Chemical Approaches for Studying Histone Modifications*

Published, JBC Papers in Press, February 10, 2010, DOI 10.1074/jbc.R109.080291
Champak Chatterjee and Tom W. Muir
From the Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York, New York 10065

Histones form the protein core around which genomic DNA is wrapped in eukaryotic chromatin. Numerous genetic studies have established that the structure and transcriptional state of chromatin are closely related to histone post-translational modifications. Further elucidation of the precise mechanistic roles for individual histone modifications requires the ability to isolate and study homogeneously modified histones. However, the highly heterogeneous nature of histone modifications in vivo poses a significant challenge for such studies. Chemical tools that have enabled biochemical and biophysical studies of site-specifically modified histones are the focus of this minireview.

Genomic DNA is stored as chromatin in the nuclei of eukaryotic cells. The fundamental repeating unit of chromatin is the mononucleosome. Each mononucleosome consists of ~147 bp of double-stranded DNA wrapped around an octameric protein complex composed of two copies each of the four core histones, H2A, H2B, H3, and H4 (Fig. 1A) (1).

The histones were first isolated and characterized in 1884 by Albrecht Kossel (2). However, it was not until 1950 that Stedman and Stedman (3) identified multiple forms of histones in the nuclei of cells and put forth the hypothesis that different cellular phenotypes in an organism may arise from the suppression of different genes by cell-specific histones. A decade later, in vitro experiments with cell-free systems demonstrated that histones were indeed inhibitory to DNA-templated RNA synthesis (4). This period also saw the discovery of histone acetylation by Phillips (5) and of histone lysine e-N-methylation by Murray (6). As the protein synthesis inhibitor puromycin did not inhibit these histone modifications, Allfrey et al. (7) suggested that acetylation and methylation were PTMs of histones. On the basis of the observation that chemical acetylation of histones greatly reduced their inhibitory effect on RNA synthesis, they put forth the prescient hypothesis that small reversible PTMs of histones could switch RNA synthesis on or off at different loci along the chromosome.

Since these early studies, a large number of histone PTMs (Fig. 1B), along with the various proteins responsible for installing (writers), removing (erasers), and binding (readers) these PTMs, have been identified (8). It is now well established that both the position and chemical property of histone modifications dictate the structure of chromatin as well as its functions in transcription, replication, and DNA repair (9). This has led to the histone code hypothesis for epigenetic control of cellular events, whereby distinct histone modifications, on one or more tails, act sequentially or in combination to bring about distinct downstream events (9). Understanding the specific roles for histone modification, either individually or in combination with other modifications, and histone interactions with chromatin-associated proteins is key to understanding the mechanisms underlying epigenetic control of cellular activity. Chemistry provides a growing arsenal of tools to study the roles for histone PTMs, and these will be discussed below.

Synthetic Peptide Strategies

The research groups of Allfrey and Merrifield first demonstrated the application of a synthetic peptide to address the substrate specificity of histone-deacetylating enzymes (HDACs) (10). A histone H4 peptide corresponding to residues 15–21 and 14C-acetylated at Lys16 was synthesized by automated solid-phase peptide synthesis. Whereas a chymotryptic peptide from H4-acetylated H4 (residues 1–37) was deacetylated by HDACs purified from calf thymus, the heptameric synthetic peptide was not. This suggested that HDACs require a minimum sequence for activity in vitro but, importantly, not the entire H4 sequence. The authors then synthesized the longer H4 peptide (residues 1–37) uniformly 14C-acetylated at Lys12 and 3H-acetylated at Lys16 and subjected it to enzymatic deacetylation (11). Interestingly, the relative amounts of [14C]acetate released from Lys12 and [3H]acetate from Lys16 were equal throughout the time course of the HDAC assay, indicating that both acetylated residues were substrates for deacetylase activity. More recently, studies with varying degrees of acetylated H4 tail peptides (residues 1–36) have revealed the multivalent engagement of multiply acetylated lysines by two tandem bromodomains in the RNA polymerase II transcription factor D (TFIID) complex protein TAFII250 (12). Acetylated H4 tail peptides have also revealed an unprecedented cooperative binding of two acetyl groups at H4 Lys5 and Lys8 by a single bromodomain module (13). Biotinylated Lys4-methylated H3 tail peptides (residues 1–20) have also been employed to identify the proteins WDR5 and BPTF as readers for the subtly different H3 Lys4-dimethylated and Lys4-trimethylated states, respectively (14).

The generation of peptide libraries has allowed high-throughput screening of histone-protein interactions in a microarray format. Toward this goal, Bedford and co-workers (15) have designed a chromatin-associated domain array (CADOR) chip containing an array of immobilized glutathione S-transferase-tagged histone-binding domains, including tudor and MBT domains, bromodomains, and chromodomains. Binding experiments with fluorophore-tagged N-terminal peptides from H3 and H4 bearing varying sites and degrees of methylation revealed novel interactions with chromodomains...
MINIREVIEW: Studying Histone Modifications

An impressive example of the combinatorial power of peptide synthesis was reported by Denu and co-workers (17), who developed a one-bead one-compound combinatorial library of 800 peptides bearing all possible permutations of the known modifications within the 21 N-terminal amino acids of histone H4. Peptide modifications included phosphorylation, acetylation, citrullination, and all possible methylation states of Lys.

An alternate approach developed by Neumann et al. (22) utilized an evolved Methanosarcina barkeri pyrrolysyl-tRNA synthetase and its cognate amber suppressor, tRNA_{Cua}, to genetically incorporate Nε-acetlysyllysine in response to the TAG codon in *E. coli* (Fig. 2A). This was employed to generate H3 acetylated at Lys, which underwent phosphorylation at Ser by the Aurora B kinase. A potential limitation of this methodology may be the well-established absence of diastereoselectivity in non-enzymatic thiol additions to Dha (20), although the local protein conformation may significantly influence the final diastereomeric ratio of the addition products (21).

Amber Suppression Strategies

An understanding of the physiological roles for histone modifications requires the ability to study them in the context of nucleosomes and chromatin. Although peptide models are often sufficient for studies of binary protein interactions, they cannot address the effects of modifications on directing *trans*-tail histone modifications, on nucleosomal or higher order chromatin structure, and on chromatin-remodeling complexes.

Amber suppression mutagenesis of proteins with orthogonal pairs of amber suppressor tRNAs and their cognate aminoacyl-tRNA synthetases could be employed to incorporate modified amino acids into full-length histones. Toward this goal, Schultz and co-workers (19) have evolved a mutant *Methanococcus janaschii* tyrosyl amber suppressor tRNA, TytMtRNA_{cua}/tyrosyl-tRNA synthetase pair to specifically incorporate (Se)-phenylselenocysteine in response to the amber TAG codon in *Escherichia coli* (Fig. 2A). Oxidative elimination of phenylselenolic acid yielded Dha (20), which underwent Michael addition with N-acetylated or N-methylated derivatives of 2-aminoethanethiol to produce the thiol-containing analogs of N-acetylated and N-methylated lysine (Fig. 2B). This methodology was employed to generate an analog of histone H3 acetylated at Lys, which underwent phosphorylation at Ser by the Aurora B kinase. A potential limitation of this methodology may be the well-established absence of diastereoselectivity in non-enzymatic thiol additions to Dha (20), although the local protein conformation may significantly influence the final diastereomeric ratios of the addition products (21).

Recently, Chin and co-workers (23) have extended their methodology to incorporate N-methyl-L-lysine into H3 at
position 9, masked as the N-ε-Boc-N-ε-methyl-l-lysine derivative, which is an efficient substrate for the M. barkeri pyrrolysyl-tRNA synthetase/tRNA_{CUA} pair (Fig. 2A). Subsequent acidolytic deprotection of the Boc group generated H3 monomethylated at Lys9 (Fig. 2B), which was demonstrated to bind an anti-H3-K9me2 antibody and HP1 (heterochromatin protein 1).

Cysteine Modification Strategies

Cys is the most convenient amino acid for selective modification because of its highly nucleophilic side chain sulfhydryl group (pK_a ∼ 8.5). Shokat and co-workers (24) have taken advantage of the unique reactivity of Cys to generate N-methylated aminoethylcysteine residues (Fig. 2C). Coupled with mutation of the single Cys in H3 (Cys110) to Ala, their methodology permitted the site-specific installation of MLAs in histone H3. Thiol-containing analogs of H3 methylated at Lys9 (H3-K_C9me2), Lys8, Lys36, and Lys79 and of H4 methylated at Lys20 were successfully recognized by methylation-specific antibodies. Furthermore, both an H3-K_C9me2 peptide (residues 1–14) and nucleosome-associated full-length H3-K_C9me2 were demonstrated to bind the heterochromatin-binding protein HP1α, which is known to bind H3-K9me1. Similar degrees of methylation of H3 Lys9 and the H3-K_C9 analog by the methyltransferase SUV39H1 demonstrated the equivalence of MLAs in biochemical assays. The ease of access to MLAs in histones also permitted structural determination of mononucleosomes where both copies of the native histones were substituted with either H4-K_C20me3 or H3-K_C79me2 (25). In either case, methylation did not significantly affect mononucleosome structure. In solution, however, sedimentation velocity analysis indicated that 12-mer nucleosomal arrays reconstituted with octamers containing H4-K_C20me3 had an enhanced propensity to form maximally folded and higher order structures with increasing Mg_2+ concentrations compared with H3-K_C79me2-containing or wild-type nucleosomal arrays. MLAs have also been applied in studies of DNA replication (26), recruitment of HDACs to histones (27), the propagation of repressive modifications (28), and cross-talk between modifications (29).

An alternate Cys-directed modification strategy developed by Davis and co-workers (30) is the oxidative elimination of Cys to Dha by the reagent O-mesitylenesulfonylhydroxylamine. Site-specific Cys mutagenesis in histones could in principle be coupled with this methodology to generate Dha, which would readily be converted to methylated and acetylated lysine analogs, similar to the methodology reported by Schultz and co-workers (19).

Native Chemical and Expressed Protein Ligation

The first report of EPL in 1998 by Muir et al. (31) marked a new avenue for employing chemistry to explore protein function. EPL extends the synthetic technique known as native chemical ligation (32), whereby two peptide halves, one bearing an N-terminal Cys and the other bearing a C-terminal thioester, are joined by a native amide bond. The peptide fragments can be obtained by solid-phase peptide synthesis employing either Boc or Fmoc (N-(9-fluorenyl)methoxycarbonyl) protecting group chemistry (33). Given the inevitable limitation of native chemical ligation by the length of the synthetic peptide halves, EPL significantly expands the range of proteins accessible for chemical modification. EPL employs an expressed protein in its N-terminal Cys and the other bearing a C-terminal thioester. Most importantly, because any desired protein modification may be introduced in the synthetic half of the ligation partners, in principle, EPL permits the incorporation of any histone modification observed in Nature.
**MINIREVIEW: Studying Histone Modifications**

Owen-Hughes and co-workers (37) employed semisynthetically tetraacetylated derivatives of H3 (Lys9, Lys14, Lys18, and Lys23) and H4 (Lys5, Lys8, Lys12, and Lys16) to generate specifically modified chromatin templates to study the interaction of various yeast remodeling complexes with differentially acetylated nucleosomes. It was shown that the ATP-dependent remodeling enzyme complex RSC (remodel the structure of chromatin) preferentially remodeled chromatin containing tetraacetylated H3, but not H4, ~16-fold faster than unmodified chromatin. Kinetic analysis revealed this to be due to a 3-fold lower $K_m$ for the tetraacetylated nucleosomes and that a single acetylation at H3 Lys14 contributed most to this effect. On the other hand, tetraacetylation of H4 inhibited nucleosome remodeling by the Isw2 enzyme by ~1.5-fold relative to unmodified nucleosomes. This was due to a reduction in the rate of ATP hydrolysis ($k_{cat}$) by Isw2. These experiments categorically demonstrated that histone modifications affect nucleosome remodeling through distinct pathways for different remodeling enzymes.

EPL has also enabled investigations of the effects of acetylation near the nucleosome dyad pseudo-symmetry axis, where key histone-DNA contacts occur (38). Results from nucleosome competitive reconstitution experiments revealed that acetylation at H3 Lys15 near the dyad axis reduced DNA binding significantly more than acetylation at H3 Lys122. However, mononucleosomes acetylated at H3 Lys122 underwent thermal repositioning about twice as fast as those acetylated at Lys115. These results suggest that the different sites of acetylation near the nucleosome dyad may have different physiological consequences *in vivo*, such as their effect on genome positioning and nucleosome assembly/disassembly. Another interesting result from this study was the observation that the Lys-to-Gln mutation that is commonly employed to mimic lysine acetylation *in vivo* showed significant differences from acetyllysine *in vitro* competitive reconstitution assays.

Our own laboratory has reported several advances toward the semisynthesis of histones for biochemical studies. Chiang *et al.* (39) have demonstrated the utility of a thiol-protected 2-hydroxy-3-mercaptopropionic acid linker (40) and a diaminobenzoic acid linker (41) to synthesize a histone H2B N-terminal peptide thioester containing pSer at position 14 as well as acetyllysines at positions 5, 11, 12, and 15. These synthetic strategies avoided epimerization at the peptide C terminus, which is known to occur during the solution-phase activation of side chain-protected peptides (42). Furthermore, a mild radical-based desulfurization methodology specific for cysteine (43) was employed to render the ligation product traceless. The phosphorylated and polyacetylated full-length H2B was probed with a commercial H2B pSer14-specific antibody. Surprisingly, the presence of multiple acetylations in the H2B tail prevented the recognition of pSer14 by this antibody. This result indicates a limitation of antibody-based ChIP (chromatin immunoprecipitation)-on-chip experiments for whole genome analysis because histones are often decorated with multiple modifications *in vivo* that may interfere with the detection of specific modifications. However, the acetyl groups did not interfere with phosphorylation of Ser14 by the human MST1 (mamma-
Semisynthesis of uH2B. A two-step ligation strategy followed by a final desulfurization step yielded native uH2B (5).

FIGURE 4. Semisynthesis of uH2B. A two-step ligation strategy followed by a final desulfurization step yielded native uH2B (5).

allowed ligation with cysteine both at the lysine side chain and within H2B. Avoiding ligation onto a sterically hindered secondary amine, which was unavoidable in the ligation auxiliary approach, led to significantly reduced reaction times and higher yields. Furthermore, substitution of UV irradiation with chemical unmasking of the second cysteine greatly facilitated sample handling. Simultaneous desulfurization of both cysteines yielded the u(G76A)H2B protein, which was recognized by a uH2B-specific antibody and by the ubiquitin-specific hydrolase UCHL3. u(G76A)H2B also stimulated robust methylation at H3 Lys79 by human DOT1L, thus demonstrating similarity to native uH2B in a nucleosomal context. Subsequent kinetic and structure-activity relationship analyses with u(G76A)H2B have revealed a non-canonical role for ubiquitin in the enhancement of the chemical step of H3 Lys79 methylation. In particular, the hydrophobic patch on the surface of ubiquitin centered around Ile44, which forms critical interactions with most helical ubiquitin-binding domains, was found to be non-essential for stimulation of human DOT1L activity. Mutagenic studies aimed at identifying the specific surface residues of ubiquitin involved in human DOT1L stimulation are currently under way in our laboratory.

Conclusions and Future Directions

Beginning with their discovery >40 years ago, histone modifications have been shown to play critical roles in directing key cellular events such as transcription activation, gene silencing, DNA damage repair, and DNA replication. Several semisynthetic methodologies for the generation of homogeneously modified histones have been developed, and these have led to investigations of the mechanistic roles for individual histone modifications in these processes. Advances in protein chemistry have made accessible synthetically challenging histone modifications, such as those found in the globular core domains of mononucleosomes and those involving large proteins such as ubiquitin and SUMO (small ubiquitin-like modifier). In the future, experiments with semisynthetic histones in our own laboratories will be aimed at the level of testing the histone code hypothesis in nucleosomal arrays bearing controlled modifications for studies of gene transcription, replication, and repair.

REFERENCES

1. Lugger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Nature 389, 251–260
2. Kossel, A. (1928) The Protamines and Histones, 1st Ed., Longmans, Green & Co., London
3. Stedman, E., and Stedman, E. (1950) Nature 166, 780–781
4. Butler, J. A. (1965) Nature 207, 1041–1042
5. Phillips, D. M. (1963) Biochem. J. 87, 258–263
6. Murray, K. (1964) Biochemistry 3, 10–15
7. Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964) Proc. Natl. Acad. Sci. U.S.A. 51, 786–794
8. Kouzarides, T. (2007) Cell 128, 693–705
9. Allis, C. D., Jenuwein, T., Reinberg, D., and Carparos, M.-L. (2007) Epige-
MINIREVIEW: Studying Histone Modifications

Hung, T., Binda, O., Champagne, K. S., Kuo, A. J., Johnson, K., Chang, H. Y., Simon, M. D., Kutateladze, T. G., and Gozani, O. (2009) Mol. Cell 33, 248–256

Bernardes, G. J., Chalker, J. M., Errey, J. C., and Davis, B. G. (2008) J. Am. Chem. Soc. 130, 5052–5053

Muir, T. W., Sondhi, D., and Cole, P. A. (1998) Proc. Natl. Acad. Sci. U.S.A. 95, 6705–6710

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Science 266, 776–779

Muralidharan, V., and Muir, T. W. (2006) Nat. Methods 3, 429–438

Shogren-Knaak, M. A., Fry, C. J., and Peterson, C. L. (2003) J. Biol. Chem. 278, 15744–15748

Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davies, J. R., and Peterson, C. L. (2006) Science 311, 844–847

He, S., Bauman, D., Davis, J. S., Loyola, A., Nishioka, K., Gronlund, J. L., Reinberg, D., Meng, F., Kelleher, N., and McCafferty, D. G. (2003) Proc. Natl. Acad. Sci. U.S.A. 100, 12033–12038

Ferreira, H., Faurs, A., and Owen-Hughes, T. (2007) J. Mol. Biol. 374, 563–579

Manohar, M., Mooney, A. M., North, J. A., Nakkula, R. J., Picking, J. W., Edon, A., Fishel, R., Poirier, M. G., and Ottesen, J. J. (2009) J. Biol. Chem. 284, 23312–23321

Chiang, K. P., Jensen, M. S., McGinty, R. K., and Muir, T. W. (2009) ChemBioChem 10, 2182–2187

George, E. A., Novick, R. P., and Muir, T. W. (2008) J. Am. Chem. Soc. 130, 4914–4924

Blanco-Canosa, J. B., and Dawson, P. E. (2008) Angew. Chem. Int. Ed. Engl. 47, 6851–6855

Goodman, M., and McGahren, W. J. (1967) Tetrahedron 23, 2031–2050

Van, Q., and Danishefsky, S. J. (2007) Angew. Chem. Int. Ed. Engl. 46, 9248–9252

Hershko, A., and Ciechanover, A. (1998) Annu. Rev. Biochem. 67, 425–479

Weake, V. M., and Workman, J. L. (2008) Mol. Cell 29, 653–663

McGinty, R., Kim, J., Chatterjee, C., Roeder, R., and Muir, T. W. (2008) Nature 453, 812–816

Chatterjee, C., McGinty, R. K., Pellois, J. P., and Muir, T. W. (2007) Angew. Chem. Int. Ed. Engl. 46, 2814–2818

McGinty, R., Kohn, M., Chatterjee, C., Chiang, K., Pratt, M., and Muir, T. W. (2009) ACS Chem. Biol. 4, 958–968

Bhaumik, S. R., Smith, E., and Shilatifard, A. (2007) Nat. Struct. Mol. Biol. 14, 1008–1016