Substrate Docking-mediated Specific and Efficient Lysine Methylation by the
SET domain–containing histone methyltransferase SETD7

Haiyang Liu1,2, Zhiwei Li2, Qingqing Yang2, Wei Liu2, Jun Wan1,2, Jianchao Li1,3,*
and Mingjie Zhang1,2,*

From the 1Division of Life Science, State Key Laboratory of Molecular Neuroscience,
Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong
Kong, China; 2Shenzhen Key Laboratory for Neuronal Structural Biology, Biomedical
Research Institute, Shenzhen Peking University-The Hong Kong University of Science
and Technology Medical Center, Shenzhen 518036, China; 3Division of Cell,
Developmental and Integrative Biology, School of Medicine, South China University
of Technology, Guangzhou 510006, China

Running title: MORN repeats docks substrates to SETD7

*To whom correspondence should be addressed to: Jianchao Li (lijch@scut.edu.cn) and
Mingjie Zhang (mzhang@ust.hk)

Keywords: SET domain–containing histone methyltransferase (SETD7), MORN
repeats, lysine methylation, docking interaction, substrate specificity, gene regulation,
protein-protein interaction, posttranslational modification, charge-charge interaction
ABSTRACT

Lysine methylation of cellular proteins is catalyzed by dozens of lysine methyltransferases (KMTs), occurs in thousands of different histone and non-histone proteins, and regulates diverse biological processes. Dysregulation of KMT-mediated lysine methylations underlies many human diseases. A key unanswered question is how proteins, non-histone proteins in particular, are specifically methylated by each KMT. Here, using several biochemical approaches, including analytical gel filtration chromatography, isothermal titration calorimetry (ITC), and in vitro methylation assays, we discovered that SET domain–containing 7 histone lysine methyltransferase (SETD7), a KMT capable of methylating both histone and non-histone proteins, uses its N-terminal membrane occupation and recognition nexus (MORN) repeats to dock its substrates and subsequently juxtapose their Lys-methylation motif for efficient and specific methylation by the catalytic SET domain. Such docking site-mediated methylation mechanism rationalizes binding and methylation of previously known substrates and predicts new SETD7 substrates. Our findings further suggest that other KMTs may also use docking-mediated substrate recognition mechanisms to achieve their catalytic specificity and efficiency.

INTRODUCTION

Methylation of lysine residues was first discovered exactly 60 years ago in bacterial flagellin by Ambler and Rees (1). For the following 40 years or so, lysine methylation was discovered to occur in abundant proteins including mammalian histones (2) and universal Ca$^{2+}$-signaling regulator calmodulin (3,4), and established as a form of enzyme-catalyzed and reversible protein post-translational modification (5,6). However, the biological relevance of lysine methylation has remained largely unclear, though specific Lys115 trimethylation of calmodulin has been implicated in regulating calmodulin target enzyme NAD kinase (6,7). The discoveries of lysine methyltransferase (KMT)-catalyzed histone lysine methylation in regulating gene transcriptions 20 years ago have transformed the protein lysine methylation research into a booming field (6,8,9). Human genome encodes more
than 100 KMTs (10,11), and these enzymes catalyze methylations on numerous proteins other than histones (6,12-15). In fact, based on data curated in Phosphosite Plus database (PhosphoSitePlus®,
www.phosphosite.org), more than 5,000 lysine methylation sites in ~2,760 human proteins have been identified mainly by proteomic-based methods (16).

One of the key challenges facing the lysine methylation research field is to understand which of these identified methylations are truly functional and what the functions of the methylations are. Directly relevant to the above question is how KMTs can specifically recognize their protein substrates. Numerous biochemical and structural studies in the past two decades have uncovered protein substrate binding by various catalytic domains of KMTs (14,17,18). The converging picture is that the catalytic domain of KMTs recognize 2-3 residues flanking both N- and C-termini of methylating Lys residue (14). If each KMT indeed only recognizes such short linear substrate recognition motifs, many KMTs would share overlapping substrates and methylation reactions would be highly promiscuous, a deduction that appears to be contradictory to numerous functional studies of lysine methylations both on histone and non-histone proteins (19). Furthermore, taking several better studied KMTs such as SETD7 (aka SET7/9 and KMT7), G9a (also known as EHMT2), and SMYD2 for examples, each of these enzymes may recognize ~20,000 methylation sites in the human proteome if simply matching with each of their optimal substrate recognition sequence. Thus, it is safe to hypothesize that KMTs contain another layer(s) of substrate recognition mechanism in addition to their catalytic domains. Identification of such KMT-substrate specificity mechanisms will not only be vital for understanding functional implications of each lysine methylation, but also be crucial for selecting KMTs as drug targets for disease therapies.

SETD7 is the first KMT that was identified to be able to methylate Lys residues both in histones and non-histone proteins (20-22). More than 40 different proteins substrates including P53, TAF10, DNMT1, Estrogen receptor α (ERα), E2F1 and Hypoxia-inducible factor
(HIF)-1α have been identified for SETD7 (22-28), but how SETD7 specifically recognizes these substrates is poorly understood. For example, a number of reported SETD7 substrates such as pRb, SIRT1, YAP and β-catenin do not contain the optimal “[RK]-[STA]-K-[pX]”-motif (29-32).

In this study, we discover that the N-terminal MORN repeats of SETD7, being highly negatively charged, can bind to positively charged proteins including a panel of transcription factors through charge-charge interaction. Importantly, we found that the MORN repeats serve as a specific substrate docking site for SETD7, thereby enhancing both the efficiency and specificity of SETD7-mediated Lys methylations. Mechanistically, binding of positively charged sequences/domains both from histone 3 and from non-histone proteins physically position their Lys methylation “[RK]-[STA]-K-[DNSTQ]”-motif for optimal methylation by the catalytic domain. Given that essentially every KMTs contains additional protein-binding domain(s) outside their catalytic core, substrate docking-mediated Lys methylations may be a common mechanism for other KMTs (or even other protein methyltransferases such as Arg methyltransferases).

RESULTS

SETD7 interacts with the DNA binding domain of PDX1 through charge-charge interaction

The crystal structures of SETD7, both in its apo- and substrate-bound forms, showed that the enzyme contains β-strand repeat domain (now known as MORN repeats) physically coupled to the catalytic SET domain (17,35,36) (Figure 1A). The role of MORN repeats in SETD7 is not known and has been proposed to stabilize the catalytic domain of the enzyme, and alterations of the MORN are known to perturb catalytic function of SETD7 (20,35). MORN repeats also exist in other proteins including MORN1-5, junctophilin1-4 and among others, but the role of MORN repeats has remained unknown. We recently discovered that the MORN repeats of MORN4 can specifically bind to a segment in the tail cargo binding domain of unconventional Myo3a (37), providing a direct clue suggesting that MORN repeats can be a protein-protein
interaction module. We therefore hypothesized that the MORN repeats of SETD7 may also function as a protein-protein interaction module, possibly serving as a substrate docking site of the enzyme.

To test the above hypothesis, we dissected the interaction of SETD7 with one of its previously reported binders, PDX1 transcription factor (38). We verified this interaction using purified recombinant full-length proteins of SETD7 and PDX1 through fast protein liquid chromatography (FPLC) (Figure 1E). Further detailed mapping experiments via isothermal titration calorimetry (ITC)-based binding assays revealed that the N terminal MORN repeats (SETD7_MORN, amino acid [aa] 1-194) of SETD7 binds to the DNA binding domain of PDX1 (PDX1_HOX, aa 149-234) with a dissociation constant (Kd) of $\sim 2.3 \mu$M, a value comparable to that of SETD7_MORN/PDX1_FL interaction or SETD7_FL/PDX1_HOX interaction (Figure 1C, 1D and S1A, B&E). Interestingly, deletion of the N terminal 51 residues dramatically diminished the binding (Figure 1C and Figure S1C; this 51-residue was not defined in the crystal structure of SETD7; (17,35)), indicating that the intact MORN repeats are required for binding to PDX1.

The PDX1_HOX is known to bind to specific DNA sequence with a high affinity (39) (Figure S1D). We then asked whether SETD7_MORN and DNA compete for binding to PDX1 using ITC and FPLC based competition experiments. ITC data showed that SETD7_MORN displayed no detectable binding to PDX1_HOX premixed with a stoichiometric amount of the specific DNA duplex (Figure 1D). Consistently, the PDX1_HOX binding DNA duplex specifically disrupted the formation of the SETD7_MORN/PDX1_HOX complex on FPLC-based analysis (Figure S2A). The above competitive bindings suggested that SETD7 and DNA bind to the overlapping positively charged region of PDX1 and the interaction between SETD7 and PDX1 is largely mediated by charge-charge interactions. Indeed, both FPLC- and ITC-based assays showed that the interaction between SETD7 and PDX1 was disrupted by raising NaCl concentration to 300 mM in the assay buffer (Figure 1C and 1E).

Negatively charged SETD7_MORN
binds to the DNA binding domains of many transcription factors

The sequence alignment of the six SETD7 MORN repeats reveals that each repeat consists of two relatively conserved β-strands and a loop in between (Figure 1F). Interestingly, the loop regions of each MORN repeats are enriched with acidic amino acids, which are distributed along one side of MORN repeats forming a highly negatively charged and concave surface (Figure 1F and 1G). This highly negatively charged surface complements the positively charged surface of PDX1 HOX domain (39). The corresponding concave surface in the MORN4 MORN repeats binds to Myo3a with K_d ~2.4 nM (37). Based on the above analysis, we speculated that other highly positively charged proteins including many transcription factors may be potential binders of SETD7.

To test this hypothesis, we selected nine transcription factors from four different classes (40) and tested their bindings to SETD7 using GST pull-down assay (Figure 1H). DNA binding domains of these transcription factors were purified as Trx-fused proteins. Strikingly, every one of these transcription factors could be pulled down by GST-SETD7_MORN in low salt concentration assay buffer (100 mM NaCl; Figure 1H and 1I). Again, all these interactions were dramatically weakened or even disrupted by raising NaCl concentration to 500 mM in the assay buffer (Figure 1H and 1I). We have verified the direct bindings of some of these transcription factors to SETD7 by FPLC by mixing each of purified DNA binding domains with SETD7_FL (Figure S3). We have also shown that a specific MyoD binding DNA duplex (41) specifically disrupted the interaction between SETD7 and MyoD (42) (Figure S2B). The N-terminal tail of histone H3 was reported to be mono-methylated at Lys4 by SETD7 and is highly positively charged. As expected, the N-terminal tail of histone H3 was also pulled down by GST-SETD7_MORN (Figure 1H and 1I). Collectively, these above biochemical data demonstrated that SETD7_MORN can function as a protein recognition module by binding to highly positively charged proteins such as histones and DNA binding domain-containing proteins through charge-charge interaction.

SETD7_MORN is required for histone H3 N terminal tail interaction and
efficient methylation

For histone H3, we hypothesized that the highly positively charged sequence C-terminal to the Lys4 methylation site can bind to the SETD7_MORN and serve to dock the histone protein to the full-length SETD7 for specific and efficient Lys4 methylation (Figure 2A). To test this hypothesis, we took the advantage of the fact that the binding of the methylation site sequence of substrates to SETD7 requires the presence of its cofactor S-adenosylmethionine (SAM) or the product S-adenosyl homocysteine (SAH) (43) (also see Figure S4A and S4B). We could study the binding of various substrate proteins with SETD7, either with or without the presence of SAH, to determine possible existence of SETD7 docking sequence and to isolate the contribution of such docking sequence to the enzyme binding and catalysis (Figure 2B).

ITC assay showed that the histone H3 methylation site peptide (H3_1-11; Figure 2A) binds to SETD7_FL with a very weak affinity (Kd ~276 μM) in the presence of SAH and the binding was undetectable in the absence of SAH (Figure S4B). Using a fluorescence-based binding assay, we determined that an elongated H3 peptide (H3_1-39) binds to SETD7_FL with a Kd value of ~0.8 μM and 1.3 μM in the presence and absence of SAH, respectively (Figure 2B, 2C and Figure S4C). The above results indicated that the residues C-terminal to the methylation sites serve as a specific and strong binding element for histone H3 to bind to SETD7_MORN, and the methylation site of H3 (i.e. H3_1-11) plays a minor role in binding to SETD7. Therefore, we defined the H3_13-39 segment of histone H3 as the SETD7 docking sequence chiefly responsible for the enzyme binding (Figure 2A). This finding is consistent with the observation that mutation of basic residues far from H3K4 in primary sequence decreases the catalytic activity of the enzyme (35). Additionally, removal of the N-terminal 51 residues of SETD7 dramatically weakened its binding to H3_1-39 (Figure 2B and 2C, blue curve), again demonstrating that the intact MORN repeats are required for SETD7 to bind to histone H3.

In vitro methylation assays were then used to investigate the role of the
docking interaction in SETD7-mediated substrate methylations. Methylation of the H3_1-11 peptide was very inefficient (Figure 2D). In contrast, SETD7 catalyzed H3_1-39 methylation with much higher efficiency (~20-fold higher than that of H3_1-11; Figure 2D). The increased methylation on H3_1-39 was not caused by methylation of additional lysine residues, as substitution of Lys4 with Arg totally abolished H3_1-39 methylation by SETD7 (Figure 2D). This finding further indicates that the docking event increased both the efficiency and specificity of the H3K4 methylation by SETD7. To further determine the steady-state kinetic parameters, we performed methylation assays at different substrate concentrations (Figure 2E). Steady-state kinetic studies showed that SETD7 catalyzed the H3_1-39 peptide with a much lower K_M value than did the H3_1-11 peptide (Figure 2E; 27.1 μM for H3_1-39, 1580 μM for H3_1-11). The k_cat value difference for the two H3 peptides was small, and the k_cat value for the H3_1-11 peptide may be with relatively large errors due to the very sluggish reactions (Figure 2E). The overall specificity, as indicated by k_cat/K_M of SETD7 towards H3_1-39, was ~18.5 folds higher than that towards H3_1-11 (Figure 2E).

We further showed that the SETD7_MORN inhibited SETD7-mediated methylation of H3_1-39, but not to H3_1-11, in a dose-dependent manner (Figure 2F). Deletion of the N terminal 51 or 110 residues of SEDT7, corresponding to the first 2 or 4 MORN repeats, significantly decreased methylation of H3_1-39 by the mutant enzymes (Figure 2G red bars). Notably, these two N-terminal truncations did not affect the catalytic activity of SETD7 towards H3_1-11 (Figure 2G black bars), suggesting that deletion of N-terminal MORN repeats did not cause an overall conformational change of SET domain of the enzyme. Taken together, the above biochemical results demonstrated that the MORN repeats of SETD7 functions as the docking site for histone H3 and thereby increase the methylation efficiency and specificity of H3K4 by the enzyme.

MORN repeats-mediated docking enhances non-histone substrate methylation

TAF10 is one of the better known
non-histone substrates for SETD7 (17,23), and it does not contain any sequence capable of binding to SETD7_MORN (data not shown). We therefore used TAF10 as a model to investigate how a docking sequence might influence Lys methylation of non-histone proteins. We constructed a chimeric protein in which the TAF10 methylation peptide was fused to the N-terminus of PDX1_HOX, with a linker length of 16 residues (5 residues from TAF10 and 11 residues from PDX1) (Figure 3A). TAF10 methylation site peptide (TAF10_P, Figure 3A) binds to SETD7 with a quite strong affinity ($K_d \approx 9.6 \mu M$; Figure 3B), as the TAF10_P has the optimal binding sequence for the catalytic domain of SETD7 (14,17). The PDX1_HOX binds to SETD7_FL or SETD7_MORN with a $K_d \approx 2-3 \mu M$ (Figure 1B). The TAF10_P/PDX1_HOX chimera has an enhanced $K_d \approx 0.35 \mu M$ (Figure 3C and 3D), showing that TAF10_P and PDX1_HOX in the chimera cooperatively bind to SETD7. Next, we measured the impact of the docking sequence on the methylation of TAF10_P by SETD7 using the in vitro methylation assay. Due to its relatively strong binding, the methylation TAF10_P by SETD7 is much more efficient than the H3_1-11 peptide (Figure 3E vs Figure 2D). The methylation efficiency of the TAF10_P/PDX1_HOX chimera was significantly higher than the TAF10_P alone. We confirmed that the improvement of the chimera methylation was not due to potential additional methylation site(s) introduced by PDX1_HOX as no methylation could be detected for a chimera mutant with its methyl acceptor Lys substituted by Arg or for PDX1_HOX only as a substrate (Figure 3E). The result in Figure 3E also indicated that SETD7-mediated methylation of the chimera is highly specific, likely due to spatial constrain of the methylation site on TAF10_P by the binding between the SETD7_MORN and PDX1_HOX.

Kinetic studies further showed that fusion of PDX1_HOX to TAF10_P lowered its $K_M$ value of the substrates from 113 $\mu M$ to 8.7 $\mu M$ (Figure 3F). The fusion of the PDX1_HOX, however, did not change the $k_{cat}$ value of TAF10_P methylation (Figure 3F), indicating that there is no obvious allosteric conformational coupling between the
MORN repeats and the SET domain in SETD7. As a result, the fusion of the docking PDX1_HOX domain increased the overall methylation efficiency and specificity (kcat/KM) of TAF10_P by ~14 fold.

Next, we investigated what might be the optimal distance between the docking sequence and the methylation site for SETD7-mediated methylation of the TAF10_P/PDX1_HOX chimera by varying the linker length between TAF10_P and PDX1_HOX. We found that the optimal linker length is around 7 amino acids, and lengthening or shortening of the linker reduced the methylation efficiency of the chimera (Figure 3G). In an extreme case, the mixture of TAF10 peptide and PDX1_HOX, which is equivalent to a chimera with an infinite length linker, displayed a similar level of methylation compared to TAF10_P alone (Figure 3G).

**MORN repeats-mediated substrate docking is a general mechanism for SETD7 to specifically methylate its substrates.**

Our above studies on histone H3 and the TAF10-PDX1 chimera point to a possibility that an optimal substrate for SETD7 should contain a MORN repeats-binding docking sequence and a SET domain-recognition methylation site (Figure 4A). We analyzed the currently identified substrates of SETD7 (28) and noticed that majority of these proteins are DNA/RNA binding proteins, each of which contains at least one positively charged DNA/RNA binding domain (Figure 4B and Table S1). We selected a transcription factor, IRF1, for further analysis. Consistent with our model, the DNA binding domain of IRF1 could bind to SETD7_MORN in a salt concentration-dependent manner (Figure 1H and 1I). The methylation reaction catalyzed by SETD7 towards the methylation site peptide only was inefficient. Inclusion of the DNA binding domain dramatically increased the methylation efficiency of IRF1 (Figure 4C and 4D). Additionally, the efficient methylation mainly occurred on Lys125, as substitution of Lys125 with Arg dramatically weakened IRF1 methylation by SEDT7 (Figure 4D) (44). The above results suggested that the MORN repeats-mediated substrate docking is a general mechanism for SETD7 to specifically and efficiently methylate its substrates.
The above SETD7 substrate recognition model prompted us to search for new substrates of the enzyme. Thousands of matches were found in the entire human proteome when searching for the SETD7 methylation site consensus motif “[RK]-[STA]-K-[DNSTQ]”. Searching a transcription factor database (45) for proteins containing this motif still returned more than 200 proteins. We further restricted the linker length between docking DNA binding domain (excluding Zinc fingers which are hard to define the boundaries) and methylation sites to 50 amino acids or less, and this narrowed the potential targets down to 21 and CDX1 is one of such transcription factors. A recombinant CDX1 containing both the positively charged DNA binding domain and methylation site was very efficiently methylated by SETD7. In contrast, very weak methylation was observed if the methylation site peptide alone was reacted with SETD7 (Figure 4E). Again, SETD7 mainly methylated one Lys in the methylation site (Lys154), and substitution of Lys154 by Arg almost blocked CDX1 methylation by SETD7 (Figure 4D).

It is noted that the methylation site of IRF1 is located C-terminal to the DNA binding domain; and the methylation sites of CDX1, the TAF10_P/PDX1_HOX chimera, and histone H3 are all located N-terminal to their MORN repeats binding domains (Figures, 2A, 3A, and 4C). This analysis suggests that the methylation site peptides of the SETD7 substrates can be situated either N- or C-terminus to their MORN repeats docking domains.

An alternative possibility is that a SETD7 MORN repeats binding protein may function as an adaptor to tether Lys methylation motif-containing proteins to be specifically and efficiently methylated by SETD7 (Figure 4I, right). To test this model, we conducted a proof-of-concept study by dividing CDX1 into two parts (Figure 4F). The DNA binding HOX domain was fused to the N terminus of a proline rich peptide from ARAP1 (denoted as CDX1-P2), and the methylation site of CDX1 was fused to the C terminus of CIN85_SH3B domain (denoted as SH3B-M). CIN85_SH3B has been shown to be able to bind to ARAP1_P2 with sub-micromolar affinity and the binding can be fine-tuned by
point mutations (33) (Figure 4G). Essentially no methylation was observed when the recombinant SH3B-M protein alone was reacted with SETD7, whereas the methylation became efficient when adaptor protein CDX1-P2 was added to the reaction system (Figure 4H). Moreover, weakening the binding between SH3B-M and CDX1-P by point mutations invariably led to decreases of methylation efficiency (Figure 4G and 4H), suggesting that the methylation efficiency was positively correlated with the binding affinity between adaptor protein and substrate protein.

Since DNA can efficiently compete with the MORN repeats for binding to the DNA binding domains of transcription factors (Figures 1D, 1H and S2), we also tested whether DNA can inhibit substrate methylation in vitro. A specific DNA sequence (IRF1-DNA, Figure S5A) which can efficiently bind to transcription factor IRF1 can dramatically inhibit IRF1 methylation by SETD7 (Figures S5A and S5B), whereas a non-specific DNA (CDX1-DNA) can neither bind to IRF1 nor inhibit IRF1 methylation (Figures S5A, S5B and S5F). Similarly, PDX1-specific and CDX1-specific DNA sequences could inhibit the TAF10_P/PDX1_HOX chimera (Fig. 3) and CDX1 (Fig. 4) methylations by SETD7, respectively (Figures S5A, S5C, S5D, S5G and S5H). Notably, a specific DNA sequence can inhibit methylation of a transcription factor with a docking site, but not a peptide containing only a methylation site (Figures S5B and S5C), further demonstrating the critical role of the MORN repeats-mediated docking interaction in substrate methylation by SETD7.

**DISCUSSION**

In this study, we demonstrate that the N-terminal MORN repeats of SETD7 functions as a protein binding module that can bind to a number of highly positively charged proteins. We further show that, via the MORN repeats-mediated substrate docking, SETD7 can efficiently and specifically methylate both histone and non-histone proteins. Our results support a model in which the SETD7_MORN binds to positively charged docking sequence situated not too far away from the Lys methylation motif (i.e. with a limited linker length) thereby facilitating specific methylation
of substrate proteins (Figure 4I, left). An alternative model is that a positively charged SETD7_MORN binding protein functions as an adaptor to tether Lys methylation motif-containing proteins to be specifically and efficiently methylated by SETD7 (Figure 4I, right). Combining the two modes of substrate docking, SETD7 is likely to be capable of specifically methylating a large set of substrates. We have manually curated over 40 proteins that have been confirmed as substrates of SETD7 in the past studies (Table S1). Among them, 69.6% are DNA/RNA binding proteins and 19.6% are binders of DNA/RNA binding proteins (Figure 4B; Table S1), supporting the SETD7 substrate recognition model proposed in Figure 4I.

Our finding that specific DNA sequences can compete with SETD7 for binding to transcription factors suggests that DNA binding may function as a regulatory mechanism for substrate accessibility of SETD7.

The substrate docking-mediated specific Lys methylation mechanism proposed for SETD7 may have general implications to other SET domain containing methyltransferases. Most of the methyltransferases contain additional protein interaction domains either N- or C-terminal to the catalytic SET domains (SMART database; http://smart.embl-heidelberg.de/). For example, apart from SET domain, NSD2 has two PWWP domains and G9a has seven ANK repeats. The PWWP domains of NSD2 and ANK repeats of G9a have been shown to bind to methylated histones, thereby tethering the enzymes to specific chromatin loci and promoting methylation of specific substrates (46–48). By analogy, the WW domain, another common protein interaction domain, in SETD2 may also function as a substrate docking site for their specific methylation by SETD2. It is also noted that most of the protein Arg methyltransferases (PRMTs) contain various protein-protein interaction modules flanking their catalytic domain. It is tempting to speculate that PRMTs may also use analogous substrate docking-mediated Lys methylation mechanism to achieve their catalytic specificities. Therefore, we caution that, when searching for substrates of protein methyltransferases, one should not be limited to the short methylation site motifs of potential substrates, as these
peptide motifs often bind to the catalytic domain of the enzymes with low affinity and high promiscuity.

EXPERIMENTAL PROCEDURES

Constructs and Protein Expression

The full-length SETD7 (NCBI Accession Number: NP_542983.3) and PDX1 (NCBI Accession Number: NP_032840.1) were PCR amplified from a mouse cDNA library. The coding sequences of DNA binding domains of IRF1 (NCBI Accession Number: NP_032416.1, aa 2-139), CDX1 (NCBI Accession Number: NP_034010.3, aa 149-217), Androgen Receptor (AR, NCBI Accession Number: NP_038504.1, aa 535-625), Estrogen Receptor (ER, NCBI Accession Number: NP_031982.1, aa 181-279), GLI3 (NCBI Accession Number: NP_032156.2, aa 431-645), CREB (NCBI Accession Number: NP_034082.1, aa 284-341), c-Myc (NCBI Accession Number: NP_001170823.1, aa 318-439), MyoD (NCBI Accession Number: NP_034996.2, aa 102-170), SOX2 (NCBI Accession Number: NP_035573.3, aa 41-123) were PCR amplified from a mouse cDNA library. The coding sequence of CIN85_SH3B was PCR amplified from the mouse CIN85 cDNA as described (33). All constructs used in this study were cloned into a home-modified pET32a vector except that Chimera constructs were cloned into a PETM.3C vector. In addition to pET32a, SETD7_1-194 was also cloned into a PGEX-6P-1 vector. All truncations, point mutations and fusion constructs were generated with the standard PCR-based mutagenesis method and confirmed by DNA sequencing.

Recombinant proteins were expressed in Escherichia coli BL21 (DE3). The N-terminal thioredoxin-His6-tagged or His6-tagged proteins were purified with a Ni Sepharose™ 6 Fast Flow column and subsequent Superdex 200 prep grade size-exclusion chromatography. GST-tagged proteins were purified with GSH-Sepharose affinity chromatography. Recombinant SETD7 proteins purified through normal procedure are contaminated with cofactor SAM (34), which leads to technical difficulties in ITC titration and in vitro methylation assay. Therefore, we incubated purified SETD7 with excess Trx tagged TAF10 peptide encompassing
methylation site at room temperature overnight to consume SAM. And then ion-exchange and size-exclusion chromatography were performed to remove cofactor product SAH and TAF10 protein.

**Analytical Gel Filtration Chromatography**

Protein samples (typically 180 µL at a concentration of 50-100 µM) were injected into an ÄKTA FPLC system with a Superdex 200 Increase 10/300 GL column (GE Healthcare) using the column buffer of 50 mM Tris-HCl pH 7.5, 100/300 mM NaCl, 1 mM EDTA and 1 mM DTT.

**Isothermal titration calorimetry (ITC) assay**

ITC measurements were performed on a MicroCal iTC200 (Malvern) at 30°C. All proteins were dissolved in a buffer containing 50 mM Tris-HCl pH 7.5, 100/300 mM NaCl, 1 mM EDTA and 1 mM DTT. When substrates (H3, TAF10 peptides and Chimera proteins) were titrated to SETD7, 300 µM SAH was included in the buffer. The concentrations of the protein in the syringe were typically 400-800 µM, while the concentrations of the protein in the cell were typically 40-80 µM. Each titration point was performed by injecting a 2 µL aliquot of the syringe sample into the cell sample at a time interval of 120 s to ensure that the titration curve returned to the baseline. The titration data were analyzed by Origin7.0 (Microcal) and fitted by the one-site binding model.

**GST Pull-Down Assay**

GST or GST-tagged proteins were first loaded to 40 µL glutathione Sepharose beads in an assay buffer (50 mM Tris pH 7.5, 100/500 mM NaCl, 1 mM DTT, and 1 mM EDTA). The GST fusion protein-loaded beads were then mixed with target proteins, and the mixtures were incubated for 2 hours at 4 °C. After extensive washing, proteins captured by affinity beads were eluted by SDS-PAGE sample buffer by boiling, resolved by 15% SDS-PAGE, and detected by Coomassie blue staining. Band intensities were analyzed using ImageJ software.

**Fluorescence Assay**

Fluorescence assays were carried out on a PerkinElmer LS-55 fluorimeter equipped with an automated polarizer at 25°C. For a typical assay, a FITC labeled peptide (~0.1-1 µM) was titrated with a
potential binding partner in 50 mM Tris (pH 7.5) buffer containing 100 mM NaCl, 1 mM DTT, 1 mM EDTA and 300 μM SAH (optional). Curves were fitted by one-site binding model using Prism 7.0 (GraphPad).

**In vitro methylation assay**

Methylation assays were performed in a continuous, high-throughput fashion using the MTase-Glo™ methyltransferase assay kit. For a typical assay, methylation reactions were initiated by addition of 20 μM substrates to a mixture containing 0.05-1 μM SETD7, 40 μM SAM, 2 × MTase-Glo reagent in a PCR tube at 1:1 ratio. The reaction buffer contains 50 mM Tris-HCL (pH 8.0), 100 mM NaCl, 3 mM MgCl₂, 1 mM DTT. After 30 min incubation at room temperature, MTase-Glo™ Detection Solution was added to the reaction at 1:1 ratio and transferred to a 96-well solid white plate. Luminescence was determined at exactly 30 min after addition of the Detection Solution using the GloMax 96 Microplate Luminometer (Promega). The kinetic parameters for the methylation of the TAF10, histone H3 peptides and Chimera protein were measured in duplicate using a MTase-Glo™ methyltransferase assay. A series of concentrations of substrates were added to a mixture containing SETD7 (0.05 μM for TAF10 peptide and Chimera protein, 0.8 μM for histone H3 peptides), 40 μM SAM, 2 × MTase-Glo reagent. After 5 or 20 min incubation, MTase-Glo™ Detection Solution was added to the reaction for another 30 min incubation. Rate plots for substrates methylation were fit to the Michaelis-Menten equation \( V = \frac{k_{\text{cat}}[E][S]}{K_M+[S]} \), using nonlinear least-squares method in Prism 7.0 (GraphPad). The \( K_M \) and \( k_{\text{cat}} \) values were calculated as the average of values derived from three separate experiments, and errors were standard deviation.

**Synthetic peptides**

Peptides of histone H3, TAF10, IRF1 and CDX1 were purchased from Shanghai GL Biochem Co., Ltd. The detailed sequences are as follow: TAF10_P, SKSKDRKYTL; H3_1-39, ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHY; H3_1-11, ARTKQTARKSTY; IRF1, RKERKSSRDTKSY; CDX1, SGKTRTKDKYRVYY. All the above peptides except for TAF10 peptide were
synthesized with a tyrosine residue added to the C terminus to determine peptide concentrations. The FITC labeled peptides were synthesized by Shenzhen PepBiotic Co., Ltd. FITC was labeled onto a lysine residue attached to the very C terminus. Peptide concentrations were determined by measuring their absorbance at 280 nm.

Acknowledgments

This research was supported by a National Key R&D Program grant of China (2016YFA0501903) from the Minister of Science and Technology of China and grants from RGC of Hong Kong (AoE-M09-12 and C6004-17G) to M.Z., the National Natural Science Foundation of China (No. 31670765 to W.L., 31700673 to H.L., and 81571043 to J.W.), the Natural Science Foundation of Guangdong Province (2016A030312016 to M.Z.), and the Shenzhen Basic Research Grants (JCYJ20160229153100269 to M.Z. and JCYJ20170411090807530 to W.L.). M.Z. is a Kerry Holdings Professor in Science and a Senior Fellow of IAS at HKUST.

Competing interests

The other authors declare that no competing interests exist.
References


34. Del Rizzo, P. A., Couture, J. F., Dirk, L. M., Strunk, B. S., Roiko, M. S., Brunzelle, J. S., Houtz,


FOOTNOTES

The abbreviations used are: MORN, membrane occupation and recognition nexus; KMT, lysine methyltransferase; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine; FPLC, fast protein liquid chromatography; ITC, Isothermal titration calorimetry; PRMT, protein Arg methyltransferase

FIGURE LEGENDS

Figure 1. SETD7_MORN binds to PDX1_HOX and other transcription factors through charge-charge interactions. A, Schematic diagram showing the domain organizations of SETD7 and PDX1. B, A table summarizing the ITC-derived binding affinities of SETD7 proteins and PDX1 proteins. C, ITC results comparing binding affinities between SETD7_MORN and PDX1_HOX in the presence of 100 mM NaCl (black) or 300 mM NaCl (red). D, ITC result showing that specific PDX1 binding DNA can disrupt the interaction between SETD7_MORN and PDX1_HOX. E, Analytical gel filtration chromatography showing the binding profiles of SETD7 and PDX1 in low salt buffer and in high salt buffer (insert). F, Sequence alignment of the six MORN repeats of SETD7 showing that the loop regions of these MORN repeats are enriched with negatively charged residues. In the alignment, totally conserved glycine residues are labeled in green, conserved hydrophobic residues are in blue, and negatively charged residues in the loop are in red. G, Surface electrostatic potential of SETD7 52-344 (PDB code: 1H3l) showing that the negatively charged residues in the loop region are distributed in one side of MORN repeats, forming a negatively charged and concave surface. H, GST pull-down assays showing the interactions between GST-SETD7_MORN and an array of DNA binding domains from different transcription factors. I, Quantification of GST pull-down results from three independent experiments. The intensity ratio of [Pull down]/[Input] was calculated to indicate the interaction strength. Error bars show the standard deviations of triplicate experiments.
Figure 2. SETD7_MORN is required for histone H3 N-terminal tail interaction and efficient methylation. A, Sequence analysis showing the methylation site and docking site of the histone H3 N-terminal region. B, SETD7_MORN is required for the histone H3 N-terminal region interaction. The schematic diagram summarizes the binding affinities of H3_1-39 with various forms of SETD7. C, Fluorescence polarization-based measurements of the bindings of H3_1-39 with various forms of SETD7. D, In vitro assay comparing SETD7-mediated methylations of various H3 N-terminal peptides. The concentration of SETD7 used in the assay was 0.5 μM. E, Michaelis-Menten plots comparing the methylation kinetics of H3_1-11 and H3_1-39 by SETD7. F, In vitro methylation assay showing that addition of the SETD7_MORN specifically inhibited SETD7 mediated methylation of H3_1-39, but not of H3_1-11. G, Progressive truncations of the MORN repeats proportionally and specifically weakened methylation of H3_1-39, but had no effect on H3_1-11. The insert shows the methylation fold changes of [H3_1-39]/[H3_1-11] by different SETD7 MORN repeat truncations. For panels D, F and G, error bars are the standard deviations of three different batches of experiments.

Figure 3. Docking-induced enhancement of TAF10 methylation by SETD7. A, Schematic diagram illustrating a chimera composed of TAF10 peptide (TAF10_P) fused with PDX1_HOX. The linker is defined as residues between the methylating Lys and Arg149 of PDX1. B&C, ITC results quantifying the binding affinities of SETD7 to TAF10_P (B) and the Chimera (C). D, ITC results summarizing the binding affinities of SETD7 to TAF10_P, PDX1_HOX, and Chimera. E, In vitro assay of TAF10_P or Chimera methylation by SETD7. The concentration of SETD7 used in the assay was 0.025 μM. F, Michaelis-Menten plots comparing the methylation kinetics of SETD7-mediated methylations of TAF10_P and the Chimera. G, Linker length between methylation site and docking site affects substrate methylation efficiency by SETD7. The detailed linker sequences are indicated in the lower panel. For panels E and G, error bars represent the standard deviations of three different batches of experiments.

Figure 4. Docking-mediated methylation is a general catalytic mechanism for specific substrate methylation by SETD7. A, Schematic diagram showing the
substrates searching strategy. Potential substrates contain a highly positively charged DNA/RNA binding domain capable of docking to the MORN repeats and a consensus methylation site with the sequence motif of “[RK]-[STA]-K-[DNSTQ]”. B, Distribution of currently identified substrates of SETD7 (See Table S1 for detailed substrates information). C, Schematic diagram showing the “Docking site + Methylation site” arrangements of IRX1 and CDX1, two selected examples of SETD7 substrates matching the searching criteria. D&E, In vitro analysis of specific methylation of IRF1 (D) and CDX1 (E) by SETD7. The concentration of SETD7 used in the assay was 0.5 μM. F, Schematic diagram showing the design of adaptor protein binding-mediated substrate methylation by SETD7. The methylation sequence of CDX1 was fused to the C-terminus of CIN85_SH3B domain (denoted as SH3B-M, Lys methylation motif-containing protein). The SH3B-M binding sequence from ARAP1 was fused to the C-terminus of CDX1_HOX domain (denoted as CDX1-P2, adaptor). G, ITC-based measurements summarizing the binding affinities of various SH3B-M proteins to different forms of CDX1-P2. H, Methylation of CDX1-P2 by SETD7 is positively correlated with its binding affinity with SH3B-M. The concentration of SETD7 used in the assay was 0.5 μM. I, Cartoon illustrating docking-enhanced substrate methylation by SETD7. Without docking, a substrate can only be recognized by the SETD7 catalytic domain and the methylation is inefficient and with low specificity (middle). A substrate can directly dock onto the SETD7_MORN (left), or use an adaptor protein to dock onto the SETD7_MORN (right) to achieve efficient methylation and high specificity. For panels D, E and H, error bars represent the standard deviations of three different batches of experiments.
Figure 1

A

MORN repeats
SETD7

1

194

SET

PDX1

149

HOX

234

SET

B

<table>
<thead>
<tr>
<th>SETD7</th>
<th>PDX1</th>
<th>K_d (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>HOX</td>
<td>3.30 ± 0.40</td>
</tr>
<tr>
<td>MORN</td>
<td>FL</td>
<td>7.75 ± 1.35</td>
</tr>
<tr>
<td>MORN</td>
<td>HOX</td>
<td>2.29 ± 0.03</td>
</tr>
<tr>
<td>FL_ΔN51</td>
<td>HOX</td>
<td>26.0 ± 7.7</td>
</tr>
</tbody>
</table>

C

D

E

F

M1  E-E-AVGEHHDDDDLPHGDFCTTVYSTD
M2  T-D-EENF-VHGEKENDG-KPFFD
M3  G-S-TLEAYY-VDDLQQGQG-VITYED
M4  G-VLQOQP-VGDELNPAQYSD-D
M5  GRLITFKQY-KDNRRHVC-WHYFD
M6  GG-SVGEVNEDEMTGKIAVVFDP

H

<table>
<thead>
<tr>
<th>Helix-turn-helix</th>
<th>Input NaCl (mM)</th>
<th>Control</th>
<th>SETD7_MORN</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF1</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>CDX1</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>Zinc-coordinating DNA-binding domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>ER</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>GLI-3</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>CREB</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>Basic domain</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>c-Myc</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>MyoD</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>HMG domain</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>SOX2</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>H3-NT</td>
<td>100</td>
<td>500</td>
<td>100</td>
</tr>
</tbody>
</table>

I

Fold change

[Pull down]/[Input]  

0  0.05  0.10  0.15  0.20

100 mM

500 mM
Figure 2

A

Methylation site (M site)  Docking site (D site)

H3_1-11: ARTKQTARKST
H3_1-39: ARTKQTARKSTGGKAPRKQLATKAAARSAPATGGVKKPH
H3_1-39 K2R: ARTRQTARKSTGGKAPRKQLATKAAARSAPATGGVKKPH

B

C

Binding to H3_1-39

D

Methylation of H3 peptides by SETD7

E

E

H3 peptide  K_d (μM)  k_cat (min^-1)  k_cat/K_d × 10^3
1-11  ~1580  0.52 ± 0.04  0.34 ± 0.07
1-39  27.1 ± 4.2  0.17 ± 0.02  6.3 ± 1.5

F

Enzyme Activity (%)

[HSETD7_MORN] (μM)

G

RLU (× 10^3)

SETD7_FL  SETD7_ΔN51  SETD7_ΔN110

Red  H3_1-39  H3_1-11

Black  H3_1-39  H3_1-11
Figure 3

<table>
<thead>
<tr>
<th>Chimera: SKSDKRKY</th>
<th>TAF10_P</th>
<th>PDX1_HOX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linker</td>
<td>138</td>
<td>254</td>
</tr>
<tr>
<td>(K2R)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D

<table>
<thead>
<tr>
<th>SETD7</th>
<th>Target</th>
<th>(K_a (\mu M))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAF10_P</td>
<td></td>
<td>9.60 ± 0.87</td>
</tr>
<tr>
<td>FL</td>
<td>HOX: 234</td>
<td>3.30 ± 0.40</td>
</tr>
<tr>
<td>Chimera</td>
<td></td>
<td>0.35 ± 0.06</td>
</tr>
</tbody>
</table>

B

\[ \text{SETD7, FL} + \text{TAF10_P} \]

\[ K_a = 9.6 \pm 0.9 \mu M \]

C

\[ \text{SETD7, FL} + \text{Chimera} \]

\[ K_a = 0.35 \pm 0.06 \mu M \]

E

F

\[ V (\mu M/\text{min}) \]

\[ \text{[Chimera/TAF10_P] (\mu M)} \]

G

Change in Chimera

linker length

Substrate \( K_a (\mu M) \) \( k_{\text{cat}} \) (min\(^{-1}\)) \( k_{\text{cat}}/K_a \)

| TAF10_P | 113 ± 9.0 | 9.1 ± 2.0 | 0.08 ± 0.007 |
| Chimera | 8.7 ± 0.4  | 9.2 ± 0.2 | 1.1 ± 0.03   |

Secondary sequences:

2aa: KDRT.............STH
7aa: KEDRTKY...........NKRTH
16aa: KEDRTK...........GAYRFTPENKAYT
23aa: KEDRTK...LKT....GAYRFTPENKAYT
39aa: KEDRTK...LKT(25)...GAYRFTPENKAYT
Figure 4

A

Potential substrate: N

Highly positively charged domains (DNA/RNA binding proteins)

Consensus motif "[RK]-[STA]-[K]-[DNSQ]"

B

60.6%

DNA/RNA-binding protein

19.8%

Binders of DNA/RNA binding protein

Others

C

IRF1

120 RERKSSSS...333 140 KTRTKDKYR...217

M&D site wt: 1 1RFL FERKSSSS...333 140 KTRTKDKYR...217

M&D site K2R: 1 1RFL FERKSSSS...333 140 KTRTKDKYR...217

D

IRF1

M site M&D site M&D site K2R

E

CDX1

M site M&D site M&D site K2R

F

SH3B-M

KTRTKDKYRVV

CDX1-M site

G

SH3B-M CDX1-P2 $K_\text{d}$ (µM)

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>WT</th>
<th>R90E</th>
<th>R90E</th>
<th>E132A</th>
<th>WT</th>
<th>R86E</th>
<th>WT</th>
<th>K2R</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>0.46</td>
<td>0.46</td>
<td>5.62</td>
<td>5.62</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
</tr>
<tr>
<td>WT</td>
<td>0.46</td>
<td>0.46</td>
<td>5.62</td>
<td>5.62</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
</tr>
<tr>
<td>R90E</td>
<td></td>
<td></td>
<td>5.62</td>
<td>5.62</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
</tr>
<tr>
<td>R90E</td>
<td></td>
<td></td>
<td>5.62</td>
<td>5.62</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
</tr>
<tr>
<td>E132A</td>
<td>16.3</td>
<td>16.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>0.46</td>
<td>0.46</td>
<td>5.62</td>
<td>5.62</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
</tr>
<tr>
<td>R86E</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>0.46</td>
<td>0.46</td>
<td>5.62</td>
<td>5.62</td>
<td>16.3</td>
<td>16.3</td>
<td>63.3</td>
<td>63.3</td>
<td>0.43</td>
</tr>
<tr>
<td>K2R</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H

CDX1-M site

SH3B-M CDX1-P2 | CDX1-P2

1 1S3B-M CDX1-P2 | CDX1-P2

2 1S3B-M CDX1-P2 | CDX1-P2

3 1S3B-M CDX1-P2 | CDX1-P2

4 1S3B-M CDX1-P2 | CDX1-P2

5 1S3B-M CDX1-P2 | CDX1-P2

6 1S3B-M CDX1-P2 | CDX1-P2

7 1S3B-M CDX1-P2 | CDX1-P2

8 1S3B-M CDX1-P2 | CDX1-P2

9 1S3B-M CDX1-P2 | CDX1-P2

10 1S3B-M CDX1-P2 | CDX1-P2

11 1S3B-M CDX1-P2 | CDX1-P2

12 1S3B-M CDX1-P2 | CDX1-P2

13 1S3B-M CDX1-P2 | CDX1-P2

14 1S3B-M CDX1-P2 | CDX1-P2

15 1S3B-M CDX1-P2 | CDX1-P2

16 1S3B-M CDX1-P2 | CDX1-P2

17 1S3B-M CDX1-P2 | CDX1-P2

18 1S3B-M CDX1-P2 | CDX1-P2

19 1S3B-M CDX1-P2 | CDX1-P2

20 1S3B-M CDX1-P2 | CDX1-P2


Substrate Docking-mediated Specific and Efficient Lysine Methylation by the SET domain –containing histone methyltransferase SETD7

Haiyang Liu, Zhiwei Li, Qingqing Yang, Wei Liu, Jun Wan, Jianchao Li and Mingjie Zhang

J. Biol. Chem. published online July 19, 2019

Access the most updated version of this article at doi: 10.1074/jbc.RA119.009630

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts