SET domain–mediated lysine methylation controls growth in lower organisms and transcriptional machinery in hosts

Chukwuazam Nwasike[1], Sinead Ewert[2], Srdan Jovanovic[2], Shozeb Haider,[2,a] and Shiraz Mujtaba[1,a]

[1]City University of New York, Medgar Evers College, Brooklyn, NY. [2]UCL School of Pharmacy, University College London, London, UK

Correspondence should be addressed to: Shozeb Haider, UCL School of Pharmacy, University College London, London, WC1N 1AX, UK. Shozeb.haider@ucl.ac.uk. Shiraz Mujtaba, City University of New York, Medgar Evers College, Brooklyn, NY 11225. smujtaba@mec.cuny.edu

Running title*: SET protein–mediated methylation*

Keywords*:* epigenetics; SET protein; methyl-transferase; chromatin; histone proteins; lysine methylation

Abstract

Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain–mediated lysine methylation, one of the major epigenetic marks, has been found to regulate chromatin-mediated gene transcription. Published studies have established, further, that methylation is not restricted to nuclear proteins but is involved in many cellular processes, including growth, differentiation, immune regulation, and cancer progression. The biological complexity of lysine methylation emerges from its capacity to cause gene activation or gene repression owing to the specific position of methylated-lysine moieties on the chromatin. Accumulating evidence suggests that despite the absence of chromatin, viruses and prokaryotes also express SET domain–containing proteins, although their functional roles remain under investigated. One possibility is that SET domain–containing proteins in lower organisms have more than one biological function, for example in regulating lower organism growth or in manipulating host transcription machinery in order to establish infection. Involvement of SET domain–containing proteins in host–pathogen interactions would require a thorough understanding of their functions. This review discusses the role of methylation in both prokaryotes and lower eukaryotes, as well as the structural complexity and functional diversity within SET domain–containing proteins.

[a] Contributed equally to the manuscript

Introduction

Lysine methylation plays a central role in regulating growth, development, and immune response in normal conditions, however, its deregulation leads to pathological situations, including cancer, inflammation and developmental disorders [1-3]. Among the 20 ribosomal amino acids, lysine is perhaps the most versatile because it can facilitate the imprinting of diverse chemical modifications, such as acetylation, butyrylation, crotonylation, formylation, hydroxylation, malonylation, methylation, propionylation, ubiquitination, succinylation, and sumoylation [4-10]. Lysine is enzymatically methylated by the addition of methyl group onto the *N*-amine group of its side chain. Functionally, methylation of lysine residues (Kme) unfold at least two forms of complexities, namely, the position-specific methylation of a lysine residue and the degree of methylation, which includes monomethylation (Kme1), dimethylation (Kme2), and trimethylation (Kme3) [11, 12]. The biochemical outcome of methylation, therefore, provides plasticity to chromatin on gene promoters for orchestrating the process of gene regulation required for specific cellular processes. Besides the substrate, the catalytic activity of a lysine methyltransferase (KMT) is dependent upon the cofactor, *S*-adenosyl-L-methionine (SAM), which serves as a donor of methyl moieties [13, 14]. Although the functional significance was unclear, methylation of lysine was originally discovered in 1959 by Ambler and Rees in the flagellin protein of *Salmonella typhimurium* [15]. Subsequently, Murray, in 1964, demonstrated that methylation is prevalent on histone proteins [16]. Taken together, although a myriad of structural, functional, and biochemical studies have elucidated the key role of the SET domain in eukaryote proteins, a role in prokaryote proteins remains to be fully investigated.

The cellular panorama of lysine methylation casts dramatically opposite functional effects, ranging from silencing to activating target genes. For instance, in eukaryotes trimethylation of lysine 27 on histone H3 (H3K27me3) triggers transcriptional silencing, while trimethylation of lysine 4 on histone H3 (H3K4me3) is an epigenetic cue for transcriptional activation [17, 18]. Additionally, methylation of lysine 9 on histone H3 (H3K9me) signals transcriptional repression, which serves as a binding site for heterochromatin protein 1 (HP1) [19]. Further, H4K20 sites can undergo monomethylation (H4K20me1), dimethylation (H4K20me2), or trimethylation (H4K20me3). Though H4K20me1 has been linked to both gene activation and repression [20], H4K20me2 and H4K20me3 are involved in cellular responses to DNA damage and cancer progression. During a cellular response, H4K20me2 serves as a binding site for the tandem Tudor domains of 53BP1, which results in a 53BP1–H4K20me2 complex [21]. On the other hand, a genome-wide investigation revealed that loss of H4K16ac and H4K20me3 is a hallmark for cancer progression [22]; for instance, loss of H4K20me3 marks, which leads to hypomethylation of DNA repetitive sequences in the human genome, is a salient characteristic of cancer cells [22]. Collectively, histone methylation in higher eukaryotes is associated with pivotal cellular functions that if deregulated have the potential to cause a wide-range of diseases.

The biological significance of lysine methylation stems from its ability to alter the nuclear environment and nucleosomal architecture. During X chromosome inactivation (Xi), the transient enrichment of enhancer of zeste homolog 2 (EED–EZH2) and H3K27me3 are characteristic of both embryonic and extraembryonic cells [23]. Subsequently, H3K27me3 triggers chromatin condensation, which represses Xi-related gene expression. During

maintenance of X chromosome silencing, the inactive X chromosome undergoes further modification in which facultative heterochromatin is formed, allowing stabilization of gene repression that is independent of *Xist* expression in somatic cells [24-26]. In a different chromatin scenario, EZH2, together with the polycomb repressor complex, imprints H3K27me3 marks on the promoters of *Hox* genes to regulate embryonic development. *Hox* antisense intergenic non-coding RNA (ncRNA) binds PRC2, which imprints H3K27me3 and thereby triggers silencing of *Hox* loci [27]. In addition to gene silencing, deregulated expression of EZH2 correlates with the aggressiveness of breast and prostate cancer progression [28, 29], as EZH2 is expressed in human mammary epithelial cell lines stimulates anchorage-independent growth and cell invasion that, together, contribute to metastasis of tumor cells. There have been suggestions that overexpression of components of the PRC2 complex could disturb the stoichiometry, resulting in non-functional EZH2 and then uncontrolled proliferation of cancer cells [28, 29].

Prokaryotes, which do not express histone proteins, have genes that encode SET domain–containing proteins. Studies have shown that lysine amino acid of histone-like proteins in *Chlamydia* can be methylated [30]. Methylation also occurs in the pilin of *Synechocystis sp*., which regulates the motility of the cell. Recently, there has been suggestions and evidence that lysine methylation is involved in optimizing bacterial adherence to their environment [31]. Like bacteria and other prokaryotes, archaea lacks a nuclear membrane and histone modifying proteins, so the presence of SET protein is puzzling.

Because there is very limited information on archaea *in vivo*, we perfoemed an evolutionary alignment to identify the archaeal homolog of eukaryotic histone modifying proteins (Fig. 1). The search was limited to archaeal homolog eukaryotic lysine methyltransferase because earlier studies provided evidence that KMT catalyzes methylation. In

Methanosarcina mazei a gene encoding a SET domain–containing protein was identified, confirming Aravind and Iyer's work [32]. Importantly, *Methanosarcina mazei* and its related species are obligate anaerobes, and are the only organisms capable of fermentation of acetate, methylamines, and methanol to methane, carbon dioxide, and ammonia [33]. The sequence alignment in Figure 1 highlights the conservation profile of the SET domain residues in the NHxxxPN motif across lower organisms and archaeal eukaryotes, while the phylogenetic tree emphasises the evolution of the complex post-SET domain containing the Zn^{2+} binding motif (CxCxxxxC) from the simpler non-Zn^{2+} binding organisms (Figure 2).

Molecular and biochemical effects of methylation

Chemical effects of methylation seem limited when compared to other posttranslational modifications such as acetylation, ubiquitination, and phosphorylation. Acetylation neutralizes the positive charge of lysine, as carbonyl dipole is added to enable interactions with bromodomain-containing proteins and deacetylases. On the other hand, phosphorylation rapidly alters the charge of proteins, while increasing the mass of the amino acid side chain, which can be detected using anti-phospho antibodies. In addition, ubiquitin and ubiquitin-like molecules increase the size of proteins—many involved in cell trafficking, transcriptional regulation, and endocytosis—and eventually leads to degradation. Methylation, taking part of relatively fewer modifications compared with other PTM's, can lead to relative increases in the mass of a protein, which can be established by mass spectrometry and anti-methylated antibodies; however, development of site- and position-specific mono-, di-, and tri-methylated antibodies is challenging. Pulse-chase experiment provided evidence that methylation regulates ubiquitination. Further, deregulation of lysine methylation can result in significant effects and

lead to serious diseases, including cancers [56-58]. Upregulation of SMYD2, for example, is associated with esophageal squamous cell carcinoma and bladder cancer [59, 60].

Viral SET protein methylates host chromatin on H3K27 to promote its replication

Viruses engage with transcriptional machinery as well as recruit chromatin-associated proteins to maintain its genome and replication inside the host's cells [34, 35]. However, *Paramecium bursaria* chlorella virus (PBCV-1) is an exception that encodes a SET domain–containing protein, namely, vSET, which methylates host (*Chlorella*) chromatin to silence the transcriptional machinery for enhancing its own replication [11].

Specifically, the transcript of PBCV-1 is detected 5–10 minutes after infection of *Chlorella* cells, which indicates that transition from host to virus transcription is rapid. Interestingly, histone H3 protein from *Chlorella* NC64A demonstrates a high level of sequence homology with human histone H3 protein. This was substantiated by an *in vitro* methylation assay, which confirmed that vSET predominantly catalyzes dimethylation, with mild levels of monomethylation and trimethylation of H3K27 sites. Immunofluoresence investigations confirmed that vSET is present in the nucleus of the *Chlorella* after infection and co-localizes with H3K27me2. The methylation target of vSET is identical to the mammalian H3K27 methyltransferase, EZH2; use of EZH2 knockdown in cells revealed that vSET could mimic the enzymatic action of EZH2 (Fig. 3). Importantly, H3K27me2 generated by vSET serves as a docking site for methyl-lysine binding protein CBX8, a component of PRC1 complex. Taken together, this was one of the first evidence clearly showing that a viral methyltransferase could provide a similar function as a mammalian protein. In addition, a recent study showed that

human cytomegalovirus targets the methyltransferase function of EZH2 to repress the transcriptional activity of GF1 [36].

BaSET regulates growth of Bacillus as well as host's inflammatory response

Human pathogens encode an extensive array of effector proteins that interact with chromatin-mediated transcription process to overcome the resistance of host's immune system [37]. Numerous studies established that toxin secreted by *Bacillus anthracis* devastates host's defense system [38, 39]. However, given severity of anthrax disease, it is highly likely that the profound immunosuppression after *B. anthracis* infection could be facilitated by additional factors. It is also surprising that an independent SET protein that methylates chromatin within eukaryotes, is encoded by the *B. anthracis* (BaSET) (Fig. 3), which lacks histone proteins encapsulating its genome [32]. Moreover, investigations using bioinformatic tools argue that SET proteins in prokaryotes are mainly of evolutionary significance (Fig. 1).

Data from a recently published study suggest that BaSET, unlike other SET proteins specifically trimethylates human histone H1 rather than histone proteins H3 and H4 [19]. To suppress host's immune response most pathogens target the NF-κB-mediated signal pathway, hence, there was a strong possibility that *B. anthracis* utilizes BaSET to repress NF-κB transcriptional functions [12]. Experiments after infecting RAW cells, which is a macrophage derived murine cell line, demonstrated that BaSET secreted by *B. anthracis* is later translocated to the nuclei of macrophages to methylate histone H1, which repress transcription of inflammatory genes [12]. For pinpointing the specific effects of BaSET on the growth curve of *B. anthracis* revealed that BaSET could regulate the septation as well as delay the growth in a SET domain protein deletion mutant of *B. anthracis* BaΔSET. On the other hand, as BaSET is

not required for spore formation and germination of heat-resistant endospore, mice infected with BaSET showed signs of symptoms and died within 4 days while mice infected with mutant BaSET survived. This shows that BaSET plays a major role in the virulence of *B. anthracis* [12].

Usually, histone H3 and H4 are the most common target for methylation by eukaryotic SET proteins, which would normally result in gene activation or repression. However, it was discovered that BaSET methylates histone H1 and mass spectroscopy data suggest that possibly lysine 8 residues could be undergoing methylation. Notably, eukaryotic methyltransferase, like EZH2 methylates more than one lysine residue. Methylation at this site could be an early strategy by the bacillus to hypermethylate host chromatin to silence inflammatory response completely, which will be further devastated by the toxin secretion. Moreover, it is also indicated that histone H1 methylation by BaSET leads to transcription repression of inflammatory genes because of substrate selectivity of BaSET which will help establish *B. anthracis* in the host during the early stages of infection [12].

cpnSET methylates histone like proteins for morphological switch

Chlamydophila pneumoniae (*Chlamydia pneumoniae*) is an obligate intracellular eubacteria, which causes acute respiratory sickness that might involve chronic inflammatory processes, such as atherosclerosis and Alzheimer's disease [40]. However, the underlying molecular mechanism after *Chlamydophila* infection, which leads to progression of chronic disease, remains unclear. Moreover, Chlamydiae exhibit a life cycle that switches between elementary bodies (EBs) and reticulate bodies (RBs). EBs represent transcriptionally inactive while RBs are transcriptionally active bodies [40]. After RBs developmental cycle, it is converted to EBs, which are infectious.

In Chlamydiae, there are two eukaryotic histone H1-like proteins, Hc1 and Hc2, which are mostly present in EBs. Hc1 and Hc2 may be important for the morphological switching between RBs and EBs states. These histone-like proteins are probably regulated by *SET* gene whose proteins are similar to eukaryotic SET domain. In one of the studies, *Chlamydia* SET protein has been shown to interact with Hc1 and Hc2 [41].

A *Chlamydia* SET domain–containing protein is composed of two distinct domains: CDC39 and SET (Fig. 3). The CDC39 domain is well conserved only in *Chlamydia* SET domain–containing proteins. In the genomes of Chlamydiae, *ftsk*, *yycl* and SET domain–containing protein genes are closely located (in that order), which could suggest that they constitute a functional operon. A series of deletion mutants of *cpn*-SET (*C. pneumoniae* SET) reveal that Hc1 and Hc2 bind to substrate-binding region, which is essential for methyltransferase activity. Further, SET protein localization, and its interactions with Hc1 and Hc2 suggest that *cpn*SET methylates Hc1 and Hc2 proteins. Together, these results indicate that *cpn*SET directed methylation of Hc1 protein induces transformation of RBs to EBs state. Furthermore, *cpn*SET also methylates both murine histone H3 *in vitro*, however, *cpn*SET was localized mainly in *Chlamydia* cells, so it is possible that either *Chlamydia* SET protein are exported into the host cell or host histone proteins are transported into *Chlamydia* cells.

The nuclear effector protein of *Chlamydia trachomatis* methylates chromatin

Chlamydia trachomatis is a pathogenic bacterium that infects an estimated 92 million people annually and can be transmitted sexually, which also leads to blindness [30]. As pathogenic gram-negative bacteria, *C. trachomatis* encodes a type III secretion system (TTS) that allows the translocation of protein across a eukaryotic host membrane. Upon biochemical characterization,

Chlamydia SET domain–containing protein CT737, together with similar proteins in other *Chlamydia* species, were designated as nuclear effector (NUE) proteins [30].

NUE proteins sequenced from six *Chlamydia* species, including human pathogens *Chlamydia trachomatis* and *C. pneumonia,* are highly conserved homologs of CT737, which is consistent with the conclusion that NUE proteins play an important role in *Chlamydia spp*. After transfection, NUE localizes in the nucleus of mammalian cells, like murine and human cells with the help of a nuclear localization signal. NUE contains type III secretion (TTS) signal, which directs NUE to enter the nucleus. In the nucleus, results show that NUE is associated with chromatin in infected cells where it methylates H2B, H3, and, to a lesser extent, H4, but it does not methylate *Chlamydia* histone like protein Hc1 [30]. When exposed at equal concentration, mammalian histone prefers NUE compared to *Chlamydia* Hc1. In addition, NUE exhibits automethylation activity [57]. This automethylation primarily occurs within the NUE sequence, which regulates NUE activity towards histone.

Human lysine methyltransferase G9a and EZH2 methylate different sites on histone proteins (Fig. 3). *Chlamydia* Hc1 and Hc2 proteins play a role in compacting DNA during transitions from RB to EB states. Hc1 acts as a substrate of *C. pneumoniae* NUE homolog in addition to histone H3; however, there were no Hc1 methylation *in vitro* studies. NUE is a methyltransferase as well as automethyltransferase, like G9a, which methylates its K239 site.

Methylation of chromatin proteins by archaeal methyltransferase (aKMT)

Archaea draw resemblance to Eukarya as it contains DNA-binding proteins, but exhibits structural and sequence differences (Fig. 3). Some of these differences can be attributed to archaea histone proteins, which lack N- and C- terminal tail, which are the primary sites for

modifications; however, all archaeal chromatin proteins other than histone proteins undergo some form of post-translational modifications [41, 42]. For instance, N- or C- terminal of architectural proteins Sul7d and Cren7 of *Sulfolobus* are monomethylated that affect gene expression and even thermal adaptation [43]. These methylation activities suggest that Archaea encodes lysine methyltransferase (KMT), which is similar to eukaryotic KMT4/DOT1 family. Accumulating evidences show that several DOT1- like proteins are present in Crearchaea and some euryarchaea despite low sequence identity with eukaryotic DOT1 proteins [60]. This DOT1 like protein, termed as aKMT, is quite similar to DOT1 protein of *Saccharomyces cerevisiae*. Nevertheless, aKMT lacks substrate recognition domain and contains a typical DOT1 family methyltransferase catalytic core. Under varied circumstances, aKMT possesses intrinsic methyltransferase activity and interaction between the aKMT and substrates resulted in a hit and run model that is similar to the other eukaryotic DOT1, proposing that aKMT has the same essential function as eukaryotic DOT1 family [43, 44]. Further, evidence reveal that aKMT undergoes automethylation and this methylation can be inhibited by substrates such as Sul7d and Cren7, proteins that are methylated in archaea.

On the other hand, there is little difference between methylated and unmethylated Sul7d and Cren7 in their ability to interact with DNA. However, lysine methylation on Sul7d has been thought to be very extensive as compared to Cren7, but a recent study suggests otherwise [45]. Cren7 is methylated at five lysine residues (K5, K11, K16, K31, and K42). Results show that aKMT4 *in vitro* methylates K11, K16, K36 but not K5 and K42 [46]. Lack of methylation activity on these lysine residues might be because of methylation efficiencies under various experimental conditions. aKMT plays a crucial role in post-translational modification of Cren7, but it is not specific to Cren7. Mutation of Cren7 into various forms showed reduction in

methylation activity but not a complete halt of methylation. Methylation was halted only in the absence of any form of Cren7.

ASHR3 SET methylates H3K36 in *Arabidopsis* to modulate cellular proliferation

In *Arabidopsis thaliana,* the E2F–retinoblastoma (Rb) pathway controls DNA replication. *Arabidopsis* encodes three E2Fs (E2Fa, E2Fb, E2Fc) that can associate with a dimerization partner to gain activity. In association with Dpa, E2Fa causes cell proliferation and increases ploidy; E2Fb acts as an activator of cell proliferation; while E2Fc functions as a repressor [47-49]. Overexpression of Rb-RELATED 1 (RBR1) protein, a transcriptional repressor, reduces E2F expression that causes loss of stem cell identity of root initials [50]. Moreover, the *Arabidopsis* gene *ASH1 RELATED 3 (ASHR3)* product has been identified as a SET domain–containing protein that functions during the cell cycle (Fig. 3) and is regulated by E2Fa/E2Fb.

Previous studies in *Arabidopsis* suggest that ATXR5 and ATXR6 encode an HKTMase, which monomethylates H3K27; but recent studies demonstrate that H3K36 sites can undergo mono and di -methylation also [51-54]. Mutation of *ASHR3* reduces H3K36me1 and to some extent H3K36me2 [54]. However, to establish that *ASHR3* can methylate two substrates lack convincing evidence. Apart from *ASHR3*, other proteins like ASHH and ASHR could potentially catalyze H3K36me [55]. These enzymes could act in other cellular or epigenetic context. Reports show that *ASHR3* may be a target for E2Fa/E2Fb, while repressing or activating cell proliferation [54].

Conclusions

The phylogenetic analysis depicts the evolutionary history of the conserved SET domain/protein (Fig. 1). The branch lengths in the tree are proportional to the differences between the subfamily

clusters. The presence of the cysteine-rich post-SET domain sequence motif (CxCxxxxC) permits the coordination of Zn^{2+} ion, and is an evolutionary divergence from the simpler organisms. In spite of this evolution, the overall architectural fold of the *S*-adenosylmethionine/ *S*-adenosylhomocysteine cofactor-binding site and the post-SET motif–containing region, which coordinates Zn^{2+} ion is highly conserved (Figure 4). The subtle structural differences in the flanking regions around the catalytic site highlight the functional diversity and ability of these proteins to differentially recognize and methylate lysine-containing substrate in different histone and histone-like proteins (Figure 3).

Emerging data have unraveled the biological role of methylation have come a long way in establishing this epigenetic phenomenon; however, further investigations are needed to understand the biological mechanisms that direct mutually exclusive modifications on a given lysine residue. Further, studies on methylation are looking outside of the nucleus to explain the consequences of methylation on non-histone and cellular proteins [61, 62]. Since prokaryotes lack histone proteins, understanding role of methylation on non-histone proteins will help to explain the role of SET domain–containing proteins in prokaryotes.

In eukaryotic organisms, lysine acetylation is mainly associated with altering the chemical landscape of chromatin and gene activation; however, lysine methylation is a complex epigenetic event, as a site-specific methylation can either lead to gene activation or to repression [63]. Furthermore, like an acetylated-lysine moiety that mainly recruits bromodomain-containing proteins, a methylated-lysine moiety recruits a wide-range of proteins that contain chromo, Tudor, PHD, MBT, PWWP, and WDR domains [64]. Therefore, to depict molecular basis of physiological selectivity of histone methylation is relatively more challenging [64]. Moreover, methylation is carried out by a stoichiometry-dependent multi-protein complex, such as PRC2

complex that contains EZH2, which methylates H3K27 site [45]. After methylation, the H3K27me site recruits PRC1 complex that ultimately leads to gene silencing [28]. Given the multi-protein nature of the PRC1 and PRC2 complexes, *in vitro* purification is not simple. Any aberration in the H3K27 methylation can lead to neurological disorders, cancer, or inflammatory disorders [65]. In this kind of pathological situation, a SET domain–containing proteins from lower organisms, which do not function as a multi-protein complex, such as PBCV-1 vSET protein, can serve as a tool to restore methylation marks on H3K27 sites [66]. Furthermore, PBCV-1 vSET protein also has a potential to act as a therapeutic agent to institute repression marks on genes that are either hyperactivated or overexpressed in certain cancers. In addition, SET domain–containing proteins from prokaryotic organisms can serve as a mechanistic probe to investigate complex dynamics of gene activation or repression that occurs during eukaryotic gene regulation.

Acknowledgements

We deeply appreciate the help from members of writing center at Medgar Evers College for their constructive comments. We are also deeply grateful to the MSEIP program at MEC for supporting C.N. and S.M. and to the NIH/NCI for supporting S.M.; S.H. is grateful for a UCL Excellence Fellowship for funding.

References

1. Hanahan, D. and R.A. Weinberg, The hallmarks of cancer. Cell, 2000. **100**(1): p. 57-70.
2. Agger, K., et al., UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature, 2007. **449**(7163): p. 731-4.
3. Crosio, C., et al., Chromatin remodeling and neuronal response: multiple signaling pathways induce specific histone H3 modifications and early gene expression in hippocampal neurons. J Cell Sci, 2003. **116**(Pt 24): p. 4905-14.

4. Choudhary, C., et al., Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science, 2009. **325**(5942): p. 834-40.
5. Chen, Y., et al., Lysine propionylation and butyrylation are novel post-translational modifications in histones. Mol Cell Proteomics, 2007. **6**(5): p. 812-9.
6. Tan, M., et al., Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell, 2011. **146**(6): p. 1016-28.
7. Wagner, S.A., et al., A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics, 2011. **10**(10): p. M111 013284.
8. Wisniewski, J.R., A. Zougman, and M. Mann, Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res, 2008. **36**(2): p. 570-7.
9. Peng, C., et al., The first identification of lysine malonylation substrates and its regulatory enzyme. Mol Cell Proteomics, 2011. **10**(12): p. M111 012658.
10. Weinert, B.T., et al., Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep, 2013. **4**(4): p. 842-51.
11. Mujtaba, S., et al., Epigenetic transcriptional repression of cellular genes by a viral SET protein. Nat Cell Biol, 2008. **10**(9): p. 1114-22.
12. Mujtaba, S., et al., Anthrax SET protein: a potential virulence determinant that epigenetically represses NF-kappaB activation in infected macrophages. J Biol Chem, 2013.
13. Qian, C., et al., Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. J Mol Biol, 2006. **359**(1): p. 86-96.
14. Qian, C. and M.M. Zhou, SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cell Mol Life Sci, 2006. **63**(23): p. 2755-63.
15. Ambler, R.P. and M.W. Rees, Epsilon-N-Methyl-lysine in bacterial flagellar protein. Nature, 1959. **184**: p. 56-7.
16. Murray, K., The Occurrence of Epsilon-N-Methyl Lysine in Histones. Biochemistry, 1964. **3**: p. 10-5.
17. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.
18. Bannister, A.J., R. Schneider, and T. Kouzarides, Histone methylation: dynamic or static? Cell, 2002. **109**(7): p. 801-6.
19. Stewart, M.D., J. Li, and J. Wong, Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. Mol Cell Biol, 2005. **25**(7): p. 2525-38.
20. Karachentsev, D., et al., PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. Genes Dev, 2005. **19**(4): p. 431-5.
21. Pannetier, M., et al., PR-SET7 and SUV4-20H regulate H4 lysine-20 methylation at imprinting control regions in the mouse. EMBO Rep, 2008. **9**(10): p. 998-1005.
22. Fraga, M.F., et al., Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet, 2005. **37**(4): p. 391-400.
23. Plath, K., et al., Role of histone H3 lysine 27 methylation in X inactivation. Science,

2003. **300**(5616): p. 131-5.
24. Wutz, A., Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet, 2011. **12**(8): p. 542-53.
25. Wutz, A., X inactivation: a histone protects from reprogramming by the frog. EMBO J, 2011. **30**(12): p. 2310-1.
26. Wutz, A., RNA-mediated silencing mechanisms in mammalian cells. Prog Mol Biol Transl Sci, 2011. **101**: p. 351-76.
27. Rinn, J.L., et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell, 2007. **129**(7): p. 1311-23.
28. Cao, R. and Y. Zhang, The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr Opin Genet Dev, 2004. **14**(2): p. 155-64.
29. Cao, R. and Y. Zhang, SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell, 2004. **15**(1): p. 57-67.
30. Pennini, M.E., et al., Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen Chlamydia trachomatis. PLoS Pathog, 2010. **6**(7): p. e1000995.
31. Lanouette, S., et al., The functional diversity of protein lysine methylation. Mol Syst Biol, 2014. **10**: p. 724.
32. Aravind, L. and L.M. Iyer, Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit. Cell Cycle, 2003. **2**(4): p. 369-76.
33. Zinder, S.H. and T. Anguish, Carbon Monoxide, Hydrogen, and Formate Metabolism during Methanogenesis from Acetate by Thermophilic Cultures of Methanosarcina and Methanothrix Strains. Appl Environ Microbiol, 1992. **58**(10): p. 3323-9.
34. Mujtaba, S. and M.M. Zhou, Anti-viral opportunities during transcriptional activation of latent HIV in the host chromatin. Methods, 2011. **53**(1): p. 97-101.
35. Mujtaba, S., et al., Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. Mol Cell, 2002. **9**(3): p. 575-86.
36. Sourvinos, G., et al., The downregulation of GFI1 by the EZH2-NDY1/KDM2B-JARID2 axis and by human cytomegalovirus (HCMV) associated factors allows the activation of the HCMV major IE promoter and the transition to productive infection. PLoS Pathog, 2014. **10**(5): p. e1004136.
37. Finlay, B.B. and G. McFadden, Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. Cell, 2006. **124**(4): p. 767-82.
38. Ezzell, J.W. and S.L. Welkos, The capsule of bacillus anthracis, a review. J Appl Microbiol, 1999. **87**(2): p. 250.
39. Turnbull, P.C., Introduction: anthrax history, disease and ecology. Curr Top Microbiol Immunol, 2002. **271**: p. 1-19.
40. Murata, M., et al., Chlamydial SET domain protein functions as a histone methyltransferase. Microbiology, 2007. **153**(Pt 2): p. 585-92.
41. Sandman, K. and J.N. Reeve, Archaeal histones and the origin of the histone fold. Curr Opin Microbiol, 2006. **9**(5): p. 520-5.
42. Forbes, A.J., et al., Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2678-83.
43. Niu, Y., et al., A prototypic lysine methyltransferase 4 from archaea with degenerate sequence specificity methylates chromatin proteins Sul7d and Cren7 in different patterns. J Biol Chem, 2013. **288**(19): p. 13728-40.

44. Frederiks, F., et al., Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. Nat Struct Mol Biol, 2008. **15**(6): p. 550-7.
45. Botting, C.H., et al., Extensive lysine methylation in hyperthermophilic crenarchaea: potential implications for protein stability and recombinant enzymes. Archaea, 2010. **2010**.
46. Chu, Y., et al., Identification and characterization of a highly conserved crenarchaeal protein lysine methyltransferase with broad substrate specificity. J Bacteriol, 2012. **194**(24): p. 6917-26.
47. De Veylder, L., et al., Control of proliferation, endoreduplication and differentiation by the Arabidopsis E2Fa-DPa transcription factor. EMBO J, 2002. **21**(6): p. 1360-8.
48. Magyar, Z., et al., Arabidopsis E2FA stimulates proliferation and endocycle separately through RBR-bound and RBR-free complexes. EMBO J, 2012. **31**(6): p. 1480-93.
49. del Pozo, J.C., et al., The balance between cell division and endoreplication depends on E2FC-DPB, transcription factors regulated by the ubiquitin-SCFSKP2A pathway in Arabidopsis. Plant Cell, 2006. **18**(9): p. 2224-35.
50. Wildwater, M., et al., The RETINOBLASTOMA-RELATED gene regulates stem cell maintenance in Arabidopsis roots. Cell, 2005. **123**(7): p. 1337-49.
51. Raynaud, C., et al., Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. Plant J, 2006. **47**(3): p. 395-407.
52. Jacob, Y., et al., ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol, 2009. **16**(7): p. 763-8.
53. Jacob, Y., et al., Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature, 2010. **466**(7309): p. 987-91.
54. Kumpf, R., et al., The ASHR3 SET-domain Protein Controls Cell Division Competence of the Meristem and Quiescent Centre of the Arabidopsis thaliana Primary Root. Plant Physiol, 2014.
55. Baumbusch, L.O., et al., The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res, 2001. **29**(21): p. 4319-33.
56. Hoffmann, I., et al., The role of histone demethylases in cancer therapy. Mol Oncol, 2012. **6**(6): p. 683-703.
57. Hoffmann, S., G. Tomasik, and Z. Polanski, DNA methylation, histone modifications and behaviour of AKAP95 during mouse oocyte growth and upon nuclear transfer of foreign chromatin into fully grown prophase oocytes. Folia Biol (Krakow), 2012. **60**(3-4): p. 163-70.
58. Truax, A.D., M. Thakkar, and S.F. Greer, Dysregulated recruitment of the histone methyltransferase EZH2 to the class II transactivator (CIITA) promoter IV in breast cancer cells. PLoS One, 2012. **7**(4): p. e36013.
59. Cho, H.S., et al., RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. Neoplasia, 2012. **14**(6): p. 476-86.
60. Cho, H.S., et al., Enhanced HSP70 lysine methylation promotes proliferation of cancer cells through activation of Aurora kinase B. Nat Commun, 2012. **3**: p. 1072.
61. West, L.E. and O. Gozani, Regulation of p53 function by lysine methylation. Epigenomics, 2011. **3**(3): p. 361-9.
62. Chang, Y., et al., Structural basis of SETD6-mediated regulation of the NF-kB network

via methyl-lysine signaling. Nucleic Acids Res, 2011. **39**(15): p. 6380-9.
63. Mujtaba, S., L. Zeng, and M.M. Zhou, Structure and acetyl-lysine recognition of the bromodomain. Oncogene, 2007. **26**(37): p. 5521-7.
64. Allis, C.D., et al., New nomenclature for chromatin-modifying enzymes. Cell, 2007. **131**(4): p. 633-6.
65. Bracken, A.P., et al., EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J, 2003. **22**(20): p. 5323-35.
66. Turgeon, N., et al., The histone H3K27 methylation mark regulates intestinal epithelial cell density-dependent proliferation and the inflammatory response. J Cell Biochem, 2013. **114**(5): p. 1203-15.

Figure legends

Figure 1. A circular evolutionary tree of selected prokaryotic and unicellular eukaryotic versions of SET protein containing organisms. The evolution of complex post-SET domains (red), containing the Zn^{2+} binding motif (CxCxxxxC) is evident from non-Zn^{2+} binding organisms (blue).

Figure 2. Multiple alignment of prokaryotic and unicellular eukaryotic SET domain containing proteins. The coloring represents the conservation profile of amino acid residues >70% (green) and >50% (yellow). The conserved SET domain residues in the NHxxxPN motif (red box) are labeled above the sequences. The post-SET domain consists of cysteine rich motif CxCxxxxC (blue box) and coordinates Zn^{2+} ion.

Figure 3. A phylogenetic analysis and structural comparison of prokaryotic and unicellular eukaryotic SET proteins described in this review. While the signature SET domain NHxxxPN motif is conserved, there are subtle structural differences in the surrounding residues, which demonstrates their ability to methylate different histone and histone-like proteins. The presence of an additional post-SET domain (CxCxxxxC) in complex organisms binds Zn^{2+} ions, and is an evolutionary divergence from the simpler form. The conserved residues in the characteristic NHxxxPN motif, detailing the interactions in the catalytic site have been illustrated. The overall fold of the SET protein is illustrated as surface. The peptide (blue) containing the substrate lysine (red cpk) has been placed in the models to give a sense of position of binding. SAM/SAH are colored in green sticks. The models of the methylation sites have been illustrated for (A) *Saccharomyces cerevisiae* (B) *Giardia lamblia* (C) *Plasmodium falciparum* (D) *Arabidopsis thaliana* (E) *Chalmydia trachomatis* (F) *Chlamydophila pneumoniae* (G) *Paramecium bursaria Chlorella Virus* (H) *Bacillus anthracis* (I) *Methanosarcina mazei*.

Figure 4. Alignment of the models, highlighting the structural conservation of the catalytic site containing the signature SET domain sequence NHxxxPN. The spatial position of the SAM/SAH cofactor is illustrated as green sticks and the presence of Zn^{2+} as white spheres.