

Probing chromatin-modifying enzymes with chemical tools

Wolfgang Fischle^{1,2,*} and *Dirk Schwarzer*^{3,*}

¹King Abdullah University of Science and Technology (KAUST), Environmental Epigenetics Program, Thuwal 23955-6900, Saudi Arabia

²Max Planck Institute for Biophysical Chemistry, Laboratory of Chromatin Biochemistry, Am Fassberg 11, 37077 Göttingen, Germany

³Interfaculty Institute of Biochemistry (IFIB), University of Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany

*Correspondence: wfischl@gwdg.de, wolfgang.fischle@kaust.edu.sa
dirk.schwarzer@uni-tuebingen.de

key words: chromatin, histone modifications, epigenetic, chemical probes, crosslinking, cosubstrate probes, CuAAC chemistry, substrate profiling, protein complex composition

Glossary section/Key words:

Epigenetics: Inheritance of gene expression profiles independent of changes in DNA sequence or content.

Histones: Small, basic proteins that package eukaryotic DNA into chromatin.

Nucleosomes: Nucleosome core particles consist of an octamer of histone proteins (2x H2A, H2B, H3, H4) around which 147 bp of DNA are wrapped. Addition of linker DNA of various length and linker histones (H1) makes up the nucleosome as the basic, repeating unit of chromatin.

DNA-Methylation: Enzymatic and none-enzymatic modification of DNA bases with methyl-groups. The methylation of C5 of cytosine is most relevant for gene regulation.

Histone modifications: Chemical, posttranslational modifications of specific amino acids of histone proteins. The most abundant modifications are acetylation and methylation of the ϵ -amino group of lysines, methylation of the guanidino group of arginines and phosphorylation of hydroxyl groups in serines and threonines.

Chemical probes: Synthetic molecules designed to interact with a protein of interest in order to study its function and activity.

Cosubstrate probes: Chemical probes that are derived from the cellular small molecules that donate modifying groups to DNA and histones such as Adenosine triphosphate (ATP), Acetyl coenzyme A (acetyl-CoA) and S-adenosyl methionine (SAM).

Substrate profiling: Proteomic approach to define the pool of target proteins of specific modifying enzymes using cosubstrate probes and mass spectrometry.

Abstract

Chromatin is the universal template of genetic information of all eukaryotic organisms. Chemical modifications of the DNA-packaging histone proteins and the DNA bases are crucial signaling events in directing the use and readout of eukaryotic genomes. The enzymes that install and remove these chromatin modifications as well as the proteins that bind these marks govern information that goes beyond the sequence of DNA. Therefore, these so called epigenetic regulators are intensively studied and represent promising drug targets in modern medicine. We summarize and discuss recent advances in the field of chemical biology that have provided chromatin research with sophisticated tools for investigating the composition, activity and target sites of chromatin modifying enzymes and reader proteins.

In all eukaryotic cells chromatin, the complex of DNA and histone proteins regulates the functional state of the genome by altering condensation states and active recruitment of regulatory factors. The basic structural unit of chromatin is the nucleosome core particle. It consists of a protein core formed by two copies each of the histone proteins H2A, H2B, H3 and H4 with 147 base pairs of DNA wrapped around.¹ Nucleosome core particles are connected by linker DNA, which can bind linker histones thereby giving rise to the fundamental repeating unit of chromatin, the nucleosome. Control of different chromatin states is mediated by a multitude of posttranslational modifications (PTMs) of the histone proteins, including methylation, acetylation and phosphorylation as well as methylation of DNA. Chromatin modifications constitute binding sites for chromatin factors thereby serving as a programming platform for integrating various cellular inputs and outputs.²⁻⁸

The enzymes that install and erase chromatin modifications as well as their binding proteins are essential for genome regulation and have entered the center stage of biomedical research as promising drug targets. Conventional biochemical methods have been instrumental for several key discoveries on chromatin regulation but appear insufficient for uncovering all aspects of the complex cross-talk between chromatin regulatory factors and cellular functions. Consequently, major efforts in the field of chemical biology have been made over the last years to develop specific and versatile tools for analyzing chromatin regulation.⁹

General strategies for studying the biochemistry of chromatin-modifications with chemical probes

The main strategy of enzyme- or binding protein chemical probes for chromatin-modifying enzymes follows the general concept of activity-based protein profiling (ABPP).^{10, 11} In the simplest setting such probes contain a reactive group, often referred to as ‘warhead’ that targets the active site of an enzyme or the binding site of a chromatin factor combined with a reporter that allows simple readout of this interaction (**Figure 1a and 1b**).^{12, 13} The ‘warheads’ are often derived from small-molecule inhibitors of enzymes and in some cases the modified amino acid. Probe-specificity is commonly mediated by integration of histone-derived peptides that serve as scaffold for presenting the reactive groups. In several cases, linkers that mediate specificity for an isoform or substrate site of the target enzyme or

chromatin modification-binding protein have been added. Photoactivatable crosslinking groups, which covalently link probe and enzyme upon UV irradiation (**Figure 1c**),^{12, 13} or ‘clickable’ groups such as alkynes or azides provide handles for analyzing the protein-probe interactions. The latter allow derivatization by bioorthogonal reactions such as Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) for linkage of probes with fluorophores, affinity tags or other biophysical labels.¹³

A second type of general probe is used for investigating the substrates of histone modifying enzymes that catalyze group-transfer reactions utilizing specific cosubstrates, such as acetyl-CoA or S-adenosyl-L-methionine (SAM). Cosubstrate probes donate unique reporter-groups to the transfer reactions, which modify the enzymes’ catalytic sites (**Figure 1b**). Since cosubstrate probes containing bulky modifications often are poor substrates for chromatin modifying enzymes ‘bump-and-hole’ strategies have been adopted. These are based on the idea of evolving complementary pairs of proteins and chemical probes in a way that the mutated enzymes only recognize cosubstrates that contain additional (bulky residues and/or electrostatic interphases) recognition moieties.¹⁴⁻¹⁷

Protein/histone methylation

Protein methyl transferases (PMTs) catalyze the transfer of the sulfonium methyl group of SAM to the side chain amino groups of lysine and arginine residues of specific protein substrates. There are more than 60 genes in the human genome that encode for protein methyl transferases. These include nine protein arginine methyltransferases (PRMTs) and more than 50 protein lysine methyltransferases (PKMTs). With the exemption of DOT1L and several METTL (methyltransferase like) homologues all PKMTs contain a characteristic, catalytic SET domain (about 130 aa; named for the founding enzymes SuVar(3-9), Ezh, Trithorax).

A particular feature of PMTs is their specific catalysis of substrate methylation to a defined degree. PKMTs mono-, di-, or tri-methylate the ϵ -amino groups of lysine residues (**Figure 2a**). While some enzymes are processive, mediating all methylation states, others catalyze only defined methylation reactions. All PRMTs are highly specific in the way these modify the ω -guanidino nitrogen of arginine residues. These mediate exclusively monomethylation, symmetric dimethylation or asymmetric dimethylation (**Figure 2b**). Single methylation events of histones have been linked to

different functional states of chromatin. For instance, di- and tri-methylation of lysine 9 of histone H3 (H3K9me_{2/3}) mediates transcriptional repression and the formation of heterochromatin. In contrast, tri-methylation of lysine 4 of H3 (H3K4me₃) is found in the promoter region of active genes. However, the mono-methylation state of the same site (H3K4me₁) is associated with gene repression. In a similar manner, differential methylation of arginine 3 of histone H4 (H4R3) by different PRMTs is linked to transcriptional activation or repression while methylation of arginine 2 of histone H3 is associated with gene silencing.

While most PKMTs were first characterized as histone methyl transferases, it is now clear that these enzymes have several if not many non-histone targets. Conversely, most of the PRMTs were initially defined outside the context of chromatin – especially in methylation of ribosomal proteins. However, most PRMTs have now been shown to also modify specific histone arginine residues.

Based on their major role in regulation gene expression and protein function it is not surprising that aberrant PMT activities are often associated with diseases, such as developmental abnormalities, neurodegenerative disorders, and cancer. Many protein methylation events are cell-type-specific, are induced by specific signaling events and occur at defined time-points in cell development and differentiation. Given the different roles of PMTs and their impact on distinct cellular processes, defining their various substrates (i.e. the total methylomes of cells or the methylome of defined PMTs) under different normal as well as disease biological conditions has become of high interest.

Many protein methylation events were initially analyzed by candidate approaches. In these, putative target factors of specific PMTs are labeled using radioactive SAM or modification sites are profiled using mass spectrometry. Also, attempts have been made to map the methylome of PMTs by reacting cellular extracts with recombinant enzymes or using libraries of putative substrates in the form of peptides or proteins spotted on arrays.^{18, 19} Due to an apparent lack of consensus sequences, it is impossible to rationalize target-recognizing patterns of PMTs. Further, the inert properties and small size of the methyl group make it difficult to directly tag or selectively bind methylation products.

Inspired by findings on DNA methyltransferases,²⁰ the first reaction that showed that PMTs could use SAM analogues as alternative cosubstrates involved PRMT1 and an *N*-mustard derivative of

SAM (**Figure 2c**). It was shown that the PMT can utilize 5'-(diaminobutyric acid)-*N*-iodoethyl-5'-deoxyadenosine ammonium hydrochloride (AAI, **3**) for labeling a histone H4 peptide containing the target arginine 3 residue. The reaction progresses via an aziridinium ion (**4**) as intermediate, in which the sulfonium-methyl group of SAM is replaced by an aziridine, and generates bisubstrate analogue **5** that serves as inhibitor of PRMT1.²¹ While the reaction is useful for generating PMT inhibitors, it does not allow characterization of novel PMT substrates.²²

To identify the substrates of PMTs, surrogates of SAM, in which the sulfonium-methyl group is replaced with functionalized chemical groups, have been developed (**Figure 1b**). These allow bioorthogonal reaction for labeling with fluorophores for visualization or linking to affinity tags for subsequent purification (**Figure 1b**). One approach made use of ketone-SAM derivative **6** for substrate modification and subsequent labeling with hydroxylamines and hydrazides (**Figure 2d**).²³

A more common strategy for the labeling and identification of targets of PMTs involves functional groups that are amenable to CuAAC.^{24, 25} Alkyl moieties with terminal alkyne groups that are enzymatically transferred to substrates from SAM analogues can be orthogonally reacted with azido-fluorescent (**7**) dyes or azido-azo-biotin (**8**) (**Figure 2e**). The latter enables enrichment of targeted proteins on streptavidin beads for subsequent identification. To this end, the azo-linker in **8** can be cleaved after mild treatment with sodium dithionite (Na₂S₂O₄) thereby greatly reducing the background of the affinity purification scheme.^{26, 27} Similarly, azide moieties that are transferred to substrates from SAM analogues can be used in strain-promoted azide-alkyne cycloaddition (SPAAC) using, for example, dibenzocyclooctyne (DIBO) derivatized fluorophores or biotin handles.

One of the first schemes using alkyne-bearing SAM analogues made use of S-(E)-prop-2-ynyl SAM (propargyl-SAM, **9**) that was shown to be used by recombinant, human H3K9 lysine methyltransferase SETDB1 for transfer of the propargyl moiety to isolated H3 for subsequent functionalization by click chemistry.²⁸ While propargyl-SAM (**9**) contains the smallest transferable handle for click chemistry, it, unfortunately, undergoes rapid decomposition at physiological pH, likely via an allene intermediate. Therefore its usability is limited. The stability issue was solved in subsequent work that replaced the sulfonium with selenium, which reduces the acidity of the methylene adjacent to the sulfonium/selenonium center (**Figure 2d**).^{29, 30} The resultant propargylic Se-

adenosyl-L-selenomethionine (ProSeAM) shows significantly improved half-life time under physiological pH and has been useful to profile substrates of a variety of PKMTs and PRMTs. Interestingly, for certain PMTs the alkyne-SeAM analogues are better cosubstrates compared to their SAM equivalents suggesting that sulfonium-to selenonium substitution can enhance reaction rates of otherwise less reactive compounds.³¹ More bulky alkyne-bearing SAM analogues are also more stable compared to propargyl-SAM. For example, S-(E)-pent-2-en-4-ynyl SAM (EnYn-SAM, **10**) can be utilized by the H3K9 methyltransferase Dim-5 from *Neurospora crassa* and the human H3K4 KMTs, MLL and MLL4.³²

The alkyne-bearing SAM analogues can only be efficiently used by a subset of PKMTs and the reaction rates are low. Although propargyl-SAM can be utilized by SETDB1, it is inert towards SET7/9, SMYD2, PRDM8, -10, and -16.²⁸ Systematic evaluation of the activities of several recombinant PMTs (PRMT1, PRMT3, CARM1, SUV39H2, SET7/9, SET8, G9a, and GLP) with different alkyne-SAM analogues (allyl-SAM, propargyl-SAM, S-(E)-pent-2-en-4-ynyl-SAM (EnYn-SAM, **10**), S-(E)-hex-2-en-5-ynyl-SAM (Hey-SAM, **11**), and 4-propargyloxy-but-2-enyl-SAM (Pob-SAM, **12**)) indicated that the bulky alkyne-SAM analogues are poor or no cosubstrates for the PMTs.³³ The limited reaction rates are likely due to steric hindrance within the SAM binding pocket.

To overcome this limitation and to obtain PMT-specific labeling of substrates a 'bump and hole'-like strategy has been adopted to the analysis of protein methylation. The applicability of this approach for the study of protein methylation was first demonstrated by the introduction of a mutation in the yeast methyltransferase RMT1 so that it would be selectively inhibited using bulky analogues of S-adenosyl homocysteine.³⁴ In the Bioorthogonal Profiling of Protein Methylation (BPPM) method put forward by the Luo laboratory,^{19, 35, 36} individual PMTs are mutated with the aim to gain the function of processing specific bulky alkyne-or azide-SAM analogues as alternative cosubstrates. In complex cellular milieu the native PMTs are incapable of binding the SAM analogues and only the engineered PMTs can utilize these engineered cosubstrates efficiently to modify their targets. After CuAAC- or SPAAC-based derivatization for affinity purification, proteins specifically methylated by the engineered PMTs can be unambiguously identified by mass spectrometry (**Figures 1b**).

The BPPM strategy has been successfully applied to define the methylome of several PTMs.

For example, it was shown that mutagenesis of conserved tyrosine residues of the PKMTs EuHMT1 (G9a) and EuHMT2 (GLP) results in substantial loss of activity. The mutant alleles process the bulky SAM analogues Hey-SAM and S-4-azidobut-2-enyl-SAM (Ab-SAM, **13**) with enzymatic efficiencies comparable to wild-type pairs.^{37, 38} Structural studies revealed that a conserved tyrosine acts as gatekeeper in the active site, preventing access of the bulky SAM analogues to a preexisting hydrophobic pocket.³⁷ The mutant EuHMT1 and EuHMT2 were also expressed in mammalian cells. The resulting cell extracts were incubated with Ab-SAM (**13**)³⁹ or Hey-SAM (**9**)³⁷ followed by SPAAC or CuAAC chemistry-based derivatization for affinity purification and proteomic analysis by mass spectrometry. BPPM strategies have also been applied to PRMT1⁴⁰ and PRMT3⁴¹ using Pob-SAM (**12**) as probe.

The limited overlap of the methylomes defined by BPPM for EuHMT1 and EuHMT2 with each other as well as generally with previously described substrates of these PMTs might imply differential interference of the bulky-SAM analogues with target selection or non-linear effects of *in vitro* labeling reactions in cell extracts compared to the *in vivo* reactions (i.e. in cell extracts the molecular interactions present *in vivo* are not recapitulated). Since the bulky SAM analogues – like SAM itself have poor permeability through cell membranes, a novel approach has been put forward to address this issue. In cells, SAM is synthesized by the transfer of methionine to ATP, a reaction that is catalyzed by different methionine ATP transferases (MATs). Wang et al. mutated human MAT2s to process membrane-permeable, bulky (E)-hex-2-en-5-ynyl homocysteine (Hey-methionine) into the corresponding Hey-SAM (**9**).⁴² Mutated MAT2 expressed in cells catalyzed the *in situ* biosynthesis of the SAM analogue. When combined with the expression of the previously described mutated EuHMT1 and EuHMT2 enzymes that are compatible with Hey-SAM efficient *in vivo* labeling of H3 could be shown.

While the approach seems applicable to the BPPM strategy for defining methylomes of PMTs, it allows another exciting *in vivo* application. The distribution of histone methylation marks over the genome is traditionally identified by chromatin immunoprecipitation (ChIP). The method involves crosslinking of histones to DNA followed by fragmentation of the genome into small pieces. Elements specifically containing particular histone methylation marks are then enriched via

immunoprecipitation using specific antibodies. Finally, underlying DNA elements can be identified via targeted PCR-based reactions or unbiased via high throughput DNA sequencing (ChIP-seq). Low specificity (i.e. via crossreactivity of antibodies or interference by adjacent sites) and low target recovery are significantly limiting the efforts to map different histone methylation marks over the genome using this method. Also, since multiple enzymes can mediate modification of a particular histone site, the ChIP approaches do not allow assignment of target sites to particular PMTs. With the cellular generation of clickable PMT substrates, the chromatin target sites of specific PMTs engineered to use the bulky-SAM analogues can be labeled. After derivatization of these elements with handles for affinity purification, this allows efficient and highly specific enrichment of chromatin regions for next generation sequencing (CliEn-seq) thereby allowing the mapping of chromatin regions modified by defined PMTs.⁴² Since expanded screens for pairs of MAT enzymes and membrane-permeable S-alkyl analogues of methionine have been conducted,⁴³ it is expected that the *in vivo* BPPM based methods can be expanded to PMTs beyond the paradigm EuHMT1 and EuHMT2.

The ‘bump and hole’ strategies for PMTs described to this point all involve derivatization of SAM on the sulfonium moiety. Yet, the Zhou laboratory showed that the viral SET-domain protein lysine methyltransferase (vSET) that is a homologue of the H3K27 methylating polycomb repressive complex 2 (PRC2) could be engineered to accommodate a 2',3'-dibenzyl SAM analogue, where the substitutions are on the ribose ring.⁴⁴

With the need for different SAM derivatives for analyzing PMT reactions, progress has been made in the synthesis of these compounds. Protection-free chemical strategies for producing large quantities of Se-adenosyl-L-selenomethionine analogues were recently described.³¹ Since SAM and many of its derivatives are not very stable and are prone to hydrolysis, the enzymatic *in situ* production of SAM analogues is promising. The enzymatic reactions also bypass the problem of racemization at the sulfonium or selenonium center that occurs in chemical synthesis. Usually, only the (*S,S*) diastereomer is biologically active, while the (*R,S*) diastereomer has been demonstrated to inhibit methyl transferases.⁴⁵ In one approach, a chemoenzymatic platform for making S- or Se-alkylated analogues of SAM using different MAT enzymes has been described.⁴⁶ Another study employed bacterial halogenases SalL and FDAS for the same purpose.⁴⁷ In both cases it could be

shown that the enzymatically produced SAM analogues can be directly used to characterize different PMT reactions.

A completely different application of chemical biology tools for analyzing PMTs is the identification of the interaction partners and potential regulators of these enzymes in physiological context. Classically, immunoprecipitation either using antibodies against the native protein or against an engineered tag are used for this purpose. However, these approaches cannot discriminate different activation states of the enzyme complexes and often produce high background. Liu et al. devised a chromatin activity-based chemoproteomic approach for defining the interactome of EuHMT1 under different cellular conditions.⁴⁸ For that, they immobilized the previously defined inhibitor UNC063826 that specifically binds the enzymatically active form of EuHMT1 to solid support using an amino propyl linker. The matrix was then used for affinity purification of EuHMT1 complexes. In combination with quantitative mass spectrometry (SILAC) this procedure allowed the differential identification of factors associated with EuHMT1 in inflammation response.

Protein/histone demethylation

Histone demethylases catalyze the removal of methyl groups from lysine (or arginine) residues in substrate proteins. Since methyl groups are chemically inert, their removal from amino-groups involves oxidative reactions. A family of more than 30 human genes that encode for proteins containing *jumonj* (Jmj) domains (KDM2 to KDM7) constitutes the main class of histone demethylases. *Jumonj* proteins belong to a larger family of non-haem iron(II) and 2-ketoglutarate (2-KG) dependent oxygenases.⁴⁹ The catalytic mechanism of lysine demethylation involves binding of the 2-OG cofactor and molecular oxygen. Oxidative decarboxylation of 2-KG results in succinate formation and release of CO₂ (**Figure 3a**). The formed highly reactive oxo-ferryl intermediate hydroxylates the methyl group of the N ϵ -side-chain of the targeted lysine. The hemiaminal intermediate is unstable and fragments into formaldehyde and a lysine side-chain decreased by a methyl group (**Figure 3a**). This reaction sequence allows demethylation of all possible lysine methylation states (mono-, di- and tri-).

While the majority of *jumonj* proteins has been classified as lysine demethylases, JMJD6 was initially described as the only known arginine demethylase.⁵⁰ However, other work classified this factor as lysyl hydroxylase of proteins associated with RNA splicing and also histones. Several other Jmj proteins are also thought to work as hydroxylases.⁵¹ Arginine methylation might not be reversible after all. Peptidyl arginine deiminases (PAD) have been shown to convert arginine residues in histones H3 and H4 to citrulline. It is unclear whether PADs work on methylated arginines or counteract arginine methylation via preemptive citrullination of target sites.⁵²

The LSD1 and LSD2 enzymes constitute the second class of lysine demethylases. These enzymes, also referred to as KDM1A and KDM1B, share homology with amino oxidases and use FAD as cofactor for demethylation of mono- and dimethylated lysines. Due to their particular reaction mechanism (**Figure 3c**) KDM1A and KDM1B are incapable of targeting tri-methylated states.

Several histone demethylases have been reported to have significant specificity towards their targeted residues. For example, free LSD1 targets H3K4me1/2, while in association with nuclear hormone receptors it targets H3K9me1/2. Proteins of the KDM6 class demethylate exclusively H3K27me and KDM3 enzymes work only on H3K9me. Other KDMs appear to be more promiscuous working on multiple sites in quite different sequence environment. For instance, the KDM7 enzyme PHF8 has been reported to demethylate H3K9me and H4K20me and the KDM4 (Jmjd2) enzymes appear to target H3K9me and H3K36me. In addition, non-histone substrates have been assigned to several KDMs. It needs to be noted, though, that the candidate site, *in vitro* reactions used to define the demethylation activities were not in all cases carefully analyzed (e.g. analysis of multiple putative substrate sites, kinetics, exact reaction product, etc.). Also, it is not in all instances clear whether the KDMs indeed are processive, i.e. removing two (KDM1) or all three (KDM2-7) methylation marks of their target lysine residues. Further, for several *jumonj* proteins no enzymatic activity has been assigned to this point. It, therefore, remains to be established what the true endogenous substrates and inferred biological functions of many KDMs are.

Given their counteracting activities of PMTs and their involvement in regulation of gene expression and other cellular processes, it is not surprising that aberrant activities and mutations of

KDMs are implicated in many diseases. Despite the important function of this class of enzymes, only very few chemical biology tools for their analysis have been reported.

A promising strategy for the identification and analysis of *jumonj* proteins targeting specific sites is based on *jumonj*-probe **14** (**Figure 3b**). Rationalizing that substrate recognition involves target as well as cofactor binding, it was shown that the KDM2 enzyme JHDM1A binds to a short peptide containing a dimethyllysine-2-KG mimic at its H3K36 target site.⁵³ By adding a benzophenone moiety for UV-induced crosslinking as well as a biotin handle, this probe could affinity purify JHDM1 from different sources, including cellular extracts. The method has the potential to systematically identify the *jumonj* proteins targeting different lysine residues as well as to define their tissue and cell-type specific interaction partners in demethylation reactions.

Other work on KDMs is based on the discovery that histone H3 peptide **15** containing a lysine 4 residue that is derivatized with a propargyl moiety function as specific inhibitors of LSD1.⁵⁴ It was shown that the N-propargylamines covalently link to the cofactor of the LSD1 aminooxidase, FAD (**Figure 3d**). While the finding has been further exploited for identifying and defining inhibitors of LSD1 (**Figure 3e**),⁵⁵ it was realized that covalent tethering of the FAD cofactor to the H3 peptide is an efficient strategy for capturing the transient LSD1-substrate complex. This not only aided in solving the X-ray crystallographic structure of the LSD1-CoREST complex⁵⁶ but also allowed studies on the interaction of LSD1 with nucleosomes.⁵⁷ In the latter the propargyl moiety was incorporated in H3 after mutagenesis of the lysine 4 residue to cysteine followed by alkylation and thioether formation, thereby generating a lysine mimic.

Binding proteins of methylated histones

In the past years numerous proteins that specifically bind methylated histones have been identified. After the initial discovery that proteins of the HP1 family bind H3K9me2/3 via a chromodomain, several more factors with homologous modules were defined. These also bind H3K9me or recognize other sites of methylation in histones. In addition, several other protein domains, including tudor, PHD, MBT, and WD 40 have been implicated in binding methylated histones. A general structural theme of these factors is the formation of hydrophobic pockets lined with aromatic residues ('aromatic cage')

for accommodating the methyl-groups. Methylation state selection is achieved by aromatic cages of different sizes as well as based on multiplicity of hydrophobic interactions. Some of the proteins containing the methyl binding domains are highly specific for particular sites of histone methylation. Others have more general recognition modes due to limited interaction with residues surrounding the methylated site. While a large number of methyl-lysine binding proteins have been characterized, only very few binding proteins of methylated arginines have been described.⁵⁸⁻⁶⁰

Many binding proteins of methylated histones were initially identified using candidate approaches on the basis of educated guesses. Also, peptide and protein arrays have been employed to define histone mark-reader pairs. More recently, systematic profiling of the interactomes of histone methylation marks was done applying synthetic peptides or nucleosome as baits in affinity purification schemes from cellular sources and in conjunction with mass spectrometry for protein identification.⁶¹ Due to the relative weak nature of the interactions between histones and their binding partners (Kd in the micromolar range), targets could be lost during the pull-down process. The weak interactions make it also difficult to distinguish direct binding partners from indirectly recruited factors (i.e. piggy packing on other proteins).

To circumvent these issues, a cross-linking-assisted and stable isotope labeling in cell culture-based protein identification (CLASPI) method has been put forward. In it, histone peptides are engineered to contain two functional handles besides the methylation mark. A photocrosslinker is installed proximal to the putative protein binding site and the C-terminus contains an alkyne moiety applicable to CuAAC chemistry (**Figures 1a**).⁶² It was first demonstrated that a benzophenone moiety appended to alanine 7 can crosslink known H3K4me3 readers, ING2 and BPTF to a H3 tail peptide. Also, functionalization of the bait via click chemistry on the alkyne group using an azide-Rhodamine fluorophore was shown.⁶³ Later, the method was expanded by derivatizing the alkyne with a biotin tag and combining it with quantitative mass spectrometry (SILAC). This way, binding proteins to H3K4me3, H3T3ph, H3K9me3 and the dual modification H3T3phK4me3 could be affinity purified and identified from cellular extracts.⁶⁴ Since the vast majority of proteins identified with H3K4me3 (>98%) appeared to be unspecific binders, the benzophenone photocrosslinker in the CLASPI

approach was recently replaced by a diazirine that improved photo-crosslinking rates, overall yields and specificities for capturing proteins on this methylation mark.⁶⁵

Protein/histone acetylation

Lysine acetyl-transferases (KATs) transfer acetyl moieties from acetyl-CoA onto the ϵ -amino group of lysine residues (**Figure 4**). Nuclear KATs can be divided into the following major families by sequence homology: Gcn5/PCAF (GNAT), MYST, p300/CBP and Rtt109.⁶⁶ The catalytic mechanisms of several KATs have been analyzed using kinetic, mutant and structural approaches. Gcn5 and Esa1 (a member of the MYST family) form ternary complexes of enzyme, acetyl-CoA and the substrate lysine residue.^{67, 68} Deprotonation of the substrate lysine by an active site Asp or Glu residue as general base appears crucial for catalysis and enables direct attack of the amine nucleophile onto the carbonyl carbon of acetyl-CoA. The tetrahedral intermediate collapses into acetylated lysine and CoA followed by the release of the products.^{67, 68} p300/CBP employs a reaction mechanism that differs from other KATs. At first the enzyme and acetyl-CoA form a complex followed by weak association of the lysine substrate, which is released immediately after acetylation.⁶⁹ This model is consistent with the broad substrate selectivity observed for p300/CBP as it does not rely on tight interactions with a specific peptide sequence of the substrate.

Besides enzymatic centers, KAT proteins commonly contain other domains that modulate function, activity, protein-protein interaction and chromatin targeting.⁷⁰ Further, most KATs form multi-subunit complexes. For example, the SAGA complex consists of Gcn5 and 10-20 additional proteins. These are important for the role of Gcn5 as transcriptional activator, which involves acetylation of nucleosomal histones.⁷¹ Isolated, recombinant Gcn5 only accepts free but not nucleosomal histones as substrates. In contrast, the additional factors in the SAGA complex mediate targeting of nucleosomes. Most of the nuclear KATs have promiscuous substrate specificities, targeting multiple sites in all four core histones. In addition, many non-histone targets of KATs have been described. A prime example for KATs with broad substrate specificity is p300/CBP with over 70 identified substrate proteins including itself.^{66, 70}

Acetylation alters the biochemical properties of target Lys residues by neutralizing a positive

charge. Besides direct consequences on protein conformation or chromatin folding, acetylated Lys residues constitute binding sites for binding modules.⁷²

Following the initial discovery on histones, proteome-wide studies have uncovered an ever-growing number of acetylated proteins.⁷³ Acetyl-lysine specific antibodies and high-performance mass spectrometry have greatly facilitated these developments. Recent findings indicate that a significant number of protein acetylation events are non-enzymatic.^{74, 75} In these cases the amino groups of accessible Lys residues react with activated acetyl-metabolites such as acetyl-CoA and acetyl-phosphate. Non-enzymatic reactions may also explain the broad range of acyl-modifications that have been identified on histone and non-histone proteins including fatty acids of various chain-length and precursors of fatty acid biosynthesis.

Distinguishing enzymatic from non-enzymatic lysine acetylation and acylation, uncovering their physiological relevance and defining how cells deal with non-enzymatic protein acylation are major challenges. Biochemical assays for analyzing protein acylation have been instrumental for characterizing modifying enzymes on single substrate basis as well as in high-throughput. New generations of enzyme- and substrate-probes represent promising tools that bypass some of the shortcomings of conventional strategies.

Determining the substrate selectivity of KATs within the complexity of acetylomes has been a major challenge (**Figure 4a**). Substrate probes have been reported that exploit the ability of some KATs to accept reactive acyl-CoA derivatives as substrate.⁷⁶ This strategy was initially employed for identifying substrates of the bacterial Gcn5/PCAF-related KAT RimL.⁷⁷ Chloroacetyl-CoA (CA-CoA, **16**) is a cosubstrate that RimL can utilize. It donates an electrophilic chloro-acetyl group to the transfer reaction, thereby enabling modification of the L12 protein in dialyzed bacterial lysates (**Figure 4b**). Downstream reaction with nucleophilic thiol-conjugated fluorophores enabled selective derivatization of chloroacetylated L12. The same approach was also applied to yeast HAT1 that exclusively chloro-acetylated its natural substrate histone H4 in a mixture of purified core histones.⁷⁷

The utility of electrophiles for downstream derivatization is, however, limited by potential side reactions with other cellular nucleophiles. This problem has been addressed with CoA analogues that donate bioorthogonal acyl-moieties to the transfer reactions. The two CoA-derivatives 4-pentynoyl

(4PY-CoA, **17**)- and 5-hexynoyl (5HY-CoA, **18**)-CoA were shown to be used as cosubstrates by the KAT p300. The reactions allowed *in vitro* alkyne-modification of histone H3 for subsequent bioorthogonal labeling by CuAAC (**Figure 4b**).⁷⁸ Further extension of this strategy to a proteomic scale by reacting nuclear extracts from HeLa cells with **17** followed by CuAAC with cleavable biotin-tag **8** enabled identification of p300 targets by LC-MS/MS.⁷⁹ Further, providing 4-pentynoate and 5-hexynoate to Jurkat T cells resulted in metabolic protein labeling, presumably by formation of **17** and **18** *in vivo* and subsequent acyl-transfer catalyzed by p300.⁷⁸

Unfortunately, other KATs including Gcn5 and MOF are more selective for the cosubstrate and do not accept clickable probes. Nevertheless, X-ray based structural insights guided ‘bump-and-hole’ strategies for both enzymes. By mutating active site residues that presumably collide with the bulky acyl moieties to either Ala or Gly residues a set of Gcn5 and MOF mutants was generated that readily utilized probes **16** - **21** as cofactors (**Figure 4b**).⁸⁰ Pairs of these mutant enzymes and cosubstrate probes allowed the selective labeling of target proteins in lysates of human 293T cells.

Probes derived from CoA have also been used for trapping KATs themselves as well as other CoA-utilizing enzymes. The probes are derived from the CoA scaffold and contain an electrophilic acceptor group, which is not transferred to the KAT substrate but reacts with a nucleophile of the enzyme. This concept has been realized with thiocarbamate sulfoxide probe **22** that was shown to transfer desthiobiotin covalently onto KATs and CoA-binding proteins, enabling subsequent enrichment on streptavidin matrices (**Figure 4c**).⁸¹

The systematic investigation of KAT multiprotein complexes that modulate activity and localization of the acetyl-transferase activity is a major challenge. Enzyme probes targeting KATs within cellular lysates appear ideally suited for this task. While probes like **22** can generally label CoA-binding enzymes with an affinity tag, more specific functional groups that are based on Lys-CoA appear better suited as KAT probes. A bisubstrate inhibitor was developed by the Cole group that inhibits p300 with nanomolar potency.⁸² It consists of CoA tethered to the ϵ -amino group of Lys-amide via a short spacer, thus simultaneously occupying substrate and cosubstrate binding pockets of the enzyme. Interestingly, embedding the Lys-CoA conjugate into a histone-derived peptide can alter the selectivity of the inhibitor. This was demonstrated for H3-CoA-20 that contained the Lys-CoA

conjugate at position K14 of a histone H3 peptide. This site is acetylated by PCAF, a KAT related to Gcn5. H3-CoA-20, but not Lys-CoA, is a selective inhibitor of PCAF.⁸² In further work, the inhibitor scaffold was extended by adding a photoactivatable crosslinker and affinity tag.⁸³ These probes enriched canonical KATs of the p300/CBP, Gcn5/PCAF and MYST family from cellular lysates overexpressing the proteins. The fact that **23** was shown to enrich endogenous Gcn5 from lysates represents an important step forward in utilizing this type of probes in characterizing cellular KATs (**Figure 4d**).⁸⁴

Protein/histone deacetylation

Protein acetyl-moieties are removed by lysine or histone deacetylases (HDACs) (**Figure 4e**). Mammalian HDACs are grouped into eleven Zn²⁺-dependent HDACs of class I, II and IV and seven mechanistically unrelated class III HDACs, commonly referred to as sirtuins.⁸⁵ In contrast to classical HDACs, sirtuins utilize NAD⁺ as cosubstrate, thereby linking catalysis to the metabolic state of the cell.⁸⁶ HDACs are key regulators of transcription by catalyzing deacetylation of nucleosomal histones.

In HDACs of class I, II and IV the central Zn²⁺ ion coordinates the Nε amide oxygen of the acetylated lysine residue and positions it in close proximity of a bound water molecule that is activated by the Zn²⁺ ion and a His-Asp pair.⁸⁷ A Tyr residue stabilizes the tetrahedral intermediate that forms after the attack of the water molecule onto the Nε carbonyl carbon. This intermediate collapses resulting in the release of acetate and the deacetylated lysine residue.⁸⁷

Like KATs, the Zn²⁺-dependent HDACs are commonly embedded in large multi-domain complexes, such as NuRD, Sin3A and CoREST, all of which contain HDAC1 and HDAC2 as catalytic subunits. The multiprotein assemblies might not only modulate HDAC activity and their recruitment to chromatin but also establish cross-talk with other chromatin modifying enzymes, like LSD1, which is also part of the CoREST complex.^{85, 88} HDACs are not limited to the regulation of chromatin. For instance, HDACs of class II are known to shuttle between the nucleus and the cytosol where several non-histone substrate proteins are localized.⁸⁵ Indeed, HDAC6 catalyzes deacetylation of α-tubulin at lysine 40, which is associated with highly stable microtubules. Due to their multiple functions HDACs have been intensively investigated and chemical biology has been instrumental not

only for the discovery of histone deacetylases (HDACs) but also for their further characterization.⁸⁹

Prior to the discovery of the first histone deacetylase it was known that natural products Trapoxin A and Trichostatin A increased the global acetylation levels of histones in tissue culture experiments, presumably by inhibition of the elusive histone deacetylases. An irreversible reaction between an active site nucleophile and the epoxide moiety of Trapoxin A appeared as the most likely mode of action of this compound.⁹⁰ The Schreiber group conducted a total chemical synthesis of Trapoxin A containing a handle for immobilization, referred to as K-Trap (**24**). In seminal work, they used this probe for isolating the first mammalian Zn²⁺-dependent HDAC from calf thymus lysates (**Figure 4f**).^{91, 92} This milestone discovery paved the way for the subsequent identification of all other mammalian Zn²⁺-dependent HDACs.⁸⁵

HDACs have been considered as potential drug targets early on since these modulate histone acetylation levels and thereby gene activity.⁹³ While most of the initially discovered natural product HDAC inhibitors could not be developed into clinical drugs, these guided the development of synthetic, small-molecule compounds targeting these enzymes. Suberoylanilide hydroxamic acid (SAHA) is among the most prominent HDAC inhibitors and has been approved for cancer therapy.⁹⁴ Structural analyses showed that the hydroxamic acid moiety of SAHA binds tightly to the catalytic Zn²⁺ ion, while the methylene linker occupies a narrow tunnel leading to the active site (**Figure 4e**).⁹³

HDACs are embedded into large multiprotein complexes and unlike most KATs, these enzymes are only poorly accessible by recombinant protein techniques, which imposes a major obstacle on HDAC research. Therefore, enzyme probes for the analysis of endogenous HDACs in cellular lysates are highly valuable tools.

Building on the SAHA scaffold the SAHA-BPyne-probe (**25**) containing a photoactivatable crosslinker and CuAAC-handle in addition to the hydroxamic acid moiety was synthesized (**Figure 4f**).^{95, 96} In extracts and living cells, this probe crosslinked to endogenous HDACs 1, 2, 3 and 6.⁹⁵ Further structure-activity relationship studies with other hydroxamate moieties underlined the utility of SAHA-BPyne as excellent tool for *in situ* labeling of HDACs and isolation of HDAC complexes.⁹⁶ Replacing the hydroxamate moiety by pimelic diphenylamide as reactive group yielded an HDAC probe that is selective for HDAC3 and allowed subsequent investigation of the role of this enzyme in

gene silencing.⁹⁷ Given the importance of HDACs as drug targets immobilized SAHA and givinostat have been used for chemoproteomic profiling of HDAC inhibitors.⁹⁸

A recent report focused on the substrate selectivity of HDAC complexes. Hydroxamic amino acids of various chain lengths were incorporated into peptide probes derived from different known acetylation sites of the human proteome.⁹⁹ Proteomic pull-down experiments defined the impact of residues in vicinity of the modification site on selectivity and composition of the recruited HDAC complexes. It was shown that the tumor suppressor protein p53-derived probe p53-AsuHd (**26**) efficiently enriched the NuRD and Sin3A complexes containing HDACs 1 and 2 from cellular lysates, while probes derived from other sequence contexts were more selective for HDACs 1 and 2 embedded in the CoREST complex.⁹⁹

Sirtuins, which share no sequence similarity with the Zn²⁺-dependent HDACs and require NAD⁺ as cofactor^{86, 100}, are insensitive to the inhibitors and probes targeting class I, II and IV HDACs. Class III HDACs play important biological roles in various processes including regulation of metabolism and ageing. Recent studies imply that sirtuins can remove a broad range of acyl-modifications from protein lysine residues with SIRT5 cleaving malonyl, succinyl and glutaryl moieties and SIRT6 targeting long-chain fatty acid modifications.¹⁰¹⁻¹⁰⁵

Enzyme probes derived from the substrate containing acetylated and acylated lysine residues as functional groups in combination with photoactivatable crosslinkers appear highly promising for investigating sirtuins. Varying acyl moieties and sequence contexts of peptide linkers in the general probe scaffold **1** resulted in a broad set of sirtuin-probes (**Figure 1a**).¹⁰⁶⁻¹⁰⁸ For instance, probe **1** was instrumental for identifying mitochondrial SIRT3 as key regulator of lysine-crotonylation (**Figure 1a**).¹⁰⁶ An extension of this concept is the development of sirtuin substrates that also block NAD⁺ binding and thereby prohibit the deacylation reaction. Peptides modified with bulky succinyl-lysines interact with the NAD⁺ binding pocket and thereby inhibit the desuccinylase Sirt5.¹⁰⁹

Sirtuin catalysis starts with the nucleophilic attack of the acyl-Lys amide oxygen onto the electrophilic C1' carbon of NAD⁺ resulting in an a-1'-O-alkylamidate intermediate **27**. Hydrolysis then releases the deacylated Lys residue and O-acetyl-ADP-ribose (**Figure 4h**).^{100, 110} It was shown that thioacetyl-lysine interferes with this reaction pathway and constitutes a mechanism-based sirtuin

inhibitor.^{100, 111} The sulfur of the thioacetyl group attacks the C1' carbon of NAD⁺, but forms the thioimidate product **28**, which cannot be processed by the enzyme (**Figure 4h**). Inhibitors of this type have supported mechanistic and structural investigations on sirtuins and have inspired new generations of sirtuin-probes. An aminoxy-modified NAD⁺ derivative was synthesized and combined with thioacetyl-peptides. During catalysis both components are linked by the stalled thioimidate intermediate thereby allowing co-purification of the sirtuin. Isolation of these complexes is facilitated by the aminoxy-moiety that can be conjugated to affinity tags by oxime ligation.¹¹² Thioacetyl-probes have further been used for mapping sirtuin substrates in high-throughput. Thiofluoroacetyl-Lys was incorporated into arrays of more than 500 peptides derived from the mitochondrial proteome. By immune-staining captured SIRT3 on the array, the amino acid sequences of preferred substrate sites of this enzyme could be defined.¹¹³

Binding proteins of acetylated histones

Bromodomains are the paradigm binding modules of acetylated protein lysine residues. 61 genes encoding bromodomain factors have been identified in the human genome.⁷² Large-scale screens with peptide-arrays and structural investigations have uncovered binding preferences of these acetyl-lysine binding modules.¹¹⁴ However, capturing bromodomain-proteins with acetylated peptides or similar probes in the context of cell extracts remains highly challenging. This is likely due to the generally low affinity for the acetylated substrate sites, which can be in the high micromolar or even millimolar range. Furthermore, competition of binding to the probes by sirtuins might further complicate affinity readouts. Based on the recent success in the development of bromodomain-specific small-molecule inhibitors, new probes have been designed. Compound JQ1 (**29**) was shown to bind the two separate bromodomains of BRD4 with affinities of 50 nM and 90 nM, respectively.¹¹⁵ By attaching an affinity tag, this small-molecule was further developed into bromodomain-probe Bio-JQ1 (**30**). Bromodomain capture combined with massive parallel DNA sequencing was used for identifying the binding sites of this drug throughout the genome of cancer cells.¹¹⁶

Histone phosphorylation

Phosphorylation of histones, like lysine acetylation and methylation was discovered in the 1960ies. However, chromatin research has not been the main driver for the development of chemical probes for phosphorylation, because the importance of this PTM in other cellular signaling networks is much better understood. Consequently, the majority of established phosphorylation probes were designed for investigating signal transduction pathways.¹¹⁷⁻¹²² These might, nevertheless, prove useful for chromatin research, as the kinases and phosphatases that play important roles in cell signaling also regulate the phosphorylation state of histones. Most phosphorylation sites have been identified on histone H3, with target sites on the N-terminal tail (Thr 3, Thr 6, Ser 10, Ser 11, and Ser 28) as well as within the globular fold. During the onset of mitosis H3S10 and later H3S28 become phosphorylated genome-wide by aurora B kinase.¹²³ Although the physiological function of this modification is still not fully understood, H3S10ph is commonly used as a marker for mitosis. Aurora B has been subjected to a ‘bump and hole’ strategy. The resulting, mutated kinase is selectively inhibited by bulky adenine analogs and utilizes bulky ATP derivatives for protein phosphorylation.¹²⁴ Use of bulky ATPgS as cosubstrate allowed thiophosphorylation of aurora B substrates. These can be enriched on electrophilic thiol-capture resins and subsequently released by oxidation. This strategy uncovered several previously unidentified aurora B substrates including histone-binding proteins.¹²⁴ A covalent inhibitor of analogue-sensitive aurora B represents another promising probe for application in chromatin research.¹²⁵ Anilinoquinazoline kinase inhibitors have been equipped with electrophilic moieties that covalently react with a cysteine residue engineered into the ATP-binding pocket of an analogue-sensitive, mutant aurora B kinase. This inhibitor can be modified with fluorophores thereby enabling direct labeling of the mutant enzyme.

Binding proteins of phosphorylated histones

Probes for binding proteins of phosphorylated histones have also been reported. Phosphorylation of H3T3 accumulates at centromeric regions of chromatin during prophase and metaphase of mitosis and is catalyzed by the kinase Haspin.¹²³ The CLASPI strategy (see section on *Binding proteins to methylated histones*) using a peptide-probe derived from histone H3 containing phosphorylated T3, a photoactivatable crosslinker and an alkyne-tag identified previously known as well as potential novel

binding proteins to this chromatin mark.¹²⁶ Phosphorylation of H3S10 has been investigated with peptide probes containing a phosphonate mimic of phosphorylated serine.¹²⁷ This probe cannot be dephosphorylated by phosphatases and bound 14-3-3 proteins from cellular lysates. It was also used to study repressive effects of H3S10ph on chromatin-factor recruitment.

DNA methylation

DNA methyltransferases (DNMTs) catalyze the transfer of the activated methyl-group of SAM to purine or pyrimidine bases in DNA (**Figure 5a**). Three classes of enzymes can be discriminated. N6-methyladenine DNMTs specifically methylate the C-6 amino group of adenines and N4-methylcytosine DNMTs specifically methylate the N-4 amino group of cytosines. In contrast to the amino-directed enzymes, C5-methylcytosine DNMTs specifically methylate the C-5 carbon of cytosine. In bacteria all three classes of DNMTs are found. There, these protect endogenous DNA from digestion and are part of the restriction modification systems of type I, II, and III. The prokaryotic DNMTs have the same recognition sequence as corresponding restriction enzymes. Methylation of bases in the targeting sequence inhibits DNA cleavage. Due to the palindromic nature of recognition sequences bacterial DNMTs mediate symmetric (i.e. on both strands) methylation. Some of the C5-methylcytosine (5mC) DNMTs have very short recognition sequences such as for example M.SssI, which works on CpG dinucleotides.

Eukaryotic cells only contain C5-methylcytosine DNMTs. In plants these target CpG, CpHpG and CpHpH sequences (where H = A, C or T). Fungal and animal C5-methylcytosine DNMTs mostly target CpG dinucleotides. Whereas *de novo* DNMTs newly methylate cytosines in early development and set up methylation patterns, maintenance DNMTs methylate DNA when only one of the strands of the DNA duplex is methylated, e.g. after replication (hemimethylated state). In contrast to *de novo* DNMTs these enzymes work throughout the lifetime of an organism. In mammalian cells, five proteins have been described to belong to the DNMT family. DNMT1 is the most abundant DNMT. It is much more active on hemimethylated DNA as compared to unmethylated DNA *in vitro*, but it is still more active at *de novo* methylation than other DNMTs. DNMT3a and DNMT3b methylate un- and hemimethylated substrates with equal efficiency. The enzymes form complexes with DNMT3L,

which contains several DNA methyltransferase motifs but is enzymatically inactive. DNMT2 (now renamed into TRDMT1) turned out not to work on DNA but what shown to target RNA in human cells.

5mC has important functions in gene regulation in eukaryotic cells, where it is especially involved in transcriptional repression via its linkage to the histone modification system (i.e. histone methylation and deacetylation). In mammals, the majority of CpG sites are methylated (>70%). However, CpG islands (i.e. regions of about 500 nt that contain higher frequency of CpG than the rest of the genome) in the promoter regions of most house keeping genes are undermethylated under normal conditions. Changes in this overall distribution of 5mC are a hallmark of cellular transformation and tumor progression.¹²⁸

The majority of work generating chemical probes for DNMTs has focused on bacterial enzymes and the sequence specific incorporation of functionalized groups into DNA. Initially, it was recognized that the side chain of methionine is dispensable for most SAM-based methyltransferases. In consequence, it was shown that the N6-methyladenine DNMT from *Thermus aquaticus* M.TaqI could use N-adenosylaziridine (5'-aziridino-5'-deoxyadenosine) **31** to transfer the adenosyl moiety to the exocyclic amino group of 2'-deoxyadenosine within the double-stranded 5'-TCGA*-3' DNA sequence (**Figure 5c and 5d**).¹²⁹ The work was later expanded to show that the C-6- and C-8-position of the adenosyl moiety of N-adenosylaziridine could be functionalized with multiple reporter groups such as biotin **32**¹³⁰ or fluorophores **33**^{131, 132} while still being utilized by the M.TaqI DNMT (**Figure 5d**). The reactions therefore allow sequence-specific methyltransferase-induced labeling of DNA (SMILing DNA).¹³³

Alternative to the linkage of bulky probes and handles to N-adenosylaziridine, azides were coupled to the C-8 position of the adenosyl moiety of this SAM analogue for subsequent functionalization. 8-Azido-5'-aziridino-5'-deoxyadenosine **34** was shown to be utilized by the prokaryotic DNMTs M.TaqI, M.EcoRI (a N6-methyladenine DNMT targeting GAA*TTC) and M.HhaI (a C5-methylcytosine DNMT targeting GCGC*) (**Figure 5d**).^{134, 135} Further, it was shown that spacing the azido group at the C-8 position of N-adenosylaziridine via methylene groups as in **35** did not impair reaction by M.TaqI.¹³⁵ The products of the labeling reactions with DNA containing the

corresponding recognition sequences of the enzymes undergo Staudinger ligation with functionalized triarylphosphines. This allowed sequence specific incorporation of biotin handles into DNA¹³⁵ as well as sequence specific chemical cleavage of DNA after treatment with Cu₂SO₄ and mercapto phenyl acetic acid (MPA) according to Haber-Weiss redox cycling via phenanthroline (OP)-Cu(I) chemistry.¹³⁶ The azido derivatives of N-adenosylaziridine should also be accessible to click chemistry, although such reactions have not been demonstrated.

A major problem of the N-adenosylaziridine derivatives is significantly reduced reactivity compared to SAM, requiring 100-fold excess of analogue in the labeling reactions. Obviously, the core structures lacking the methionine recognition moiety present in SAM are less efficient in binding the DNMTs (or other methyltransferases). Solving this issue, N-mustard analogues of SAM were synthesized that incorporate this amino acid functionality at the 5' -position of the ribose sugar (5' (diaminobutyric acid)-N iodoethyl-5' -deoxyadenosine) precursor **36** (**Figure 5d**). These compounds generate highly reactive aziridinium ions at physiological pH and are utilized by the *M.TaqI*, *M.EcoRI*, *M.HhaI* and *M.SssI* DNMTs.¹³⁷ Further, replacing the diaminobutyric acid part of the N-mustard by a propargyl moiety allowed functionalization of specific DNA sequences by CuAAC chemistry after reaction with prokaryotic DNMTs. Lastly, the derivatization of adenine by azido groups on C-8 for subsequent functionalization has been combined with the N-mustard approach for efficient interactions of the analogues with DNMTs and improved labeling reactions.¹³⁸

Since the aziridinium-based analogues form strong product inhibitors during the methyltransferase-catalyzed reactions, derivatives with extended carbon chains (e.g. alkyl-, alkenyl- and alkynyl-groups) replacing the methyl group at the sulfonium center of SAM were analyzed. SAM analogues with allyl and propargyl moieties function as efficient cofactors for all three classes of DNMTs.²⁰ Subsequent work showed that appending bulky moieties to the propargylic side chain, doubling its size is tolerated by *M.TaqI* but requires a 'bump and hole' strategy for *M.HhaI*. Via an amino functionality linked to the propargyl moiety of **37** it is possible to couple N-hydroxysuccinimide (NHS)-activated reporters to the DNA after the labeling reaction (**Figure 5d**).¹³⁹

Besides the targeted labeling of DNA (and protein) chemical biology approaches have also been put forward for the capture and identification of methyltransferases. The demethylated

cosubstrate product of SAM-based reactions is S-adenosyl-L-homocysteine (SAH). The thioether function in SAH is chemically much more stable than the activated sulfonium center in SAM, and SAH is a general competitive product inhibitor of methyl transferases. SAH analogues with different attachment points (N6 and C8 of the adenine base) for two functionalities, biotin and photoactivatable crosslinkers have been synthesized. These compounds, including **38**, could be efficiently used to capture DNMTs from complex cell lysates in affinity purification schemes (**Figure 5d**).¹⁴⁰ Other work has been building on known inhibitors of DNMT1. Via introducing functionalized groups such as propargyl moieties to these compounds it is anticipated that the enzymes can be specifically targeted for further chemical reactions.¹⁴¹

DNA demethylases

The mechanisms of demethylation of 5mC have been unclear for a long time. Recent work has established a mechanism that has some similarity to the demethylation of proteins. The three members of the ten-eleven translocation (TET1, 2, 3) family of proteins have been shown to work as methylcytosine directed non-haem iron(II) and 2-KG dependent oxygenases. The general reaction mechanism of oxidative decarboxylation of 2-KG and activation of molecular oxygen via an oxo-ferryl intermediate is analogous to the *jumonji* proteins (**Figure 5e**). However, the resulting 5-hydroxymethylcytosine (5hmC) is stable in contrast to the heminal formed with methylated amino groups. TET enzymes can, however, oxidize 5hmC further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).¹⁴²

There is still some debate how the oxidation of 5mC ultimately results in restoration of the unmodified cytosine at its original position. First of all, 5hmC, 5fC and 5caC biochemically and physiologically cause responses different from 5mC. These are also thought to ‘dilute out’ the 5mC signal as the marks are not propagated by DNA maintenance methylation. Another route might involve decarboxylation of 5caC by decarboxylases or removal of the oxidized states of 5mC via DNMTs. 5hmC can also be deaminated by AID/Apobec enzymes to give 5-hydroxymethyluracil. While these mechanisms might be at place in certain cells and at certain stages, a more general route of demethylation involves base excision repair (BER). Both the deamination and the oxidation

products are targeted by thymine DNA glycosylase (TDG). The resulting abasic site is then excised by the BER machinery, followed by replacement of the nucleotide with dCTP by DNA polymerase and ligation to repair the nick.¹⁴²

Given the relative recent discoveries of the TET enzymes and their catalytic activity, it is not surprising that only few chemical biology approaches have so far been designed to target these important epigenetic regulators. A major focus has been on mapping sites of oxidized 5mC in DNA. Based on the reactivity of hydrazines towards aldehydes a chemical probe for 5fmc was designed. In combination with dye (i.e. acridine) that can intercalate the DNA duplex it was shown that 9-(hydrazinylmethyl)acridine (HMA) selectively modifies 5fC in a covalent bond. The formed hydrazone inhibits base excision by TDG, polymerase bypass as well as gene expression.¹⁴³ Additional work expanded this strategy to develop a probe that can quantitatively detect the symmetry of 5fC loci. For this purpose, the 5fC-targeting hydrazine moiety was conjugated through a linker with a pyrene fluorophore. It was shown that the two neighboring pyrene groups of the ligand produce distinct dimer fluorescence emission with large Stokes shift if the 5fC sites are located in a symmetric form of 5fCpG.¹⁴⁴ 5fC containing DNA can also be detected via selective biotinylation after oxime-ligation using biotin-hydroxylamines. The highly efficient reaction allowed subsequent enrichment of 5fC DNA for analysis by mass spectrometry.¹⁴⁵ A recent report further exploited the chemical properties of the aldehyde group of 5fC. Modification with 2,3,3-trimethylindole derivatives in an aldol-like reaction yields chromophores with characteristic photo-chemical properties that could facilitate simple readouts of enzymatic 5fC formation.¹⁴⁶

Perspective

As we have discussed, a multitude of recently developed chemical tools supports the analysis of the complex cross-talk between chromatin modifying enzymes and chromatin modifications. Chromatin-directed research has paved the way for the study of major protein posttranslational modification events. Yet, it has become clear that these are not restricted to histone proteins but are involved in different cellular signaling pathways. While chemical probes that target chromatin-modifying enzymes enable detailed analysis of the composition of multiprotein complexes and their distribution,

cosubstrate-based-probes have been used for defining modification events and substrate sites within the complex environments of proteomes. Defining the context-dependent reactions that transduce signals to and from chromatin in cellular development and differentiation as well as the many signaling events orchestrated by enzymes that were initially characterized in chromatin context but which have other substrates is a major challenge where chemical biology approaches will be invaluable. Future developments need to define probes that mimic the cellular targets of chromatin-modifying enzymes more closely and that can be delivered and used within the complex environment of living cells. Also, more versatile toolboxes that allow the high-throughput analysis of groups of enzymes instead of single catalysis events need to be set up. Lastly, strategies need to be considered that enable temporal (i.e. kinetics) and spatial (i.e. different subcellular environments) analysis as the current approaches do not discriminate between many putative modification events. Based on the pace the field of chemical biology has taken within chromatin research during the last years, it is likely that these and other challenges can be met within the near future.

Acknowledgment

This work was funded by the priority program SPP1623 of the Deutsche Forschungsgemeinschaft (W.F. and D.S., Grants: FL 1513/3-1 and SCHW 1163/4-1) as well as intramural programs of the King Abdullah University of Science and Technology (KAUST) and the Max Planck Society (W.F.).

References

1. Kornberg, R. D. (1977) Structure of Chromatin, *Annual Review of Biochemistry* 46, 931-954.
2. Fischle, W., Mootz, H. D., and Schwarzer, D. (2015) Synthetic histone code, *Curr Opin Chem Biol* 28, 131-140.
3. Fischle, W. (2008) Talk is cheap-cross-talk in establishment, maintenance, and readout of chromatin modifications, *Gene Dev* 22, 3375-3382.
4. Holt, M., and Muir, T. (2015) Application of the Protein Semisynthesis Strategy to the Generation of Modified Chromatin, *Annu Rev Biochem*.
5. Jenuwein, T., and Allis, C. D. (2001) Translating the histone code, *Science* 293, 1074-1080.

6. Muller, M. M., and Muir, T. W. (2015) Histones: at the crossroads of Peptide and protein chemistry, *Chemical reviews* 115, 2296-2349.
7. Schwarzer, D. (2010) Chemical tools in chromatin research, *Journal of peptide science : an official publication of the European Peptide Society* 16, 530-537.
8. Kouzarides, T. (2007) Chromatin modifications and their function, *Cell* 128, 693-705.
9. Cole, P. A. (2008) Chemical probes for histone-modifying enzymes, *Nat Chem Biol* 4, 590-597.
10. Niphakis, M. J., and Cravatt, B. F. (2014) Enzyme inhibitor discovery by activity-based protein profiling, *Annu Rev Biochem* 83, 341-377.
11. Sanman, L. E., and Bogoy, M. (2014) Activity-based profiling of proteases, *Annu Rev Biochem* 83, 249-273.
12. Nomura, D. K., Dix, M. M., and Cravatt, B. F. (2010) Activity-based protein profiling for biochemical pathway discovery in cancer, *Nat Rev Cancer* 10, 630-638.
13. Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry, *Annu Rev Biochem* 77, 383-414.
14. Bishop, A. C., Shah, K., Liu, Y., Witucki, L., Kung, C. Y., and Shokat, K. M. (1998) Design of allele-specific inhibitors to probe protein kinase signaling, *Curr Biol* 8, 257-266.
15. Schreiber, S. L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry, *Bioorgan Med Chem* 6, 1127-1152.
16. Belshaw, P. J., Schoepfer, J. G., Liu, K. Q., Morrison, K. L., and Schreiber, S. L. (1995) Rational Design of Orthogonal Receptor-Ligand Combinations, *Angewandte Chemie-International Edition in English* 34, 2129-2132.
17. Bishop, A. C., Ubersax, J. A., Petsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., Shimizu, E., Tsien, J. Z., Schultz, P. G., Rose, M. D., Wood, J. L., Morgan, D. O., and Shokat, K. M. (2000) A chemical switch for inhibitor-sensitive alleles of any protein kinase, *Nature* 407, 395-401.
18. Carlson, S. M., and Gozani, O. (2014) Emerging technologies to map the protein methylome, *J Mol Biol* 426, 3350-3362.
19. Luo, M. (2015) Current Methods for Methylome Profiling, 187-217.

20. Dalhoff, C., Lukinavicius, G., Klimasauskas, S., and Weinhold, E. (2006) Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases, *Nat Chem Biol* 2, 31-32.
21. Osborne, T., Roska, R. L., Rajski, S. R., and Thompson, P. R. (2008) In situ generation of a bisubstrate analogue for protein arginine methyltransferase 1, *J Am Chem Soc* 130, 4574-4575.
22. Yao, Y., Chen, P., Diao, J., Cheng, G., Deng, L., Anglin, J. L., Prasad, B. V., and Song, Y. (2011) Selective inhibitors of histone methyltransferase DOT1L: design, synthesis, and crystallographic studies, *J Am Chem Soc* 133, 16746-16749.
23. Lee, B. W., Sun, H. G., Zang, T., Kim, B. J., Alfaro, J. F., and Zhou, Z. S. (2010) Enzyme-catalyzed transfer of a ketone group from an S-adenosylmethionine analogue: a tool for the functional analysis of methyltransferases, *J Am Chem Soc* 132, 3642-3643.
24. Palomo, J. M. (2012) Click reactions in protein chemistry: from the preparation of semisynthetic enzymes to new click enzymes, *Org Biomol Chem* 10, 9309-9318.
25. Patterson, D. M., Nazarova, L. A., and Prescher, J. A. (2014) Finding the right (bioorthogonal) chemistry, *Acs Chem Biol* 9, 592-605.
26. Yang, Y. Y., Grammel, M., Raghavan, A. S., Charron, G., and Hang, H. C. (2010) Comparative Analysis of Cleavable Azobenzene-Based Affinity Tags for Bioorthogonal Chemical Proteomics, *Chem Biol* 17, 1212-1222.
27. Verhelst, S. H. L., Fonovic, M., and Bogoy, M. (2007) A mild chemically cleavable linker system for functional proteomic applications, *Angew Chem Int Edit* 46, 1284-1286.
28. Binda, O., Boyce, M., Rush, J. S., Palaniappan, K. K., Bertozzi, C. R., and Gozani, O. (2011) A chemical method for labeling lysine methyltransferase substrates, *Chembiochem* 12, 330-334.
29. Bothwell, I. R., Islam, K., Chen, Y., Zheng, W., Blum, G., Deng, H., and Luo, M. (2012) S-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation, *J Am Chem Soc* 134, 14905-14912.
30. Willnow, S., Martin, M., Luscher, B., and Weinhold, E. (2012) A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases, *Chembiochem* 13, 1167-1173.

31. Bothwell, I. R., and Luo, M. (2014) Large-scale, protection-free synthesis of Se-adenosyl-L-selenomethionine analogues and their application as cofactor surrogates of methyltransferases, *Org Lett* 16, 3056-3059.
32. Peters, W., Willnow, S., Duisken, M., Kleine, H., Macherey, T., Duncan, K. E., Litchfield, D. W., Luscher, B., and Weinhold, E. (2010) Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click labeling, *Angew Chem Int Ed Engl* 49, 5170-5173.
33. Wang, R., Ibanez, G., Islam, K., Zheng, W., Blum, G., Sengelaub, C., and Luo, M. (2011) Formulating a fluorogenic assay to evaluate S-adenosyl-L-methionine analogues as protein methyltransferase cofactors, *Mol Biosyst* 7, 2970-2981.
34. Lin, Q., Jiang, F., Schultz, P. G., and Gray, N. S. (2001) Design of allele-specific protein methyltransferase inhibitors, *J Am Chem Soc* 123, 11608-11613.
35. Islam, K. (2015) Allele-specific chemical genetics: concept, strategies, and applications, *Acs Chem Biol* 10, 343-363.
36. Wang, R., and Luo, M. (2013) A journey toward Bioorthogonal Profiling of Protein Methylation inside living cells, *Curr Opin Chem Biol* 17, 729-737.
37. Islam, K., Chen, Y. L., Wu, H., Bothwell, I. R., Blum, G. J., Zeng, H., Dong, A. P., Zheng, W. H., Min, J. R., Deng, H. T., and Luo, M. K. (2013) Defining efficient enzyme-cofactor pairs for bioorthogonal profiling of protein methylation, *Proceedings of the National Academy of Sciences of the United States of America* 110, 16778-16783.
38. Islam, K., Zheng, W., Yu, H., Deng, H., and Luo, M. (2011) Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling, *Acs Chem Biol* 6, 679-684.
39. Islam, K., Bothwell, I., Chen, Y., Sengelaub, C., Wang, R., Deng, H., and Luo, M. (2012) Bioorthogonal profiling of protein methylation using azido derivative of S-adenosyl-L-methionine, *J Am Chem Soc* 134, 5909-5915.
40. Wang, R., Zheng, W., Yu, H., Deng, H., and Luo, M. (2011) Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues, *J Am Chem Soc* 133, 7648-7651.

41. Guo, H., Wang, R., Zheng, W., Chen, Y., Blum, G., Deng, H., and Luo, M. (2014) Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues, *Acs Chem Biol* 9, 476-484.
42. Wang, R., Islam, K., Liu, Y., Zheng, W., Tang, H., Lailler, N., Blum, G., Deng, H., and Luo, M. (2013) Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells, *J Am Chem Soc* 135, 1048-1056.
43. Wang, R., Zheng, W., and Luo, M. (2014) A sensitive mass spectrum assay to characterize engineered methionine adenosyltransferases with S-alkyl methionine analogues as substrates, *Anal Biochem* 450, 11-19.
44. Li, J., Wei, H., and Zhou, M. M. (2011) Structure-guided design of a methyl donor cofactor that controls a viral histone H3 lysine 27 methyltransferase activity, *J Med Chem* 54, 7734-7738.
45. Khani-Oskouee, S., Jones, J. P., and Woodard, R. W. (1984) Stereochemical course of the biosynthesis of 1-aminocyclopropane-1-carboxylic acid. I. Role of the asymmetric sulfonium pole and the alpha-amino acid center, *Biochem Biophys Res Commun* 121, 181-187.
46. Singh, S., Zhang, J., Huber, T. D., Sunkara, M., Hurley, K., Goff, R. D., Wang, G., Zhang, W., Liu, C., Rohr, J., Van Lanen, S. G., Morris, A. J., and Thorson, J. S. (2014) Facile chemoenzymatic strategies for the synthesis and utilization of S-adenosyl-(L)-methionine analogues, *Angew Chem Int Ed Engl* 53, 3965-3969.
47. Thomsen, M., Vogensen, S. B., Buchardt, J., Burkart, M. D., and Clausen, R. P. (2013) Chemoenzymatic synthesis and in situ application of S-adenosyl-L-methionine analogs, *Org Biomol Chem* 11, 7606-7610.
48. Liu, C., Yu, Y., Liu, F., Wei, X., Wrobel, J. A., Gunawardena, H. P., Zhou, L., Jin, J., and Chen, X. (2014) A chromatin activity-based chemoproteomic approach reveals a transcriptional repressome for gene-specific silencing, *Nat Commun* 5, 5733.
49. Johansson, C., Tumber, A., Che, K., Cain, P., Nowak, R., Gileadi, C., and Oppermann, U. (2014) The roles of Jumonji-type oxygenases in human disease, *Epigenomics* 6, 89-120.
50. Chang, B., Chen, Y., Zhao, Y., and Bruick, R. K. (2007) JMJD6 is a histone arginine demethylase, *Science* 318, 444-447.

51. Bottger, A., Islam, M. S., Chowdhury, R., Schofield, C. J., and Wolf, A. (2015) The oxygenase Jmjd6--a case study in conflicting assignments, *Biochem J* 468, 191-202.
52. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why, *Mol Cell* 33, 1-13.
53. Marholz, L. J., Chang, L., Old, W. M., and Wang, X. (2015) Development of substrate-selective probes for affinity pulldown of histone demethylases, *Acs Chem Biol* 10, 129-137.
54. Culhane, J. C., Szewczuk, L. M., Liu, X., Da, G., Marmorstein, R., and Cole, P. A. (2006) A mechanism-based inactivator for histone demethylase LSD1, *J Am Chem Soc* 128, 4536-4537.
55. Culhane, J. C., Wang, D., Yen, P. M., and Cole, P. A. (2010) Comparative analysis of small molecules and histone substrate analogues as LSD1 lysine demethylase inhibitors, *J Am Chem Soc* 132, 3164-3176.
56. Yang, M., Culhane, J. C., Szewczuk, L. M., Gocke, C. B., Brautigam, C. A., Tomchick, D. R., Machius, M., Cole, P. A., and Yu, H. (2007) Structural basis of histone demethylation by LSD1 revealed by suicide inactivation, *Nat Struct Mol Biol* 14, 535-539.
57. Pilotto, S., Speranzini, V., Tortorici, M., Durand, D., Fish, A., Valente, S., Forneris, F., Mai, A., Sixma, T. K., Vachette, P., and Mattevi, A. (2015) Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation, *Proceedings of the National Academy of Sciences of the United States of America* 112, 2752-2757.
58. Fischle, W. (2012) One, two, three: how histone methylation is read, *Epigenomics* 4, 641-653.
59. Musselman, C. A., Lalonde, M. E., Cote, J., and Kutateladze, T. G. (2012) Perceiving the epigenetic landscape through histone readers, *Nat Struct Mol Biol* 19, 1218-1227.
60. Patel, D. J., and Wang, Z. (2013) Readout of epigenetic modifications, *Annu Rev Biochem* 82, 81-118.
61. Nikolov, M., and Fischle, W. (2013) Systematic analysis of histone modification readout, *Mol Biosyst* 9, 182-194.
62. Li, X., and Li, X. D. (2015) Chemical proteomics approaches to examine novel histone posttranslational modifications, *Curr Opin Chem Biol* 24, 80-90.

63. Li, X., and Kapoor, T. M. (2010) Approach to Profile Proteins That Recognize Post-Translationally Modified Histone "Tails", *Journal of the American Chemical Society* 132, 2504-+.
64. Li, X., Foley, E. A., Molloy, K. R., Li, Y., Chait, B. T., and Kapoor, T. M. (2012) Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions, *J Am Chem Soc* 134, 1982-1985.
65. Yang, T., Liu, Z., and Li, X. D. (2015) Developing diazirine-based chemical probes to identify histone modification 'readers' and 'erasers', *Chem. Sci.* 6, 1011-1017.
66. Carrozza, M. J., Utley, R. T., Workman, J. L., and Cote, J. (2003) The diverse functions of histone acetyltransferase complexes, *Trends Genet* 19, 321-329.
67. Berndsen, C. E., Albaugh, B. N., Tan, S., and Denu, J. M. (2007) Catalytic mechanism of a MYST family histone acetyltransferase, *Biochemistry-Us* 46, 623-629.
68. Rojas, J. R., Trievel, R. C., Zhou, J. X., Mo, Y., Li, X. M., Berger, S. L., Allis, C. D., and Marmorstein, R. (1999) Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide, *Nature* 401, 93-98.
69. Liu, X., Wang, L., Zhao, K. H., Thompson, P. R., Hwang, Y., Marmorstein, R., and Cole, P. A. (2008) The structural basis of protein acetylation by the p300/CBP transcriptional coactivator, *Nature* 451, 846-850.
70. Wang, L., Tang, Y., Cole, P. A., and Marmorstein, R. (2008) Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function, *Curr Opin Struct Biol* 18, 741-747.
71. Lee, K. K., and Workman, J. L. (2007) Histone acetyltransferase complexes: one size doesn't fit all, *Nat Rev Mol Cell Bio* 8, 284-295.
72. Filippakopoulos, P., and Knapp, S. (2012) The bromodomain interaction module, *Febs Lett* 586, 2692-2704.
73. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014) The growing landscape of lysine acetylation links metabolism and cell signalling, *Nat Rev Mol Cell Bio* 15, 536-550.

74. Wagner, G. R., and Hirschey, M. D. (2014) Nonenzymatic Protein Acylation as a Carbon Stress Regulated by Sirtuin Deacylases, *Molecular Cell* 54, 5-16.
75. Simic, Z., Weiwad, M., Schierhorn, A., Steegborn, C., and Schutkowski, M. (2015) The ϵ -Amino Group of Protein Lysine Residues Is Highly Susceptible to Nonenzymatic Acylation by Several Physiological Acyl-CoA Thioesters, *Chembiochem* 16, 2337-2347.
76. Yang, Y. Y., and Hang, H. C. (2011) Chemical approaches for the detection and synthesis of acetylated proteins, *Chembiochem* 12, 314-322.
77. Yu, M., de Carvalho, L. P. S., Sun, G. X., and Blanchard, J. S. (2006) Activity-based substrate profiling for Gcn5-related N-acetyltransferases: The use of chloroacetyl-coenzyme A to identify protein substrates, *Journal of the American Chemical Society* 128, 15356-15357.
78. Yang, Y. Y., Ascano, J. M., and Hang, H. C. (2010) Bioorthogonal Chemical Reporters for Monitoring Protein Acetylation, *Journal of the American Chemical Society* 132, 3640-+.
79. Yang, Y. Y., Grammel, M., and Hang, H. C. (2011) Identification of lysine acetyltransferase p300 substrates using 4-pentynoyl-coenzyme A and bioorthogonal proteomics (vol 21, pg 4976, 2011), *Bioorg Med Chem Lett* 21, 6613-6613.
80. Yang, C., Mi, J., Feng, Y., Ngo, L., Gao, T., Yan, L., and Zheng, Y. G. (2013) Labeling lysine acetyltransferase substrates with engineered enzymes and functionalized cofactor surrogates, *J Am Chem Soc* 135, 7791-7794.
81. Hwang, Y., Thompson, P. R., Wang, L., Jiang, L., Kelleher, N. L., and Cole, P. A. (2007) A selective chemical probe for coenzyme A-requiring enzymes, *Angew Chem Int Ed Engl* 46, 7621-7624.
82. Lau, O. D., Kundu, T. K., Soccio, R. E., Ait-Si-Ali, S., Khalil, E. M., Vassilev, A., Wolffe, A. P., Nakatani, Y., Roeder, R. G., and Cole, P. A. (2000) HATs off: Selective synthetic inhibitors of the histone acetyltransferases p300 and PCAF, *Molecular Cell* 5, 589-595.
83. Montgomery, D. C., Sorum, A. W., and Meier, J. L. (2014) Chemoproteomic profiling of lysine acetyltransferases highlights an expanded landscape of catalytic acetylation, *J Am Chem Soc* 136, 8669-8676.

84. Montgomery, D. C., Sorum, A. W., Guasch, L., Nicklaus, M. C., and Meier, J. L. (2015) Metabolic Regulation of Histone Acetyltransferases by Endogenous Acyl-CoA Cofactors, *Chem Biol*.
85. Yang, X. J., and Seto, E. (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men, *Nat Rev Mol Cell Bio* 9, 206-218.
86. Houtkooper, R. H., Pirinen, E., and Auwerx, J. (2012) Sirtuins as regulators of metabolism and healthspan, *Nat Rev Mol Cell Bio* 13, 225-238.
87. Grozinger, C. M., and Schreiber, S. L. (2002) Deacetylase enzymes: biological functions and the use of small-molecule inhibitors, *Chem Biol* 9, 3-16.
88. Ahringer, J. (2000) NuRD and SIN3 - histone deacetylase complexes in development, *Trends Genet* 16, 351-356.
89. Minoshima, M., and Kikuchi, K. (2015) Chemical Tools for Probing Histone Deacetylase (HDAC) Activity, *Analytical Sciences* 31, 287-292.
90. Kijima, M., Yoshida, M., Sugita, K., Horinouchi, S., and Beppu, T. (1993) Trapoxin, an Antitumor Cyclic Tetrapeptide, Is an Irreversible Inhibitor of Mammalian Histone Deacetylase, *J Biol Chem* 268, 22429-22435.
91. Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p, *Science* 272, 408-411.
92. Taunton, J., Collins, J. L., and Schreiber, S. L. (1996) Synthesis of natural and modified trapoxins, useful reagents for exploring histone deacetylase function, *Journal of the American Chemical Society* 118, 10412-10422.
93. Gallinari, P., Di Marco, S., Jones, P., Pallaoro, M., and Steinkuhler, C. (2007) HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics, *Cell Res* 17, 195-211.
94. Marks, P. A., and Breslow, R. (2007) Dimethyl sulfoxide to vorinostat: Development of this histone deacetylase inhibitor as an anticancer drug, *Nat Biotechnol* 25, 84-90.
95. Salisbury, C. M., and Cravatt, B. F. (2007) Activity-based probes for proteomic profiling of histone deacetylase complexes, *Proc Natl Acad Sci U S A* 104, 1171-1176.

96. Salisbury, C. M., and Cravatt, B. F. (2008) Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes, *Journal of the American Chemical Society* 130, 2184-2194.
97. Xu, C., Soragni, E., Chou, C. J., Herman, D., Plasterer, H. L., Rusche, J. R., and Gottesfeld, J. M. (2009) Chemical probes identify a role for histone deacetylase 3 in Friedreich's ataxia gene silencing, *Chem Biol* 16, 980-989.
98. Bantscheff, M., Hopf, C., Savitski, M. M., Dittmann, A., Grandi, P., Michon, A. M., Schlegl, J., Abraham, Y., Becher, I., Bergamini, G., Boesche, M., Delling, M., Dumpelfeld, B., Eberhard, D., Huthmacher, C., Mathieson, T., PoECKel, D., Reader, V., Strunk, K., Sweetman, G., Kruse, U., Neubauer, G., Ramsden, N. G., and Drewes, G. (2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes, *Nat Biotechnol* 29, 255-265.
99. Dose, A., Sindlinger, J., Bierlmeier, J., Bakirbas, A., Schulze-Osthoff, K., Einsele-Scholz, S., Hartl, M., Essmann, F., Finkemeier, I., and Schwarzer, D. (2015) Interrogating Substrate Selectivity and Composition of Endogenous Histone Deacetylase Complexes with Chemical Probes, *Angew Chem Int Ed Engl*.
100. Chen, B., Zang, W., Wang, J., Huang, Y., He, Y., Yan, L., Liu, J., and Zheng, W. (2015) The chemical biology of sirtuins, *Chem Soc Rev* 44, 5246-5264.
101. Feldman, J. L., Baeza, J., and Denu, J. M. (2013) Activation of the protein deacetylase SIRT6 by long-chain fatty acids and widespread deacylation by mammalian sirtuins, *J Biol Chem* 288, 31350-31356.
102. Du, J. T., Zhou, Y. Y., Su, X. Y., Yu, J. J., Khan, S., Jiang, H., Kim, J., Woo, J., Kim, J. H., Choi, B. H., He, B., Chen, W., Zhang, S., Cerione, R. A., Auwerx, J., Hao, Q., and Lin, H. N. (2011) Sirt5 Is a NAD-Dependent Protein Lysine Demalonylase and Desuccinylase, *Science* 334, 806-809.
103. Jiang, H., Khan, S., Wang, Y., Charron, G., He, B., Sebastian, C., Du, J., Kim, R., Ge, E., Mostoslavsky, R., Hang, H. C., Hao, Q., and Lin, H. (2013) SIRT6 regulates TNF- α secretion through hydrolysis of long-chain fatty acyl lysine, *Nature* 496, 110-113.

104. Peng, C., Lu, Z., Xie, Z., Cheng, Z., Chen, Y., Tan, M., Luo, H., Zhang, Y., He, W., Yang, K., Zwaans, B. M., Tishkoff, D., Ho, L., Lombard, D., He, T. C., Dai, J., Verdin, E., Ye, Y., and Zhao, Y. (2011) The first identification of lysine malonylation substrates and its regulatory enzyme, *Mol Cell Proteomics* 10, M111 012658.
105. Tan, M., Peng, C., Anderson, K. A., Chhoy, P., Xie, Z., Dai, L., Park, J., Chen, Y., Huang, H., Zhang, Y., Ro, J., Wagner, G. R., Green, M. F., Madsen, A. S., Schmiesing, J., Peterson, B. S., Xu, G., Ilkayeva, O. R., Muehlbauer, M. J., Braulke, T., Muhlhausen, C., Backos, D. S., Olsen, C. A., McGuire, P. J., Pletcher, S. D., Lombard, D. B., Hirschey, M. D., and Zhao, Y. (2014) Lysine glutarylation is a protein posttranslational modification regulated by SIRT5, *Cell Metab* 19, 605-617.
106. Bao, X., Wang, Y., Li, X., Li, X. M., Liu, Z., Yang, T., Wong, C. F., Zhang, J., Hao, Q., and Li, X. D. (2014) Identification of 'erasers' for lysine crotonylated histone marks using a chemical proteomics approach, *Elife* 3.
107. Kalesh, K. A., and Tate, E. W. (2014) A succinyl lysine-based photo-cross-linking peptide probe for Sirtuin 5, *Org Biomol Chem* 12, 4310-4313.
108. Liu, Z., Yang, T., Li, X., Peng, T., Hang, H. C., and Li, X. D. (2015) Integrative chemical biology approaches for identification and characterization of "erasers" for fatty-acid-acylated lysine residues within proteins, *Angew Chem Int Ed Engl* 54, 1149-1152.
109. Roessler, C., Nowak, T., Pannek, M., Gertz, M., Nguyen, G. T. T., Scharfe, M., Born, I., Sippl, W., Steegborn, C., and Schutkowski, M. (2014) Chemical Probing of the Human Sirtuin 5 Active Site Reveals Its Substrate Acyl Specificity and Peptide-Based Inhibitors, *Angew Chem Int Edit* 53, 10728-10732.
110. Smith, B. C., and Denu, J. M. (2007) Sir2 deacetylases exhibit nucleophilic participation of acetyl-lysine in NAD(+) cleavage, *Journal of the American Chemical Society* 129, 5802-+.
111. Fatkins, D. G., Monnot, A. D., and Zheng, W. P. (2006) N-epsilon-thioacetyl-lysine: A multi-facet functional probe for enzymatic protein lysine N-epsilon-deacetylation, *Bioorg Med Chem Lett* 16, 3651-3656.

112. Cen, Y. N., Falco, J. N., Xu, P., Youn, D. Y., and Sauve, A. A. (2011) Mechanism-based affinity capture of sirtuins, *Org Biomol Chem* 9, 987-993.
113. Smith, B. C., Settles, B., Hallows, W. C., Craven, M. W., and Denu, J. M. (2011) SIRT3 Substrate Specificity Determined by Peptide Arrays and Machine Learning, *Acs Chem Biol* 6, 146-157.
114. Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J. P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Muller, S., Pawson, T., Gingras, A. C., Arrowsmith, C. H., and Knapp, S. (2012) Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family, *Cell* 149, 214-231.
115. Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Felletar, I., Philpott, M., Munro, S., McKeown, M. R., Wang, Y. C., Christie, A. L., West, N., Cameron, M. J., Schwartz, B., Heightman, T. D., La Thangue, N., French, C. A., Wiest, O., Kung, A. L., Knapp, S., and Bradner, J. E. (2010) Selective inhibition of BET bromodomains, *Nature* 468, 1067-1073.
116. Anders, L., Guenther, M. G., Qi, J., Fan, Z. P., Marineau, J. J., Rahl, P. B., Loven, J., Sigova, A. A., Smith, W. B., Lee, T. I., Bradner, J. E., and Young, R. A. (2014) Genome-wide localization of small molecules, *Nat Biotechnol* 32, 92-+.
117. Bishop, A. C., Buzko, O., and Shokat, K. M. (2001) Magic bullets for protein kinases, *Trends Cell Biol* 11, 167-172.
118. Dar, A. C., and Shokat, K. M. (2011) The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling, *Annu Rev Biochem* 80, 769-795.
119. Fang, Z., Grutter, C., and Rauh, D. (2013) Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features, *Acs Chem Biol* 8, 58-70.
120. Koch, A., and Hauf, S. (2010) Strategies for the identification of kinase substrates using analog-sensitive kinases, *Eur J Cell Biol* 89, 184-193.
121. Schwarzer, D., and Cole, P. A. (2005) Protein semisynthesis and expressed protein ligation: chasing a protein's tail, *Curr Opin Chem Biol* 9, 561-569.

122. Shen, K., Hines, A. C., Schwarzer, D., Pickin, K. A., and Cole, P. A. (2005) Protein kinase structure and function analysis with chemical tools, *Biochim Biophys Acta* 1754, 65-78.
123. Rossetto, D., Avvakumov, N., and Cote, J. (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events, *Epigenetics* 7, 1098-1108.
124. Hengeveld, R. C. C., Hertz, N. T., Vromans, M. J. M., Zhang, C., Burlingame, A. L., Shokat, K. M., and Lens, S. M. A. (2012) Development of a Chemical Genetic Approach for Human Aurora B Kinase Identifies Novel Substrates of the Chromosomal Passenger Complex, *Molecular & Cellular Proteomics* 11, 47-59.
125. Koch, A., Rode, H. B., Richters, A., Rauh, D., and Hauf, S. (2012) A Chemical Genetic Approach for Covalent Inhibition of Analogue-Sensitive Aurora Kinase, *Acs Chem Biol* 7, 723-731.
126. Li, X., Foley, E. A., Kawashima, S. A., Molloy, K. R., Li, Y., Chait, B. T., and Kapoor, T. M. (2013) Examining post-translational modification-mediated protein-protein interactions using a chemical proteomics approach, *Protein Sci* 22, 287-295.
127. Klingberg, R., Jost, J. O., Schumann, M., Gelato, K. A., Fischle, W., Krause, E., and Schwarzer, D. (2015) Analysis of Phosphorylation-Dependent Protein Protein Interactions of Histone H3, *Acs Chem Biol* 10, 138-145.
128. Cedar, H., and Bergman, Y. (2012) Programming of DNA Methylation Patterns, *Annual Review of Biochemistry, Vol 81* 81, 97-117.
129. Pignot, M., Siethoff, C., Linscheid, M., and Weinhold, E. (1998) Coupling of a nucleoside with DNA by a methyltransferase, *Angew Chem Int Edit* 37, 2888-2891.
130. Pljevaljcic, G., Schmidt, F., Scheidig, A. J., Lurz, R., and Weinhold, E. (2007) Quantitative labeling of long plasmid DNA with nanometer precision, *Chembiochem* 8, 1516-1519.
131. Pljevaljcic, G., Pignot, M., and Weinhold, E. (2003) Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA, *J Am Chem Soc* 125, 3486-3492.

132. Schmidt, F. H., Huben, M., Gider, B., Renault, F., Teulade-Fichou, M. P., and Weinhold, E. (2008) Sequence-specific Methyltransferase-Induced Labelling (SMILing) of plasmid DNA for studying cell transfection, *Bioorg Med Chem* 16, 40-48.
133. Pljevaljcic, G., Schmidt, F., and Weinhold, E. (2004) Sequence-specific methyltransferase-induced labeling of DNA (SMILing DNA), *ChemBiochem* 5, 265-269.
134. Comstock, L. R., and Rajske, S. R. (2004) Efficient synthesis of azide-bearing cofactor mimics, *J Org Chem* 69, 1425-1428.
135. Comstock, L. R., and Rajske, S. R. (2005) Conversion of DNA methyltransferases into azidonucleosidyl transferases via synthetic cofactors, *Nucleic Acids Res* 33, 1644-1652.
136. Comstock, L. R., and Rajske, S. R. (2005) Methyltransferase-directed DNA strand scission, *J Am Chem Soc* 127, 14136-14137.
137. Weller, R. L., and Rajske, S. R. (2006) Design, synthesis, and preliminary biological evaluation of a DNA methyltransferase-directed alkylating agent, *ChemBiochem* 7, 243-245.
138. Mai, V., and Comstock, L. R. (2011) Synthesis of an azide-bearing N-mustard analogue of S-adenosyl-L-methionine, *J Org Chem* 76, 10319-10324.
139. Lukinavicius, G., Lapiene, V., Stasevskij, Z., Dalhoff, C., Weinhold, E., and Klimasauskas, S. (2007) Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG), *J Am Chem Soc* 129, 2758-2759.
140. Dalhoff, C., Huben, M., Lenz, T., Poot, P., Nordhoff, E., Koster, H., and Weinhold, E. (2010) Synthesis of S-adenosyl-L-homocysteine capture compounds for selective photoinduced isolation of methyltransferases, *ChemBiochem* 11, 256-265.
141. Zhu, B., Ge, J., and Yao, S. Q. (2015) Developing new chemical tools for DNA methyltransferase 1 (DNMT 1): a small-molecule activity-based probe and novel tetrazole-containing inhibitors, *Bioorg Med Chem* 23, 2917-2927.
142. Kohli, R. M., and Zhang, Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation, *Nature* 502, 472-479.

143. Xu, L., Chen, Y. C., Nakajima, S., Chong, J., Wang, L., Lan, L., Zhang, C., and Wang, D. (2014) A Chemical Probe Targets DNA 5-Formylcytosine Sites and Inhibits TDG Excision, Polymerases Bypass, and Gene Expression, *Chem Sci* 5, 567-574.
144. Xu, L., Chen, Y. C., Chong, J., Fin, A., McCoy, L. S., Xu, J., Zhang, C., and Wang, D. (2014) Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG duplex content during TET-dependent demethylation, *Angew Chem Int Ed Engl* 53, 11223-11227.
145. Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C. A., Hagemeyer, C., and Carell, T. (2011) The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA, *Angew Chem Int Edit* 50, 7008-7012.
146. Samanta, B., Seikowski, J., and Hobartner, C. (2015) Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA, *Angew Chem Int Ed Engl*.

Figure legends

Figure 1: General design of probes for chromatin modifying enzymes. a) Probes targeting enzymes and PTM-binding proteins contain a reactive group that binds to the active site, a specificity-mediating linker that often consist of histone-tail peptides and an attachment site for reporter groups. A photoactivatable crosslinker can be added for covalent attachment of the probes. b) Probes targeting substrate sites of chromatin modifying enzymes utilize derivatives of cosubstrates that donate a unique chemical group to the transfer reactions. c) Reaction of benzophenone mediated photo-crosslinking.

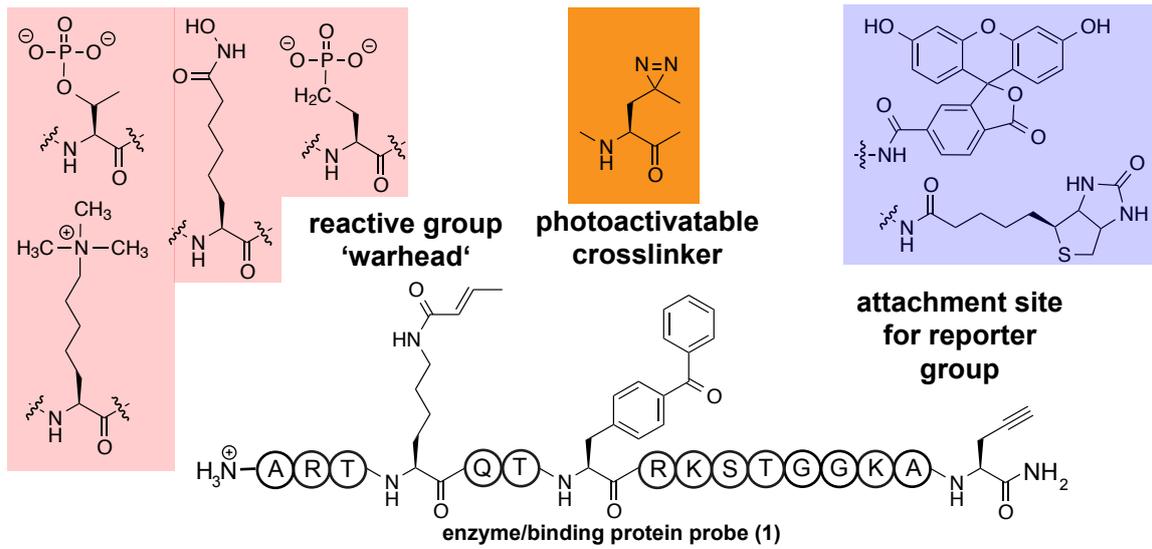
Figure 2: Probing protein methylation. a) Reactions catalyzed by protein lysine methyl-transferases (PKMT) and b) protein arginine methyl-transferase (PRMT). c) Proposed mechanism of arginine labeling by the *N*-mustard derivate SAM analogue 5'-(diaminobutyric acid)-*N*-iodoethyl-5'-deoxyadenosine ammonium hydrochloride (AAI). d) Substrate probes for protein methyl-transferases that donate clickable groups to the transfer reactions. e) Reporter groups for subsequent derivatization of modified PKMT and PRMT substrate sites.

Figure 3: Probing protein demethylation. a) Proposed reaction mechanism of *jumonji* lysine-demethylases. b) Structure of an enzyme probe targeting JMJD. c) Proposed reaction mechanism of LSD1 lysine-demethylase. d) Mechanism-based LSD1 inhibition by propargylamines. e) Enzyme probe targeting LSD1.

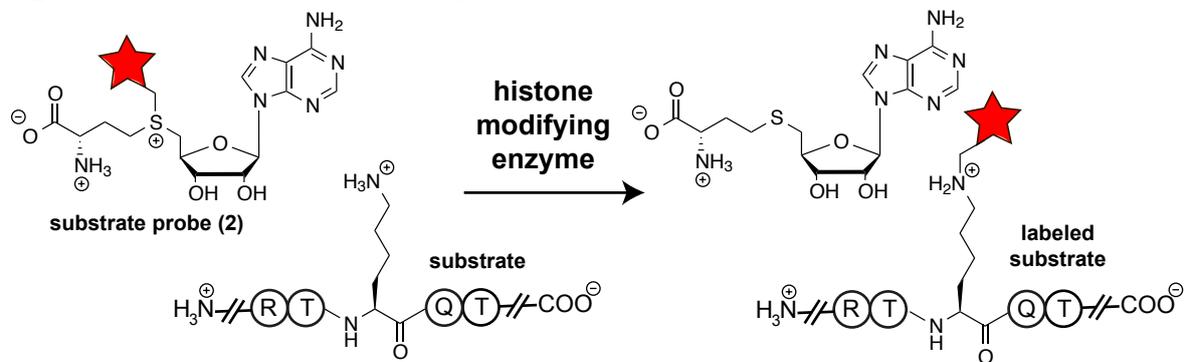
Figure 4: Probing protein acylation and deacylation. a) Reaction of lysine acetyl-transferases (KATs). b) Acyl-coenzyme A derivatives that have been used as KAT substrate probes. c) Probe for selective labeling of Coenzyme A binding proteins. d) Enzyme probe targeting Gcn5. e) Reaction catalyzed and inhibition of Zn²⁺-dependent histone deacetylases (HDACs) by hydroxamic acid inhibitors. f) Chemical probes targeting HDACs. g) Reaction catalyzed by sirtuin-type lysine-deacylases. h) Thioacetyl-Lysine is a mechanism-based inhibitor of sirtuins that binds covalently to the NAD⁺ cofactor. The reaction is similar to the regular deacetylation reaction but the thioimidate intermediate cannot be further processed by sirtuins. i) Structures of bromodomain inhibitor JQ1 and the derived probe Bio-JQ1.

Figure 5: Probing DNA methylation. a) Proposed reaction mechanism of DNA methyl-transferases (DNMTs) that methylate C-5 of cytosine. b) General scaffold of N-adenosylaziridine probes for DNMTs and c) their proposed mechanism. d) Selected structures of enzyme probes for DNMTs and DNMT-substrate labeling. e) Proposed reaction mechanism of TET catalyzed oxidative demethylation of 5mC.

a) general design of probes for enzyme and binding protein probes



b) general concept of cosubstrate probes



c) photoactivatable crosslinker

