Histone variants: Deviants?

Rohinton T. Kamakaka and Sue Biggins

Division of Basic Sciences,
Fred Hutchinson Cancer Research Center,
Seattle, WA 98109, USA

UCT/ NIH,
Bldg. 18T, Rm. 106,
18 Library Dr., Bethesda, MD 20892
Ph. No. 301 402 8317
Fax No. 301 402 1323
e. mail. Rohinton@helix.nih.gov
Abstract

All eukaryotic chromosomes are composed of chromatin, the protein-DNA complex fundamental to genome packaging, function and regulation. The building block of chromatin is the core nucleosome, a histone octamer wrapped with DNA. Although the major histones are some of the most highly conserved proteins known, there are variants that have specialized localization and species-distribution patterns. Recently, much progress has been made in understanding the histone variants. Here we focus on the known localization patterns and functions for the variants. In addition, we review variant assembly into chromatin, the structure of the variant chromatin, and the known post-translational modifications that occur on the variants.

Introduction

Approximately two meters of human diploid DNA are packaged into the cell’s nucleus with a volume of approximately 1000 cubic microns. This compaction is achieved by protein-mediated folding of DNA. Chromatin, the nucleoprotein complex found in the nucleus, has approximately twice the protein mass as DNA (Butler 1983) and half of this mass is the highly basic histones, H1, H2A, H2B, H3 and H4.

At the first level of packaging, the DNA is wrapped around histones to form a beaded chain. Each bead is referred to as a core nucleosome and contains an octamer of two molecules of each of the core histones H2A, H2B, H3 and H4 with two turns of DNA wrapped around the proteins (reviewed in (Luger 2003; Khorasanizadeh 2004)). These core histones all contain a conserved C-terminal histone fold domain and unique N-terminal tails. The four core histones interact in pairs via a “handshake motif” with two H3/H4 dimers interacting together to form a tetramer while the two H2A/H2B dimers associate with the H3/H4 tetramer in the presence of DNA. Multiple electrostatic, hydrophobic and hydrogen bonds at the interface of these sub-complexes are required for
nucleosome formation. The N-terminal tails of the histones do not significantly participate in the nucleosome structure but are involved in interactions with other proteins and nucleosomes. In multi-cellular eukaryotes, one molecule of histone H1 associates at the position where the DNA enters and exits the nucleosome core, thus sealing the two turns of DNA. The nucleosome filament is then folded into a 30 nm fiber mediated in part by nucleosome-nucleosome interactions, and this fiber is probably the template for most nuclear processes. Additional levels of compaction enable these fibers to be packaged into the small volume of the nucleus.

The packaging of DNA into nucleosomes and chromatin positively or negatively affects all nuclear processes in the cell. While nucleosomes have long been viewed as stable entities, there is a large body of evidence indicating that they are highly dynamic (reviewed in (Kamakaka 2003)), capable of being altered in their composition, structure, and location along the DNA. Enzyme complexes that either post-translationally modify the histones or alter the position and structure of the nucleosomes carry out these functions. There are a wide variety of post-translational modifications that occur on histones, such as phosphorylation, methylation, acetylation and ubiquitylation (Iizuka and Smith 2003) and these modifications affect the properties of the histones. Moreover, chromatin remodeling complexes contain ATPase subunits and are known to either slide nucleosomes, replace histones, or alter the histone-DNA interactions (Tsukiyama 2002; Langst and Becker 2004). A third way to modulate chromatin is via incorporation of histone variants. Here we provide an overview of the best-characterized histone variant functions and the ways that they can alter chromatin to facilitate various cellular processes.

An introduction to the Variants

In most organisms, there are multiple copies of the histone genes encoding for the
major histone proteins. These genes are highly similar in sequence, expressed primarily during the S-phase of the cell cycle, and code for the bulk of the cellular histones. While histones are among the slowest evolving proteins known, there are non-allelic variants of the major histones that can have significant differences in primary sequence. These variants are usually present as single copy genes that are expressed constitutively or in specific cell types. Some variants have distinct biophysical characteristics that are thought to alter the properties of nucleosomes. Furthermore, variants often localize to specific regions of the genome. These observations have led to the suggestion that the histone variants have specialized functions regulating chromatin dynamics.

**Who are the variants?**

The similarity between the major histone subtypes and the variants can range from almost no amino acid differences to extremely divergent changes. Because the phylogeny of the variants was comprehensibly reviewed recently (Malik and Henikoff 2003), we refer the reader to this review for an understanding of the evolution of the variants.

**Histone H1**

Histone H1 has numerous sequence variants such as $H1^0$, $H5$ and the **sperm** and **testis** specific variants. Most of the sequence differences between the major histone subtype and the variants occur in the non-globular N- and C-terminal tail domains of these proteins. The abundance of these variants fluctuates in different cell types as well as during the cell cycle, differentiation and development (reviewed in (Cole 1987; Brown 2001; Parseghian and Hamkalo 2001; Brown 2003)). Furthermore, the major histones and variants have distinct biophysical properties (Cole 1987; Ramakrishnan 1997) and different distribution patterns in the genome (Parseghian and Hamkalo 2001). Based on these observations, it has been suggested that the H1 variants have specific functions,
although tests of this prediction have uncovered only subtle functional differences (Brown et al. 1996; Shen and Gorovsky 1996; Lin et al. 2000; Alami et al. 2003).

**Histone H2A**

Amongst the core histones, H2A has the largest number of variants, including **H2A.Z**, **MacroH2A**, **H2A-Bbd**, **H2AvD**, and **H2A.X** (Ausio and Abbott 2002; Redon et al. 2002; Fernandez-Capetillo et al. 2004) (see Figure1). Some H2A variants, like H2A.Z, are conserved through evolution (Jackson et al. 1996), while others, such as MacroH2A (Pehrson and Fuji 1998) and H2A-Bbd (Chadwick and Willard 2001), are restricted to vertebrates or mammals. The H2A variants are distinguished from the major H2A histones by their C-terminal tails that diverge in both length and sequence as well as in their genome distribution. Macro-H2A localizes predominantly to the inactive X-chromosome (Costanzi and Pehrson 1998) while H2A-Bbd localizes to the active X-chromosome and autosomes (Chadwick and Willard 2001). H2A.X and H2A.Z are constitutively expressed and localize throughout the genome, although H2A.Z shows some enrichment in intergenic regions. Interestingly, the major H2A proteins in *S. cerevisiae* and *S. pombe* are more similar to the mammalian H2A.X variant than the mammalian major H2A subtypes (Malik and Henikoff 2003) (see Figure1). In *Drosophila*, a single variant called H2AvD has sequence characteristics of both H2A.X and H2A.Z (Redon et al. 2002). Because the *Drosophila* protein likely encompasses the separate functions ascribed to both H2A.Z and H2A.X in mammals, care needs to be taken in comparing the functions of variants between species.

**Histone H2B**

Histone 2B is markedly deficient in variants. The few that have been documented completely replace the major H2B subtypes and appear to have very specialized functions in chromatin compaction and transcription repression during gametogenesis (reviewed in (Poccia and Green 1992; Green et al. 1995)). Unlike the major H2B subtypes, the sperm
specific H2B in sea urchins has a long N-terminal tail that is highly charged. This tail assists in the condensation of chromatin fibers suggesting that this variant may play a role in packaging the chromatin in the sperm. There are additional H2B variants that are developmental stage specific but their specific role is unclear.

Histone H3

There are two major histone H3 variants in mammalian cells, called centromeric H3 (CenH3) and H3.3 (Ahmad and Henikoff 2002a; Malik and Henikoff 2003) (see Figure2). CenH3 is a conserved essential protein that binds to centromeres, the DNA locus that directs formation of the kinetochore protein structure that mediates chromosome segregation in eukaryotes. Despite similarity in the histone fold domain, all CenH3 proteins have highly divergent N-terminal tails. H3.3 is the least divergent variant, containing only 4 amino acid differences compared to H3 in Drosophila. However, unlike the major H3 histones, it is expressed throughout the cell cycle and localizes to transcriptionally active regions of the chromosome (Ahmad and Henikoff 2002b). Similar to H2A.X, the major H3 proteins in S. cerevisiae (Ahmad and Henikoff 2002b) are more similar to mammalian H3.3 than H3.

Histone H4

Histone H4 is one of the slowest evolving proteins and there appear to be no known sequence variants of histone H4. However, there are H4 genes that are constitutively expressed throughout the cell cycle that encode for proteins that are identical in sequence to the major H4 (Akhmanova et al. 1996). The reason for a lack of sequence variants is not clear.

Are variants deviant in behavior? Functions of the variants

Most of the studies aimed at elucidating the functions of the histone variants are based on the correlation between the localization of the variant and the activity of the
locus, or on analyses of phenotypes associated with the loss of the variant.

**Transcriptional activation and repression**

Several histone H1 variants appear to have roles in transcription, particularly in repression during differentiation (Poccia and Green 1992; Doenecke et al. 1994; Buttinelli et al. 1999). One example is histone H5 in chicken erythrocytes. This variant is deposited into chromatin during the terminal stages of erythrocyte differentiation and its deposition coincides with global transcriptional repression (Wagner et al. 1977). The H5 variant is depleted from active genes in vivo and the presence of this variant represses transcription initiation in vitro (reviewed in (Paranjape et al. 1994)).

The Macro-H2A variant is also thought to be involved in transcriptional repression. This variant localizes to the inactive X-chromosome (Costanzi and Pehrson 1998), and while binding does not initiate X inactivation (Mermoud et al. 1999), current models do suggest that the C-terminal tail of Macro-H2A can repress transcription by blocking access to transcription factors and coactivators (Perche et al. 2000; Angelov et al. 2003; Abbott et al. 2004).

In contrast, H2A-Bbd lacks a significant C-terminal tail and it has been postulated that the lack of such a tail may destabilize the nucleosome, thus aiding in ease of nucleosome displacement during transcription (Angelov et al. 2004; Bao et al. 2004; Gautier et al. 2004). This role is consistent with the localization of H2A-Bbd to the active X-chromosome and autosomes (Chadwick and Willard 2001).

H2A.Z has also been linked to transcription. In *S. cerevisiae*, genetic interactions have been observed between H2A.Z and numerous transcription cofactors (Santisteban et al. 2000; Adam et al. 2001; Hwang et al. 2003; Krogan et al. 2003a; Krogan et al. 2003b; Zhang et al. 2004) while in *Tetrahymena*, H2A.Z is present in the transcriptionally active macronucleus and only appears in the micronucleus when this organelle becomes transcriptionally active (Allis et al. 1986). Although H2A.Z is generally distributed...
throughout the genome (Dhillon and Kamakaka 2000; Leach et al. 2000; Krogan et al. 2003a; Kobor et al. 2004; Mizuguchi et al. 2004), elevated amounts are found in the intergenic regions of some genes (Santisteban et al. 2000; Krogan et al. 2003a; Kobor et al. 2004; Mizuguchi et al. 2004), consistent with a role in gene regulation (see figure 3).

H2A.Z is necessary for the efficient induction of specific genes, localizes to the promoters of these genes when they are inactive and is subsequently lost from these promoters upon induction (Leach et al. 2000; Santisteban et al. 2000; Krogan et al. 2003a; Larochelle and Gaudreau 2003; Kobor et al. 2004). Furthermore, the absence of this variant leads to compromised recruitment of TBP and RNA polymerase II and delayed induction of the gene (Larochelle and Gaudreau 2003; Santisteban and Smith 2004). It is thus likely that the presence of the variant enables promoters to be rapidly induced, while the deposition of H2A.Z at promoters probably occurs as transcription terminates and nucleosomes reform at the promoter, since promoters are usually nucleosome free when the genes are active (Boeger et al. 2004; Lee et al. 2004).

Despite the role for H2A.Z in transcription, it is a non-essential gene in yeast, only necessary for the activation of 214 genes and the repression of 107 genes as assayed by DNA microarray analysis (Meneghini et al. 2003). Therefore, H2A.Z localization to intergenic regions does not reflect an essential role in transcription of all genes, probably due to functional redundancy with other chromatin remodeling and modifying complexes (Santisteban et al. 2000; Adam et al. 2001) as well as other functions that are described below.

Surprisingly, results from multi-cellular organisms indicate that H2A.Z may also be involved in transcriptional repression at heterochromatin. Immunofluorescence analyses in mammalian cells indicate that H2A.Z co-localizes with foci containing the heterochromatic protein HP1α and unacetylated histones H3 and H4 (Leach et al. 2000; Rangasamy et al. 2003; Rangasamy et al. 2004). Furthermore, depletion of H2A.Z is
accompanied by the loss of HP1α from chromosomal arms. Because the absence of acetylated histones and the presence of HP1 are characteristics of transcriptionally inactive heterochromatin, this variant may play a role in transcription repression. It needs to be noted that although HP1 has been shown to localize to active genes and function in transcriptional activation (Piacentini et al. 2003), it is also possible that H2A.Z has roles in both transcriptional repression and activation. Whether the observed differences between yeast and mammals are due to the different assays used in these studies or reflect fundamental differences between these organisms is not known and will need to be resolved.

The H3.3 histone variant also plays a role in transcription. One of the distinguishing features of this variant is that it is constitutively expressed during the cell cycle (though there is increased expression during S-phase) and is deposited into chromatin outside of S-phase. H3.3 is present in genes that are either poised for transcription, or are actively transcribed. It is widely accepted that nucleosome disruption occurs during nuclear processes such as transcription and DNA repair, and the consequent loss of histones need to be replaced. Because the *Drosophila* H3.3 variant is deposited at transcriptionally active loci like the rDNA outside of S-phase (Ahmad and Henikoff 2002b), H3.3 may serve to replace H3 at active genes as nucleosomes reform behind the transcribing polymerase (see figure4). Although H3.3 is less abundant than the major histone H3, there is a sufficient amount of this variant to bind all of the active genes in the nucleus (Henikoff et al. 2004). In this context, it is significant to note that in yeast, where the major histone is H3.3, the vast majority of its genome is packaged in a transcriptionally active or competent state (Lohr and Hereford 1979). Replacement of the major H3 with the variant could potentially mark active genes and aid in future rounds of transcription initiation as well as allow histone modifications to be changed due to removal of the histone.
Heterochromatic Barriers

Some regions of the chromatin are transcriptionally inactive, or "silenced". In yeast, silencing is achieved by the binding of a complex of repressor proteins that spreads along the chromatin that is silenced. The silenced chromatin domains are restricted from spreading along the DNA by the presence of barrier elements (Donze and Kamakaka 2002). H2A.Z, which was initially isolated as a weak suppressor of a silencing defect in budding yeast (Dhillon and Kamakaka 2000), was subsequently shown to be enriched in regions adjacent to the silenced domains and function in parallel with barrier elements to block the spread of silencing (Meneghini et al. 2003). Consistent with its role in transcription activation in \textit{S. cerevisiae}, current models suggest that H2A.Z containing chromatin is an unfavorable substrate for the binding of silencing proteins (Kimura et al. 2002; Suka et al. 2002; Krogan et al. 2003a; Meneghini et al. 2003; Kobor et al. 2004; Zhang et al. 2004).

Genome Stability

Some histone variants contribute to genome stability by regulating the fidelity of chromosome segregation or the efficiency of DNA replication and repair. The CenH3 variant is required for accurate chromosome segregation in every organism examined (Stoler et al. 1995; Figueroa et al. 1998; Howman et al. 2000; Takahashi et al. 2000; Blower and Karpen 2001; Oegema et al. 2001; Goshima et al. 2003). There are two major functions that CenH3 is likely to fulfill at centromeres-first it has been proposed that CenH3 is the epigenetic mark that specifies the site of kinetochore formation. This is supported by the observation that all active centromeres contain CenH3 whereas inactive centromeres do not (Warburton et al. 1997; Ouspenski et al. 2003). However, CenH3 does not appear to be sufficient for centromere identity because mistargeting of CenH3 to euchromatin causes some, but not all kinetochore proteins to mislocalize with it (Van Hooser et al. 2001). Therefore, additional mechanisms must assist CenH3 in specifying
the site of kinetochore formation. An idea that was recently proposed is that histone modifications specific for centromeric chromatin could also aid in propagating centromere identity (Sullivan and Karpen 2004).

The other major function for CenH3 is in directing assembly of the proteinaceous kinetochore structure. In worms, CenH3 depletion leads to a kinetochore null phenotype where most kinetochore proteins examined were mislocalized (Oegema et al. 2001). Consistent with this, CenH3 depletion experiments and CenH3 null mice exhibited altered localization of many kinetochore proteins (Howman et al. 2000; Blower and Karpen 2001). CenH3 is essential for the specialized centromeric chromatin structure in budding and fission yeast (Meluh et al. 1998; Takahashi et al. 2000), and directly or indirectly interacts with many kinetochore proteins (Van Hooser et al. 2001) (see figure5). Taken together, these data suggest that the kinetochore protein binding sites in CenH3 combined with the underlying chromatin structure created by CenH3 nucleosomes create an environment favorable for kinetochore assembly. There may also be additional CenH3 functions that have not yet been fully explored, such as recruitment of the cohesion complex that holds sister chromatids together, positioning of the mitotic spindle and cytokinesis (Tanaka et al. 1999; Glowczewski et al. 2000; Zeitlin et al. 2001b).

Histone H2A.Z also regulates genomic integrity. It is present in pericentric chromatin in numerous organisms (Rangasamy et al. 2003; Krogan et al. 2004; Rangasamy et al. 2004), and H2A.Z mutant cells exhibit defects in DNA repair and genome instability (Carr et al. 1994; Madigan et al. 2002; Krogan et al. 2003a; Rangasamy et al. 2003). Genetic and biochemical interactions have also been observed between H2A.Z and proteins involved in chromosome segregation, replication and repair (Krogan et al. 2003a; Rangasamy et al. 2003; Dhillon et al. 2004) and the role of H2A.Z in these processes is just beginning to be explored.

Genome stability also requires the H2A.X variant. Double strand breaks that
occur during replication, recombination or DNA rearrangement must be repaired. While H2A.X is expressed throughout the cell cycle and deposited all over the chromosomes, it is preferentially phosphorylated by the ATM and ATR kinases at sites that flank double stranded breaks (Rogakou et al. 1999) and this phosphorylation is essential to recruit many components of the DNA damage response to these sites (Paull et al. 2000). Although H2A.X has not been shown to directly mediate DNA repair, it is important for suppressing oncogenic translocations and tumor formation (Celeste et al. 2003). It is possible that H2A.X phosphorylation helps retain repair proteins at the site of damage, or facilitates interactions between chromosomes that are important for DNA repair.

**How do variants find their way home? Assembly of variant nucleosomes**

The variety of localization patterns and functions exhibited by the histone variants leads to the question of how the variants assemble into the proper chromosomal loci. In dividing cells, the synthesis and deposition of the major histone subtypes occurs during S-phase and is coupled with replication (see figure4). The deposition of these histones is dependent upon CAF-1, a chromatin assembly factor that deposits histones H3 and H4 during replication (reviewed in (Loyola and Almouzni 2004)). Other proteins like Nap1, Sp4, HIRA and Hif1 can also deposit histones into chromatin though they do so independent of DNA replication. The expression of some histone variants like H3.3, H2A.X and H2A.Z, is constitutive and until quite recently, very little was known about the factors required for the assembly of these histone variants into chromatin. However, new studies in several labs isolated the histone variants H2A.Z and H3.3 as components of separate soluble complexes that are most likely involved in depositing these variants into chromatin outside of S-phase. None of the other variants have been purified with chromatin remodeling factors that assist in their assembly, so future studies will be aimed at determining whether these factors use known modes of assembly or new ones.
**H2A.Z Assembly**

H2A.Z has been identified in two complexes (Krogan et al. 2003a; Kobor et al. 2004; Mizuguchi et al. 2004): one contains the H2A/H2B histone chaperone/assembly protein Nap1 and the other contains a Swi/Snf like ATPase called Swr1. Whether these two complexes function together or in separate pathways needs to be determined, but the incorporation of H2A.Z is dramatically reduced in the absence of Swr1 in vivo (Krogan et al. 2003a; Kobor et al. 2004; Mizuguchi et al. 2004). Swr1 can mediate the exchange of H2A with H2A.Z in nucleosomes in an ATP-dependent manner in vitro (Mizuguchi et al. 2004), and the Swr1 containing complex is likely involved in transcription-dependent deposition of this variant since there is considerable overlap in the genes that are misregulated in cells lacking either H2A.Z or Swr1 (Kobor et al. 2004; Mizuguchi et al. 2004; Zhang et al. 2004). Unfortunately, it is not clear whether the Swr1-mediated deposition of this variant is a cause or a consequence of transcription. It is interesting to note that H2A.Z deposition also depends on Yaf9, a component of both the Swr1 and the NuA4 histone acetyltransferase complexes (Zhang et al. 2004). Since acetylation is correlated with active transcription, it raises the question of whether H2A.Z is targeted to chromatin that is acetylated or vice versa.

**H3.3 Assembly**

As mentioned previously, there are only minor differences between major H3 and H3.3. Three out of four changes reside in the histone fold domain. While converting any one of these three residues in *Drosophila* H3.3 to that in major H3 does not affect the replication independent deposition of H3.3, changing any one of the residues in major H3 to its counterpart in H3.3 allows major H3 to be incorporated into chromatin outside of S-phase (Ahmad and Henikoff 2002b).

This sequence specificity likely reflects interactions with different assembly factors. While mammalian major H3 histones are deposited by the CAF-1 replication-
coupled chromatin assembly complex, H3.3 incorporation requires the HIRA complex (Tagami et al. 2004) that mediates deposition in a replication-independent manner in vitro (Ray-Gallet et al. 2002). The sites of H3.3 deposition outside of S-phase are transcriptionally active loci suggesting that the HIRA complex may use transcription-coupled deposition to replace major H3 with H3.3 (Ahmad and Henikoff 2002b). The presence of histone H3.3 instead of the major H3 subtypes in *S. cerevisiae* may explain the ability of this organism to survive in the absence of CAF-1. Despite the identification of the H3.3 assembly complex, it is still unclear how HIRA targets H3.3 to transcriptionally active genes.

**CenH3 Assembly**

Because other histones and variants are deposited by chaperones, it has been assumed that a specific loading complex will exist for CenH3. However, a CenH3 loading factor has not yet been identified, presumably because the low levels of soluble CenH3 make it difficult to identify interacting factors. The only chromatin-related proteins that are known to have a role in CenH3 centromere localization are the RbAp46/48 proteins that are components of a number of complexes including CAF-1 (for review, see (Loyola and Almouzni 2004)). In fission yeast, the RpAp46/48 homolog Mis16 and the conserved Mis18 protein form a complex that regulates kinetochore function, and defects in RpAp46/48 in both pombe and mammalian cells result in CenH3 mislocalization from the centromere (Hayashi et al. 2004). Because a direct interaction with CenH3 has not yet been detected, it is still unclear if these proteins directly mediate CenH3 loading or instead alter centromeric chromatin structure to allow CenH3 deposition. Mutants in the fission yeast Mis16 and Mis18 proteins affect the acetylation state of the centromere, suggesting a potential link between CenH3 localization and acetylation.

Several other chromatin-assembly complexes have roles in *S. cerevisiae*
kinetochore function and cause CenH3 to mislocalize to euchromatin. The CAF-1 and HIRA complexes have overlapping roles in kinetochore function (Sharp et al. 2002), and mutants in the Spt4 transcription factor and the RSC chromatin-remodeling complex have kinetochore defects (Tsuchiya et al. 1998; Hsu et al. 2003; Crotti and Basrai 2004). Proteins from each of these sub-complexes localize to kinetochores, and mutants in all complexes alter the centromeric chromatin structure and lead to chromosome missegregation phenotypes (Sharp et al. 2002; Hsu et al. 2003; Crotti and Basrai 2004). Although CenH3 mislocalizes to euchromatin in some of these mutants, the centromeric localization of CenH3 is unaffected. However, subtle changes, such as a shift in CenH3 nucleosomal positioning, have not yet been assayed in the mutant strains.

Recent data also suggest a potential link between CenH3 localization, kinetochore function, and transcription. In fission yeast, a putative GATA transcriptional factor, \textit{ams2}, binds to the central centromere region where CenH3 localizes (Chen et al. 2003). Cells defective in \textit{ams2} have an altered centromeric chromatin structure with reduced levels of CenH3 at centromeres and defects in kinetochore function. Whether this protein is modulating these effects through recruitment of kinetochore components or via transcription is not clear. Transcription has not been detected at the core centromeres where CenH3 localizes, probably because transcription can abolish kinetochore function (Hill and Bloom 1987). However, one could imagine that transcription could create an open chromatin environment that is more permissive for CenH3 assembly or alternatively, remove H3 nucleosomes to allow replacement by CenH3 nucleosomes. The presence of Spt4 at centromeres could help in this process (Crotti and Basrai 2004) since this complex is involved in chromatin assembly and transcription elongation (Winston 2001). Although intriguing, the isolation of transcription factors that affect kinetochore function may also be indirect due to transcription defects elsewhere, or due to additional functions for the proteins that are not related to transcription.
CenH3 is normally deposited only during S phase though it is not known if its deposition is replication coupled or not (Shelby et al. 2000; Pearson et al. 2004). Because ectopically expressed CenH3 fusion proteins can localize to the centromere at all cell cycle stages, CenH3 localization does not strictly depend on DNA replication (Sullivan et al. 1994; Shelby et al. 2000; Ahmad and Henikoff 2002a). However, it is not known if the exogenously expressed CenH3 uses the same mechanism of deposition as endogenous CenH3.

It has been suggested that the timing of centromere replication and spatial restriction within heterochromatin may aid in the localization of the variant (Ahmad and Henikoff 2001). However, recent evidence showed that the centromere is replicated at the same time as euchromatic regions that contain the canonical H3 (Shelby et al. 2000; Sullivan and Karpen 2001; Blower et al. 2002), so CenH3 deposition cannot be controlled solely by a restricted time of DNA replication.

The inheritance of CenH3 during centromere duplication has recently been investigated and it was found that budding yeast CenH3 is completely replaced during S-phase (Pearson et al. 2004). When synthesis of a fluorescently tagged CenH3 protein was repressed, the centromere-bound tagged protein was completely replaced by the endogenous protein in S-phase suggesting that yeast use a dispersive mode of CenH3 replication. Similar experiments need to be performed in multi-cellular eukaryotes to determine if this is a conserved mode of CenH3 duplication.

The kinetochore structure is also important for CenH3 localization. Mutants in the budding yeast Ndc10 kinetochore protein completely abolish the localization of all kinetochore proteins including CenH3 (He et al. 2001). In addition, the fission yeast proteins Mis6, Mis15, Mis16, Mis17, Mis18, and Sim4 that bind to the central centromere are all required for CenH3 localization (Takahashi et al. 2000; Pidoux et al. 2003; Hayashi et al. 2004). While it is possible that these proteins play a direct role in
CenH3 localization, it is just as likely that they help assemble a proper chromatin structure for CenH3 binding or help stabilize its binding following deposition.

A number of key questions about CenH3 localization remain-first, it is still completely unclear how CenH3 is targeted to the centromere although a number of factors appear to be involved. If a loading factor is isolated, it will be important to determine whether it exchanges CenH3 for H3 nucleosomes, similar to the mechanism used for H2A.Z localization by the Swr1 complex. In addition, it will be interesting to determine how the loading factor specifically recognizes the centromere and not the euchromatin. Other questions include how CenH3 is prevented from spreading beyond the centromere and the role of modifications in its deposition and localization.

**How deviant are variants? Structure of variant nucleosomes**

The large number of histone variants leads to the question of how many different nucleosome structures exist, and whether the structural alterations can account for differences in function and localization. Crystal structures of H1 variants (Cerf et al. 1994) and nucleosomes containing some histone variants have been solved (Suto et al. 2000; White et al. 2001). While there is no single unifying theme that characterizes all of these structures, the basic structure of the variant is similar to the structure of the major histone subtype.

The canonical core histones bind tightly to the DNA via arginine side chains and there are also numerous hydrogen bonds and water-mediated protein-DNA interactions between the canonical histones and DNA (reviewed in (Luger 2003; Khorasanizadeh 2004)). Most of these residues and the basic histone-DNA contacts are conserved in the variants. While there are no sequence specific interactions between the core histone side chains and the DNA bases for the major histone subtypes, it will be interesting to see if a variant such as CenH3, which has some DNA targeting specificity, also lacks interactions
between the histones and the bases. The two biggest changes due to the presence of variants appear to be in the stability of the nucleosome and the residues of the nucleosome that are exposed.

**Variant nucleosome surface residues**

One key finding of the structural studies is that variant nucleosomes have changes on the exposed surface. Macro-H2A has an extensive C-terminal tail that likely extends away from the nucleosome and imparts an asymmetrical structure to the variant nucleosome that may be important for transcriptional repression (Allen et al. 2003; Abbott et al. 2004). While the overall structure of H2A.Z nucleosomes is similar to the major H2A structure (Suto et al. 2000), two of the most striking differences are the presence of an extended acidic patch on the nucleosome surface and a novel divalent cation-binding pocket. These changes on the surface of the nucleosome could alter protein-nucleosome and nucleosome-nucleosome interactions as well as the higher order folding of the chromatin.

**Variant nucleosome stability**

Crystallography and various biophysical studies also indicate that there are changes in the stability of variant nucleosomes. FRET experiments with fluorescence donor and acceptor pairs attached at different locations in a nucleosome suggest that the overall binding of the H2A.Z/H2B dimer to the H3/H4 tetramer is slightly stabilized (Park et al. 2004). Recent data suggests that the CenH3/H4 tetramer (Black et al. 2004) is more compact and rigid than a H3/H4 tetramer and may also be more stable. The additional rigidity may be important to resist the microtubule pulling forces at the centromere during mitosis. Similarly, the Macro-H2A nucleosomes may also be more stable (Changolkar and Pehrson 2002; Abbott et al. 2004) though additional biophysical studies will be required to fully understand the differences between the variant and major histone subtypes. Also, the in vivo consequences of a more stable variant nucleosome
scattered among the canonical nucleosomes are hard to predict.

In contrast to the other variant nucleosomes, the H2A-Bbd nucleosome structure may be weaker. In the absence of DNA it is unable to form a stable histone octamer and the H2A-Bbd nucleosome organizes only 118 bp of DNA rather than the 147 bp around the histone core (Bao et al. 2004). While these nucleosomes are not very mobile, they are less stable and more accessible to transcription factors (Angelov et al. 2004; Bao et al. 2004; Gautier et al. 2004). Therefore, it is likely that the structural alterations in the H2A-Bbd nucleosome lead to a weaker nucleosome structure that facilitates gene activation.

**Variant nucleosome composition**

A final observation is that certain variants are unable to coexist with the major histone subtypes. For example, the structure of the H2A.Z containing nucleosome suggests that H2A and H2A.Z may not co-exist in the same nucleosome. Similarly, the CenH3, Macro-H2A and H2A-Bbd nucleosomes may also be homotypic as CenH3 interacts with H2A, H2B, and H4, but not H3 in vivo (Shelby et al. 1997; Blower and Karpen 2001; Westermann et al. 2003), and nucleosome reconstitution experiments with either Macro-H2A or H2A-Bbd showed they completely replace H2A (Angelov et al. 2003; Angelov et al. 2004; Gautier et al. 2004). Although these variants likely exist only in homotypic nucleosomes, it is possible that certain variants (like H3.3) will exist in heterotypic nucleosomes along with the canonical histones.

An interesting observation that arises from the crystallographic analysis of the *S. cerevisiae* nucleosome is that different variant histones may co-exist in the same nucleosome. The major H2A and H3 histones in *S. cerevisiae* are most similar to the mammalian histone variants H2A.X and H3.3, respectively. The yeast nucleosome structure has revealed that H3.3 coexists in the same nucleosome as H2A.X (White et al. 2001). Therefore, nucleosome alterations could potentially come from combinations of
variants in addition to specific changes associated with a single variant.

**Higher Order Structures**

The details about variant nucleosome structures lead to the question of how the changes in structure affect the higher order chromatin structure. This is especially important because it has long been believed that histone variants may exert their effects via changes in the higher order packaging of chromatin. While earlier studies convincingly showed that histone H1 variants condensed chromatin to a greater extent compared to the major H1 subtypes (reviewed in (Thomas 1984)), only recently have such analyses been extended to core histone variants.

**H2A.Z chromatin**

Two independent studies used positioned arrays of 12 nucleosomes and recombinant histones (H2A and H2A.Z) to analyze the folding of chromatin fibers. In one study, the presence of H2A.Z helped facilitate the folding of the nucleosomal filament into the 30nm fiber as a function of divalent cations (Fan et al. 2002) while in a second independent study, H2A.Z containing arrays were consistently less folded as a function of monovalent salt (Abbott et al. 2001). Although it is hard to reconcile the results given that both studies used similar histones and DNA, it is possible that the differences observed are due to the different cations utilized, especially since H2A.Z has a divalent cation-binding pocket. While these studies need further clarification, an important caveat is that there is no in vivo evidence of long arrays of nucleosomes containing either one or another type of variant. Thus, the effects of having a few H2A.Z variant nucleosomes located among nucleosomes containing the major histones have not yet been explored.

**CenH3 chromatin**

While it was originally thought that CenH3 forms a linear array at centromeres, recent microscopy on extended chromatin fibers showed that the linear relationship between CenH3 and the major H3 is not exclusionary (Blower and Karpen 2001).
Instead, arrays of the major H3 nucleosomes are dispersed throughout the CenH3-rich chromatin in flies and humans (see figure5). This suggests that the higher order structure assembles in a manner that causes the CenH3 nucleosomes to orient together to form the base of the kinetochore and exclude the H3 nucleosomes. This is consistent with the holocentric chromosomes of C. elegans that contain both CenH3 and H3 nucleosomes in a linear array and also orient CenH3 nucleosomes to form the kinetochore along the length of the chromosome (Buchwitz et al. 1999). Whether similar structures form with other histone variants such as H2A.Z is not known.

What makes a variant a variant? Specificity within the variants

As discussed above, variant nucleosomes may exert their effects via changes in the stability of the nucleosome and chromatin fiber, or by changes in the surface residues that are available for interactions with cellular proteins. Several studies have begun to dissect the sequences in each variant that are crucial for function and may therefore lead to these changes.

Domains in H2A variants

For the histone H2A variants, the C-terminal tail appears to distinguish their specific functions. The invariant SQE motif in the tail of H2A.X is crucial for function because it is the site of reversible phosphorylation that occurs in response to double strand breaks (Reviewed in (Redon et al. 2002)).

Like H2A.X, the C-terminal of H2A.Z is important for its function in S. cerevisiae and Drosophila. Domain swaps in Drosophila between specific H2A.Z sequences and the major H2A identified the C-terminal docking domain that interacts with histone H4 as required for viability (Clarkson et al. 1999). However, the underlying basis for the lethality seen with the C-terminal swaps in Drosophila is not yet clear. In S. cerevisiae, the C-terminal tail likely recruits transcription factors because in vitro binding
studies indicate that this domain interacts with the general transcription machinery (Adam et al. 2001; Larochelle and Gaudreau 2003). In addition, it is required for growth on medium containing galactose (Adam et al. 2001) and for the localization of H2A.Z to the promoter of the GAL1 gene, suggesting it plays a role in recruiting various factors to the regulatory regions of these genes.

In contrast to the studies in Drosophila and yeast, in Tetrahymena the lysines in the N-terminus of H2A.Z are important for function (Ren and Gorovsky 2001). While the differences between Tetrahymena (Ren and Gorovsky 2001), Drosophila (Clarkson et al. 1999) and S. cerevisiae (Adam et al. 2001) appear striking at first glance, it should be noted that these studies are not directly comparable. It is not clear if converting all the N-terminal lysine residues to arginine in Drosophila or S. cerevisiae would have any phenotype and conversely, the phenotype of deleting the C-terminal tail in Tetrahymena is also not known.

For the Macro-H2A variant, both the histone fold and long C-terminal tail have roles in transcriptional repression. While the histone fold domain prevents nucleosome sliding and is responsible for assembly into the nucleosome, the tail interferes with transcription factor binding (Perche et al. 2000; Angelov et al. 2003). Therefore, both domains in Macro-H2A appear to inhibit transcription using different mechanisms.

Domains in H3 variants

Numerous studies have dissected the CenH3 residues required for targeting to centromeres and for proper kinetochore function. While both the CenH3 N-terminal tail and histone fold domain are essential for function, only the histone fold domain contains the centromere targeting information (Sullivan et al. 1994; Shelby et al. 1997; Keith et al. 1999). Although there are few distinguishing features among the various CenH3 histone fold domains, studies have led to the conclusion that loop I and helix II are critical for CenH3 centromere targeting in flies and mammals (Shelby et al. 1997; Keith et al. 1999;
Vermaak et al. 2002; Black et al. 2004). These same regions also confer the more rigid structure to the CenH3/H4 tetramer (Black et al. 2004), suggesting this feature may be important for targeting CenH3.

CenH3 might have some DNA binding specificity because the residues necessary for targeting are predicted to map to H3/DNA contact sites. This may seem surprising since centromeres are one of the most rapidly evolving sequences in the genome (Schueler et al. 2001). However, the observation that the *Drosophila* CenH3 loop I undergoes adaptive evolution is consistent with the idea that CenH3 has DNA binding specificity that is evolving along with the centromeric sequence (Henikoff et al. 2001). Although α-satellite DNA is a hallmark of mammalian centromeric DNA, CenH3 is also found at neocentromeres that completely lack α-satellite DNA (Lo et al. 2001a; Lo et al. 2001b). Therefore, any targeting specificity must arise from a unique secondary structure instead of primary sequence requirements.

The essential N-terminal tails of all the CenH3 variants diverge in sequence and in length ranging from 27 to greater than 400 residues. The only organism where the essential residues have been mapped is in budding yeast where a 33 amino acid N-terminal domain (END) is sufficient to provide the essential function (Keith et al. 1999; Chen et al. 2000). Although the N-terminal domains are not necessary for centromere targeting, they are required for the binding of other kinetochore proteins. In budding yeast, the END domain binds to the Ctf19 kinetochore protein while in mammalian cells, this domain appears to recruit CENP-C (Chen et al. 2000; Van Hooser et al. 2001). The END domain does not have known homology to other proteins or other CenH3 N-terminal tails, and the divergence in CenH3 N-terminal sequence and length is likely due to the differences in the kinetochore proteins that are recruited by each CenH3 (see below).

Unlike the other variants, which possess domains that are distinct from the major
histones, in *Drosophila* H3.3, only four amino acids distinguish this variant from the major histones and three of these residues reside in the histone fold domain (Ahmad and Henikoff 2002b). These residues likely specify the mode of deposition of this variant as discussed above.

**Can you alter a variant? Modifications of the variants**

Histones are post-translationally modified in vivo by a host of different enzymes and these alterations subtly but specifically alter the characteristics of these proteins in chromatin. There are numerous functions ascribed to core histone modifications. The presence of specific modifications may act as a histone code functioning to recruit proteins that recognize the modified residue. The charge neutralization that accompanies certain modifications may reduce the strength of histone-DNA interaction, allowing the chromatin to “breathe” thus facilitating various processes. Alternatively, changes in modifications may aid variant deposition into chromatin or be required to evict the variant from chromatin. Changes could also affect the higher order structure of chromatin that might subsequently affect binding of non-histone proteins to chromatin. While the presence of variant nucleosomes likely affects the characteristics of the chromatin fiber, modifications of the variants may also modulate chromatin dynamics. Studies on histone modification of the variants are few, but essentially suggest that modifications are also important for the proper functioning of these proteins.

**H1 modifications**

One function of histone modifications is thought to be in helping mediate deposition and removal of the protein from chromatin. The same appears to be true for the variants since histone H1 variants are phosphorylated on their tails during deposition into chromatin and prior to their removal from chromatin (Wagner et al. 1977; Poccia and Green 1992; Green et al. 1995; Dou et al. 1999).
H2A modifications

Histone acetylation is a common modification associated with the major core histones and this modification also decorates the H2A variants. The N-terminal tail of H2A.Z is reversibly acetylated on lysine residues 4, 7, 10, 13, 16 and 21 in *Tetrahymena* (Ren and Gorovsky 2001). Replacing all six lysine residues in the N-terminus with arginine results in lethality whereas a single lysine at any of these residues is sufficient for viability. Thus, modulation of the charge of these residues and the ability to dissociate the tail from the DNA appears to be critical for cell survival rather than the sequence context in which it occurs. Intriguingly, the residues that are modified in the major histones are absent in H2A-Bbd and it is not clear if this protein is modified. While phosphorylation of H2A.X is important for its function, it is unclear if other modifications affect its role in DNA repair and recombination.

H3 modifications

A study on the modifications associated with histone H3.3 showed that this variant in *Drosophila* Kc cells has the same modification as those associated with the major H3 at active genes (McKittrick et al. 2004). H3.3 was methylated on lysine 4 and 79 and acetylated on lysine 9, 14, and 18. Whether these modifications act to recruit specific proteins to variant nucleosomes, or facilitate the deposition or removal of variant nucleosomes is not known. Given that this variant is almost indistinguishable from the major histone H3 and localizes to active chromatin, it is possible that some of the previously published data on the distribution of histone modifications may have actually analyzed this variant.

In mammalian cells, CenH3 is phosphorylated on serine 7 in vivo and in vitro by the Aurora protein kinases (Zeitlin et al. 2001b; Kunitoku et al. 2003). Ser7 phosphorylation is first observed at prophase and it disappears during anaphase (Zeitlin et al. 2001a; Zeitlin et al. 2001b). While a CenH3 mutant where the Ser7 residue is altered
to alanine or glutamate is still targeted to the centromere, it has a dominant negative phenotype that results in a cytokinesis defect that may be due to altered localization of the "chromosomal passenger complex" that regulates cytokinesis and other mitotic functions (Zeitlin et al. 2001b; Carmena and Earnshaw 2003). In addition, the CenH3 mutant exhibits defects in kinetochore function and chromosome alignment at prometaphase (Kunitoku et al. 2003). that could also be due to altered chromosomal passenger complex activity and/or localization. It is not clear how many other CenH3 phosphorylation sites exist in vivo, and further work will be needed to determine any additional roles for phosphorylation.

Intriguingly, modifications of variants may also play an important role in regulating their localization. The CenH3 variant is polyubiquitinated and degraded by ubiquitin-mediated proteasome-dependent proteolysis in budding yeast (Collins et al. 2004). When yeast CenH3 is stabilized by multiple mutations, it localizes to euchromatin in addition to centromeres. Therefore, one function of CenH3 ubiquitination appears to be to restrict CenH3 localization to centromeres by targeting excess CenH3 for degradation.

**Future Directions**

Studies on the histone variants will continue to shed light on their functions. Although exciting progress has been made, many questions still remain. Do all variant nucleosomes have differences only in surface exposed regions or do alterations in DNA binding regions also exist? What are the effects of variant nucleosomes dispersed amongst major histone containing nucleosomes? How many modifications occur to variants and what are the corresponding functions? How are variants targeted to specific chromosomal regions? What proteins interact with variants? The future promises to answer many of these questions as well as raise new ones.
Acknowledgements

We gratefully acknowledge David Clark, Kim Collins, Orna Cohen-Fix, Suzanne Furuyama, Jim Kadonaga, Harmit Malik, Ben Pinksy, Toshi Tsukiyama and members of the Kamakaka laboratory for reading the manuscript and critical discussions. S.B. is supported by grants from the National Institutes of Health and the Beckman Young Investigator program and RTK is supported by an NIH intramural grant.

Figure Legends

Figure 1. Sequence alignments of specific H2A variants
A Sequence alignment of H2A.X variants and the major H2A subtypes.
B Sequence alignments of H2A.Z variants

The predicted secondary structure of the histones is shown below the sequence and asterisks mark the conserved residues. The conserved SQ motif in the C-terminal tail of H2A.X is shown in red.

Figure 2. Sequence alignments of specific H3 variants
A Sequence alignment of H3.3 variants and the major H3 subtypes.
B Sequence alignments of CenH3 variants

The asterisks indicate the conserved residues and the four amino acid differences between Drosophila H3 and H3.3 are marked in red. The predicted secondary structure of the histones is shown below the sequence.

Figure 3.
A speculative model for the function of H2A.Z in the activation of the GAL1 gene.

H2A.Z containing nucleosomes (red circles) are enriched at the promoter of the
gene when it is inactive and is lost from the promoter upon activation.

Simultaneously H2A.Z facilitates the recruitment of TFIID and RNA pol II as well as the serine2 phosphorylation of the CTD of Pol II.

Figure 4.

A schematic diagram of the different modes of incorporating variant (red) nucleosomes into chromatin.

Figure 5.

CenH3 and kinetochores

CenH3 and H3 nucleosomes exist in linear arrays on the chromatin fiber. The CenH3 nucleosomes must be packaged into a higher order chromatin structure that allows CenH3 to cluster on the outer face of the chromosome. These CenH3 nucleosomes facilitate the binding of proteins that form the kinetochore structure. Microtubules that emanate from the spindle poles capture the kinetochores and lead to chromosome segregation.

Table 1

Histone variants and their functions
References


265-97.


Ren, Q. and M.A. Gorovsky. 2001. Histone H2A.Z acetylation modulates an essential
charge patch. Mol Cell 7: 1329-35.


Replication Coupled Histone Deposition

Transcription Coupled Histone Replacement

Histone Exchange
Linear arrays of H3 and CenH3 nucleosomes are packaged so CenH3 is clustered on the outer face of the chromosome.

Kinetochore proteins assemble onto centromeric nucleosomes.

Microtubules capture kinetochores.
<table>
<thead>
<tr>
<th>VariantSpecies</th>
<th>Species</th>
<th>Chromatin effect</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>H10</td>
<td>Mouse</td>
<td>Chromatin condensation</td>
<td>Transcription repression</td>
</tr>
<tr>
<td>H5</td>
<td>Chicken</td>
<td>Chromatin condensation</td>
<td>Transcription repression</td>
</tr>
<tr>
<td>SpH1</td>
<td>Sea Urchin</td>
<td>Chromatin condensation</td>
<td>Chromatin packaging</td>
</tr>
<tr>
<td>H1t</td>
<td>Mouse</td>
<td>Open chromatin</td>
<td>Histone exchange, recombination?</td>
</tr>
<tr>
<td>MacroH2A</td>
<td>Vertebrate</td>
<td>Condensed chromatin</td>
<td>X-chromosome inactivation</td>
</tr>
<tr>
<td>H2ABbd</td>
<td>Vertebrate</td>
<td>Open chromatin</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H2A.X</td>
<td>Ubiquitous</td>
<td>Condensed chromatin</td>
<td>DNA repair/recombination/ transcription repression</td>
</tr>
<tr>
<td>H2A.Z</td>
<td>Ubiquitous</td>
<td>Open/closed chromatin</td>
<td>Transcription activation/repression, chromosome segregation</td>
</tr>
<tr>
<td>SpH2B</td>
<td>Sea Urchin</td>
<td>Chromatin condensation</td>
<td>Chromatin packaging</td>
</tr>
<tr>
<td>CenH3</td>
<td>Ubiquitous</td>
<td></td>
<td>Kinetochore formation/function</td>
</tr>
<tr>
<td>H3.3</td>
<td>Ubiquitous</td>
<td>Open chromatin</td>
<td>Transcription</td>
</tr>
</tbody>
</table>