Chromatin is the universal template of genetic information in all eukaryotic cells. This complex of DNA and histone proteins not only packages and organizes genomes but also regulates gene expression. A multitude of posttranslational histone modifications and their combinations are thought to constitute a code for directing distinct structural and functional states of chromatin. Methods of protein chemistry, including protein semisynthesis, amber suppression technology, and cysteine bioconjugation, have enabled the generation of so-called designer chromatin containing histones in defined and homogeneous modification states. Several of these approaches have matured from proof-of-concept studies into efficient tools and technologies for studying the biochemistry of chromatin regulation and for interrogating the histone code. We summarize pioneering experiments and recent developments in this exciting field of chemical biology.

Addresses
1 Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany
2 Institute of Biochemistry, University of Muenster, 48149 Muenster, Germany
3 Interfaculty Institute of Biochemistry, University of Tübingen, 72076 Tübingen, Germany

Corresponding authors: Fischle, Wolfgang (wfischl@gwdg.de), Mootz, Henning D (henning.mootz@uni-muenster.de) and Schwarzer, Dirk (dirk.schwarzer@uni-tuebingen.de)

Chromatin and posttranslational histone modifications

The physiological template of genetic information in all eukaryotic cells is chromatin, the complex of DNA and histone proteins. As signaling platform, chromatin integrates a variety of internal and external cellular inputs. These direct distinct local and global structural and functional states of chromatin, thereby controlling gene expression [1].

In the repeating unit of chromatin, the nucleosome, DNA is wrapped around an octamer of histone proteins (two copies each of the core histones H2A, H2B, H3, and H4 and one copy of linker histone H1, Figure 1a) [1]. By various levels of folding, the originating primary chromatin fiber can be organized into various structural arrangements, including highly condensed mitotic chromosomes (Figure 1b). While the basic architecture of nucleosomes is the same throughout the genome, posttranslational modifications (PTMs) of histones are central means of increasing the biochemical divergence that regulates chromatin structure and function [2,3]. Histone PTMs are structurally diverse and include methylation, acetylation and ubiquitinylation of lysine, as well as phosphorylation of serine and threonine residues (Figure 1c). More than 150 histone modification sites have been identified in different experimental systems. Major sites of modification cluster within the unstructured regions of the N-terminal histone tails that vary between 10 and 35 amino acids in length. These are protruding out from the nucleosome core (Figure 1a).

Research over the past years has focused on defining the distribution, biochemistry and cellular function of individual and combinations of histone marks. A ‘histone code’ hypothesis has been put forward that defines chromatin as a dynamic programming platform, which integrates internal and external cellular signals [2,3]. Histone modifying enzymes are considered the writers of the histone code. Chromatin factors possessing PTM-binding domains serve as histone code readers and execute regulatory functions upon recruitment. Finally, dedicated PTM-removing enzymes are the erasers of histone PTMs [4–7].

Deciphering the complex cross-talk between histone PTMs, chromatin factors and gene regulation is a major research challenge [8]. While the accessibility of ‘designer chromatin’ composed of histones with defined and homogeneous modification states has been a central bottleneck in chromatin research, novel chemical approaches are providing powerful tools enabling important discoveries of the rules of the histone code [4–7,9–12].

Tools for designer chromatin

Chromatin research has benefited strongly from peptide chemistry. Especially, the availability of building blocks of modified amino acids has enabled the synthesis of histone peptides with defined modification patterns. Such histone-derived peptides have been instrumental for many key discoveries on the readers, writers and eraser of histone PTMs. However, histone-peptides can only recapitulate a minor fraction of the complex chromatin structure (Figure 1). Consequently, major efforts have been undertaken to site-specifically introduce...
modification marks into full-length histones, nucleosomes, arrays of nucleosomes and even into chromatin of living cells. These approaches draw from the technological repertoire of modern protein chemistry [4–7].

**Selective modification of cysteine residues**

The most widely used strategy for generating modified histones is the selective conversion of cysteine residues into mimics of modified amino acids. The chemical properties of the cysteine thiol group are unique among the proteinogenic amino acids. Selective alkylation of this soft nucleophile has been used in traditional biochemistry and this approach has recently experienced a renaissance for site-specific introduction of mimics of PTMs. Since natural cysteine residues are absent in core histones with exception of a single residue in H3, this method is particularly attractive for chromatin chemistry. In a pioneering report site-directed mutagenesis was used to introduce cysteine at sites of methylated lysine residues in otherwise Cys-free mutant histone H3 as well as histone H4 [13]. Upon treatment with 2-chloro-ethyl-methylanmonium or bromo-ethyl-methylanmonium compounds (1) the respective methyl-thialysine (2) residues were generated in high yields (Figure 2a). Downstream experiments showed that these so-called methyl-lysine analogs (MLAs) mimicked properties of methylated lysine in peptide and nucleosomal context by recruiting chromatin readers.

In a similar fashion methyl-arginine analogs (4) were generated [14]. In this case α,β-unsaturated amidines (3) served as alkylation agents. Extending this approach to acetyl-thialysine (6) turned out much more challenging [15]. The treatment of cysteine with 2-bromoethyl-acetalamine (5) or with more reactive N-acetyl-aziridine, were either poor yielding or led to undesired side-products. Alkylation with methyliocarbonyl-aziridine (7) converted cysteine residues into methyliocarbonyl-thialysine (8), which mimicked some aspects of acetylated lysine [15]. Efficient formation of acetyl-thiaLys (6) was finally achieved with N-vinyl-acetamidine (9), however not by direct conjugation but via the radical thiol-ene
Chemical approaches for the generation of histones possessing defined modification patterns. (a) Modification of cysteine residues engineered into recombinant histones can be modified into mimics of amino acids carrying posttranslational modifications. (b) Recombinant modified histones are accessible by amber suppression technology. (c) Protein semisynthesis, here: native chemical ligation, affords homogenously modified histones by ligation of synthetic peptide-thioesters and globular folds of histones that possess an N-terminal cysteine residue. (d) Approaches for site-specific ubiquitination of lysine residues by NCL using thiol auxiliaries and 5-mercaptolysine. Both approaches can be traceless when the auxiliary is removed by UV irradiation or when the thiol is desulfurized.
reaction [16]. A promising alternative to the direct modification is the selective elimination of the cysteine thiol resulting in dehydroalanine (10), which can serve as acceptor for thiol-containing nucleophiles (11) in Michael reactions [17].

Despite the applicability of this approach it needs to be emphasized that thialysine and its derivatives are isosteric mimics but differ from lysine with respect to the $pK_a$ of side-chain amine as well as the molecular geometry at the thioether bond. Consequently, quantitative studies on methyllysine reader domains and lysine demethylases have uncovered significant differences between natural substrates and mimics in terms of binding and catalytic properties [18–20].

Site-specific histone ubiquitylation by disulfide formation represents a further approach useful for coupling small proteins to histones. Via an isopeptide bond the C-terminus of the 76 aa ubiquitin polypeptide is naturally linked to the ε-amino group of lysine. While the formed disulfide moieties differ significantly from the native lysine modification, this method is highly efficient in terms of synthesis. The disulfide-bound ubiquitin moieties were shown to mimic the natural modification state well [21]. Disulfides are not stable under reducing conditions, which can complicate reaction procedures. Nevertheless, the simple erasure of this modification mark by reducing agents might be useful in specific experimental settings.

Amber suppression technology

A powerful alternative approach to chemical side chain modification is direct genetic encoding of modified amino acids by the amber suppression technology (Figure 2b). An ‘empty’ stop codon (commonly the amber codon) is read during ribosomal translation by an extra tRNA that is co-expressed together with its cognate aminoacyl-tRNA synthetase (RS). The latter one is obtained by genetic engineering and protein evolution to recognize the desired amino acid [22]. The most widely used RS for chromatin-relevant modifications, PylRS, is responsible for esterification of the 22nd proteinogenic amino acid pyrrollysine to its cognate tRNA. It displays relaxed substrate selectivity and is able to process selected Nε-acyl lysine derivatives without engineering [23]. The repertoire of encoded non-natural amino acids has expanded vastly in recent years. Well-known examples are fluorescent amino acids, photo-crosslinkers and amino acids that carry special functional groups, which are otherwise absent from proteins [24]. The aldehydes, alkynes, azides or additional thiols enable downstream modification with probes or modifications including ubiquitin and SUMO [25–28]. Also, encoded amino acids that carry PTMs, such as acetyllysine, various acyllysines (propionyllysine, butyryllysine and crotonyllysine) derivatives and a protected mono-methyllysine have been already successfully introduced into histones [23,29–32]. A further important modified amino acid that was recently encoded is phospho-serine; in this case by hijacking an O-phosphoseryl-RS involved in cysteiny–tRNA synthesis of archaea [33,34*].

Protein semisynthesis

Protein semisynthesis is a collective term for chemoselective ligation strategies of synthetic peptides and recombinant proteins, which allow reconstitution of whole proteins. Chemical synthesis of one of these peptide fragments enables the introduction of structural variations with virtually no restrictions regarding their nature and number [35,36]. Native chemical ligation (NCL) and the related expressed protein ligation (EPL) are based on the selective ligation of thioesters with peptides or proteins possessing an N-terminal Cys residue (Figure 2c) [35,36].

In recent years NCL-based ligation approaches were applied to all of the four core histones. This includes strategies for introducing different PTMs in the tails of histones H3 and H4 as well as elaborate ligation schemes for site-specific coupling of ubiquitin at the C-terminus of histone H2B [37–43]. In the latter, a photo-cleavable thiol auxiliary (12) mediated NCL between an ubiquitin-thioester and the respective lysine side chain in a peptide derived from H2B (Figure 2d) [41]. In this case the C-terminal glycine residue of ubiquitin was linked to the lysine residue of the H2B peptide. The corresponding ubiquitin thioester was truncated by this C-terminal glycine, thereby placing the ligation site within the ubiquitin protein (Figure 2d). Subsequent cleavage of the auxiliary and uncaging of an N-terminal cysteine of the H2B peptide by UV light established the native isopeptide connection with ubiquitin and enabled a second NCL reaction with a recombinant N-terminal histone fragment containing a C-terminal thioester.

An alternative approach for chemically attaching the ubiquitin PTM to a lysine side chain employs the non-natural amino acid δ-mercaptolysine (13) [44*]. The additional thiol in the side chain enables NCL with the Nε amino group of this lysine derivative, which also establishes an isopeptide bond with ubiquitin thioesters (Figure 2d) [43]. A potential problem of NCL-like ligation schemes arises from the thiol-containing residue that mediates the reaction and remains in the ligation products as mutation. However, efficient and specific desulfurization strategies have been developed for removing the thiol moieties after the ligation [45]. Finally, it should be noted that NCL-like strategies have been applied for the total chemical synthesis of histones, which provides chemical control over each residue of the synthetic proteins [43,46,47].

In contrast to the chemical ligation approaches to reconstitute the peptide backbone of a protein from two fragments, sortase-mediated ligation and protein
**trans**-splicing by split inteins represent chemo-enzymatic technologies [48–51]. The bacterial transeptidase sortase A recognizes a highly conserved so-called sorting motif consisting of the amino acid sequence LPxTG (X: any amino acid) in one of the ligation partners. The enzyme cleaves the motif at the threonine residue under concomitant formation of a sortase-bound thioester, which is ligated to a peptide or protein possessing an N-terminal Gly residue [49,51]. Histone H3 contains an amino acid sequence, which differs in a single residue (APxTG) from the sorting motif and this sequence is located at the interface between tail and the globular fold. Protein evolution approaches have been put forth to generate a sortase ligator that ligates the H3 sequence rather than the native sorting motif, thereby enabling traceless semisynthesis of this histone [52].

Protein **trans**-splicing mediates the ligation between two polypeptides that bear at their C-termini and N-termini the N-terminal and C-terminal halves of a split intein (peptideN-inteinN and inteinC-peptideC) [48,50]. Since the intein parts remove themselves during the reaction and leave behind only a single Cys, Ser or Thr residue at the splice junction, the reaction can be designed in a traceless fashion. The other key advantage of split inteins is the inherent affinity between the two halves, thereby enabling protein reconstitution at low concentrations (down to nanomolar). Combined with the selectivity and bioorthogonality of the reaction, split inteins are highly attractive for applications in a cellular context [53,54].

**Designer chromatin in biochemistry and biology**

The availability of histones of defined modification status has enabled a level of **in vitro** biochemical investigation of chromatin structure and function that would not have been accessible by conventional approaches. Classically, chromatin fragments have been isolated from cells or mononucleosomes and oligonucleosomal arrays were reconstituted from recombinant DNA and purified histone proteins. While chromatin reconstitution using salt dialysis based methods is still central to **in vitro** chromatin biochemistry, the undefined and mixed nature of PTMs on histones purified from cells has in the past limited the mechanistic investigation of the effects of post-translational marks.

Mechanistically, histone PTMs can act in two non-exclusive ways. On one side, these may directly organize chromatin architecture (i.e. ‘compactness’ vs ‘openness’ of chromatin) by modulating internucleosomal contacts [55]. On the other side, the proteins specifically interacting with particular histone PTMs (readers) are thought to mediate downstream functions [6]. The establishment of designer chromatin carrying defined histone PTMs has made major contributions to our understanding of these processes [4–7].

In a landmark study it was shown that a single modification of histone H4 (H4K16ac) that is implicated in the dosage compensation process in flies impedes chromatin higher order folding (i.e. the transition of primary chromatin fibrils into folded structures) when uniformly incorporated into nucleosomal arrays (Figure 3a) [56]. Structural analysis and modeling implied the modulation of a contact of the highly positively charged H4 tail with an acidic patch within the folded core of H2A in this process. Surprisingly, it was reported that a neighboring acetylation mark at H4K20ac had very little effect on array condensation, indicating that only residues in close proximity of H4K16 are important for the interaction with H2A [32].

While ubiquitylation of the C-terminus of H2B (H2BK120ub) has been shown to have a similar, overall inhibiting effect on chromatin folding, the mechanism here appears to be distinct and mediated by the addition of the, compared to histones (100–150 aa), large and bulky ubiquitin moiety (76 aa) [57]. Other work also showed that H2BK120ub controls the binding of a large number of factors to chromatin [58].

Compared to H4K16ac, acetylation of a different site within the central, folded part of H3, H3K56ac was shown to have no direct impact on chromatin folding. Here, the effect appears to be on the stability (i.e. dynamics of interaction of histones with DNA) as well as on the enzyme-directed positioning of nucleosomes [59].

Different sites of histone lysine methylation have also been analyzed. For example, it was shown that dimethylation of lysine 79 of histone H3 (H3K79me2) locally alters the nucleosomal surface, whereas trimethylation of lysine 20 of H4 (H4K20me3) facilitates higher-order chromatin folding [60]. The examples illustrate the complexity of the histone code with the same PTMs on different histone sites having similar effects but also different PTMs on the same histone site having distinct effects.

Besides the analysis of the molecular impact of histone modifications, the chemical biology approaches of designer chromatin have enabled a new level of understanding the binding of proteins to histone marks. A large number of factors specifically recognizing histone PTMs has been identified and characterized on the level of histone peptide–protein domain complexes. These studies have provided fruitful insights into the modes of specific interaction. However, when comparing the set of factors that binds a given histone PTM in the context of a peptide with the set of factors that interacts with such marks in chromatin context on a proteome-wide level,
Application of designer chromatin. (a) The folding properties of nucleosome arrays can be analyzed in vitro. Single modifications of histones impair the condensation behavior. (b) Nucleosomes containing histones in distinct modification states can be used as baits for recovering the corresponding reader proteins from cell extracts. Mass spectrometry in combination with specifically labeled cell lysates allows the identification and quantification of nucleosome-bound reader proteins. (c) Synthetic histone tails can be targeted to nucleosomal arrays by PNA–DNA interactions. The modification state of the attached tails can modulate downstream modification of adjacent nucleosomes.
there is limited overlap (Figure 3b) [61*,62*,63]. The results imply that the availability and presentation of histone tails and their modifications is more complex than previously assumed.

Biochemical and structural work has begun to dissect the exact contributions of DNA, other histones and chromatin folding on the interaction properties of defined histone PTMs. For example, it was shown that trimethylation of lysine 9 in H3 (H3K9me3) directs alternative modes of highly dynamic interaction of heterochromatin protein hHP1β with the nucleosome [64]. Also, the parameters of binding of the PHF1 protein to the trimethylation mark of lysine 36 of H3 (H3K36me3) in nucleosomal context could be worked out. There, it could be shown that this interaction enhances nucleosome accessibility [65].

Finally, designer chromatin templates have enabled the study of the molecular effects and consequences of proteins interacting with histone modifications. For example, for the yeast protein Swi6 a nucleosome bridging mechanism depending on H3K9me3 that might be involved in the assembly and spreading of silenced, condensed (i.e. hetero-) chromatin was suggested [66].

**Chromatin chemistry is going complex**

While the use of designer chromatin for in vitro chromatin biochemistry is now becoming more and more routine, recent developments are extending the chemical toolbox towards more complex and in vivo chromatin research. For example, conjugates of modified histone tails and peptide nucleic acids (PNAs) for targeted delivery of PTMs were recently described [67]. In order to mediate regulatory functions, histone PTMs need to propagate through larger assemblies of nucleosomes. Prominent examples of this type of ‘modification spreading’ include the heterochromatin mark H3K9me3, which apparently propagates through recruitment of the heterochromatin binding protein 1 (HP1) and histone methyltransferase Su(var)3–9 that methylates H3K9 on neighboring nucleosomes. A similar mechanism is considered for H3K27me3, which is introduced by the polycomb repressive complex 2 (PRC2) that by itself is regulated by the modification status of H3K27. Histone tail peptides with either PRC2 activating or repressing modification marks were targeted to single nucleosomes or arrays of nucleosomes via conjugated PNAs directed against nearby DNA sequences. The close proximity of the activating and repressing marks were sufficient of either promoting or repressing H3K27 methylation on neighboring nucleosomes although the synthetic tails were only linked to the proximal DNA and not the nucleosomes (Figure 3c) [67].

A further major challenge for deciphering the histone code lies in uncovering the substrate selectivity of chromatin readers, writers and erasers, which need to be probed in high throughput based on the vast number of potential ‘histone codons’. The first library of designer nucleosomes representing a large number of combinations of histone PTMs was recently presented [68*]. In this approach each species of nucleosomes was identified by a unique DNA ‘bar code’ sequence. After affinity purification, for example via binding to a chromatin factor, next-generation sequencing of the DNA barcodes allowed qualitative and quantitative characterization of the interaction with the distinct modification patterns of the nucleosomes from the library. Modulation of the activity of histone code writers by preexisting modification marks could be analyzed in a similar way by using modification specific antibodies in the pull-down step.

Chromatin of living cells has been largely inaccessible to chemical modification approaches. In two recent studies this ‘in vivo barrier’ of chromatin chemistry has been breached. Amber suppression technology was used to establish non-natural crosslinking amino acids in chromatin of living cells [69**]. These investigations allowed the dissection of events that initiate chromatin condensation in mitoses mediated by orchestrated writing and erasing of histone marks. Another approach to manipulate chromatin in cells is based on the high affinity of split inteins [70**]. An intein-N fragment was expressed as a fusion to core of the histone H2B. The H2B-intein-N construct was efficiently assembled into chromatin in a mammalian cell line despite the size of the intein-N part (102 aa). The complementary intein-C fragment (35 aa) was prepared by solid-phase peptide synthesis together with a short tag sequence for detection by immunoblotting or fluorescence. Intracellular delivery of the intein-C-tag peptide was accomplished by conjugation to a cell-penetrating-peptide (CPP) and specific transfer of the tag sequence to H2B by protein trans-splicing was observed, thus opening a new way for direct chemical modification of chromatin. Using the same experimental design, ubiquitin as a PTM of the H2B C-terminal tail could be installed on chromatin. However, this experiment was performed with isolated nuclei, thus circumventing the barrier of the plasma membrane [70**].

**Conclusion**

Protein semisynthesis, amber suppression technology and MLAs have matured from proof-of-concept studies to versatile tools. These experimental strategies are now frequently used in chromatin research. Each method provides advantages and drawbacks in terms of convenience, flexibility (i.e. number and types of modifications) and in vivo applicability.

Installing MLAs and related amino acids mimics into recombinant histones appears straightforward and is widely used. However, the mimics do not cover all chemical properties of the natural modifications and this chemistry is not applicable in living cells. Protein semi-synthesis by chemical or chemo-enzymatic means is the
most flexible approach in terms of number and kinds of modification that can be introduced into a histone protein. However, it cannot always be used in a traceless manner. Also, it is mostly limited to in vitro studies, if not combined with techniques like microinjection for potential cellular applications. Intracellular protein trans-splicing is the only exception reported so far, as it can be applied in vivo when combined with cell penetrating peptides (CPPs) for delivery of the synthetic fragment [54]. Finally, modified amino acids can be efficiently incorporated inside cells without further alteration of the histone sequence using genetic code expansion. However, this approach is limited by the nature of available unnatural amino acids and the remaining difficulty to simultaneously incorporate different PTMs.

While we expect these methods to mature further, chemical control over the modification states of chromatin in living cells is still one of the major challenges in the field. The first steps to introduce chromatin alterations on a global scale are very promising in this regard. However, the selective introduction of defined modification patterns in locally restricted chromatin regions, such as individual promoters, remains a long-term objective. With the focus on new in vivo tools, it appears likely that chemical biology will pave the way for many more discoveries of the rules of the histone code.

Acknowledgments

We are grateful for financial support by the priority program SPP 1623 of the Deutsche Forschungsgemeinschaft (DFG). We apologize to the authors of the many recent interesting reports in the field of chromatin chemical biology that could not be covered in this review due space restrictions.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

● of special interest
• of outstanding interest


First genetically encoded phospho-amino acid


45. Method for the generation of a native isopeptide bond of the C-terminus of ubiquitin to a lysine residue.


64. See annotation to Ref. [61].


First library of designer nucleosomes for high-throughput chromatin research.


One of so far only two examples of in vivo chromatin chemistry.


See annotation to Ref. [69*].