# Reading signals on the nucleosome with a new nomenclature for modified histones

Bryan M Turner

**The tails of the four core histones are exposed on the nucleosome surface, where they are subject to a variety of enzyme-catalyzed, post-translational modifications. Modifications, singly or in combination, provide a source of information that can be used for signal transduction during ongoing processes, such as transcription, or as heritable epigenetic marks. A nomenclature is presented that allows patterns of histone modification to be clearly and unambiguously specified and that should facilitate discussion of their functional roles.**

In 1977 Klug and colleagues published a structure[1] of the nucleosome core particle based on X-ray crystallographic analysis. It was and remains a thing of beauty and wonder. The structure not only explained the earlier electron microscopic and nuclease digestion data that led Kornberg[2] to propose a repeating histone-DNA structure for chromatin, but also provided us with a rock-solid foundation on which all subsequent chromatin research has been based. Important details have been added over the years as technology has advanced and resolution improved, but nothing essential has changed. Also in 1977, Bradbury published a short paper entitled "Histone nomenclature"[3]. It brought order to the various names given to the same histone proteins purified by different workers and set out the histone nomenclature that we use today, namely H2A, H2B, H3 and H4 for the four core histones (**Fig. 1**), and H1 for the major linker histone. Given the growth of interest in histones after the realization of their fundamental role in chromatin, it came just in the nick of time.

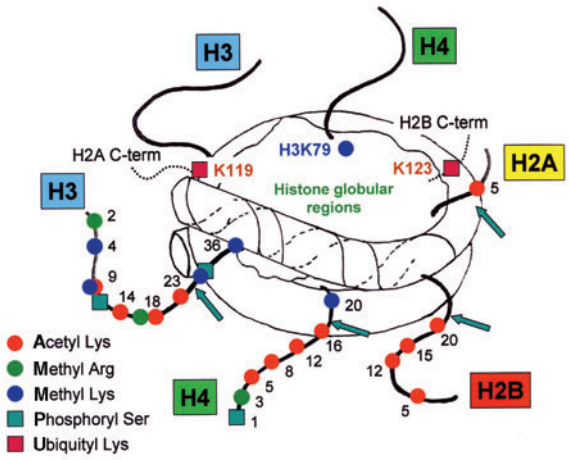


**Figure 1** Histone modifications on the nucleosome core particle. The nucleosome core particle, showing six of the eight core histone N-terminal tail domains and two C-terminal tails. The core particle consists of two copies each of histones H2A, H2B, H3 and H4, around which are wrapped 146 base pairs of DNA. The structure shown is based on the original, low-resolution X-ray structure[1]. The histone tails are not amenable to X-ray analysis and are probably unstructured; they are shown fully extended. Sites of post-translational modification are indicated by defined colored symbols (lower left). Residue numbers are shown for each modification. H3 Lys9 can be either acetylated or methylated. The C-terminal tail domains of one H2A molecule and one H2B molecule are shown (dashed lines) with sites of ubiquitylation at H2A Lys119 (most common in mammals) and H2B Lys123 (most common in yeast). Modifications are shown on only one of the two copies of histones H3 and H4 and only one tail is shown for H2A and H2B. Sites marked by blue arrows are susceptible to cutting by trypsin in intact nucleosomes. The cartoon is a compendium of data from various organisms, some of which may lack particular modifications (adapted from ref.18).

One of many remarkable things about the nucleosome core particle is that it is essentially the same in all eukaryotes. It is one of the most highly conserved structures known. This unchanging nature was consistent with the particle's fundamental role in packaging DNA into the nucleus, but raised the question of whether this was the nucleosome's sole purpose in life, or whether it played some more active role in chromatin function. Indeed, the possibility was considered that nucleosomes, despite their tight association with DNA, were 'transparent' to the transcription apparatus. Evolutionarily, this makes good sense; a highly conserved packaging structure such as the nucleosome core particle is an integral part of the genome's molecular architecture, and mechanisms of transcription, replication and repair, among others, would be expected to evolve to deal with it as efficiently as possible.

However, even from the earliest days in this field, there were indications that the core particle had at least the potential to be more than a passive DNA packaging device. Core histone variants were identified that were expressed in

*The author is in the Chromatin and Gene Expression Group, Institute of Biomedical Research, University of Birmingham Medical School, Birmingham B15 2TT, UK.*
*e-mail: b.m.turner@bham.ac.uk*

**Table 1 The Brno nomenclature for histone modifications**

| Modifying group | Amino acid(s) modified | Level of modification | Abbreviation for modification[a] | Examples of modified residues[b] |
|---|---|---|---|---|
| Acetyl- | Lysine | mono- | ac | H3K9ac |
| Methyl- | Arginine | mono- | me1 | H3R17me1 |
| | Arginine | di-, symmetrical | me2s | H3R2me2s |
| | Arginine | di-, asymmetrical | me2a | H3R17me2a |
| | Lysine | mono- | me1 | H3K4me1 |
| | Lysine | di- | me2 | H3K4me2 |
| | Lysine | tri- | me3 | H3K4me3 |
| Phosphoryl- | Serine or threonine | mono- | ph | H3S10ph |
| Ubiquityl- | Lysine | mono-[c] | ub1 | H2BK123ub1 |
| SUMOyl- | Lysine | mono- | su | H4K5su[d] |
| ADP ribosyl- | Glutamate | mono- | ar1 | H2BE2ar1 |
| | Glutamate | poly- | arn | H2BE2arn[d] |

[a]The use of lowercase letters for the modifications helps distinguish them from either amino acids (identified by their single-letter codes) or histones (such as H2A), for which letters are always uppercase. [b]The nomenclature starts from the left with the histone, then the residue, then the modification. In cases where the modified residue is not known, or not relevant, the modification should follow the histone, for example H4ac and H2Bar1. Multiple modifications can be accommodated by simply extending the listing (for example, H3K4me3K9acS10ph…) for a long as necessary. Because each individual modified residue begins with the uppercase letter specifying the amino acid, and because the modifications themselves are all designated by lowercase letters, the use of commas or dots to separate the individual modified residues in a 'word' specifying multiple modifications is not necessary. On occasion, the presence of an unmodified residue may be an essential component of an information-bearing combination of residues; in this case the residue should be inserted without additions (for example, H3K9S10ph) to indicate H3 unmodified at Lys9 and phosphorylated at Ser10. [c]Polyubiquitylated histones are designated ubn. [d]Hypothetical at present.

different cell types, such as developing male germ cells, and that inevitably changed the fine structure of the core particle. There were also puzzling features of the core particle itself, particularly the N-terminal tail domains of the four core histones. These are highly conserved and yet largely unstructured (hence invisible by X-ray crystallography), and seemed to play no role in maintaining the structure of the particle itself. Why were they there? Also, the earliest histone sequencing studies had revealed consistent and frequent modifications to defined amino acids on the histone N-terminal tails, particularly acetylation and methylation of selected lysines[4]. Did these have a function? In fact, a possible link between chromatin function and histone tail modifications was revealed at a very early stage. In 1964, Allfrey and colleagues presented evidence that altered levels of histone acetylation and methylation were associated with changed rates of transcription[5]. This association was actively explored and circumstantial evidence (for and against) accumulated, but it was almost a quarter of a century before a direct and conclusive link between upregulated transcription and increased levels of histone acetylation on the chicken α-globin gene was demonstrated by a new approach in which antibodies to acetyllysine were used to immunoprecipitate acetylated chromatin[6].

Interest then focused on the mechanisms by which histone modifications (primarily acetylation at that time) actually exerted their effects. Was it through a general loss of positive charge leading to an 'opening up' of chromatin higher-order structure and improved accessibility for transcription factors, or was it through more specific and targeted modifications to defined residues, which would then exert their own functional effects? The way forward was again through a new antibody-based approach, in this case the preparation and use of modification-specific antibodies, reagents capable of recognizing defined amino acids modified in specific ways on individual histones[7]. Such antisera provided the first evidence for an association between a specific histone tail modification and chromatin function, with the demonstration, by immunostaining, that the transcriptionally hyperactive male X chromosome in *Drosophila melanogaster* is uniquely marked by H4 acetylated at Lys16, a modified isoform not found on the autosomes or the female X chromosomes[7]. On the basis of this association, it was proposed that modifications to individual histone amino acids provided, singly or in combination, 'nucleosome surface markers' that could exert defined effects on chromatin function. These markers would operate not by directly altering the structure of the nucleosome, but instead by selectively binding nonhistone proteins, which in turn would exert a functional effect[7,8]. Over the past ten years or so, a wealth of experimental evidence, much of it antibody-based, has lent support to this hypothesis. Biochemical approaches in many laboratories have revealed not only the families of enzymes that put specific histone modifications in place but, crucially, proteins that bind with remarkable selectivity to modified residues and are associated with defined changes in chromatin structure and function[9,10]. Antibody-based approaches have also been largely responsible for revealing the rich variety of histone tail modifications disposed across the nucleosome surface and, often, their functional implications. For example, altered patterns of histone modification have been shown to be an integral part of cyclical changes in ongoing transcriptional activity[11]. Thus, the signaling function of the nucleosome and its modified forms is now a well-established principle. Nonetheless, in most cases the proteins that read these histone modification signals

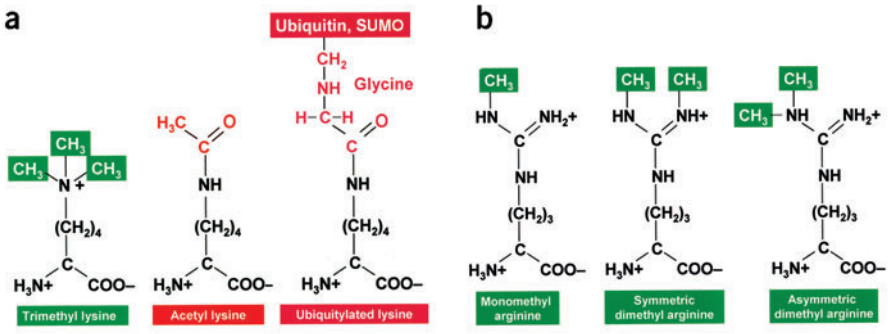


**Figure 2** Some alternative modifications to lysine and arginine residues. (**a**) Lysines can accommodate one, two or three methyl groups (left), always retaining a positive charge; a single acetate group (middle), in which case the positive charge is lost; or can form an isopeptide bond with the C-terminal glycine of ubiquitin or SUMO (right), which also neutralizes the positive charge. (**b**) Arginines can be modified with a single methyl group (left), or with two methyl groups that can be arranged symmetrically (middle) or asymmetrically (right). All forms retain a positive charge.

remain to be defined, as do the molecular mechanisms by which their information is transmitted to the transcription machinery, or to other functional processes[12].

The diversity of the various modifications to histones (**Table 1**) and the positions of these modifications along the N-terminal tails of the four core histones (**Fig. 1**) are remarkable. In addition to the old favorites of acetylation, methylation and phosphorylation, we are now aware of more complex modifications involving addition of nucleotides (ADP-ribosylation) or the proteins ubiquitin or SUMO (small ubiquitin-related modifier)[13]. Even among the simpler modifications there are complications; lysines can be mono-, di- or trimethylated, in addition to being acetylated or ubiquitylated (**Fig. 2a**), whereas arginines can carry one or two methyl groups, of which the latter can be symmetrically or asymmetrically distributed (**Fig. 2b**). But do these chemical details really matter? Remarkably, it seems that they do[10]. To give just one example, H4 trimethylated at Lys20 is located specifically at pericentric heterochromatin in mammalian cells, whereas H4 mono- or dimethylated at the same residue is not[14]. It is likely that the different levels of modification are put in place by different enzymes[10], and recognized by different binding proteins. The latter remains to be proven, but as antibodies can be made that readily distinguish the different methylation states[10,14,15], other binding proteins can potentially do the same.

In recent years, it has been proposed that histone tail modifications may have a role that reaches far beyond their involvement in signal transduction during ongoing chromatin functions such as transcription or replication. It has been suggested that the modifications, singly or in combination, constitute a code that acts in concert with the underlying genetic code itself to determine chromatin function and particularly long-term patterns of gene expression[9,16,17]. This epigenetic histone code is proposed to play a role in determining and stabilizing gene expression patterns from one cell generation to the next (cellular memory), and must therefore be heritable, a property that distinguishes it from modifications involved only in signaling during transcription or other ongoing processes[18]. Given that patterns of histone modification that constitute a heritable epigenetic code may be present alongside those involved in more immediate aspects of chromatin function, possibly even on the same nucleosome or histone tail, telling them apart presents a major technical and intellectual challenge.

It is clear that the histone modifications available on the nucleosome surface provide, in combination, an almost infinite source of variability that can be used for signal transduction and potentially for epigenetic inheritance. However, dealing with the enormous complexity of these histone modifications presents some practical challenges, one of which is to ensure that the coded information, real or potential, is presented in a consistent and coherent format. The nomenclature outlined in **Table 1** attempts to do this by providing a logical and unambiguous way of representing both the individual chemical modifications (**Table 1**, column 4) and their locations on specified residues on the histone tails (as shown in **Fig. 1**). These can be regarded as the letters of a histone alphabet. An individual modified amino acid is uniquely defined by a two-letter word specifying the residue and the modification (examples are shown in **Table 1**, column 5). Twelve letters are required to represent the known chemical modifications, and 26 additional letters specify the modifiable residues shown in **Figure 1**. The latter is an underestimate and additional, rarely modified residues are already popping up, but the size of our histone alphabet is likely to remain closer to the Roman alphabet than the Chinese one. The 12 chemical modifications shown in column 4 of **Table 1** will stay the same from one organism to another, whereas the modifiable amino acids will sometimes vary, at least in their numbering. Specific combinations of modifications can be defined by adding the appropriate letters to make a longer word (such as H3K9acS10ph). It is likely that longer words will be both necessary and biologically relevant. In its simplest form, a combinatorial histone code could be read by binding proteins that are sensitive to combinations of two adjacent modifications[19]. Although this may be appropriate for signaling purposes, more subtle or longer-term effects on chromatin function are likely to involve more numerous or more widely distributed sets of modifications—that is, longer words. For example, three (modified) residues on the H3 N-terminal tail have recently been shown[20] to influence binding of the heterochromatin protein HP1.

It seems clear that even if the cell uses only a small fraction of the potential information content of histone modifications, we will understand it only if our record-keeping is as clear as that of the cell itself. Hopefully a standard nomenclature and an agreed alphabet are a step in the right direction.

ACKNOWLEDGMENTS

The nomenclature set out here was devised following the first meeting of the Epigenome Network of Excellence (NoE), at the Mendel Abbey in Brno, Czech Republic. For this reason, it can be referred to as the Brno nomenclature. Epigenome is a consortium of 25 research groups across Europe, plus associated members and NET teams, working in epigenetic research and funded by the European Union. The network is coordinated and directed by T. Jenuwein (Institute of Molecular Pathology, Vienna). The nomenclature presented here was finalized following wide discussions after the meeting, including consultations with the seven members of our scientific advisory board, all from the United States. I am grateful to all those who took part for their helpful, constructive and insightful comments, but particularly to P. Becker and A. Peters for a crucial early suggestion. Details of the network and its members are available at http://www.epigenome-noe.net/.

1. Finch, J.T. *et al.* Structure of the nucleosome core particle of chromatin. *Nature* **269**, 29–36 (1977).
2. Kornberg, R.D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868–871 (1974).
3. Bradbury, E.M. Histone nomenclature. *Methods Cell Biol.* **16**, 179–181 (1977).
4. Van Holde, K.E. *Chromatin* (Springer, New York, 1988).
5. Allfrey, V.G., Faulkner, R. & Mirsky, A.E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. USA* **51**, 786–794 (1964).
6. Hebbes, T.R., Thorne, A.W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7**, 1395–1403 (1988).
7. Turner, B.M., Birley, A.J. & Lavender, J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* **69**, 375–384 (1992).
8. Turner, B.M. Decoding the nucleosome. *Cell* **75**, 5–8 (1993)
9. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
10. Lachner, M., O'Sullivan, R.J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116**, 2117–2124 (2003).
11. Métivier, R. *et al.* Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
12. Schreiber, S.L. & Bernstein, B. Signaling network model of chromatin. *Cell* **111**, 771–778 (2002).
13. Shiio, Y. & Eisenman, R.N. Histone sumoylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. USA* **100**, 13225–23230 (2003).
14. Schotta, G. *et al.* Silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
15. Peters, A.H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
16. Turner, B.M. Histone acetylation and an epigenetic code. *BioEssays* **22**, 836–845 (2000).
17. Strahl, B. & Allis, C.D. The language of covalent histone modifications *Nature* **403**, 41–45 (2000).
18. Turner, B.M. Cellular memory and the histone code. *Cell* **111**, 285–291 (2002).
19. Fischle, W., Wang, Y. & Allis, C.D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
20. Mateescu, B. *et al.* Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. *EMBO Rep.* **5**, 490–496 (2004).