Crystal Structure of Cardiac-specific Histone Methyltransferase SmyD1 Reveals Unusual Active Site Architecture*

Received for publication, July 24, 2010, and in revised form, September 26, 2010. Published, JBC Papers in Press, October 12, 2010, DOI 10.1074/jbc.M110.168187

Nualpun Sirinupong‡, Joseph Brunzelle†, Jun Ye‡, Ali Pirzada‡, Lindsey Nico‡, and Zhe Yang‡1

From the ‡Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201 and the §Advance Photon Source, Argonne National Laboratory, Argonne, Illinois 60439

SmyD1 is a cardiac- and muscle-specific histone methyltransferase that methylates histone H3 at lysine 4 and regulates gene transcription in early heart development. The unique domain structure characterized by a “split” SET domain, a conserved MYND zinc finger, and a novel C-terminal domain (CTD) distinguishes SmyD1 from other SET domain containing methyltransferases. Here we report the crystal structure of full-length SmyD1 in complex with the cofactor analog sinefungin at 2.3 Å. The structure reveals that SmyD1 folds into a wrench-shaped structure with two thick “grips” separated by a large, deep concave opening. Importantly, our structural and functional analysis suggests that SmyD1 appears to be regulated by an autoinhibition mechanism, and that unusually spacious target lysine-access channel and the presence of the CTD domain both negatively contribute to the regulation of this cardiovasculary relevant methyltransferase. Furthermore, our structure also provides a structural basis for the interaction between SmyD1 and cardiac transcription factor skNAC, and suggests that the MYND domain may primarily serve as a protein interaction module and cooperate SmyD1 with skNAC to regulate cardiomyocyte growth and maturation. Overall, our data provide novel insights into the mechanism of SmyD1 regulation, which would be helpful in further understanding the role of this protein in heart development and cardiovascular diseases.

Regulation of multipotent cardiac progenitor cell expansion and subsequent differentiation into cardiomyocytes, smooth muscle, or endothelial cells is a fundamental aspect of cardiovascular biology and cardiac regenerative medicine (1). But the roles of epigenetic modifications, especially histone methylation, in cardiovascular biology and early heart development still remain poorly understood. Recent studies identified that SmyD1, a heart- and muscle-specific histone lysine methyltransferases (HKMT), is a key transcriptional regulator essential for cardiomyocyte differentiation in mouse and myogenesis in zebrafish (2, 3). Further studies showed that SmyD1 expression is regulated by serum response factor as well as myogenin through their direct interaction with the SmyD1 promoter sequence, whereas normal expression of Hand2, a transcription factor required for right ventricular development, is dependent upon SmyD1 in cardiomyocyte precursors (2, 4). These data clearly suggested that SmyD1 is involved in a transcriptional network that is essential for the development of secondary heart field lineage (1). There is also emerging evidence that this protein is associated with cardiovascular diseases. Overexpression of SmyD1 was found in the explanted hearts of patients with end stage heart failure, and was correlated with dysregulation of several ion channels (5). In addition, it was reported that SmyD1 is essential for establishing heart repolarization gradient and normal cardiac contraction, suggesting that SmyD1 might contribute to abnormal heartbeat rhythm (6). These observations thus implied that SmyD1 could be a new drug target for improving treatment in these heart disorders.

SmyD1 belongs to a subfamily of SET containing proteins that consists of five members (SmyD1–5) (7). In vitro studies have shown that SmyD1, -2, and -3 specifically methylate histone H3 lysine 4 (3, 7, 8), suggesting that they may function as a transcriptional activator. However, several reports indicated that SmyD1 functions as a transcriptional repressor in cardiomyocyte maturation and heartbeat rhythm regulation and its repressive activity is dependent on the recruitment of class I histone deacetylase (2, 6). Interestingly, the evidence showed that SmyD1 and other members of the SmyD family methylate histone H3 only to a very limited extent, but their activity can be significantly stimulated by the molecular chaperon HSP90 (3, 7, 8). The exact mechanism of this activity enhancement is unknown, but it was suggested that a conformational change might occur upon HSP90 binding (9). Besides histone methylation, some members of this protein family can methylate non-histone proteins. SmyD2 represses p53 function via methylation at Lys770, whereas SmyD3 methylates Lys831 of vascular endothelial growth factor receptor 1 essential for tumorigenesis (10, 11).

All the characterized HKMTs contain a conserved SET domain with only one exception (12). The SET domain, about 130 amino acids, is the catalytic motif responsible for lysine methylation. The crystal structures of several HKMTs from other SET protein families have been reported (13–16). These structures revealed that the SET domain shares a conserved α/β topology that is completely different from the canonical AdeMet-dependent methyltransferase fold, although both domains use S-adenosylmethionine (AdoMet)2 as a methyl group donor. However, SmyD1 reveals an unusual active site architecture characterized by a “split” SET domain, a conserved MYND zinc finger, and a novel C-terminal domain.

* This work was supported by the American Heart Association.

The atomic coordinates and structure factors (code 3N71) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

† To whom correspondence should be addressed: 540 E. Canfield St., Detroit, MI 48201. E-mail: zyang@med.wayne.edu.

2 The abbreviations used are: AdoMet, S-adenosylmethionine; CTD, C-terminal domain.
**Crystal Structure of SmyD1**

Distinct from other HKMTs, SmyD family proteins have a unique “split” SET domain, which is divided by the MYND domain, a conserved zinc finger motif, into the S-sequence and the core SET domain (2). The significance of this split feature is unknown, but it has been shown that the S-sequence plays a repressive role in SmyD3 activity (9). The existence of two additional conserved domains, MYND and CTD, also distinguishes SmyD proteins from other HKMTs. The MYND domain, a C4C2HC-type zinc finger often present in transcriptional regulators such as leukemogenic fusion protein AML1/ETO and tumor suppressor BS69 (17, 18), has been identified as a protein-protein interaction module, mediating specific interaction with a proline-rich sequence. In SmyD1, this domain is required for interaction with cardiac transcription factor skNAC, and it may function as an adaptor that brings SmyD1 and skNAC together to regulate muscle development (19). Function of the novel C-terminal domain (CTD) remains unknown, but because CTD is immediately preceded by the SET domain whose C-terminal region is essential for catalysis (14, 20), we speculated that this domain might function as a regulatory motif to modulate histone methylation. However, no structure has been reported for any member of the SmyD family, and the exact mechanism of SmyD1 regulation is not clear. To reveal the molecular architecture of SmyD1 and provide insights into the mechanism of SmyD1 regulation, here we report a 2.3-Å crystal structure of full-length SmyD1 in complex with the cofactor analog sinefungin. The crystal structure reveals the significant differences between the active sites of SmyD1 and other HKMTs, shows that the CTD domain interferes with substrate histone binding, and also allows the identification of the putative skNAC binding site.

**EXPERIMENTAL PROCEDURES**

**Protein Preparation**—Mouse Smyd1 was cloned into the pSUMO vector (LifeSensors), with an N-terminal His$_6^{-}$SUMO tag. Recombinant Smyd1 was then transformed into Escherichia coli strain BL21-CodonPlus(DE3)-RIP (Stratagene) for protein expression. The transformants were grown to an $A_{600}$ of 0.4 at 37 °C in 2 liters of LB medium, and then induced with 0.1 mM isopropyl thio-$eta$-d-galactoside overnight at 15 °C. The cells were harvested, and lysed by French Press. The soluble fraction was then subjected to a series of chromatography purifications by an AKTA purifier system (GE Healthcare), including His tag affinity purification by the Hitrap HP column, ion-exchange chromatography with the HiTrap Q HP column, and gel filtration chromatography with the Superdex 200 column. The His$_6^{-}$SUMO tag was cleaved off with yeast SUMO protease 1 immediately after affinity purification. SmyD1 proteins were finally purified to apparent homogeneity and concentrated to 10–20 mg/ml in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM $eta$-mercaptoethanol, and 5% glycerol.

**Crystallization and Data Collection**—Prior to crystallization, SmyD1 (10 mg/ml) was incubated with 2 mM sinefungin at 4 °C for 2 h. The binary complex of SmyD1-sinefungin was then crystallized by hanging drop vapor diffusion at 20 °C, with 12% PEG8000, 2.5% glycerol, 200 mM NDSB-205, 100 mM MES (pH 6.5). Crystals typically appeared within 1 day and achieved their full size in 1 week. X-ray diffraction data from single crystals were collected at beamline 21IDD at the Advanced Photon Source (Argonne, IL), and then were processed and scaled using the program HKL2000 (26). The crystals belong to the rhombohedral space group R3 with unit cell dimensions of $a = b = 170.5$ Å, $c = 57.1$ Å and contain one molecule in the asymmetric unit.

**Structure Determination and Refinement**—The crystal structure of SmyD1 was solved by the multiple wavelength anomalous diffraction method using three intrinsic zinc ions. Initial phases were obtained from zinc anomalous data sets collected at three different wavelengths (inflection, remote, and peak). The program SOLVE (27) was able to identify all three zinc sites, and the solution has an overall $Z$-score of 35.6 and a figure of merit of 0.49 in the resolution range 20–30 Å. After density modification with the program RESOLVE (27), the resulting electron density map is interpretable, displaying sufficient features of the SmyD1 structure. After merging the phases from RESOLVE with a 2.3-Å resolution native data set using CAD of CCP4 suite (28), automated model building was carried out by RESOLVE, which was able to build 67% of the protein residues including side chains. The model was then completed and improved by alternating cycles of manual model building and refinement using COOT (29) and CNS (30). The final refined model, which is well ordered with the exception of the last 22 residues, has working and free $R$-values of 20.9 and 25.6%, respectively. All figures of three-dimensional representations of the SmyD1 structure were made with PyMOL.

**Site-directed Mutagenesis**—SmyD1 mutants were prepared using the Plusion Site-directed Mutagenesis Kit (New England Biolabs) according to the manufacturer’s instructions. Briefly, point mutations were introduced with the phosphorylated primers using PCR, in which pSUMO-SmyD1 (see above) containing the wild-type SmyD1 gene was used as PCR template. The amplified linear PCR product was then circularized in a ligation reaction with Quick T4 DNA Ligase (New England Biolabs). After transforming into competent E. coli DH5$\alpha$ cells, the plasmid sequences were verified by DNA sequencing. Subsequently, all SmyD1 mutants were expressed and purified using the same procedures as described for the wild-type protein.

**Histone Methyltransferase Assay**—Histone methyltransferase assay using the purified proteins was carried out similarly to that described previously (3). Briefly, SmyD1 proteins (5 µg) were incubated with 2 µg of recombinant histone H3 (New England Biolabs) and methyl donor AdoMet (New England Biolabs) at 20 °C overnight in a mixture of 20 µl of reaction buffer containing 50 mM Tris-HCl (pH 8.5), 100 mM NaCl, 10 mM DTT. Set7/9, which was purified using the same procedures as SmyD1 proteins, was used as a positive control. The reactions were analyzed by Western blotting using an antibody against mono-, di-, and trimethylated H3K4 (Millipore). To determine enzymatic activities, immunoblots were quantified based on chemiluminescence using a CCD gel imager (UVP Chemidoc).
Biotinylated Histone Peptide Pulldown Assay—To investigate the interaction between SmyD1 and substrate histone, a biotinylated histone peptide pulldown assay was performed. Histone H3 peptide used for the pulldown assay was purchased from AnaSpec, containing amino acids 1–21 with a biotin label at its C terminus. In the pulldown assay, 2 μg purified proteins were incubated with 1 μg of biotinylated H3 peptide at 4 °C overnight in 20 μl of binding buffer (50 mM Tris-HCl, pH 8.5, 100 mM NaCl, 10 mM DTT, 0.1% Triton X-100, 0.1% BSA), and then the resulting protein-peptide complexes were pulled down with the streptavidin beads (Pierce). After washing five times in the binding buffer, the beads were subjected to SDS-PAGE and then analyzed by Coomassie Blue staining or Western blotting using anti-SmyD1 antibody (Abcam).

Coordinates—Atomic coordinates and structure factor files have been deposited in the Protein Data Bank, accession code 3N71.

RESULTS

Wrench-shaped SmyD1 Structure—The crystal structure of SmyD1 in complex with sinefungin has been determined at 2.3 Å (Table 1). The structure reveals that SmyD1 folds into five distinct structural domains, with overall dimensions of ~100 × 50 × 50 Å (Fig. 1). The shape of the SmyD1 structure resembles an open-ended wrench, where the two thick “grips” are separated by a large, deep concave opening. The N-terminal grip is formed by 4 domains. The SET domain, which is located in the middle of this grip, is surrounded by the MYND, SET-I, and post-SET domains. A remarkable feature of the SmyD1 structure is the presence and location of the CTD domain. This domain forms both the C-terminal grip and the “handle” of the wrench-shaped structure. Notably, the putative histone binding site is located at the bottom of the concave opening, suggesting that both the N- and C-terminal grips are involved in substrate recognition.

Split SET Domain with Highly Variable Flanking and Insertion Regions—The SET domain is an evolutionarily conserved motif that catalyzes lysine methylation. SmyD1 features a unique split SET domain characterized by two separated segments (Fig. 2), the S-sequence (residues 1–49) and the core SET domain (residues 181–258), divided by MYND and SET-I domains. The potential importance of this split feature has been implied by previous studies showing that the lack of S-sequence dramatically enhanced the SmyD3 activity in cancer cells (9). However, our structure reveals that despite the split, the above two segments still come together to form a conserved SET domain fold, consisting of one central 3_10 helix (3_10) and 10 ß-strands (ß1–ß10) that are arranged into 4
antiparallel β-sheets (Fig. 1). Moreover, contrary to its role in SmyD3, we observed that the SmyD1 S-sequence appears essential for protein structural integrity, participates in the formation of all the SET domain β-sheets, and also makes numerous contacts with sinefungin (Fig. 4A). Our point mutation studies on the S-sequence confirmed these observations, which will be addressed later in this article (Fig. 4C).

The SET domain alone, however, is not sufficient for lysine methylation, three other domains are required, including N- and C-terminal flanking domains (pre-SET and post-SET) as well as insertion SET-I domain. The structures of these domains are highly variable in the known SET proteins but they occupy similar positions and play similar roles in these enzymes. It is clear that the post-SET and SET-I domains are involved in cofactor binding and substrate binding. Variations in the sequence and structure of these domains play a key role in determining the substrate specificity of different SET enzymes (15). The exact function of pre-SET is still unclear, but it has been proposed that this domain might function to stabilize SET domain fold or provide an extended histone binding site (14). Interestingly, SmyD1 does not contain a pre-SET domain, and its polypeptide chain begins with the SET motif. Our structure reveals that the post-SET domain, which is immediately downstream of the SET domain, is a small cysteine-rich region consisting of 3 short α-helices (αE, αF, and αG) that are organized around a single zinc ion (Fig. 1). The zinc ion is coordinated by 4 highly conserved cysteine residues: Cys274, Cys276, Cys279 from post-SET, and Cys208 from the SET domain. This zinc ion appears important for the folding of the post-SET domain, and also tethers this domain to the SET domain. As a result of this tethering, the post-SET domain lies close to the active site. We observed that the loop connecting helices αE and αF from the post-SET domain is placed near the cofactor and contributes significantly to sinefungin interaction (Fig. 4), whereas the C-terminal end of helix αE is positioned to participate in the formation of the substrate binding cleft (Fig. 5B).

The insertion region between SET domain strands β5 and β8 is typically referred as SET-I domain (Fig. 1). SmyD1 has a unique, large insertion region (residues 50–180) that consists of two domains, MYND followed by a helix bundle (βB, 310–1, βC, and βD). The equivalent region in Set7/9 or Dim-5, however, contains only 1 or 2 small helices of 15–20 residues. As seen in Fig. 1, the helix bundle and MYND domain pack against the same β-sheet (β4, β10, and β11) but each on the opposite face. As the result of this structural arrangement, these two domains are held apart by the SET domain to confer different functions. We observe that the helix bundle, not MYND, contributes to cofactor and substrate binding. Specifically, the last two helices (αC, αD) of the bundle appear important for recognition of the H3 N-terminal residues (Fig. 5B), whereas the loop preceding the 310 helix makes exten-
sive contacts with the cofactor (Fig. 4A). Interestingly, this bulky SET-I domain or helix bundle results in a nearly closed cofactor binding pocket (Fig. 4B), and the possible significance of this feature will be discussed in more detail later.

**MYND Domain Mediates Specific Interaction with skNAC**—MYND, a conserved zinc finger motif, has been identified to mediate specific interaction with a proline-rich sequence (19). Our structure reveals that MYND consists of one kinked α helix (αA) and two antiparallel β-strands (β6 and β7) that are organized around 2 zinc ions (Fig. 3A). The two zinc centers are characterized by 7 cysteine residues and a single histidine arranged in a C4C2HC format. As shown in Fig. 1, MYND is part of the N-terminal grip of the wrench-shaped SmyD1 structure, forms direct contacts with the catalytic SET domain, but does not contribute residues to either substrate or cofactor binding. The latter observation is consistent with previous findings showing that the MYND deletion has little effect on the activity of SmyD2 or SmyD3 (7, 9).
Crystal Structure of SmyD1

These data thus indicate that MYND may primarily function as a protein-protein interaction module, and cooperate SmyD1 with skNAC to regulate cardiomyocyte growth and maturation (19).

SmyD1 interacts with skNAC through a proline-rich sequence containing a "PPLIP" motif (19), where the middle leucine plays a central role in the binding (19). To identify the putative skNAC binding site and gain insight into how SmyD1 MYND recognizes the PXLXP motif, we performed structural modeling based on the solution structure of the MYND domain of AML1/ETO in complex with a proline-rich peptide (17) (Fig. 3B). Although the two MYND domains share only 30% sequence identity, their structures are highly superimposed with a backbone root mean square deviation of 0.67 over 40 residues. The high structural similarity suggests that these two domains may use a similar mode of recognition in binding to a proline-rich sequence.

The modeling reveals that the proline-rich peptide is located in a shallow, fully exposed surface groove that appears readily accessible by skNAC (Fig. 3C). One side of the groove is formed by a loop connecting β6 and β7, and the other side by residues from the N-terminal half of helix αA (Fig. 3B).

The structural comparison suggests that 3 conserved residues in this groove are important for the SmyD1 and skNAC interaction. In particular, the side chain of Trp83 packs against the target lysine-access channels of SmyD1, Set7/9, and Dim-5 (Fig. 5, surface representation of lysine-access channels illustrates that the channel in SmyD1 is more spacious than Set7/9 or Dim-5. Sinefungin and AdoHcy are shown with C-NH3 amine colored in blue and sulfur atom in yellow, respectively). E, hydrophobic cluster and hydrogen bonding interaction between the CTD and SET domains in branch II. Residues are colored according to the domain in which they reside, and hydrogen bond is indicated as dashed line.

These two domains may use a similar mode of recognition in binding to a proline-rich sequence.

As described earlier, the CTD domain, which forms both the C-terminal grip and the handle of the wrench-shaped SmyD1 structure, is separated from the rest of the protein by a deep groove (Fig. 1). As a result, CTD does not make extensive contacts with the other domains in SmyD1, and the domain interfaces are mostly mediated by one edge of the CTD superhelix, which involves only 3 loops. The loop between αI...
charged α-amino group is recognized by a trigonal array of hydrogen bonds with the main chain carbonyl oxygens of Lys17 and Arg19 and the amide Oδ of Asn206. In the middle of sinefungin, the C-NH3 amine group, which is in place of the S-CH3 sulfonium of AdoMet, engages in a hydrogen bond with the backbone oxygen of Gly202. The similar interaction is expected in the case of AdoMet, which might contribute to enzymatic function by destabilizing the active methyl group. Collectively, the overall cofactor-binding mode of SmyD1 is structurally conserved with other SET enzymes and serves to orient the methyl group of AdoMet into the methyltransfer pore during catalysis.

However, several striking differences are observed between the cofactor binding pockets of SmyD1 and other SET proteins. The most variable is the interaction with the ribose moiety of sinefungin. In the structures of other SET proteins, the ribose hydroxyls of the cofactor are either solvent-exposed, such as in Set7/9, or hydrogen bonded with a single neighboring residue, as is observed in Dim-5. Due to the bulky SET-I domain, however, SmyD1 makes more extensive contacts with the ribose moiety than other SET proteins (Fig. 4A). The imidazole nitrogen of His135 from the SET-I domain forms a hydrogen bond with the ribose 3′-hydroxyl group, whereas the side chain of Gln133 extends over the cofactor binding pocket and makes van der Waals contacts with the most ribose moiety. Another significant variation is observed in the vicinity of the carboxylate moiety of sinefungin. In SmyD1, the carboxylate moiety of sinefungin forms a salt bridge interaction with the guanidinium group of Arg19 from the S-sequence. In contrast, Lys294 in Set7/9 (Arg238 in Dim-5), which plays an analogous role in the electrostatic interaction with the cofactor, resides in a region that corresponds to the SmyD1 310 helix. More significantly, in SmyD1, this arginine residue (Arg19) extends its side chain across the cofactor binding pocket to make unique interactions with the SET-I domain, including a salt bridge interaction with the carboxylate group of Asp131 and also a hydrogen bond to the backbone carboxyl oxygen of this aspartate residue.

Collectively, these structural differences result in a nearly closed cofactor-binding pocket in SmyD1, where the cofactor is more buried than in Set7/9 or Dim-5 (Fig. 4B). Given that Set7/9 shows much better activity than SmyD1 (Fig. 4C), we had hypothesized that this closed conformation might not be well suited for rapid exchange between the cofactor and its product during catalysis, which might in part account for the low activity of SmyD1. In contrast, however, the following mutation studies suggest that the interaction between SET-I and the cofactor has little effect on SmyD1 activity and that the split S-sequence is required for the methylation.

To elucidate the functional importance of the nearly closed cofactor binding site, we generated a series of mutations that are designed to weaken cofactor binding and help the cofactor-product exchange (Fig. 4C). We found that mutation of His135 or Gln133 to alanine, which disrupts the hydrogen bond to the cofactor ribose group or weakens the van der Waals interaction between SmyD1 and sinefungin, respectively, has only modest effects on the protein activity. These results sug-
Crystal Structure of SmyD1

gest that the interaction with SET-I is not essential for high affinity binding of AdoMet, and may not affect the rapid exchange of the cofactor and its product. In addition, to clarify the role of the S-sequence in SmyD1, we performed mutation studies on several highly conserved residues. These residues are: Lys17, which packs against the adenosine moiety of sine-fungin; Arg19, which makes the electrostatic interaction with the cofactor carboxylate group; and Gly18 and Gly20, which are located in the β1-β2 loop. We found that substitution of Lys17 or Arg19 to alanine completely abolishes the activity of the enzyme. Double mutation of G18A and G20A, which might change the S-sequence backbone structure and disrupt the hydrogen bond network in cofactor binding, also eliminates the H3 methylation by SmyD1, a result that is inconsistent with that observed in SmyD3 (9). These data thus indicate that the split S-sequence is an integral component of the SET domain and is essential for SmyD1 cofactor binding.

Substrate Binding Site with Unusual Architecture—Our co-crystallization of SmyD1 with histone H3 peptides has been unsuccessful. This is probably due to the relatively low binding affinity between SmyD1 and H3 as compared with other SET proteins such as Set7/9 (Fig. 6A). However, inspection of the SmyD1 surface clearly reveals a deep Y-shaped cleft with one branch (branch I) that connects through a narrow channel to the cofactor binding pocket, suggesting that branch I is the putative substrate binding site (Fig. 5A). To gain insight into SmyD1 substrate recognition and also understand why this enzyme binds very weakly to the substrate, we compared the SmyD1 structure with the H3-bound structure Set7/9 and Dim-5, which methylate Lys4 and Lys9 in H3, respectively. The structural comparison suggests that SmyD1 shares the broadly similar H3 binding mode to Set7/9 and Dim-5, but with dramatic differences in details.

As shown in Fig. 5B, the modeled H3 peptide, which is based on the Set7/9 ternary structure (21), binds at branch I of the SmyD1 Y-shaped cleft, possibly forming a parallel β-sheet with strand βB. Notably this hybrid β-sheet-binding mode is conserved among the known structures of SET proteins. In general, histone recognition by SET proteins can be divided into interactions with 3 regions: the residues N- and C-terminal to target lysine, and the target lysine itself (15). Interactions with H3 N-terminal residues have been shown to be critical for substrate recognition (22). As revealed by the structural comparison, Thr3 in the H3 N terminus is possibly recognized by SmyD1 Thr183 through a side chain-side chain hydrogen bonding interaction, similar to Set7/9 Ser268. However, these two proteins apparently differ in interacting with H3 Arg5. In the SmyD1-H3 model, the side chain of H3 Arg5, which points to a negatively charged surface region in the SET-I domain, appears able to engage in hydrogen bonding or salt-bridge interactions with a cluster of glutamate residues (Glu142, Glu145, and Glu146). In sharp contrast, this arginine salt-bridge interactions with a cluster of glutamate residues

SET proteins generally do not engage in extensive interactions with the C-terminal side of target lysine, which is consistent with the fact that the residues in this region do not significantly contribute to substrate recognition (22). However, due to the presence of the unique CTD domain, substantial differences in H3 C-terminal interactions may exist between SmyD1 and other SET proteins. The SmyD1-H3 model shows that the H3 C terminus sterically clashes with the CTD inner surface (Fig. 5B), suggesting that this domain might interfere with H3 binding or the H3 peptide may adopt a different conformation when binding to SmyD1. However, the CTD domain, on the other hand, appears to effectively extend the SmyD1 substrate binding site (branch I), with two possible extended binding clefts (branches II and III) that run toward the opposite directions (Fig. 5A). Branch II is formed by the CTD domain on one side and the post-SET domain on the other, and branch III by the CTD and SET domains. The existence of these two clefts implicates that CTD might be important for H3 binding and mediate interactions with the residues C-terminal to H3 Lys4.

To elucidate the function of the CTD domain, we generated a C-terminal truncated SmyD1 (ΔCTD) that ends with residue Asp294. Interestingly, our functional studies showed that the CTD truncation markedly increases both H3 binding and methylation by SmyD1 (Fig. 6), suggesting that this domain may interfere with SmyD1-histone interaction and plays a repressive role in the enzyme catalysis. To further confirm this idea, we performed mutation studies on two conserved residues in branch II: Tyr381 in the CTD domain and Asp254 from the loop preceding the post-SET domain. The aromatic ring of Tyr381 participates in a hydrophobic cluster with Leu256, Leu351, Leu384, and Tyr385, whereas the side chain of Asp254 makes a hydrogen bond to the side chain of Tyr385 in this cluster (Fig. 5E). These interactions appear to be important for holding the CTD and SET domains together and maintaining the conformation of branch II. Based on that, we speculate that breaking these interactions would reduce structural restraints on the CTD domain and might cause it to swing out to expose the substrate binding site. Interestingly, substitution of Asp254 or Tyr381 by alanine, which eliminates the hydrogen bonding interaction or might disrupt the hydrophobic cluster, respectively, significantly increases the H3 binding affinity as well as the enzymatic activity of SmyD1 (Fig. 6B). The results further indicate that CTD has a negative effect on SmyD1 activity, and that the presence of this domain may interfere with efficient histone binding.

In SET enzymes, the target lysine generally is recognized by hydrophobic and aromatic residues that form a narrow channel leading to the cofactor binding pocket of the proteins (Fig. 5C). The structural comparison reveals that the lysine access channel in SmyD1 is broadly similar to that in Set7/9 and Dim-5 with an almost circular channel opening (Fig. 5D). Most of the residues in the SmyD1 channel are well aligned with the corresponding residues in Set7/9 and Dim-5, serving to orient target lysine in their active site for methyl transfer. Specifically, Tyr252, which resides in the loop preceding the post-SET domain, is highly conserved across all classes of SET enzymes, with its side chain orientation remarkably similar to Tyr335 in Set7/9 and Tyr283 in Dim-5. Tyr270, which is located in the post-SET domain, is superimposed with Tyr317 in Set7/9 and Trp318 in Dim-5 despite that the aromatic rings of
these residues are orientated differently. In addition, the side chain of Phe to that residues in SmyD1 strand β8 occupies the same position as that of Leu in Set7/9 and Phe in Dim-5.

The structural comparison, however, also revealed some unusual features in the SmyD1 lysine-access channel. As shown in Fig. 5D, the lysine-access channel in SmyD1 is more spacious than Set7/9 and Dim-5, primarily because some bulky aromatic residues in Set7/9 or Dim-5 are replaced by small hydrophobic ones in SmyD1 (Fig. 5C). In Set7/9, Tyr, which is structurally conserved with Dim-5 Phe and hydrogens bonds to the side chain of the target lysine, is substituted by Val in SmyD1. In addition, the highly conserved Tyr in Set7/9 or Tyr in Dim-5 is replaced by Leu from the SmyD1 3-helix. This substitution is surprising because mutation of Tyr to either phenylalanine or alanine completely abolished both H3 binding and methylation by Set7/9 (21). Based on these observations, we hypothesize that the large lysine-access channel of SmyD1 is unable to effectively interact with or position target lysine during methyl transfer, affecting its activity. Interestingly, substitution of Val by tyrosine, a mutation that would presumably make the similar hydrogen bond to target lysine as Set7/9, results in a significant increase in H3 binding and also enhances SmyD1 methylation by about 3-fold (Fig. 6). These data thus indicate that the spacious lysine-access channel contributes to the weak substrate binding and the low enzymatic activity of SmyD1.

**DISCUSSION**

Recent studies have provided increasing evidence that SmyD family proteins function together with HSP90 in transcriptional regulation and epigenetic inheritance (23). For example, SmyD3 associates with HSP90 both in vitro and in vivo, and this association enhances the SmyD3 activity and is critical for Nkx2.8 regulation in cancer proliferation (8, 9). It was also suggested that HSP90 might control muscle cell differentiation in zebrafish by facilitating the SmyD1-mediated methylation (3, 24). Based on these findings, SmyD proteins can thus be characterized as HSP90-dependent histone methyltransferases, especially after recent studies showed that SmyD2 activity can also be increased by HSP90 binding (7); however, the mechanism of this activity enhancement remains unknown. Our data have established that the low enzymatic activity of SmyD1 is associated with the weak histone binding affinity (Fig. 6), implicating that HSP90, a molecular chaperone, might exert its activity-enhancing effect through altering SmyD1 conformation and increasing its ability to bind to the substrate. Interestingly, the CTD domain, which interferes with efficient SmyD1-histone interaction, may have the potential to undergo a hinge movement as suggested by our structure studies. Therefore, we propose a model that interaction with HSP90 might drive the CTD domain to swing out, fully exposing the substrate binding cleft situated in a deep groove with one side formed by the CTD domain, a process that might mimic the effects of the CTD truncation that we have shown significantly increased the H3 binding and methylation by SmyD1. Further research is required to support this proposed mechanism and to determine whether HSP90 can induce conformational changes in SmyD1 that involve the CTD domain.

The SmyD1 structure also provides insights into how MYND recognizes cardiac transcription factor skNAC, and reveals that the binding site in MYND is solvent-exposed and readily accessible. Although significance of the SmyD1-skNAC interaction in muscle development remains to be determined, similar kinetics of induction and localization of the two proteins during myogenesis in cultured cells clearly suggest an associated role for these proteins in this process (19). Given skNAC has been implicated in activating transcription in a sequence-specific manner (25), this protein may be responsible for targeting SmyD1 to specific loci, working in concert with SmyD1 in controlling gene transcription in myogenesis. Previous studies, however, showed that MYND alone is not sufficient for SmyD1 binding to skNAC, and both the intact MYND domain and the adjacent S-sequence are required, suggesting that these SmyD1 domains cooperate with one another to create a structural motif essential for the interaction (19). As revealed by our modeling studies, the S-sequence does not directly interact with the PXLXP peptide, but the N terminus of this peptide projects toward the S-sequence (Fig. 3C). Moreover, in the SmyD1 structure, MYND and the S-sequence form a continuously positively charged surface, and the boundary of these two domains with only a few interactions is characterized by grooves and cavities. Therefore, it can be predicted that the S-sequence with the positively charged surface might interact with the N-terminal residues beyond the PXLXP motif that are rich in acidic residues (19). Given the importance of the S-sequence in the methylation (Fig. 4C), it is unknown whether this presumed interaction can affect SmyD1 structure and function, which in turn may alter histone methylation profiles in heart and muscle development. Further structure studies between SmyD1 and skNAC should shed light on these issues.

**REFERENCES**

1. Kwon, C., Qian, L., Cheng, P., Nigam, V., Arnold, J., and Srivastava, D. (2009) *Nat. Cell Biol.* 11, 951–957
2. Gottlieb, P. D., Pierce, S. A., Sims, R. J., Yamagishi, H., Weihe, E. K., Harries, J. V., Maika, S. D., Kuziel, W. A., King, H. L., Olson, E. N., Nakagawa, O., and Srivastava, D. (2002) *Nat. Genet.* 31, 25–32
3. Tan, X., Rotllant, J., Li, H., De Deyne, P., and Du, S. J. (2006) *Proc. Natl. Acad. Sci. U.S.A.* 103, 2713–2718
4. Li, D., Niu, Z., Yu, W., Qian, Y., Wang, Q., Li, Q., Yi, L., Luo, J., Wu, X., Wang, Y., Schwartz, R. J., and Liu, M. (2009) *Nucleic Acids Res.* 37, 7059–7071
5. Borlak, J., and Thum, T. (2003) *FASEB J.* 17, 1592–1608
6. Costaolini, D. L., Arruda, E. P., Agarwal, P., Kim, K. H., Zhi, Y., Zhu, W., Lebel, M., Cheng, C. W., Park, C. Y., Pierce, S. A., Guerciof, A., Pollevick, G. D., Chan, T. Y., Kabir, M. G., Cheng, S. H., Husain, M., Antzelevich, C., Srivastava, D., Gross, G. J., Hui, C. C., Backx, P. H., and Bruneau, B. G. (2005) *Cell* 123, 347–358
7. Abu-Farha, M., Lambert, J. P., Al-Madhoun, A. S., Elisma, F., Skerjanc, I. S., and Figeys, D. (2008) *Mol. Cell Proteomics* 7, 560–572
8. Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., Yagyu, R., and Nakamura, Y. (2004) *Nat. Cell Biol.* 6, 731–740
9. Silva, F. P., Hamamoto, R., Kunizaki, M., Tsuge, M., Nakamura, Y., and Furukawa, Y. (2008) *Oncogene* 27, 2686–2692
10. Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., Kubicek, S., Opravil, S., Jenuwein, T., and Berger, S. L. (2006) *Nature* 444, 629–632
Crystal Structure of SmyD1

11. Kunizaki, M., Hamamoto, R., Silva, F. P., Yamaguchi, K., Nagayasu, T., Shibuya, M., Nakamura, Y., and Furukawa, Y. (2007) Cancer Res. 67, 10759–10765
12. Sawada, K., Yang, Z., Horton, J. R., Collins, R. E., Zhang, X., and Cheng, X. (2004) J. Biol. Chem. 279, 43296–43306
13. Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., and Cheng, X. (2003) Mol. Cell 12, 177–185
14. Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J., and Xiao, B. (2002) Cell 111, 105–115
15. Couture, J. F., Collazo, E., Brunzelle, J. S., and Trievel, R. C. (2005) Genes Dev. 19, 1455–1465
16. Xiao, B., Jing, C., Kelly, G., Walker, P. A., Muskett, F. W., Frenkel, T. A., Martin, S. R., Sarma, K., Reinberg, D., Gamblin, S. J., and Wilson, J. R. (2005) Genes Dev. 19, 1444–1454
17. Liu, Y., Chen, W., Gaudet, J., Cheney, M. D., Roudaia, L., Cierpicki, T., Klet, R. C., Hartman, K., Laue, T. M., Speck, N. A., and Bushweller, J. H. (2007) Cancer Cell 11, 483–497
18. Ansieau, S., and Leutz, A. (2002) J. Biol. Chem. 277, 4906–4910
19. Sims, R. J., 3rd, Weihe, E. K., Zhu, L., O’Malley, S., Harriss, J. V., and Gottlieb, P. D. (2002) J. Biol. Chem. 277, 26524–26529
20. Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., and Cheng, X. (2002) Cell 111, 117–127
21. Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., and Gamblin, S. J. (2003) Nature 421, 652–656
22. Couture, J. F., Collazo, E., Hauk, G., and Trievel, R. C. (2006) Nat. Struct. Mol. Biol. 13, 140–146
23. Ruden, D. M., Xiao, L., Garfinkel, M. D., and Lu, X. (2005) Hum. Mol. Genet. 14, R149–155
24. Du, S. J., Li, H., Bian, Y., and Zhong, Y. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 554–559
25. Yotov, W. V., and St-Arnaud, R. (1996) Genes Dev. 10, 1763–1772
26. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
27. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) Acta Crystallogr. D Biol. Crystallogr. 58, 1948–1954
28. Collaborative Computational Project Number 4 (1994) Acta Crystallogr. D Biol. Crystallogr. 50, 670–673
29. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
30. Brüger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. D Biol. Crystallogr. 54, 905–921
31. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Bioinformatics 23, 2947–2948