP-A^{Cnp1} nucleosome sample but only very faintly in the H3 sample (Figure 3-15). For both samples the extent of cross-linking was low, but was similar to cross-linking efficiencies seen in other systems (Camerini-Otero and Felsenfeld, 1977). As disulphide bonds typically span just ~ 2 Å, rather than a true representation of the percentage of H3 / CENP-A^{Cnp1} dimers within the population the most likely explanation for this low efficiency of disulphide formation is the distance (expected to be ~ 8 Å in *S. pombe* based on homology modelling) between the two cysteines in fission yeast, which then requires flexibility in the structure to allow the formation of these bonds. To help span the ~ 8 Å distance between the two H3 / CENP-A^{Cnp1} cysteines a cysteine-specific cross-linker containing a spacer arm of 10.2 Å, BMDB (Thermo Scientific), was trialled but no cross-links were observed (not shown). One explanation for this may be that the 280 Da cross-linker was unable to access the buried cysteines, perhaps because the spacer arm length of 10.2 Å was too bulky for the ~ 8 Å distance between cysteines. With Drosophila nucleosomes the oxidation method have provided much better cross-linking efficiencies (Zhang *et al*., 2012), suggesting it may still be possible to improve the oxidation cross-linking efficiency in these reactions with increased incubation time, or else by moving the cysteines so they were closer together. Nevertheless, CENP-A^{Cnp1} is the only histone within the CENP-A^{Cnp1} nucleosome to contain cysteine residues, so the presence of a CENP-

A^{Cnp1} : CENP-A^{Cnp1} specific cross-link indicates that at least a proportion of *in vitro* CENP-A^{Cnp1} nucleosomes contain two copies of CENP-A^{Cnp1}.

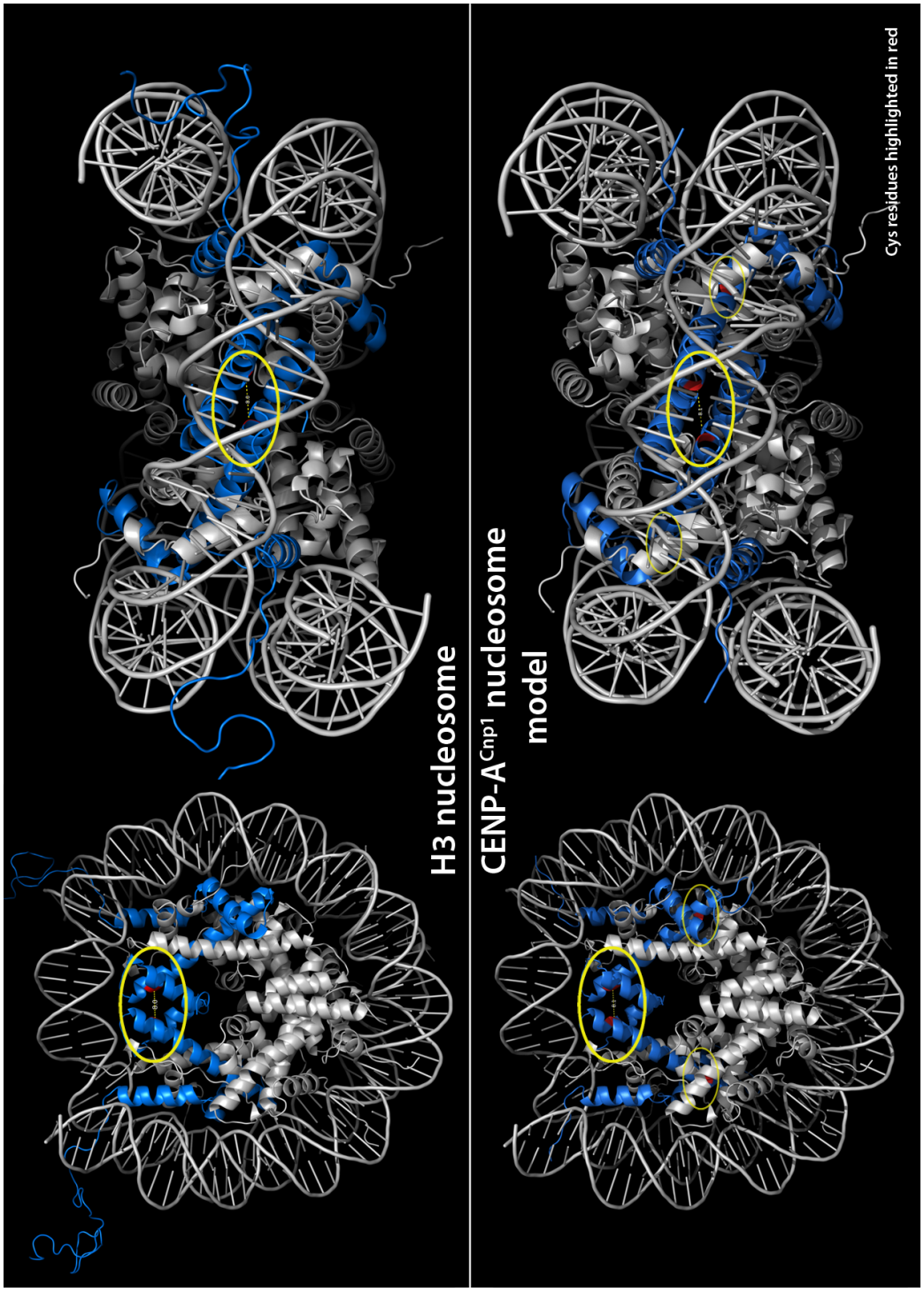


Figure 3-14 – Location of cysteine residues

The location of the cysteine residues in H3 (upper panel) and CENP-A^{Cnp1} (lower panel) nucleosomes are highlighted in red and circled by yellow ellipses for clarity. H3 contains only a single cysteine residue, whilst CENP-A^{Cnp1} contains two, one of which is in a structurally similar location to that of

the H3 cysteine. Homology based modelling predicts the location of cysteine residues within the four-helix bundles of both H3 and CENP-A^{Cnp1} nucleosomes (both ~ 8 Å apart) that could be used to form disulphide bonds that specifically link closely interacting CENP-A^{Cnp1} (and H3) histones. If CENP-A^{Cnp1} nucleosomes are hemisomes, the single copy of CENP-A^{Cnp1} would not be able to form these disulphide bonds. The H3 nucleosome structure shown here is actually the crystal structure of the Xenopus H3.1 nucleosome (Davey *et al.*, 2002) but was used for of the very high degree of sequence conservation with *S. pombe* H3 nucleosomes. The CENP-A^{Cnp1} nucleosome model uses the same H3 crystal structure but has replaced H3 with a CENP-A^{Cnp1} histone structure created by homology based modelling against the crystal structure of the human CENP-A nucleosome (Tachiwana *et al.*, 2011a).

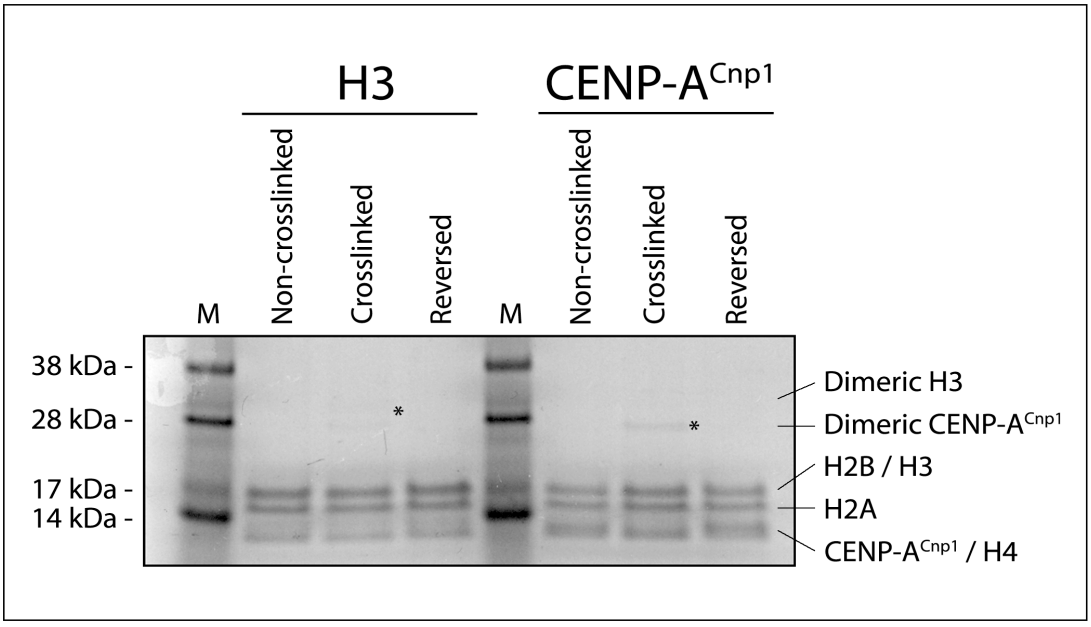


Figure 3-15 - Cysteine cross-linking of H3 and CENP-A^{Cnp1} nucleosomes
Non-reducing SDS-PAGE gel of H3 and CENP-A^{Cnp1} nucleosome samples cross-linked by disulphide bond formation. The disulphide-specific band visible in the CENP-A^{Cnp1} sample can only form between the two C96 residues of a CENP-A^{Cnp1} dimer (see Figure 3-14). Reversal of the cross-

linking was achieved by incubation with 10 mM DTT before loading on the gel.

Discussion

In vitro assembled nucleosomes have characteristics of *in vivo* nucleosomes and are octameric

In this chapter untagged, recombinant *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes were prepared *in vitro* for the first time, demonstrating that CENP-A^{Cnp1} nucleosome-like particles can form without the need for non-nucleosomal components (such as Scm3). These *in vitro* nucleosomes showed a number of properties identified for CenH3 nucleosomes *in vivo*, namely that they have canonical diameters and predicted molecular weights (Figure 3-6) (Tachiwana *et al*., 2011a) and protect less DNA than H3 nucleosomes (Figure 3-8) (Conde e Silva *et al*., 2007; Dalal *et al*., 2007), thus suggesting that these nucleosomes act as *in vitro* mimics of their *in vivo* counterparts.

The stoichiometric composition of the CENP-A nucleosome is difficult to determine unambiguously, as evidenced by the absence of definitive biochemical data to date. This is particularly true for distinguishing between hemisomes and octamers as they would contain the same components at the same ratios. Determination of nucleosome size was chosen as the most appropriate method to distinguish between octameric and hemisomal stoichiometries and classically there are two paths to achieve this, cross-linking and sedimentation rate (hydrodynamic mobility), both of which have been employed here in a number of guises.

In keeping with the roughly equivalent hydrodynamic diameters of ~ 10 nm for H3 nucleosomes and 11.5 nm for CENP-A^{Cnp1} nucleosomes that were observed by dynamic light scattering, the mobilities of CENP-A^{Cnp1} and H3 nucleosomes through native PAGE gels were also found to be equivalent (Figure 3-9), suggesting that recombinant CENP-A^{Cnp1} nucleosomes share the octameric stoichiometries of H3 nucleosomes. Using a different biochemical approach, the stoichiometry of subunits within recombinant H3 and CENP-A^{Cnp1} nucleosomes was also determined by three methods of cross-linking. Both H3 and CENP-A^{Cnp1} nucleosomes were found to cross-link up to the expected molecular weight of octameric complexes (Figure 3-10), consistent with the predicted molecular weight of the unfixed complexes determined by dynamic light scattering. Supporting this result, a more detailed cross-linking approach used mixed heavy and light labelled nucleosomes to identify cross-links linking two copies of each histone within a nucleosome, strongly suggesting both H3 and CENP-A^{Cnp1} nucleosomes contain more than one copy of H4 (Figure 3-13). Confirmation for the presence of two copies of CENP-A^{Cnp1} within a nucleosome was then demonstrated by the formation of disulphide bridges that specifically link two directly interacting CENP-A^{Cnp1} proteins (Figure 3-15). Taken together, these experiments strongly suggest these *in vitro* prepared nucleosomes have octameric stoichiometries. The question then is whether CENP-A^{Cnp1} nucleosomes are also octameric *in vivo*.

Whilst the fixation of nucleosomes for the purpose of stabilisation is common, experiments here couple cross-linking with SDS-PAGE as a method to estimate the molecular weight of nucleosomes. The results clearly show complete cross-linking of both H3 and CENP-A^{Cnp1} nucleosomes extends to

the molecular weights of octamers rather than hemisomes (Figure 3-10). The possibility of "over cross-linking" two hemisomes was considered, but is not consistent with the presence of intermediate molecular weight bands between that of hemisome and octamer corresponding to the expected molecular weights of hexameric and septameric complexes (Figure 3-5). Furthermore, the low (nM) nucleosome densities make inter-particle interactions unlikely and is inconsistent with the prominent nature of the octameric, hexameric and septameric bands. Finally, over cross-linking of a samples consisting of hemisomes would be expected to form aggregates in ~ 50 kDa steps extending well above the hemisome band, yet even with a > 20-fold increase in the concentration of fixative the major complex remains that of a histone octamer (Figure 3-11). In addition, the apparent molecular weight of CENP-A^{Cnp1} core particles observed by cross-linking is also in excellent agreement with that estimated by (unfixed) dynamic light scattering and by the mobility of H3 and CENP-A^{Cnp1} nucleosomes through native PAGE gels (Figures 3-6 and 3-8).

Oxidation of nucleosomes to form disulphide bridges between the two H3 / CENP-A^{Cnp1} histones of H3 and CENP-A^{Cnp1} nucleosomes (respectively) was an important experiment confirming the presence of two copies of CENP-A^{Cnp1} within these recombinant nucleosomes (Figure 3-15). Attempts to improve the extent of cross-linking using longer-reaching cysteine-reactive cross-linkers failed to show cross-links over a wide titration range (not shown). Whilst these cross-linkers had been used previously for this purpose (Bowman *et al*., 2011) those experiments were performed using histone tetramer (without DNA) rather than a full nucleosome used here and so it is

assumed that access to the cysteines was limited by the wrapping of DNA and the physical bulk of these larger cross-linkers. Oxidation induced disulphide bond formation has the advantage of access to these sites but is typically limited to much shorter distances than those suggested by structural homology in the *S. pombe* four helix bundles. The very presence of disulphide bonds in this experiment indicates two CENP-A^{Cnp1} histones must interact closely, as is observed within the CENP-A four-helix bundle in the CENP-A nucleosome crystal structure (Tachiwana *et al*., 2011a) but which would be absent from the hemisome model. One may argue that this interaction could also result from a transient interaction of two hemisomes, which is a possibility when this experiment is considered in isolation. However the nucleosomes used in these experiments were at low (nM) concentration and in the context of data from the other experiments presented here, which suggest these CENP-A^{Cnp1} nucleosomes are octameric complexes, these data therefore support an octameric model in which each recombinant CENP-A^{Cnp1} nucleosome contains two copies of CENP-A^{Cnp1}.

It must be emphasised that these experiments only indicate that recombinant CENP-A^{Cnp1} nucleosomes are octameric, and similar analyses would need to be performed on CENP-A^{Cnp1} nucleosomes from cells before firm conclusions could be drawn for the stoichiometry of histones within CENP-A^{Cnp1} nucleosomes *in vivo*. However, with only ~60 CENP-A^{Cnp1} nucleosomes per diploid cell (Lando *et al*., 2012) it is difficult to obtain sufficient material for the analytical methods used here.

Chapter 4

CENP-A^{Cnp1} nucleosomes display distinct measurable physical properties

Introduction

H3 nucleosomes are octameric protein complexes that tightly wrap ~ 147 bp of DNA, functioning both to inhibit access to this DNA and to compact it. In all eukaryotes H3 nucleosomes consist of two copies of the highly conserved histone proteins H2A, H2B, H3 and H4, wrapped ~ 1.7 times with ~ 147 bp of DNA (Luger *et al*., 1997). Variations in nucleosome function are enabled either through post-translational modifications of the histones (typically in the exposed amino-terminal tails) or by replacement of a particular histone type with a histone variant. However, whilst functionally distinct, even these histone variants appear to impose only slight structural differences from their canonical nucleosome (Tachiwana *et al*., 2011b; Thakar *et al*., 2009). It is therefore surprising that one variant nucleosome, the centromere-specific nucleosome containing the H3-variant CenH3, should be subject to such a variety of different models concerning the stoichiometry of its constituents (see Figure 4-1 for an overview).

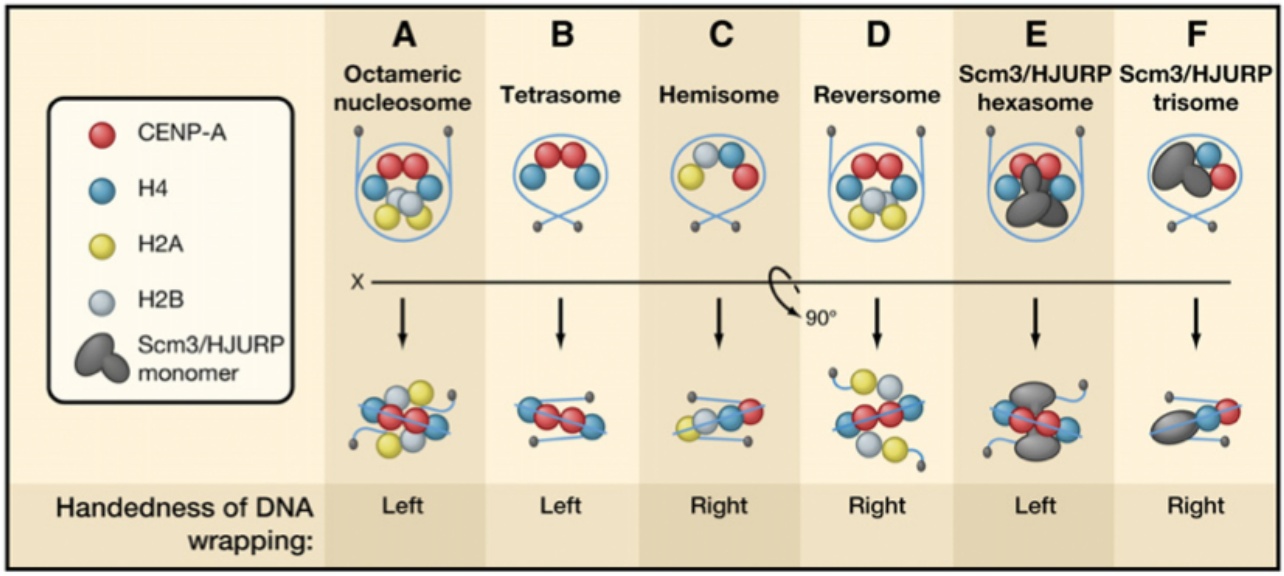


Figure 4-1 – CenH3 nucleosome models

Representation of current models for CenH3, highlighting the differences in nucleosome stoichiometry, components and DNA wrapping direction. Handedness of DNA wrapping refers to the induction of positive supercoiling for right handed wrapping and negative supercoiling for left handed wrapping. Adapted from Black *et al* 2011 (Black and Cleveland, 2011).

Perhaps one reason for this glut of models is that seemingly contradictory models are often proposed, even for the same nucleosome type in the same organism, with little overlap in experimental approaches. For example, the proposal of the hemisome model in flies and humans is predominantly based on atomic force microscopy (AFM) imaging (Dimitriadis *et al.*, 2010) - a technique that appears solely in papers supporting the hemisome model - whilst data supporting the octameric model, also (but not only) in humans, is largely based on biochemical analysis of nucleosomes prepared *in vitro*

(Tachiwana *et al*., 2011a), with little cross-over of assays and methodology applied between different studies supporting either model. The notable exception here is that in both models, MNase-protection data support the conclusion that CenH3 nucleosomes wrap DNA more loosely than H3 nucleosomes at the DNA entry / exit sites (Dalal *et al*., 2007; Sekulic *et al*., 2010; Tachiwana *et al*., 2011a).

What is required is a clear crossover and consistency in methods used to derive these models. In an attempt to resolve the discrepancies that lead to disparate models for the composition of CenH3 nucleosomes this chapter will focus on the two most prominent models in current literature - the hemisome and octamer models. The previous chapter detailed the biochemical characterisation of *in vitro* prepared H3 and CENP-A^{Cnp1} nucleosomes from *S. pombe* as octameric nucleosomes. Initial analyses from cross-linking and mass spectrometry suggested that Drosophila CENP-A^{CID} nucleosomes might only contain a single CENP-A^{CID} due to an absence of observable CENP-A^{CID} : CENP-A^{CID} cross-links (Dalal *et al*., 2007). Although a lack of cross-linkable residues has also been offered in explanation of these data (Black and Bassett, 2008). In addition, AFM was used to measure nucleosome height and found CENP-A^{CID} nucleosomes to be shorter than H3 nucleosomes (Dalal *et al*., 2007). Since then, "short" nucleosomes, as measured by AFM, is seen to indicative of hemisomes. However, no experiment has yet been performed to show that octameric CenH3 nucleosomes are the same height as octameric H3 nucleosomes. As AFM provides the predominant data set supporting the hemisome model this chapter will test the hypothesis that octameric CENP-A^{Cnp1} nucleosomes

appear shorter in height than octameric H3 nucleosomes by AFM, with the aim of evaluating the reliability of AFM as a technique for determining nucleosome component stoichiometry.

Atomic force microscopy

Atomic force microscopy (AFM) can be utilised for imaging, manipulating and characterising particles at sub-nanometer resolution. An atomic force microscope essentially comprises of a sharp tip (typically between $1-5$ nm in diameter for nucleosome samples) at the end of a cantilever that scans line by line over samples immobilised on a surface, like a finger over braille (Figure 4-2). Deflections in the tip are sensed by a laser reflected from the top surface of the cantilever to a photosensitive quadrant that distinguishes movement of the tip in 2 dimensions, up / down (vertical) and left / right (roll) - although movement of the tip is largely restricted to variations in the vertical axis. In addition, feedback mechanisms are employed that move the tip up and down to maintain a constant force between the tip and the surface when moving over variations in the surface topography. Within the setup of the AFM used here acoustic, feedback and mechanical noise typically induced a ~ 0.18 nm error in height readings, meaning that an atomically flat surface (in this case freshly cleaved mica) was observed to show ~ 0.18 nm average standard deviation in height over four 2.5 μm^2 surfaces (Figure 4-3).

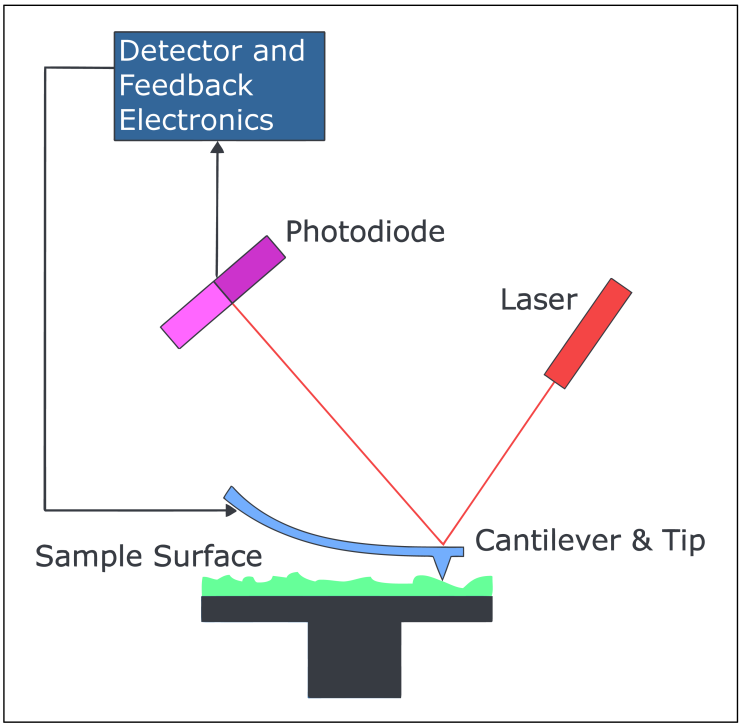
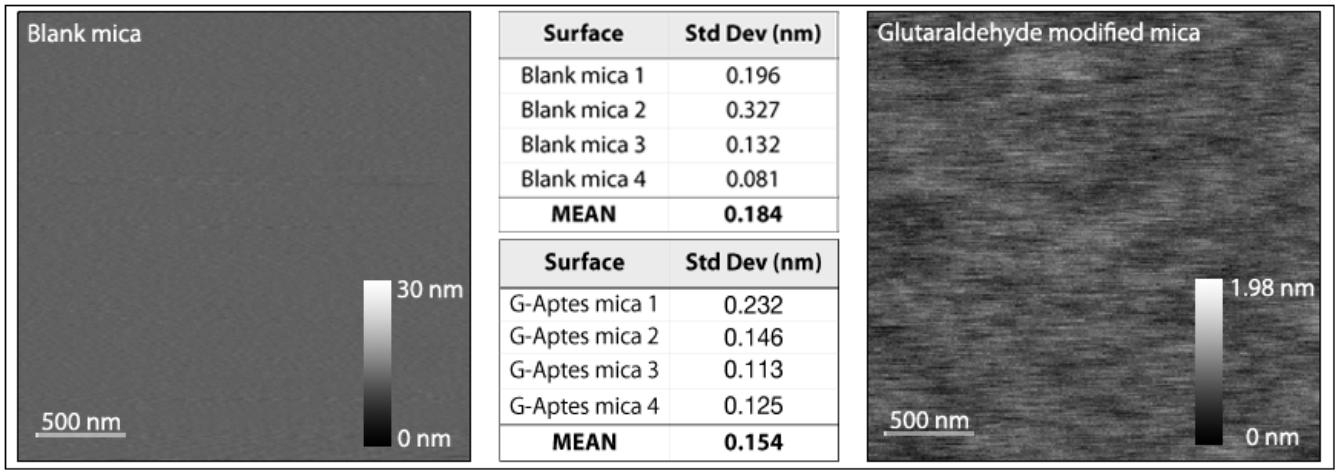


Figure 4-2 Principles of AFM setup

Atomic force microscopy (AFM) is capable of measuring the topography of objects in a sample at sub-nanometer resolution. Briefly, a sample of interest is deposited onto a surface and a nanometer scale tip attached to a piezoelectrically controlled cantilever scans back and forth across the surface in progressive lines. Variations in sample or surface height cause deflections in the cantilever that are detected by the altered reflection of a laser from the top surface of the tip by a photosensitive quadrant. Feedback loops ensure the force between the tip and the surface remains constant as the height of the sample varies during scanning. Image adapted from *http://en.wikipedia.org/wiki/File:Atomic_force_microscope_block_diagram.svg* and used under a creative commons license.



Surface	Std Dev (nm)
Blank mica 1	0.196
Blank mica 2	0.327
Blank mica 3	0.132
Blank mica 4	0.081
MEAN	**0.184**

Surface	Std Dev (nm)
G-Aptes mica 1	0.232
G-Aptes mica 2	0.146
G-Aptes mica 3	0.113
G-Aptes mica 4	0.125
MEAN	**0.154**

Figure 4-3 - Height variations for a blank and modified mica surfaces measured by AFM

The average height variation from four freshly cleaved mica surfaces was 0.184 nm over 2.5 μm^2 and 0.154 nm for glutaraldehyde-modified mica. Example images of blank mica (left) and glutaraldehyde modified APTES mica (right) are shown without subtracting the average background levels (11.52 nm for blank mica and 0.644 nm for glutaraldehyde modified mica) to ensure small surface variations are clearly visible.

Freshly cleaved mica is most commonly used as a surface for AFM as its crystalline structure provides an atomically flat surface over large areas with little preparation, yet mica alone is negatively charged and so repels DNA (Pastré *et al.*, 2003). A solution to this is to first silanise the mica with aminopropyltriethoxysilane (APTES) and then functionalise the surface with glutaraldehyde, according to the protocol of Wang and colleagues (Wang *et al.*, 2002). This glutaraldehyde modified mica surface was found to not significantly alter the surface roughness (~ 0.15 nm standard deviation)

(Figure 4-3) and has allowed measurements of nucleosome heights in previous studies (Dalal *et al*., 2007).

For the study of nucleosome heights using AFM the most common method of scanning is known as "tapping" mode AFM (Bui *et al*., 2012; Dalal *et al*., 2007; Dimitriadis *et al*., 2010). In tapping mode, the AFM cantilever is oscillated near its resonance frequency above the surface and then gradually lowered until intermittent contact between the tip and the surface reduces the amplitude of the oscillations by a set percentage. The greater the dampening of these tip oscillations the greater the force applied to the surface. Whilst these settings may differ slightly between samples the goal is to apply the minimum force required to establish contact between the tip and the surface. Indeed, tapping mode is often preferred over the alternative "contact" mode imaging due to the reduced tip forces required for tapping mode that allow for less destructive imaging (Zhong *et al*., 1993). In tapping mode, the recorded height is derived from variations in the oscillation amplitude that result from a combination of forces between tip from the surface, including electrostatic, dipole-dipole and Van der Waals forces (Butt, 1991). Additionally, although samples are dried under argon before imaging, ambient humidity creates a liquid layer that coats and permeates samples, so the apparent heights recorded by AFM "in air" do not always reflect the true crystallographic heights of samples, and are generally smaller (Andrew Downes and David Dryden, personal communications). Differences in sample hydration, sample compression, buffer composition, tip properties and tip force are also commonly used to explain differences in the apparent heights of samples between experiments (Müller and Engel, 1997).

CenH3 nucleosome height measurements by AFM

To date, four papers have measured the height of CenH3 nucleosomes by AFM, with all measurements performed by the same research group. Each paper draws the conclusion that CenH3 nucleosomes appear as hemisomes relative to H3 nucleosomes, representing the entirety of data supporting the hemisome model (Dalal *et al*., 2007; Dimitriadis *et al*., 2010; Wang *et al*., 2008).

In initial observations Dalal *et al*, 2007, measured samples of fly CENP-A^{CID} and bulk (H4) nucleosomes obtained by IP of CENP-A^{CID} or H4-Biotin from partial micrococcal nuclease digests of unfixed chromatin that were washed in 0.35 M salt (Dalal *et al*., 2007). Due to the partially digested nature of these arrays they contained both H3 and CENP-A^{CID} nucleosomes, at a ratio of ~ 0.3 : 1 (H3 : CENP-A^{CID}). Nevertheless, the vast majority of nucleosomes along the CENP-A^{CID}-containing arrays appeared half the height of the bulk nucleosomes (~ 1 nm and ~ 1.8 nm respectively) (Figure 4-4 A), this led to the proposal of the hemisome model, in which the CENP-A^{CID} nucleosome contained just one copy of each histone.

Subsequent analysis of CENP-A^{CID} nucleosomes by AFM was primarily aiming to address difficulties identifying particles as CENP-A^{CID} particles from background noise or other H3-containing nucleosomes (Wang *et al*., 2008). Wang *et al*, 2008 utilised an AFM tip functionalised with CENP-A^{CID} antibody that gave a recognition signal specific for CENP-A^{CID}. Though instead of the mixed nucleosome arrays used previously (Dalal *et al*., 2007), the samples used in this study were first fixed with glutaraldehyde (to an unknown degree) then the core histone complex dissociated from DNA by elution from hydroxyapatite in high (2 M) salt. The salt concentration was

reduced and CENP-A^{CID} core particles (without DNA) were immuno-precipitated. The "bulk" control sample used (serving as a proxy for H3 core particles) was simply the unbound fraction of the CENP-A^{CID} IP, which appears to contain a large range of protein complexes (also observed in the CENP-A^{CID} sample to a lesser degree) that presumably contain histone aggregates following the dialysis from the high to low salt environment as well a whole range of other DNA binding complexes. Perhaps as a consequence of an extraction technique that only really measures the CENP-A^{CID} histone core complex (and even then with some degree of uncertainty over it's integrity), the height values differ somewhat from the previous analysis by the same research group (Dalal *et al.*, 2007) (Figure 4-4 B). The authors state:

> *"CenH3 [CENP-A^{CID}] core particles display a tight distribution between 1 and 2 nm, consistent with their proposed hemisomal organization. In contrast, unbound core particles display a broad distribution of heights [1.6 nm - 6 nm] that are on average about twice as tall as CenH3 core particles".*

It is a concern that the particles with heights less than 1.6 nm were arbitrarily classified as non-nucleosomal particles or debris *only* in the unbound "bulk" samples - whereas the CENP-A^{CID} particles are clearly recorded down to at least ~ 1.15 nm (Figure 4-4 B). This discrepancy in the analysis, in addition to the undefined parameters for the particles chosen for measurement limits the relevance of these data for making conclusions about the relative component stoichiometry of CENP-A^{CID} and H3 nucleosomes.

Further analysis of chromatin from human cell lines (HeLa and HEK) concluded CENP-A nucleosomes are also smaller than H3 nucleosomes

(Dimitriadis *et al*., 2010). CENP-A nucleosomes were again immuno-precipitated from partial micrococcal nuclease digests of unfixed chromatin and so are likely to contain a mixture of CENP-A and H3 nucleosomes (Blower *et al*., 2002). The origins of the "bulk" control nucleosomes were not stated, but presumably represent the input chromatin sample from which the CENP-A IP was taken. Analysis of the AFM images employed automated particle filters to discount "non-nucleosomal" particles whose circularity was less that 0.5, diameter less than 10 nm and height greater than 10 nm or less than 0.5 nm (the height given for DNA alone). Dimitriadis *et al*, 2010 provide much data for H3 and CENP-A nucleosome heights, from both HeLa and HEK cell lines, from intact chromatin, ~ 1 kb fragments and predominantly 150 – 600 bp fragments, all from both low (50 mM) and medium (350 mM) salt extractions. However, it is confusing that much of the data discussed does not appear to match with the data shown in figures, particularly with reference to the frequency histograms. This may arise because the values stated frequently refer to the mean height and standard deviations of sub-populations within the particles counted (that are not otherwise shown) rather than the modal or median measurements. Using the few frequency histograms shown by Dimitriadis *et al*, 2010, our own inference of the modal nucleosome height values suggest the low salt HeLa bulk sample ranges from ~ 2 – 5 nm with a modal height of ~ 2.8 nm, and ranges from ~ 1 – 3.5 nm for the low salt HeLa CENP-A sample, with a modal height of ~ 1.6 nm (Figure 4-4 C). Thus the conclusions of Dimitriadis *et al*, 2010 are that, as in flies, human CENP-A nucleosomes appear half the height of H3 nucleosomes by AFM. The increased range of nucleosome heights observed

in Dimitriadis *et al*, 2010 is likely a reflection of the increased number of particles counted (~ 2000 here, compared with 100 in Dalal *et al*, 2007). In addition, despite the particle filters, it is likely that the use of automated analysis lead to a higher proportion of false positive "nucleosome" particles being included in their data analysis than when nucleosomes are distinguished by eye. For example, in previous AFM measurements DNA appears ~ 0.5 nm in height (Dalal *et al*., 2007), so overlapping DNA fibres are likely to appear a similar height to that proposed for hemisomes. Overlapping DNA fibres would also have a diameter (above 0.5 nm) of less than 10 nm, and so would fulfil the requirements for analysis by the automated particles filters employed by Dimitriadis *et al*, 2010. Further, it is conceivable that free DNA may have been more apparent in the CENP-A sample than in the bulk sample due to the preferential dissociation of CENP-A nucleosome particles arising from an inherent instability of unfixed CenH3 nucleosomes (Camahort *et al*., 2009).

In the most recent publication (Bui *et al*., 2012) CENP-A chromatin from human cell lines are again analysed and the composition stoichiometry of CENP-A chromatin is examined throughout the cell cycle. The authors propose that CENP-A nucleosomes are hemisomal throughout the cell cycle, except for late during G1 phase and S phase, where they briefly become octameric before splitting back into hemisomes, presumably following transit of the replication machinery. This model is based on AFM data that measures the height of immunoprecipitated CENP-A chromatin throughout the cell cycle and observes them to be ~ 1.8 nm during G2, M and G1 but ~ 2.7 nm during late G1 / S phase (Figure 4-4 D). It is surprising they fail to mention in

the text that in the supplemental data provided the height of immunoprecipitated H3 nucleosomes at G1/S and G2 *also* shows a similar ~ 0.75 nm increase in height at S phase (relative to a ~ 0.9 nm increase for CENP-A nucleosomes), with G2 phase H3 nucleosomes clearly showing two distinct populations (Figure 4-4 E). In light of this data it would seem that any cell cycle-related change to CENP-A nucleosomes is similarly affecting H3 nucleosomes, and because both CENP-A and H3 appear to change to a similar extent one explanation may be that CENP-A nucleosomes are not changing in stoichiometric composition but instead these data reflect a cell cycle-specific change in the properties of the CENP-A and H3 nucleosomes that alters their apparent height measured by AFM. Intriguingly, in the same paper, Bui *et al* (2012) identify specific post-translational modifications of CENP-A and H4, CENP-A Lys 124 acetylation and H4 Lys 79 acetylation, that become apparent at G1/S phase, although it is not yet clear whether this H4 modification is also found in H3 nucleosomes. However, it raises the possibility that post-translational modification of nucleosomes may provide a cell cycle-specific change in nucleosome structure or stability that alters the apparent height of nucleosomes measured by AFM. It therefore seems equally plausible that previous observations of the decreased height of CENP-A nucleosomes (relative to H3 nucleosomes) is a consequence of different physical properties between CENP-A and H3 nucleosomes rather than a difference in their stoichiometric composition. The above issues raise concerns over the designation of particles with reduced height, measured using AFM, as hemisomes.

This chapter attempts to resolve these issues by measuring the height of defined H3 and CenH3 nucleosomes. Recombinant nucleosomes of defined stoichiometric composition are utilised to determine if reduced height is a reliable reporter of nucleosome stoichiometric composition, or if it instead represents an intrinsic property of octameric CenH3 nucleosomes.

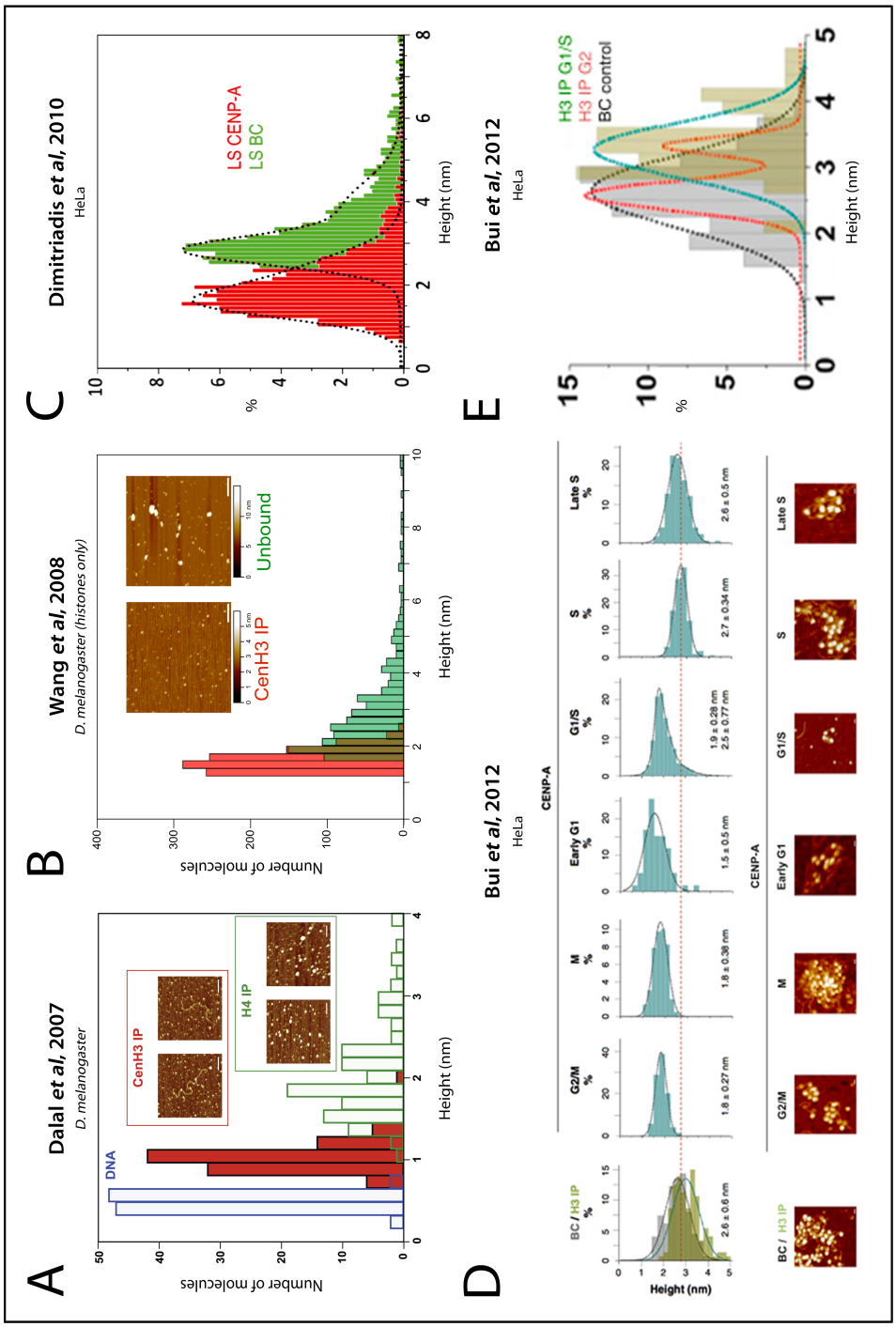


Figure 4-4 – Comparison of existing CenH3 AFM data

Summary of existing AFM data comparing apparent heights of H3 and CENP-A nucleosomes. Figures adapted from Dalal *et al* (2007) (a), Wang *et al* (2008) (b), Dimitriadis *et al* (2010) (c) and Bui *et al* (2012) (d and e). See text for a detailed analysis of each image.

The four studies that propose the existence of hemisomal CenH3 nucleosomes (described above) used CenH3 nucleosome populations from flies and human cell lines. One drawback of these *ex vivo* studies is the level of contamination present within these samples (see purifications within Dimitriadis et al, 2010). Contamination from non-nucleosomal molecules, partially disassembled nucleosomes and other particulate debris is the likely cause of the large range of particle sizes observed in each of these studies which, even with the use of automated particle selection constraints makes clear identification of nucleosome height difficult. A second form of contamination present in the above studies is cross-contamination of H3 nucleosomes present within the CenH3-containing nucleosome arrays due to the limited nuclease digestion and interspersed nature of CenH3 and H3 nucleosomes in an array (Blower et al., 2002). Again, this contamination may mask the true heights of CenH3 and H3 nucleosomes, making it difficult to determine if CenH3 nucleosomes are reduced in height relative to H3 nucleosomes when bulk or H3 nucleosomes are seen to span over twice the range of CenH3 nucleosomes and both populations contain multiple sub-populations (Figure 4-4).

In Chapter 3, recombinant *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes were biochemically characterised and shown to form octameric complexes. Because the stoichiometric composition of these *in vitro* assembled nucleosomes have been determined by a range of techniques these samples provide defined material with which to test whether CENP-A^{Cnp1} nucleosomes appear smaller than H3 nucleosomes using AFM for some reason other than a true difference in their stoichiometric composition.

Additionally, the exceptional purity afforded by *in vitro* nucleosome assembly would likely tighten the observed range of particle sizes by minimising sample contamination inherent to *ex vivo* nucleosome samples.

Results

Octameric CENP-A^{Cnp1} nucleosomes appear smaller than H3 nucleosomes using AFM

Nucleosome samples reconstituted as described in the previous chapter were deposited onto freshly cleaved mica surfaces functionalised with APTES and glutaraldehyde as described by Wang *et al* (2002). Briefly, nucleosomes samples at ~0.1 nM concentrations in TE were dropped onto the freshly prepared surfaces and allowed to bind, at room temperature, for 5 minutes before being washed twice with water and gently dried under a clean Argon stream.

Initial attempts to image mono-nucleosomes by AFM were hindered by particles with a range of sizes that were also present in blank samples. In addition, pits in the surfaces also prevented the clear identification of true nucleosomal particles (Figure 4-5). These contaminating particles are similar to those observed in previous AFM studies (Dalal *et al*., 2007; Dimitriadis *et al*., 2010; Wang *et al*., 2002; 2008) and are generally thought to be dust particles (Dr Andrew Downes, personal communication). Particles persisted through multiple wash steps (not shown), suggesting they had become fixed to the glutaraldehyde surface on contact. In order to allow proper identification of nucleosomes, the protocol was altered so that H3 or CENP-

A^{Cnp1} octamers were reconstituted onto a 5 kb repeat array of DNA containing 25 x 197 bp 601 sequence (Routh *et al*., 2008) and subsequently used for AFM imaging. The use of arrays in a low salt environment meant that the assembled nucleosomes adopt the classical beads-on-a-string structure and were easily distinguished by eye from non-nucleosomal debris / particles (Figure 4-6).

Images were collected for multiple reconstitutions on such arrays for both H3 and CENP-A^{Cnp1} nucleosomes. To avoid ascertainment bias an automated analysis protocol was built within the ImageJ software that allowed the collection of a large number of nucleosomes. In order to follow the previous filters used by Dimitriadis *et al* (2010) filters were set to count particles with a circularity > 0.5 (where 1 is a perfect circle, and 0 a straight line), a height < 10 nm and a diameter > 5 nm. The only difference being the diameter filter chosen in the analysis described was intentionally less than that of a full nucleosome (10 nm) as thresholding the mask image to DNA height - a requirement to isolate nucleosomes from the array - necessarily reduced the apparent diameter of the nucleosomes. The full method used is described in Chapter 2 (materials and methods). Notably, nucleosomes were not found to be taller than ~ 4 nm, suggesting that nucleosomes lay with their largest surface "flat" on the surface, like a coin on a table, likely because this is the most energetically favourable position.

Comparing particle heights over a number of images showed large ranges in the heights of both H3 and CENP-A^{Cnp1} nucleosomes that were not consistent between images (Figure 4-7). Inspection of the particles selected by the automated imaging software consistently showed that the software

incorrectly selected a significant proportion of non-nucleosomal particles for inclusion in the analysis. These errant particles appear to primarily be overlapping DNA fibres; material that would otherwise be included using the same filters appointed by Dimitriadis *et al* (2010). Identification of these particles as non-nucleosomal was clear by visual inspection because nucleosomes were observed in the typical beads on a string entities whereas these non-nucleosomal particles were detached from DNA fibres, overlapping DNA fibres clearly stemmed from two distinct DNA fibres and compacted nucleosome arrays or aggregates formed complexes with large areas (Figure 4-6). To avoid problems associated with automated selection presumed nucleosomes were selected manually. In order to limit any potential bias in the particles designated as nucleosomes all nucleosome-like particles that were pre-selected using the filters described above and which met with visual criteria within each image were included in the analysis. Manual rejection of non-nucleosomal particles from the analysis significantly improved the consistency of the nucleosome heights collected (Figure 4-7). Whilst this selection of nucleosomal particles would ideally be performed as a double-blind task this type of analysis remains arguably more appropriate than that utilised previously (Dimitriadis *et al*., 2010) because the same particle filters were employed but with additional filtering to eliminate false-positives.

As previously reported (Dalal *et al*., 2007), the modal DNA height was measured in each image as an internal control to allow comparison of heights between images. Typically it was found that the recorded DNA heights varied between 0.4 − 0.7 nm (Figure 4-7). This range is common when

imaging in air and thought to be due to changes in the environmental humidity, to which AFM is particularly susceptible (Andrew Downes and David Dryden, personal communication). Attempts were made to perform AFM measurements in liquid phase, which is performed in buffer so not affected by changes in environmental humidity and typically obtains heights that are closer to their crystallographic norms due to proper sample hydration (Wang *et al*., 2002). However, our Veeco Explorer AFM (Bruker) was not designed for liquid AFM and attempts to image nucleosomes resulted in only poor resolution scans and were therefore not used. Instead, as an internal control, recorded nucleosome heights were adjusted by normalising to a DNA height of 0.5 nm for each image (the average modal DNA height). This internal levelling of DNA heights was performed by addition or subtraction rather than by division because the cause of the variations in DNA height were thought to be defined changes in the thickness of the hydration layer rather than a proportional change in particle size (Andrew Downes, personal communication). Thus the adjustment is effectively offsetting changes in humidity between images. Whilst this adjustment was not strictly necessary, as essentially identical differences in nucleosome heights recorded without such levelling (Figure A2-1), the adjustment was nevertheless included as an internal control as it increases confidence and is good scientific practise.

The collected data show the heights of the recombinant CENP-A^{Cnp1} nucleosomes measured here (mode ~ 0.94 nm, median ~1.05, n = 223) were lower than those observed for H3 nucleosomes (mode ~ 1.58, median ~ 1.74 nm, n = 338) (Figure 4-8). Using the hypothesis that CENP-A^{Cnp1} nucleosomes

appear smaller than H3 nucleosomes the probability associated with a one-tailed student's t-test is 5.54 $x10^{-87}$, making this a highly significant difference in nucleosome heights. These data mirror the height difference previously reported between H3 and CenH3 for both human and fly nucleosomes extracted from cells (Dalal *et al*., 2007; Dimitriadis *et al*., 2010). The height measurements are particularly close to those observed for CENP-A^{CID} and H3 nucleosomes prepared from *D. melanogaster* S2 cells (Dalal *et al*., 2007). Of particular interest is that the difference in apparent (modal) heights between H3 and CenH3 nucleosomes remains surprisingly well conserved between these *in vitro* recombinant *S. pombe* nucleosomes and human and fly nucleosomes extracted from cells. When expressed as a percentage of the H3 nucleosome heights the CENP-A^{Cnp1} nucleosome height is 60 % for the *in vitro* recombinant nucleosomes described here. This compares well with the 54 % in flies (Dalal *et al*., 2007), and 58 % in humans (Dimitriadis *et al*., 2010). This consistency of the height differences between the nucleosomes prepared here and those extracted from two different organisms suggests that CenH3 nucleosomes may possess a conserved feature that results in the reduction of height observed by AFM. Importantly, the cross-linking and hydrodynamic mobility assays detailed in Chapter 3 showed that the *in vitro* assembled H3 and CENP-A^{Cnp1} *S. pombe* nucleosomes have an octameric composition. This raises the intriguing possibility that previous studies where it was concluded that CenH3 nucleosomes are hemisomes, based on reduced height in AFM, may instead be measuring some intrinsic distinct physical property of CenH3 nucleosomes that causes them to appear smaller by AFM than H3 nucleosomes. If this is correct then CenH3 nucleosomes are likely to be

octameric rather than hemisomal. This finding calls into question the conclusion that CenH3 nucleosomes with reduced height (in AFM measurements) are hemisomes.

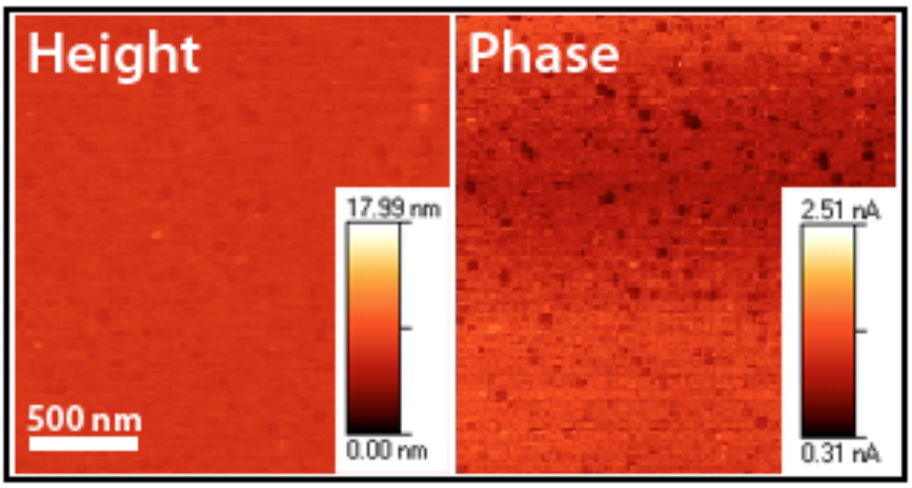


Figure 4-5 Non-nucleosomal AFM background

Image shows the background noise, imaging surface and dust particles detected by AFM in both height and phase modes for a buffer only sample. Pits in the surface are the result of an uneven deposition of APTES and glutaraldehyde whilst the particles are thought to be dust (Andrew Downes, personal communication) and are most apparent in the phase image as this mode detects changes in the tip interaction with the surface. These surface features are similar to those seen in other AFM studies (Bui *et al.*, 2012; Dalal *et al.*, 2007; Dimitriadis *et al.*, 2010; Wang *et al.*, 2002; 2008) but make imaging of mono-nucleosomes difficult and prone to artefacts.

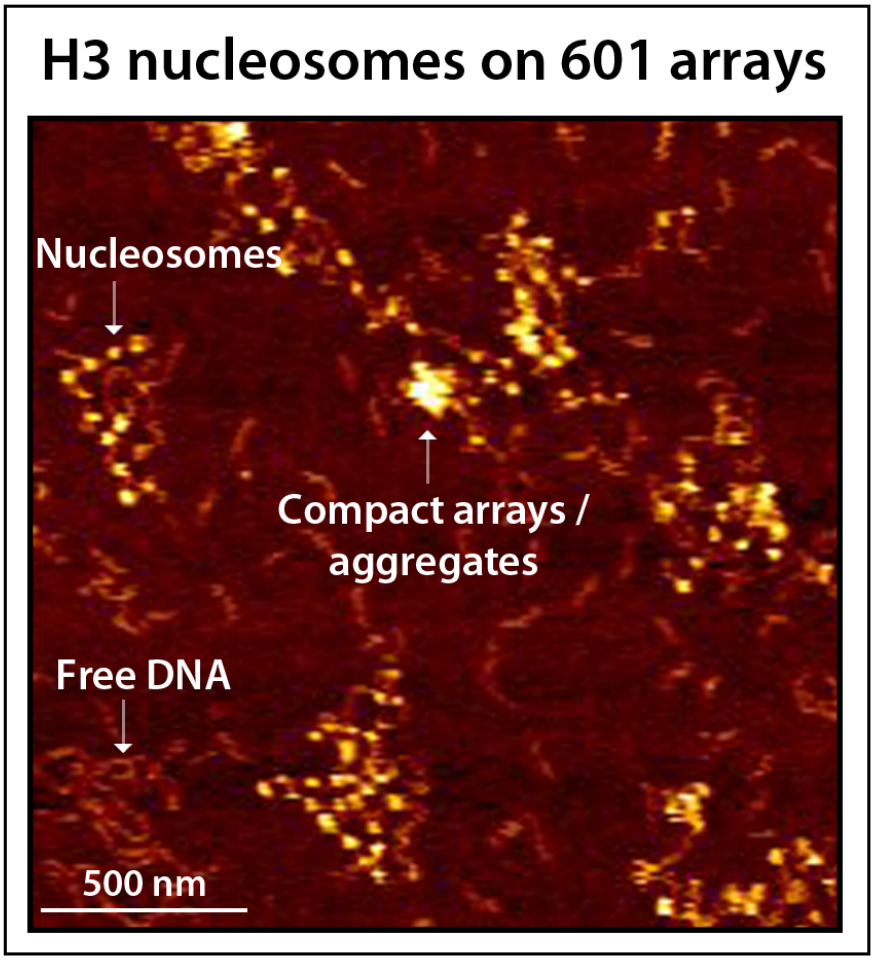
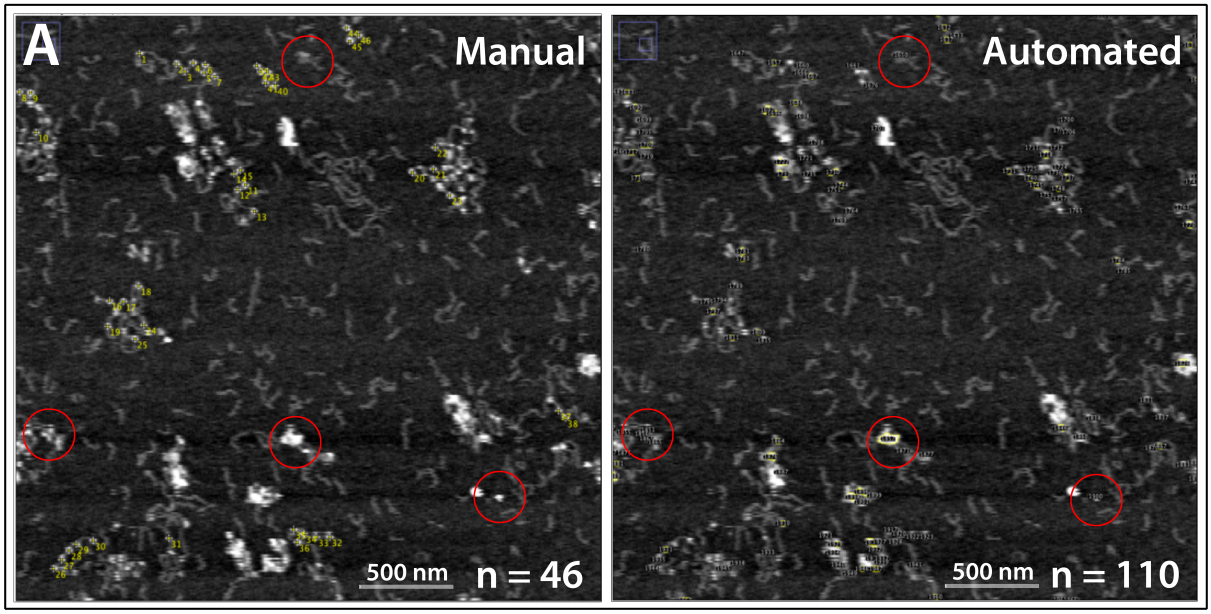


Figure 4-6 – AFM image of H3 nucleosomes on 601 arrays

H3 nucleosomes on 601 arrays are detected in AFM images in the typical beads-on-a-string confirmation. Compact nucleosomes or aggregates and free DNA are also seen. Shorter DNA fragments in this image are the result of competitor ~ 200 bp DNA used during the reconstitution and typically do not form nucleosomes.


A
Manual
500 nm
n = 46
Automated
500 nm
n = 110

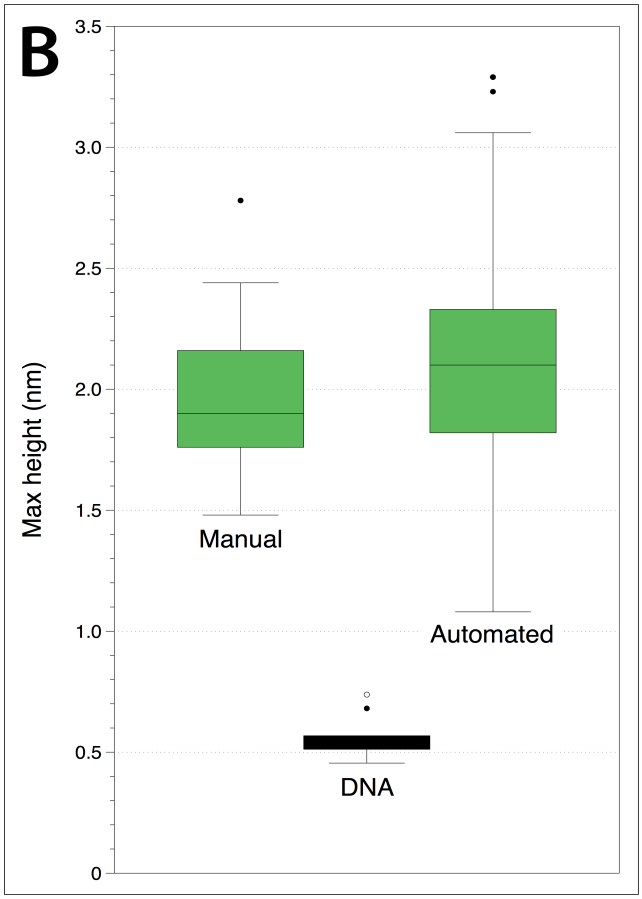

B
Max height (nm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Manual
Automated
DNA

Figure 4-7 - Range of H3 nucleosome heights before and after manual filtering of non-nucleosomal particles

Automated selection of nucleosomes based on the criteria described in the text includes a large number of false-positive, non-nucleosomal particles in the analysis (a) which gives rise to a large range in apparent nucleosome heights (b). Manual identification of nucleosomes within an image removes ambiguous particles that may either be tightly clustered, overlapping other particles, the result of overlapping DNA strands or not associated with DNA (red circles in (a)) and the recorded nucleosome heights are consequently more consistent (b). Nucleosome heights summarised here were collected from *in vitro* assembled recombinant human H3 nucleosome sample, reconstituted onto a 19 x array of the 197 bp 601 nucleosome positioning sequence. Automated software counted 110 particles, whilst additional manual visual filtering counted 46 particles. The average height of DNA was recorded for 10 DNA fragments for each image and the distribution of all heights are shown without levelling of the DNA height. The central line within each box plot marks the median, the outer edges of each box represents the first and third interquartile ranges and the whiskers mark the range. Statistical outliers are shown here as single dots.

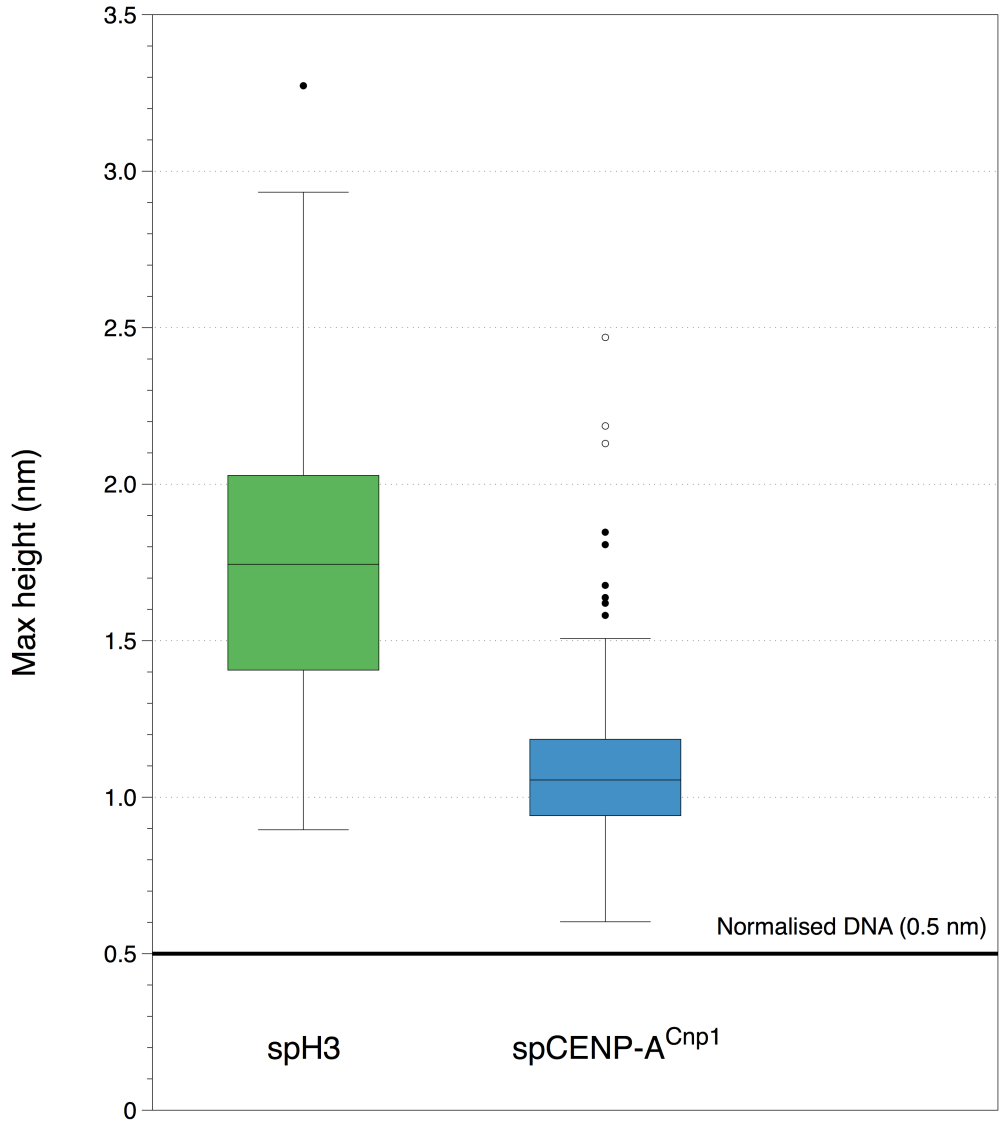


Figure 4-8 - Apparent heights of H3 and CENP-A^{Cnp1} nucleosomes by atomic force microscopy

H3 and CENP-A^{Cnp1} nucleosome arrays were imaged by AFM and the heights of individual nucleosome particles recorded using ImageJ. As an internal control, the average DNA height was calculated for each image and the nucleosome heights normalised for a DNA height of 0.5 nm. The central line within each box plot marks the median, the outer edges of each box represents the first and third interquartile ranges and the whiskers mark the range. Statistical outliers are shown here as single dots. A one-tailed

student's t-test finds CENP-A^{Cnp1} nucleosomes to be smaller than H3 nucleosomes by a highly significant margin, with the probability that the reduction in height of CENP-A^{Cnp1} nucleosomes is due to chance = 5.54 x 10^{-87}, see text for details). H3 modal value = 1.58 nm, CENP-A^{Cnp1} modal value = 0.94 nm.

The diameters of H3 and CENP-A^{Cnp1} nucleosomes are equivalent

It is well known that nucleosomes appear less than their crystallographic heights by AFM. This is typically explained by a combination of factors such as the absorption of salts into the surface, compression and dehydration (Bussiek *et al*., 2007; Dalal *et al*., 2007). It was therefore possible that CENP-A^{Cnp1} nucleosomes are simply compressed more easily under the force of the AFM tip than H3 nucleosomes. If nucleosomes were compressed from above by the AFM tip then an expected outcome would be that the diameter (or circumference) of the nucleosomes might bulge to compensate and conserve the volume of the nucleosome. If CENP-A^{Cnp1} nucleosomes were compressed more than H3 nucleosomes as a result of the tip force then this expansion of the diameter would presumably increase proportionally. To test whether this occurs, measurements of the diameter and circumference of the same CENP-A^{Cnp1} and H3 nucleosomes from which the height measurements were taken above were made using the SPIP software program (Image Metrology) and sorted into 20 bins. Charts plotting the relative frequency of each bin show that both H3 and CENP-A^{Cnp1} nucleosomes appear to have modal diameters of ~ 22.5 nm and modal perimeters of ~ 90 nm, suggesting the difference in height is unlikely due to distinct degrees of compression (Figure 4-9). The

dimensions obtained here by AFM are larger than crystallographic nucleosome diameters (~ 10 nm) due to dilation caused by the finite AFM tip size (Bui *et al.*, 2012).

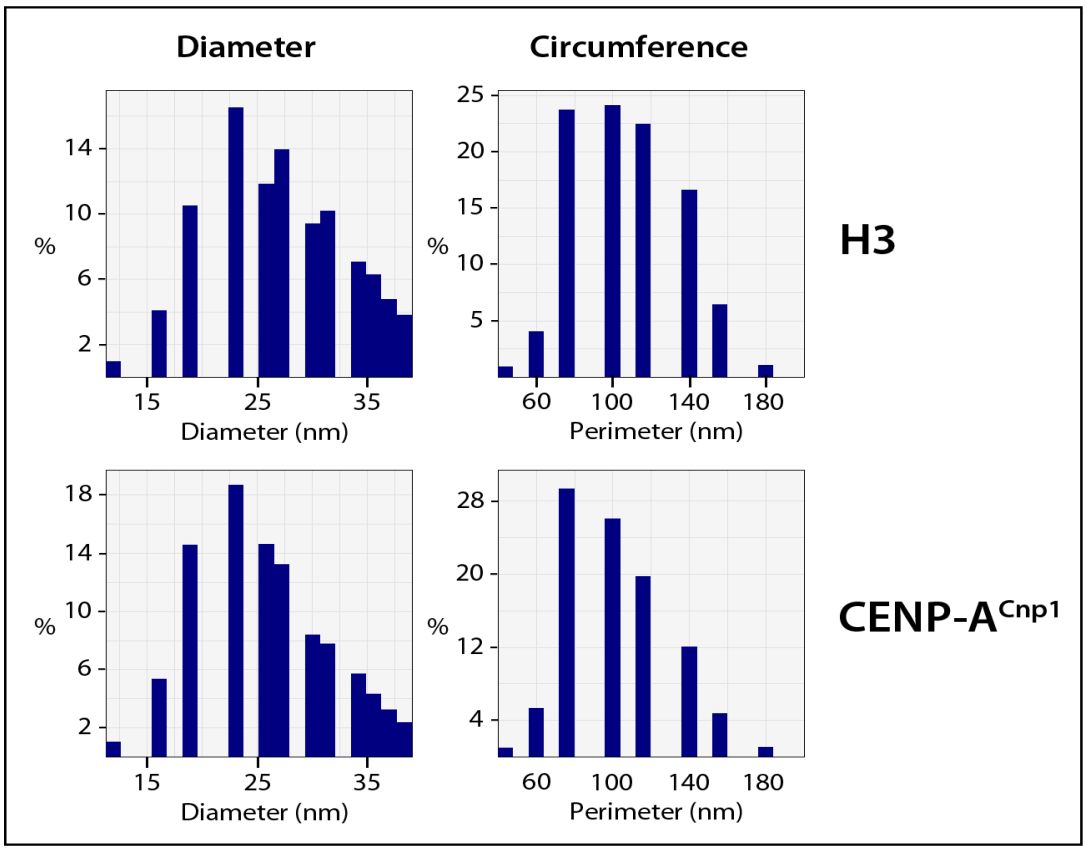


Figure 4-9 - Comparison of H3 and CENP-A^{Cnp1} nucleosome diameters

Two measures of nucleosome footprint, diameter and circumference, are compared between H3 and CENP-A^{Cnp1} nucleosomes. The dimensions were calculated for the same *S. pombe in vitro* reconstituted nucleosomes for which the height measurements were obtained, using SPIP software, and sorted into 20 bins. Charts plot the relative frequency of each bin to show the distribution of sizes. Both H3 and CENP-A^{Cnp1} nucleosomes appear to have modal diameters of ~ 22.5 nm and modal perimeters of ~ 90 nm. These dimensions are larger than crystallographic nucleosome diameters (~ 10 nm) due to dilation caused by the finite tip size of the AFM.

A consistent size difference is observed between recombinant H3 and CenH3 nucleosomes from distinct sources

All published analyses of CenH3 and H3 nucleosome height measurements has utilised chromatin extracted from fly or human cells. In the section here entitled "Octameric CENP-A^{Cnp1} nucleosomes appear smaller than H3 nucleosomes using AFM" analysis showed that recombinant *S. pombe* nucleosomes exhibit a similar height difference. To reconcile this it is possible that *S. pombe* nucleosomes have different properties than human nucleosomes. Thus our analysis was extended to nucleosomes assembled *in vitro* using recombinant human H3 or CENP-A and other core histones. Recombinant human nucleosomes were assembled on the same 5 kb arrays of 19 x 197 bp 601 DNA used for *S. pombe* nucleosomes (section entitled "Octameric CENP-A^{Cnp1} nucleosomes appear smaller than H3 nucleosomes using AFM")(human nucleosome reconstitutions performed by A. Guse in A. Straight's laboratory, Stanford). To assess the relative stoichiometries of these recombinant human CENP-A and H3 nucleosomes arrays were digested to mono-nucleosomes and separated by native PAGE gel (performed by A. Guse in A. Straight's laboratory, Stanford). Digestion of these nucleosome arrays with the restriction enzyme Ava1 cuts between each 601 repeat, leaving single nucleosomes centred on a ~ 200 bp 601 sequence. Both the H3 and CENP-A mono-nucleosomes have similar hydrodynamic mobilities that is consistent with octameric nucleosomes (Figure 4-10 A). As a further test of their stoichiometric composition, nucleosomes were fixed with a 5000-fold molar excess of BS(PEG)$_5$, the DNA digested with Benzonase nuclease and the histone core particles analysed by silver-stained SDS-PAGE.

Both H3 and CENP-A nucleosomes fixed up to similar octameric molecular weight complexes of ~ 100 kDa (Figure 4-10 B). Taken together, the migration patterns and cross-linking data support the conclusion that these human H3 and CENP-A nucleosomes assembled from recombinant histones have octameric stoichiometries.

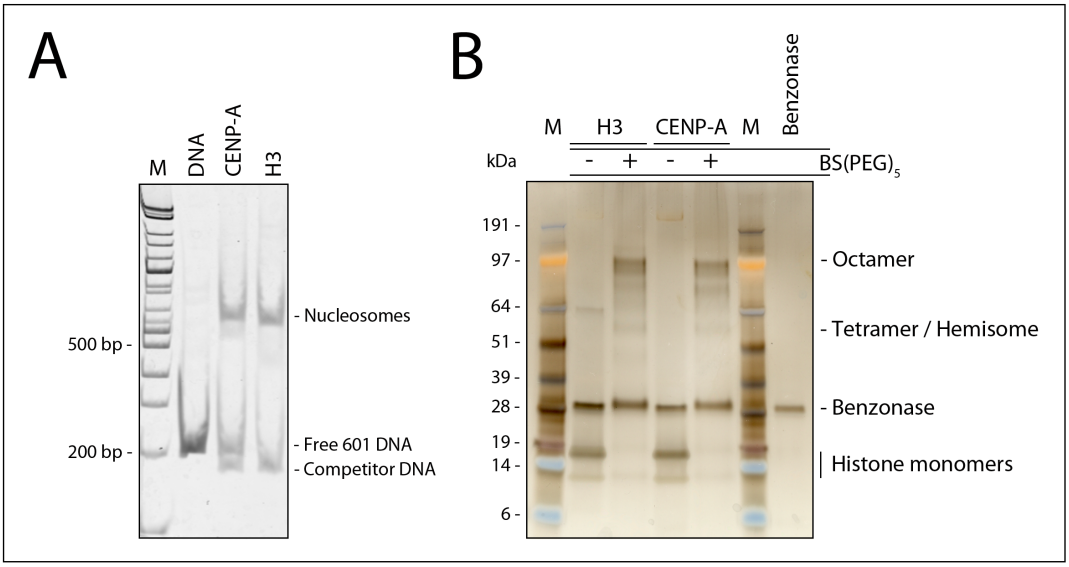


Figure 4-10 - *In vitro* assembled human nucleosomes appear octameric
In vitro reconstituted arrays of human H3 and CENP-A nucleosomes on 601 DNA were digested to mono-nucleosomes using the restriction digest enzyme Ava1 and run on a 5 % native PAGE gel (a). The gel shows nucleosome bands from the H3 and CENP-A arrays, along with unsaturated 601 sites and competitor DNA used in the reconstitution. The DNA was post-stained with Sybr Green 1. In (b), undigested arrays were fixed for 2 h with a 5000 x molar excess of $BS(PEG)_5$ (a primary amine cross-linker), DNA was digested by the addition of Benzonase and the samples run on a silver-stained SDS-PAGE gel. *In vitro* reconstituted human H3 and CENP-A nucleosomes on 601 arrays were a kind gift from Aaron Straight's laboratory and the array digestion in (a) was performed by A. Guse in A. Straight's laboratory (Stanford).

Following the characterisation of these recombinant human H3 and CENP-A nucleosomes they were prepared for measurement using AFM as described for *S. pombe* nucleosomes in the section entitled "Octameric CENP-A^{Cnp1} nucleosomes appear smaller than H3 nucleosomes using AFM". Using the same AFM set up and identical parameters including manual "approval" of selected particles 238 H3 nucleosomes and 239 CENP-A nucleosomes were counted. Similar to the measurements obtained with *S. pombe* nucleosomes, again, recombinant CENP-A nucleosomes were found to be significantly shorter than human recombinant H3 nucleosomes (Figure 4-11). As with the *S. pombe* data, the probability that CENP-A nucleosomes appear smaller than H3 nucleosomes by chance is again a highly significant 2.18×10^{-19}, using a one-tailed student's t-test. Interestingly, whilst recombinant human H3 nucleosomes were 0.62 nm higher and CENP-A nucleosomes were 0.46 nm higher than their *S. pombe* counterparts, CENP-A nucleosomes still show essentially the same 36 % reduction in height relative to H3 nucleosomes (*c.f.* 40 % in *S. pombe*) (Table 4-1). The ~ 40 % reduction in the height of both *S. pombe* and human *in vitro* assembled CENP-A nucleosomes relative to H3 nucleosomes observed here is consistent with the published heights obtained for fly and human nucleosomes (Dalal *et al*., 2007; Dimitriadis *et al*., 2010). This strongly suggests that this difference in apparent height is a conserved intrinsic property of CENP-A nucleosomes and that the *in vitro* assembled nucleosomes exhibit similar characteristics to their *ex vivo* counterparts with respect to measurements with AFM.

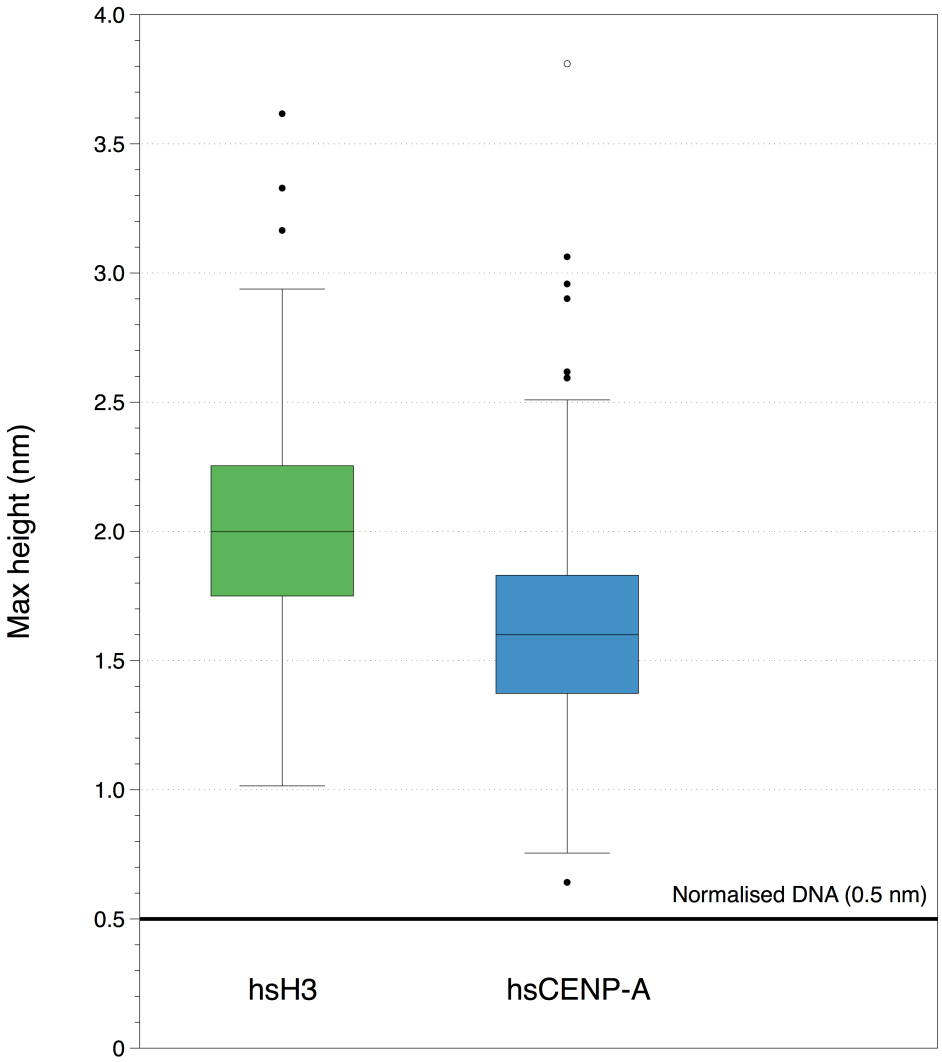


Figure 4-11 - Apparent heights of *in vitro* assembled human H3 and CENP-A nucleosomes

Recombinant human H3 and CENP-A nucleosomes assembled on arrays of 601 DNA were imaged by AFM as described in Chapter 2 (Materials and methods). The difference in modal human nucleosome heights is ~ 0.8 nm (H3 mode = 2.2 nm, CENP-A mode = 1.4 nm), compared with ~ 0.6 nm for *S. pombe* nucleosomes. Again, the central line within each box plot marks the median, the outer edges of each box represents the first and third interquartile ranges and the whiskers mark the range. Statistical outliers are

shown here as single dots. A one-tailed student's t-test shows the probability that these *in vitro* assembled CENP-A nucleosomes are shorter than H3 nucleosomes by chance is 2.18 $x10^{-19}$, making this a highly significant difference.

		Flies	Human	Human		S. pombe	Human
				G2	S	*In vitro*	
H3 nucleosome height (nm)	Mode	1.90	2.81	2.6	3.35	1.58	2.20
	Median	-	-	-	-	1.74	2.00
	Mean	-	-	-	-	1.74	2.00
	SD	-	-	-	-	0.41	0.41
CENP-A nucleosome height (nm)	Mode	1.03	1.64	1.8	2.7	0.94	1.40
	Median	-	-	-	-	1.05	1.60
	Mean	-	-	-	-	1.09	1.65
	SD	-	-	0.27	0.34	0.23	0.43
Difference (H3 - CENP-A)		0.87	1.17	0.8	0.6	0.63	0.79
CENP-A as % of H3		54	58	72	81	60	64
Reference		Dalal *et al* 2007	Dimitriadis *et al* 2010	Bui *et al* 2012		This study	

Table 4-1 - Summary of nucleosome heights recorded by AFM

Summary of AFM nucleosome height measurements collected from *in vitro* assembled chromatin compared with published AFM nucleosome height measurements for *D. melanogaster* and human CenH3 and H3 nucleosomes extracted from cells. Modal values were not provided in published height measurements, but were extrapolated as accurately as possible from the data provided.

Discussion

The objective of the analysis described in this chapter was to assess the height of nucleosome particles assembled *in vitro* and shown by other means to likely be octameric, composed of two subunits each of H3 or CenH3 plus

H4, H2A and H2B. The measurement of the height of *in vitro* prepared nucleosomes with defined octameric composition showed that CENP-A^{Cnp1} nucleosomes exhibit a marked ~ 0.6 nm reduction in apparent (modal) height compared to H3 nucleosomes, which is approximately a 40 % reduction in height (Table 4-1). A similar apparent height difference was also observed using recombinant human nucleosomes (Figure 4-11). The conclusion, therefore, is that CenH3 nucleosomes appear smaller than H3 nucleosomes using AFM not because they are hemisomes with just one copy of each histone, but because of some intrinsic physical characteristic of CenH3 nucleosomes that causes them to appear smaller than H3 nucleosomes.

The reason for this difference in height remains unknown, but may represent a difference in structure or response to the AFM tip force. It has previously been suggested that recombinant H3 nucleosomes appear shorter using AFM depending on the amount of DNA wrapped around the nucleosome (Bussiek *et al*., 2007). In their analysis Bussiek *et al* (2007) found that arrays of recombinant *X. lavis* H3 nucleosomes analysed by AFM fell into two distinct height classes. Upon closer inspection, those nucleosomes exhibiting the smaller heights were most often at the end of the arrays and, significantly, wrapped less DNA. Published analyses, along with data presented in Figure 3-8 show that in general CenH3 nucleosomes protect 10-30 bp less DNA than H3 nucleosomes (Dalal *et al*., 2007; Tachiwana *et al*., 2011a). Thus it is possible that the CenH3 nucleosomes appear smaller in height using AFM simply because DNA is wrapped less tightly at the entry / exit points of these nucleosomes. This is entirely consistent with the suggestion that post-translational modification of nucleosomes may change

their apparent height throughout the cell cycle (Bui *et al*., 2012), as observed by AFM, since modifications are known to alter the stability or structure of the nucleosome such that DNA is wrapped differently (Li *et al*., 1993; Panchenko *et al*., 2011).

The difference in height observed between human and *S. pombe* nucleosomes (~ 0.62 nm for H3, 0.46 nm for CenH3) was not wholly unexpected given the range of apparent nucleosome heights recorded between different species, and even within the same species (Allen *et al*., 1993; Dalal *et al*., 2007; Dimitriadis *et al*., 2010; Wang *et al*., 2008) (Table 4-1). Such differences are generally assumed to reflect differences in imaging conditions, such as the force applied by the tip, tip to tip variability, the extent of sample dehydration and the roughness of the surface onto which the nucleosomes are deposited (Bussiek *et al*., 2007; Dalal *et al*., 2007). However, in the analyses presented here the samples containing *S. pombe* or human nucleosomes were similarly prepared for AFM measurement and imaged alongside each other. If differences in DNA wrapping around nucleosomes does affect the nucleosome heights recorded by AFM, then one possible explanation may be that the difference in height detected between *S. pombe* and human nucleosomes reflects differences in their relative affinity for 601 DNA. It is worth noting that the 601 DNA was evolved through iterative rounds of a SELEX experiment to preferentially wrap chicken histones (Lowary and Widom, 1998) so it may interact more extensively with H3 nucleosomes. Additionally, because water is known to bridge many of the interactions between the DNA and histone core of nucleosomes (Davey *et al*., 2002) it is possible that any difference in nucleosome affinity for 601

DNA is compounded by the dehydration of nucleosomes that occurs during AFM imaging. It would therefore be interesting to compare these nucleosomes under liquid AFM conditions and also to compare the relative affinity of human and *S. pombe* nucleosomes for 601 and other DNA sequences.

Chapter 5

Sequence preferences of CENP-A^{Cnp1} and H3 nucleosomes *in vitro*

Introduction

The physical bending of ~ 147 bp of DNA around nucleosomes effectively occludes binding sites within that sequence. In the context of a gene, nucleosome occupancy can alter gene expression by blocking access to transcription factor binding sites (Richmond and Davey, 2003; Wolffe and Kurumizaka, 1998; Wyrick *et al*., 1999). To avoid nucleosomes blocking particularly important binding sites DNA sequences surrounding these sites have seemingly evolved to be refractory to wrapping nucleosomes (Field *et al*., 2009). This effect is most apparent at the promoter regions of genes (Sekinger *et al*., 2005), where a ~200 bp region directly upstream of a transcriptional start site (known as the −1 nucleosome position) is often maintained as a nucleosome-free region by the DNA sequence rather than by nucleosome remodelling complexes (Zhang *et al*., 2011). This ensures these sites are not occluded and remain available independent of the activity status of the adjacent gene. The constant availability of such sites is particularly useful at genes requiring a quick response to external stimuli (Sekinger *et al*., 2005).

DNA sequences can evolve so that they preferentially promote or prevent the assembly of nucleosomes because not all sequences are equally suited to bending around a nucleosome. As DNA wraps around the histone core it interacts with the histones at each of the 14 sites where the minor groove faces inwards (Richmond and Davey, 2003). Ensuring these interaction sites align optimally requires a combination of DNA under- and over-twisting, leading (in part) to bulging and stretching of the DNA respectively (Figure 1-5). Stretching appears to be confined to just four locations around the nucleosome, at two and five double-helical turns in each direction from the nucleosome centre (the ± 2 and ± 5 sites) and is achieved by an increased twist at these sites such that the double helix completes a turn in 9.4, rather than the usual 10.5, base pairs (Vasudevan *et al*., 2010).

Due to the energy constraints of over-twisting DNA, certain sequences are preferred at these stretched sites. Poly(dA:dT) tracts naturally exhibit a relative narrowing of the minor groove resulting in the required increase in twist (Alexeev *et al*., 1987), making such tracts energetically favourable at the ± 2 and ± 5 sites (and to a lesser extent at each of the histone : DNA contact points (Segal *et al*., 2006)) but similarly unfavourable outside of these sites (Vasudevan *et al*., 2010). This enforced pattern of localised DNA distortion as sequences wrap around a nucleosome therefore defines sequences that are correspondingly favourable and unfavourable to nucleosome occupancy (Segal *et al*., 2006). The synthetic '601' sequence was enriched for in a SELEX experiment that repeatedly selected from synthetic derivatives of canonical alpha satellite repeat sequence to evolve a sequence which has a high preference for the assembly of H3 nucleosomes in *in vitro* reconstitutions

(Lowary and Widom, 1998). The resulting '601' DNA fragment contains the expected 5 bp periodicity of poly(dA:dT) and poly(dC:dG) dinucleotides (particularly over the H3:H4 tetramer region) that would be energetically favourable for DNA over- and under-twisting around a nucleosome and was correspondingly shown to produce particularly stable nucleosomes (Thåström *et al*., 2004c) (Thåström *et al*., 2004a).

When the promoter associated −1 nucleosome-free region of genes are assessed, it is the inclusion of relatively A/T -rich sequences that are unfavourable to nucleosome formation and that are largely responsible in keeping these regions nucleosome-free (Mai *et al*., 2000). The DNA sequences associated with eukaryotic centromeres are relatively A / T -rich. *S. pombe* centromeric central domain regions are ~ 72 % A/T compared with ~ 65 % across the whole genome, whilst *S. cerevisiae* centromeres are ~ 86 % A/T, compared with ~ 61 % across the whole genome. The increased A/T content of centromeres relative to the rest of the genome predicts that these centromeres are relatively unfavourable to H3 nucleosomes, yet CenH3 nucleosomes are found exclusively at centromeres (Kaplan *et al*., 2009; Locke *et al*., 2010; Warburton *et al*., 1997). Thus the question is why CenH3 nucleosomes are preferentially assembled at centromeres. Given the altered structure of the CENP-A histone octamer (Tachiwana *et al*., 2011a) it is possible that CenH3 nucleosomes have evolved an altered sequence preference which differs to that of H3 nucleosomes. CenH3 nucleosomes may have co-evolved with the A/T-rich sequence of centromeres, which are generally predicted to be unfavourable for the assembly of canonical H3 nucleosomes. In recent years the contribution of DNA sequence to directing

H3 nucleosome location has received a great deal of interest, but the extent to which DNA directs CenH3 nucleosome location has not been investigated in depth. In this chapter the aim is to test whether recombinant *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes exhibit any differences in their preferred sites of assembly on *S. pombe* genomic DNA. For this purpose sequence preference is defined as the consensus sequence for nucleosome positioning. This definition is in line with current discussions concerning H3 nucleosome positioning (Kaplan *et al*., 2009; 2010; Zhang *et al*., 2009; 2010).

Results

CENP-A^{Cnp1} nucleosomes are more stable on centromeric sequences than euchromatic sequences

Examination of centromeric sequences in a variety of eukaryotes indicates that they are relatively A/ T -rich (Koch, 2000; Lomiento *et al*., 2008) and therefore predicted to be unfavourable to canonical nucleosome occupancy (Segal *et al*., 2006). If CENP-A^{Cnp1} nucleosomes have an altered nucleosome sequence preference so that they associate with distinct DNA sequences from H3 nucleosomes then one hypothesis is that CENP-A^{Cnp1} nucleosomes may have adapted so they are better suited to assemble on centromeric sequences than H3 nucleosomes. It is also possible that CENP-A^{Cnp1} nucleosomes may still be capable of occupying the same sequences as H3 nucleosomes, but would be more stable than H3 nucleosomes on a 'poor' nucleosome forming sequence, such as a centromeric sequence. If this was the case then it might allow CenH3 nucleosomes to persist on centromeric sequences whereas H3

nucleosomes would be relatively unstable and may be lost. To study the lower limits of nucleosome stability recombinant *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes were compared on regular "euchromatic" and "centromeric" DNA sequences from regions over which CENP-A^{Cnp1} nucleosomes are known to assemble *in vivo*.

Previously, MNase released H3 and CENP-A^{Cnp1} nucleosomes have been sequenced following immunoprecipitation from fixed chromatin (ChIP-Seq) (Lando *et al*., 2012). Six different 200 bp sequences were chosen from these datasets where well-defined H3 and CENP-A^{Cnp1} nucleosome positioned peaks were evident (sequences identified by H. Berger, Allshire research group). Three H3 peaks were chosen as representative euchromatic nucleosome sequences from the genome and three CENP-A^{Cnp1} peaks from the central domain regions were chosen as representative centromeric sequences - the exact sequences and their locations can be found in Appendix 3. These six fragments were amplified by PCR with Ava1 sites engineered into both primers, were cut with Ava1, and a small aliquot of the DNA end-labelled by filling in with P^{32} dCTP. Nucleosomes reconstituted using a 1:10 spike of radio-labelled DNA to cold DNA were diluted over a 200-fold range of final concentrations (from 10 nM to 0.05 nM) using a simple buffer defined by Thastrom *et al* (2004), containing 0.01 % Igepal CA-630 (formerly known as Nonidet NP-40) and 50 mM NaCl. Analysis of the samples by native PAGE separates nucleosomes from free DNA. Subsequent phorphorimaging of the radio-labelled DNA allowed quantification of the fraction of DNA that had assembled into nucleosomes. The stability of nucleosomes assembled on different DNA sequences is then defined as the

concentration at which half of the nucleosomes present at the highest concentration have disassembled. This is measured as the concentration at the inflection point of a sigmoidal function fitted to the data by a least-squares regression. Whilst dilution experiments such as these cannot be used to identify the free energy of histone-DNA binding they do report on kinetic (or effective) nucleosome stability, which is arguably a more biologically relevant measure of stability (Thåström *et al*., 2004b).

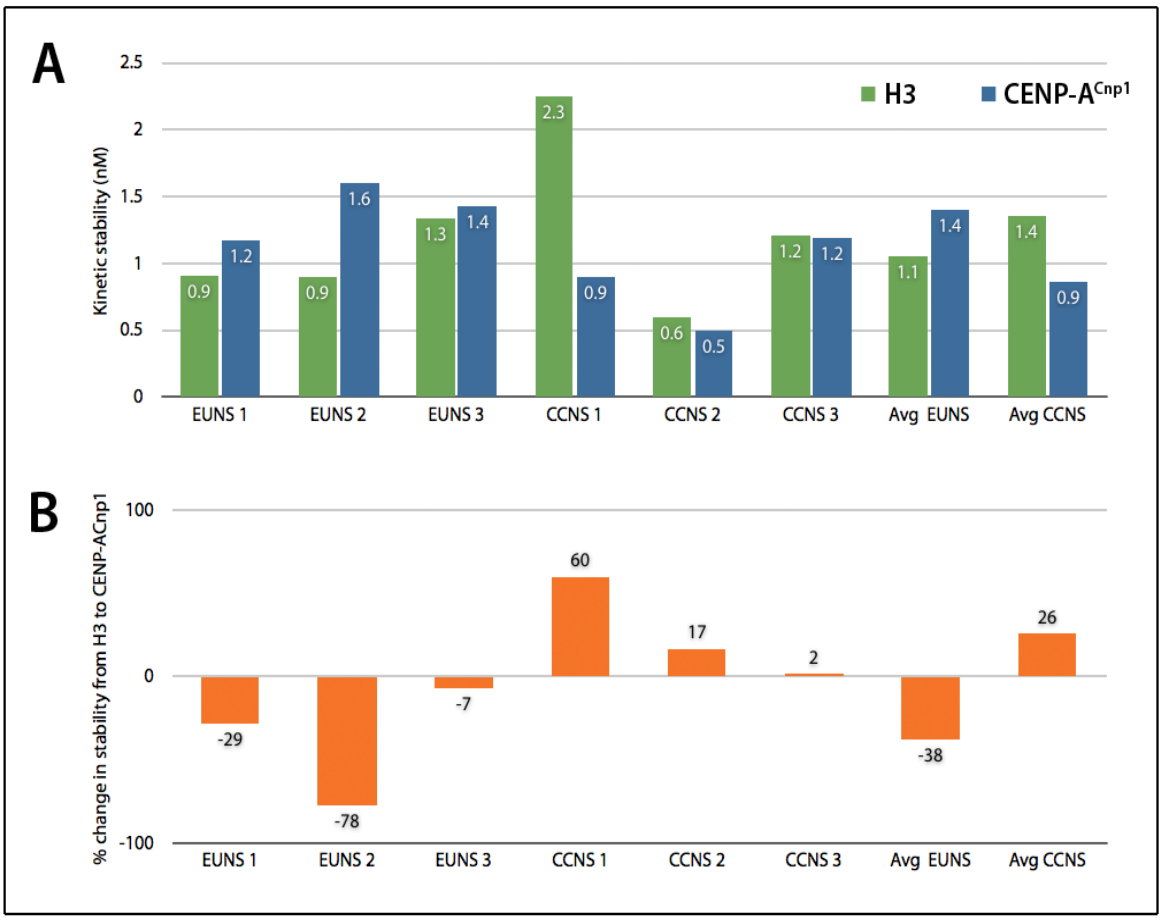


Figure 5-1 - Kinetic stability of *S. pombe* recombinant H3 and CENP-A^{Cnp1} nucleosomes assembled on different DNA substrates *in vitro*

CENP-A^{Cnp1} and H3 nucleosomes were reconstituted from recombinant *S. pombe* histones onto three euchromatic (EUNS) and three centromeric (CCNS) radiolabelled DNA sequences. As nucleosomes are diluted over a 200-fold range between 10 – 0.05 nM they increasingly dissociate and the relative ratio of nucleosomal to free DNA decreases (Gottesfeld and Luger, 2001). To determine the concentration at which 50 % of the nucleosomes remain intact (relative to the 10 nM concentration) the diluted nucleosomes were run on native PAGE gels to separate nucleosomal from free DNA. The ratios of which were used to fit a sigmoidal curve of best fit to the dilution-driven dissociation, with the inflection point defining the concentration at

which 50 % of the nucleosomes remain intact. This concentration represents the kinetic stability of the H3 or CENP-A^{Cnp1} nucleosomes on these sequences and is plotted here (a) along with the average (mean) stability across the three EUNS and CCNS sequences. Higher nM values for kinetic stability refer to less stable nucleosomes. See Chapter 2 (material and methods) for a full description of this process. For comparison, the lower graph (b) plots the relative difference in stability between CENP-A^{Cnp1} and H3 nucleosomes on each sequence, such that negative values represent a sequence on which H3 forms more stable nucleosomes than CENP-A^{Cnp1}.

The results show that recombinant *S. pombe* CENP-A^{Cnp1} nucleosomes are, on average, ~ 1.6 x more stable on these centromeric sequences than on the euchromatic sequences tested, whilst H3 nucleosomes are only ~ 0.8 x as stable (Figure 5-1). On the euchromatic sequences, H3 nucleosomes are ~ 1.3 x more stable than CENP-A^{Cnp1} nucleosomes. Whilst on the centromeric sequences, CENP-A^{Cnp1} nucleosomes were ~ 1.6 x more stable than H3 nucleosomes. However, only a small number of sequences were tested here for H3 and CENP-A^{Cnp1} nucleosome stability, ideally this experiment would be repeated to increase confidence in the data. The relative stabilities of H3 and CENP-A^{Cnp1} nucleosomes vary significantly between the different DNA sequences used, with equal stability on CCNS 3, but a 78 % difference on EUNS 2. Furthermore, whilst on average H3 nucleosomes are more stable on euchromatic DNA, H3 nucleosome stability on CCNS 2 is twice that of H3 nucleosomes on EUNS 3. Thus not all sequences are equal in terms of nucleosome stability and distinctions cannot be as broad as a simple

comparison between centromeric and euchromatic sequences. Nevertheless, overall the analysis of this limited set of centromeric and euchromatic sequences suggests that H3 and CENP-A^{Cnp1} nucleosomes require different DNA sequences for optimum nucleosome stability *in vitro*; and for CENP-A^{Cnp1} nucleosomes, these sequences with better stability are associated with the regions of centromeres over which CENP-A^{Cnp1} nucleosomes are formed *in vivo*. Since the stability of nucleosomes can vary greatly on different sequences, an unbiased approach is required to determine the inherent sequence preferences of H3 and CENP-A^{Cnp1} nucleosomes. For this purpose the sequences selected from *S. pombe* genomic DNA in large scale assembly of H3 and CENP-A^{Cnp1} nucleosomes needs to be examined.

CENP-A^{Cnp1} and H3 nucleosome sequence preferences differ

High throughput sequencing allows the sequences associated with nucleosomes to be assessed on a much larger scale. Such methods have been used to determine the sequence preference of H3 nucleosomes (Kaplan *et al*., 2009; Locke *et al*., 2010; Segal *et al*., 2006). CENP-A^{Cnp1} or H3 nucleosomes were reconstituted onto *S. pombe* genomic DNA at low density (10 % occupancy saturation) to test whether the sequences selected for the assembly of *S. pombe* CENP-A^{Cnp1} nucleosomes *in vitro* differ significantly from those selected by H3 nucleosomes. Assembly at a low histone : DNA ratio was used to ensure that nucleosomes had a large excess of DNA on which to form, promoting an element of selectivity whereby nucleosomes are formed most frequently on the most favourable sites for assembly.

Previously, similar conditions have been used to reconstitute chicken erythrocyte or *D. melanogaster* histones onto *S. cerevisiae* genomic DNA at an occupancy level of 40 % (Kaplan *et al*., 2009) or 33 % (Zhang *et al*., 2009) respectively, meaning there was a 2.5-fold or 3-fold excess of genomic DNA relative to the molar amount of histone octamer (assuming a nominal DNA wrapping length of 147 bp). To discourage low affinity background variations and promote high selectivity of nucleosomes for their preferential sequences the conditions implemented here utilise an even lower nucleosome occupancy level than that implemented by Kaplan *et al* (2009) and Zhang *et al* (2009).

Preliminary assays were performed to gauge the efficiency of nucleosome assembly and suggested a 10 % occupancy level gave the most appropriate trade-off between lower occupancy levels that give sequence selectivity and the quantity of wrapped DNA required for analysis. In this trial experiment, a 10 % occupancy level meant a 10-fold molar excess of DNA (nominal length of 147 bp) to histone octamer, such that the reconstitution should result in one nucleosome for every 10 potential nucleosome sites across the genome. For reference, the genomic DNA used here was purified from the "972" wildtype h- strain of *S. pombe* (Leupold, 1958).

To guard against the possibility that 10 % occupancy was not sufficiently selective to allow for clear identification of strong nucleosome peaks in the data, nucleosomes were also assembled at two higher occupancy levels. *S. pombe* genomic DNA was assembled into H3 and CENP-A^{Cnp1} nucleosomes at 20 % and 40 % occupancy levels in parallel. By comparing nucleosome peaks across these three samples it was envisaged that it would be possible

to identify trends in the data; for example, preferential sequences might become progressively enriched as selectivity increased, whilst spurious sub-optimal (background) sequences would be diminished.

This experiment is designed to identify sequences that H3 and CENP-A^{Cnp1} nucleosomes prefer to occupy (those that form the most stable nucleosomes), rather than the sequences that histone octamers might prefer to reconstitute onto. By 'heat-shifting' the nucleosomes after reconstitution, a process by which nucleosomes are incubated at 37 C for 2 h, nucleosomes are able to reposition themselves by sliding along the DNA (Muthurajan *et al.*, 2004). This heat-shift ensures nucleosomes adopt the most thermodynamically favourable positions within a region of DNA.

Once repositioned nucleosomes were digested with MNase so that mono-nucleosomal fragments predominated in the population. Nucleosomal DNA was then purified by treatment with Proteinase K and Phenol extraction. The resulting DNA was separated on agarose gels and a region of mostly mono-nucleosome sized fragments (100 – 200 bp) was extracted using a Qiagen gel extraction kit (Figure 5-2). The resulting DNA, pooled from three experimental replicas, was processed for paired-end library preparation and Illumina sequencing by the commercial company BGI (http://www.genomics.cn/en/index).

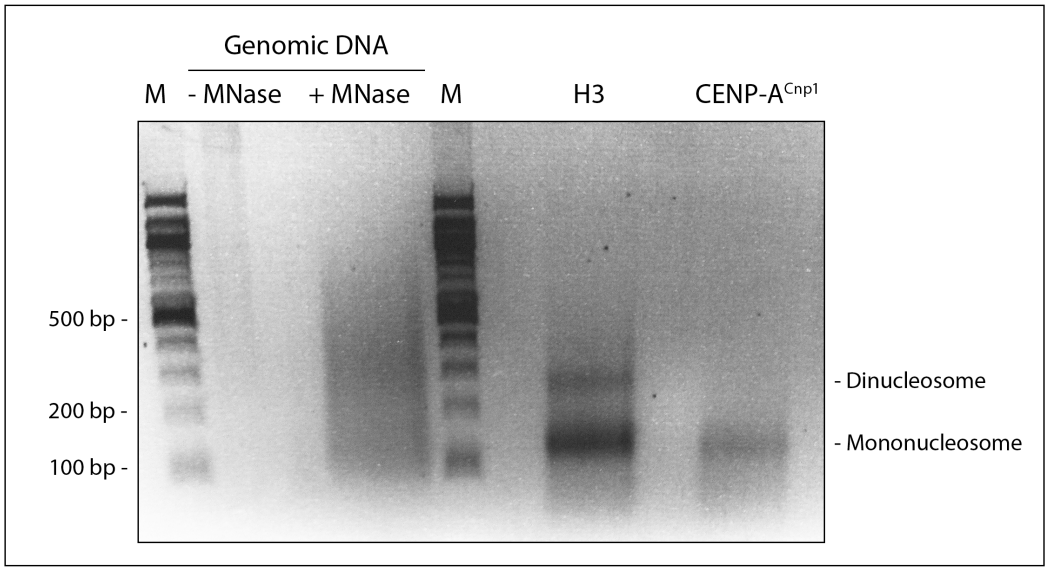


Figure 5-2 - Purification of S. pombe H3 and CENP-ACnp1 nucleosomal DNA for sequencing

S. pombe H3 and CENP-A^{Cnp1} nucleosomes reconstituted onto high molecular weight *S. pombe* genomic DNA at 10, 20 and 40 % occupancy levels were digested with MNase to release individual nucleosomes. Purified nucleosomal DNA was separated in agarose gels to isolate the mono-nucleosome fragments (100 – 200 bp), which were extracted from the gel, pooled between three separate reconstitutions and prepared for Illumina sequencing. This figure shows an example agarose gel for one of the 10 % occupancy samples of H3 and CENP-A^{Cnp1} nucleosomes. As an input control, naked genomic DNA was digested with much less MNase to obtain average fragment sizes around 150 bp. The genomic DNA extracted from the gel was between 100 – 200 bp.

There are two principal measures of nucleosome location with respect to DNA, nucleosome occupancy and nucleosome positioning (Pugh, 2010) (Figure 5-3). Nucleosome occupancy refers to the probability that a given

region is wrapped anywhere within a nucleosome and is a measure of nucleosome density. Nucleosome positioning is the probability that a given region aligns with a defined nucleosomal feature, typically the dyad axis. This distinction is important when comparing nucleosome locations between two datasets because high nucleosome occupancy does not necessarily correlate with high nucleosome positioning, and *vice versa*. Comparing nucleosome positioning peaks between two samples with large standard deviations (low nucleosome positioning) is of less statistical relevance because the probability that a given region is occupied in both samples may be significantly lower. Therefore, whilst nucleosome occupancy is more likely to be of biological significance (as a occluded nucleotide is occluded no matter where it sits within a nucleosome) when comparing the overlap of nucleosome locations between two samples it is more accurate to compare samples with strong nucleosome positioning.

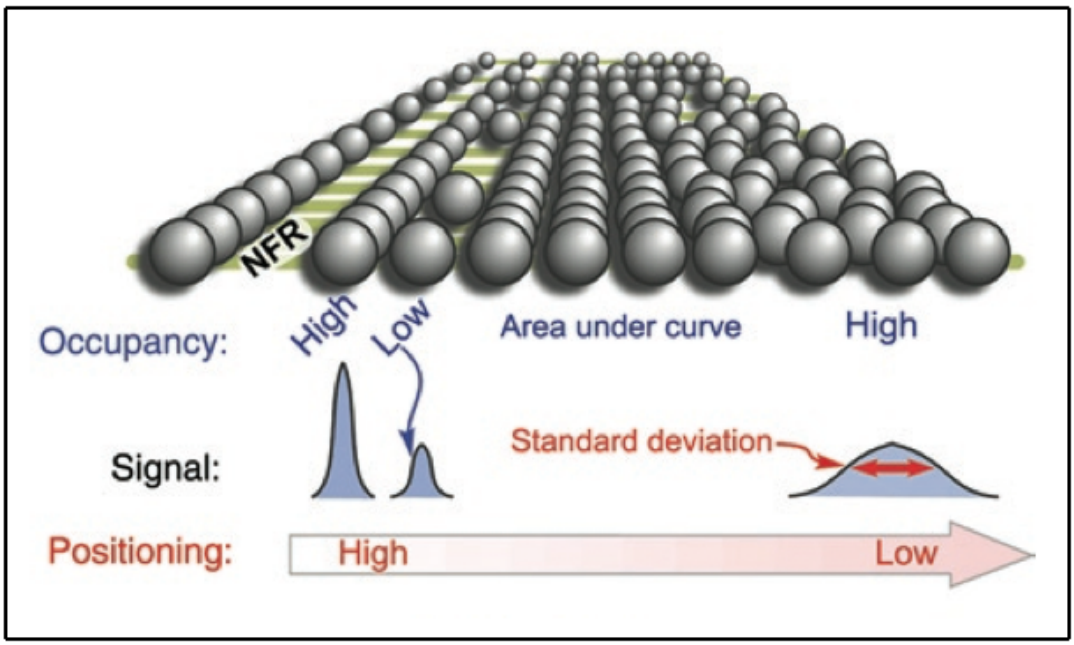


Figure 5-3 - Illustration highlighting the differences between nucleosome occupancy and nucleosome positioning

Nucleosome occupancy refers to the density of nucleosomes for a certain region within a population. When the positions of a particular nucleosomal feature, typically the dyad axis, are mapped to the DNA sequence from a population of nucleosomes (represented here as the signal peaks) nucleosome occupancy at a given region corresponds to the area under the signal peak. Nucleosome positioning is a measure of how well the nucleosomes are aligned and is represented here by the standard deviation of the peak. Figure adapted from Pugh *et al* (2010).

There has been much debate concerning the influence of DNA sequence on nucleosome location, with some confusion seemingly arising from the interchangeable use of the terms "occupancy" and "positioning" (Pugh, 2010; Zhang *et al*., 2010). The current consensus, based mainly on the analysis of H3 nucleosomes on *S. cerevisiae* genomic DNA appears to be that DNA

sequence accounts for ~ 20 % of *in vivo* nucleosome positioning, with a correlation of ~ 0.3 (where 1 is a perfect correlation and 0 is no correlation) between *in vivo* and *in vitro* nucleosome positioning (Stein *et al*., 2010; Zhang *et al*., 2010). These values for positioning were determined using a 23 bp window centred on the dyad axis. Theoretically, positioning should be measured for a single nucleotide, but often a larger window of 23 bp, or even as much as 40 bp, is used to allow for error in the definition of the central dyad axis - typically due to uneven MNase digestion from the two ends of a nucleosome (Zhang *et al*., 2010). In the analysis performed in this chapter, the use of paired-end libraries enabled direct mapping of both the 5′ and 3′ ends of each fragment, in essence mapping both the nucleosome entry and exit points, as previously described (Ercan *et al*., 2011). This allows for greater confidence in defining the dyad axis as the central nucleotide between the 5′ and 3′ ends and so here positioning is defined as an alignment within a window of the central 7 nucleotides for each fragment. The assumption was, however, that MNase had digested evenly at both ends. Although the distribution of fragment sizes obtained was a tight range for each of the 6 samples (H3 and CENP-A^{Cnp1} at either 10, 20 or 40 % occupancy levels), to exclude the possibility that potential sub-populations of uneven digestion (fragments smaller or larger than the modal fragment size) may confound analysis the fragments were mapped in discrete size bins (performed by H. Berger, R. Allshire research group) (Figure 5-4). As all fragment sizes show essentially the same pattern across the genome uneven MNase digestion was not considered to be a significant factor and consequent analysis used the full range of fragments collected.

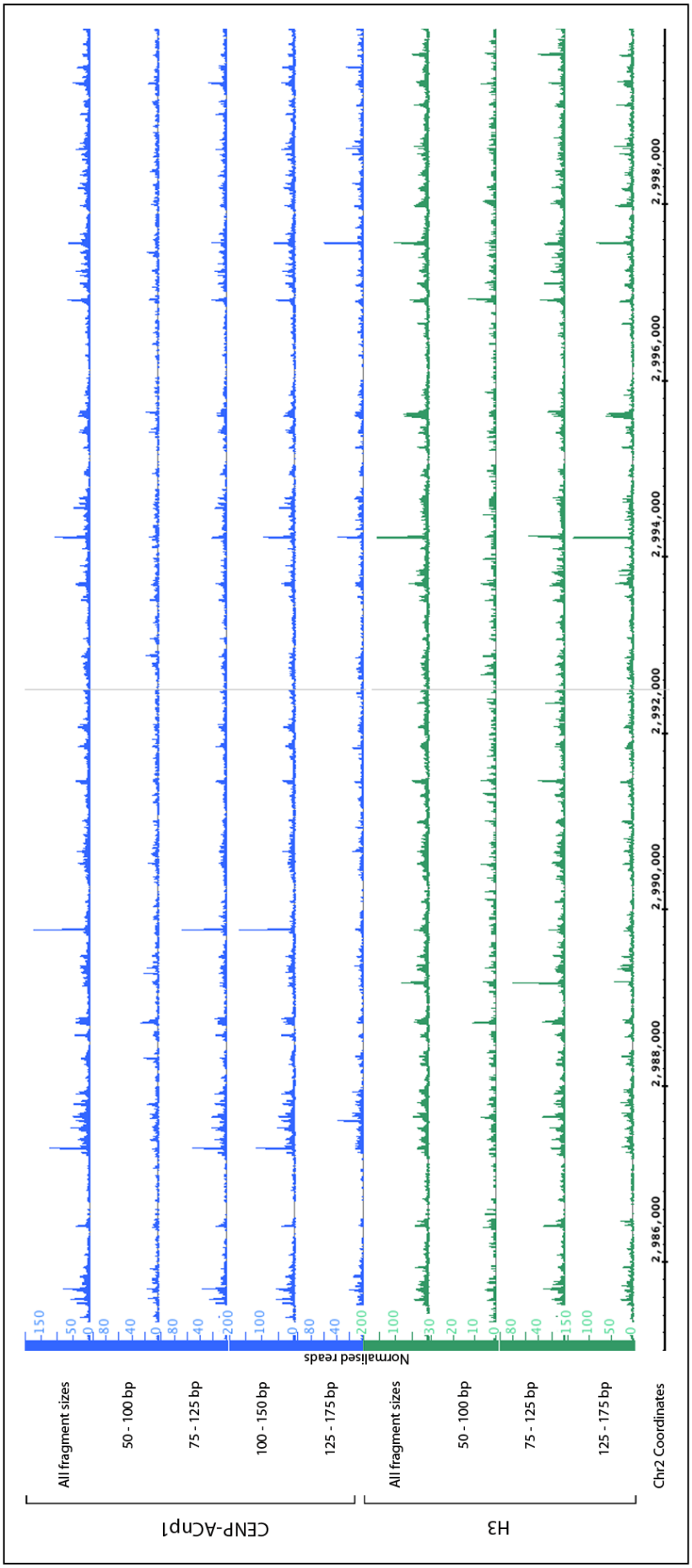

CENP-ACnp1
H3
All fragment sizes
50 - 100 bp
75 - 125 bp
100 - 150 bp
125 - 175 bp
All fragment sizes
50 - 100 bp
75 - 125 bp
125 - 175 bp
Normalised reads
Chr2 Coordinates
2,986,000
2,988,000
2,990,000
2,992,000
2,994,000
2,996,000
2,998,000

Figure 5-4 - Mapped nucleosome positioning reads of recombinant H3 and CENP-A^{Cnp1} nucleosomes assembled onto *S. pombe* genomic DNA
Normalised positioning reads from both H3 and CENP-A^{Cnp1} *in vitro* reconstitutions onto *S. pombe* genomic DNA at 10 % occupancy are shown (top rows in green and blue, respectively). For comparison, a range of fragment length bins arbitrarily spread over the range of read lengths are also shown for both H3 and CENP-A^{Cnp1} samples. Image shows data for a randomly chosen ~ 15 kb region of chromosome 2, between bases ~ 2,985,000 – 3,000,000.

Perhaps the most immediately striking feature of the nucleosome positioning data is how well H3 and CENP-A^{Cnp1} positioning correlates throughout the genome (Figure 5-4). The spearman's correlation coefficient across 6 random genomic locations is 0.71, for the data generated from the 10 % occupancy sample. Contrary to our initial hypothesis, this data appears to suggest that H3 and CENP-A^{Cnp1} nucleosomes have surprisingly similar sequence preferences for wrapping DNA. Of course it is plausible that the 10 % occupancy parameter is simply not stringent enough to reveal subtle but significant differences between the two nucleosome types. However, comparing trends across the 40, 20, and 10 % data also shows no significant change in nucleosome positioning (not shown - performed by N. Toda, R. Allshire research group). This indicates that there is no observable increase in sequence selectivity between 40 and 10 % occupancy. Thus this again suggests that there is no significant difference in the sequence positioning preferences of H3 and CENP-A^{Cnp1} nucleosomes.

However, on closer inspection it is noticeable that although the location of the peaks correlate well between all H3 and CENP-A^{Cnp1} samples, the relative peak heights often differ. This difference suggests that H3 and CENP-A^{Cnp1} nucleosomes may indeed differ in their relative preference for particular DNA sequences, albeit subtly.

To examine differences between the sequence preference of H3 and CENP-A^{Cnp1} nucleosomes with greater sensitivity the top 5 % of H3 and CENP-A^{Cnp1} nucleosome peaks (representing the best nucleosome positioning sequences) were chosen for a comparative analysis of their sequence preference. Because the ability of any DNA sequence to assemble a nucleosome is essentially determined by the positioning of poly(dA:dT) and poly (dC:cG) dinucleotides around the nucleosome (Locke *et al*., 2010; Segal *et al*., 2006) (described above), differences in these frequencies are likely to represent differences in the consensus sequence preference of each nucleosome type. In mapping the relative frequency of AA/AT/TA/TT and CC/CG/GC/GG dinucleotides (performed by H. Berger, R. Allshire research group) it became clear these most strongly positioned CENP-A^{Cnp1} and H3 nucleosomes (the top 5 %) clearly differ in their DNA sequence preference (Figure 5-5). For comparison, the dinucleotide frequencies of H3 nucleosomes from data collected here match well with those of *in vivo* H3 nucleosome sequencing data previously collected in our lab (H. Berger, unpublished personal communication), particularly over the 40 bp surrounding the dyad axis, where the biggest differences are observed between *in vitro* assembled H3 and CENP-A^{Cnp1} nucleosomes.

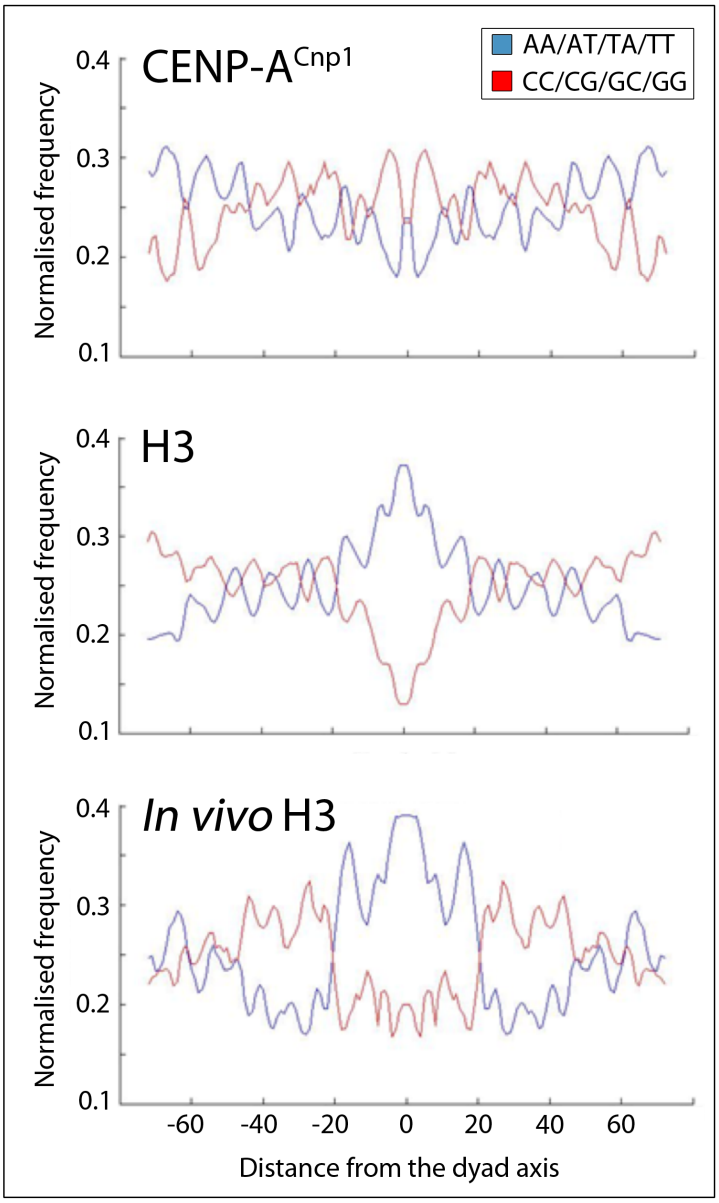


Figure 5-5 - Dinucleotide frequency around nucleosome positioning sequences for H3 and CENP-A^{Cnp1} nucleosomes assembled *in vitro*

To compare the sequence preference of nucleosomes the top 5 % of nucleosome positioning peaks were aligned and the relative frequencies of AA/AT/TA/TT and CC/CG/GC/GG dinucleotides calculated across each base pair. The graphs here show these relative frequencies for CENP-A^{Cnp1} (top) and H3 (middle) nucleosomes from the *in vitro* experiment described above as well as from existing *in vivo* H3 positioning data (H. Berger, unpublished personal communication) for comparison (bottom).

To determine the extent to which nucleosome stability influences nucleosome location on genomic DNA in these *in vitro* reconstitutions a comparison was made between the nucleosome occupancy and stability for each of the six specific sequences tested initially in nucleosome stability assays (Figure 5-6). If nucleosome stability is a major determinant of final nucleosome location then one would expect the stability to match well with nucleosome occupancy on that sequence. The average occupancy value was calculated for the central 100 bp of the 200 bp within the sequences used for the stability assays. This gave an estimation, in arbitrary units, of the probability that a nucleosome was located within that 200 bp sequence. To compare these data with the stabilities calculcated both data sets were normalised internally. This was achieved by normalising to the value of the EUNS 1 sequence. For ease of comparison the data are presented separately for H3 (Figure 5-6 A) and CENP-A^{Cnp1} (Figure 5-6 B) nucleosomes. For H3 nucleosomes the relative stability and occupancy values correlate rather well over most of the sequences tested, suggesting that stability is a major factor determining nucleosome occupancy at these sites. However, for CENP-A^{Cnp1} nucleosomes, stability and occupancy match well over (euchromatic) EUNS sequences but differ significantly over the (centromeric) CCNS sequences. This might indicate that stability is not the major determinant of nucleosome occupancy on centromeric sequences. Why CENP-A^{Cnp1} nucleosome occupancy does not correlate with nucleosome stability on centromeric sequence is unknown, but it is not due to the nature of the DNA sequence itself since H3 nucleosome occupancy profiles correlate with their stability on the same sequences. Thus it would appear to be a specific feature of the

interaction between CENP-A^{Cnp1} nucleosomes and the centromeric DNA sequences tested. It is possible that this finding highlights a general disparity between sites favouring CENP-A^{Cnp1} nucleosome reconstitution and nucleosome stability. Such an effect may be more pronounced at centromeres because large stretches of the sequence are homopolymers, which are known to resist nucleosome occupancy (Field *et al.*, 2008). Inhibiting the ability of nucleosomes to reposition after reconstitution might account for the reduction in occupancy observed for *in vitro* assembled CENP-A^{Cnp1} nucleosomes on the same centromeric sequences that formed relatively stable CENP-A^{Cnp1} nucleosomes in initial nucleosome stability assays.

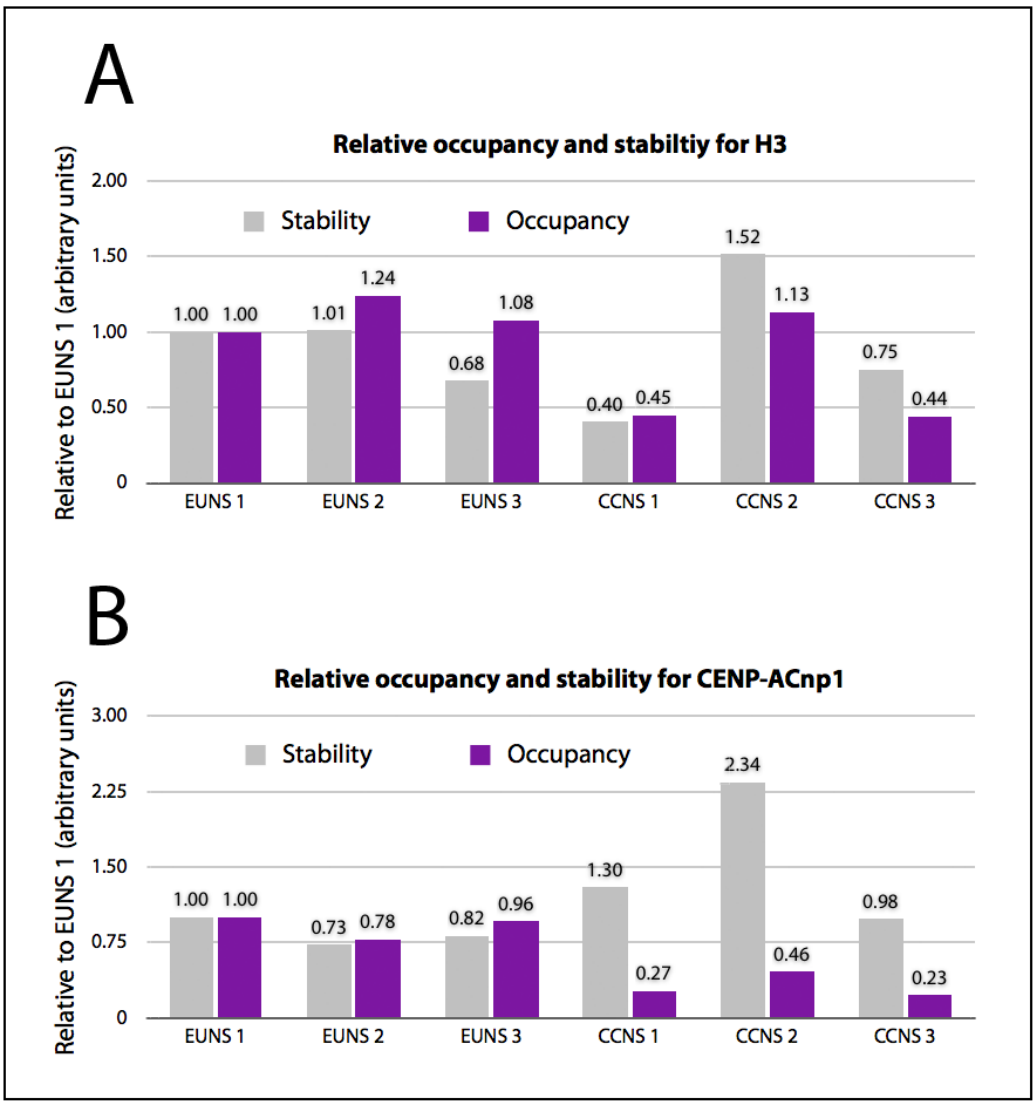


Figure 5-6 - Comparison of in vitro assembled nucleosome stability and occupancy

Nucleosome occupancy was estimated from the genome-wide *in vitro* reconstitution data for H3 and CENP-A^{Cnp1} nucleosomes at each of the representative euchromatic (EUNS) and centromeric (CCNS) sites used in the specific nucleosome stability assay presented in (Figure 5-1). To enable comparison between the relative stability and occupancy at each site stability and occupancy values were normalised to their EUNS 1 values. Data for H3 (a) and CENP-A^{Cnp1} (b) nucleosomes are shown separately to aid comparison between the occupancy and stability at each site.

When the *in vitro* reconstitution data presented here is compared with *in vivo* datasets derived from CENP-A^{Cnp1} ChIP-Seq experiments (Allshire laboratory - in publication) the low spearman's correlation score of 0.17 for positioning within the CENP-A^{Cnp1} domain of centromeres 1, 2 and 3 suggests that primary DNA sequence itself has relatively little influence in the positioning of CENP-A^{Cnp1} nucleosomes within cells (Figure 5-3). However, for this purpose, comparing nucleosome positions may not be the most meaningful measure. As described above, well positioned nucleosomes have a high probability that nucleosomes will be precisely aligned at that location within a given population. Thus the comparison of nucleosome positioning between two datasets accurately reflects the correlation of exact nucleosome locations between the two samples. In contrast, nucleosome occupancy is a measure of nucleosome density at a given site but gives little information about the precise location of nucleosomes. In the data presented here the standard deviations of nucleosome read peaks, which is a measure of nucleosome positioning, are well within the size of a single nucleosome for both the *in vivo* extracted and *in vitro* assembled samples as a measure of nucleosome location. Thus, occupancy can reliably be compared between samples. For this reason, occupancy measurements were compared for sequences associated with *in vitro* assembled and *in vivo* extracted CENP-A^{Cnp1} nucleosomes. In this case the spearman's correlation score is 0.4, which is a moderate correlation. However, the similarity between sequencing data for H3 and CENP-A^{Cnp1} *in vitro* reconstitutions means that in many cases the DNA is equally likely to assemble an H3 nucleosome in the same position.

Discussion

In this chapter the contribution of DNA sequence to CENP-A^{Cnp1} nucleosome positioning has been assessed. Illumina sequencing has allowed for the high resolution mapping of H3 and CENP-A^{Cnp1} nucleosomes assembled *in vitro* at low density onto *S. pombe* genomic DNA. The low density of these reconstitutions ensures that nucleosomes are free select from a large pool of DNA sequences such that the better positioned nucleosomes represent the most energetically stable sequences for nucleosomes to occupy, within the constraints of their local energy landscape.

Nucleosome stability assays were also used to compare H3 and CENP-A^{Cnp1} nucleosome stability across three centromeric, and three euchromatic DNA sequences (Figure 5-1). On average, CENP-A^{Cnp1} nucleosomes were ~ 1.6 x less stable on euchromatic sequences than centromeric sequences. Correspondingly, H3 nucleosomes were, on average ~ 1.3 x more stable than CENP-A^{Cnp1} nucleosomes on these euchromatic DNA sequences and ~ 1.6 x less stable than CENP-A^{Cnp1} nucleosomes on the centromeric sequences tested. Although, the significance of these comparisons is limited by the small number of DNA sequences that can be analysed using this technique. Nevertheless, these analyses suggests that H3 and CENP-A^{Cnp1} nucleosome sequence preferences differ. Previously, an inherent instability has been proposed as a functional role for H3.3 nucleosomes, which were also shown to be less stable than canonical H3 nucleosomes extracted from a chicken 6C2 cell line. H3.3 nucleosomes are thus proposed to promote the accessibility of enhancer and transcribed regions through the facilitated removal of these inherently unstable nucleosomes (Jin and Felsenfeld, 2007). Similarly, if data

presented here are representative of a general trend of nucleosome stability across the genome, then this reduction in the stability of CENP-A^{Cnp1} nucleosomes on euchromatic sequences (relative to that of H3 nucleosomes on the same sequences) might facilitate the removal of CENP-A^{Cnp1} nucleosomes from euchromatic regions following incorporation of CENP-A^{Cnp1} at non-centromeric regions.

The occupancy values of 10, 20 and 40 % used for the reconstitution experiments described here represent the molar ratio of input DNA (nominally at 147 bp) to histone octamer. The actual occupancy levels of nucleosomes after reconstitution are likely to be significantly lower due to inefficiency in the reconstitutions, as not all histone octamer will form nucleosomes. In our hands, reconstitutions of H3 and CENP-A^{Cnp1} mono-nucleosomes onto specific sequences appear to be roughly equivalent in efficiency, with the percentage of input DNA reconstituted into nucleosomes typically about 60 % for both H3 and CENP-A^{Cnp1} nucleosomes (Figure 3-5). Therefore both H3 and CENP-A^{Cnp1} samples are expected to have comparable occupancies at each of the stated occupancy levels. It is worth noting that the occupancy levels used here are up to 4 x more selective than similar experiments conducted previously (Segal *et al*., 2006; Zhang *et al*., 2009) and that data generated here reveals a 4-fold difference in nucleosome occupancy has no significant effect on either H3 or CENP-A^{Cnp1} nucleosome positioning. Thus small differences in occupancy between H3 and CENP-A^{Cnp1} nucleosome samples are unlikely to alter the data.

Many factors are known to influence nucleosome positioning *in vivo* (Segal and Widom, 2009), so it was perhaps unsurprising to observe little

correlation between the exact positioning of H3 and CENP-A^{Cnp1} nucleosomes assembled *in vitro* and those extracted from cells. Nevertheless, the correlation between CENP-A^{Cnp1} nucleosome occupancy *in vitro* and *in vivo* (spearman's correlation score of 0.4 across the three centromeres) suggests centromeric DNA sequence alone influences nucleosome location *in vivo*. A conclusion that is in excellent agreement with those of previous similar studies of H3 nucleosomes (Segal *et al*., 2006; Zhang *et al*., 2009). The observation that *in vivo* CENP-A^{Cnp1} nucleosome occupancy correlates with the nucleosome occupancy of both H3 and CENP-A^{Cnp1} nucleosomes assembled *in vitro* also fits with the strong correlation (spearman's correlation score of 0.71) between *in vitro* assembled CENP-A^{Cnp1} and H3 nucleosome positioning observed here. Thus *S. pombe* genomic DNA appears sufficient to define the *in vivo* nucleosome occupancy of a given region but is largely unable to define the exact positioning of nucleosomes and does not appear to distinguish between H3 or CENP-A^{Cnp1} nucleosome occupancy, at least at the genome-wide scale.

Whilst the positions of *in vitro* assembled H3 and CENP-A^{Cnp1} nucleosomes were strongly correlated there were also differences. Examination of the relative peak heights demonstrates that whilst both nucleosome types tended to occupy the same positions, they often did so to varying degrees. Together with stability data showing subtle differences in the stabilities of H3 and CENP-A^{Cnp1} nucleosomes on the same specific sequences (Figure 5-1), differences in the average dinucleotide frequencies were observed in the top 5% of positioned sequences (Figure 5-5), suggesting that *in vitro* assembled H3 and CENP-A^{Cnp1} nucleosomes do have

subtle differences in their sequence preferences. As both nucleosome types appear to select the same sequences, but to different degrees, it would be informative to repeat the whole-genome *in vitro* reconstitutions performed here in a competitive manner. With both H3 and CENP-A^{Cnp1} octamers in the same reconstitution one would expect the probability that a given site forms either a H3 or CENP-A^{Cnp1} nucleosome to be a function of their relative stabilities for that sequence. A decreased stability of CENP-A^{Cnp1} nucleosomes outside centromeres would thus be predicted to produce a larger difference between CENP-A^{Cnp1} and H3 nucleosome positioning than was observed here. Also, CENP-A^{Cnp1} nucleosomes may better match their *in vivo* positions as H3 nucleosomes might be expected to occupy the less favourable CENP-A^{Cnp1} nucleosome sites. This experiment would be informative for determining the sequence preference of CENP-A^{Cnp1} nucleosomes since the competition of CENP-A^{Cnp1} nucleosomes with H3 nucleosomes effectively increases the stringency above that used here.

One possibility for why a difference in nucleosome preference was only apparent when analysing the top 5 % of positioning peaks is that factors limiting nucleosome repositioning, such as long stretches of A/T nucleotides, caused an underrepresentation of the most stable sequences. In this type of experiment, nucleosomes reconstitute by first loading the H3 : H4 / CENP-A^{Cnp1} : H4 tetramer and later binding the H2A : H2B dimers. This means that the initial selection of whether a nucleosome forms or not is defined by the tetramer, and may differ from the preference of the full nucleosome. To address this issue, this experiment had included a nucleosome repositioning step through the established method of heat shifting (Muthurajan *et al.*, 2004).

Yet even with heat shifting, nucleosomes are still confined to move within a local area that is defined energetically by steric hindrance and sequences unfavourable to nucleosome occupancy. What are observed are simply the most favourable nucleosome positions within this local area, and so it is possible that at low density the most preferable sequences for stable nucleosomes are under-represented. Whilst at higher nucleosome densities the background noise is increased, as nucleosomes are forced to occupy sub-optimal sequences by steric hindrance. Perhaps a more thorough investigation of nucleosome stability would be to repeat the genome-wide reconstitution experiment but then to stress nucleosomes, say with increasing levels of salt, and use Illumina sequencing to identify consesus sequences that retain nucleosomes with increasing stress. Such and experiment would provide a genome-wide analysis of nucleosome stability. It could also start with a higher level of nucleosome occupancy to ensure a significant number of reads are collected for the majority of nucleosome sites and yet would be resistant to steric hindrance noise as poorly positioned nucleosomes should be lost at relatively low stress levels.

Chapter 6

Discussion

Nucleosomes containing the histone variant CenH3 are essential for the function of all eukaryotic centromeres. Moreover, CenH3 nucleosomes clearly act as the epigenetic mark that defines the location of active regional centromeres (Barnhart *et al*., 2011). Thus it is important to understand how CenH3 nucleosomes are specifically incorporated at centromeres and how they differ from canonical nucleosomes. To address these questions, data presented in this thesis has detailed for the first time the *in vitro* assembly of CENP-A^{Cnp1} and canonical H3 nucleosomes using recombinant *S. pombe* histones, thereby enabling these nucleosomes to be studied using a range of biochemical and biophysical techniques. Using *in vitro* assembled CenH3 and H3 nucleosomes, from both *S. pombe* and humans, atomic force microscopy measurements revealed that some intrinsic feature of octameric CenH3 nucleosomes leads them to consistently register a lower height than octameric H3 nucleosomes. Reduced height was previously observed for CenH3 nucleosomes extracted from *D. melanogaster* and human cells and proposed to be indicative of a tetrameric structure for CenH3 nucleosomes. In addition, the sequencing of *S. pombe* CENP-A^{Cnp1} and H3 nucleosomes assembled on *S. pombe* genomic DNA has enabled a comparison of the inherent sequence preferences of these distinct nucleosome types. This

represents the first investigation of how DNA sequence might influence the genomic location of a non-canonical nucleosome.

CenH3 nucleosome component stoichiometry

The observation that centromeric CENP-A^{Cse4} nucleosomes in budding yeast appear to lack or contain low levels of H2A and H2B first led to the hypothesis that CenH3 nucleosomes may exhibit an atypical structure (Mizuguchi *et al*., 2007). It was suggested that a unique structure for CenH3 nucleosomes may distinguish active centromeres as the site for kinetochore establishment (Dalal *et al*., 2007). There have since arisen a number of models in which distinct compositions and stoichiometries of various subunits in CenH3 nucleosomes from different species have been proposed. The predominant models are centred around octameric and tetrameric versions of the nucleosome (Black and Cleveland, 2011). Defining which model is correct *in vivo* has proved difficult however, principally due to the low abundance of CenH3 nucleosomes and their relative instability (Camahort *et al*., 2009; Lando *et al*., 2012). The majority of data supporting an octameric CenH3 model has used nucleosomes reconstituted *in vitro*, typically from recombinantly produced histones, which have the advantages of sample purity and quantity that enable detailed biochemical and biophysical analysis of nucleosome component stoichiometries. However, the disadvantage of these assays is that *in vitro* assembled nucleosomes do not necessarily represent the true state of these nucleosomes *in vivo*. Particularly since the method used to produce nucleosomes *in vitro* was originally designed to purify octameric H3 nucleosomes (Luger *et al*., 1999). However, data supporting the tetrameric hemisome model has predominantly used

AFM to measure the heights of nucleosomes extracted from cells, but typically without cross-linking so it remains possible that in this extracted chromatin these nucleosomes also do not reflect their *in vivo* component stoichiometries (Camahort *et al*., 2009).

The observation that CenH3 nucleosomes appear lower in height than H3 nucleosomes extracted from cells using AFM represents the majority of data put forward to support the hemisome model. However, the conclusion that a lower height is indicative of a tetrameric CenH3 nucleosome structure is largely unsubstantiated and relies on the assumption that nucleosomes with a lower height have a different stoichiometry, such as a hemisome. There is no direct evidence to show that CenH3 nucleosomes are hemisomes. An alternative explanation is that CenH3 octameric nucleosomes have distinct biophysical properties resulting in them registering lower heights when measured using AFM. Thus one of the aims of the analyses presented here was to test whether biochemically defined octameric CenH3 nucleosomes produced *in vitro* appeared the same height as octameric H3 nucleosomes when measured using AFM. The first step towards this was the production of both recombinant H3 and CenH3 nucleosomes *in vitro* and the characterisation of their components and histone stoichiometries. Chapter 3 details the characterisation of what is thought to be the first *in vitro* assembly of *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes. Production of these recombinant CENP-A^{Cnp1} nucleosomes *in vitro* required a number of changes to the protocol most commonly used for the production of nucleosomes *in vitro* (Luger *et al*., 1999), namely the inclusion of a ~ 50 % molar excess of histones H2A and H2B relative to histones CENP-A^{Cnp1} and H4, and the

inclusion of a non-detergent sulfobetaine. These alterations to the standard method were found to significantly increase the solubility of CENP-A^{Cnp1} and H4 during refolding and therefore all future attempts to produce recombinant CENP-A^{Cnp1} nucleosomes *in vitro* should include these alterations. However, these alterations were found to be dispensable for the efficient production of recombinant *S. pombe* H3 nucleosomes. Lowering the salt concentration from 2 M to 0.5 M for the refolding of the CENP-A^{Cnp1} histone octamer further increased the solubility of CENP-A^{Cnp1} and H4. However, increasing the salt concentration to 2 M after refolding, to allow reconstitution of the histone octamer onto DNA caused the complex to precipitate. Thus it may be useful to further investigate the effect of salt concentration on CENP-A^{Cnp1} octamer refolding, perhaps using a salt titration within a concentration range that would allow for subsequent reconstitution into nucleosomes. Such a range is likely to be greater than 1.2 M salt, as this is approximately the salt concentration at which the H3 : H4 tetramer binds to DNA (Khrapunov *et al*., 1997).

A range of biochemical techniques were used to characterise the component stoichiometry of the recombinant nucleosomes assembled *in vitro*. Dynamic light scattering showed both the H3 and CENP-A^{Cnp1} nucleosome samples were mono-disperse, with equivalent hydrodynamic particle diameters of ~ 10 nm. This measurement is in accordance with the diameters of homologous human nucleosomes in respective H3 and CENP-A crystal structures, but also consistent with expected diameters of a hemisome (Tachiwana *et al*., 2011a; 2011b). Ideally, to confirm the octameric nature of these complexes they would be analysed by multi-angle light scattering or

analytical ultracentrifugation, both of which are capable of determining the absolute molar mass of complexes and so able to distinguish between octameric and tetrameric histone complexes. Additionally, to confirm their identity and purity a quantitative mass spectrometry analysis of the reconstituted nucleosomes would be required. Although the purity and identity of individual histones was confirmed by MALDI-TOF mass spectrometry, and reconstituted nucleosomes appeared to show equivalent quantities of each of the expected histones by SDS-PAGE.

Micrococcal nuclease (MNase) protection assays determined that the amount of DNA protected from nuclease digestion by CENP-A^{Cnp1} nucleosomes varied to a greater extent than for H3 nucleosomes but was on average 10 bp less than for H3 nucleosomes (~ 134 bp for CENP-A^{Cnp1} and ~ 144 bp for H3). A result that is consistent with both the conclusions drawn from the analysis of *in vitro* CENP-A nucleosomes (Tachiwana *et al*., 2011a) and *in vivo* CENP-A^{CID} nucleosomes (Dalal *et al*., 2007). The increased variation in the length of DNA protected by CENP-A^{Cnp1} nucleosomes compared to that protected by H3 nucleosomes likely represents a transient binding of these terminal 5 bp at the regions where DNA enters / exits from the nucleosome. The kinetics of this transient binding could potentially be measured using FRET between fluorophores attached to one end of the DNA and to the alpha-helix region of CenH3 / H3. If the DNA were stably bound then a close interaction between the DNA terminus and the alpha-helix region of CenH3 / H3 would provide a strong FRET signal, whilst if the DNA were transiently bound or unbound then the signal would be reduced

or absent. H3 nucleosomes could act as a positive control for stably bound DNA and should give a strong FRET signal.

Both the ~ 10 nm diameter and reduced MNase protection observed here for CENP-A^{Cnp1} nucleosomes assembled *in vitro* are consistent with both octameric and hemisome models of CenH3 nucleosomes. To help determine whether these *in vitro* assembled CENP-A^{Cnp1} (and H3) nucleosomes were tetrameric or octameric they were cross-linked with BS(PEG)$_5$ and subsequently analysed by SDS-PAGE. The migration of cross-linked H3 and CENP-A^{Cnp1} nucleosomes, and also of recombinant human H3 and CENP-A nucleosomes (in Chapter 4), through SDS-PAGE gels was consistent with that of an octameric complex. Moreover, coupling cross-linking to analysis by mass spectrometry revealed a number of cross-links between two copies of the same histone type in H3 nucleosomes and also for histone H4 in CENP-A^{Cnp1} nucleosomes. These data suggest both recombinant H3 and CENP-A^{Cnp1} nucleosomes are octameric when assembled *in vitro*, but it remains possible that this result could also represent the transient association of CENP-A^{Cnp1} hemisomes into octameric complexes. However, this alternative explanation seems unlikely since dynamic light scattering showed these samples to be monodisperse. Coupling cross-linking to analysis by mass spectrometry found fewer cross-links were observed for CENP-A^{Cnp1} nucleosomes than for H3 nucleosomes, making this analysis somewhat inconclusive, therefore it would be useful to trial a range of different cross-linking agents, or else to repeat the same cross-linking experiment again in order to collect data on a greater number of cross-links. However, the oxidisation of CENP-A^{Cnp1} nucleosomes did create a disulphide bond

between two CENP-A^{Cnp1} molecules. This is consistent with the proposed location of cysteine residues in the histones of an octameric nucleosome (Tachiwana *et al*., 2011a), and suggests that *in vitro* assembled CENP-A^{Cnp1} nucleosomes contain two copies of CENP-A^{Cnp1}. These cross-linking data are consistent with similar observations made elsewhere for recombinant CenH3 and H3 nucleosomes assembled *in vitro* from *S. cerevisiae* and human histones and observed *in vivo* for *D. melanogaster* CENP-A^{CID} nucleosomes (Camahort *et al*., 2009; Tachiwana *et al*., 2011a; Zhang *et al*., 2012). Taken together these analyses suggest that the CENP-A^{Cnp1} and H3 recombinant nucleosomes assembled here have octameric stoichiometries.

The main evidence for hemisomal, tetrameric CenH3 nucleosomes is based on the reduced height of CenH3 nucleosomes relative to H3 nucleosomes, as measured by AFM. This conclusion is based on the assumption that H3 and CenH3 nucleosomes respond equally to AFM height measurement assays and that reduced height equates to altered stoichiometry. In Chapter 4, AFM was used to measure the heights of octameric CENP-A^{Cnp1} and H3 nucleosomes assembled *in vitro*, with the conclusion that *in vitro* assembled recombinant octameric CENP-A^{Cnp1} nucleosomes from *S. pombe* also show a reduction in height relative to octameric H3 nucleosomes. Further, a similarly lower height was observed for octameric CENP-A nucleosomes assembled *in vitro* from recombinant human histones. This suggests that some conserved feature of *in vitro* assembled CenH3 nucleosomes cause them to appear lower in height than H3 nucleosomes using AFM. The differences in height observed here between H3 and CenH3 nucleosomes are similar to those observed for

CenH3 nucleosomes extracted from *D. melanogaster* and human cells (Bui *et al*., 2012; Dalal *et al*., 2007; Dimitriadis *et al*., 2010). Thus the analyses presented suggest that CenH3 nucleosomes extracted from *D. melanogaster* and human cells might also have octameric component stoichiometries, but register lower heights using AFM due to some intrinsic biophysical property conferred by CenH3.

Future challenges include the identification of the conserved property of CenH3 nucleosomes that causes them to appear to have reduced height relative to H3 nucleosomes. One interesting possibility is to test whether the relative unwrapping of DNA at the entry / exit sites of CenH3 nucleosomes compared with H3 nucleosomes can account for the apparent difference in height. AFM data from Bussiek *et al*, (2007) found a positive correlation between the height of chicken H3 nucleosomes assembled *in vitro* and the length of DNA each nucleosome wrapped. Data presented for *in vitro* assembled *S. pombe* nucleosomes in Chapter 3 of this study, and also for *Drosophila* (Dalal *et al*., 2007) and human (Conde e Silva *et al*., 2007; Tachiwana *et al*., 2011a) nucleosomes, have shown that CenH3 nucleosomes consistently wrap less DNA than H3 nucleosomes. Thus it is possible that CenH3 nucleosomes appear lower in height than H3 nucleosomes using AFM because they wrap less DNA. One prediction then is that H3 and CenH3 histone core particles that lack DNA should register similar heights using AFM. Histone core particles are unstable in the low salt conditions required for AFM imaging in air, but it should also be possible to use liquid AFM to compare the heights of CenH3 and H3 core particles in a buffer containing 2 M salt. Recently, the replacement of the arginine 49 in human

H3 with the lysine found at that location in CENP-A (H3 R49K) was shown to largely account for the difference in DNA wrapping at the entry / exit sites of the H3 and CENP-A nucleosomes (Conde e Silva *et al*., 2007; Panchenko *et al*., 2011). Thus, a second possible approach to determine whether DNA wrapping contributes to the reduced height observed for CenH3 nucleosomes relative to H3 nucleosomes is to use AFM to compare the heights of octameric nucleosomes containing CENP-A, CENP-A K52R, H3 and H3 R49K. If DNA wrapping influences the apparent height of nucleosomes in AFM measurements then the H3 R49K mutant would be expected to decrease the apparent height of H3 nucleosomes relative to the height of wildtype H3 nucleosomes. Reciprocally, the CENP-A K52R mutation might result in increased wrapping and, consequently, an increase in height.

The main conclusion of the analyses presented is that *in vitro* assembled CenH3 nucleosomes appear to have a reduced height relative to H3 nucleosomes using AFM. The similarity of these data from octameric nucleosomes with those previously collected for H3 and CenH3 nucleosomes extracted from *D. melanogaster* and human cells strongly implies that CenH3 nucleosomes might also be octameric *in vivo*.

An emerging model is that the stoichiometry of CenH3 nucleosomes is dynamic throughout the cell cycle (Bui *et al*., 2012; Shivaraju *et al*., 2012). In human cells, the height of CENP-A nucleosomes appears to increase in late G1, dropping back down again after S phase (Bui *et al*., 2012). Data collected here is consistent with the lower of these heights representing an octameric (*in vitro*) CENP-A nucleosome height, making it unclear whether this change

in height represents a real change in stoichiometry, seemingly suggesting a nucleosome with a histone stoichiometry above that of an octamer. One possibility is that this change is instead caused by cell cycle-dependent modifications to the nucleosome. A post-translational modification or binding partner, for example, might alter how the nucleosome interacts with the AFM tip, leading to an increased apparent height without necessitating an increase in histone stoichiometry. Such an explanation is also consistent with the observation that the same increase in height is apparent for H3 nucleosomes throughout the cell cycle.

If CenH3 nucleosomes have component stoichiometries equal to that of canonical H3 nucleosomes then other features of CenH3 nucleosomes must distinguish them from H3 nucleosomes, enabling CenH3 to specify centromere location. Some specific factors that bind CENP-A and ensure assembly at centromeres, and which promote kinetochore assembly once it is in place have now been identified (Black *et al*., 2004; 2007b; Guse *et al*., 2011).

Sequence preferences of recombinant *S. pombe* CENP-A^{Cnp1} and H3 nucleosomes

By tightly wrapping ~ 147 bp of DNA, nucleosomes effectively occlude wrapped sequences from other DNA binding factors. Occlusion of transcription factor binding sites by nucleosomes has been shown to influence gene expression (Richmond and Davey, 2003; Wolffe and Kurumizaka, 1998; Wyrick *et al*., 1999). The location of nucleosomes is controlled through the combined effects of both dynamic factors, such as the

local concentrations of transcription factors, remodelling complexes, other nucleosomes, genome methylation, histone variants and modifications, as well as static effectors such as the intrinsic affinities of nucleosomes, transcription factors and other DNA-binding proteins for DNA (Segal and Widom, 2009). Since nucleosome location is regulated by the combination of such a range of factors, it is surprising that nucleosome occupancy is so closely correlated with the base pair composition of the DNA, which is effectively a measure of the intrinsic nucleosome affinity for that sequence (Kaplan *et al*., 2010; Locke *et al*., 2010). For H3 nucleosomes, nucleosome occupancy correlates positively with the frequency of GC dinucleotides (Locke *et al*., 2010). *S. pombe* centromeric central domain regions contain ~ 72 % A/T compared with ~ 65 % across the whole genome, whilst *S. cerevisiae* centromeres contain ~ 86 % A/T, compared with ~ 61 % across the whole genome. Thus the increase in AT content found at centromeres might be expected to lead to a relative reduction in the density of H3 nucleosomes. In *Drosophila* and *Arabidopsis* it has been proposed that centromere-specific CenH3 nucleosomes have evolved to occupy centromeric sequences by adaptive evolution (Cooper and Henikoff, 2004). In the analyses presented in Chapter 3, MNase digestions of *S. pombe* nucleosomes assembled *in vitro* showed that CENP-A^{Cnp1} nucleosomes protect ~10 bp less DNA than H3 nucleosomes, a difference that might to result from an altered nucleosome stability. This led to the hypothesis that *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes differ in their stability on different sequences, and that this may lead to differences in the location of CenH3 versus H3 nucleosomes when assembled *in vitro* onto *S. pombe* genomic DNA.

The stability of recombinant CENP-A^{Cnp1} and H3 nucleosomes was measured for nucleosomes assembled *in vitro* onto three centromeric and three euchromatic DNA sequences from *S. pombe.* Whilst the data suggested that CENP-A^{Cnp1} nucleosomes were more stable on the centromeric sequences tested than on the euchromatic sequences, both CENP-A^{Cnp1} and H3 nucleosome stability varied significantly depending on the specific DNA sequence on which they were assembled. Thus, a genome-wide approach was used to compare the sequence preferences of H3 and CENP-A^{Cnp1} nucleosomes assembled *in vitro* on *S. pombe* genomic DNA.

The locations of *S. pombe* H3 and CENP-A^{Cnp1} nucleosomes assembled *in vitro* onto the *S. pombe* genome were sequenced and mapped. Contrary to the initial hypothesis, the positions of both H3 and CENP-A^{Cnp1} were correlated strongly across the genome. Moreover, the occupancy of both H3 and CENP-A^{Cnp1} nucleosomes on centromeric DNA *in vitro* matched with the occupancy of CENP-^{Cnp1} nucleosomes observed *in vivo*. This suggests that although centromere sequence alone is sufficient to define sites for nucleosome occupancy *in vivo*, it is insufficient to distinguish between H3 and CENP-A^{Cnp1} nucleosome occupancy at these sites. This observation contrasts with the initial measurements of nucleosome stability on specific sequences that suggested CENP-A^{Cnp1} nucleosomes were more stable on centromeric sequences than euchromatic sequences. A key assumption in these comparisons between *in vitro* and *in vivo* nucleosome location is that nucleosome location *in vitro* is determined solely by nucleosome sequence preference. This is because confounding factors that are inherent to *in vitro* reconstitutions, such as preferential sites for nucleosome reconstitution, are

not relevant for defining nucleosome location *in vivo* (Segal and Widom, 2009). In Chapter 5, analysis of nucleosome occupancy over the same six regions that were tested previously for nucleosome stability revealed that CENP-A^{Cnp1} nucleosome occupancy at the three centromeric sites was reduced relative to occupancy at three euchromatic sites tested. This contrasted with analysis of nucleosome stability on these same sites which suggested that CENP-A^{Cnp1} nucleosomes were more stable on these centromeric sequences than on the euchromatic sites. This difference between nucleosome occupancy and nucleosome stability appeared specific for CENP-A^{Cnp1} nucleosomes and was only apparent at the three centromeric sites tested. These analyses suggest that CENP-A^{Cnp1} nucleosome occupancy at these centromeric regions was not determined by nucleosome stability alone. It was proposed that sequences that preferentially associate with CENP-A^{Cnp1} during nucleosome assembly differ from the sequences that are preferential for the formation of stable CENP-A^{Cnp1} nucleosomes, with respect to these centromeric sequences. For H3 nucleosomes, the site of assembly appears to be determined by the tetramer of H3 : H4, whilst nucleosome stability is defined by that of the whole nucleosome (Thåström *et al*., 2004a). Thus, differences between preferential sequences for the assembly and stability of CENP-A^{Cnp1} nucleosomes are intriguing, yet preferential sequences for *in vitro* assembly are unlikely to have functional consequences *in vivo*, where nucleosomes are instead formed with the aid of assembly factors.

To further investigate the question of whether CENP-A^{Cnp1} and H3 nucleosomes exhibit differences in their occupancies as a result of differences

in their inherent nucleosome stabilities, a further genome-wide experiment could be performed. In this experiment nucleosomes assembled *in vitro* onto *S. pombe* genomic DNA at high densities would be stressed with increasing concentrations of salt or urea and subsequently sequenced and mapped to determine the relative change in nucleosome locations with increasing stress. Such analyses would allow comparison of the relative stabilities of H3 and CENP-A^{Cnp1} nucleosomes across the whole genome, while minimising the influence of inherent reconstitution preferences. Consequently it is anticipated that this would facilitate the comparison of *in vitro* nucleosome positions with those observed *in vivo*.

To our knowledge, these data represent the first example of a comparison between the sequence preferences of canonical and non-canonical nucleosomes. The observation that CENP-A^{Cnp1} and H3 nucleosomes occupy the same sequences when assembled *in vitro* suggest that DNA sequences within the *S. pombe* genome do not specify a preference for occupancy by either H3 or CENP-A^{Cnp1} nucleosomes. This implies that the specific localisation of CenH3 to the centromere, and the interspersed pattern of H3 and CenH3 nucleosomes observed at the centromeres of humans and *Drosophila* (Blower *et al*., 2002), are not directed by an inherent preference of CenH3 nucleosomes for a particular DNA sequence or feature. This observation is consistent with the observation that tethering of CENP-A^{CID} to lac operator arrays is sufficient to recruit the stable ectopic assembly of untethered CENP-A^{CID} and the conclusion that regional centromere location is epigenetically defined (Mendiburo *et al*., 2011). If DNA sequence alone is not sufficient to distinguish between the positioning of H3 and CENP-A^{Cnp1}

nucleosomes then specific protein factors may establish the interspersed pattern of H3 and CenH3 nucleosomes observed at the centromeres of humans and *Drosophila* (Blower *et al*., 2002). Candidates for such factors are HJURP (Scm3 in yeast) the CenH3-specific chromatin assembly factor that is also capable of targeting hybrid CENP-A/H3 histones to centromeres (Dunleavy *et al*., 2009; Foltz *et al*., 2009; Sanchez-Pulido *et al*., 2009). Whilst the epigenetic plasticity of centromeres means that no specific DNA sequence motif is required for HJURP targeting, it remains possible that HJURP preferentially targets to sites with features associated with centromeres, such as enriched A/T content and/or a unique chromatin environment (Choo, 2001; Sullivan and Karpen, 2004). Moreover, it is possible that other factors such as transcription and heterochromatin act to distance canonical H3 nucleosomes and promote CenH3 assembly. It is also conceivable that histone post-translational modifications affect assembly factors such as HJURP. The genome-wide assembly assay used here could similarly be used to test whether other factors such as HJURP preferentially assemble CENP-A^{Cnp1} nucleosomes at DNA sequences with particular features. The assay could even be altered such that nucleosome repositioning on the DNA is disfavoured by exclusion of the heat-shifting step. More generally, the effect of nucleosome remodelling factors could be tested, similar to previous assays for H3 nucleosomes (Zhang *et al*., 2011), by incubating assembled nucleosomes with either cell extracts (± ATP) or with individual remodelling factors to determine whether there are CenH3-specific remodelling factors.

In conclusion, the data presented provide an explanation for the distinct difference in height observed between CenH3 and H3 nucleosomes and suggest that CenH3 nucleosomes are octameric. Genome-wide analyses of CENP-A^{Cnp1} and H3 nucleosomal sequences indicates that primary DNA sequence has little influence on the selection of CenH3 nucleosomes by centromeric DNA and instead suggest that assembly or remodelling processes establish the patterns of H3 and CenH3 nucleosomes observed *in vivo*.

Appendix 1

Cross-linked peptides identified by mass spectrometry

Validated peptides from cross-linked *in vitro* H3 nucleosome sample							
Protein 1	**Peptide 1**	**Start Peptide**	**Protein 1 link**	**Protein 2**	**Peptide 2**	**Start Peptide**	**Protein 2 link**
H2A	KGNYAQR	37	0	H2A	KGNYAQR	37	0
H2A	SAKAGLAFPVGR	19	2	H2A	MSGGKSGGKAAVAKSAQSRSAK	0	19
H2A	SAKAGLAFPVGR	19	2	H2B	SVTKYSSSAQ	116	7
H2A	SAKAGLAFPVGR	19	2	H3	KLPFQR	64	0
H2A	SAKAGLAFPVGR	19	2	H3	TKQTAR	3	0
H2A	SAQSRSAKAGLAFPVGR	14	7	Cnp1	SLMAEPGDPIPRPR	4	0
H2A	LLRKGNYAQR	34	3	Cnp1	VTIMQRDMQLARR	103	1
H2B	SAAEKKPASKAPAGKAPR	1	14	H2B	APRDTMoxKSADK	16	7
H2B	LILPGELAKHAVTEGTK	99	15	H2B	SVTKYSSSAQ	116	7
H2B	SVTKYSSSAQ	116	3	H2B	LILPGELAKHAVTEGTK	99	8
H2B	LILPGELAKHAVTEGTK	99	15	H3	TKQTAR	3	0
H4	KTVTSLDVVYSLKR	79	12	H2A	KGNYAQR	37	0
H4	DNIQGITKPAIR	24	7	H2A	SAKAGLAFPVGR	19	2
H4	TVTSLDVVYSLKR	80	11	H2B	LAAYNKK	79	5
H4	KTVTSLDVVYSLKR	79	12	H4	DAVTYTEHAKR	68	9
H4	KTVTSLDVVYSLKR	79	1	H4	KILR	20	0
H4	KTVTSLDVVYSLK	79	1	H4	KILR	20	0
H4	TVTSLDVVYSLK	80	3	H4	DAVTYTEHAKRK	68	9
H4	KTVTSLDVVYSLKR	79	12	H4	KILR	20	0
H4	KTVTSLDVVYSLKR	79	1	H4	KILR	20	0
H4	KTVTSLDVVYSLKR	79	1	H4	DAVTYTEHAKR	68	9
H4	TVTSLDVVYSLKR	80	0	H4	DAVTYTEHAKRK	68	9
H4	RKTVTSLDVVYSLKR	78	5	H4	DAVTYTEHAK	68	5
H4	KTVTSLDVVYSLKR	79	1	H4	DAVTYTEHAKR	68	9
H4	KTVTSLDVVYSLKR	79	1	H4	DAVTYTEHAKR	68	9
H4	KTVTSLDVVYSLK	79	10	H4	DAVTYTEHAKR	68	9
H4	DNIQGITKPAIR	24	7	H4	ISALVYEETR	46	1
H4	TVTSLDVVYSLKR	80	11	H4	DAVTYTEHAKRK	68	9
H4	KTVTSLDVVYSLKR	79	12	H3	KLPFQR	64	0
H4	KTVTSLDVVYSLK	79	10	H3	KLPFQR	64	0
H4	DAVTYTEHAKR	68	9	H3	EIAQDFKTDLR	73	7
H4	KTVTSLDVVYSLKR	79	12	H3	EIAQDFKTDLR	73	7
H4	KTVTSLDVVYSLKR	79	4	H3	KLPFQR	64	0
H4	DNIQGITKPAIR	24	7	H3	KLPFQR	64	0
H4	KTVTSLDVVYSLKR	79	1	H3	EIAQDFKTDLR	73	6
H4	DNIQGITKPAIR	24	7	H3	VTIQPKDMQLAR	117	5
H4	KTVTSLDVVYSLKR	79	4	H3	KLPFQR	64	0
H4	DNIQGITKPAIR	24	7	H3	KLPFQR	64	0
H4	KTVTSLDVVYSLK	79	10	H3	KLPFQR	64	0
H4	DNIQGITKPAIR	24	7	H3	VTIQPKDMoxQLAR	117	5
H4	DNIQGITKPAIR	24	7	H3	VTIQPKDMoxQLAR	117	5
H4	KTVTSLDVVYSLKR	79	12	H3	EIAQDFKTDLR	73	7
H4	TVTSLDVVYSLK8	80	9	Cnp1	STDLLIQR10	37	1
H3	EIAQDFKTDLR	73	7	H2B	LILPGELAKHAVTEGTK	99	15
H3	EIAQDFKTDLR	73	7	H4	DAVTYTEHAKR	68	9
H3	EIAQDFKTDLR	73	7	H4	KTVTSLDVVYSLKR	79	12
H3	VTIQPKDMQLAR	117	5	H4	ISALVYEETR	46	5
H3	EIAQDFKTDLR	73	7	H4	KTVTSLDVVYSLKR	79	12
H3	EIAQDFKTDLR	73	6	H4	DAVTYTEHAKR	68	9
H3	EIAQDFKTDLR	73	7	H4	RKTVTSLDVVYSLK	78	2
H3	VTIQPKDMQLAR	117	5	H4	ISALVYEETR	46	1
H3	EIAQDFKTDLR	73	7	H4	DAVTYTEHAKR	68	9
H3	EIAQDFKTDLR	73	7	H4	KTVTSLDVVYSLKR	79	12
H3	EIAQDFKTDLR	73	7	H4	KTVTSLDVVYSLKR	79	12
H3	EIAQDFKTDLR	73	6	H4	RKTVTSLDVVYSLKR	78	13
H3	EIAQDFKTDLR	73	7	H4	DAVTYTEHAKR	68	9
H3	EIAQDFKTDLR	73	7	H4	DAVTYTEHAKR	68	9
H3	EIAQDFKTDLR	73	6	H4	DAVTYTEHAKR	68	9
H3	EIAQDFKTDLR	73	7	H3	KLPFQR	64	0
H3	VTIQPKDMQLAR	117	5	H3	VTIQPKDMQLAR	117	5
H3	EIAQDFKTDLR	73	6	H3	KLPFQR	64	0
H3	TKQTAR	3	1	H3	KSTGGK	9	1
H3	KSTGGK	9	2	H3	TKQTAR	3	1
H3	YQKSTELLIR	54	4	H3	KLPFQR	64	0
H3	VTIQPKDMQLAR	117	5	H3	KLPFQR	64	0
H3	EIAQDFKTDLR	73	7	H3	TKQTAR	3	3
H3	QLASKAAR	19	4	H3	QTARKSTGGKAPR	5	9
H3	VTIQPKDMoxQLAR	117	5	H3	VTIQPKDMoxQLAR	117	5
H3	STELLIR	57	1	Cnp1	KYQR	33	0
H3	YQKSTELLIR	54	2	Cnp1	RVTIMQR	102	2
H3	STELLIR	57	1	Cnp1	KYQR	33	1

Validated peptides from cross-linked *in vitro* CENP-A^{Cnp1} nucleosome sample							
Protein 1	**Peptide 1**	**Start Peptide**	**Protein 1 link**	**Protein 2**	**Peptide 2**	**Start Peptide**	**Protein 2 link**
H2A	KGNYAQR	37	0	H2B	GK8NR10	29	1
H2A	KGNYAQR	37	0	H2B	LAAYNK8K8	79	5
H2A	KGNYAQR	37	0	H2B	LAAYNK8K8	79	5
H2A	KGNYAQR	37	3	H2B	LAAYNK8K8	79	5
H2A	KGNYAQR	37	0	H2B	K8STISSR10	85	0
H2A	KGNYAQR	37	0	H2B	LAAYNK8K8	79	5
H2A	KGNYAQR	37	0	H2B	LAAYNK8K8	79	5
H2A	KGNYAQR	37	0	H2B	GK8NR10	29	1
H2A	KGNYAQR	37	0	H2B	GK8NR10	29	1
H2A	SAKAGLAFPVGR	19	2	H2B	SVTK8YSSSAQ	116	7
H2A	KGNYAQR	37	0	H2B	GK8NR10	29	1
H2A	KGNYAQR	37	0	H2B	GK8NR10	29	1
H2A	KGNYAQR	37	0	H2B	LAAYNK8K8	79	5
H2A	SAKAGLAFPVGR	19	2	H2B	LILPGELAK8HAVTEGTK8	99	15
H2A	KGNYAQR	37	0	H4	DAVTYTEHAK8R10	68	9
H2A	KGNYAQR	37	0	H3	TK8QTAR10	3	1
H2A	SAKAGLAFPVGR	19	2	H3	EIAQDFK8TDLR10	73	7
H2B	LAAYNK8K8	79	5	H2A	KGNYAQR	37	0
H2B	IATEASK8LAAYNK8K8	72	12	H2A	KGNYAQR	37	0
H2B	IATEASK8LAAYNK8	72	6	H2A	KGNYAQR	37	3
H2B	SVTK8YSSSAQ	116	7	H2A	SAKAGLAFPVGR	19	2
H2B	LILPGELAKHAVTEGTK	99	8	H4	DAVTYTEHAK8R10	68	9
H2B	KSTISSR	85	0	H4	LFLENVIR10DAVTYTEHAK8	60	13
H2B	KSTISSR	85	0	H4	LFLENVIR10DAVTYTEHAK8	60	13
H2B	KSTISSR	85	5	H4	DAVTYTEHAK8R10	68	9
H3	EIAQDFK8TDLR10	73	7	H4	KILR	20	0
H3	EIAQDFK8TDLR10	73	6	H4	KILR	20	0
H3	EIAQDFK8TDLR10	73	7	H4	DAVTYTEHAKR	68	9
H4	DAVTYTEHAK8R10	68	9	H2A	KGNYAQR	37	0
H4	TVTSLDVVYSLK8R10	80	11	H2A	KGNYAQR	37	3
H4	DAVTYTEHAKR	68	9	H2B	K8STISSR10	85	0
H4	DAVTYTEHAK8R10K8	68	9	H2B	KSTISSR	85	0
H4	K8TVTSLDVVYSLK8R10	79	1	H2B	KSTISSR	85	0
H4	DAVTYTEHAKR	68	9	H2B	K8STISSR10	85	0
H4	DAVTYTEHAKRK	68	9	H2B	IATEASK8LAAYNK8	72	10
H4	DAVTYTEHAK	68	5	H4	DAVTYTEHAK8R10K8TVTSLDVVYSLK8	68	21
H4	TVTSLDVVYSLKR	80	3	H4	DAVTYTEHAK8R10K8	68	9
H4	RKTVTSLDVVYSLK	78	2	H3	EIAQDFK8TDLR10	73	7
H4	TVTSLDVVYSLKR	80	0	H3	TK8QTAR10	3	3
H4	DNIQGITK8PAIR10	24	7	H3	VTIQPKDMoxQLAR	117	5
H4	DNIQGITK8PAIR10	24	7	H3	VTIQPKDMoxQLAR	117	5
H4	KILR	20	0	H3	TK8QTAR10	3	1
H4	DAVTYTEHAK8R10K8	68	9	H3	QTARKSTGGKAPR	5	9
Cnp1	EISSEFVANFSTDVGLR	53	11	H4	K8ILR10	20	0
Cnp1	KSLMAEPGDPIPRPR	3	0	Cnp1	SLMAEPGDPIPR10PR10K8K8	4	14

Table A1-1 - Cross-linked peptides identified by mass spectrometry between a heavy-labelled and an unlabelled histone

H3 and CENP-A^{Cnp1} nucleosomes assembled from an equal mixture of heavy-labelled (labelled with N15 and C13 Arginine and Lysine) or unlabelled versions of the appropriate histone. Nucleosomes were fixed with the

primary amine cross-linker $BS(PEG)_5$ and subject to analysis by mass spectrometry. Peptides determined by manual validation to contain cross-links between one heavy-labelled histone and one unlabelled histone are shown here for the H3 nucleosome sample (above) and the CENP-A^{Cnp1} nucleosome sample (below). Peptides containing cross-links between one heavy- and one unlabelled version of the same histone are highlighted.

Appendix 2

AFM data without DNA height levelling

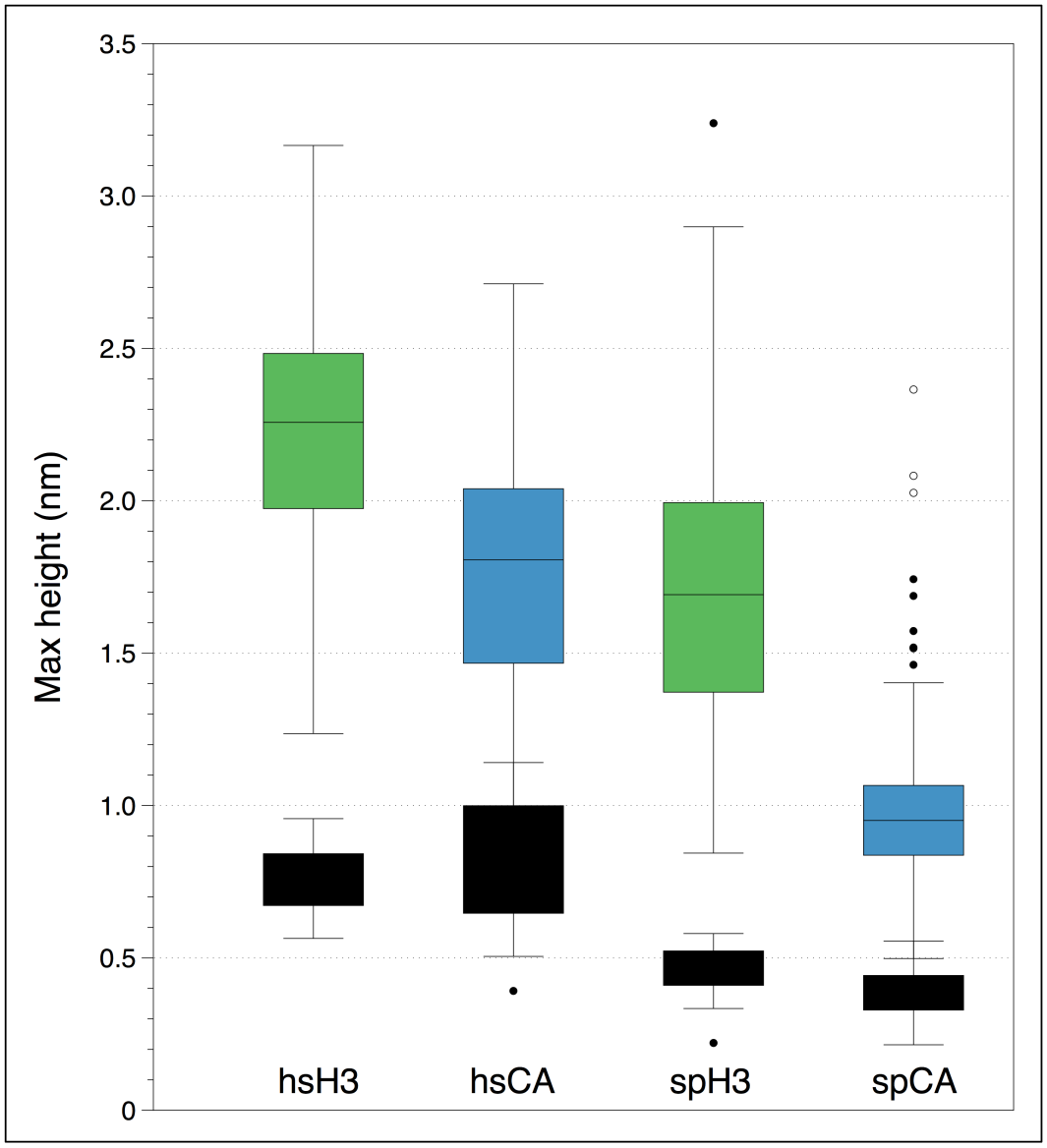


Figure A2-1 - AFM data without DNA height levelling

The height of H3 (green) and CenH3 (blue) nucleosomes assembled *in vitro* from human (hs) and *S. pombe* (sp) recombinant histones measured using

AFM as in Chapter 4 but without the DNA levelling. The results are essentially identical to those shown for *S. pombe in vitro* assembled nucleosomes in (Figure 4-8) and for human *in vitro* assembled nucleosomes in Figure 4-11. For each sample the distribution of DNA heights is also represented by the black boxes. The central line within each box plot marks the median, the outer edges of each box represents the first and third interquartile ranges and the whiskers mark the range. Statistical outliers are shown as single dots.

Appendix 3

Sequences and locations of "EUNS" and "CCNS" loci

Six different 200 bp sequences were chosen from these existing datasets where well defined H3 and CENP-A^{Cnp1} nucleosome positioned peaks were evident (sequences identified by H. Berger, Allshire research group) (Lando *et al*., 2012). Three H3 peaks were chosen as representative euchromatic nucleosome sequences from the genome and three CENP-A^{Cnp1} peaks from the central domain regions were chosen as representative centromeric sequences. These representative euchromatic and centromeric nucleosome sequences were labelled here as EUNS 1-3 and CCNS 1-3, respectively. The genomic location and nucleotide sequence of each is presented below.

Name: "EUNS 1"

Approximate location: *ckil*

Chromosome: 2

Location: 4072599 - 4072798

GTAGATCATATTAGCATTTTTACTGTTTGGTCGACCGATCAAAAAATT
ATCAGGCTTAATGTCTCGATAAACGAGACTCTTCTCATGGATAGATTG
CACTCTTGCCAACATCTGTTTTGCGGCCATAGCAACTGTTTTTACACTA

AATTTACGGCCGCAGAGATCGAGCAAATCTTCTAAACTTGGCCCAAG
AAGGTCAA

Name: "EUNS 2"

Approximate location: *tif211*

Chromosome: 1

Location: 3166836 - 3167035

TTGTTAGGAATGTACTAACCAGGATTAAGGATACATCGATCTTTCAAA
GCGCCGTGTTAGCCCTGAGGATGTTGTAAAGTGCGAAGAACGCTTCA
ACAAATCTAAGGCTGTTCATTCCATTATGCGCCATATTGCTGAGAAGC
ATAATGTTCCTTTGGAGACCATGTATACCACAATTGGATGGCCATTAT
ACCGCAAAT

Name: "EUNS 3"

Approximate location: SPBC24C6.03

Chromosome: 2

Location: 2323946 - 2324145

CTGCAGTCGCTCATGTAACCAAAGATGCAAAGGGTTTAGTGTGGCCA
TCGTCGATAGCTCCTTGGAAAGTACTAGTTGTTCCTACTTCTGATAAT
CATATTCAGAGCGCTGAAACGGTTTACGATGCCACGGCTAACGTTGT
AGGCTTTGATAATGTGCTGTTGGAGGACAGACAAAATAGAGCTTTTG
GATATAAAATG

Name: "CCNS 1"

Approximate location: Central domain of centromere 1

Chromosome: 1

Location: 3769282 - 3769481

ACTATATTGCAAGTCAAAGTTGGTATTAGTTCATCGTCCTCCGTTTGG
CAGGAAGGCCGTTTCTTTATGGTTAACATATCCGTAGTTTATAACTAA
ATAAGACCAGTTAAGCGGTATTATCACGTAACTTTTACTGTATTTATT
GAATATAGCTTTCAAAAATGTGGATGACTGATTGTCCGATAACCCTG
CATAAAATG

Name: "CCNS 2"

Approximate location: Central domain of centromere 2

Chromosome: 2

Location: 1624909 - 1625108

AAGTTCGAAAAATCGTGCACATTTGTGAAAAGGTTTAGCTCTACCTAT
CGCAAATAAATGCTGCATCAAACGCAATTGCTGCTATGTGGGTCTGT
ATTGCCTCTCCCTTGCCAGTAATGTGTATTTCATCTTGTTTATTTGTTTT
CTTATCGAAATTAACTTAAGCTTTGTCCAAATGCTAATTATTAACCAC
TTTGGGT

Name: "CCNS 3"

Approximate location: Inner most repeat region of centromere 2

Chromosome: 2

Location: 1628304 - 1628503

GAAAGTTTAAGTTCATTACTTAAAAAATAACAAAACGTTGTTCACTG
AAACCTCATGTGGTTTTAAAGTACCTATCTAAAAGCAGTTTCATAAAC
GATCTGGTACTACCTAATTGTATTATTGTTTCGCCCTAAAAGTAAACG

GTAAGCACTTTGCAACGATTACCGGTTTATAAAAATTATTCCATGGCA
TGATTAATT

References

Agudo, M., Abad, J.P., Molina, I., Losada, A., Ripoll, P., and Villasante, A. (2000). A dicentric chromosome of Drosophila melanogaster showing alternate centromere inactivation. Chromosoma *109*, 190–196.

Ahmad, K., and Henikoff, S. (2002). Histone H3 variants specify modes of chromatin assembly. Pnas *99 Suppl 4*, 16477–16484.

Alexeev, D.G., Lipanov, A.A., and Skuratovskii IYa (1987). Poly(dA).poly(dT) is a B-type double helix with a distinctively narrow minor groove. Nature *325*, 821–823.

Allen, M.J., Dong, X.F., O'Neill, T.E., Yau, P., Kowalczykowski, S.C., Gatewood, J., Balhorn, R., and Bradbury, E.M. (1993). Atomic force microscope measurements of nucleosome cores assembled along defined DNA sequences. Biochemistry *32*, 8390–8396.

Allshire, R.C., Nimmo, E.R., Ekwall, K., Javerzat, J.P., and Cranston, G. (1995). Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev *9*, 218–233.

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

Amor, D.J., Bentley, K., Ryan, J., Perry, J., Wong, L., Slater, H., and Choo, K.H.A. (2004). Human centromere repositioning "in progress." Pnas *101*, 6542–6547.

Anderson, J.D., and Widom, J. (2001). Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Mol Cell Biol *21*, 3830–3839.

Bancaud, A., Wagner, G., Conde e Silva, N., Lavelle, C., Wong, H., Mozziconacci, J., Barbi, M., Sivolob, A., Le Cam, E., Mouawad, L., *et al*. (2007). Nucleosome chiral transition under positive torsional stress in single chromatin fibers. Mol Cell *27*, 135–147.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Bao, Y., Konesky, K., Park, Y.-J., Rosu, S., Dyer, P.N., Rangasamy, D., Tremethick, D.J., Laybourn, P.J., and Luger, K. (2004). Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. Embo J *23*, 3314–3324.

Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black,

B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol *194*, 229–243.

Black, B., and Bassett, E.A. (2008). The histone variant CENP-A and centromere specification. Current Opinion in Cell Biology *20*, 91–100.

Black, B.E., and Cleveland, D.W. (2011). Epigenetic centromere propagation and the nature of CENP-a nucleosomes. Cell *144*, 471–479.

Black, B., Brock, M.A., Bédard, S., Woods, V.L., and Cleveland, D.W. (2007a). An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. Pnas *104*, 5008–5013.

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

Black, B., Jansen, L.E.T., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V., and Cleveland, D.W. (2007b). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell *25*, 309–322.

Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved organization of centromeric chromatin in flies and humans. Dev Cell *2*, 319–330.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612–616.

Bowman, A., Ward, R., El-Mkami, H., Owen-Hughes, T., and Norman, D.G. (2010). Probing the (H3-H4)2 histone tetramer structure using pulsed EPR spectroscopy combined with site-directed spin labelling. Nucleic Acids Res *38*, 695–707.

Bowman, A., Ward, R., Wiechens, N., Singh, V., El-Mkami, H., Norman, D.G., and Owen-Hughes, T. (2011). The histone chaperones Nap1 and Vps75 bind histones H3 and H4 in a tetrameric conformation. Mol Cell *41*, 398–408.

Bui, M., Dimitriadis, E.K., Hoischen, C., An, E., Quénet, D., Giebe, S., Nita-Lazar, A., Diekmann, S., and Dalal, Y. (2012). Cell-Cycle-Dependent Structural Transitions in the Human CENP-A Nucleosome *In vivo*. Cell *150*, 317–326.

Bussiek, M., Müller, G., Waldeck, W., Diekmann, S., and Langowski, J. (2007). Organisation of nucleosomal arrays reconstituted with repetitive African green monkey α-satellite DNA as analysed by atomic force microscopy. Eur Biophys J *37*, 81–93.

Butt, H.J. (1991). Electrostatic interaction in atomic force microscopy. Biophysj *60*, 777–785.

Camahort, R., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., and Gerton,

J.L. (2007). Scm3 is essential to recruit the histone h3 variant cse4 to centromeres and to maintain a functional kinetochore. Mol Cell *26*, 853–865.

Camahort, R., Shivaraju, M., Mattingly, M., Li, B., Nakanishi, S., Zhu, D., Shilatifard, A., Workman, J.L., and Gerton, J.L. (2009). Cse4 is part of an octameric nucleosome in budding yeast. Mol Cell *35*, 794–805.

Camerini-Otero, R.D., and Felsenfeld, G. (1977). Histone H3 disulfide dimers and nucleosome structure. Pnas *74*, 5519–5523.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Carone, D.M., Longo, M.S., Ferreri, G.C., Hall, L., Harris, M., Shook, N., Bulazel, K.V., Carone, B.R., Obergfell, C., O'Neill, M.J., *et al*. (2009). A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma *118*, 113–125.

Carroll, C.W., Silva, M.C.C., Godek, K.M., Jansen, L.E.T., and Straight, A.F. (2009). Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat Cell Biol *11*, 896–902.

Chadwick, B.P., and Willard, H.F. (2001). 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. (2002). Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. J Cell Biol *157*, 1113–1123.

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.

Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.-C., Nilges, M., *et al*. (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. Embo J *29*, 717–726.

Cho, U.-S., and Harrison, S.C. (2011). Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. Pnas *108*, 9367–9371.

Choi, E.S., Strålfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. J Biol Chem *286*, 23600–23607.

Choo, K.H. (2001). Domain organization at the centromere and neocentromere. Dev Cell *1*, 165–177.

Church, G.M., Gao, Y., and Kosuri, S. (2012). Next-generation digital information storage in DNA. Science *337*, 1628.

Coffman, V.C., Wu, P., Parthun, M.R., and Wu, J.-Q. (2011). CENP-A exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast. J Cell Biol *195*, 563–572.

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

Cooper, J.L., and Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol *21*, 1712–1718.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367–1372.

D'Arcy, S., and Luger, K. (2011). Understanding histone acetyltransferase Rtt109 structure and function: how many chaperones does it take? Curr Opin Struct Biol *21*, 728–734.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007). Tetrameric Structure of Centromeric Nucleosomes in Interphase Drosophila Cells. PLoS Biol *5*, e218.

Davey, C.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.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni-Pettinotti, G. (2007). Histone chaperones: an escort network regulating histone traffic. Nat Struct Mol Biol *14*, 997–1007.

de Murcia, G., Huletsky, A., Lamarre, D., Gaudreau, A., Pouyet, J., Daune, M., and Poirier, G.G. (1986). Modulation of chromatin superstructure induced by poly(ADP-ribose) synthesis and degradation. J Biol Chem *261*, 7011–7017.

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

Dimitriadis, E.K., Weber, C., Gill, R.K., Diekmann, S., and Dalal, Y. (2010). Tetrameric organization of vertebrate centromeric nucleosomes. Pnas *107*, 20317–20322.

Ding, R., McDonald, K.L., and McIntosh, J.R. (1993). Three-dimensional reconstruction and analysis of mitotic spindles from the yeast, Schizosaccharomyces pombe. J Cell Biol *120*, 141–151.

Dirksen, E.H.C., Pinkse, M.W.H., Rijkers, D.T.S., Cloos, J., Liskamp, R.M.J.,

Slijper, M., and Heck, A.J.R. (2006). Investigating the dynamic nature of the interactions between nuclear proteins and histones upon DNA damage using an immobilized peptide chemical proteomics approach. J. Proteome Res. *5*, 2380–2388.

Donham, D.C., Scorgie, J.K., and Churchill, M.E.A. (2011). The activity of the histone chaperone yeast Asf1 in the assembly and disassembly of histone H3/H4-DNA complexes. Nucleic Acids Res *39*, 5449–5458.

Dunleavy, E.M., Pidoux, A.L., Monet, M., Bonilla, C., Richardson, W., Hamilton, G.L., Ekwall, K., McLaughlin, P.J., and Allshire, R.C. (2007). A NASP (N1/N2)-related protein, Sim3, binds CENP-A and is required for its deposition at fission yeast Centromeres. Mol Cell *28*, 1029–1044.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

Earnshaw, W.C., and Migeon, B.R. (1985). Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. Chromosoma *92*, 290–296.

Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma *91*, 313–321.

Elgin, S.C. (1996). Heterochromatin and gene regulation in Drosophila. Curr Opin Genet Dev *6*, 193–202.

Ercan, S., Lubling, Y., Segal, E., and Lieb, J.D. (2011). High nucleosome occupancy is encoded at X-linked gene promoters in C. elegans. Genome Res *21*, 237–244.

Ferreira, H., Somers, J., Webster, R., Flaus, A., and Owen-Hughes, T. (2007). Histone tails and the H3 alphaN helix regulate nucleosome mobility and stability. Mol Cell Biol *27*, 4037–4048.

Field, Y., Fondufe-Mittendorf, Y., Moore, I.K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J.D., Widom, J., and Segal, E. (2009). Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nat Genet *41*, 438–445.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comp Biol *4*, e1000216.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., *et al*. (2010). Systematic Protein Location Mapping Reveals Five Principal

Chromatin Types in Drosophila Cells. Cell *143*, 212–224.

Fischer, T., Cui, B., Dhakshnamoorthy, J., Zhou, M., Rubin, C., Zofall, M., Veenstra, T.D., and Grewal, S.I.S. (2009). Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. Pnas *106*, 8998–9003.

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell *137*, 472–484.

Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., and Cleveland, D.W. (2006). The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol *8*, 458–469.

Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cell *12*, 17–30.

Furuyama, S., and Biggins, S. (2007). Centromere identity is specified by a single centromeric nucleosome in budding yeast. Pnas *104*, 14706–14711.

Furuyama, T., and Henikoff, S. (2009). Centromeric nucleosomes induce positive DNA supercoils. Cell *138*, 104–113.

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D.P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep.

Gassmann, R., Rechtsteiner, A., Yuen, K.W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Monen, J., *et al*. (2012). An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. Nature *484*, 534–537.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A Role for Snf2-Related Nucleosome-Spacing Enzymes in Genome-Wide Nucleosome Organization. Science *333*, 1758–1760.

Gottesfeld, J.M., and Luger, K. (2001). Energetics and affinity of the histone octamer for defined DNA sequences. Biochemistry *40*, 10927–10933.

Gould, H.J., Cowling, G.J., Harborne, N.R., and Allan, J. (1980). An examination of models for chromatin transcription. Nucleic Acids Res *8*, 5255–5266.

Gracey, L.E., Chen, Z.-Y., Maniar, J.M., Valouev, A., Sidow, A., Kay, M.A., and Fire, A.Z. (2010). An *in vitro*-identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome *in vivo*. Epigenetics Chromatin *3*, 13.

Guse, A., Carroll, C.W., Ben Moree, Fuller, C.J., and Straight, A.F. (2011).

In vitro centromere and kinetochore assembly on defined chromatin templates. Nature 1–7.

Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J Biol Chem *281*, 559–568.

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell *118*, 715–729.

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., and Spector, D.L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell *107*, 727–738.

Henikoff, S., Ahmad, K., Platero, J.S., and van Steensel, B. (2000). Heterochromatic deposition of centromeric histone H3-like proteins. Pnas *97*, 716–721.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet *9*, 15–26.

Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science *293*, 1098–1102.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell *10*, 303–315.

Hewawasam, G., Shivaraju, M., Mattingly, M., Venkatesh, S., Martin-Brown, S., Florens, L., Workman, J.L., and Gerton, J.L. (2010). Psh1 is an E3 ubiquitin ligase that targets the centromeric histone variant Cse4. Mol Cell *40*, 444–454.

Hirota, T., Lipp, J.J., Toh, B.-H., and Peters, J.-M. (2005). Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature *438*, 1176–1180.

Ioshikhes, I.P., Albert, I., Zanton, S.J., and Pugh, B.F. (2006). Nucleosome positions predicted through comparative genomics. Nat Genet *38*, 1210–1215.

Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y., and Takahashi, K. (2008). Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science *321*, 1088–1091.

Jansen, L.E.T., Black, B., Foltz, D.R., and Cleveland, D.W. (2007). Propagation of centromeric chromatin requires exit from mitosis. J Cell Biol *176*, 795–805.

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

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519–1529.

Joglekar, A.P., Bouck, D.C., Finley, K., Liu, X., Wan, Y., Berman, J., He, X., Salmon, E.D., and Bloom, K.S. (2008). Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres. J Cell Biol *181*, 587–594.

Johnson, L., Mollah, S., Garcia, B.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Jacobsen, S.E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res *32*, 6511–6518.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev *19*, 295–310.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Leproust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., *et al*. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010). Nucleosome sequence preferences influence *in vivo* nucleosome organization. Nat Struct Mol Biol *17*, 918–920.

Ketel, C., Wang, H.S.W., Mcclellan, M., Bouchonville, K., Selmecki, A., Lahav, T., Gerami-Nejad, M., and Berman, J. (2009). Neocentromeres form efficiently at multiple possible loci in Candida albicans. PLoS Genet *5*, e1000400.

Khrapunov, S.N., Dragan, A.I., Sivolob, A.V., and Zagariya, A.M. (1997). Mechanisms of stabilizing nucleosome structure. Study of dissociation of histone octamer from DNA. Bba *1351*, 213–222.

Kim, S.-M., Dubey, D.D., and Huberman, J.A. (2003). Early-replicating heterochromatin. Genes Dev *17*, 330–335.

Koch, J. (2000). Neocentromeres and alpha satellite: a proposed structural code for functional human centromere DNA. Hum Mol Genet *9*, 149–154.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lagana, A., Dorn, J.F., De Rop, V., Ladouceur, A.-M., Maddox, A.S., and Maddox, P.S. (2010). A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. Nat Cell Biol *12*, 1186–1193.

Lando, D., Endesfelder, U., Berger, H., Subramanian, L., Dunne, P.D., McColl, J., Klenerman, D., Carr, A.M., Sauer, M., Allshire, R.C., *et al*. (2012). Quantitative single-molecule microscopy reveals that CENP-A(Cnp1)

deposition occurs during G2 in fission yeast. Open Biol *2*, 120078.

Lavelle, C., Recouvreux, P., Wong, H., Bancaud, A., Viovy, J.-L., Prunell, A., and Victor, J.-M. (2009). Right-handed nucleosome: myth or reality? Cell *139*, 1216–1217; authorreply1217–authorreply1218.

Lawrimore, J., Bloom, K.S., and Salmon, E.D. (2011). Point centromeres contain more than a single centromere-specific Cse4 (CENP-A) nucleosome. J Cell Biol *195*, 573–582.

Leupold, U. (1958). Studies on Recombination in Schizosaccharomyces pombe. Cold Spring Harb Symp Quant Biol *23*, 161–170.

Li, W., Nagaraja, S., Delcuve, G.P., Hendzel, M.J., and Davie, J.R. (1993). Effects of histone acetylation, ubiquitination and variants on nucleosome stability. Biochem J *296 (Pt 3)*, 737–744.

Liu, S.-T., Rattner, J.B., Jablonski, S.A., and Yen, T.J. (2006). Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. J Cell Biol *175*, 41–53.

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.

Locke, G., Tolkunov, D., Moqtaderi, Z., Struhl, K., and Morozov, A.V. (2010). High-throughput sequencing reveals a simple model of nucleosome energetics. Pnas *107*, 20998–21003.

Lomiento, M., Jiang, Z., D'Addabbo, P., Eichler, E.E., and Rocchi, M. (2008). Evolutionary-new centromeres preferentially emerge within gene deserts. Genome Biol *9*, R173.

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

Loyola, A., and Almouzni-Pettinotti, G. (2007). Marking histone H3 variants: how, when and why? Trends in Biochemical Sciences *32*, 425–433.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell *24*, 309–316.

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

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution.

Methods Mol Biol *119*, 1–16.

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

Luger, K., and Richmond, T.J. (1998b). DNA binding within the nucleosome core. Curr Opin Struct Biol *8*, 33–40.

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

Ma, Y., Cai, S., Lu, Q., Lu, X., Jiang, Q., Zhou, J., and Zhang, C. (2008). Inhibition of protein deacetylation by trichostatin A impairs microtubule-kinetochore attachment. Cell Mol Life Sci *65*, 3100–3109.

Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A.B. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol *176*, 757–763.

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

Mai, X., Chou, S., and Struhl, K. (2000). Preferential accessibility of the yeast his3 promoter is determined by a general property of the DNA sequence, not by specific elements. Mol Cell Biol *20*, 6668–6676.

Maison, C., and Almouzni-Pettinotti, G. (2004). HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Biol *5*, 296–305.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838–849.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Pnas *101*, 1525–1530.

Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A., and Heun, P. (2011). Drosophila CENH3 Is Sufficient for Centromere Formation. Science *334*, 686–690.

Mizuguchi, G., Wisniewski, J., Wei, D., Wei, D., and Wu, C. (2011). Nonhistone Scm3 Binds to AT-Rich DNA to Organize Atypical Centromeric Nucleosome of Budding Yeast. Mol Cell *43*, 369–380.

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

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.

Müller, D.J., and Engel, A. (1997). The height of biomolecules measured with the atomic force microscope depends on electrostatic interactions. Biophysj *73*, 1633–1644.

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

Okada, M., Okawa, K., Isobe, T., and Fukagawa, T. (2009). CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. Mol Biol Cell *20*, 3986–3995.

Okada, T., Ohzeki, J.-I., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., and Masumoto, H. (2007). CENP-B controls centromere formation depending on the chromatin context. Cell *131*, 1287–1300.

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. Pnas *88*, 3734–3738.

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

Pastré, D., Piétrement, O., Fusil, S., Landousy, F., Jeusset, J., David, M.-O., Hamon, L., Le Cam, E., and Zozime, A. (2003). Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study. Biophysj *85*, 2507–2518.

Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol *10*, 886–895.

Peters, A.H.F.M., Mermoud, J.E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., and Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat Genet *30*, 77–80.

Pidoux, A.L., and Allshire, R.C. (2004). Kinetochore and heterochromatin domains of the fission yeast centromere. Chromosome Res *12*, 521–534.

Pidoux, A.L., Choi, E.S., Abbott, J.K.R., Liu, X., Kagansky, A., Castillo, A.G., Hamilton, G.L., Richardson, W., Rappsilber, J., He, X., *et al*. (2009). Fission yeast Scm3: A CENP-A receptor required for integrity of subkinetochore chromatin. Mol Cell *33*, 299–311.

Pidoux, A.L., Richardson, W., and Allshire, R.C. (2003). Sim4: a novel fission

yeast kinetochore protein required for centromeric silencing and chromosome segregation. *161*, 295–307.

Polizzi, C., and Clarke, L. (1991). The chromatin structure of centromeres from fission yeast: differentiation of the central core that correlates with function. J Cell Biol *112*, 191–201.

Pugh, B.F. (2010). A preoccupied position on nucleosomes. Nat Struct Mol Biol *17*, 923.

Ranjitkar, P., Press, M.O., Yi, X., Baker, R., Maccoss, M.J., and Biggins, S. (2010). An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. Mol Cell *40*, 455–464.

Ray-Gallet, D., Quivy, J.-P., Silljé, H.W.W., Nigg, E.A., and Almouzni, G. (2007). The histone chaperone Asf1 is dispensable for direct de novo histone deposition in Xenopus egg extracts. Chromosoma *116*, 487–496.

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N.A., Adams, P.D., Jansen, L.E.T., *et al*. (2011). Dynamics of histone H3 deposition *in vivo* reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol Cell *44*, 928–941.

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., *et al*. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

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

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

Robinson, P.J.J., and Rhodes, D. (2006). Structure of the “30 nm” chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol *16*, 336–343.

Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. J Cell Biol *146*, 905–916.

Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., Desalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. BMC Genomics *10*, 143.

Routh, A., Sandin, S., and Rhodes, D. (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Pnas *105*, 8872–8877.

Saffery, R., Wong, L.H., Irvine, D.V., Bateman, M.A., Griffiths, B., Cutts, S.M.,

Cancilla, M.R., Cendron, A.C., Stafford, A.J., and Choo, K.H. (2001). Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. Pnas *98*, 5705–5710.

Samel, A., Cuomo, A., Bonaldi, T., and Ehrenhofer-Murray, A.E. (2012). Methylation of CenH3 arginine 37 regulates kinetochore integrity and chromosome segregation. Pnas *109*, 9029–9034.

Sanchez-Pulido, L., Pidoux, A.L., Ponting, C.P., and Allshire, R.C. (2009). Common ancestry of the CENP-A chaperones Scm3 and HJURP. Cell *137*, 1173–1174.

Saxena, A., Saffery, R., Wong, L.H., Kalitsis, P., and Choo, K.H.A. (2002). Centromere proteins Cenpa, Cenpb, and Bub3 interact with poly(ADP-ribose) polymerase-1 protein and are poly(ADP-ribosyl)ated. J Biol Chem *277*, 26921–26926.

Schäfer, G., McEvoy, C.R.E., and Patterton, H.-G. (2008). The Saccharomyces cerevisiae linker histone Hho1p is essential for chromatin compaction in stationary phase and is displaced by transcription. Pnas *105*, 14838–14843.

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., and Willard, H.F. (2001). Genomic and genetic definition of a functional human centromere. Science *294*, 109–115.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet *25*, 335–343.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772–778.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol Cell *18*, 735–748.

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

Shelby, R.D., Monier, K., and Sullivan, K.F. (2000). Chromatin assembly at kinetochores is uncoupled from DNA replication. J Cell Biol *151*, 1113–1118.

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.

Shivaraju, M., Camahort, R., Mattingly, M., and Gerton, J.L. (2011). Scm3 is a centromeric nucleosome assembly factor. J Biol Chem *286*, 12016–12023.

Shivaraju, M., Unruh, J.R., Slaughter, B.D., Mattingly, M., Berman, J., and Gerton, J.L. (2012). Cell-cycle-coupled structural oscillation of centromeric

nucleosomes in yeast. Cell *150*, 304–316.

Slattery, S.D., Moore, R.V., Brinkley, B.R., and Hall, R.M. (2008). Aurora-C and Aurora-B share phosphorylation and regulation of CENP-A and Borealin during mitosis. Cell Cycle *7*, 787–795.

Stein, A., Takasuka, T.E., and Collings, C.K. (2010). Are nucleosome positions *in vivo* primarily determined by histone-DNA sequence preferences? Nucleic Acids Res *38*, 709–719.

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.

Stoler, S., Rogers, K., Weitze, S., Morey, L., Fitzgerald-Hayes, M., and Baker, R.E. (2007). Scm3, an essential Saccharomyces cerevisiae centromere protein required for G2/M progression and Cse4 localization. Pnas *104*, 10571–10576.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Strålfors, A., Walfridsson, J., Bhuiyan, H., and Ekwall, K. (2011). The FUN30 chromatin remodeler, Fft3, protects centromeric and subtelomeric domains from euchromatin formation. PLoS Genet *7*, e1001334.

Su, D., Hu, Q., Li, Q., Thompson, J.R., Cui, G., Fazly, A., Davies, B.A., Botuyan, M.V., Zhang, Z., and Mer, G. (2012). Structural basis for recognition of H3K56-acetylated histone H3-H4 by the chaperone Rtt106. Nature *483*, 104–107.

Suka, N., Suka, Y., Carmen, A.A., Wu, J., and Grunstein, M. (2001). Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. Mol Cell *8*, 473–479.

Sullivan, B.A., and Willard, H.F. (1998). Stable dicentric X chromosomes with two functional centromeres. Nat Genet *20*, 227–228.

Sullivan, B.A., Blower, M.D., and Karpen, G.H. (2001). Determining centromere identity: cyclical stories and forking paths. Nat Rev Genet *2*, 584–596.

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 *11*, 1076–1083.

Sun, X., Wahlstrom, J., and Karpen, G. (1997). Molecular structure of a functional Drosophila centromere. Cell *91*, 1007–1019.

Tachiwana, H., Kagawa, W., and Kurumizaka, H. (2012). Comparison between the CENP-A and histone H3 structures in nucleosomes. Nucleus *3*, 6–11.

Tachiwana, H., Kagawa, W., Osakabe, A., Kawaguchi, K., Shiga, T., Hayashi-Takanaka, Y., Kimura, H., and Kurumizaka, H. (2010). Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. Pnas *107*, 10454–10459.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., *et al*. (2011a). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 1–6.

Tachiwana, H., Osakabe, A., Shiga, T., Miya, Y., Kimura, H., Kagawa, W., and Kurumizaka, H. (2011b). Structures of human nucleosomes containing major histone H3 variants. Acta Crystallogr D Biol Crystallogr *67*, 578–583.

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.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Mol Biol Cell *3*, 819–835.

Takahashi, K., Chen, E.S., and Yanagida, M. (2000). Requirement of Mis6 Centromere Connector for Localizing a CENP-A-Like Protein in Fission Yeast. Science *288*, 2215–2219.

Takahashi, K., Takayama, Y., Masuda, F., Kobayashi, Y., and Saitoh, S. (2005). Two distinct pathways responsible for the loading of CENP-A to centromeres in the fission yeast cell cycle. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *360*, 595–606–discussion606–7.

Takayama, Y., Sato, H., Saitoh, S., Ogiyama, Y., Masuda, F., and Takahashi, K. (2008). Biphasic incorporation of centromeric histone CENP-A in fission yeast. Mol Biol Cell *19*, 682–690.

Talbert, P.B., Ahmad, K., Almouzni, G., Ausió, J., Berger, F., Bhalla, P.L., Bonner, W.M., Cande, W.Z., Chadwick, B.P., Chan, S.W., *et al*. (2012). A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin *5*, 7.

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, K., Tomschik, M., Ausió, J., and Zlatanova, J. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry *48*, 10852–10857.

Thåström, A., Bingham, L.M., and Widom, J. (2004a). Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. J Mol Biol *338*, 695–709.

Thåström, A., Gottesfeld, J.M., Luger, K., and Widom, J. (2004b). Histone-

DNA binding free energy cannot be measured in dilution-driven dissociation experiments. Biochemistry *43*, 736–741.

Thåström, A., Lowary, P.T., and Widom, J. (2004c). Measurement of histone-DNA interaction free energy in nucleosomes. Methods *33*, 33–44.

Tomonaga, T., Matsushita, K., Yamaguchi, S., Oohashi, T., Shimada, H., Ochiai, T., Yoda, K., and Nomura, F. (2003). Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. Cancer Res *63*, 3511–3516.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J Cell Sci *114*, 3529–3542.

Vasudevan, D., Chua, E.Y.D., and Davey, C.A. (2010). Crystal structures of nucleosome core particles containing the "601" strong positioning sequence. J Mol Biol *403*, 1–10.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell *87*, 95–104.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1998). Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. Curr Biol *8*, 96–108.

Voullaire, L.E., Slater, H.R., Petrovic, V., and Choo, K.H. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am. J. Hum. Genet. *52*, 1153–1163.

Wang, H., Bash, R., Yodh, J.G., Hager, G.L., Lohr, D., and Lindsay, S.M. (2002). Glutaraldehyde modified mica: a new surface for atomic force microscopy of chromatin. Biophysj *83*, 3619–3625.

Wang, H., Dalal, Y., Henikoff, S., and Lindsay, S. (2008). Single-epitope recognition imaging of native chromatin. Epigenetics Chromatin *1*, 10.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., *et al*. (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol *7*, 901–904.

Waterborg, J.H. (1990). Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa. J Biol Chem *265*, 17157–17161.

Weaver, B.A.A., Silk, A.D., Montagna, C., Verdier-Pinard, P., and Cleveland, D.W. (2007). Aneuploidy acts both oncogenically and as a tumor suppressor. Cancer Cell *11*, 25–36.

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.

Wieland, G., Orthaus, S., Ohndorf, S., Diekmann, S., and Hemmerich, P. (2004). Functional complementation of human centromere protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. Mol Cell Biol *24*, 6620–6630.

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

Wolffe, A.P., and Kurumizaka, H. (1998). The nucleosome: a powerful regulator of transcription. Prog. Nucleic Acid Res. Mol. Biol. *61*, 379–422.

Wong, H., Victor, J.-M., and Mozziconacci, J. (2007a). An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. PLoS ONE *2*, e877.

Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., *et al*. (2007b). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160.

Woodland, H.R., and Adamson, E.D. (1977). The synthesis and storage of histones during the oogenesis of Xenopus laevis. Dev Biol *57*, 118–135.

Wu, B., Mohideen, K., Vasudevan, D., and Davey, C.A. (2010). Structural insight into the sequence dependence of nucleosome positioning. Structure *18*, 528–536.

Wu, R.S., and Bonner, W.M. (1981). Separation of basal histone synthesis from S-phase histone synthesis in dividing cells. Cell *27*, 321–330.

Wyrick, J.J., Holstege, F.C., Jennings, E.G., Causton, H.C., Shore, D., Grunstein, M., Lander, E.S., and Young, R.A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature *402*, 418–421.

Zeitlin, S.G., Barber, C.M., Allis, C.D., Sullivan, K.F., and Sullivan, K. (2001a). Differential regulation of CENP-A and histone H3 phosphorylation in G2/M. J Cell Sci *114*, 653–661.

Zeitlin, S.G., Shelby, R.D., and Sullivan, K.F. (2001b). CENP-A is phosphorylated by Aurora B kinase and plays an unexpected role in completion of cytokinesis. J Cell Biol *155*, 1147–1157.

Zhang, W., Colmenares, S.U., and Karpen, G.H. (2012). Assembly of Drosophila centromeric nucleosomes requires CID dimerization. Mol Cell *45*, 263–269.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core

histone tails. Genes Dev *15*, 2343–2360.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. Nat Struct Mol Biol *16*, 847–852.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2010). Evidence against a genomic code for nucleosome positioning Reply to [ldquo] Nucleosome sequence preferences influence *in vivo* nucleosome organization [rdquo]. Nat Struct Mol Biol *17*, 920–923.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977–980.

Zhong, Q., Inniss, D., Kjoller, K., and Elings, V.B. (1993). Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. Surface Science Letters *290*, L688–L692.

Zhou, Z., Feng, H., Zhou, B.-R., Ghirlando, R., Hu, K., Zwolak, A., Miller Jenkins, L.M., Xiao, H., Tjandra, N., Wu, C., *et al*. (2011). Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. Nature *472*, 234–237.