Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins

Shutong Xu1,2,†, Chen Zhong1,†, Tianlong Zhang1, and Jianping Ding1,*

1 State Key Laboratory of Molecular Biology and Research Center for Structural Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
2 Graduate School of Chinese Academy of Sciences, 320 YueYang Road, Shanghai 200031, China
* Correspondence to: Jianping Ding, E-mail: jpding@sibs.ac.cn

The SET- and myeloid-Nervy-DEAF-1 (MYND)-domain containing (Smyd) lysine methyltransferases 1–3 share relatively high sequence similarity but exhibit divergence in the substrate specificity. Here we report the crystal structure of the full-length human Smyd2 in complex with S-adenosyl-L-homocysteine (AdoHcy). Although the Smyd1–3 enzymes are similar in the overall structure, detailed comparisons demonstrate that they differ substantially in the potential substrate-binding site. The binding site of Smyd3 consists mainly of a deep and narrow pocket, while those of Smyd1 and Smyd2 consist of a comparable pocket and a long groove. In addition, Smyd2, which has lysine methyltransferase activity on histone H3-lysine 36, exhibits substantial differences in the wall of the substrate-binding pocket compared with those of Smyd1 and Smyd3, which have activity specifically on histone H3-lysine 4. The differences in the substrate-binding site might account for the observed divergence in the specificity and methylation state of the substrates. Further modeling study of Smyd2 in complex with a p53 peptide indicates that mono-methylation of p53-Lys372 might result in steric conflict of the methyl group with the surrounding residues of Smyd2, providing a structural explanation for the inhibitory effect of the SET7/9-mediated mono-methylation of p53-Lys372 on the Smyd2-mediated methylation of p53-Lys372.

Keywords: Smyd2, Smyd proteins, SET family, histone lysine methyltransferase, epigenetics

Introduction

Histone methylation, one of the major forms of histone modification, has been shown with extensive studies to play important roles in the modulation of chromatin structure, transcriptional regulation, and hence development and disease pathogenesis (Martin and Zhang, 2005; Bhauik et al., 2007). Although histone methylation involves both lysine and arginine residues, it occurs mainly at lysine residues at the N-terminal region of histones H3 and H4 (H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20) (Cheng et al., 2005). The enzymes responsible for methylation of all of these sites except H3K79 belong to the SET family of protein lysine methyltransferases (PKMTs).

Within the SET family, five members share a unique feature with the SET domain being split into two segments by the insertion of a myeloid-Nervy-DEAF-1 (MYND) domain, which constitute a SET- and MYND-domain containing (Smyd) protein subfamily (Gottlieb et al., 2002; Hamamoto et al., 2004; Brown et al., 2006). Smyd1, Smyd2, and Smyd3 have been proven to possess methyltransferase activities but have different expression profiles and histone substrate specificities, even though they share a relatively higher sequence similarity among themselves (~30%) than with the other identified SET family methyltransferases (~20%). Smyd1 (also known as Bop), which was the first to be studied, has been found to be expressed in the skeletal and cardiac muscle tissues and cytotoxic T cells and is able to methylate specifically H3K4; Smyd3 is universally expressed and also has a specific methyltransferase activity toward H3K4; Smyd2 is expressed in a wide variety of tissues with the highest level in the heart and brain (Hwang and Gottlieb, 1997; Gottlieb et al., 2002; Brown et al., 2006; Hamamoto et al., 2006; Tan et al., 2006). In contrast to Smyd1 and Smyd3, Smyd2 is unique for its ability to di-methylate H3K36; it might also be able to methylate H3K4 in the presence of Hsp90α (Brown et al., 2006; Abu-Farha et al., 2008). The knowledge about the other two Smyd proteins, namely Smyd4 and Smyd5, is very scarce, and so far there is no evidence showing that Smyd4 or Smyd5 harbors PKMT activity.

In Saccharomyces cerevisiae, di-methylation of H3K36 is present at high level in open reading frames and 3’ untranslated regions and has been shown to correlate with the ‘on’ state of gene transcription (Bannister et al., 2005; Rao et al., 2005).
However, it is also shown that di- and tri-methylation of H3K36 can recruit Rpd3S to deacetylate coding regions, which suppresses spurious intragenic transcription (Carrozza et al., 2005; Keogh et al., 2005; Li et al., 2009). Recently, the function of Smyd2 in the regulation of gene transcription has been examined. Overexpression of Smyd2 in 293T cells results in predominantly up-regulation of genes (Abu-Farha et al., 2008). In addition, Smyd2 is able to interact with RNA polymerase II and helicase HELZ, and in Smyd2 conditional knockout heart, the majority of the differentially expressed transcripts are down-regulated; intriguingly, most of these transcripts encode proteins associated with the translation process such as subunits of cytosolic and mitochondrial ribosomes, suggesting that Smyd2 is involved in the regulation of protein translation (Diehl et al., 2010).

Additionally, it has also been demonstrated that Smyd2 exerts PKMT activity towards non-histone proteins including the tumor suppressor p53 and retinoblastoma (Rb) (Huang et al., 2006; Saddic et al., 2010). Smyd2 can mono-methylate p53 at Lys570, which probably decreases the DNA-binding efficiency of p53 (Huang et al., 2006). Recently, Smyd2 has been found to be able to methylate Rb both in vitro and in vivo at Lys866, a highly conserved lysine residue, which provides a binding site for the methyl-binding domain of transcriptional repressor L3MBTL1; and the methylation level of Rb Lys860 is regulated in various biological processes and in response to DNA damage, suggesting that this newly identified modification of Rb might be important for its functions (Saddic et al., 2010).

Given the ability of Smyd2 to modulate gene transcription and the functional activities of p53 and Rb, a link between Smyd2 and cancers can be expected. Recently, overexpression of Smyd2 is found in esophageal squamous cell carcinoma and has an independent reverse correlation with the survival rate of the patients, suggesting that Smyd2 might be a prognostic marker and a therapeutic target of this cancer (Komatsu et al., 2009). Therefore, the structural study of Smyd2 may provide valuable information not only for the molecular basis underlying the substrate binding and specificity and the catalytic mechanism, but also for the design of inhibitors against this potential therapeutic target. Here we report the crystal structure of the full-length human Smyd2 in complex with the cofactor product S-adenosyl-L-homocysteine (AdoHcy) at 2.8 Å resolution. Structural analysis reveals the differences between Smyd2 and the other Smyd proteins and provides insights into their substrate divergence.

### Results and discussion

#### Overall structure

The crystal structure of the full-length human Smyd2 in complex with the cofactor product AdoHcy was solved at 3.0 Å resolution with the single-wavelength anomalous dispersion (SAD) method and refined to 2.8 Å resolution, yielding an R-factor of 24.6 and a free R-factor of 29.9 (Table 1 and Figure 1A). There are two molecules in an asymmetric unit which assume almost identical overall structure with a root-mean-square deviation (RMSD) of 0.4 Å for all Ca atoms, and molecule A, which has more detectable residues and better electron density was used in structural analysis and discussion. As shown in Figure 1A, the N-terminal region of Smyd2 comprises the SET domain (residues 1–46 and 97–243), the MYND domain (residues 47–96), and the post-SET domain (residues 244–271). This region consists of a mixed structure of α-helices (α1–α6), β-strands (β1–β12), and long extended loops. The C-terminal region contains mainly a tetra-trico-peptide repeat (TPR) domain (residues 272–433) consisting of seven α-helices (α8–α14).

The overall structure of Smyd2 is similar to those of human Smyd3 (Srirupong et al., 2011; Xu et al., 2011) and murine Smyd1 (Srirupong et al., 2010) with RMSDs of 3.4 and 3.8 Å for all Ca atoms, respectively. However, there are notable conformational differences between the N- and C-terminal regions among these enzymes. The N- and C-terminal regions of the three enzymes can be superimposed well separately: superposition of the N-terminal regions yields an RMSD of 1.8 Å between Smyd2 and Smyd3 and 2.3 Å between Smydi and Smyd3, and superposition of the C-terminal regions yields an RMSD of 2.4 Å between Smyd2 and Smyd3 and 2.0 Å between Smyd1 and Smyd3 (Figure 1B). Relative to the C-terminal region of Smyd3, the equivalent regions of Smyd2 and Smyd1 rotate about 16° and 23° away from the N-terminal region (Figure 1B). Accompanying the domain rotation, significant differences are observed at the interface between the N- and C-terminal regions, which may account in part for a unique feature of the potential substrate-binding site of Smyd3 (see discussion later). Detailed comparisons indicate that several factors may contribute to these conformational differences (Supplementary Figures S2 and S3). The α7–α8 loops of Smyd2 and Smyd1 which link the

### Table 1 Crystallographic data and refinement statistics.

<table>
<thead>
<tr>
<th>Diffraction data</th>
<th>Zn-peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.2825</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>a, b, c (Å)</td>
</tr>
<tr>
<td></td>
<td>α, β, γ (°)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0–2.8 (2.9–2.8)*</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>191382</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>25466</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>7.5 (6.9)</td>
</tr>
<tr>
<td>Average l/σ(l)</td>
<td>14.8 (2.8)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.4 (97.8)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>14.3 (66.9)</td>
</tr>
</tbody>
</table>

*Numbers in parentheses represent the highest resolution shell; bRmerge = ∑hkl|Ihkl|−<Ihkl>|/|∑hkl|Ihkl|; cRmerge = ∑hkl|Fobs|−|Fcalc|/∑hkl|Fobs|.
N-terminal region and the C-terminal TPR domain contain 11 and 10 residues, respectively, while the equivalent α8–α9 loop of Smyd3 contains only four residues (Supplementary Figure S2). The variation in the length of the connecting loop among the three Smyd proteins causes slight conformational differences of the linker between the N- and C-terminal regions which could be propagated to the following structure elements of the TPR domain, resulting in the positional differences of the TPR domain. In addition, the α14–α15 loop of the TPR domain in Smyd3 is found to make intense interactions with the N-terminal region compared with its equivalent in Smyd1 and Smyd2, which could be due to the substitution of a Pro residue (Pro^{432} of Smyd1 and Pro^{421} of Smyd2) on this loop with a Ser residue (Ser^{409}). In Smyd1 and Smyd2, the Pro residue might

Figure 1 Structure of the Smyd2–AdoHcy complex. (A) Overall structure of the Smyd2–AdoHcy complex. Top: a schematic representation of the full-length Smyd2. The N-terminal SET domain (residues 1–46 and 97–243), the MYND domain (residues 47–96), the post-SET domain (residues 244–271), and the C-terminal region (residues 272–433) are colored in green, yellow, magenta, and blue, respectively. Bottom: two views of the overall structure of the Smyd2–AdoHcy complex. The domains are colored accordingly and the secondary structure elements are labeled. The cofactor product AdoHcy is shown with a ball-and-stick model and colored in cyan. Two zinc ions bound in the MYND domain and one in the post-SET domain are shown with spheres and colored in gray. (B) Comparison of the overall structures of Smyd1 (blue), Smyd2 (green), and Smyd3 (salmon). The structures are superimposed based on the N-terminal region and the respective rotation angles of the C-terminal regions of Smyd1 and Smyd2 relative to that of Smyd3 are marked. (C) A representative simulated-annealing omit F_{o}–F_{c} electron density map (contoured at 2.0σ level) for the cofactor product AdoHcy. (D) A detailed comparison of the cofactor binding mode of Smyd2 (left panel) with that of Smyd3 (right panel, PDB code 3OXF). In Smyd2, the cofactor is bound at the cofactor-binding pocket, interacting with surrounding residues and an SO_{4}^{2−} ion. The cofactor, the SO_{4}^{2−} ion, and the surrounding residues are shown with ball-and-stick models. The hydrogen bonds are indicated with black dashed lines. For the residues interacting with the cofactor and the SO_{4}^{2−} ion, the contributing moieties are shown. Depending on whether the side chains or the main chains are involved, these residues are labeled in red and black, respectively.
restrain the conformation of the loop to point away from the N-terminal region, while in Smyd3, with a more flexible Ser, the loop can be easily oriented toward the N-terminal region, allowing for the flanking structural elements to make interactions with the N-terminal region. In Smyd3, the main chain and side chain of Ser{eq}^{409}\) each forms a hydrogen bond with Asp{eq}^{217}\), and the neighboring Glu{eq}^{407}\) forms two salt bridges with Arg{eq}^{274}\). Nevertheless, since both the N- and C-terminal regions of all three Smyd proteins are involved in crystal packing, it cannot be completely excluded that the crystal packing might also have an impact on the conformation of the overall structure.

Structural comparisons of the SET domain of Smyd2 with those of the other SET methyltransferases with known structures, including the H3K4 specific enzymes SET7/9 (Jacobs et al., 2002; Wilson et al., 2002; Kwon et al., 2003; Xiao et al., 2003; Couture et al., 2006) and MLL1 (Southall et al., 2009), the Smyd family members Smyd1 (Sirinupong et al., 2010) and Smyd3 (Sirinupong et al., 2011; Xu et al., 2011), and others such as Rubisco large subunit methyltransferase (LSMT) (Trievel et al., 2002), indicate that the SET domain of Smyd2 is most similar to those of Smyd3 (Xu et al., 2011) and Smyd1 (Sirinupong et al., 2010) (RMSDs of 1.6 and 1.9 Å, respectively, for 184 Cx atoms) than to the others (an RMSD of 2.5–2.7 Å for about 100 Cx atoms). The post-SET domain of Smyd2 is also cysteine-rich and is bound with a Zn{eq}^{2+}\) ion in a mode similar to that of Smyd3.

The TPR domain of Smyd2 works together with the SET and post-SET domains to form the substrate-binding site (Figure 1A, see discussion later), which is similar to that of Smyd3. In analogy, it would also be required for the substrate binding and the enzymatic activity of Smyd2 and might mediate the interactions of Smyd2 with Hsp90\(x\). The MYND insert forms a relatively independent structure, consisting mainly of a long extended loop, a short β-sheets (β5 and β6), and a distorted α-helix (α1) (Supplementary Figure S1A), and two Zn{eq}^{2+}\) ions are bound by the Cys zinc-chelating motif. The structure of the MYND domain in Smyd2 is similar to that in Smyd3. The electrostatic potential surfaces of the MYND domains of the Smyd1–3 enzymes were compared (Supplementary Figure S1B). In general, the MYND domain of Smyd2 exhibits a positively charged surface, and so does that of Smyd1. It has been shown that in Smyd3, Arg{eq}^{66}\) which is located in the center of a positively charged surface patch is important for the DNA binding of the MYND domain (Xu et al., 2011). In Smyd1 and Smyd2, the equivalent residue (Lys{eq}^{66}\) is also located within a positively charged patch at a position similar to that of Arg{eq}^{66}\). These results imply that the MYND domains of Smyd1 and Smyd2 might also be involved in the regulation of their histone lysine methyltransferase (HKMT) activities via a mechanism similar to that of Smyd3 (Xu et al., 2011).

Cofactor-binding site

In the Smyd2–AdoHcy complex, AdoHcy is bound at a cleft formed by the β1–β2 and nα–α5 loops and the β9 strand of the SET domain and the α5–α6 loop of the post-SET domain with well-defined electron density (Figure 1A and C). It takes a U-shape configuration and forms intense interactions with the surrounding residues (Figure 1D). Specifically, the aromatic side chain of Phe{eq}^{260}\) makes a π–π stacking interaction with the adenine ring of AdoHcy, and the main-chain carbonyl of His{eq}^{207}\) forms a hydrogen bond with the N6 atom of the ring. Additionally, an SO{eq}^{2-}\) ion is bound near AdoHcy to form a hydrogen bond with the N1 atom of the adenine ring and two extra hydrogen bonds with the side chains of Asp{eq}^{263}\) and Lys{eq}^{17}\). The ribose moiety is recognized by the side-chain N4 of His{eq}^{337}\) via hydrogen bonds with both the 2’- and 3’-hydroxyls, and by the main-chain carbonyl of Tyr{eq}^{298}\) via a hydrogen bond with the 3’-hydroxyl. The amide group of the homocysteine moiety is hydrogen-bonded to the main-chain carbonyls of Lys{eq}^{17}\) and Arg{eq}^{19}\) and the side chain of Asn{eq}^{206}\), and the carboxyl group forms hydrogen bonds with the main-chain amide and the side-chain amine of Arg{eq}^{19}\).

Although AdoHcy in the Smyd2 complex binds at an equivalent position and assumes a similar conformation as AdoHcy (PDB code 3O XF) (Xu et al., 2011) and S-adenosyl-L-methionine (AdoMet) (PDB code 3MEK) in the Smyd3 complexes, notable differences are observed in the cofactor binding mode between Smyd2 and Smyd3 (Figure 1D). In Smyd3, the phenol hydroxyl of Tyr{eq}^{126}\) forms a hydrogen bond with the carboxyl group of the homocysteine moiety of AdoHcy; while in Smyd2, the side chain of Arg{eq}^{19}\) occupies the position of the side chain of Tyr{eq}^{124}\) and correspondingly makes a similar interaction. In Smyd3, the side chain of Asn{eq}^{132}\) forms two hydrogen bonds with the 2’- and 3’-hydroxyls of the ribose moiety; while in Smyd2, the side chain of His{eq}^{337}\) makes the corresponding interactions. Intriguingly, in the Smyd3–AdoMet complex (PDB code 3MEK), a water molecule is found at a position similar to that of the SO{eq}^{2-}\) ion in the Smyd2–AdoHcy complex to interact with both Asp{eq}^{263}\) and the adenine moiety of AdoMet. It is possible that in the absence of high concentration of SO{eq}^{2-}\), the SO{eq}^{2-}\) ion observed in the Smyd2-AdoHcy structure might be replaced with a water molecule which would mediate the indirect interactions between Asp{eq}^{263}\) and the cofactor. These results suggest that the post-SET domain of Smyd2, which contains Phe{eq}^{260}\) and Asp{eq}^{263}\), might also contribute to the cofactor binding as its equivalent of Smyd3.

Substrate-binding site and lysine channel

It is common that methyltransferases and the associated proteins may recognize and bind various substrates (Cheng et al., 2005; Suganuma and Workman, 2010). The Smyd enzymes not only show divergent substrate specificities, but also exhibit differences in the methylation state of the substrate. Both Smyd1 and Smyd3 harbor HKMT activities on H3K4 (Hamamoto et al., 2004; Tan et al., 2006). For Smyd1, the methylation state of the substrate is unclear as the antibody used to detect methylated H3K4 could not distinguish mono-, di-, and tri-methylated lysine (Tan et al., 2006). Smyd3 has been demonstrated to specifically di- or tri-methylate H3K4 (Hamamoto et al., 2004), indicating that Smyd3 can use mono- and di-methylated H3K4 as substrates. In agreement with the experimental results, our structural data show that the lysine channel in Smyd3 is sufficient to accommodate a tri-methylated H3K4 (Xu et al., 2011). In contrast to Smyd3, Smyd2 catalyzes mono-, but not di- or tri-methylation of pS3 (Huang et al., 2006) and Rb (Saddic et al., 2010). As for the histone substrate, it has been shown that Smyd2 can di-methylate H3K36; however, it was not examined whether it can tri-methylate H3K36 (Brown et al., 2006). It has been shown that in the presence of Hsp90\(x\), Smyd2 has an HKMT
activity on H3K4 (Abu-Farha et al., 2008). However, this observation needs further validation and the methylation state of the modified H3K4 needs to be examined.

To investigate the structural basis of the observed divergence in the specificity and methylation state of the substrates among the Smyd enzymes, comparisons of the potential substrate-binding site and the lysine channel of Smyd2 with those of Smyd1, Smyd3, and other SET PKMTs were performed. Superposition of the Smyd2 structure with other PKMT structures, particularly the structure of pea Rubisco LSMT in complex with a free ε-N-methyl-lysine (MeLys) (Trievel et al., 2003), reveals the lysine channel of Smyd2. As the cofactor-binding site and the substrate-binding site should be at the opposite ends of the lysine channel, the potential substrate-binding site could also be identified. In all of the three Smyd proteins, there is a narrow and deep pocket on the surface formed by structural elements of the SET, post-SET, and TPR domains that extends to the lysine channel and thus could be the potential substrate-binding site (Figure 2A). The size of the pocket varies in Smyd1–3, and additionally, there is a striking difference in the structure surrounding the pocket among these enzymes. In Smyd2, the potential substrate binding pocket is located in the center of a long groove which could provide additional binding space for the substrate. A similar surface groove also exists in Smyd1 (Sirinupong et al., 2010). In contrast, the pocket in Smyd3 is merely a concave depression (Xu et al., 2011).

Detailed analyses of the Smyd2 and Smyd3 structures indicate that the observed divergence in the substrate-binding site could contribute to the conformational difference of the N- and C-terminal regions between the two enzymes, in particular, the conformational difference between the α8–α9 loop of the C-terminal TPR domain of Smyd2 and the equivalent α9–α10 loop of Smyd3 (the α9–α10 loop of Smyd3 is displaced by about 10 Å relative to the α8–α9 loop of Smyd2) (Figure 2B). On the α9–α10 loop of Smyd3, Trp300 forms a hydrogen bond with Ser44, and Trp300 makes hydrophobic interactions with Trp302 of helix α10 and Val67 of the N-terminus. Together, Ala298, His299, Trp300, and Lys251 of the α9–α10 loop and several flanking residues fill the space corresponding to the vacancy between the α8–α9 loop and the N-terminal region in Smyd2, hence preventing the formation of a similar groove on the surface.

In Smyd3, several acidic residues are located at the opening of the substrate-binding pocket, and single or combined mutations of two of these residues, namely Asp244 of the SET domain and Asp312 of the TPR domain, significantly impair the HKMT activity (Xu et al., 2011). The equivalent residues are Asp354 and Glu359 in Smyd1, and Asp362 and Tyr346 in Smyd2, respectively, implying that the acidic property of the opening of the binding pocket, especially at the position of the conserved Asp might be important for the substrate binding of these Smyd enzymes. Further examination of the wall of the substrate-binding pocket reveals more differences among Smyd1–3. As shown in Figure 2C, Glu187, His193, Tyr197, Arg253, and Ser257 of Smyd2 which are located at 4–7 Å away from the lysine channel are replaced with Cys186, Glu192, Phe216, Gln252, and Gln256 in Smyd3, and Ser185, Gln191, Phe216, Gln265, and Gln269 in Smyd1, respectively. Apparently, Smyd1 and Smyd3 show more similarity in this region compared with Smyd2.

In contrast to the differences in the potential substrate-binding site, the lysine channel which connects the substrate-binding site to the cofactor-binding site shows little variance among the Smyd enzymes. The lysine channel of Smyd2 is formed mainly by several conserved hydrophobic and aromatic residues. In particular, the base of the channel where the methyl transfer reaction takes place is surrounded by Phe184, Asn182, and Tyr240 which are all conserved in Smyd3 and Smyd1, and additionally Met205 which is substituted with Leu204 in Smyd3 or Val204 in Smyd1 (Figure 2D). Since the lysine channel of Smyd2 has similar size and electrostatic property as those of Smyd3 and Smyd1, the divergence of the substrate specificity in these Smyd enzymes is unlikely due to the minor difference in the lysine channel. Taken together, the observed differences in the substrate specificity and in the methylation state of the substrates among the Smyd proteins are most likely due to their divergence in the substrate-binding site, particularly in the wall of the substrate-binding pocket. The residues forming the wall of the substrate-binding pocket are involved in the recognition and binding of the substrate, and the specific interactions between the enzyme and the substrate could determine the position and conformation of the substrate. The divergence in the substrate-binding site might lead to subtle differences in the position and conformation of the ε-amine of the substrate lysine in the lysine channel, which set limits for the methylation state of the substrate and thus determine the substrate specificities.

Implication of the binding of Smyd2 to p53

It has been reported that Smyd2 mono-methylates Lys370 of p53, which suppresses the p53-mediated transcriptional regulation probably through its function in the dissociation of p53 from DNA (Huang et al., 2006). Intriguingly, mono-methylation of an adjacent residue, namely Lys372 can inhibit the methylation of Lys370 by Smyd2, which is caused partially by blocking the binding of Smyd2 to p53 (Huang et al., 2006). To understand the molecular basis of these observations, a p53 peptide encompassing residues 369–374 (NH2-LKSK(Me)KG-COOH) was manually docked into the substrate-binding pocket of Smyd2 based on superposition of the structure of SET7/9 in complex with the p53 peptide (PDB code 1XQH) (Chuikov et al., 2004) onto the structure of the Smyd2–AdoHcy complex, and further processed with subtle positional adjustment. In this complex model, the p53 peptide can be accommodated in the substrate-binding pocket of Smyd2 with the side chain of Lys370 inserting into the lysine channel and that of the methylated Lys372 into an adjacent channel (Figure 3A). Superposition of this structural model with the structure of the Rubisco LSMT in complex with MeLys shows that the ε-amine of Lys370 of the p53 peptide is located at a position similar to that of MeLys in the Rubisco LSMT structure (Figure 3B). The methyl group of the methylated Lys372 is confined in a narrow space and has steric hindrance with both the main chain and side chain of Asn380 and potentially with the side chain of Leu191 (Figure 3C). This provides an explanation for the abolishment of the Smyd2 binding to p53 by SET7/9-mediated methylation at Lys372 and hence for the inhibitory effect of mono-methylation of Lys372 on Smyd2-mediated methylation at Lys370.
Figure 2 Structures of the substrate-binding pocket and the lysine channel. (A) Electrostatic potential surface of the substrate-binding pocket in Smyd2 (left panel), Smyd3 (middle panel, PDB code 3OXF), and Smyd1 (right panel, PDB code 3N7I). Superposition of each of the Smyd structures with the pea Rubisco LSMT–MeLys complex (PDB code 1POY) reveals the position of the lysine channel, and MeLys in the superposed structure of Rubisco LSMT is shown for positional reference. In Smyd1 and Smyd2, the potential substrate-binding site consists of a deep pocket and a long groove across the surface, while in Smyd3, it is only comprised of a deep and narrow pocket. The surface charge distribution is displayed as blue for positive, red for negative, and white for neutral. The potential substrate-binding sites are indicated with yellow dotted curves and the substrate-binding pockets with green solid circles. MeLys is shown with a ball-and-stick model and colored in yellow. (B) Comparison of the α8–α9 loop of Smyd2 (green, left panel) and the α9–α10 loop of Smyd3 (salmon, right panel) and the surrounding structural elements, which partially accounts for the differences in the substrate-binding site in the two enzymes. The residues on the α9–α10 loop of Smyd3 including Ala298, His299, Trp300, and Lys301 and the residues of the N-terminal region in the vicinity including Lys42, Gly43, Ser44, Arg45, Gly46, and Val47 are shown with ball-and-stick models. The equivalent residues of Smyd2 are also shown. (C) A close-up view of the wall of the substrate-binding pocket. The wall of the substrate-binding pocket of Smyd2 (green) which is located between the opening of the substrate-binding pocket and the lysine channel is compared with those of Smyd3 (salmon, left panel) and Smyd1 (blue, right panel). The differed residues are shown with ball-and-stick models. (D) A close-up view of the lysine channel. The lysine channel of Smyd2 (green) is compared with those of Rubisco LSMT (yellow, left panel) and Smyd3 (salmon, right panel). The residues forming the lysine channels are shown with ball-and-stick models and colored accordingly.
Structure of human lysine methyltransferase Smyd2

Figure 3 Modeling study of the binding of Smyd2 to p53. (A) A surface presentation of the structure model of the Smyd2–p53 peptide complex. Modeling study of the binding of Smyd2 to p53 was performed by docking the p53 peptide derived from the superimposed structure of SET7/9 in complex with a p53 peptide containing mono-methylated Lys372 (PDB code 1XQH) into the substrate-binding pocket of the Smyd2–AdoHcy structure. AdoHcy is shown with a ball-and-stick model and colored in cyan. The p53 peptide is shown with a ribbon presentation and colored in magenta. The side chains of Lys370 and methyl-Lys372 are shown with ball-and-stick models and are indicated. The surface charge distribution of Smyd2 is displayed in transparency to better view the location of the docked p53 peptide. (B) A stereo view of the lysine channels in the Smyd2–p53 complex and the Rubisco LSMT–MeLys complex (PDB code 1P0Y). In the structure model of the Smyd2–p53 complex, the side chain of Lys370 of the p53 peptide (magenta) and the cofactor (cyan) are shown with ball-and-stick models. For simplicity, only MeLys in the LSMT–MeLys complex is shown with a ball-and-stick model in yellow. (C) A stereo view of the potential steric conflict between methyl-Lys372 of the p53 peptide and residues Asn380 and Leu191 in the structure model of the Smyd2–p53 complex. The color coding is the same as in B.

Materials and methods

Protein expression and purification

The full-length human Smyd2 gene was amplified with PCR from HEK293 cDNA and was subcloned into the pET-28b vector (Novagen) with a His6 tag at the C-terminus. The constructed plasmid was transformed into Escherichia coli BL21(DE3) Codon Plus strain (Novagen). The bacterial cells were grown in LB medium at 37°C to A600 of 0.6, and then induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside at 16°C for 24 h. The cells were collected by centrifugation at 6000 g, and lysed in a lysis buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 1 mM phenylmethylsulfonyl fluoride) on ice by sonication. The cell lysate was precipitated by centrifugation at 18000 g for 30 min, and the supernatant was used for protein purification.

The human Smyd3 protein was purified by affinity chromatography using an Ni2+-NTA column (Qiagen) equilibrated with a binding buffer (20 mM Tris–HCl, pH 8.0, 300 mM NaCl, and 5 mM β-mercaptoethanol). The column was washed with the binding buffer supplemented with 25 mM imidazole, and then the target protein was eluted with the binding buffer supplemented with 200 mM imidazole. The protein sample was further purified by gel filtration using Superdex 200 16/60 column (Amersham Biosciences) and then was stored in the storage buffer containing 20 mM Tris–HCl, pH 8.0, 100 mM NaCl, and 1 mM DTT.

Crystallization, diffraction data collection, and structure determination

The purified Smyd2 protein was concentrated to 5–10 mg/ml in the storage buffer and then incubated with 600 μM AdoHcy for 1 h. Crystallization was performed using the hanging drop vapor diffusion method. Crystals belonging to space group C2 were obtained at 4°C with an equal volume of the protein solution and the reservoir solution containing 0.1 M Bis–Tris, pH 6.5, 0.2 M Li2SO4, and 25% polyethylene glycol 3350. The diffraction data were collected from a flash-cooled crystal at 100 K at the anomalous peak wavelength of zinc (1.2825 Å) at beamline 17 U of Shanghai Synchrotron Radiation Facility, China. The SAD diffraction data were collected to 2.8 Å resolution and were processed, integrated, and scaled together using HKL2000 (Otwinowski and Minor, 1997). The statistics of the diffraction data are summarized in Table 1.

The structure of Smyd2 was solved using the SAD method implemented in program AutoSol of the Phenix suite (Adams et al., 2010). The zinc SAD data set was cut off at 3.0 Å resolution, and six zinc atoms were unambiguously identified from the SAD phased map, indicating that there are two Smyd2 molecules in an asymmetric unit. AutoSol automatically built a partial model of about 500 polyalanines out of 866 residues and a complete model was built based on improved electron density maps after iterative cycles of alternating manual building with COOT (Emsley and Cowtan, 2004) and refinement of the model with Phenix (Adams et al., 2010). The final structure refinement was accomplished with REFMAC5 (Murshudov et al., 1997). There is an AdoHcy molecule bound at a cleft formed by the SET domain and the post-SET domain with well-defined electron density. In addition, there is an SO42− ion bound near AdoHcy at the active site, which is probably introduced at the crystallization process due to the presence of high concentration of Li2SO4 in the crystallization solution. The stereochemical geometry of the structure was analyzed using PROCHECK (Laskowski et al., 1993). The figures were generated using Pymol (http://www.pymol.org). The statistics of the structure refinement and the quality of the final structure model are also summarized in Table 1.

Protein data bank accession code

The structure of Smyd2 in complex with AdoHcy has been deposited with the RCSB Protein Data Bank under accession code 3RIB.

Supplementary material

Supplementary material is available at Journal of Molecular Cell Biology online.
Acknowledgements

We thank the staff members at the Shanghai Synchrotron Radiation Facility of China for technical supports in diffraction data collection, and other members of our group for helpful discussion.

Funding

This work was supported by the grants from the Ministry of Science and Technology of China (2007CB914302 and 2011CB966301), the National Natural Science Foundation of China (30730028), and the Chinese Academy of Sciences (SIBS2008002), and the Science and Technology Commission of Shanghai Municipality (10C1416500). C.Z. gratefully acknowledges the support of the SA-SIBS scholarship program.

Conflict of interest: none declared.

References


Li, B., Jackson, J., Simon, M.D., et al. (2009). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. J. Biol. Chem. 284, 7970–7976.


