Biochimica et Biophysica Acta xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

Molecular basis for substrate recognition by lysine methyltransferases and demethylases ☆

Paul A. Del Rizzo, Raymond C. Trievel *

Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 13 January 2014
Received in revised form 16 May 2014
Accepted 9 June 2014
Available online xxxx

Keywords:
Chromatin
Transcription
Histone lysine methylation
Lysine methyltransferase
Lysine demethylase
Substrate specificity

ABSTRACT

Lysine methylation has emerged as a prominent covalent modification in histones and non-histone proteins. This modification has been implicated in numerous genomic processes, including heterochromatinization, cell cycle progression, DNA damage response, DNA replication, genome stability, and epigenetic gene regulation that underpins developmental programs defining cell identity and fate. The site and degree of lysine methylation is dynamically modulated through the enzymatic activities of protein lysine methyltransferases (KMTs) and protein lysine demethylases (KDMs). These enzymes display distinct substrate specificities that in part define their biological functions. This review explores recent progress in elucidating the molecular basis of these specificities, highlighting structural and functional studies of the methyltransferases SUV4-20H1 (KMT5B), SUV4-20H2 (KMT5C), and ATXR5, and the demethylases UTX (KDM6A), JMJD3 (KDM6B), and JMJD2D (KDM4D). We conclude by examining these findings in the context of related KMTs and KDMs and by exploring unresolved questions regarding the specificities and functions of these enzymes. This article is part of a Special Issue entitled: Methylation Multifaceted Modification — looking at transcription and beyond, edited by Dr. Johnathan Whetstine.

1. Introduction

Lysine methylation in histones was first reported by Murray and by Allfrey and colleagues in the mid-1960s [1,2]. In subsequent decades, relatively little progress was made toward understanding of the biological significance of histone lysine methylation and the identities of the methyltransferases responsible for this modification. In 2000, this situation changed dramatically with the discovery of the first histone-specific lysine methyltransferases (KMTs), mammalian SUV39H1 (KMT1A) and its fission yeast homolog Clr4 (KMT1), which belong to the **S**uppressor of Variegation – **E**nhancer of Zeste – **T**rithorax (SET) domain family of transcriptional regulators [3]. This discovery ushered in the identification of myriad SET domain KMTs that methylate the N-terminal tails of histone H3 at K4, K9, K27, and K36 and histone H4 at K20, as well as the non-SET domain KMT DOT1L that methylates K79 in the globular core of histone H3 [4,5]. Methylation at these sites has been implicated in host of genomic processes including epigenetic gene regulation, DNA replication, DNA repair, and maintenance of genome integrity [5]. Further studies demonstrated that many SET domain KMTs catalyze site-specific methylation of lysine residues in non-histone proteins, including transcription factors and other chromatin modifying enzymes, illustrating that lysine methylation is a widespread post-translational modification in signal transduction [5–7]. In addition to their site specificity, SET domain KMTs can catalyze mono-, di-, or trimethylation of the lysine ε-amine group, an enzymatic property commonly termed product specificity [8–10]. Both the site and state of methylation is important for recognition by effector proteins bearing methyllysine binding domains that elicit the functions associated with methylation of histones and other proteins [11,12]. Thus, the substrate and product specificities of KMTs define the biological roles of these enzymes.

The discovery of the KMTs and correlative studies investigating the functions of lysine methylation raised intriguing questions regarding the existence of lysine demethylases (KDMs) that enzymatically reverse this modification. Despite reports dating back to the 1960s and 1970s of purified cellular extracts displaying KDM activity [13–15], the existence and sequence identities of these enzymes remained enigmatic. The questions surrounding enzymatic lysine demethylation were finally resolved in 2004 with the discovery of first protein lysine-specific demethylase, LSD1 (KDM1), a subunit of the BRAF-35, Co-REST, CtBP, and NURD corepressor complexes that selectively demethylates mono- and dimethylated K4 in histone H3 (H3K4me1/2) [16–20]. The following year marked the identification of the H3K36me2-specific demethylase JHMD1, the first member of the Jumonji C (JmjC) family of Fe(II)- and 2-oxoglutarate (2-OG)-dependent KDMs [21]. Subsequent

☆ This article is part of a Special Issue entitled: Methylation Multifaceted Modification — looking at transcription and beyond, edited by Dr. Johnathan Whetstine.

* Corresponding author at: University of Michigan Medical School, Department of Biological Chemistry, 1150 West Medical Center Drive, 5301 Medical Science Research Building III, Ann Arbor, MI 48109, USA. Tel.: +1 734 647 0889; fax: +1 734 763 4581.
E-mail address: rtrievel@umich.edu (R.C. Trievel).

http://dx.doi.org/10.1016/j.bbagrm.2014.06.008
1874-9399/

studies led to the identification of several subfamilies of JmjC KDMs that display different methylation site specificities for H3K4, H3K9, H3K27, or H3K36 and also exhibit distinct methylation state specificities toward mono-, di-, or trimethyllysines [22–27]. The identification of these enzymes has defined a new paradigm wherein KMTs and KDMs function coordinately to dynamically regulate protein lysine methylation status (Fig. 1) [4,23].

Following these discoveries, structural and functional studies of multiple SET domain KMTs and JmjC KDMs have yielded a wealth of knowledge with respect to their catalytic mechanisms and the molecular bases of their substrate recognition, expanding our understanding of the biological roles of these enzymes [10,28–35]. These studies have also proven foundational in the development of small molecule inhibitors as molecular probes to interrogate their cellular functions and as potential lead compounds in drug design to treat cancer and other diseases linked to aberrant KMT and KDM function [36–39]. This review offers a topical survey of recent structural and functional studies of KMTs and KDMs, focusing on the H4K20 methyltransferases SUV4-20H1 and SUV4-20H2 (KMT5B and KMT5C) [40,41], the H3.1K27 methyltransferase ATXR5 [42], the H3K27 demethylases UTX (KDM6A) [43] and JMJD3 (KDM6B) [44], and the H3K9-specific demethylase JMJD2D (KDM4D) [45].

2. SUV4-20H1 and SUV4-20H2

Methylation of H4K20 is associated with diverse nuclear processes, including transcriptional silencing, DNA damage response, DNA replication, and maintenance of genome integrity [46,47]. In metazoans, H4K20 methylation is primarily mediated by two KMTs: 1) SET8 (also known as PR-SET7 and KMT5A) [48–50] and 2) SUV4-20, which possesses two homologs in mammals termed SUV4-20H1 and SUV4-20H2 [51]. Biochemical and structural studies have established that SET8 is an H4K20 monomethyltransferase [52,53], whereas SUV4-20H1/2 were first reported to function as H4K20 di- and trimethyltransferases that associate with pericentric heterochromatin [51]. However, subsequent studies have defined a more nuanced relationship between SET8 and SUV4-20 in establishing H4K20 methylation states. RNA interference (RNAi)-mediated knockdown and null mutations of SUV4-20 in *Drosophila* resulted in a dramatic decrease in H4K20me2 and H4K20me3 coupled with an accumulation of H4K20me1, whereas RNAi knockdown of SET8 led to a marked increase in unmodified H4K20 [54,55]. Consistent with these findings, a conditional double knockout mouse for the *Suv4-20h1* and *Suv4-20h2* genes exhibited a substantial decrease of H4K20me2 and H4K20me3 with a concomitant increase in H4K20me1 [56]. Together, these findings have led sequential model of H4K20 methylation in which SET8 is predominantly responsible for generating H4K20me1 and SUV4-20 then methylates H4K20me1 to produce H4K20me2 and H4K20me3 [47]. Notably, this model is predicated upon the unusually monomethyllysine substrate specificity of SUV4-20H1/2, distinguishing these enzymes from other SET domain KMTs that methylate unmodified lysines.

To elucidate the molecular basis for this unusual substrate specificity, Southall et al. and Wu et al. determined crystal structures of the catalytic domain of SUV4-20H1/2 in complex with cognate substrates and products [40,41]. The structure of the ternary complex of mouse SUV4-20H2 bound to AdoHcy and an H4K20me2 peptide will be described here, as it provides key insights into substrate recognition and the catalytic mechanism of these KMTs [40]. The SUV4-20H2 catalytic domain comprises a central SET domain flanked by an N-terminal domain (nSET), an inserted motif (iSET) that bisects the SET domain, and a C-terminal region composed of a PostSET domain (denoted as cSET) (Fig. 2A). The SET domain of mouse SUV4-20H2 adopts a canonical β-sheet topology that is characteristic of this methyltransferase family [10,30]. The N-terminal domain is composed of a four α-helical bundle that packs against the β-sheet fold of the SET domain. Following the SET domain is a C-terminal PostSET domain that contains a four cysteine Zn-binding cluster, which has been previously described in the structures of other SET domain KMTs. The iSET region consists of a meandering loop that divides the SET domain, contrasting the iSET regions of other SET domain KMTs that frequently consist of one or more α-helices. Overall, the structure of SUV4-20H2 ternary complex is highly homologous to that of mouse SUV4-20H1 bound to AdoMet [40], as well as to structures of human SUV4-20H1 and SUV4-20H2 [41]. A comparison of the SUV4-20 crystal structure with that of SET8 [52,53] illustrates the overall homology shared by their SET domains, whereas the structures of the nSET, iSET, and cSET regions diverge substantially between the enzymes (Fig. 2A).

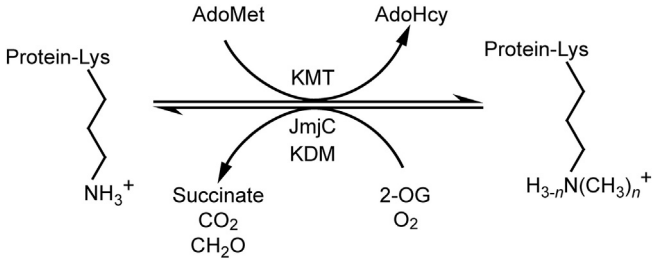

Fig. 1. Dynamic regulation of protein lysine methylation by KMTs and JmjC KDMs.

The crystal structure of the SUV420H2·AdoHcy·H4K20me2 complex provides a conceptual framework for understanding its substrate specificity and how its recognition of the H4K20 site differs from that of SET8 [40]. The H4K20me2 peptide binds in an extended conformation within the substrate binding cleft formed predominantly by the SET domain (Fig. 2A). A parallel β-sheet interaction stabilizes this binding mode through hydrogen bonding between K20 and L22 in the H4K20me2 peptide and the β-strand that demarcates one side of the substrate binding cleft in SUV4-20H2 (Fig. 2A & B). The backbone atoms of R17, H18, and V21 also participate in hydrogen bonding with residues in the substrate binding cleft, further stabilizing the peptide's conformation (Fig. 2B). Specificity for the H4K20 site appears to be mediated in part through recognition of the aliphatic side chains of V21 and L22 through hydrophobic contacts with Tyr164, Trp174, and Pro193 in SUV4-20H2. Surprisingly, most of the polar side chains in the H4K20me2 peptide are predominantly solvent-exposed and do not engage in hydrogen bonding with SUV4-20H2, with the exception of the carboxylate of D24 that is recognized through hydrogen bonds to Ser165 and Arg167. This binding mode contrasts sharply with recognition of the H4K20 site by SET8 which engages in a complex pattern of hydrogen bonds with the side chains of R17, R18, and R23, in addition to hydrogen bonding between the H18 imidazole group and the AdoHcy ribose moiety [52,53]. Despite their differences in H4K20 site specificity, SET8 and SUV4-20H2 share an apparent similarity with respect to the recognition of L22 through the binding of its side chain in a hydrophobic pocket within their respective substrate binding clefts [40,52,53].

Both Southall et al. and Wu et al. investigated the determinants that confer specificity for monomethyllysine in the SUV4-20 KMTs [40,41]. In methyltransferase assays using H4K20, H4K20me1, and H4K20me2 peptides, mouse SUV4-20H1/2 displayed strong preferences toward the H4K20me1 substrate compared to the other substrates, corroborating prior in vivo studies of SUV4-20 specificity for H4K20me1 [40]. In contrast, human SUV4-20H1/2 exhibited modest three-fold differences in their specificity for H4K20me1 over H4K20 and were inactive against H4K20me2 [41]. This variation in specificity may be due to differences in the activities of the mouse and human homologs or to different assay conditions. A comparison of the structures of SET8·AdoHcy·H4K20me1 and SUV4-20H2·AdoHcy·H4K20me2 peptide complexes offers an explanation for the monomethyllysine substrate specificity of the SUV4-20 KMTs. In both ternary complexes, the methylated K20 side chain is bound in an extended conformation in the lysine binding channel through van der Waals interactions with aliphatic part of the lysyl side chain and hydrogen bonding to its ε-amine group (Fig. 2C). In

Please cite this article as: P.A. Del Rizzo, R.C. Trievel, Molecular basis for substrate recognition by lysine methyltransferases and demethylases, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbagrm.2014.06.008