Primer

Histones and histone modifications

Craig L. Peterson and Marc-André Laniel

Imagine trying to stuff about 10,000 miles of spaghetti inside a basketball. Then, if that was not difficult enough, attempt to find a unique one inch segment of pasta from the middle of this mess, or try to duplicate, untangle and separate individual strings to opposite ends. This simple analogy illustrates some of the daunting tasks associated with the transcription, repair and replication of the nearly 2 meters of DNA that is packaged into the confines of a tiny eukaryotic nucleus. The solution to each of these problems lies in the assembly of the eukaryotic genome into chromatin, a structural polymer that not only solves the basic packaging problem, but also provides a dynamic platform that controls all DNA-mediated processes within the nucleus.

The basic unit of chromatin is the nucleosome core particle, which contains 147 bp of DNA wrapped nearly twice around an octamer of the core histones. The histone octamer is composed of a central heterotetramer of histones H3 and H4, flanked by two heterodimers of histones H2A and H2B. Each nucleosome is separated by 10–60 bp of 'linker' DNA, and the resulting nucleosomal array constitutes a chromatin fiber of ~10 nm in diameter. This simple 'beads-on-a-string' arrangement is folded into more condensed, ~30 nm thick fibers that are stabilized by binding of a linker histone to each nucleosome core (note that linker histones are not related in sequence to the core histones). Such 30 nm fibers are then further condensed *in vivo* to form 100–400 nm thick interphase fibers or the more highly compacted metaphase chromosome structures. This organization of DNA into chromatin fibers hinders its accessibility to proteins that must 'read' and/or copy the nucleotide base sequence, and consequently such structures must be dynamic and capable of regulated unfolding–folding transitions.

Each of the core histones has a related globular domain that mediates histone–histone interactions within the octamer, and that organizes the two wraps of nucleosomal DNA. Each histone also harbors an amino-terminal 20–35 residue segment that is rich in basic amino acids and extends from the surface of the nucleosome; histone H2A is unique in having an additional ~37 amino acid carboxy-terminal domain that protrudes from the nucleosome. These histone 'tails' do not contribute significantly to the structure of individual nucleosomes nor to their stability, but they do play an essential role in controlling the folding of nucleosomal arrays into higher-order structures. Indeed, *in vitro* removal of the histone tails results in nucleosomal arrays that cannot condense past the beads-on-a-string 10 nm fiber. Although the highly basic histone tails are generally viewed as DNA-binding modules, their essential roles in tail-mediated chromatin folding also involve inter-nucleosomal histone–histone interactions.

Post-translational modifications of histones: encoding or patterning?

Histones are subject to an enormous number of post-translational modifications, including acetylation and methylation of lysines (K) and arginines (R), phosphorylation of serines (S) and threonines (T), ubiquitylation and sumoylation of lysines, as well as ribosylation (Figure 1; Table 1). Adding to the complexity is the fact that each lysine residue can accept one, two or even three methyl groups, and an arginine can be either mono- or di-methylated. The majority of these post-translational marks occur on the amino-terminal and carboxy-terminal histone tail domains, although more and more examples of modifications within the central domains of the histones have been identified (Figure 1). Given the number of new modification sites that are identified each year, it seems likely that nearly every histone residue that is accessible to solvent may be a target for post-translational modification.

Why should histones be the target for so much enzymatic activity? Given that chromatin is the physiological template for all DNA-mediated processes, histone modifications are likely to control the structure and/or function of the chromatin fiber, with different modifications yielding distinct functional consequences. Indeed, recent studies have shown that site-specific combinations of histone modifications correlate well with particular biological functions (see Table 1). For instance, the combination of H4 K8 acetylation, H3 K14 acetylation, and H3 S10 phosphorylation is often associated with transcription. Conversely, tri-methylation of H3 K9 and the lack of H3 and H4 acetylation correlates with transcriptional repression in higher eukaryotes. Particular patterns of histone modifications also correlate with global chromatin dynamics, as diacetylation of histone H4 at K4 and K12 is associated with histone deposition at S phase, and phosphorylation of histone H2A (at S1 and T119) and H3 (at T3, S10 and S28) appear to be hallmarks of condensed mitotic chromatin.

These and other observations have led to the idea of a histone modification 'code' which might be read by various cellular machineries. The term 'code' may be a misnomer, however, as it implies that a particular combination of histone marks will always dictate the same biological function. By analogy, the genetic code is always the same no matter which gene is analyzed, in any cell type or tissue: TAG always means STOP. In the case of histone modifications, however, there are clear exceptions — a particular mark or set of marks

can have different or even opposite biological consequences. For instance, the generally inhibitory H3 K9 methylation can in some cases be associated with actively transcribed genes, and histone acetylation can be inhibitory rather than stimulatory for transcription. Thus, rather than a histone code there are instead clear patterns of histone marks that can be differentially interpreted by cellular factors, depending on the gene being studied and the cellular context.

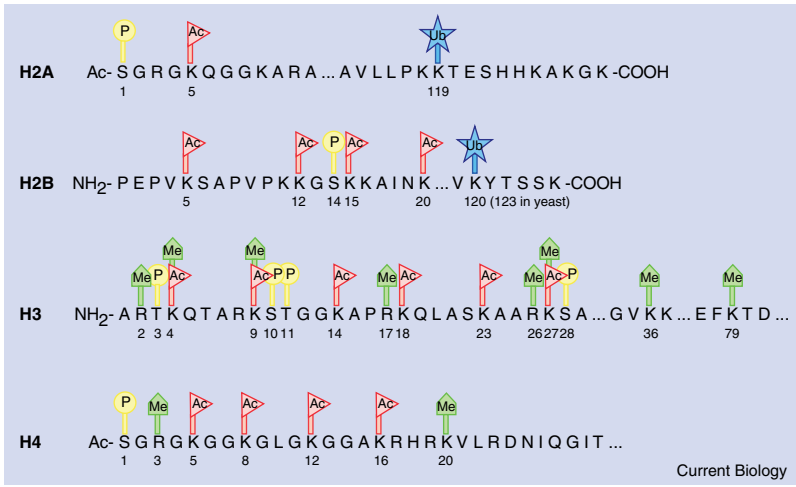

Figure 1. Post-translational modifications of the core histones.
The colored shapes represent known post-translational modifications of the core histones. The histone tails can be methylated at lysines and arginines (green pentagons), phosphorylated at serines or threonines (yellow circles), ubiquitylated (blue stars) and acetylated (red triangles) at lysines.

Patterning chromatin: targeting the enzymes

Although histone modifications have been studied for over 30 years, the identification of the histone modifying enzymes themselves remained elusive until the first nuclear histone acetyltransferase (HAT), a *Tetrahymena* homolog of yeast Gcn5, was identified in 1996. *In vivo* studies in yeast had previously characterized Gcn5 as a transcriptional co-activator protein, and thus its identification as a HAT solidified the view that histone modifications directly regulate transcription. Subsequently, a variety of other transcriptional co-activators, such as CBP/p300 were found to have intrinsic HAT activity, and many co-repressors, such as Rpd3, were found to have histone deacetylation (HDAC) activity. Histone modification enzymes are now organized into large HAT, HDAC, histone methyltransferase (HMT) and histone kinase families (see Table 1).

The precise combination of locus-specific histone modifications is due to the combined effects of targeting histone modifying enzymes to specific loci, as well as to the inherent substrate specificity of the enzymes themselves. In the case of transcription, it is clear that targeting of histone modifications is achieved by direct interactions between histone modifying enzymes and DNA sequence-specific transcriptional regulators. For instance, the yeast HAT complex SAGA interacts with the transcriptional activation domains of a variety of yeast gene-specific activator proteins, and these interactions target HAT activity to specific promoter regions *in vivo*. Likewise, unliganded nuclear hormone receptors interact with HDAC complexes, such as NCoR and SMRT, which direct histone deacetylase activity to target genes and contribute to subsequent gene repression. In addition to targeting via gene-specific regulators, the yeast Set1 and Set2 HMTs are found associated with RNA polymerase II holoenzymes, directing histone H3 K4 or K36 methylation, respectively, during transcriptional elongation.

Targeting histone modifications is not unique to transcriptional control, as DNA repair and centromeric heterochromatin use distinct mechanisms to generate novel patterns of histone marks. In the case of DNA repair, the DNA lesion itself seems to play a central role in targeting histone modifications. For instance, the DNA damage checkpoint kinase ATM (Mec1p in yeast) is recruited to a DNA double strand break where it phosphorylates histone H2A (in yeast) or the histone H2A variant, H2AX (in mammals). Likewise, the human STAGA HAT complex contains DNA binding subunits that recognize DNA backbone-distorting base adducts, targeting histone H3 acetylation activity to sites of nucleotide excision repair.

A quite different strategy uses small noncoding RNAs to target histone H3 K9 methylation to chromatin surrounding mammalian and fission yeast centromeres. These centromeric regions are characterized by repetitive DNA sequences that are transcribed at low levels. The resulting double-stranded RNAs provide substrates for processing by the RNA interference (RNAi) machinery which produces small, 21–23 nucleotide RNAs. Recent studies have shown that an intact RNAi pathway is essential for targeting H3 K9 methylation to centromeric chromatin, and furthermore that these small RNAs actually associate with several chromatin components. The resulting novel ribonucleoprotein complex ultimately targets the Clr4p HMT to centromeric repeats, via either RNA–RNA (nascent centromeric transcripts) or RNA–DNA homologous pairing. Subsequent histone methylation leads to recruitment of proteins such as Heterochromatin Protein 1 (HP1), which directs formation of highly condensed, heterochromatin structures required for centromere function.

Table 1. A current view of histone modifications.

Modification	Histone	Site	Enzyme	Possible function
Acetylation	H2A	K4 (*S. cerevisiae*)	Esa1	Transcriptional activation
		K5 (mammals)	Tip60	Transcriptional activation
			p300/CBP	Transcriptional activation
		K7 (*S. cerevisiae*)	Hat1	?
			Esa1	Transcriptional activation
	H2B	K5	ATF2	Transcriptional activation
		K11 (*S. cerevisiae*)	Gcn5	Transcriptional activation
		K12 (mammals)	p300/CBP	Transcriptional activation
			ATF2	Transcriptional activation
		K16 (*S. cerevisiae*)	Gcn5	Transcriptional activation
			Esa1	
		K15 (mammals)	p300/CBP	
			ATF2	Transcriptional activation
		K20	p300	Transcriptional activation
	H3	K4	Esa1	Transcriptional activation
			Hpa2	?
		K9	?	Histone deposition
			Gcn5	Transcriptional activation
			SRC-1	Transcriptional activation
		K14	Gcn5, PCAF	Transcriptional activation
			Esa1, Tip60	Transcriptional activation
				DNA repair
			SRC-1	Transcriptional activation
			Elp3	Transcription elongation
			Hpa2	?
			hTFIIIC90	RNA polymerase III transcription
			TAF1	RNA polymerase II transcription
			Sas2	Euchromatin?
			Sas3	Transcriptional activation/elongation?
			p300	Transcriptional activation
		K18	Gcn5 (SAGA/STAGA complex)	Transcriptional activation
				DNA repair
			p300, CBP	DNA replication
				Transcriptional activation
		K23	Gcn5 (SAGA/STAGA complex)	Transcriptional activation
			Sas3	DNA repair
			p300, CBP	Transcriptional activation/elongation?
				Transcriptional activation
		K27	Gcn5	Transcriptional activation
	H4	K5	Hat1	Histone deposition
			Esa1, Tip60	Transcriptional activation
				DNA repair
			ATF2	Transcriptional activation
			Hpa2	?
			p300	Transcriptional activation
		K8	Gcn5, PCAF	Transcriptional activation
			Esa1, Tip60	Transcriptional activation
				DNA repair
			ATF2	Transcriptional activation
			Elp3	Transcription elongation
			p300	Transcriptional activation
		K12	Hat1	Histone deposition
				Telomeric silencing
			Esa1, Tip60	Transcriptional activation
				DNA repair

Table 1. A current view of histone modifications (continued).

Modification	Histone	Site	Enzyme	Possible Function
		K12	Hpa2	?
		K16	Gcn5	Transcriptional activation
			MOF (*D. melanogaster*)	Transcriptional activation
				Transcriptional activation
			Esa1 (yeast), Tip60 (mammals)	DNA repair
			ATF2	Transcriptional activation
			Sas2	Euchromatin
Methylation	H3	K4	Set1 (yeast)	Permissive euchromatin (di-Me)
			Set9 (vertebrates)	Active euchromatin (tri-Me)
				Transcriptional elongation/memory (tri-Me)
				Transcriptional activation
			MLL, Trx	Transcriptional activation
			Ash1 (*D. melanogaster*)	Transcriptional activation
		K9	Suv39h, Clr4	Transcriptional silencing (tri-Me)
				DNA methylation (tri-Me)
			G9a	Transcriptional repression
				Imprinting
			SETDB1	Transcriptional repression (tri-Me)
			Dim-5, Kryptonite	DNA methylation (tri-Me)
			Ash1 (*D. melanogaster*)	Transcriptional activation
		R17	CARM1	Transcriptional activation
		K27	Ezh2	Transcriptional silencing
				X inactivation (tri-Me)
		K36	Set2	Transcriptional elongation
				Transcriptional repression?
		K79	Dot1p	Euchromatin
				Transcriptional elongation / memory
	H4	R3	PRMT1	Transcriptional activation
		K20	PR-Set7	Transcriptional silencing (mono-Me)
			Suv4-20h	Heterochromatin (tri-Me)
			Ash1 (*D. melanogaster*)	Transcriptional activation
		K59	?	Transcriptional silencing?
Phosphorylation	H2A	S1	?	Mitosis
			?	Chromatin assembly?
			MSK1	Transcriptional repression
		T119	NHK-1	Mitosis
		S129 (S. cerevisiae)	Mec1	DNA repair
		S139 (mammalian H2AX)	ATR, ATM, DNA-PK	DNA repair
	H2B	S14 (vertebrates)	Mst1	Apoptosis
		S33 (D. melanogaster)	TAF1	Transcriptional activation
	H3	T3	?	Mitosis
		S10	Aurora-B kinase	Mitosis, meiosis
			MSK1, MSK2	Immediate-early activation
			Snf1	Transcriptional activation
		T11 (mammals)	Dlk/ZIP	Mitosis
		S28 (mammals)	Aurora-B kinase?	Mitosis
			MSK1, MSK2	Immediate-early activation
	H4	S1	?	Mitosis
Ubiquitylation	H2A	K119 (mammals)	HR6A,B?	Spermatogenesis
	H2B	K120 (mammals)	HR6A,B?	Meiosis
		K123 (S. cerevisiae)	Rad6	Transcriptional activation
				Euchromatin
	H3	?	?	Spermatogenesis
Sumoylation	H4	?	Ubc9	Transcriptional repression

Overlaid on top of these locus-specific marks is the genome-wide, bulk chromatin modifications that may control the day-to-day folding dynamics of chromosomes. For instance, newly synthesized histones that are deposited after passage of replication forks in S phase are enriched in acetylated isoforms of histones H3 and H4, and the formation of condensed chromosomes in mitosis is associated with phosphorylation of histones H3 and H2A.

In addition to these marks linked to the cell cycle, there appears to be a constant battle among HATs and HDACs on a global, nontargeted level that maintains a baseline, equilibrium level of histone acetylation throughout the genome. Histone deacetylase inhibitors, such as trichostatin or sodium butyrate, disrupt this equilibrium, leading to a general increase in bulk histone acetylation. Such genome-wide activities of histone modifying enzymes likely act in concert with the cell-cycle-linked changes in bulk chromatin to enhance the general dynamic nature of eukaryotic chromosomes.

Patterning chromatin: controlling enzyme substrate specificity

Recruitment of a histone modifying enzyme to the right place at the right time is only the first step in establishing a combination of histone marks that may direct a biological outcome. The second step in this process revolves around the specificity of the enzyme for individual histone tails and for specific histone residues (Table 1). For example, yeast and human Gcn5 and human PCAF preferentially acetylate lysine residues within the histone H3 amino-terminal tail, at K9 and K14. In contrast, the yeast and human NuA4 HAT complexes preferentially acetylate K4, K8, K12 and K16 of histone H4. Even more extreme specificity is seen with HMTs. For instance, the HMT Set7/9 is restricted to mono-methylation of histone H3 at K4, whereas the Dim-5 HMT is a tri-methylase specific for H3 K9. Thus, recruitment of different HATs or HMTs can result in distinct combinations of histone modifications.

Cross-talk among different histone marks can also have a profound effect on enzyme activity. For instance, ubiquitylation of H2B K123 by the E2 ubiquitin conjugating enzyme Rad6 is required for subsequent di-methylation of H3 K4 by Set1p or H3 K79 by Dot1p. Prior histone marks can also inhibit subsequent modifications. For instance, H3 S10 phosphorylation inhibits subsequent H3 K9 methylation, and of course H3 K9 methylation can also block acetylation of this same residue. An excellent example of even more complex cross-talk is exemplified during p53-dependent transcriptional activation *in vitro*. In this case methylation of H4 R3 by protein arginine methyltransferase 1 (PRMT1) stimulates CBP-p300 acetylation of H4 K5, K8, K12 and K16, which in turn promotes the methylation of H3 R2, R17 and R26 by another PRMT family member, CARM1. Thus, positive and negative crosstalk ultimately generates the complex patterns of gene or locus-specific histone marks associated with distinct chromatin states.

Patterns of histone modifications: what happens next?

Once a pattern of histone modifications is established at a target locus, what do they do? Many older models proposed that histone modifications might directly influence either the structure of individual nucleosomes or the folding dynamics of nucleosomal arrays. Indeed one common misconception is that histone modifications that alter the charge of a residue, such as lysine acetylation or serine phosphorylation, will disrupt histone–DNA interactions leading to ‘open’ or ‘active’ chromatin structures. There is not actually much evidence for such models. For example, the histone H3 tail contains 13 positively charged amino acids, and thus acetylation of one to four residues will only yield a 10–30% decrease in positive charge, levels that are unlikely to perturb ionic interactions with DNA.

Consistent with this view, *in vivo* laser crosslinking studies have shown that histone hyperacetylation does not release tails from DNA, and nucleosomes that harbor >12 acetates per octamer wrap DNA normally *in vitro* and have hydrodynamic properties that are nearly identical to unmodified nucleosomes. Although it is true that histone hyperacetylation does disrupt the folding dynamics of nucleosomal arrays *in vitro*, even in this case 6–12 acetates per nucleosome are required. Although most site-specific patterns of histone modifications have yet to be generated and tested *in vitro*, the prevailing view is that these histone marks may not alter nucleosomal dynamics by themselves.

In contrast to the lack of evidence pointing to direct changes in chromatin structure, there is now a wealth of examples where specific histone modifications control the binding of nonhistone proteins to the chromatin fiber. These nonhistone proteins then elicit the function that is associated with a particular histone mark. A hallmark of many proteins that bind to histone tails is the presence of small histone binding modules. For example, some chromodomains bind to methylated lysines, whereas bromodomains specify binding to acetylated lysines. Furthermore, these modules often bind to only a particular modified histone residue. For example, the chromodomain within HP1 interacts specifically with a dimethylated K9 of histone H3, whereas the chromodomain of the Polycomb protein binds to a dimethylated K27 of histone H3. In contrast, the binding of bromodomains to different acetylated lysines does not show as much specificity. For instance, acetylation of K8 within histone H4 can promote the recruitment of the ATP-dependent chromatin remodeling enzyme, human SWI/SNF — via a bromodomain within the Brg1 subunit — but a similar bromodomain within the Swi2 subunit of the yeast

SWI/SNF complex interacts with a broader range of acetylated H3 and H4 tails.

The interactions of bromodomains and chromodomains with modified tails is also subject to modification crosstalk — the modification of adjacent residues can positively or negatively regulate binding. Thus, in many ways histone tails can be viewed as complex protein–protein interaction surfaces that are regulated by numerous post-translational modifications. Furthermore, it is clear that the overall constellation of proteins bound to each tail plays a primary role in dictating the biological functions of that chromatin domain.

Chromatin fiber heterogeneity: histone variants

Throughout this primer we have described chromatin (in its simplest form) as a linear array of canonical nucleosomes. An *in vivo* chromatin fiber, however, is actually an extremely heterogeneous nucleoprotein filament, even at the nucleosome level. First and foremost, in addition to canonical nucleosomes, *in vivo* chromatin arrays also contain novel types of nucleosome that harbor one or more variant isoforms of the core histones. For instance, nucleosomes assembled at yeast and mammalian centromeres contain a histone H3 variant, Cse4/CENP-A, which is essential for centromere function or assembly. Another histone H3 variant, H3.3, replaces canonical histone H3 during transcription, generating a mark of the transcription event. Several variants of histone H2A have also been identified. The macro-H2A variant is restricted to metazoans and functions in X chromosome inactivation, while H2AZ (also known as H2A.F/Z or H2AvD) is found in all eukaryotes. Surprisingly, H2AZ is required for one or more essential roles in chromatin structure that cannot be replaced by *bona fide* histone H2A.

In most cases, how histone variants alter nucleosome structure or change the folding properties of nucleosomal arrays is not known. It is also not clear how many of these variant nucleosomes are localized to specific DNA sequences; for example, why are CENP-A-containing nucleosomes found only at centromeres? Notable exceptions include the deposition of H3.3 to chromatin of RNA polymerase II transcribed genes via a novel replication-independent assembly complex, and the exchange of H2AZ for canonical H2A via the ATP-dependent SWR1 complex.

Once a histone variant is targeted to a specific locus, there is the potential for creation of novel chromatin domains that have distinct regulatory properties. For instance, the amino-terminal tail of CENP-A lacks the phosphorylation and acetylation sites that are normally modified in histone H3 at transcriptionally active regions. Thus, CENP-A might produce islands of unmodified histone H3 that help to maintain centromeric chromatin in its condensed, inactive state. In contrast, the histone H3.3 variant contains an amino-terminal tail that is virtually identical to that of histone H3, and thus it seems likely that many of the transcription-associated marks that have been attributed to histone H3 are likely also occurring on the histone H3.3 variant.

In the case of H2AZ, biochemical studies suggest that nucleosomal arrays containing H2AZ may only partially compact, resisting formation of large 100–400 nm fibers and thereby facilitating transcription. Thus, incorporation of histone variants into chromatin fibers might enhance chromosome dynamics by creating domains of chromatin with novel properties.

Concluding remarks

Histone variants, distinct patterns of posttranslational modifications of histones, and histone tail binding proteins all contribute to establishment of various 'open' or 'closed' chromatin domains that have specialized folding properties and biological functions. Some of these domains can be propagated through DNA replication and mitosis, guaranteeing the inheritance of chromatin states to progeny. Histone lysine methylation may play a central role in the stability of these chromatin states, as to date no enzymes are known that catalyze lysine demethylation. Furthermore, several nonhistone proteins, such as HP1 or the PRC1 polycomb complex, not only bind to methylated histone lysines, but also recruit the methylase itself, thus providing a means for templating new histone methylation events — for example, following replication fork passage — or for spreading the domain to adjacent nucleosomes.

How 'open' states are propagated through cell divisions is not clear, especially as histone lysine acetylation or serine phosphorylation can be rapidly reversed by HDACs or histone phosphatases. Future studies will no doubt continue to identify the functional and biochemical properties of new chromatin domains as well as to elucidate the principles that govern their maintenance and propagation.

Further reading

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

Elgin, S.C.R., and Grewal, S.I.S. (2003). Heterochromatin: silence is golden. Curr. Biol. *13*, R895–R898.

Fischle, W., Wang, Y., and Allis, C.D. (2003). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475–479.

Fischle, W., Wang, Y., and Allis, C.D. (2003). Histone and chromatin cross-talk. Curr. Opin. Cell Biol. *15*, 172–183.

Korber, P., and Horz, W. (2004). SWRred not shaken; mixing the histones. Cell *117*, 5–7.

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

Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
E-mail: craig.Peterson@umassmed.edu