An increasingly complex code
Hugh T. Spotswood and Bryan M. Turner

Chromatin and Gene Expression Group, Anatomy Department, University of Birmingham Medical School, Birmingham, United Kingdom

The nucleosome core particle is the fundamental unit of chromatin structure in all eukaryotes. It comprises eight core histones (two each of H2A, H2B, H3, and H4) around which are wrapped 146 bp of DNA, and its structure has been defined at 2.8 Å resolution by x-ray crystallography (1). Since it was first described, it has been assumed that the function of the core particle is to package DNA into the interphase cell nucleus and metaphase chromosomes. Wrapping DNA around the histone octamer in left-handed supercoils results in an approximately sevenfold reduction in its length. Although this is only a small fraction of the several-thousand-fold length reduction required for compaction into metaphase chromosomes, it is presumably an essential first step that enables higher-order structures to assemble.

The nucleosome’s rather mundane packaging job became more interesting with the realization that compaction of DNA into chromatin is a crucial element in eukaryotic gene regulation (2). Appropriately located nucleosomes can hinder assembly of transcription initiation complexes and polymerase progression, and over recent years, whole families of chromatin-remodeling enzymes have been identified whose primary purpose is to reorganize histone-DNA interactions so as to facilitate (or repress) transcription (3). These enzymes disrupt the nucleosome in an ATP-dependent manner that often results in movement of the histone core relative to the DNA, thereby exposing recognition sequences for DNA-binding activators or repressors. However, even these insights still left the nucleosome as a passive inhibitor of transcription, something to be moved aside so transcription could proceed.

The final transformation of the nucleosome into a truly exciting and colorful character has come through studies on the effects and properties of other enzymes that act on chromatin. It has been known for many years that the histone N-terminal tails are exposed on the surface of the nucleosome and that selected amino acid residues are subject to a variety of enzyme-catalyzed, posttranslational modifications. These include acetylation of lysines, phosphorylation of serines, and methylation of lysines and arginines. The locations of the histone N-terminal tails in the nucleosome and the residues that can be modified are shown in Figure 1.

The function of these modifications, and indeed of the tails themselves, is the focus of much current attention, centered largely on the possibility that the nucleosome, with its modified tail domains, is not just a humble packer of DNA, but a carrier of epigenetic information that determines both how genes are expressed and how their expression patterns are maintained from one cell generation to the next.

An epigenetic code
The first association between a histone tail modification and a particular functional state of chromatin came with the demonstration that transcriptionally active chromatin fractions are enriched in acetylated histones (4, 5). Subsequently, regions of transcriptionally silent constitutive and facultative heterochromatin were found to be underacetylated (6), consistent with the idea that acetylation of the histone tails, with the associated loss of positive charge and weaker DNA binding, somehow caused chromatin to become more “open” (or less “condensed”) and thereby more conducive to transcription. While such charge-mediated structural changes are likely to contribute to altered patterns of gene expression, it has now become clear that the story is much more complex and that the functional effects of tail modifications depend on the specific amino acids that are modified.

An early indication that modification of specific tail residues was linked to chromatin functional states came from immunostaining of Drosophila polytene chromosomes. Unlike in mammals, where female cells inactivate one of their two X chromosomes to avoid unbalanced expression of X-linked genes, Drosophila equalizes the levels of X-linked gene products between XY males and XX females by transcribing X-linked genes in males at twice the rate seen in females. Analyses using residuespecific antibodies show that H4 acetylated at lysine 16 (H4Ac16) occurs almost exclusively on the transcriptionally hyperactive male X chromosome. In addition, the Drosophila H4 lysine 12 remains acetylated in centric heterochromatin, whereas lysines 5, 8, and 16 are all underacetylated (7). These observations led to the suggestion that the histone N-terminal tails constitute nucleosome surface markers that can be recognized by nonhistone proteins in a modification-dependent manner to alter the functional state of chromatin (7). More
importantly, these observations also raised the possibility that tail modifications not only contribute to the ongoing up- and downregulation of transcription but also carry information regarding the potential for transcription (or lack of it) in response to specific factors. In other words, the nucleosome may have a role not only in DNA packaging, but also in the transmission from one cell generation to the next of epigenetic information regarding transcriptional potential (8, 9).

The amount of epigenetic information that can be carried in the histone tails is remarkable. For example, there are 50 different acetylated isoforms of the four core histones (H2B, H3, and H4) that have been identified (H2A has two). These isoforms can be modified further by methylation of selected lysines and arginines (H3 and H4) and phosphorylation of serine (H3, H4, H2B). Furthermore, methylation can involve attachment of one, two, or three methyl groups (see below), and there are other modifications, such as ubiquitination and ADP-ribosylation (10). The total number of possible histone isoforms, carrying different combinations of tail modifications, that can mark the nucleosome surface, runs into many thousands. This has given rise to the idea that the tail modifications constitute a histone code (11, 12) or epigenetic code (13), which is set and maintained by tail-modifying and -demodifying enzymes and read by nonhistone proteins.

The histone tail modifications are likely to act in concert with a rather more widely known mediator of chromatin structure and gene expression, namely methylation of cytosine residues in CpG dimers through the action of DNA methyltransferases. Long-term silencing, as found in imprinted genes or the female inactive X chromosome, is generally associated with relatively high levels of CpG methylation. The mechanism(s) by which CpG methylation leads to gene silencing remains to be defined, but there is evidence that in at least some situations, histone modifications are involved. The methyl DNA-binding protein MeCP2 can bind histone deacetylases, thereby targeting them to methylated DNA and resulting in local histone deacetylation and suppression of transcription (14). Conversely, experiments in the filamentous fungus Neurospora crassa have shown that DNA methylation is dependent upon methylation of histone H3 lysine 9 (15).

**Dynamics of histone modification**

The modifications to the N-terminal histone tails in chromatin are put in place and maintained by the action of families of modifying and demodifying enzymes. For example, most histone acetates are undergoing a continuous cycle of acetylation and deacetylation and turn over with halflives ranging from just a few minutes to several hours (16). Thus, the level of acetylation of a particular lysine in a particular histone is likely to be a steady state maintained by the ongoing activities of histone acetyltransferases (HATs) and deacetylases (HDACs). While this makes continuous demands on the cell’s energy resources, it also offers an advantage to those genes whose expression may need to be up- or downregulated relatively quickly, such as those that respond to hormonal cues or growth factor stimulation. Removal or inhibition of either HATs or HDACs will precipitate a rapid change in acetylation, and hence transcription or transcriptional potential. Conversely, where levels of gene expression are to be stably maintained over long periods, patterns of histone modification are stabilized by removing (or inhibiting) the HATs and HDACs targeted to that genomic region. In the case of the imprinted gene U2af1-rs1 in mice, differential acetylation levels on the maternal and paternal alleles are maintained dynamically early in development but become static (operationally defined by resistance to HDAC inhibitors) as development proceeds (17).

Lysine methylation seems to be an exceptionally stable modification. Early studies showed that turnover of histone methyl groups was even slower than turnover of the histones themselves (see, for example, ref. 18). No conclusive evidence has yet been found for histone-demethylating enzymes, which may indeed not exist. Removal of methylated histones may simply occur passively through postreplication chromatin assembly and replacement of old, methylated histones with new, unmethylated ones. However, the possibility remains that local methylation patterns may be more dynamic and may involve novel mechanisms for removal of methylated tails, such as proteolysis (19) or histone replacement uncoupled from DNA replication (20).

**Setting the code**

Among the various tail modifications, setting patterns of histone phosphorylation might seem relatively...
straightforward in that only a few residues are typically involved (Figure 1). However, the same modification has been implicated in two different functions. Phosphorylation of H3S10, precisely targeted to control regions, accompanies activation of immediate-early genes, while a genome-wide phosphorylation of the same residue occurs just prior to entry into mitosis (21, 22). Different kinases are involved in each case (22). At first sight, it seems odd that the same modification is associated with both progression through mitosis (a time of chromatin condensation and transcriptional silence) and gene induction. Perhaps H3S10 phosphorylation is involved in the chromatin decondensation and reinitiation of transcription that accompany progression from mitosis into G1.

In theory, the setting of specific patterns of acetylation (i.e., residue-specific modifications) could reflect the specificities of HATs or HDACs. Both HATs and HDACs constitute extensive enzyme families and are often found as a catalytic subunit in multiprotein complexes whose other components confer genomic targeting and other capabilities (23, 24). In vitro assays have shown specificity, or at least substrate preferences, for most HATs. The most extreme example is provided by the Drosophila enzyme MOF, which acetylates H4 specifically at lysine 16 and is responsible for the acetylation of this residue exclusively on the male X chromosome (ref. 25 and references therein). The HDACs so far tested show less specificity for individual histones or residues, though there are certainly preferences (ref. 26 and references therein). It is, of course, difficult to be sure that in vitro assays provide a true reflection of specificity in vivo. Both HATs and HDACs can show quite different specificities, or levels of catalytic activity, depending on whether they are assayed against peptide, whole histone, or chromatin substrates. Linking in vitro specificity to in vivo effects, as in the case of MOF, is reassuring.

Histone methyltransferases (HMTs) are specific for either lysine or arginine and are the subject of two excellent recent reviews (27, 28). There are five arginine HMTs (four of which are confirmed as enzymes). All have the same, highly conserved catalytic domain and S-adenosyl methionine–binding region, but little else in the way of known protein domains. The enzymes studied so far show varying degrees of specificity for particular histone arginines. All can attach one or two methyl groups to a single arginine, but they differ in whether the two methyl groups are distributed symmetrically (i.e., one on each of the two terminal nitrogens) or asymmetrically (i.e., both on the same nitrogen) (27).

Lysine HMTs all contain a SET domain, although not all SET domain proteins have HMT activity (29). Database searches reveal 73 SET domain proteins in humans, of which about one-third have been characterized and grouped into four families on the basis of sequence homology and relationship to the SET domain proteins in yeast (28). Several have confirmed HMT activity, and some are highly specific. For example, four members of the SUV39 family methylate only H3 lysine 9 (30–32). Importantly, this specificity has been linked to a particular chromatin state, namely, formation of heterochromatin.

The mammalian enzymes SUV39H1 and SUV39H2 have been shown to locate to constitutive heterochromatin, which also, by immunofluorescent labeling, appears to be relatively rich in H3 methylated at lysine 9 (33, 34). The protein HP-1 is also enriched in heterochromatin, where it plays a crucial, although as-yet unspecified, role in heterochromatin assembly. HP-1 has recently been shown to bind, through its chromodomian, specifically to H3 peptides methylated at lysine 9 (35, 36). This is the first example in which the enzymes that put a specific modification in place, the protein that reads it, and the functional effects that result are all known, at least in outline. As always, there are complications, as discussed below, and much remains to be worked out. In particular, we need to understand how the appropriate HMTs are targeted to defined genomic regions. Lysine HMTs are rich in known protein domains (28), some of which may have a targeting function. Intriguingly, some SET domain proteins also have a methyl DNA–binding domain, which would be predicted to target them to methylated CpG, and which may play a role in establishing the transcriptionally silent state usually associated with DNA methylation.

Modifications can interact, both in determining whether or not a particular residue can be modified and in providing a meaningful code. For example, H3 tail peptides phosphorylated at S10 are not methylated at lysine 9 by SUV39H1 in vitro (29). Conversely, peptides methylated at lysine 9 are poor substrates for phosphorylation of serine 10 by the Ipl1/aurora kinase. Unsurprisingly, acetylation of lysine 9 prevents methylation of the same residue by SUV39H1 (29), but it is interesting that the homologous enzyme in Drosophila has been shown to exist in a protein complex containing dHDAC1 (37). The deacetylase activity in the complex may serve both to clear away any potentially inhibitory acettes and to create a globally underacetylated chromatin context in which the H3 methylation signal can be most appropriately read.

Attention to detail and creating a context

The lysine ε-amino group can be mono-, di-, or trimethylated, and all three methylation levels occur at high frequency in vivo. In mammalian cultured cells, the ratio of mono-, di-, and trimethylated lysines in H3 has been estimated as 1:2:1, with the overall methylation of H3 lysines 4 and 9 being close to 100% (18). Whether different enzymes are responsible for putting the different levels of methylation in place remains to be determined, but it is clear from recent results that these extremely detailed differences are significant in vivo. Studies in yeast using antisera that distinguish between H3 dimethylated and H3 trimethylated at lysine 4 show that only the trimethylated isoform, not its dimethylated counterpart, is consistently associated with transcriptionally active chromatin (38).

One of the most consistent and widespread associations between histone modification and chromatin structure is the general underacetylation of all four core histones in constitutive and facultative heterochro-
The inactive X retains H3 dimethylated at lysine 9 but lacks H3 dimethylated at lysine 4 (40) but retains H3 dimethylated at lysine 9 at levels at or above those found elsewhere in the genome (41). Thus, Xi exemplifies both extremes of the histone code, namely the general deacetylation of all four core histones and the specific retention of H3 dimethylated at lysine 9. It also illustrates the great stability of epigenetic signals carried by the histone tails. The general underacetylation of Xi, the dimethylation of H3 lysine 4, and the selective retention of H3 dimethylated at lysine 9 are all maintained in human x hamster somatic cell hybrids, in which Xi is maintained in an almost completely foreign cellular environment (Figure 2).

It may be that the general deacetylation of all four histones provides a molecular or structural context in which the specific methylation signal can be most effectively read. Perhaps deacetylated tails encourage binding of a protein that recognizes H3 dimethylated at lysine 9, just as selective deacetylation of some H4 tail lysines encourages binding of the proteins Sir2 and Sir3 that assemble heterochromatin-like structures in yeast (42). This hypothetical protein remains to be identified. It seems not to be HP-1, which has not been detected on Xi (43). The fact that HP-1 binds to constitutive centric heterochromatin, but not to Xi, even though both are underacetylated and retain H3 dimethylated at K9, suggests that there is an additional determinant of HP-1 binding.

**Reading the code**

Antibodies can distinguish modified tails, combinations of modifications, and even lysines carrying two or three methyl groups. Other proteins appear to make similar discriminations. The clearest example may be HP-1’s ability to bind specifically, but with different affinities, to H3 peptides di- or trimethylated at lysine 9 (35, 36). Also, certain bromodomains (sequence elements found in proteins associated with maintenance of transcriptionally active states) bind preferentially to acetylated tail peptides in vitro (44, 45). Recent results have added a new twist to the question of how tail modifications might be read by implicating the histone tails in targeting of chromatin-remodeling enzymes. Firstly, immobilized H3 tail peptides lacking lysine methylation can remove components of the chromatin remodeling complex NuRD from crude cell extracts in pull-down assays. Remarkably, peptides methylated at lysine 4 lose this NuRD-binding ability, while peptides methylated at lysine 9 do not (46). It seems that binding of the NuRD complex to the nucleosome requires a specific region of the H3 tail and is methylation-sensitive.

Secondly, a genetic approach has provided evidence that the selective accumulation of H4 acetylated at lysine 16 on the *Drosophila* male X chromosome counteracts the condensing effect of the chromatin-remodeling ATPase ISW1 (47). Related work shows that the site on chromatin recognized by ISW1 consists of the H4 tail bound to DNA, and that acetylation of H4 at lysine 16 reduces the ability of ISW1 to interact productively with its substrate (47, 48). Together, these results provide a plausible mechanism by which H4 acetylated at lysine 16 can lead to decondensation and transcriptional upregulation of the *Drosophila* male X chromosome. The results also provide an important example of how signals on the nucleosome surface, read by nonhistone proteins, may require a combination of both DNA and the appropriately modified histone tail.

To take full advantage of the information potential of histone tail modifications, the histone code must encompass combinations of modifications. Ideally, combinations will involve not only modifications on the same tail, but those on different tails as well. Recent studies in yeast have provided a dramatic example of how modification of one histone tail can influence that of another and have brought an additional player, the 76–amino acid peptide ubiquitin, into the limelight.
Ubiquitin can be attached to specific lysines in the C-terminal tails of H2A (lysine 119) and H2B (lysine 120) in higher eukaryotes, and to H2B lysine 123 alone in yeast (49). By a combination of genetic and biochemical assays, Sun and Allis (50) have shown that yeast mutants that cannot ubiquitination H2B show no detectable methylation of lysine 4 in the N-terminal tail of H3. This is true whether the ubiquitination defect is due to mutation of the enzyme responsible (RAD6) or to substitution of lysine 123 of H2B with arginine. RAD6 is involved in many cellular processes, including DNA repair and gene silencing. The mechanism that links H2B ubiquitination and H3 methylation remains uncertain, though one can hazard a guess that ubiquitination somehow enhances the ability of the SET1 histone methylase to access its chromation substrate. It also remains to be determined whether this, or similar, interactions occur in higher eukaryotes.

Manipulating the code: new routes to therapy
Like all enzymes, the enzyme families that modify core histones or reorganize chromatin structure are susceptible to inhibition by appropriately designed reagents. Since these enzymes regulate patterns of gene expression by targeting selected genes or chromosome regions, the therapeutic potential of such inhibitors is enormous, particularly for treating cancers or other diseases showing aberrant patterns of gene expression. The enzymes most closely studied to date are the HDACs, for which a wide range of inhibitors is now available. Some of these agents have proved to be remarkably effective at selectively inhibiting the growth of human tumor cells, and several are now in clinical trials against a variety of tumor types (51). Unfortunately, we know rather little about the in vivo roles of the various deacetylases and their inhibitor sensitivities. The different members of the deacetylase family are likely to have very different functions in vivo, which provides a strong incentive to identify more selective inhibitors. Inhibitors specific for the NAD-dependent deacetylase Sir2 have recently been identified from screens of relatively small numbers of compounds using an in vivo assay (52). Initial experiments showed no increase in histone acetylation in cultured cells treated with these inhibitors, suggesting that acetylated histones are not a major substrate of the Sir2 enzyme family in vivo (52).

When considering therapeutic approaches, it is worth bearing in mind that factors other than enzyme inhibitors (natural or synthetic) can influence patterns of histone modification in vivo. The situation for histone acetylation is summarized in Figure 3. Metabolic factors include intranuclear levels of acetyl-CoA for HAT activity, NAD for Sir2 deacetylases (ref. 52 and references therein), and ATP for deacetylation of chromatin substrates by at least some HDACs (26). The HDAC inhibitor butyric acid occurs in the large intestine in millimolar concentrations that may be influenced by diet (53). HDAC activity can also be influenced indirectly by targeting of proteins that modify or interact with the HDACs themselves. For example, HDACs are phosphorylated in vivo by protein kinases such as CK2 in ways that influence both their catalytic activity and their ability to interact with other proteins in functional complexes (54, 55). This links HDAC function to intracellular signaling pathways. Also, the stress response (heat shock) protein HSP70 has recently been shown to be present in HDAC complexes, where it may play a role in the ATP-dependent deacetylation of chromatin substrates (22). Levels of HSP70 and other stress response proteins rise in response to a variety of toxins and other adverse environmental conditions, and it will be of interest to determine to what extent this influences HDAC activity in vivo. Finally, patterns of acetylation are inevitably influenced by, and exert an effect on, other histone tail modifications. These interactions, some of which are discussed above, will make it difficult to predict the functional effects of even highly specific enzyme inhibitors.

Pressing questions
With new results appearing at such a pace, prediction is a risky business. However, there remain some fairly obvious gaps in our knowledge. Firstly, how is histone methylation removed? Is there a histone demethylase, or are methylation levels regulated by histone replacement or even removal of N-terminal tails by proteolysis? What is the significance of H3 isoforms mono- or di-methylated at lysines 4 and 9, and how do we reconcile the frequency and stability of histone methylation with a regulatory role in gene expression? We also need more information on the relatively neglected histones H2A and H2B: The importance of H2B ubiquitination has recently been noted, and H2B phosphorylation has been associated with the chromatin reorganization that accompanies apoptosis (56), but studies on these histones still lag behind those on H3 and H4. Finally, we need to refine our ideas about the histone code and what it does, and to continue to search for specific examples of how it operates. The “epigenome” provides an attractive target for those seeking new work for soon-to-be-redundant DNA sequencing factories, but revealing the secrets of such a complex, dynamic, and multifaceted entity presents a challenge even more daunting than that of the genome itself.

Acknowledgments
The authors thank Darren White for advice on the preparation of illustrations. H.T. Spotswood is supported by a Prize Studentship from the Wellcome Trust.


