

Histone modifications: Combinatorial complexity or cumulative simplicity?

Steven Henikoff*

Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024, Seattle, WA 98109-1024

Posttranslational modifications of histones have attracted enduring interest ever since it was realized that histones are hyperacetylated on lysines at actively transcribed genes (1). Because it neutralizes the charge on a lysine, acetylation was thought to reduce interactions with DNA phosphates, making the DNA more accessible for active processes such as transcription (2). In recent years, this simple charge neutralization model has been succeeded by a complex alternative: the histone code, in which combinations of different histone modifications specify alternative chromatin states (3). In this issue of PNAS, Dion *et al.* (4) present a critical test of these competing models.

The concept of a histone code was introduced in the early 1990s by Turner (5), based on seminal studies of the involvement of histone lysine acetylation in a dosage compensation process. In flies, histone H4 is hyperacetylated on lysine-16 (K16) on the male, but not the female, X chromosome, a feature that was later shown to be instrumental in the process whereby the male X chromosome is 2-fold up-regulated to compensate for being hemizygous (6). Thus, H4 K16 appears to be dedicated to the process of X chromosome dosage compensation in *Drosophila*. The subsequent discovery that histone acetyltransferases and deacetylases are components of activator or repressor complexes (7), and the realization that histone methyltransferases provide potentially enormous combinatorial complexity (8), led to elaborations of Turner's original concept and considerable excitement in the chromatin field.

One would think that an abundance of evidence underlies this paradigm shift from a model based on structural properties to an information-based code. However, experiments claimed to support the histone code hypothesis might also be accommodated by structural alternatives. For example, histone H3 K9 methylation is regarded as an epigenetic "mark" for heterochromatin because it provides a platform for binding by heterochromatin-associated protein 1 (HP-1) (8). It is not known whether the modification step occurs on the chromatin template, which could be interpreted

as "writing" a code, or instead on soluble H3 before assembly, in which case H3 K9 methylation is just a prerequisite for assembly of a heterochromatic structure. In at least one case, H3 K9 is found to be methylated within a soluble nucleosome assembly complex, which favors the structural model (9).

To experimentally distinguish the histone code from structural alternatives, such as charge neutralization and nucleosome assembly, the concept needs to be rigorously defined. It is worth noting that the term "code," as commonly understood, involves a translation machine. The Morse code, a computer code, and the genetic code each have a translation machine in the form of a telegraph, a program, or a ribosome. However, the binding of HP-1 to methylated H3 K9 in heterochromatin does not constitute a translation machine, because "reading" is synonymous to simply "binding," with only one bound state, like a telegraph key that can only read a single dot. Multiple inputs or outputs are needed for a nontrivial code. These must be distinct, not simply cumulative: one Morse code dot reads "e" and two read "i," not "e" and "ee." Thus, what is needed to distinguish a nontrivial histone code from cumulative alternatives like charge neutralization is the demonstration of distinct outputs using different combinations of input components.

In the case of H3 K9 methylation, a recent study in *Arabidopsis* provides compelling evidence that different combinations of input components can provide distinct outputs. Genetic studies indicate that methylation by the DNA methyltransferase, CHROMOMETHYLASE3 (CMT3), is maintained by the action of an H3 K9 methyltransferase (10, 11). Recently, Lindroth *et al.* (12) showed that the CMT3 enzyme requires methylation of both H3 K9 and H3 K27 for avid binding to an H3 N-terminal tail peptide *in vitro*. Thus, there are multiple states of this histone tail, and CMT3 binds effectively to only one of them. *In vivo*, this would result in a binary output, the methylation of DNA cytosine bases by CMT3.

In some other cases, histone modifications might act only indirectly, complicating interpretations of cause and effect. For example, ubiquitylation of H2B

K123 leads to the methylation of H3 K4 in budding yeast, which contributes to increased transcription (13). Although it remains possible that this is a direct interaction between modified histone lysines within single nucleosomes, an attractive alternative hypothesis is that ubiquitylated H2B at the promoter facilitates transcription, resulting in modification of nucleosomes in the body of the gene (14). In this case, the combinatorial read-out is not necessarily evidence for a code but rather would simply be one of many consequences of transcriptional activation.

Budding yeast provides an attractive model system for a critical test of the histone code hypothesis. Histone lysines are often differentially acetylated, especially in the vicinity of promoters (2). Although chromatin immunoprecipitation (ChIP) analysis using microarrays has revealed that acetylations on all four core histones are mostly strongly correlated with one another, a sensitive statistical analysis detected differences (15). These differences were suggested to reflect specific combinations of acetylated histone lysines that contribute to transcriptional regulation. If so, then loss of individual sites of lysine acetylation should impact transcription genome-wide in a combinatorial manner. Alternatively, if acetylation only neutralizes charge, then substituting lysines for a similar residue that cannot be acetylated should have simple cumulative effects on transcription.

Dion *et al.* (4) used microarrays to profile expression of yeast genes for a series of strains in which one, two, or three lysines on the N-terminal tail of histone H4 have been substituted with arginine (K-to-R), thus preventing acetylation while retaining the positive charge. Dion *et al.* used hierarchical cluster analysis to identify similarly regulated genes in the different mutant lines. If effects on expression result from charge effects, then similar groupings should be found for all four cases in which only a single lysine was substituted with arginine, for all combinations

See companion article on page 5501.

*E-mail: steveh@fhcrc.org.

© 2005 by The National Academy of Sciences of the USA

of two substitutions and for all combinations of three substitutions; by and large, these similar groupings were found. In addition, changes in gene expression levels attributable to K-to-R substitutions at positions 5, 8, and 12 were indistinguishable for all $\approx 1,200$ affected genes (Fig. 1). Only K16-to-R16 did not fully conform to a cumulative model, with $\approx 10\%$ of affected genes showing expression changes that were discordant with expression changes seen for K-to-R substitutions at the other three positions.

The general interchangeability of sites of histone H4 acetylation provides support for the charge neutralization model. In addition, Dion *et al.* (4) found that coregulated genes comprise novel nonrandom clusters along each of the chromosomes. Clusters of coregulated genes have been described in *Drosophila* (16), and charge clusters resulting from extensive histone acetylation provide a possible structural basis.

Exceptions to the general rule of interchangeability described by Dion *et al.* (4) are consistent with studies of Sir3 and Bdf1, two proteins that bind to H4 tails and depend on K16 acetylation to silence or activate transcription (17, 18). As is the case for *Drosophila* dosage compensation, H4 K16 acetylation appears to be a special modification in yeast. However, a single special modification is not a code, and we seem to be no farther along in establishing a non-trivial code for transcriptional regulation than Turner was over a decade ago (5). Nevertheless, by directly testing predictions of the histone code hypothesis, Dion *et al.* (4) have provided strong evidence for the impact of charge on gene expression, and this should focus re-

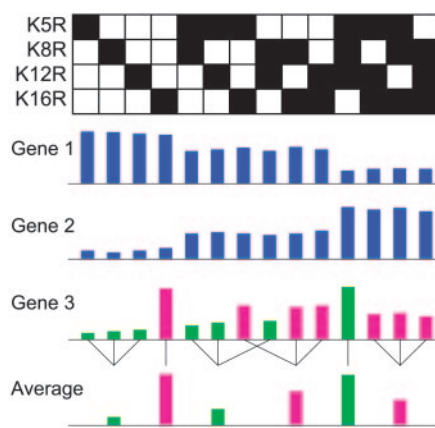


Fig. 1. Testing the histone code hypothesis. Dion *et al.* (4) constructed all combinations in which one, two, or three of the four histone lysine residues were changed to arginine (black rectangles) and measured expression levels (vertical bars). For gene 1, monotonic decreases in expression are seen with increases in the number of K-to-R changes regardless of the particular residue changed. For gene 2, monotonic increases in expression are seen. For gene 3, there are significant incremental expression changes attributable to K16 (magenta bars), calculated by comparing the K16 value to the average of K5, -8, and -12 values (green bars). This finding indicates that changing K16 to R has an effect on gene 3 that is discordant with changes of K5, -8, and -12 to R.

newed attention on charge-dependent interactions (7).

There are indications that the findings of Dion *et al.* (4) will generalize to other histone modifications and other organisms. For example, a ChIP study using microarrays in *Drosophila* showed striking correlations between H4 acetylation and several modifications on H3, including di- and trimethylation of K4 and a core modification, dimethylation of K79 (19). The correlations between different modifications were similar in

magnitude to correlations between replicate samples for the same modification, leaving little if any room for modification differences to contribute to biological processes. Strong correlations between tail modifications were also reported in human cells (20). That study also revealed that at some sites trimethylated H3 K4 was better correlated with transcriptional starts than dimethylated H3 K4, an observation that might reflect the action of a recently discovered mono- and dimethyl H3 K4 demethylase within a transcriptional corepressor complex (21). Therefore, although specific modifications contribute to gene expression in important ways, the existence of a combinatorial histone code for gene expression is in doubt.

The histone code was proposed and popularized before revolutionary findings that small double-stranded RNAs are often involved in heterochromatin formation and DNA methylation (22) with corresponding changes in histone modification. In addition, it appears that nucleosomes can be replaced as part of the transcriptional process, so that coordinated histone patterns at active genes might result from turnover during transcriptional initiation and elongation (23). Transiently hyperacetylated nucleosomes are evicted from yeast promoters upon transcriptional induction (24), consistent with a role for acetylation in mobilizing nucleosomes rather than marking them. The rapid turnover of acetylation, the eviction and transcription-coupled assembly of nucleosomes, and the presence of nucleosome remodeling subunits in histone replacement complexes (23–26) all support the notion that modifications facilitate the assembly and disassembly of chromatin. These processes provide attractive testable mechanisms for gene regulation.

1. Allfrey, V. G., Faulkner, R. M. & Mirsky, A. E. (1964) *Proc. Natl. Acad. Sci. USA* **51**, 786–794.
2. Wade, P. A., Pruss, D. & Wolffe, A. P. (1997) *Trends Biochem. Sci.* **22**, 128–132.
3. Strahl, B. D. & Allis, C. D. (2000) *Nature* **403**, 41–45.
4. Dion, M. F., Altschuler, S. J., Wu, L. F. & Rando, O. J. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 5501–5506.
5. Turner, B. M. (1993) *Cell* **75**, 5–8.
6. Gilfillan, G. D., Dahlsveen, I. K. & Becker, P. B. (2004) *FEBS Lett.* **567**, 8–14.
7. Hansen, J. C., Tse, C. & Wolffe, A. P. (1998) *Biochemistry* **37**, 17637–17641.
8. Jenuwein, T. & Allis, C. D. (2001) *Science* **293**, 1074–1080.
9. Sarraf, S. A. & Stancheva, I. (2004) *Mol. Cell* **15**, 595–605.
10. Malagnac, F., Bartee, L. & Bender, J. (2002) *EMBO J.* **21**, 6842–6852.
11. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. (2002) *Nature* **416**, 556–560.
12. Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., *et al.* (2004) *EMBO J.* **23**, 4286–4296.
13. Sun, Z. W. & Allis, C. D. (2002) *Nature* **418**, 104–108.
14. Ezhkova, E. & Tansey, W. P. (2004) *Mol. Cell* **13**, 435–442.
15. Kurdistani, S. K., Tavazoie, S. & Grunstein, M. (2004) *Cell* **117**, 721–733.
16. Spellman, P. T. & Rubin, G. M. (2002) *J. Biol.* **1**, 5.
17. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. & Grunstein, M. (1995) *Cell* **80**, 583–592.
18. Matangasombut, O. & Buratowski, S. (2003) *Mol. Cell* **11**, 353–363.
19. Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., *et al.* (2004) *Genes Dev.* **18**, 1263–1271.
20. Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., III, Gingeras, T. R., *et al.* (2005) *Cell* **120**, 169–181.
21. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J. R., Cole, P. A., Casero, R. A. & Shi, Y. (2004) *Cell* **119**, 941–953.
22. Lippman, Z. & Martienssen, R. (2004) *Nature* **431**, 364–370.
23. Henikoff, S., Furuyama, T. & Ahmad, A. (2004) *Trends Genet.* **20**, 320–326.
24. Reinke, H. & Horz, W. (2003) *Mol. Cell* **23**, 3468–3476.
25. Hanlon, S. E. & Lieb, J. D. (2004) *Curr. Opin. Genet. Dev.* **14**, 697–705.
26. Waterborg, J. H. (2002) *Biochem. Cell Biol.* **80**, 363–378.