Chemical approaches for studying histone modifications

Champak Chatterjee and Tom W. Muir

From the Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York, NY10065

Address correspondence to: Tom W. Muir, Ph.D., 1230 York Avenue, Flexner Hall Room 359A, New York, NY 10065; E-mail: muirt@rockefeller.edu

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 homogenously 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 review.

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 about 147 base-pairs of double-stranded DNA wrapped around an octameric protein complex comprised of two copies each of the four core histones, H2A, H2B, H3, and H4 (Figure 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 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 (3). 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 ε-N-methylation by Murray (6). As the protein synthesis inhibitor Puromycin did not inhibit these histone modifications, Allfrey et al. suggested that acetylation and methylation were post-translational modifications (PTMs) of histones (7). Based on 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 (Figure 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 their 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 in the following sections.

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 K16 was synthesized by automated solid-phase peptide synthesis. While a chymotryptic peptide from [3H]-acetylated H4 (residues 1-37) was deacetylated by HDACs purified from calf thymus, the hepatameric 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 K12 and [3H]-acetylated at K16 and subjected it to enzymatic deacetylation (11).
Interestingly, the relative amounts of [14C]-acetyl released from lysine 12 and [3H]-acetyl from K16 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 TAF$_{II}$250 (12). Acetylated H4 tail peptides have also revealed an unprecedented cooperative binding of two acetyl groups at H4 K5 and K8 by a single bromodomain module (13). Biotinylated H3 K4 methylated tail peptides (residues 1-20) have also been employed to identify the proteins WDR5 and BPTF as readers for the subtly different H3 K4 dimethylated and 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 have designed a chromatin-associated domain array (CADOR) chip containing an array of immobilized glutathione-S-transferase tagged histone-binding domains including tudor, bromo, chromo, and malignant brain tumor (MBT) domains (15). Binding experiments with fluorophore-tagged N-terminal peptides from H3 and H4 bearing varying sites and degrees of methylation revealed novel interactions with chromo, tudor and MBT domains from various chromatin associated proteins. Rathert et al. have utilized SPOT synthesis to generate arrays of as many as 420 mutant H3 tail-peptides (residues 1-21) and tested the substrate-specificity of the H3 K9 methyltransferase Dim-5 from Neurospora crassa (16). Results from these assays suggested an important role for the residues Thr 11 and Gly 12 in the H3 tail toward conferring specificity for Dim-5 activity and its discrimination against other lysines.

An impressive example of the combinatorial power of peptide synthesis was reported by Denu and co-workers 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 (17). Peptide modifications included phosphorylation, acetylation, citrullination and all possible methylation states of Lys and Arg. From the initial library, 512-members were used to elucidate the binding preferences of the double tudor domain of the human demethylase JMJD2A (hJMJD2A) for the H4 tail. Interestingly, binding hits with various combinations of modifications revealed a rheostat-like continuum of binding affinities for hJMJD2A from 1 µM to 1 mM. Finally, a SAMDI (self-assembled monolayer for matrix-assisted laser desorption-ionization) mass spectrometric assay was applied by Gurard-Levin and Mrksich to characterize the activity of HDAC8 toward H4 tail peptides. Their results indicated the importance of both distal residues (residues 16-19) as well as those immediately adjacent to acetylated K12 for HDAC8 activity (18).

**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. While 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 in full-length histones. Toward this goal, Schultz and coworkers have evolved a mutant Methanococcus jannaschii tyrosyl amber suppressor tRNA, Tyr $Mf$tRNA$_{ACU}$/tyrosyl-tRNA synthetase (MfTyrRS) pair to site-specifically incorporate (Se)-phenylselenocysteine in response to the amber TAG codon in E. coli (Figure 2A) (19). Oxidative elimination of phenylselenic acid yielded dehydroalanine (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 (Figure 2B). This methodology was employed to generate an analog of histone H3 acetylated at K9 that underwent phosphorylation at Ser 10 by the Aurora B kinase. A potential limitation of this methodology may be the well-established absence of diastereoselectivity in non-enzymatic

---

The text continues here.
thiol additions to Dha (20), although the local protein conformation may significantly influence the final diastereomeric ratios of the addition products (21).

An alternate approach developed by Neumann et al. utilized an evolved Methanosarcina barkeri pyrrolysyl-tRNA synthetase (MbPyRS) and its cognate amber suppressor, MbtRNA_{CUA} to genetically incorporate N-\(\varepsilon\)-acetyllysine in response to the TAG codon in E.coli (Figure 2A) (22). This was employed to generate H3 acetylated at K56 that was incorporated into mononucleosomes and nucleosomal arrays. Förster Resonance Energy Transfer (FRET) based experiments with fluorophore labeled nucleosomes permitted the direct observation of DNA unwrapping. Acetylation was found to increase the extent of DNA unwrapping within the last turn of DNA on the nucleosome core by 7-fold. It also accelerated nucleosomal repositioning by remodeling complexes about 20% over unmodified nucleosomes. However, acetylation did not affect the ATP-dependent H2A/H2B dimer transfer from mononucleosomes or higher order chromatin structure formation. Thus, the overall effects of H3 K56 acetylation on nucleosome structure and stability are fairly subtle and are manifest at the level of DNA breathing.

Recently Chin and co-workers have extended their methodology to incorporate N-\(\varepsilon\)-methyl-L-lysine in H3 at position 9, masked as the N-\(\varepsilon\)-tert-butyloxy-carbonyl-N-\(\varepsilon\)-methyl-L-lysine derivative which is an efficient substrate for the M. barkeri pyrrolysyl-tRNA synthetase/tRNA_{CUA} pair (Figure 2A) (23). Subsequent acidolytic deprotection of the tert-Butyloxy-carbonyl (Boc) group generated H3 mono-methylated at K9 (Figure 2B) which was demonstrated to bind an anti-H3 K9me\(_2\) antibody and heterochromatin protein 1 (HP1).

Cysteine modification strategies

Cys is the most convenient amino acid for selective modification due to its highly nucleophilic side-chain sulfhydryl group (pKa \(~8.5\)). Shokat and co-workers have taken advantage of the unique reactivity of Cys to generate N-methylated aminomethylcysteine residues (Figure 2C) (24). Coupled with mutation of the single Cys in H3 (C110) to Ala, their methodology permitted the site-specific installation of methyl lysine analogues (MLAs) in histone H3. Thiol containing analogues of H3 methylated at K9 (H3 K9me) or H4 methylated at K20 were successfully recognized by methylation-specific antibodies. Furthermore, both an H3 K9me\(_2\) peptide (residues 1-14) as well as nucleosome associated full-length H3 K9me\(_2\) were demonstrated to bind the heterochromatin binding protein HP1\(\alpha\) which is known to bind H3 K9me\(_2\). Similar degrees of methylation of H3 K9 and the H3 Kc9 analogue 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 Kc20me\(_3\) or H3 Kc79me\(_2\) (25). In either case, methylation did not significantly affect mononucleosome structure. In solution, however, sedimentation velocity analysis of 12-mer nucleosomal arrays reconstituted with octamers containing H4 Kc20me\(_3\) had an enhanced propensity to form maximally folded and higher order structures with increasing Mg\(^{2+}\) concentration than either H3 Kc79me\(_2\) 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 crosstalk between modifications (29).

An alternate Cys-directed modification strategy developed by Davis and co-workers is the oxidative elimination of Cys to Dha by the reagent O-mesitylenesulfonylhydroxylamine (MSH) (30). 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 et al (19).

Native Chemical and Expressed Protein ligation

The first report of expressed protein ligation (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 (NCL) (32) whereby two peptide halves, one bearing an N-terminal Cys and the other bearing a C-terminal thioester are joined by means of a native amide bond. The peptide fragments can be obtained by solid-phase peptide synthesis employing either Boc
or Fluorenylmethoxycarbonyl (Fmoc) protecting group chemistry (33). Given the inevitable limitation of native chemical ligation by the length of the synthetic peptide halves, expressed protein ligation (EPL) significantly expands the range of proteins accessible for chemical modification. EPL employs an expressed protein in its C-terminal thioester form, obtained by thiolysis of a C-terminally fused intein, rather than by synthetic means. The expressed protein thioester undergoes trans-thioesterification when reacted with a second peptide/protein bearing an N-terminal Cys (Figure 3). A subsequent S-to-N acyl-shift generates a native amide bond between the two halves leading to the full-length target. EPL may also be reversed to enable the ligation of a C-terminal expressed protein half, with an N-terminal synthetic peptide thioester. Most importantly, since 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.

Site-specific histone modification by EPL

Shogren-Knaak et al. first reported the use of EPL to generate the Xenopus laevis histone H3 bearing a phosphoserine residue (pSer) at position 10 (34). Nucleosomal arrays reconstituted with H3 pSer10 were efficiently remodeled by the yeast SWI/SNF remodeling complex demonstrating that semi-synthetic H3 pSer10 did not drastically affect nucleosome structure. These arrays were also used to probe the substrate specificity of the histone acetyl transferase (HAT) Gcn5 and revealed its different activities when presented with peptide versus nucleosomal substrates. This indicates the need for investigating the mechanisms of histone-modifying enzymes on intact nucleosomes rather than peptide substrates.

In another study, site-specifically K16-acetylated histone H4 was generated (35). This was incorporated in nucleosomal arrays and its effect on chromatin compaction determined by sedimentation velocity analysis during ultracentrifugation. The single acetylation of H4 K16 inhibited 30-nm fiber formation in nucleosomal arrays, similar to the absence of the histone tail. These results suggested a structural role for acetylated H4 K16 in establishing transcriptionally active euchromatic regions by decondensing chromatin.

EPL has permitted the generation of multiply modified histones with as many as three acetyl groups at positions K5, K8 and K12 in histone H4 and five acetyl groups at K4, K9, K14, K18 and K23 in H3 (36). An additional desulfurization step was added to these syntheses, converting the Cys introduced for ligation to a native Ala in the final step and rendering the ligation traceless. Histones are particularly amenable to chemical desulfurization since only a single Cys (Cys110 in H3) is present in the four core histones in higher eukaryotes, and Cys is altogether absent from yeast histones. Surprisingly, substituting wild-type H3 with semi-synthetic pentaacetylated H3 did not interfere with RSF-complex mediated assembly of a chromatinized plasmid. Additionally, H3-H4 tetramers generated with pentaacetylated H3 were also demonstrated to be a substrate for the HDAC Sir1 and subsequently for the methyltransferase G9a.

Owen-Hughes and co-workers employed semi-synthetically tetra-acetylated derivatives of H3 (K9, K14, K18, K23) and H4 (K5, K8, K12, K16) to generate specifically modified chromatin templates in order to study the interaction of various yeast remodeling complexes with differentially acetylated nucleosomes (37). It was shown that the ATP-dependent remodeling enzyme, RSC, preferentially remodeled chromatin containing tetra-acetylated H3, but not H4, about 16-fold faster than unmodified chromatin. Kinetic analysis revealed this to be due to a 3-fold lower $K_m$ for the tetra-acetylated nucleosomes and that a single acetylation at H3 K14 contributed most to this effect. On the other hand, tetra-acetylation of H4 inhibited nucleosome remodeling by the Isw2 enzyme by about 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 K115 near the dyad axis reduced DNA binding significantly more than H3 K122 acetylation. However, mononucleosomes acetylated
at H3 K122 underwent thermal repositioning about twice as fast as those acetylated at K115. 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 acetylysine in in vitro competitive reconstitution assays.

Our own laboratory has reported several advances toward the semi-synthesis of histones for biochemical studies. Chiang et al. have demonstrated the utility of a thiol-protected 2-hydroxy-3-mercaptopropionic acid linker (39) and a dianimobenzoic acid linker (40) to synthesize a histone H2B N-terminal peptide thioester containing pSer at position 14 as well as acetyl lysines at positions 5, 11, 12 and 15 (41). 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-on-chip experiments for whole genome analysis since 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 mammalian sterile twenty-like (Mst1) kinase, suggesting an unusual mechanism for H2B recognition by this enzyme.

One of the most dramatic post-translational modifications of proteins is their conjugation with the small protein ubiquitin. Ubiquitylation is undertaken by a family of E1-E3 ligases that activate the C-terminus of ubiquitin and catalyze its condensation with specific lysine side-chain ε-amines in target proteins (44). Unlike its typical role in proteasome assisted degradation, the ubiquitylation of histones is associated with DNA damage repair, and both transcription elongation and repression (45). Ubiquitylation of H2B occurs at K123 in yeast (K120 in humans) and has been linked to transcription elongation and trans-tail methylation of the H3 residues K4 and K79, through genetic studies (45). However, the small fraction (1-2%) of ubiquitylated H2B (uH2B) and the heterogeneity of histone modifications in vivo posed a serious challenge toward the isolation of uH2B for biochemical studies aimed at identifying its mechanistic role in H3 methylation.

We have developed several semi-synthetic approaches for the site-specific ubiquitylation of histones. McGinty et al. have reported an EPL strategy to access uH2B (46) which employs traceless peptide ubiquitylation (47). In this synthetic scheme (Figure 4), the H2B protein was divided into a short synthetic C-terminal fragment (1) and recombinant N-terminal thioester (4). In the first step, ubiquitylation of the H2B C-terminal peptide (1) was accomplished with a photolytically removable ligation auxiliary that was coupled to the side chain of the residue corresponding to Lys120 in full-length H2B. The ligation auxiliary acted as an N-terminal cysteine surrogate and facilitated EPL with a recombinant ubiquitin thioester lacking its C-terminal residue Gly76 (2). Upon ligation, photolysis of the auxiliary with UV irradiation yielded native uH2B peptide (3) and simultaneously released a photo-protected Cys at the N-terminus of the H2B peptide. In a second ligation step, the ubiquitylated peptide was reacted with the recombinant H2B thioester (4) to generate full-length uH2B(A117C). In a final step, the Cys was desulfurized to yield native uH2B (5). Biochemical assays of chemically ubiquitylated mononucleosomes with the histone lysine methyltransferase, hDot1L, revealed that ubiquitylation directly stimulates intranucleosomal methylation of H3 K79 (46). This was the first direct biochemical evidence of crosstalk between PTMs on different histones.

Very recently, we reported an alternate methodology for histone ubiquitylation that bypasses the need for the synthetically challenging ligation auxiliary (48). This was achieved by incorporating the mutation Gly76Ala at the ubiquitin C-terminus which allowed ligation with cysteine at both 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 by 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 UCH-L3. u(G76A)H2B also stimulated robust methylation at H3 K79 by hDot1L thus demonstrating similarity with 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 K79 methylation. In particular, the hydrophobic patch on the surface of ubiquitin centered around Ile 44 which forms critical interactions with most helical ubiquitin binding domains was found to be non-essential for stimulation of hDot1L activity. Mutagenic studies aimed at identifying the specific surface residues of ubiquitin involved in hDot1L stimulation are currently underway in our laboratory.

Conclusions and Future Directions

Beginning with their discovery more than forty 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 semi-synthetic methodologies for the generation of homogenously 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. In future, experiments with semi-synthetic 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


**Figure Legends**

**Figure 1. Histones and their post-translational modifications.** *A*, Mononucleosome structure with histone tails protruding from the core. Generated from PDB code 1KX5 with PyMoL. Colors used for individual histones are the same as in *B*. *B*, Schematic representation of human histone tails and their modifications. Some modifications at the histone C-terminus and globular core are also shown. ac = acetyl, Cit = citrull, me = methyl, ph = phosphoryl, pr = propionyl, rib = ADP-ribosyl, Ub = ubiquityl, groups. Adapted from (49).

**Figure 2. Amber suppression and cysteine-specific modification techniques.** *A*, *In vivo* amber suppression methodology to incorporate PTMs in histones. 1 = (Se)-phenylselenocysteine, 2 = N-ε-tert-butyloxycarbonyl-N-ε-methyl-L-lysine, 3 = N-ε-acetyl-L-lysine, MeLys = methyl lysine, AcLys = acetyl lysine. *B*, Conversion of 1 and 2 to modified lysine analogs after incorporation in histones. *C*, *In vitro* cysteine-modification strategy for generating methyllysine analogues (MLAs).

**Figure 3. Expressed Protein Ligation.** Multiple strategies for incorporating synthetic peptides into full-length proteins.

**Figure 4. Semi-synthesis of ubiquitylated histone H2B.** A two-step ligation strategy followed by a final desulfurization step yielded native ubiquitylated H2B, 5.
Figure 1.
A. 

\[
\begin{align*}
1 & : \text{H}_2\text{N} - \text{Se} - \text{OH} \\
2 & : \text{H}_2\text{N} - \text{Me} - \text{Boc} - \text{OH} \\
3 & : \text{H}_2\text{N} - \text{Ac} - \text{NH} - \text{N} - \text{Me} - \text{H}
\end{align*}
\]

B. 

\[
\begin{align*}
\text{R} = \text{Boc} & : \text{MeLys} \\
\text{R} = \text{Se} & : \text{AcLys/MeLys analogues}
\end{align*}
\]

R₁ = acetyl, R₂ = R₃ = H  
R₁ = R₂ = Me, R₃ = H  
R₁ = R₂ = R₃ = Me

C. 

\[
\begin{align*}
\text{Cl} - \text{NH}_2\text{HCl} & \xrightarrow{\text{Base}} \text{Lys analogue} \\
\text{Cl} - \text{NHMeHCl} & \xrightarrow{\text{Base}} \text{MeLys analogue} \\
\text{Cl} - \text{NMe}_2\text{HCl} & \xrightarrow{\text{Base}} \text{Me}_2\text{Lys analogue} \\
\text{Br} - \text{NMe}_3\text{Br}^- & \xrightarrow{\text{Base}} \text{Me}_3\text{Lys analogue}
\end{align*}
\]

Figure 2.
Figure 3.
Figure 4.