It takes a PHD to interpret histone methylation

Yi Zhang

Covalent modifications of histones play an important role in regulating chromatin structure and function, probably by serving as docking sites for effector proteins. The discovery that PHD fingers of two different proteins recognize trimethyl-Lys4 of histone H3 supports and extends this notion.

DNA in eukaryotic cells is packaged with core histones and other chromosomal proteins to form chromatin. The basic repeating unit of chromatin, the nucleosome, is composed of two copies of each of the four core histones, H2A, H2B, H3 and H4, wrapped by 146 base pairs (bp) of DNA[1]. The core histones are rife with covalent modifications, including acetyl, phosphoryl, ubiquitin and methyl groups[2], that can serve as docking sites for effector proteins to mediate downstream events[3,4]. Four recent studies in *Nature*[5–8] demonstrate that the plant homeodomain (PHD) finger can specifically recognize and interpret the trimethylation of Lys4 on histone H3 (H3K4; one-letter amino acid abbreviations are used here for histone residues).

Although chromatin had been regarded as a static structural entity whose major function was to store genetic information, more recent studies have revealed its dynamic nature. Changes in chromatin structure are largely mediated by ATP-dependent chromatin remodeling and covalent histone modifications. In general, histone acetylation at lysine residues, regardless of their position, correlates with gene activation. In contrast, histone methylation can lead to either transcriptional activation or repression, depending on the site of methylation[9]. Even when considering a specific lysine residue, the biological consequence of methylation may differ depending on whether it is mono-, di- or trimethylated. How can the same modification mediate different biological outcomes? One idea is that the outcome of a particular modification is mediated by the downstream effector proteins that recognize and 'interpret' specific changes[3,4]. This hypothesis has gained support with the identification of protein modules that specifically recognize and bind a particular modified lysine residue. For example, bromodomains have been demonstrated to recognize acetylated lysine residues[10,11]. Similarly, chromodomains of heterochromatin protein-1 (HP1) and Polycomb have been shown to recognize and bind methylated H3K9 and H3K27, respectively[12–15]. Now, the protein module named the PHD finger has been shown to specifically recognize trimethylated H3K4 (H3K4me3)[5–8] (**Fig. 1**). This new discovery not only uncovers a novel function of the PHD finger, but also provides new insights into the function of H3K4 methylation.

Genome-wide analysis has revealed that H3K4 trimethylation is preferentially associated with the transcription start site of active genes[16,17]. A protein that binds H3K4me3 was isolated using a peptide pull-down assay[8] and was identified as BPTF[18], the largest subunit of the nucleosome-remodeling factor NURF[19,20]. NURF is an ATP-dependent chromatin remodeling factor known to be involved in the expression of a large number of genes[19,20]. The affinity of BPTF for H3K4me3 is evolutionarily conserved, as the fly counterpart, NURF301, by itself or in complex with other NURF components has similar binding affinity for H3K4me3. Deletion studies identified the bromodomain-proximal PHD finger of BPTF/NURF301 as responsible for the H3K4me3 binding. The H3K4me3-binding activity of BPTF is functionally important, as manipulation of H3K4me3 abundance through small interfering RNA (siRNA)-mediated knockdown of WDR5, an important modulator of genome-wide H3K4me3 methylation[21], impairs the recruitment of the NURF complex to specific target genes. Consistent with a role in recognition and interpretation of H3K4me3 modification, morpholino depletion of the *Xenopus laevis* BPTF homolog phenocopies many of the developmental defects observed when WDR5 is depleted. Furthermore, the phenotypes can be largely rescued by expression of wild-type *Xenopus* BPTF, but not by H3K4me3 recognition–defective mutants, which strongly supports an essential function of H3K4me3 recognition by BPTF *in vivo*.

Li *et al.* solved the structure of the bromodomain-proximal PHD finger of BPTF in complex with an H3K4me3 peptide[5]. The structure revealed that the H3K4me3 peptide interacts with BPTF through an antiparallel β-sheet formed on the surface of the PHD finger. Two structural features are believed to be crucial for the specificity of recognition. First, the PHD finger forms two deep binding channels that are separated by a conserved tryptophan, Trp32. R2 and K4me3 of H3 respectively occupy the two recognition pockets (**Fig. 1b**). Notably, separation of the two binding channels makes them uniquely suitable for R2 and K4me3 recognition, as the two amino acids are separated by T3, which is not the case for R8-K9me3 or R26-K27me3. Second, an aromatic cage formed by Tyr10, Tyr17, Tyr23 and Trp32 of BPTF directly binds K4me3 and contributes to the specific recognition of the trimethyllysine (**Fig. 1b**). Thus, the two studies provide both biological and structural evidence that the BPTF PHD finger is involved in reading the H3K4me3 mark.

PHD fingers, which comprise about 60 amino acid residues and belong to the C4HC3-type zinc-finger class, are a common structural motif found in all eukaryotes[22]. A SMART search of the human genome identifies more than 300 PHD domains in 181 proteins. Given the large number of proteins that contain the PHD finger, it is not surprising that additional PHD finger proteins may also recognize H3K4me3. An independent study from Shi *et al.* identified the PHD finger of the inhibitor of growth-2 (ING2) tumor-suppressor protein as a module that binds H3K4me3 and H3 dimethyl-Lys4 (H3K4me2)[7]. ING2 is a component of the repressive mSin3A–HDAC complex and has been shown to have an important role in the cell cycle, apoptosis and DNA repair[23]. Previous studies indicate that recruitment of ING2 to chromatin in response to DNA damage involves binding of the ING2 PHD finger to the lipid signaling molecule phosphatidylinositol-5-phosphate[24]. Experiments seeking additional binding partners of the ING2 PHD finger identified histone H3. Shi *et al.* observed that, although the ING2 PHD finger binds robustly to native mononucleosomes, it does not bind mononucleosomes reconstituted from recombinant histones, raising the intriguing possibility that covalent modifications on H3 might be a requirement for the interaction. Peptide pull-down studies identified the modification to be H3K4me3. Consistent with

The author is in the Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA.
e-mail: yi_zhang@med.unc.edu

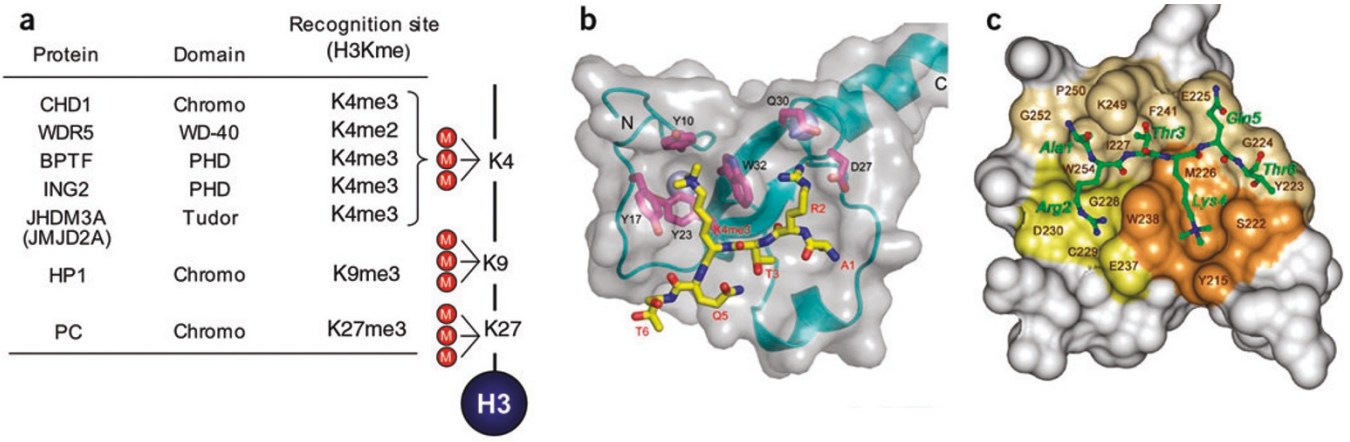


Figure 1 Recognition of methylated lysines in H3. (**a**) Known protein modules that recognize methyllysine residues (K4, K9, K27) on H3. Not all of the modules shown are specific for H3K4me3. (**b**,**c**) Structures of the PHD fingers of BPTF (**b**; provided by H. Li and D.J. Patel) and ING2 (**c**; provided by T.G. Kutateladze) in complex with H3K4me3 peptides. Amino acid residue numbers as in the original papers[5,6].

a role for the ING2 PHD finger in H3K4me3 binding, an intact ING2 PHD finger is required for efficient histone deacetylation by the ING2-containing mSin3A–HDAC complex. The H3K4me3-binding activity of the ING2 PHD finger is biologically relevant, as manipulation of H3K4me3 abundance through siRNA–mediated knockdown of WDR5, a global H3K4me3 regulator, decreases ING2 occupancy on the cyclin D1 promoter. More importantly, knockdown of ING2 results in upregulation of cyclin D1, which can be rescued by wild-type ING2, but not by an H3K4me3 binding–defective ING2 mutant. Collectively, these data establish a role for the ING2 PHD finger in H3K4me3 recognition and cyclin D1 repression.

In addition to the biochemical and biological evidence mentioned above, a companion paper from Peña *et al.* reveals the structural basis for recognition of H3K4me3 by the ING2 PHD finger[6]. The structure, which includes two channels separated by Trp238 of ING2, respectively, accommodating R2 and K4me3 of H3 (**Fig. 1c**), is very similar to that of H3K4me3 in complex with the BPTF PHD finger[5]. However, a notable difference is that the aromatic cage involved in K4me3 recognition consists of only two aromatic residues (Tyr215 and Trp238), instead of four (**Fig. 1c**).

Given that the PHD fingers of ING-family proteins are highly conserved, all ING family members can probably recognize H3K4me3. Indeed, NMR and tryptophan-fluorescence spectroscopic studies have shown that all PHD fingers of ING-family proteins have similar affinities for the H3K4me3 peptide[6]. However, it is likely that only a small subset of PHD finger proteins can bind H3K4me3, because the crucial residues for H3K4me3 recognition are not conserved in many of the PHD finger proteins. Indeed, an alignment of all the PHD fingers of human proteins has revealed only about ten PHD fingers outside of the ING family that contain these crucial residues. Whether conservation of the residues involved in H3K4m3 binding alone is sufficient for the interaction remains to be established. Determining the physiologically relevant ligand(s) of other PHD fingers is also an important challenge for future studies.

In addition to the PHD fingers, recent studies have revealed at least three modules that can also bind H3K4me (**Fig. 1a**). For example, the chromodomain of chromo-ATPase/helicase-DNA-binding-1 (CHD1) has been reported to bind H3K4me[25,26]. Cocrystal structure studies have revealed that the two chromodomains of human CHD1 cooperate to interact with one H3K4me tail[27]. Similar to ING2, methylammonium recognition by the human CHD1 double chromodomain involves two aromatic residues instead of the three-aromatic-residue cage used by the chromodomains of the HP1 and Polycomb proteins. Furthermore, the WD-40 repeat protein WDR5 mentioned above has also been reported to have specificity for H3K4me2 (ref. 21). Unlike the chromodomains and PHD fingers that recognize H3K4me using an aromatic cage, the specificity of WDR5 for H3K4me2 is conferred by unconventional hydrogen bonds between the two ζ-methyl groups of K4me2 and the carboxylated oxygen of Glu322 in WDR5 (ref. 28). Finally, a screen of a protein-domain array has revealed that the tudor domains of the histone demethylase JHDM3A/JMJD2A can also bind H3K4me[29,30]. A cocrystal structure shows that an aromatic cage, composed of three aromatic residues from both tudor domains, is responsible for H3K4me3 recognition, and the binding specificity is conferred by side chain interactions involving residues from the first tudor domain[31]. Thus, the double tudor domain of JHDM3A functions as an interdigitated H3K4me3-binding domain.

Studies mentioned above indicate that the H3K4me mark can be recognized by at least four different protein modules. Although these protein modules have very little sequence similarity, the structural features that confer their specificity for H3K4me have two common features. First, they all form an aromatic cage responsible for H3K4me recognition. Second, the three H3 residues preceding K4, particularly R2, make multiple contacts with the protein modules and contribute to binding specificity. These studies provide strong support for the general view that histone modifications serve as binding sites for effector proteins to mediate downstream effects and that every residue in a histone may make contributions in determining how upstream signaling information is 'read' by these effectors.

Other modifications may also affect reading. For example, does R2 methylation or T3 phosphorylation in H3 matter? Given that multiple protein modules can recognize the same modification, the biological readout of a particular modification is likely to be more complicated than previously thought. For example, H3K4me3 was previously regarded as an epigenetic mark for gene activation[9]. However, the new studies indicate that H3K4me3 can also signal transcription repression, as an intact PHD finger of ING2 is required for binding and repression of the cyclin D1 gene by the ING2 complex[7]. Similarly, although the new study examining NURF association with H3K4me3 is mainly focused on its role in *HOX* gene activation[8], a genome-wide analysis of NURF mutants has also revealed that close to equal numbers of genes are upregulated and downregulated by histone methylation[19]. The bifunctional effect

of histone methylation on transcription is not limited to H3K4me3 as a well-known silencing mark, H3K9me3, has been recently linked to gene activation[32]. It is clear that methylation of a particular lysine residue *per se* does not dictate transcriptional outcome. Instead, this is determined by the effector protein or protein complex that binds the methylated residue.

If the above notion is correct, identification and characterization of the effector proteins is crucial for understanding the function of histone modifications. In this sense, the studies described here provide new insights into the function of H3K4 methylation[5–8], but also raise several important questions. (i) How do proteins with modules capable of binding H3K4me3 find their target genes, given that so many different proteins can bind H3K4me3? (ii) How much does the H3K4me3-binding property contribute to target gene recognition? (iii) What other factors are also important for target gene recognition? (iv) Is the moderate difference in binding affinity of PHD fingers for H3K4me2 and H3K4me3 *in vitro* relevant *in vivo*? These questions are pertinent, because NURF, for example, can be recruited to specific target genes through interaction with sequence-specific transcription factors, such as the ecdysone nuclear receptor (EcR)[19]. Interestingly, EcR can also directly recruit an H3K4 methyltransferase coactivator complex[33]. It seems that EcR is upstream in the regulatory pathway whereas NURF and the H3K4 methyltransferase recruitment are downstream. It is likely that recruitment of NURF to a specific gene promoter is mainly determined by sequence-specific transcription factors. Once recruited, it is stabilized by binding to H3K4me3, a mark enriched in many active gene promoters[16,17].

The H3K4me-binding proteins described above are involved in diverse biological processes, including transcriptional regulation, the cell cycle and apoptosis, yet they recognize the same modification. Many of the questions mentioned above could be addressed by a chromatin immunoprecipitation–coupled genomic microarray approach, which would reveal the colocalization of H3K4 methylation sites and specific H3K4me-binding proteins. The relative importance of H3K4me3 in the recruitment of the binding protein could be evaluated by using a similar approach, but analyzing cells that lack H3K4me3, such as cells depleted of WDR5. As both BPTF and ING2 function in complexes with other associated proteins *in vivo*, understanding the potential effect of the associated proteins on H3K4me3 recognition is also important. Answers to many of the above questions are probably forthcoming.

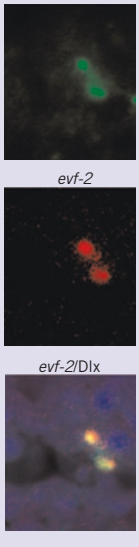
1. Kornberg, R.D. & Lorch, Y. *Cell* **98**, 285–294 (1999).
2. Jenuwein, T. & Allis, C.D. *Science* **293**, 1074–1080 (2001).
3. Strahl, B.D. & Allis, C.D. *Nature* **403**, 41–45 (2000).
4. Turner, B.M. *Bioessays* **22**, 836–845 (2000).
5. Li, H. *et al. Nature*, advance online publication 21 May 2006 (doi:10.1038/nature04802).
6. Peña, P.V. *et al. Nature*, advance online publication 21 May 2006 (doi:10.1038/nature04814).
7. Shi, X. *et al. Nature*, advance online publication 21 May 2006 (doi:10.1038/nature04835).
8. Wysocka, J. *et al. Nature*, advance online publication 21 May 2006 (doi:10.1038/nature04815).
9. Martin, C. & Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
10. Dhalluin, C. *et al. Nature* **399**, 491–496 (1999).
11. Jacobson, R.H., Ladurner, A.G., King, D.S. & Tjian, R. *Science* **288**, 1422–1425 (2000).
12. Fischle, W. *et al. Genes Dev.* **17**, 1870–1881 (2003).
13. Jacobs, S.A. & Khorasanizadeh, S. *Science* **295**, 2080–2083 (2002).
14. Min, J., Zhang, Y. & Xu, R.M. *Genes Dev.* **17**, 1823–1828 (2003).
15. Nielsen, P.R. *et al. Nature* **416**, 103–107 (2002).
16. Bernstein, B.E. *et al. Cell* **120**, 169–181 (2005).
17. Schneider, R. *et al. Nat. Cell Biol.* **6**, 73–77 (2004).
18. Barak, O. *et al. EMBO J.* **22**, 6089–6100 (2003).
19. Badenhorst, P. *et al. Genes Dev.* **19**, 2540–2545 (2005).
20. Tsukiyama, T. & Wu, C. *Cell* **83**, 1011–1020 (1995).
21. Wysocka, J. *et al. Cell* **121**, 859–872 (2005).
22. Bienz, M. *Trends Biochem. Sci.* **31**, 35–40 (2006).
23. Doyon, Y. *et al. Mol. Cell* **21**, 51–64 (2006).
24. Gozani, O. *et al. Cell* **114**, 99–111 (2003).
25. Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R. III & Grant, P.A. *Nature* **433**, 434–438 (2005).
26. Sims, R.J., III *et al. J. Biol. Chem.* **280**, 41789–41792 (2005).
27. Flanagan, J.F. *et al. Nature* **438**, 1181–1185 (2005).
28. Han, Z. *et al. Mol. Cell* **22**, 137–144 (2006).
29. Kim, J. *et al. EMBO Rep.* **7**, 397–403 (2006).
30. Klose, R.J. *et al. Nature*, advance online publication 28 May 2006 (doi:10.1038/nature04853).
31. Huang, Y., Fang, J., Bedford, M.T., Zhang, Y. & Xu, R.M. *Science* **312**, 748–751 (2006).
32. Vakoc, C.R., Mandat, S.A., Olenchock, B.A. & Blobel, G.A. *Mol. Cell* **19**, 381–391 (2005).
33. Sedkov, Y. *et al. Nature* **426**, 78–83 (2003).

Noncoding RNAs and homeodomains get together

Non-coding RNAs (ncRNAs) exist in many flavors and facilitate a variety of molecular processes, from X chromosome inactivation to splicing to translational regulation. In a recent paper by Kohtz and colleagues (*Genes Dev.* **20**, 1470-1484, 2006), a ncRNA is directly implicated in gene regulation by homeodomain transcription factors during vertebrate brain patterning.

The authors isolated a non-coding transcript, *evf-2*, corresponding to an ultraconserved region (> 90% conserved from fish to humans) that lies between the homeodomain transcription factors Dlx-5 and Dlx-6. *evf-2* is a novel splice form of *evf-1*, a non-coding RNA previously isolated from specific regions of the brain, and is a polyadenylated, single stranded RNA. Although the authors do not exclude the possibility of an *evf-2*–encoded peptide, the RNA contains few open reading frames, suggesting that it is indeed a ncRNA.

evf-2 expression is responsive to sonic hedgehog (shh), a signaling molecule that plays a major role in developmental patterning, as viral overexpression of shh in mouse forebrains results in increased *evf-2* expression. In addition, expression of a reporter carrying the *dlx-5/6* region is increased dose-dependently by the *evf-2* RNA, in a fashion dependent upon the Dlx-2 homeodomain protein. This data is consistent with *evf-2* acting together with Dlx-2 to regulate transcription.



Reporter-based experiments suggest that the effect of *evf-2* on Dlx-2 transcription is specific to the *dlx-5/6* enhancer and cell line specific, perhaps suggesting tissue-specificity. Moreover, the effect of *evf-2* on gene expression is strongest in combination with Dlx-2; *evf-2* has less effect on regulation by other Dlx proteins and little effect on other homeodomain proteins.

Further experiments using the *dlx-5/6* reporter assay indicate that *evf-2* does not repress known Dlx-2 inhibitors, so the possibility of a direct interaction between the ncRNA and the homeodomain protein was tested. Dlx-2 forms a complex with *evf-2* in cells, and *evf-2* can be detected in immunoprecipitates of Dlx-family proteins from embryonic nuclear extracts, suggesting an *in vivo* interaction. In addition fluorescent *in situ* hybridization detects two *evf-2* foci that colocalize with Dlx-2 in cell nuclei within a specific region of the developing mouse forebrain (see picture). While further experiments are required, these data tantalizingly suggest a direct interaction between the two factors.

It is known that fly *Rox* RNAs upregulate male X chromosome transcription, while the *SRA* ncRNA increases steroid receptor gene transcription. Further work will reveal whether there is a common mechanistic basis for ncRNA activity in these different systems and whether other genomic regions contain ncRNAs that regulate transcription of nearby genes.

Sabbi Lall