doi:10.1016/j.jmb.2006.03.006 J. Mol. Biol. (2006) 359, 86–96

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structural Insights of the Specificity and Catalysis of a Viral Histone H3 Lysine 27 Methyltransferase

Chengmin Qian[1], Xueqi Wang[1], Karishma Manzur[1], Sachchidanand[1] Amjad Farooq[1], Lei Zeng[1], Rong Wang[2] and Ming-Ming Zhou[1]*

[1]*Department of Molecular Physiology and Biophysics Mount Sinai School of Medicine New York University, One Gustave L. Levy Place, New York, NY 10029, USA*

[2]*Department of Human Genetics, Mount Sinai School of Medicine, New York University One Gustave L. Levy Place New York, NY 10029, USA*

**Corresponding author*

SET domain lysine methyltransferases are known to catalyze site and state-specific methylation of lysine residues in histones that is fundamental in epigenetic regulation of gene activation and silencing in eukaryotic organisms. Here we report the three-dimensional solution structure of the SET domain histone lysine methyltransferase (vSET) from *Paramecium bursaria* chlorella virus 1 bound to cofactor *S*-adenosyl-L-homocysteine and a histone H3 peptide containing mono-methylated lysine 27. The dimeric structure, mimicking an enzyme/cofactor/substrate complex, yields the structural basis of the substrate specificity and methylation multiplicity of the enzyme. Our results from mutagenesis and enzyme kinetics analyses argue that a general base mechanism is less likely for lysine methylation by SET domains; and that the only invariant active site residue tyrosine 105 in vSET facilitates methyl transfer from cofactor to the substrate lysine by aligning intermolecular interactions in the lysine access channel of the enzyme.

Keywords: structure; vSET; lysine 27

Introduction

Site-specific histone modifications including acetylation, methylation, phosphorylation and ubiquitination are known to control chromatin-directed gene activation or silencing in eukaryotic genomes.[1] Histone lysine methylation, with the exception of lysine 79 of histone H3, has been shown to be catalyzed exclusively by the conserved SET domain family proteins,[2] which are originally identified in *Drosophila* suppressor of variegation (Su(var)3-9),[3] enhancer of zeste (E(z))[4] and trithorax;[5] hence the name. Lysine methylation is more complex than lysine acetylation, as a lysine can be subject to mono-, di- or tri-methylation. As suggested by the histone code hypothesis,[6,7] acting in a combinatorial or sequential manner with different histone modifications on one or multiple histones, position and state-specific lysine methylation of histones that is accomplished by SET domain histone lysine methytransferases (HKMTs) in a particular biological context specifies unique functional consequences.[8,9] For instance, in cell proliferation, H3-K4 di-methylation by Set1 correlates with basal transcription, whereas H3-K4 tri-methylation is observed at fully activated promoters;[10] recruitment of H3-K9 HKMTs including G9a induces gene repression in euchromatin.[11] During cell differentiation, extended H3-K27 di and/or tri-methylation by the *Drosophila* Polycomb group E(z)-Esc complex[12,13] or by its mammalian Ezh-Eed counterpart[14,15] marks for long-term gene silencing. A combination of H3-K9 di and H3-K27 tri-methylation is associated with inactive X chromosome.[16,17] Moreover, regional H3-K9 tri-methylation by Suv39h at transcriptionally inert chromatin domains is regarded as a hallmark of constitutive heterochromatin.[8] Finally, recent studies show that SET domain proteins can also catalyze lysine methylation of cellular proteins. For example, SET9 has been shown to methylate p53 at Lys373[18] and Taf10,[19] each of which directly leads to

Present address: A. Farooq, Department of Biochemistry and Molecular Biology, University of Miami Miller School of Medicine, 1011 NW 15th Street, Miami, FL 33136, USA.

Abbreviations used: vSET, SET domain histone lysine methyltransferase from *Paramecium bursaria* chlorella virus 1; SAH, *S*-adenosyl-L-homocysteine; SAM, *S*-adenosyl-L-methionine; H3-K27me, mono-methylated K27 H3 peptide; HKMT, histone lysine methytransferases; NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum coherence.

E-mail address of the corresponding author: ming-ming.zhou@mssm.edu

physiological subsequences in the distinct cellular processes.

The high degree of modification complexity and coding potential of histone lysine methylation in epigenetic control of chromatin biology may explain the existence of an unusually large family of SET domain-containing proteins, which contain more than 700 members (with more than 100 members in human).[20] Consistent with their high substrate specificity, SET domain HKMTs show overall low sequence similarity and the residues at the active site are not all conserved. Despite the conserved structural fold of the core SET domain, as shown in several recently reported three-dimensional structures of SET domains,[21–23] there is still no consensus understanding of the SET domain lysine methytransferases' catalytic mechanism; and the structural and molecular bases for different lysine methylation sites (e.g. H3-K27) in histones are not well understood.

Within this extensive family there is a small subclass of SET domain-containing proteins that are encoded by viruses and bacteria,[24,25] which may be important for epigenetic regulation of pathogens and/or host transcription at the chromatin level. Unlike eukaryotic SET domain HKMTs that have pre and post-SET motifs flanking the core SET domain, which are highly variable in amino acid sequences but important for HKMT activity,[21] these viral and bacterial SET domain proteins contain only the core SET domain. Particularly, the prototypical SET domain protein, which is encoded by *Paramecium bursaria* chlorella virus 1 (PBCV-1) and we refer to as vSET, contains only 119 residues and is possibly the smallest known SET domain protein.[24] Moreover, in contrast to the eukaryotic SET domains that usually function as a monomer, vSET is dimeric in solution and exhibits high specificity for methylation of lysine 27 on histone H3. Here, we report the three-dimensional solution structure of vSET in complex with cofactor product *S*-adenosyl-L-homocysteine (SAH) and mono-methylated K27 H3 peptide (H3-K27me), and provide biochemical characterization of its substrate specificity (methylation site and state) and the catalytic mechanism of histone lysine methylation.

Results and Discussion

The 3D structure of the vSET/SAH/H3-K27me ternary complex

To gain the structural insights into substrate recognition and catalytic mechanism, we attempted to solve the three-dimensional structure of vSET in complex with a histone H3 peptide containing mono-methylated K27 (H3-K27me, residues 13–33) and SAH, a product of cofactor *S*-adenosyl-L-methionine (SAM) after donating the methyl group to the substrate. Such a complex mimics that of the enzyme and methyl-donor cofactor and substrate (H3-K27), which by themselves cannot form a stable complex because of enzymatic reaction. As revealed by 2D ^{1}H–^{13}C heteronuclear single quantum coherence (HSQC) spectra, vSET can form a stable ternary complex with SAH or the H3-K27me peptide, but not with either ligand alone (Figure 1(a)). The formation of a stable ternary complex is supported by observation of intermolecular nuclear Overhauser effects (NOEs) between the protein and ligands in 3D $^{13}C/^{15}N$-filtered, ^{13}C-edited NOESY spectra collected with $^{13}C/^{15}N$-labeled vSET bound to unlabeled SAH and H3-K27me peptide (Figure 1(b)).

The three-dimensional structure of vSET in complex with SAH and histone H3-K27me peptide was determined with a total of 2711 NOE-derived distance and dihedral angle restraints obtained by heteronuclear multidimensional NMR methods,[26] which include 155 inter-subunit constraints and 65 intermolecular constraints between vSET and the H3 peptide or cofactor. The overall complex structure of vSET is well-defined by NMR data except for the C-terminal residues 110–119 that exhibited high mobility in solution as indicated by the significantly reduced or negative backbone $\{^{1}H\}$–^{15}N heteronuclear NOEs as compared to the structural core (Figure 2(d)). An ensemble of 20 lowest energy NMR structures is shown in Figure 2(a), and the corresponding NMR structure statistics are listed in Table 1.

Unlike the monomeric structures of eukaryotic SET domain HKMTs,[21] vSET is dimeric in solution.[24] Each subunit of the vSET dimer comprises the conserved structural fold of the SET domain, consisting of two domains (Figure 2(b)). Domain I is an antiparallel β-barrel. Domain II, as an insert in the domain I sequence, is comprised of a three-stranded open-faced sandwich that forms the majority of the dimer interface with its counterpart, knitting together the two subunits in a head-to-head fashion. The H3-K27me peptide binds in an extended conformation in a surface-exposed groove crossing β6 and β8, whereas SAH is surrounded by a cluster of highly conserved residues in the SET domain family, including Asn69, His70 and Asn75 in the α2/β8 loop, and Glu100 and Tyr105 in the C-terminal sequence (Figure 2(b) and (c); see blow). The overall structure of the vSET ternary complex remains to be compact and is not significantly different from its free form[24] with RMSD of 1.6 Å for the secondary structural elements (Figure 2(d)) except that the C-terminal tail becomes more structurally ordered. This local structural restriction is demonstrated by an increase in backbone $\{^{1}H\}$–^{15}N heteronuclear NOEs of the C-terminal residues 115–119 upon complex formation (Figure 2(e)).

Substrate recognition

The new structure provides the structural basis for vSET's high substrate specificity for H3-K27 over H3-K9 (with greater than 20-fold difference in the HMT activity),[24] despite their similar amino acid residues flanking the corresponding lysine. Both

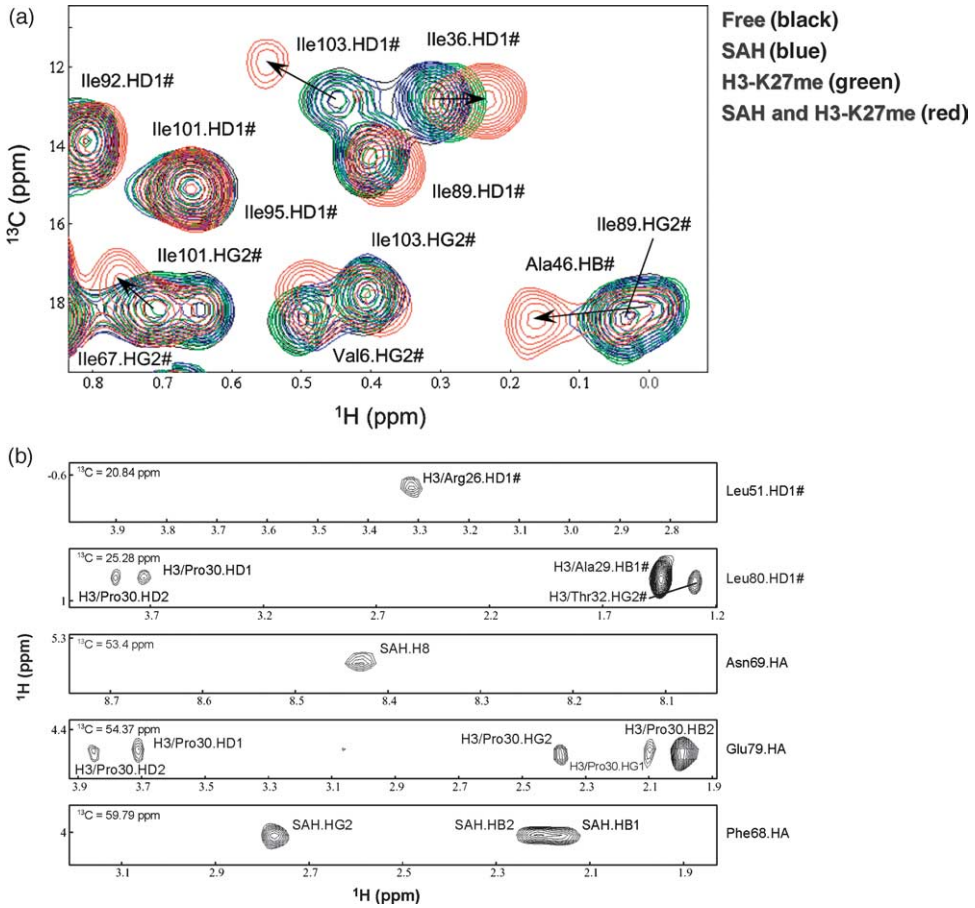

Figure 1. NMR spectra of vSET in tertiary complex association with cofactor SAH and H3-K27me peptide. (a) Binding of SAH and H3-K27me peptide to vSET, as shown by 2D ^{1}H–^{13}CHSQC spectra of the protein in free (black), or with SAH (blue) or H3-K27me peptide (green) alone or with SAH and H3-K27me (red). The molar ratio of vSET:SAH:H3-K27me peptide was kept at 1:5:5. (b) Representative 2D strips of the 3D ^{13}C/^{15}N-filtered, ^{13}C-edited NOESY, showing intermolecular NOEs of the ternary complex, i.e. between vSET and H3-K27me peptide and vSET and SAH.

sites contain a AR**K**S(APA)TGG sequence, where K is Lys9 or Lys27, and the APA motif is present only at Lys27 but not at Lys9 (Figure 3(a)). The H3-K27me peptide binds in a shallow surface-exposed groove defined by Leu51, Phe52 of β6 and His78, Glu79 and Leu80 of β8 on one side, and the C-terminal Tyr105 and Tyr109 on the other side (Figure 3(b) and (c)). Selective recognition of H3 at Lys27 (0) by vSET is consistent with the observation of the strong intermolecular NOEs observed in the ^{13}C/^{15}N-filtered, ^{13}C-edited NOESY spectrum between Ala29 (+2) and Pro30 (+3) of the H3 peptide and Glu79 and Leu80 of vSET. These interactions explain our results that the Pro30-to-Ala mutation resulted in nearly a fourfold reduction in the H3-K27 methylation by vSET, and that insertion of the APA motif between Ser10 and Thr11 in a H3 peptide substrate produced Lys9 methylation to a level nearly as high as that of methylation at Lys27 by vSET.[24] The new structure together with the mutagenesis data concludes that vSET is a specific H3-K27 lysine methyltransferase through its selective recognition of the APA motif in histone H3.

The structure also reveals how other H3 residues flanking both sides of Lys27 contribute to vSET specificity at H3-K27. Particularly, Arg26 (−1) of H3 interacts with Leu51 and Ser53 (Figure 3(b)), whereas the mono-methylated Lys27 has numerous intermolecular NOEs with Tyr50, Phe52 and Tyr105 that together form a lysine access channel at the enzyme active site (Figure 3(d); see below). Moreover, the main-chain amide and side-chain hydroxyl of Ser28 (+1) of H3 possibly form hydrogen bonds, respectively, with the backbone carbonyl of Tyr105 and the side-chain hydroxyl of Ser53 of vSET. These interactions collectively play an important role in the selective H3-K27 recognition by vSET, as phosphorylation of H3 at Ser28 almost completely abolished the H3-K27 methylation activity by vSET,[24] which likely resulted from disruption of these hydrogen bond interactions by the negatively charged phosphate group on Ser28. This is reminiscent of the crystal structure of Dim-5 SET domain/SAH/H3-K9 peptide complex,[27] in which the side-chain hydroxyl of Ser10 of H3-K9 peptide forms a hydrogen bond with Asp209 of Dim-5.

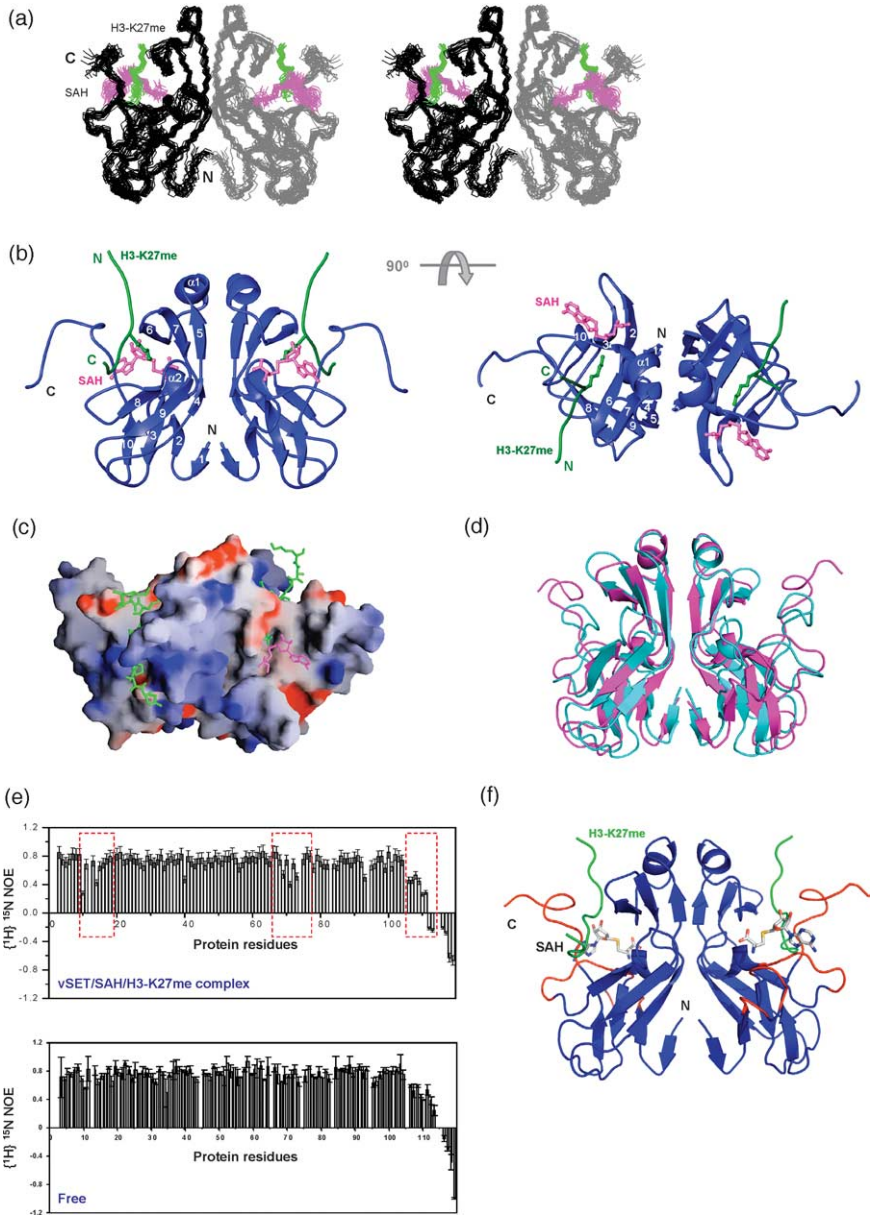

Figure 2. The three-dimensional structure of the tertiary complex of vSET with cofactor SAH and H3-K27me peptide. (a) Stereoview of superimposition of backbone atoms (N, C^{α} and C′) of 20 lowest energy NMR structures of the vSET/SAH/H3-K27me. For clarity, only residues A25–A31 of the H3 peptide are depicted in the final structures. (b) Ribbon diagrams of the vSET/SAH/H3-K27me tertiary complex structure, shown in front view as in (a) (left panel) and top view with 90° rotation (right panel). vSET, SAH and H3-K27me peptide are color-coded in blue, green and pink, respectively. The peptide residues shown in this structure are T22-G33. (c) Surface electrostatic potential representation of the vSET protein in tertiary complex with SAH and H3-K27me peptide, depicted in a front view as in (b). (d) Comparison of the apo and ternary complex structures of vSET. The structures were superimposed with the secondary structural elements only. (e) The backbone $\{^1H\}$–^{15}N heteronuclear NOEs of vSET in the free form (lower panel) and in the tertiary complex with SAH and H3-K27me peptide (top panel). Error bars represent the standard deviation of NOE values measured in three data sets. (f) The regions of vSET that exhibited increased backbone dynamics upon ternary complex formation, as indicated by reduced $\{^1H\}$–^{15}N heteronuclear NOEs in (e), are highlighted in red in the Ribbon diagram of the vSET structure.

Differential modifications of site-specific amino acids in histones have been proposed to be achieved by "local switch" models that modifications of two adjacent residues in histones can be mutually exclusive.[28] For instance, a "methyl/phos" switch may be operational at Lys9/Ser10 in H3, where Lys9 methylation serves as a mark for heterochromatin formation (gene repression) and Ser10 phosphorylation for mitosis (gene activation); likewise at Lys27/Ser28 in H3, where Lys27 methylation is functionally linked to Polycomb group-mediated gene silencing[12–14] and Ser28 phosphorylation is implicated in mitosis.[29] Our new structural insights of vSET, as well as those reported for Dim-5 SET

Table 1. NMR structural statistics for vSET/SAH/H3-K27me

Total experimental restraints	2711	
Total NOE distance restraints	2607	
Intramolecular	2387	
Ambiguous	545	
Unambiguous		
Manually assigned	1641	
ARIA assigned	201	
Intra-residue	926	
Inter-residue		
Sequential $\|i-j\|=1$	401	
Medium $1<\|i-j\|\leq 4$	257	
Long range $\|i-j\|>4$	813	
Intermolecular		
Manually assigned	220	
Between vSET subunits	155	
Between vSET/H3-K27me peptide	34	
Between vSET/SAH	31	
Hydrogen bond restraints	50	
Dihedral angle restraints	54	
Final energies (kcal/mol)[a]		
E_{TOT}	542.9±13.5	
E_{NOE}	78.1±14.5	
E_{DIH}	4.3±0.8	
E_{LJ}	−1233.2±43.2	
Ramachandran plot (%)	Full molecule[b]	Secondary structure[c]
Most favorable regions	53.3±4.4	84.5±3.0
Additionally allowed regions	38.1±3.9	15.5±2.4
Generously allowed regions	6.1±2.1	0.0±0.0
Disallowed regions	2.5±1.3	0.0±0.0
RMSDs of atomic coordinates (Å)		
Backbone	0.84±0.14	0.66±0.16
Heavy atoms	1.22±0.13	1.03±0.15

[a] Based upon 20 lowest energy-minimized structures. The Lennard-Jones potential was not used during any refinement stage. None of these final structures exhibit NOE-derived distance restraint violations greater than 0.5 Å or dihedral angle restraint violations greater than 5°.
[b] Residues 1–101.
[c] Residues 2–3, 6–9, 23–25, 28–31, 42–47, 50–54, 57–61, 64–67, 76–79, 88–91, 94–96 and 99–101.

domain,[27] offer the structural basis for such possible local cross-talk among different modifications in histones that govern chromatin-mediated gene expression or silencing.

Substrate lysine and cofactor binding

The substrate lysine is completely buried in a channel at the enzyme active site formed by Tyr50, Leu51, Phe52 and Ser53 of β6, His78 of β8, and Tyr105 and Tyr109 of the C-terminal sequence (Figure 3(d)). The cofactor SAH is bound at the other end of the channel, with a sulfur atom positioned to the tip of the methyl group of the mono-methylated K27. Note that this lysine access channel is not present in the free enzyme (Figure 4(a) *versus* (b)), and only formed upon substrate and cofactor binding with restructuring side-chains of several residues in the C-terminal sequence (Figure 2(d)). The construction of the active site conformation through cooperative binding of substrate lysine and cofactor is consistent with our NMR titration data (Figure 1(a)), which show that the ^{1}H–^{13}CHSQC spectrum of vSET exhibited major chemical shift perturbations only when both the substrate and cofactor were bound to the enzyme. Such active site structural rearrangement was reported in the Dim-5 SET domain[30] with its C-terminal Trp318 that forms a part of the lysine access channel upon substrate binding (Figure 4), but not in SET7/9[27] or Rubisco LMST[31] SET domain lysine methyltransferases. In the latter structures, the lysine-binding channel is preformed in the enzyme.

The bound cofactor adapts a U-shaped conformation *via* possible hydrogen bonds with main-chain atoms of residues in the NHSxxPN motif (residues 68–74, where x is a non-conserved amino acid) in β8 and GGxG motif (residues 14–17) in the β2/β3 loop, which are highly conserved in the SET domain family (Figure 3(d)). The compact conformation of SAH and the geometry of its binding cleft are also seen in other SET domain structures, but distinct from protein arginine methyltransferases that also use SAM as methyl donor.[23] As revealed in vSET, the polar edge of the adenine base is within the hydrogen bond distance to the backbone atoms of His70, i.e. the exocyclic NH2 of N6 and the aromatic N7 interacting with the carbonyl and amide groups of His70, respectively. Moreover, C8 of the adenine ring possibly makes van der waals contacts to Tyr105, whereas the carboxylate group of homocysteine moiety interacts with the hydroxyl group of Tyr50. In addition, NOEs were observed between the indole ring of Trp110 and the ribose of SAH. Collectively, these interactions likely position the labile methyl group of the cofactor in the pore pointing to the ε-amino nitrogen of the substrate lysine. Finally, the structure explains that the major loss of the enzymatic activity by point mutation of the conserved Asn69, His70 or Glu100 to alanine[24] was likely due to disruption of the hydrogen bonds or van del Waals contacts between SAH and vSET resulting from these mutations.

Direction participation of Tyr109 in the formation of the lysine access channel and observation of intermolecular NOEs from Trp110 to SAH explain how the C-terminal sequence becomes more structurally ordered upon the complex formation, as indicated by the increased backbone $\{^{1}H\}$–^{15}N heteronuclear NOE values (Figure 2(d)). Most notably, upon cofactor and substrate binding the backbone $\{^{1}H\}$–^{15}N heteronuclear NOEs exhibited a significant decrease for residues 10–17 of the β2-β3 loop and residues 68–76 in the α2/β8 loop that together constitute the cofactor SAH binding site, as well as residues 106–113 that form one side of the lysine access channel at the enzyme active site (Figure 2(e) and (f)). These NMR data support the notion that these vSET active site residues are structurally flexible during methylation reaction, thus accommodating the dynamic binding and releasing of cofactor and substrate at the enzyme active site. It is worth noting that Tyr109 in vSET is reminiscent of Trp318 in Dim-5 or Tyr337 in SET7/9 (Figure 4(b) *versus* (c) and (d)). The former forms a part of the lysine access channel through structural

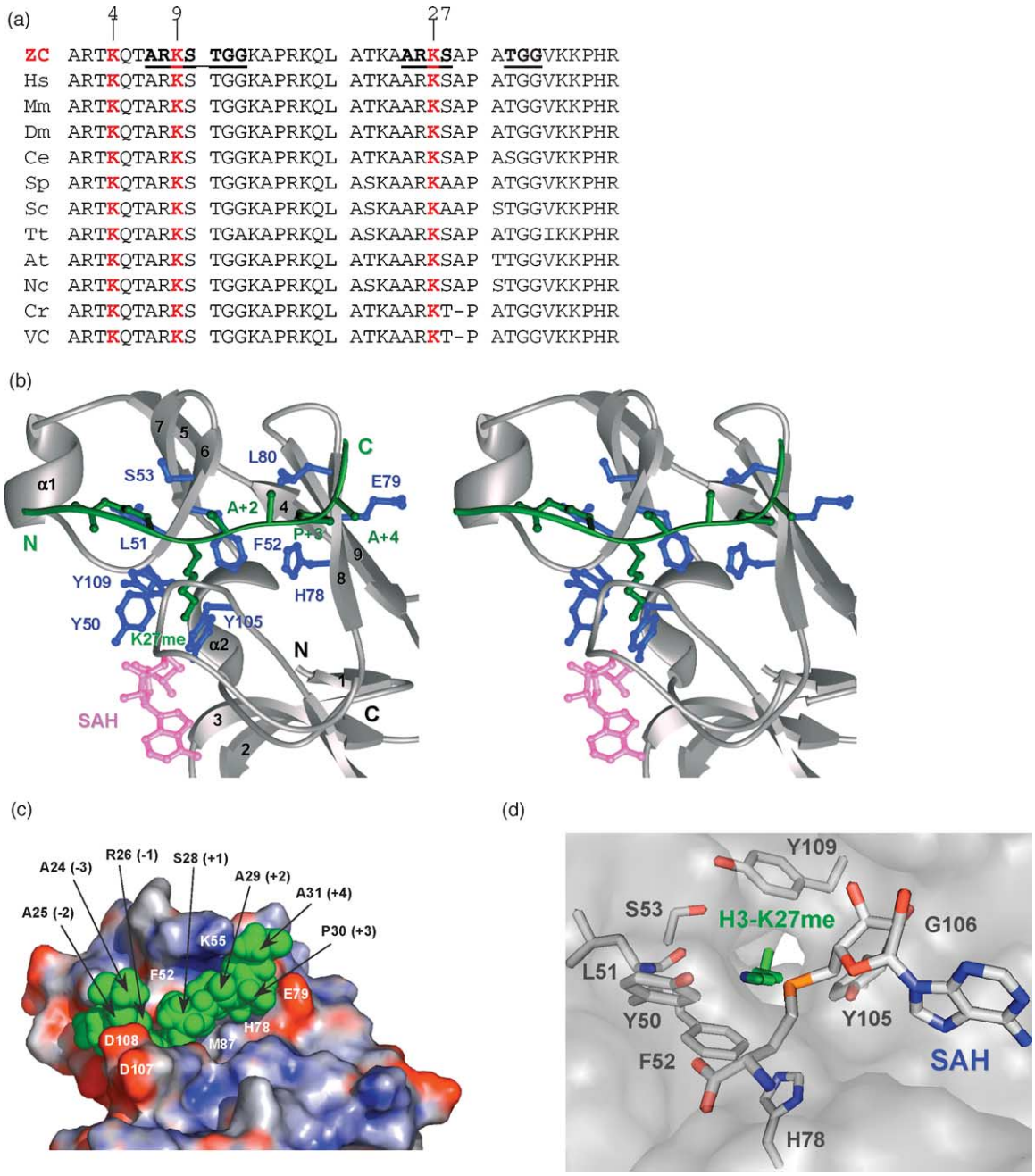

Figure 3. The structural basis of substrate specificity of vSET. (a) Sequence comparison of histone H3 (residues 1–40) between zoochlorella (*ZC*) and other eukaryotes including *Homo sapien*, *Mus musculus*, *Xenopus laevis*, *Caenorabditis elegans* and *Chlamydomonas reinhardtii*. Histone H3 K27 (in red) is conserved among all these histones, and residues flanking K27 are conserved as well, except for in *C. reinhardtii*. The sequences that are similar at K9 and K27 are underlined. (b) Stereoview of a ribbon diagram of the vSET/SAH/H3-K27me tertiary complex structure, highlighting the H3 peptide substrate recognition by the enzyme. The cofactor SAH and the H3 peptide are color-coded in pink and green, whereas the active site residues of vSET are highlighted in blue. (c) Surface representation of the H3-K27me peptide-binding site in vSET. The peptide residues are shown in green in a space-filled model. (d) The substrate lysine access channel and the cofactor SAH-binding site at the active site of vSET. Key residues at the enzyme active site are shown in stick formation and the substrate lysine side-chain is shown in green.

arrangement in Dim-5 upon substrate binding,[30] whereas the latter is already positioned in the pre-formed lysine access channel in SET7/9.[27] Moreover, in the SET8 SET domain complex structure, H4-K20 peptide substrate residue His18 (−2) appears to fulfill such a role in the formation of the substrate lysine access channel (Figure 4(e)).[32]

Molecular determinants of H3-K27 methylation by vSET

The geometry, shape, and type of amino acids that comprise the substrate lysine access channel in the active site have been suggested to confer methylation multiplicity in the SET domain.[27,30]

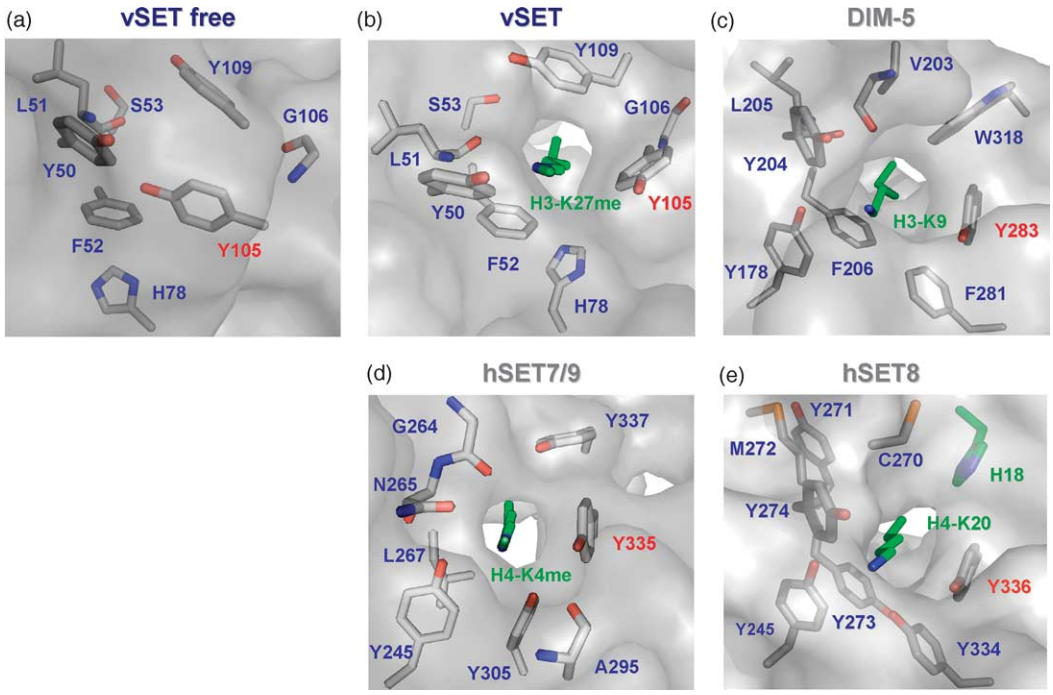

Figure 4. The molecular determinants of H3-K27 methylation by vSET. The geometry and molecular environment of the lysine access channel in SET domains is shown for (a), vSET in the apo form; (b) vSET in the ternary complex; (c) DIM-5; (d) Set7/9; and (e) SET8. The SET domain protein residues are color-coded in gray, blue and red for carbon, nitrogen and oxygen, respectively, whereas the residues of the histone peptide substrates including substrate lysine are color-coded in green for carbon and nitrogen.

We analyzed methylation kinetics of vSET using MALDI-TOF mass spectrometry to assess its methylation multiplicity. As shown in Figure 5(a), vSET catalyzes consecutive mono, di and tri-methylation of a H3-K27 peptide (residues 15–33) with rate constants of 0.047 min^{-1}, 0.015 min^{-1} and 0.005 min^{-1}, respectively (Figure 5(b)). Our steady-state kinetics study showed that the apparent K_m and k_{cat} for the overall H3-K27 methylation by vSET are 276.2($\pm$20.7) μM and 1.32($\pm$0.03) s^{-1}, which are comparable to those of Dim-5, SET7/9 and SET8. The former catalyzes lysine tri-methylation, whereas the latter two catalyze only mono-methylation. Except the sole invariant Tyr105 that corresponds to Tyr283 in DIM-5, Tyr335 in SET7/9 or Tyr336 in SET8, vSET shares additionally Tyr50 and Phe52 with Tyr204 and Phe206 at the corresponding positions in Dim-5 (Figure 4(b) and (c)), consistent with their similarity in methylation multiplicity. Moreover, His78 is positioned similarly in the lysine access channel in vSET as Phe281 in Dim-5.

Mutational analysis of the active site residues in vSET helps explain their contributions to the methylation multiplicity (Figure 6(a) and (b)). For

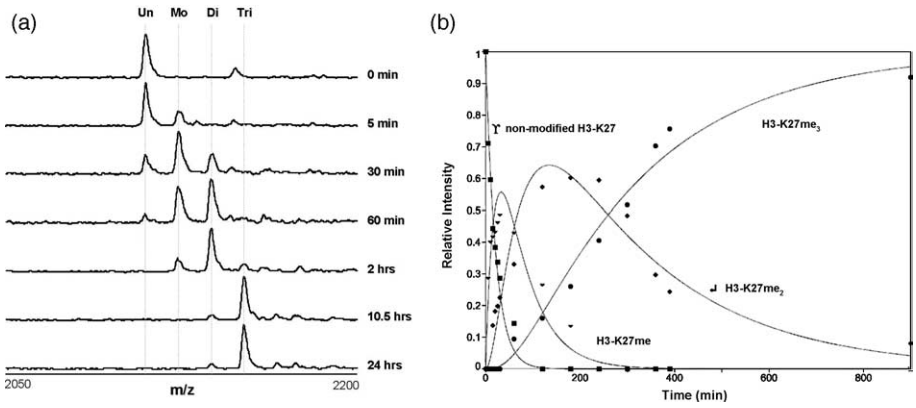

Figure 5. Kinetics of H3-K27 methylation by vSET. (a) Mono, di and tri-methylation of K27 in a histone H3 peptide (residues 13–33) by vSET, as shown by mass spectrometry at representative time points during the reaction. (b) Kinetics of K27 methylation of the same histone H3 peptide by vSET over the entire time-course of the reaction, as assessed by mass spectrometric analysis of the H3-K27 methylation products.

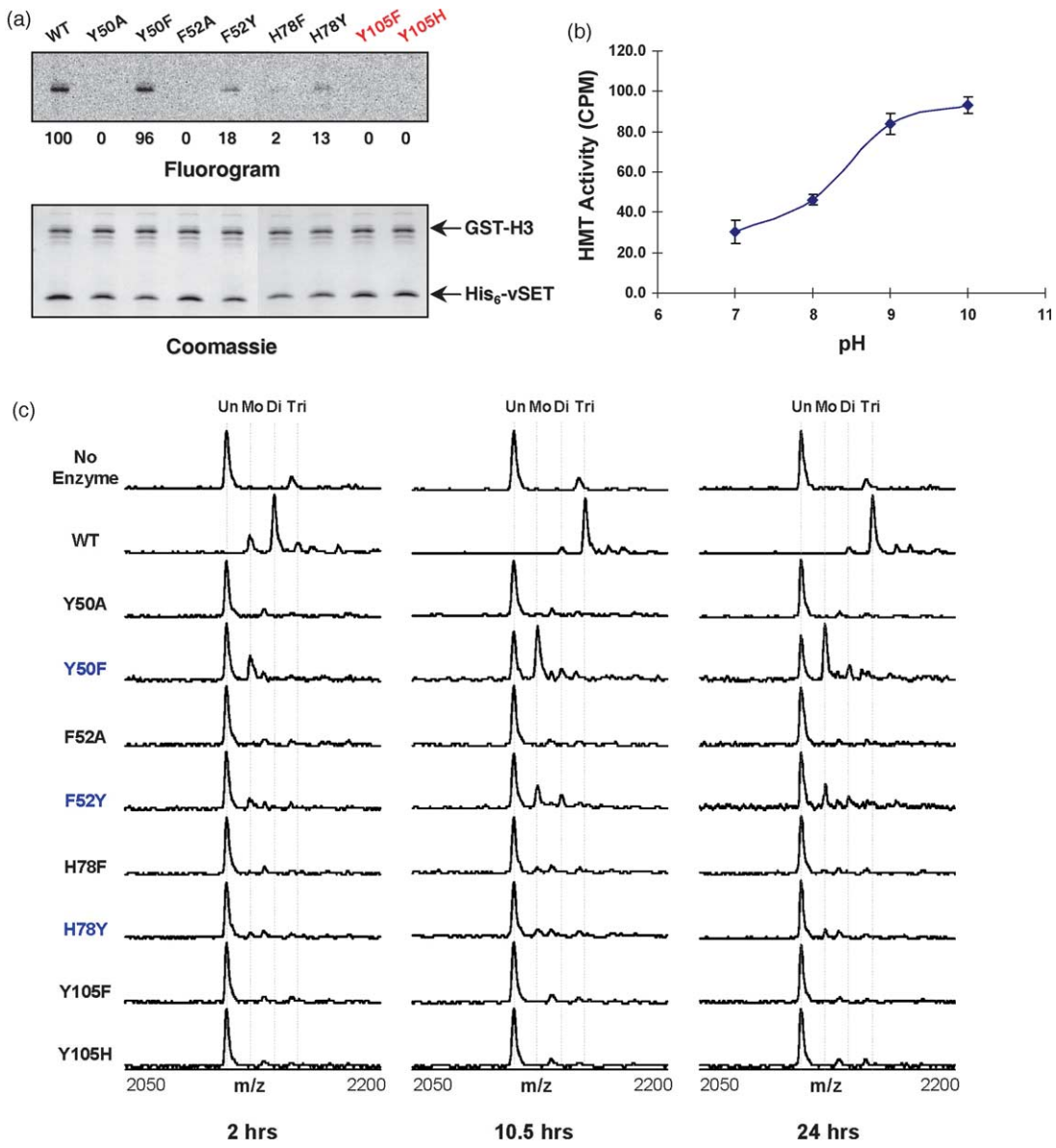

Figure 6. Mutational analysis of vSET active site residues. (a) Mutation analysis of vSET active site residues. *In vitro* HKMT activity of vSET and its mutants measured using GST-histone H3 N terminus (residues 1–57). Relative equal amounts of GST-H3 and vSET proteins used in the assay are shown in an SDS-PAGE gel (lower panel). Signals in the fluorogram represent HKMT activity of vSET for GST-H3 (upper panel). Phosphorescence values are normalized against wild-type vSET signal (lane 1). (b) Mass spectrometric analysis of the histone H3-K27 peptide (residues 13–33) by the wild-type and various active site mutants of vSET, as shown at reaction times of 2, 10.5 and 24 hours. (c) pH dependence of the H3-K27 methylation activity by vSET using an *in vitro* assay as described in (a).

example, mutation of Tyr50 or Phe52 to an Ala caused a nearly complete loss of methylation activity, whereas their more conservative amino acid mutations showed that Y50F became largely a mono-methylase for H3-K27 with little, if any, di or tri-methylation activity, and that F52Y can only mono and di but not tri-methylate H3-K27. Notably, the Y50F mutation nearly abolished di and tri-methylation of H3-K27, but did not affect the mono-methylation, as the apparent kinetic parameters for the latter reaction, i.e. K_m of 427.4($\pm$44.4) μM and k_{cat} of 1.36($\pm$0.09) s^{-1}, are similar to those of the wild-type vSET. On the other hand, the kinetic parameters for F52Y were substantially worsened with severely compromised reaction rate and K_m of 3.9($\pm$1.3) mM, 14-fold higher than the wild-type. Finally, mutation of His78 to a Phe or Tyr resulted in a nearly complete loss of the H3-K27 methylation activity except for a little mono-methylation activity (Figure 6(b)). The mutation effects of His78 are consistent with the reported mutational results of the corresponding Phe281 in Dim-5, which showed that F281Y exhibited only mono and di but not tri-methylation of the target H3-K27.[30] Similarly, mutations of Tyr305 to Phe in SET7/9[27] or Tyr334 to Phe in SET8,[33] which are positioned similarly to that of His78 in vSET or Phe281 in Dim-5, has been shown to alter methylation multiplicity from sole

mono-methylation to mono and di-methylation. Taken together, these results suggest that Tyr50, Phe52 and His78 play an important role in determining the methylation multiplicity of vSET by defining the geometry and size of the narrow lysine access channel and by maintaining a network of molecular interactions between the enzyme, cofactor and the substrate lysine.

Insights into the catalytic mechanism of H3-K27 methylation by vSET

The network of molecular interactions revealed in the structures of the ternary complexes of several SET domains supports a notion that methyl transfer from SAM to the ε-amino of the target lysine likely proceeds by a direct in-line S_N2 nucleophilic attack.[22] While it is believed that only a deprotonated substrate lysine bearing a free lone pair of electrons is capable of the nucleophilic attack on the SAM methyl group, there is no consensus in understanding on the catalytic mechanism. Two possible mechanisms have been discussed in the literature. The first is of a general base catalysis, which predicts that an active site residue deprotonates substrate lysine before methyl-transfer; and the second suggests that pK_a of the substrate lysine is decreased upon incorporating into the lysine access channel at the active site. The supporting evidence for either mechanism is not conclusive. For instance, the only invariant residue Tyr335 in SET7/9 or Tyr283 in Dim-5 SET domain that could act as the general base for lysine deprotonation is separated away from the ε-ammonium group of the target lysine by greater than 3–4 Å in the crystal structures,[22,27] indicating a lacking of hydrogen bonding between them. The latter would be required for a general base catalysis.

In an effort to probe the catalytic mechanism, we performed mutational analysis of the invariant tyrosine in vSET. Notably, mutation of the corresponding Tyr105 to Ala resulted in a complete loss of H3-K27 methylation activity, although this mutant retained ~50% activity in cofactor binding.[24] More conservative mutations of Tyr105 to Phe or His, which had little effect on the protein structure (data not shown), also showed a complete loss of lysine methylation activity (Figures 5(a) and 6(c)), implying that the phenoxyl group of this tyrosine plays an essential role in the catalysis, which is other than substrate or cofactor binding as suggested for Tyr50, Phe52 and His78. Interestingly, while the pH dependence of the H3-K27 methylation activity of the wild-type vSET exhibited an apparent pK_a of pH 9.0 (Figure 6(c)), the Y105F or F105H mutant exhibited little activity at pH 8.0 or above pH 9.0, arguing that this invariant Tyr105 does not likely act as a general base that deprotonates substrate lysine. Collectively, these results argue that the general base mechanism is less likely, and instead that Tyr105 may function to provide a charge–dipole interaction between its phenoxyl oxygen and the positively charged sulfur of SAM, or may interact with and align the amino nitrogen of the substrate lysine to the methyl carbon of SAM to facility the S_N2 methyl transfer. Such a mechanism, which needs to be further validated experimentally, is reminiscent of what has been shown for glysine *N*-methyltransferase.[34] This mechanism may be consistent with the notion that the target lysine becomes deprotonated while incorporating into the lysine access channel at the active site, as a deprotonated lysine exists in a minor fraction in solution at pH < 10. The formation of this lysine access channel is likely coordinated with substrate and cofactor-binding induced structural re-ordering of vSET that involves its C-terminal tail (residues 106–119), which agrees with the low turnover numbers and high pH optima for the SET domain histone lysine methyltransferases.

Conclusion

We solved the three-dimensional solution structure of the dimeric viral histone lysine methyltransferase vSET in a ternary complex with an H3-K27me peptide and the cofactor SAH. We further investigated the structural basis of substrate specificity and methylation multiplicity. The ternary complex structure explains the basis of the substrate specificity of vSET for histone H3 at K27 but not K9 through recognition of the APA motif (H3 residues 29–31) that is present at the former but not the latter site. Structural re-ordering of the C-terminal segment of vSET upon enzyme/substrate/cofactor complex formation plays an important role in establishing a narrow substrate lysine access channel at the active site, in which the structural geometry and dynamics of the molecular interaction network define the target lysine methylation multiplicity. Moreover, our structure-based mutagenesis and enzyme kinetics analyses support the notion that a general base mechanism appears less likely for the lysine methylation catalysis by SET domains, and instead that incorporation of the substrate lysine into the lysine access channel through interactions with the active site residues possibly helps its deprotonation prior to methyl transfer from the methyl donor cofactor SAM. Our results further suggest that in addition to the other conserved amino acid residues at the active site, the only invariant residue Tyr105 in vSET may be responsible for facilitating the methyl transfer from SAM to the substrate lysine by aligning intermolecular interactions between them at the substrate lysine access channel in the enzyme.

Experimental Procedures

Sample preparation

The recombinant vSET from *Paramecium bursaria* chlorella virus 1 was prepared using the procedure as described.[24] Uniformly ^{15}N and $^{15}N/^{13}C$-labelled

proteins were prepared by growing bacteria in a minimal medium containing $^{15}NH_4Cl$ with or without [$^{13}C_6$]-glucose. A uniformly $^{15}N/^{13}C$-labelled and fractionally deuterated protein was prepared by growing the cells in 75% ^2H_2O. NMR samples contained ~0.5 mM protein in a 50 mM phosphate buffer (pH 6.5) containing 700 mM NaCl, 300 mM urea, 0.1 mM EDTA and 5 mM β-mercaptaethanol (ME) in $H_2O/^2H_2O$ (9/1) or ^2H_2O.

NMR structure determination

The protein backbone assignments were made by using 3D triple-resonance spectra of HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and (H)C(CO)NH-TOCSY recorded on uniformly $^{13}C/^{15}N$-labelled and fractionally deuterated vSET domain in a tertiary complex with a unlabelled H3-K27me peptide (residues 13~33) and co-factor SAH with molar ratio of 1:5:5. The side-chain assignments were obtained using HCCH-TOCSY and HCCH-COSY spectra. The H3 peptide and cofactor SAH resonances were assigned using $^{13}C/^{15}N$-filtered and *J*-resolved NOESY and TOCSY spectra recorded on the $^{13}C/^{15}N$-labelled protein with unlabelled H3 peptide and SAH, and ROESY and TOCSY spectra recorded on the unlabelled free H3 peptide and free cofactor SAH. Interproton distance restraints were obtained from ^{15}N or ^{13}C-edited 3D NOESY or 2D homonuclear NOESY spectra. $^3J_{HN,H\alpha}$ coupling constants measured from 3D HNHA data were used to determine backbone ϕ-angle restraints. The intermolecular NOEs between labelled protein and unlabelled H3 peptide and SAH were obtained by 3D ^{13}C-F_1 edited, $^{13}C/^{15}N$-F_3 filtered NOESY spectra. The intermolecular NOEs used in defining the dimer interface of vSET were obtained as described.[24]

Structures of vSET/H3-K27me peptide/SAH ternary complex were calculated using distance geometry simulated annealing protocol with X-PLOR.[35] Initial protein structure calculations were performed with manually assigned NOE-derived distance restraints. Hydrogen-bond distance restraints, generated from the $H/^2H$ exchange data, were added at a later stage of structure calculations for residues with characteristic NOE patterns. The converged structures were used for the iterative automated NOE assignment by ARIA[36] that integrates with X-PLOR for refinement. In the dimer structure calculation, only one subunit was subjected to experimental distance and angular restraints. The non-crystallography symmetry constraints and distance-based symmetry constraints have been applied for the final dimer structure calculation.[37] The quality of calculated structures was assessed with Procheck-NMR (Table 1).

Mutagenesis and histone methyltransferase assay

Mutant vSET proteins were prepared using the QuikChange site-directed mutagenesis kit (Stratagene). The presence of appropriate mutations was confirmed by DNA sequencing. The *in vitro* histone methyltransferase assay was carried out using a procedure similar to that reported.[24,38] Briefly, the enzymatic reaction was carried out in a 20 mM Tris buffer (pH 8.0) containing 20 mM KCl, 10 mM $MgCl_2$, 10 mM β-ME, 10 μg of histone H3 (Roche), and 300 nCi of *S*-adenosyl-[^{14}C-*methyl*]-L-methionine (Amersham) for 0.5–1 h at 25 °C. The reaction products were separated by SDS-PAGE and visualized by Coomassie staining and fluorography. Methylation reaction by vSET was also performed using histone peptide substrates followed by mass spectrometry analysis.

Coordinate deposition

The restraints used to calculate the structure and the chemical shift assignments have been deposited in the BMRB under the accession number 6997. The structure coordinates of the vSET/SAH/H3-K27me ternary complex have been deposited in the RCSB Protein Data Bank under the accession code 2G46.

Acknowledgements

We acknowledge the use of the NMR facility at the New York Structural Biology Center for this study. S. is a recipient of an American Foundation for AIDS Research (amfAR) Postdoctoral Fellowship. This work was supported by grants from the National Institutes of Health to M.-M.Z.

References

1. Jenuwein, T. & Allis, C. D. (2001). Translating the histone code. *Science*, **293**, 1074–1080.
2. Jenuwein, T. (2001). Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266–273.
3. Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. & Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13**, 3822–3831.
4. Jones, R. S. & Gelbart, W. M. (1993). The *Drosophila* Polycomb-group gene enhancer of zeste contains a region with sequence similarity to trithorax. *Mol. Cell. Biol.* **13**, 6357–6366.
5. Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V. & Harte, P. J. (1995). The *Drosophila* trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech. Dev.* **52**, 209–223.
6. Strahl, B. D. & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, **403**, 41–45.
7. Turner, B. M. (2002). Cellular memory and the histone code. *Cell*, **111**, 285–291.
8. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. (2003). An epigenetic road map for histone lysine methylation. *J. Cell. Sci.* **116**, 2117–2124.
9. Bannister, A. J., Schneider, R. & Kouzarides, T. (2002). Histone methylation: dynamic or static? *Cell*, **109**, 801–806.
10. Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S. *et al*. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl Acad. Sci. USA*, **99**, 8695–8700.
11. Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M. *et al*. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791.

12. Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. & Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, **111**, 185–196.
13. Muller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B. *et al*. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell*, **111**, 197–208.
14. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P. *et al*. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039–1043.
15. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905.
16. Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H. *et al*. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science*, **300**, 131–135.
17. Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C. & Allis, C. D. (2002). Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nature Genet.* **30**, 73–76.
18. Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S. *et al*. (2004). Regulation of p53 activity through lysine methylation. *Nature*, **432**, 353–360.
19. Kouskouti, A., Scheer, E., Staub, A., Tora, L. & Talianidis, I. (2004). Gene-specific modulation of TAF10 function by SET9-mediated methylation. *Mol. Cell.* **14**, 175–182.
20. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA*, **95**, 5857–5864.
21. Trievel, R. C. (2004). Structure and function of histone methyltransferases. *Crit. Rev. Eukaryot. Gene Expr.* **14**, 147–169.
22. Xiao, B., Wilson, J. R. & Gamblin, S. J. (2003). SET domains and histone methylation. *Curr. Opin. Struct. Biol.* **13**, 699–705.
23. Cheng, X., Collins, R. E. & Zhang, X. (2005). Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 267–294.
24. Manzur, K. L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A. W., Sachchidanand & Zhou, M. M. (2003). A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nature Struct. Biol.* **10**, 187–196.
25. Van Etten, J. L. (2003). Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* **37**, 153–195.
26. Clore, G. M. & Gronenborn, A. M. (1994). Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 249–363.
27. Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U. & Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell.* **12**, 177–185.
28. Fischle, W., Wang, Y. & Allis, C. D. (2003). Binary switches and modification cassettes in histone biology and beyond. *Nature*, **425**, 475–479.
29. Strahl, B. D. & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, **403**, 41–45.
30. Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G. *et al*. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature*, **421**, 652–656.
31. Trievel, R. C., Flynn, E. M., Houtz, R. L. & Hurley, J. H. (2003). Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. *Nature Struct. Biol.* **10**, 545–552.
32. Xiao, B., Jing, C., Kelly, G., Walker, P. A., Muskett, F. W., Frenkiel, T. A. *et al*. (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev.* **19**, 1444–1454.
33. Couture, J. F., Collazo, E., Brunzelle, J. S. & Trievel, R. C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev.* **19**, 1455–1465.
34. Takata, Y., Huang, Y., Komoto, J., Yamada, T., Konishi, K., Ogawa, H. *et al*. (2003). Catalytic mechanism of glycine N-methyltransferase. *Biochemistry,* **42**, 8394–8402.
35. Brunger, A. T. (1993). *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR* (version 3.1 edit.), Yale University Press, New Haven, CT.
36. Nilges, M. & O'Donoghue, S. (1998). Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spectrosc.* **32**, 107–139.
37. Nilges, M. (1993). A calculation strategy for the structure determination of symmetry dimers by ^{1}H NMR. *Proteins: Struct. Funct. Genet.* **17**, 297–309.
38. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M. *et al*. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 593–599.

Edited by P. Wright

(Received 20 January 2006; received in revised form 1 March 2006; accepted 2 March 2006)
Available online 20 March 2006