SET8 Recognizes the Sequence RHRK<sup>20</sup>VLRLDN within the N Terminus of Histone H4 and Mono-methylates Lysine 20*

Yinliang Yin, Changdong Liu, Sau Na Tsai, Bo Zhou, Sai Ming Ngai, and Guang Zhu

From the §Department of Biochemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, the ‡Department of Biology, The Chinese University of Hong Kong, Shatin, Hong Kong, and the ¶Department of Physics, The Eastern China Normal University, Shanghai 200062, China

Received for publication, February 14, 2005, and in revised form, June 13, 2005
Published, JBC Papers in Press, June 17, 2005, DOI 10.1074/jbc.M501691200

Methylation of lysine 20 in histone H4 has been proven to play important roles in chromatin structure and gene regulation. SET8 is one of the methyltransferases identified to be specific for this modification. In this study, the minimal active SET domain of SET8 has been mapped to the region of amino acids 195–352. This region completely retains the same methylation activity and substrate specificity as the full-length SET8. The SET domain recognizes a stretch of specific amino acid sequence around lysine 20 of H4 for its methylation activity. Methylation assays with N terminus mutants of H4 that contain deletions and single alanine or glutamine substitutions of charged residues revealed that SET8 requires the sequence RHRK<sup>20</sup>VLRLDN for methylation at lysine 20. The individual mutation of any charged residue in this sequence to alanine or glutamine abolished or greatly decreased levels of methylation of lysine 20 of H4 by SET8. Interestingly, mutation of lysine 16 to alanine, arginine, glutamine, or methionine did not affect methylation of lysine 20 by the SET domain. Mass spectrometric analysis of synthesized H4 N-terminal peptides modified by SET8 showed that SET8 selectively mono-methylates lysine 20 of H4. Taken together, our results suggested that the coordination between the amino acid sequence RHRK<sup>20</sup>VLRLDN and the SET domain of SET8 determines the substrate specificity and multiplicity of methylation of lysine 20 of H4.

Chromosomal nucleosomes in eukaryotic cells consist of an octamer core of histones H2A, H2B, H3, and H4, as well as 146 bp of DNA, which wraps around the histone core. The majority of the histone protein residues fold inside the histone core, whereas their unstructured N-terminal tails stretch outside the nucleosome (1). Covalent modifications of histones, such as acetylation, methylation, and phosphorylation, have been demonstrated to play important roles in the regulation of chromatin structure and gene activity (2–8). Acetylation of histone lysines has been proven to be associated with euchromatin formation and maintenance, as well as gene activation (9, 10). Five lysine residues in the N termini of histones H3 and H4 have been identified to be the main target sites of methylation: lysines 4, 9, 27, and 36 of histone H3 and lysine 20 of histone H4. To date, methylation of H3 lysines has been extensively studied, and the role of tri-methylation of lysine 9 in heterochromatin formation and gene repression has been well documented (4, 6, 11–13). Structures of SET domains of several histone H3 methyltransferases have been elucidated (14–21).

Lysine 20 of H4 has been shown to be methylated to the mono-, di-, or tri-methylation forms in vivo (22–25). Human PR-Set7 and SET8 were first reported to be two methyltransferases responsible for methylation of lysine 20 of H4 (26–28). Suvar4-20h1 and Suvvar20h2 were then identified to be specific for tri-methylation of lysine 20 of H4 (25). In yeast, Set9 was reported to be able to mono-, di-, and tri-methylate lysine 20 of H4 (29). Most methyltransferases specifically methylate a certain lysine residue within either H3 or H4. However, there are also some methyltransferases that can methylate lysines at different sites. In Drosophila, Ash1 can methylate lysine residues 4 and 9 of H3, as well as lysine 20 of H4 (30). Notwithstanding all the studies on histone methyltransferases, the issue of substrate specificity, how a methyltransferase specifically recognizes a lysine residue for methylation, is not well understood and requires more investigation.

The methylation states of lysine 20 of H4 have been demonstrated to be biologically significant. Tri-methylation of H4 lysine 20 is related to heterochromatin formation and gene repression (25). However, correlations between methylation states and the responsible enzymes have not been completely established. Although it has been reported that PR-Set7, which is very similar to SET8 in its primary sequence, is likely to be a mono-methyltransferase, fluorographic data, due to poor sensitivity, do not completely rule out the possibility that PR-Set7 can di- or tri-methylate lysine 20 of H4 (25, 31).

For the purpose of promoting our understanding of the substrate specificity of SET8 and to determine unambiguously the methylation states of lysine 20 of H4 catalyzed by SET8, we investigated the methylation activities of the minimal active SET domain of human SET8, which spans residues 195–352. Specifically, we investigated the ability of the SET domain, which retains the substrate specificity and full enzymatic activities of SET8, to methylate recombinantly expressed histone H4 N-terminal (H4NT)<sup>1</sup> fusion proteins and a synthetic H4NT peptide. The products of H4NT peptide methylation were directly subjected to mass spectrometric analysis. In this study, we demonstrated that for methylation of lysine 20, the SET

---

* This work was supported by grants from the Research Grants Council of Hong Kong (Grant HKUST6139/03M) and High Impact Research Area grant from The Hong Kong University of Science and Technology. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported by an industrial fund from the GreaterChina Technology Group Ltd.

‡ To whom correspondence should be addressed. Tel.: 852-2358-8705; Fax: 852-2358-1552; E-mail: gzhu@ust.hk.

---

1 The abbreviations used are: H4NT, H4 N terminus or H4 N terminal; AdoMet, S-adenosyl-l-methionine; [3H]AdoMet, S-adenosyl-l-[methyl-3H]methionine; Trx, thioredoxin; GST, glutathione S-transferase; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight.
domain recognizes the core sequence RHRK^{20}VLRDN within the N terminus of H4. Through mass spectrometric analysis, we provided direct evidence showing that SET8 specifically mono-methylates lysine 20 of H4. Intriguingly, we found that the mutation of lysine 16 of H4 to alanine, arginine, glutamine, or methionine did not affect the mono-methylation of lysine 20 of H4 by the SET domain of SET8. This result implies that methylation of lysine 20 of H4 by SET8 is not related to the residue lysine 16 of H4.

MATERIALS AND METHODS

Construction of Plasmids—PCR primers were designed to amplify the full-length SET8 gene, as well as a selection of its mutants and truncated forms, using a plasmid containing SET8 cDNA as a template. As shown in Fig. 1A, PCR products were inserted into pGex-4T-1 (Amersham Biosciences) to express GST-tagged SET8, ΔN100SET8, ΔN145SET8, ΔN174SET8, ΔN194SET8, and ΔN204SET8. For the histone H4 N terminus (H4NT), the N-terminal deletion mutants and the single alanine, arginine, glutamine, or methionine substitution mutants shown in Figs. 2A and 3, A and C, primers were designed to generate PCR fragments for cloning into pET32a (a modified version of the pET32a vector, Novagen) or pGex-4T-1 to construct expression plasmids. All of the expression plasmids used in this study were verified by DNA sequencing.

Recombinant Protein Expression—The expression plasmids were transformed into E. coli BL21 (DE3). GST-tagged fusion proteins of SET8 and its mutants were expressed by incubating the bacteria at 18 °C overnight. After induction, bacterial lysates were applied to glutathione-Sepharose columns and washed thoroughly with a washing buffer containing 10 mM KH$_2$PO$_4$ or K$_2$HPO$_4$, and 100 mM NaCl, at pH 7.8. The tag-free SET8s and mutant proteins were expressed similarly in E. coli BL21 (DE3), except that cultures were incubated at 37 °C for 4 h after induction. GST-tagged H4NT and H4NTΔ4 (K^{194}RHRK^{20}VLRDN^{20}) fusion proteins were purified with glutathione-Sepharose affinity chromatography as above but finally eluted with the washing buffer containing 10 mM glutathione. The His-Trx-tagged H4NT and mutant proteins were purified by nickel-nitrilotriacetic acid agarose affinity chromatography. After desalting, all of the H4NT and mutant fusion proteins were used as substrates in methylation assays.

Methylation Assay—Methylation assays were carried out according to the protocol published previously (32) with slight modifications. Free histones (containing the natural four core histones) or free histone H4 (Roche Applied Science) and Trx- or GST-tagged H4NT and mutants were incubated at 37 °C for 4 h after induction. GST-tagged H4NT and H4NTΔ4 (K^{194}RHRK^{20}VLRDN^{20}) fusion proteins were purified with glutathione-Sepharose affinity chromatography as above but finally eluted with the washing buffer containing 10 mM glutathione. The His-Trx-tagged H4NT and mutant proteins were purified by nickel-nitrilotriacetic acid agarose affinity chromatography. After desalting, all of the H4NT and mutant fusion proteins were used as substrates in methylation assays.

Mass Spectrometric Analysis of Methylation Products—To characterize the composition of the SET8 methylation product, a 20-mer H4NT peptide, G$_2$KGGARKRKLVR DINQGIT$_{20}$ (Roche Applied Science) was synthesized. Its calculated molecular mass is 2204 Da. Non-radioactive H4NT and Trx-tagged H4NT (Amersham Biosciences) served as the donor of the methyl group in the methylation assays. The H4NT peptide was incubated with various amounts of ΔN194SET8 for 25 min or other time intervals as specified. Samples from the various methylation treatments were subjected to MALDI-TOF mass spectrometry analysis using the Applied Biosystems 4700 proteomics system (Foster City, CA). For acquisition of mass spectra, 0.5-μl aliquots of samples were dispensed onto the MALDI plate followed by 0.5 μl of a-cyano-4-hydroxycinnamic acid matrix solution (a solution of a-cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% trifluoroacetic acid). All spectra were acquired in positive-ion reflector mode and were accumulated from 500 laser shots. Des-Arg$^1$-bradykinin, angiotensin II, Glu$^2$-fibrinopeptide B, and adrenocorticotropin hormone fragments 18–39 and 7–39 were used as external standards for calibration of MALDI-TOF mass spectrometry spectra.

RESULTS

The Minimal Active SET Domain of SET8 Spanns Residues 195–352 and Retains Full Methylation Activity and Substrate Specificity—It has been reported that SET8 is a methyltransferase specific for methylation of lysine 20 of histone H4, which recognizes free H4NT, nucleosomal, and free histone H4 as substrates (27), and that its active SET domain is confined to the C-terminal portion comprising residues 175–352. To further localize and characterize the SET domain of SET8, methylation assays were performed with recombinant full-length SET8 and a series of SET domain-containing mutants, using free histone H4 as a substrate. Results indicated that amino acids 195–352 of the SET domain of SET8 retained the full activity of the wild type SET8. As shown in Fig. 1B, similar concentrations of full-length SET8, ΔN100SET8, ΔN145SET8, ΔN174SET8, and ΔN194SET8 have similar methylation activity toward histone H4. The continued N-terminal deletion of SET8 to residue 204 led to an almost complete abolition of methylation activity.

As indicated in Fig. 1C, at low concentrations, SET8 and ΔN194SET8 exerted poor methylation activity toward H4. When their concentrations were increased, both SET8 and ΔN194SET8 exhibited strong activity toward free histone H4. This result suggested that the region between amino acids 195 and 352 could be regarded as the minimal active SET domain of SET8. To as substrate preference, ΔN194SET8 can only methylate histone H4, but not H3, H2A, or H2B, indicating that the minimal SET domain of SET8 retains the substrate specificity of full-length SET8 (Fig. 1D). Therefore, in the following experiments to characterize the SET8 substrate and product specificity, ΔN194SET8 was used in place of SET8 as the methyltransferase for methylation of lysine 20 of H4.

The SET Domain Recognizes the Sequence Around Lysine 20 in H4NT for Methylation—To study the substrate specificity of SET8, methylation assays were conducted using recombinant Trx- or GST-tagged H4NT fusion proteins of different lengths. The compositions of the recombinant H4NT fusion proteins are shown in Fig. 2A. In H4NTΔ1a, the first 10 residues were removed; in H4NTΔ1b, residues 11–30 were deleted, including lysine 20; in H4NTΔ2a, residues 11–19 were deleted; in H4NTΔ2b, residues 1–10 and 21–30 were deleted; in H4NTΔ3a, residues 21–30 were deleted; in H4NTΔ3b, residues 1–19 were deleted; and in H4NTΔ4, only residues 16–25 were present. The lysine 20 methylation site was retained in all of the mutants, with the exception of H4NTΔ1b. ΔN194SET8 exhibited similar levels of activity toward Trx-tagged H4NT, as it did toward free histone H4 (Fig. 2B). Of the H4NT deletion mutants, only H4NTΔ1a could be methylated. As expected, since the SET domain can only specifically methylate lysine 20, the H4NT Δ1b peptide (without lysine 20) cannot be methylated. However, methylation was not observed for mutants H4NTΔ2a, D2b, D3a, or D3b either, in which the residues directly connected to the N-terminal or C-terminal side of lysine 20 were deleted (Fig. 2C). In contrast, the H4NTΔ4, which contains only the 10 central residues, could still be methylated by ΔN194SET8 (Fig. 2D). These results suggested that sequences flanking Lys$_{20}^{20}$ of H4 are involved in specific recognition by the SET domain of SET8.

The Sequence RHRKVLRDN of H4NT Is Required for Lysine 20 Methylation—To further study the SET8 substrate specificity, a series of H4NT single alanine, arginine, glutamine, or methionine substitution mutants was generated for testing as substrates in methylation assays. Each of the positively or negatively charged residues within H4NT was mutated in turn (Fig. 3A). Results demonstrated that the mutation of Arg$^3$, Lys$^5$, Lys$^8$, Lys$^{12}$, Lys$^{16}$, or Glu$^{27}$ to Ala did not significantly
affect methylation of lysine 20 of H4 by the SET domain. However, the mutation of either Arg17 or His18 to alanine led to a complete loss of methylation of lysine 20. Similarly, the mutation of Arg19, Arg23, or Asp24 to alanine greatly decreased levels of methylation of lysine 20 (Fig. 3B). These results confirmed that the residues within the core sequence RHRK20VLRDN of H4NT are required for recognition and methylation of lysine 20 by the SET domain of SET8.

To check the relationship between lysine 16 and methylation...
DISTRICT

Distinct methylation states of certain histone lysine residues, or combinations of methylation states of several lysine residues, may exert completely different physiological effects. The factors leading to specific methylation of histone lysine residues by methyltransferases appear to be complex. Different methyltransferases may recognize the same lysine site and catalyze the formation of the same methylation product, such as the human SUV39H1 and SETDB1, both of which tri-methylate lysine 9 of H3 (34, 35). Different enzymes may target the same residue but produce different methylation states, such as SET8 and PR-Set7, which mono-methylate lysine 20 of H4 (26–28), whereas Suv4-20h1 and Suv4-20h2 tri-methylate it (25). A single enzyme can also target different lysine sites for methylation, such as Ash1, which recognizes lysines 4 and 9 of H3 as well as lysine 20 of H4 (30). Most methyltransferases catalyze a single methylation process, such as the human SET8, which mono-methylates lysine 20 of H4. However, Set9 in yeast can methylate H4 lysine to the mono-, di-, or tri-methylated forms. Thus, there are two issues relating to the methylation specificity of methyltransferases for histone lysine residues: (a) product specificity, in that one, two, or three methyl groups are added to the target lysine residues and (b) substrate specificity, which concerns the selection of lysine residue targeted for methylation.

Based on structural studies of the H3 lysine 4 methyltransferase Set7/9 (14, 19), H3 lysine 9 methyltransferase DIM-5 (15, 21), and Rubisco large subunit lysine 14 methyltransferase LSmT (36, 37) and mutation studies of the conserved residue Phe or Tyr residues of the SET domains of some lysine methyltransferases (21, 33), product specificity may be determined...
FIG. 4. Mass spectrometric analysis of methylation of synthesized H4NT peptides by ΔN194SET8. Unlabeled AdoMet was used as the methyl group donor in this experiment. The molar ratio between the H4NT peptide and AdoMet is 1:3. In the left column, A is for unmethylated H4NT peptide, whereas B–E are for methylation reaction mixtures containing H4NT peptide and ΔN194SET8 at molar ratios of 12:1, 40:1, 12:1, and 6:1, respectively. Reactions were incubated at 30 °C for 25 min. The right column is for methylation reaction mixtures containing H4NT peptide and ΔN194SET8 at a molar ratio of 12:1. F–J are for mixtures incubated at 30 °C for 1, 5, 25, and 125 min and overnight, respectively. The molecular mass of the unmethylated H4NT peptide of 20 residues, 2204 Da, and the mono-methylated H4NT peptide, 2218 Da, are marked above the peaks representing unmethylated and mono-methylated H4NT peptides, respectively. The minor peaks, which come from impurities and are not specific for un-, mono, di-, or tri-methylated H4NT peptide, are marked by arrows. A.I., arbitrary intensity unit.
mainly by the space and shape of the methyltransferase catalytic center, especially the lysine access channel (19, 21, 37–39). However, little is understood regarding substrate specificity or methylation site specificity, which is of special significance, as methylation site specificity is the central consideration of the “histone code” hypothesis (2, 40).

For the histone methyltransferases identified in earlier studies, the N-terminal flanking regions of the SET domain or the pre-SET domains were thought to be important for methylation activity and to contribute to the substrate specificity. Our results demonstrated that the SET domain alone retains the full activity and substrate specificity of SET8 and that the N-terminal region of the SETs is not involved in catalysis and substrate differentiation. This did not support the hypothesis that the pre-SET domain participates in determining substrate specificity as the SET8 does not have pre-SET and post-SET domain yet it recognizes the specific sequence within H4NT.

By aligning the primary sequences and comparing tertiary structures, the SET domains of most methyltransferases can be divided into three regions: SET-N, SET-C, and SET-I (38, 40). The SET-N and SET-C regions are relatively consistent in length and have patches consisting of both conserved and highly variable residues. In contrast, the residues within the SET-I regions are poorly conserved, and their lengths can range from tens to hundreds of amino acids. It has been suggested that the SET-I region determines substrate specificity, due to its variability in length and sequence (38). This argument does not appear convincing because some enzymes have dramatically different SET-I regions yet share the same substrate specificity, such as human SUV39H1 and SETDB1.

Comparing the peptide sequence surrounding lysine 20 of H4 (RHRK^{30}VLRDN) with the analogous regions surrounding lysine 4 (ARTK^{34}Q), lysine 9 (QARK^{38}SA), and lysine 27 (QARK^{27}ST) of H3 and lysine 14 of the Rubisco large subunit N-terminus (VGFK^{43}AGV), there appears to be little amino acid sequence similarity among methylation target sequences. In contrast to the other methylation regions mentioned above, the methylation target sequence of H4NT (RHRK^{30}VLRDN) contains many more charged and large residues. Single mutation of any of the charged residues within RHRK^{30}VLRDN either abolished or greatly reduced methylation of lysine 20 of H4 by the SET domain of SET8. This suggests that the methylation process may depend upon specific interactions between the SET domain and the amino acid residues around the target lysine. Therefore, we favor a model in which the substrate specificity and activity of the methylation process is determined by the interactions formed between the variable residues within the SET-N and SET-C regions of SET domain and the residues around the target lysine site, especially the charged residues flanking the target lysine (40).

Establishing the relationships between the covalent modification of histone N-terminal tail peptides and their resulting biological effects is an intriguing scientific problem, especially the subtle interplay between these modifications. It was reported that the acetylation of lysine 16 of H4 inhibits methylation of lysine 20 of H4, and there appeared to be an inverse correlation between these two modifications (26). Our results showed that the mutation of lysine 16 of H4 to alanine, arginine, glutamine, and methionine did not affect the mono-methylation of lysine 20 of H4 (Fig. 3), suggesting that methylation of lysine 20 by SET8 is not directly related to the residue lysine 16 of H4NT.

Although the SET domain retained the full methylation activities of SET8, it failed to stably bind to lysine 20 of H4NT in vitro (data not shown). We postulate that this is due to the “one-step” mono-methylation reaction mediated by the SET domain. This reaction could happen during the ping-pong collision of the target lysine residue and the methyl group within the SET domain-AdoMet complex. In contrast, for tri-methylation, the target lysine residue (to be tri-methylated) would have to remain tightly bound to the SET domain during the cofactor exchange. This is important because only after the first methyl group is added to the target lysine will the second and third methyl groups be added (40). After this study had been submitted for publication, a report describing the structure of SET-domain-H4NT-AdoMet complex appeared (Xiao et al. (41)). The reported structural results (41) are in good agreement with our biochemical and mass spectrometric studies described here.

Acknowledgments—We thank Dr. Rui-ming Xu for providing the plasmid containing the full-length cDNA of SET8 and Dr. Yi Zhang and Dr. Timothy J. Richmond for providing us with the plasmids containing cDNAs of histones H2A, H2B, H3, and H4. We also thank Dr. David Smith and Dr. Zhenguo Wu for proofreading the manuscript.

REFERENCES
SET8 Recognizes RHRKVLRDH and Mono-methylates H4-Lys