A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3

Karishma L. Manzur, Amjad Farooq, Lei Zeng, Olga Plotnikova, Alexander W. Koch, Sachchidanand and Ming-Ming Zhou

Published online 3 February 2003; doi:10.1038/nsb898

Site-specific lysine methylation of histones by SET domains is a hallmark for epigenetic control of gene transcription in eukaryotic organisms. Here we report that a SET domain protein from *Paramecium bursaria* chlorella virus can specifically di-methylate Lys27 in histone H3, a modification implicated in gene silencing. The solution structure of the viral SET domain reveals a butterfly-shaped head-to-head symmetric dimer different from other known protein methyltransferases. Each subunit consists of a Greek-key antiparallel β-barrel and a three-stranded open-faced sandwich that mediates the dimer interface. Cofactor *S*-adenosyl-L-methionine (SAM) binds at the opening of the β-barrel, and amino acids C-terminal to Lys27 in H3 and in the flexible C-terminal tail of the enzyme confer the specificity of this viral histone methyltransferase.

Evolutionarily conserved SET domains were originally identified in three *Drosophila* proteins: Suppressor of variegation (Su(var)3-9)[1], Enhancer of zeste (E(z))[2] and Trithorax[3]. Presently, SET domains are found in >360 genes in organisms ranging from virus to human listed in the SMART database[4]. Recent advances in genetic and biochemical studies have revealed that SET domains in many proteins function as histone methyltransferases (HMTases) to methylate lysine residues at specific positions within the N termini of histones[5–7]. Site-specific lysine methylation on histone tails in combination with other post-translational modifications, such as acetylation and phosphorylation, is referred to as the 'histone code' that controls a broad spectrum of chromatin-based biological processes[8,9]. These epigenetic control processes include heterochromatin formation[10–13], heterochromatic gene silencing[14], euchromatic transcriptional repression[15,16] and activation[17], initiation and maintenance of X chromosome inactivation[18], triggering of DNA methylation[19] and programmed DNA elimination[20]. Notably, most of the SET domain proteins show remarkably high substrate specificity[6,7]. For instance, *Drosophila* Su(var)3-9 (ref. 21) and *Schizosaccharomyces pombe* Clr4 (ref. 12) methylate H3 only at Lys9, yeast SET1 (ref. 22) and SET7/9 (refs. 23,24) methylate H3 at Lys4, whereas human G9a[25] can methylate H3 at both Lys9 and Lys27. Also, *Drosophila* SET8 methylates specifically histone H4 at Lys20 (ref. 26). Furthermore, distinct mono-, di- or tri-methylation states of a given lysine in histone tails, modified by SET domain HMTases, have recently been shown to be linked to different epigenetic processes *in vivo*[27,28]. In an effort to understand the structural and molecular basis of catalysis of SET domain HMTases, we report here the first three-dimensional solution structure and substrate specificity of a SET domain protein (here we refer to as vSET) encoded by *Paramecium bursaria* chlorella virus 1 (PBCV-1).

SET domain of *Paramecium bursaria* chlorella virus

PBCV-1 is the prototype of a family of large, icosahedral, double-stranded DNA–containing viruses that are known to replicate in certain unicellular, eukaryotic chlorella-like green algae, particularly zoochlorellae[29]. DNA sequence analysis of PBCV-1 reveals that this giant virus contains a large 330 kb genome of 376 protein-encoding genes[30]. The SET domain-containing PBCV-1 protein consists of 119 amino acids and represents the smallest known SET domain–containing protein in the SET domain family. vSET lacks the cysteine-rich pre-SET and post-SET motifs flanking the conserved core SET domain, which are required for HMTase activity in various SET domain proteins including human SUV39H1 (ref. 21). In addition, vSET has an alanine instead of a conserved arginine in the SET-domain signature motif **RFΦNHSCXPN**, where the absolutely or highly conserved residues are highlighted by an underline or in bold, respectively (**Φ** is a hydrophobic residue, and X is any amino acid) (Fig. 1). This arginine has been shown to be important for HMTase activity of several SET domains; substitution of a histidine at the corresponding position in SUV39H1 enhanced its activity by ~20-fold[21] and mutation of the arginine in G9a to histidine abolished its HMTase activity[25].

Methylation of Lys27 of histone H3 by vSET

To investigate possible enzymatic activity of vSET, we cloned, expressed and purified the recombinant vSET (see Methods). *In vitro* HMTase assays using histones H2A, H2B, H3 and H4 as potential substrates showed that vSET is capable of transferring ^{14}C-labeled methyl group(s) from *S*-adenosyl-[^{14}C-methyl-]-L-methionine (SAM) selectively to histone H3 but not to the other histones (Fig. 2*a*). To determine whether H3 methylation by vSET is specific, we conducted the HMTase assay with synthetic peptides derived from the N-terminal sequence in H3 that contain known methylation sites. Notably, methyl transfer catalyzed by vSET takes place only with the H3 peptide containing residues 15–30 but not with another H3 peptide consisting of residues 1–20 (Fig. 2*b*, substrates 1 and 2, respectively). This difference in HMTase activity is directly related to the ability of vSET to bind the former but not the latter H3 peptide, as

Structural Biology Program, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, One Gustave L. Levy Place, New York, New York 10029, USA.

Correspondence should be addressed to M.-M.Z. *e-mail: zhoum@inka.mssm.edu*

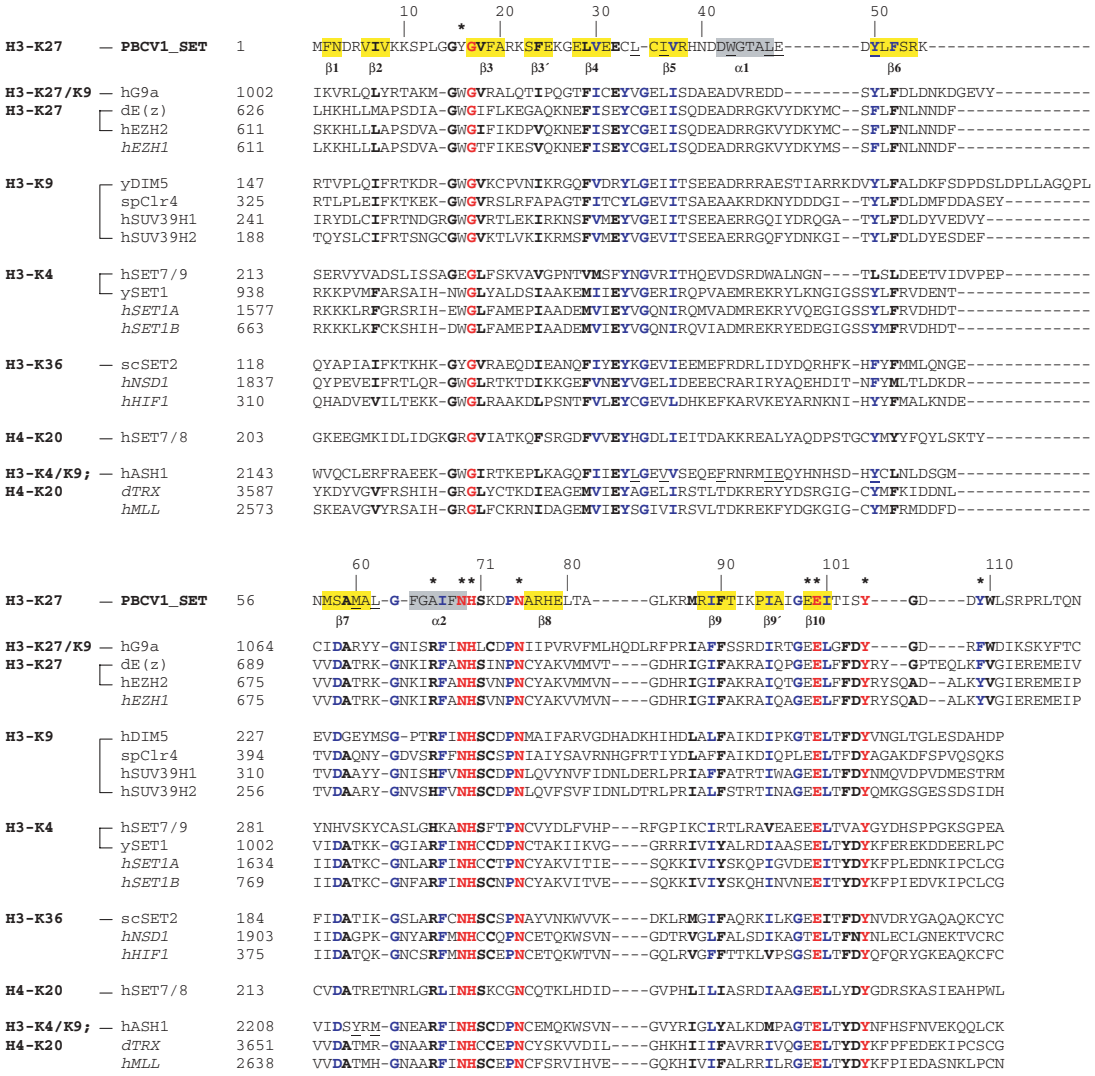

Fig. 1 Structure-based sequence alignment of SET domains. The sequences were aligned on the basis of the NMR-derived structure of the SET domain from *P. bursaria* chlorella virus 1. SET domain proteins are grouped on the basis of sequence similarities[7] and the known substrate specificity towards different lysine methylation sites of histone H3 and H4. The SET domains with known site-specific histone lysine methylation activity are marked by a dash or a bracket, and SET domains with unknown specificity are shown in italics. Residues numbers are indicated along the sequences. Residue numbers of vSET are indicated above its sequence. Absolutely conserved residues are red. Highly conserved residues are black (bold) or blue, with blue letters indicating a higher degree of conservation. Residues of vSET that showed intermolecular NOEs between the subunits in the dimer are underlined, and residues subjected to mutagenesis are indicated by asterisks. Residues in hASH1 that are similar to corresponding vSET residues located in the dimer interface are underlined.

demonstrated by NMR (Fig. 2*c*), underscoring its remarkably high selectivity even towards peptide substrates. Moreover, vSET showed little methyltransferase activity towards the same H3 peptide when it is already di-methylated at Lys27 or phosphorylated at Ser28 (Fig. 2*b*, substrates 3 and 4, respectively), suggesting that vSET can selectively methylate Lys27 in H3. This conclusion was confirmed by a comparision of NMR spectra of the vSET-methylated H3 peptide and a synthetically prepared H3 peptide containing di-methylated Lys 27 (data not shown). Notably, methylation of Lys27 in H3 has been functionally linked to Polycomb group-mediated gene silencing[28,31,32], and phosphorylation at Ser28, which we show here to inhibit methylation at Lys27, has been implicated in the control of mitosis[33]. To determine additional HMTase specificity of vSET, we monitored the enzymatic reaction by reversed-phase HPLC and MALDI-TOF mass spectrometry analysis. These studies show that vSET can mono- and di-methylate, but not tri-methylate, Lys27 of the H3 peptide (Fig. 2*d*). Taken together, these results argue that this PBCV-1 SET domain protein is a methyltransferase specific to Lys27 of histone H3.

Interestingly, amino acids flanking Lys9 and Lys27 are strikingly similar — both contain a ARKS(APA)TGG sequence, where K is Lys9 or Lys27 and the APA motif is present only at Lys27 but not at Lys9. This sequence similarity in H3 may explain why G9a methylates both Lys9 and Lys27 in H3 (ref. 25). In an effort to determine whether the APA motif is important for the substrate specificity of vSET, we prepared an H3 peptide substrate (residues 19–33) containing a P30A mutation. Remarkably, this point mutation caused a nearly four-fold reduction in methylation at Lys27 in the mutant H3 peptide substrate by vSET (Fig. 2*e*, substrates 1 and 2). Furthermore, insertion of the APA motif between Ser10 and Thr11 in a H3 peptide substrate resulted in a

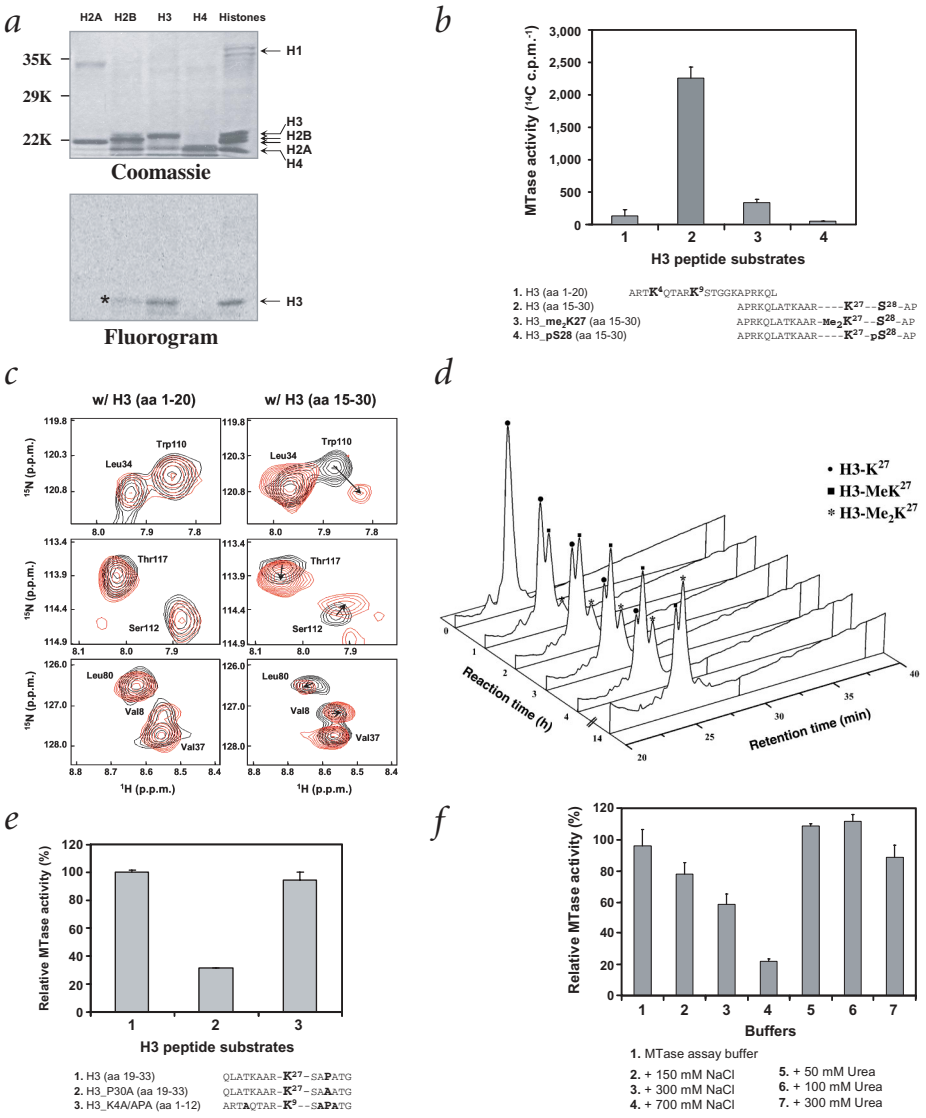

Fig. 2 Histone methyltransferase activity of vSET. ***a***, HMTase activity of vSET for free histones measured in the *in vitro* HMTase assay (lower panel). Relative amount of free histones used in the HMTase assay are shown in an SDS-PAGE gel (upper panel). When judged on the basis of molecular weight, the weak signal in the H2B lane in the fluorogram (indicated by an asterisk, lower panel) is probably due to H3 contamination in the H2B protein sample obtained from the manufacturer. ***b***, HMTase activity of vSET measured against histone H3 peptides in a reaction condition similar to that of (*a*). The amino acid sequences of H3 peptides are shown below the enzyme activity plot, where Me_2K and pS stand for Nε-di-methylated lysine and phosphorylated serine, respectively. ***c***, vSET binding to H3 peptides. Superimposition of 2D $^1H/^{15}N$ HSQC spectra of vSET shows representative protein backbone resonances in the absence (black) and the presence (red) of H3 peptide containing residues 1–20 (left column) or 15–30 (right column). ***d***, HPLC analysis of the enzyme kinetics of vSET. The HPLC stacking plot depicts time-dependent formation of mono- and di-methylated–Lys27 H3 peptides (residues 15–30) catalyzed by vSET. Elution peaks were confirmed by MALDI-TOF mass spectrometry analysis. HPLC peaks corresponding to the non-, mono- and di-methylated–Lys27 H3 peptides are marked by solid dot, solid square and asterisk, respectively. ***e***, Effect of amino acid changes of the histone H3 sequence on the lysine methylation by vSET. The reaction condition was similar to that of (*b*). Lys4 in substrate 3 was changed to Ala to ensure a direct comparison of methylation between Lys9 and Lys27. ***f***, HMTase activity of vSET in the presence of varying amounts of urea and sodium chloride, measured using the H3-peptide substrate (amino acids 15–30) as described (see Methods).

level of methylation of Lys9 comparable to that of Lys27 by vSET (Fig. 2*e*, substrates 1 and 3). Lys4 in the H3 peptide substrate 3 was changed to Ala to ensure a direct comparison of methylation between Lys9 and Lys27 (Fig. 2*e*). These results strongly suggest that residues C-terminal to Lys27 in H3, including Pro30 of the APA motif, dictate the highly specific Lys27 methylation by vSET.

The amino acid sequence of histone H3 of zoochlorellae is presently not available. However, H3 sequences from the green algae *Chlamydomonas reinhardtii* and *Volvox carteri* show >90% sequence identity in the N terminus (residues 1–40) to H3 proteins from various organisms, including amino acids flanking Lys27. Thus, on the basis of highly specific HMTase activity of vSET shown here, it is reasonable to postulate that the *in vivo* function of this viral SET domain protein is to methylate histones of the host algae to silence gene transcription. This plausible mechanism may explain the observation that 75% of gene transcription in the host zoochlorellae is blocked within one hour after infection by PBCV-1 (ref. 29).

Structure determination of vSET dimer

In an effort to perform structural analysis of vSET by NMR, we optimized buffer conditions and found that the NMR spectra of the protein were excellent in the presence of 700 mM NaCl and 300 mM urea. The low amount of urea used in the NMR buffer enhanced protein stability and the high concentration of salt significantly improved the quality of the NMR spectra (data not shown). Although the MTase activity of vSET is sensitive to buffer ionic strength, the low amount of urea used in the NMR buffer does not hinder its enzymatic activity (Fig. 2*f*). Under such NMR buffer conditions, gel filtration chromatography analysis revealed that the protein elutes with an apparent molecular weight (M_w) of 25.1 kDa, which is consistent with a dimeric form of vSET (calculated M_w of a dimeric vSET is 27.2 kDa; Fig. 3*a*).

To assess whether the dimerization of vSET is dependent upon the buffer conditions, we performed sedimentation experiments by analytical ultracentrifugation at protein concentrations much lower than those in the NMR structural study. Under the NMR buffer conditions described above, sedimentation equilibrium experiments (Fig. 3*b*) yield a molar mass of 27.3 ± 0.2 kDa for vSET, confirming its dimeric structure at low protein concentrations (~1 μM; Table 1). The dimeric state of vSET does not change in the absence of urea or at the physiological salt concentration because both sedimentation coefficient and molecular

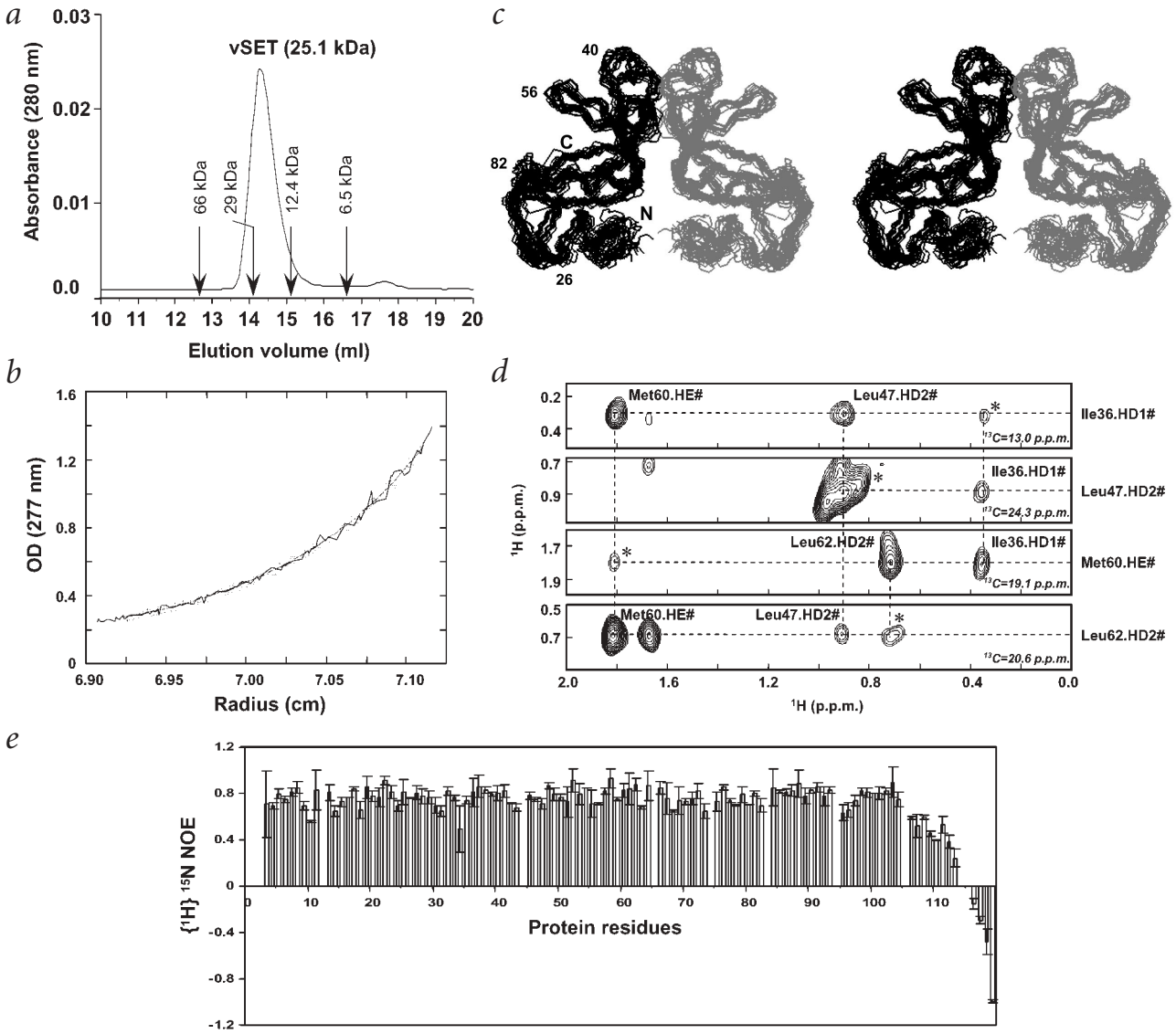

Fig. 3 Structure determination of the vSET dimer. ***a***, Gel-filtration chromatography analysis of vSET dimer formation in solution. Protein standards used in the experiment were bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome *c* (12.4 kDa) and aprotinin (6.5 kDa), as indicated in the chromatogram. vSET was eluted as a protein of 25.1 kDa, which is consistent with its dimeric form. ***b***, Typical sedimentation equilibrium profile of vSET as obtained from analytical ultracentrifugation experiments. A M_w of 27.3 ± 0.2 kDa was calculated from the data, indicating that the protein exists predominantly as a dimer in the NMR buffer. ***c***, Stereo view of the backbone atoms (N, Cα and C′) of 20 superimposed NMR-derived structures of the vSET dimer. For clarity, structures of the adjacent subunits of the dimer are colored in black and gray. ***d***, Representative 2D strips of the 3D ^{13}C/^{15}N-filtered, ^{13}C-edited NOESY spectrum showing intermolecular NOEs observed between the adjacent subunits of the dimer. The NOESY was collected with a 1:1 ration of ^{13}C/^{15}N-labeled/nonlabeled protein sample of vSET. Peaks indicated by asterisks are diagonal peaks. The ^{13}C chemical shift of the 2D strips is indicated in the bottom right corner of each spectrum. ***e***, The backbone {^{1}H}-^{15}N heteronuclear NOEs of the protein residues. Error bars represent the standard deviation (s.d.) of NOE values measured in three data sets.

mass are virtually the same (Table 1). Taken together, these results prove that vSET indeed exists as a dimer in solution.

We determined the three-dimensional structure of vSET with a total of 2,059 NOE-derived distance and dihedral angle restraints obtained from heteronuclear multidimensional NMR methods[34,35]. The structure taken from an ensemble of 20 lowest energy NMR structures (Fig. 3*c*) confirms that the enzyme forms a dimer in solution. Dimer formation of vSET is supported by numerous intermolecular NOEs obtained in 3D ^{13}C/^{15}N-filtered and ^{13}C-edited NOESY spectra collected with a ^{13}C/^{15}N-labeled/nonlabeled protein sample (1:1 ratio of labeled to nonlabeled protein; Fig. 3*d*). The overall structure of vSET is well defined by NMR data, except for the C-terminal residues 105–119, which are not structured (Table 2). These C-terminal residues have almost no long-range NOEs to the rest of the protein and show significantly reduced or negative backbone {^{1}H}-^{15}N heteronuclear NOEs, which is indicative of high mobility in solution (Fig. 3*e*).

Structure of vSET dimer

The structure of vSET is a two-fold symmetric, butterfly-shaped dimer (Fig. 4*a*). Each subunit of the dimer consists of two domains. Domain I is an antiparallel β-barrel with an unusual Greek-key topology[36]. Domain II, which is an insert of domain I in protein sequence, consists of a three-stranded open-faced sandwich with the classical C-centered overhand +2x, –1 topo-

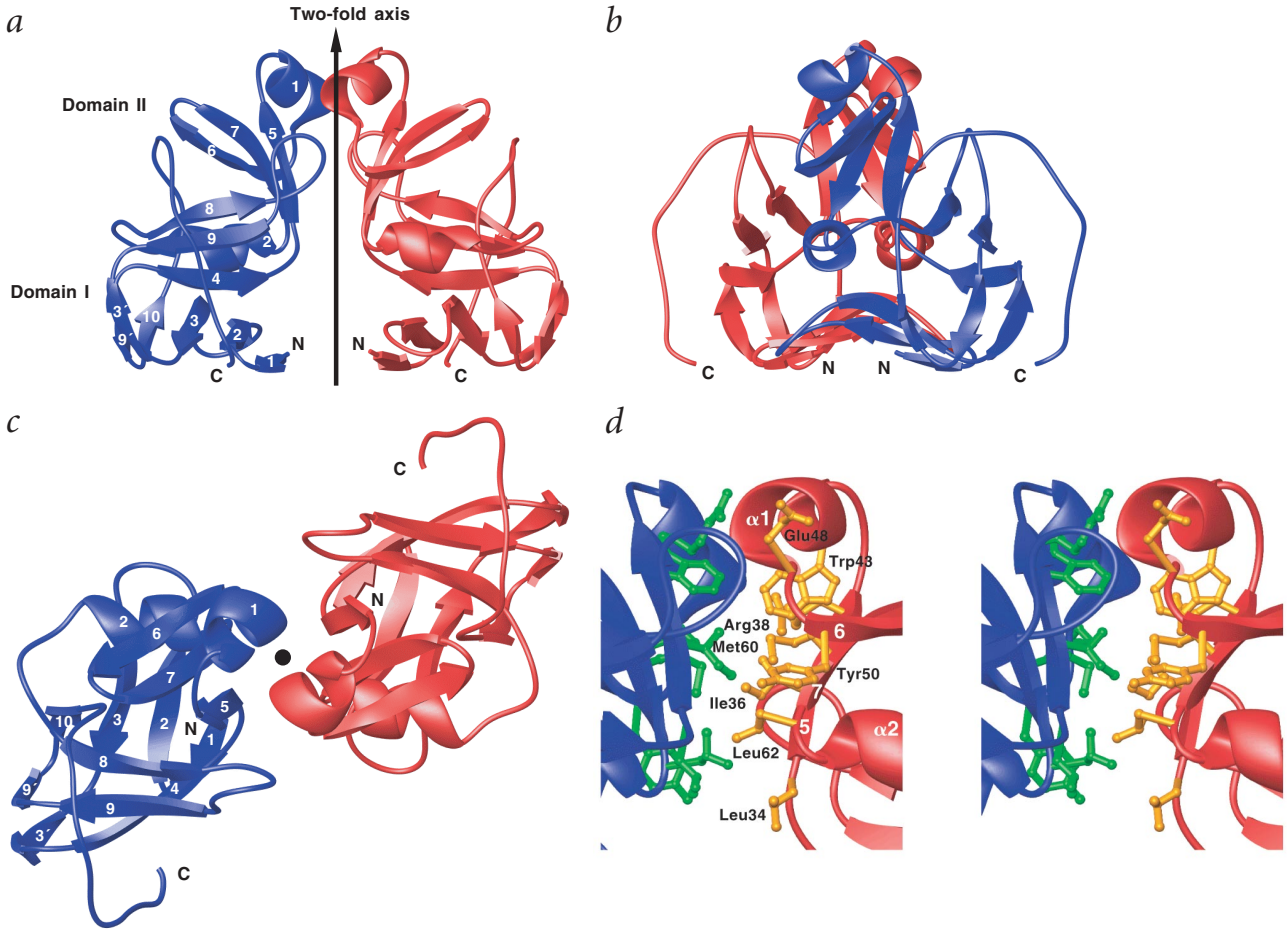

Fig. 4 Three-dimensional structure of the vSET dimer. ***a***, Stereo view (front view) of ribbon depiction of the averaged minimized NMR structure of the vSET dimer. Orientation of the structure is similar to that of Fig. 3*b*. The two adjacent subunits of the dimer are colored in red and blue. Domain I of the antiparallel β-barrel and domain II of the open-faced β-sandwich are indicated. ***b***, Side view of the vSET dimer. ***c***, Top view of the vSET dimer looking down the two-fold axis. ***d***, The dimer interface showing side chain contacts between the adjacent subunits. The side chains are color-coded in orange and green for the subunits in red and blue, respectively.

logy (β5-α1-β6-β7)[36] and is followed by an interdomain-connecting α-helix (α2). Domain II–domain II interactions between the two subunits mediate almost all of the direct subunit–subunit contacts in the dimer (Fig. 4*b*,*c*).

The dimer is knit together in a head-to-head fashion between the two open-faced β-sandwiches of the pairing subunits (Fig. 4*c*). The amphipathic α1 helices of the two subunits are packed against one another in an antiparallel manner with a ~60° crossing angle. Hydrophobic and aromatic side chains of Leu34, Ile36, Trp43, Leu47, Tyr50, Met60 and Leu62 on one subunit form two-fold axis–related interactions with their counterparts on an adjacent subunit (Fig. 4*d*). These intersubunit interactions are complemented by electrostatic interactions formed between Arg38 in β5 on one subunit and Glu48 and Asp49 in the α1–β6 loop on the other (Fig. 4*d*).

The overall subunit structure of the dimer is compact; however, the C-terminal sequence is structurally flexible. Strikingly, the sequence consisting of residues 102–105 that connects the flexible C-terminal segment and the last structural element β10 threads underneath the α2–β8 loop, forming a rather unusual knot-like structure (Fig. 4*a*,*c*). Formation of this knot-like structure is well supported by observation of numerous long-range NOEs in the protein. Particularly, Ile101 and Ile103 in the C-terminal sequence have extensive interactions with Met57, Phe68 and Ile89, which together constitute a hydrophobic core within the β-barrel structure of domain I. The residues that comprise the knot structure reside in the two most conserved sequence motifs in SET domains (see below), suggesting that this unique knot structure is likely conserved in SET domain family.

The active site of vSET HMTase

The conserved SET-domain signature sequence in vSET (69-NHSKDPN-75) is located in a loop connecting α2 and β8 (Fig. 5*a*). Notably, this signature sequence is in spatial proximity to the Gly-Gly-Tyr-Gly sequence (residues 14–17) in the β2–β3

Table 1 Sedimentation analysis of the PBCV-1 SET domain

	Buffer composition[1]	M_W (kDa)[2]	$s_{20,w}$ (S)[3]
A	300 mM urea, 700 mM NaCl	27.3 ± 0.2	2.40 ± 0.04
B	no urea, 300 mM NaCl	26.7 ± 0.3	2.37 ± 0.04
C	no urea, 150 mM NaCl	26.9 ± 0.2	2.36 ± 0.05

[1]The protein sample buffers used for sedimentation experiments consist of 50 mM sodium phosphate pH 6.5, containing 0.1 mM EDTA and 5 mM β-ME with varying amounts of NaCl and urea as indicated.
[2]The protein concentration of vSET used for sedimentation equilibrium experiments was 1.0–1.3 μM.
[3]The protein concentration of vSET used for sedimentation velocity experiments was ~2.0 μM.

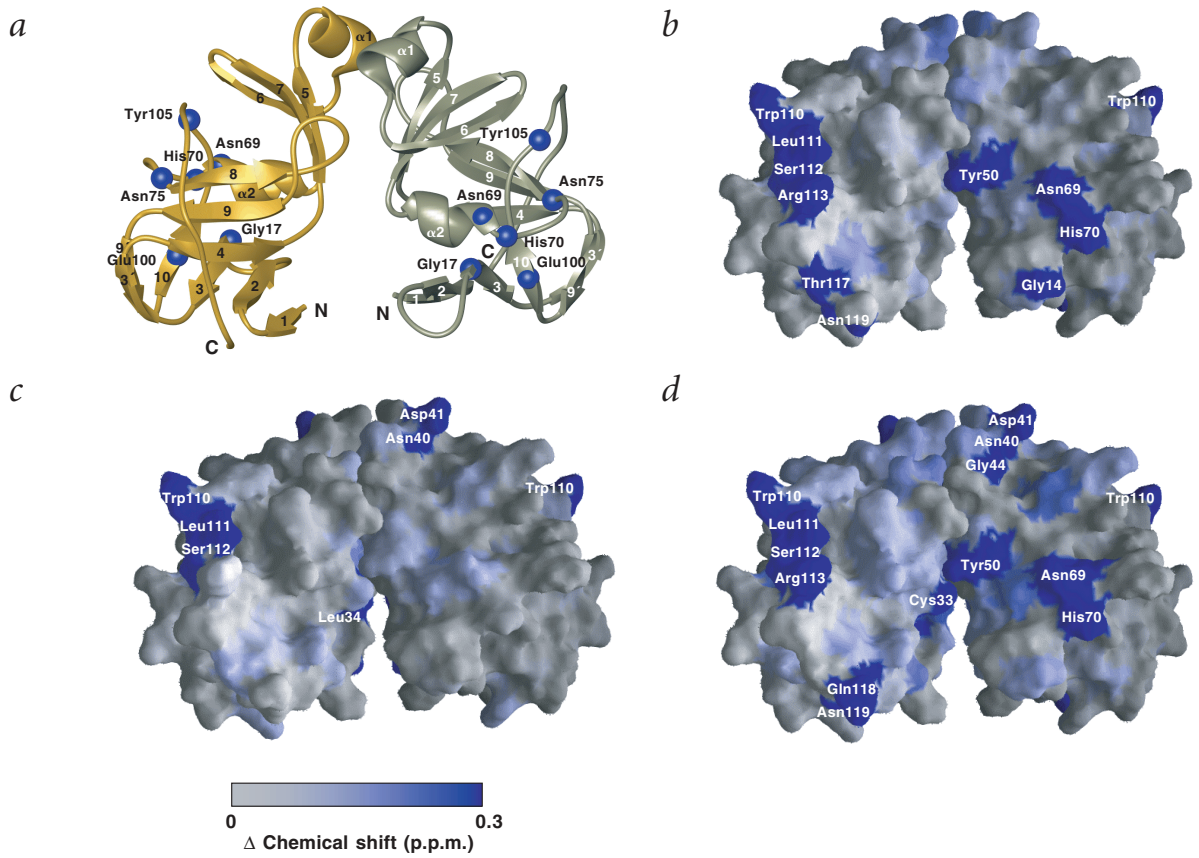

Fig. 5 SAM and substrate recognition of vSET. ***a***, Ribbon diagram of vSET dimer depicting six absolutely conserved residues in SET domains. The blue spheres represent Cα positions of these conserved residues. Molecular surface views of vSET highlighting protein residues that showed major chemical shift perturbations upon binding to ***b***, cofactor SAM; ***c***, H3-K27 peptide substrate (residues 15–30); or ***d***, to both ligands. The protein/ligand molar ratio was kept at 1:10. The protein residues are color-coded according to the extent of chemical shift changes from gray to blue, with residues that showed significant line broadening upon ligand binding colored in blue. The results reflect the sum of the chemical shift changes of the backbone amide ^{1}H and ^{15}N resonances as observed in the ^{15}N-HSQC spectra.

loop (Fig. 5*a*), which is believed to be the conserved SAM-binding motif present in protein methyltransferases[37,38]. Moreover, all six absolutely conserved residues in SET domains — corresponding to Gly17, Asn69, His70, Asn75, Glu100 and Tyr105 in vSET — are clustered on the solvent-exposed side of the unusual knot-like structure at the opening of the β-barrel in each subunit (blue spheres, Fig. 5*a*). Although the molecular basis of the catalysis of SET domains remains to be explained, mutation of the conserved His70 to different amino acids at the corresponding sites in the homologous SET domains almost completely abolished the HMTase activity[21,23,24,26]. Thus, the clustering of these functionally conserved residues suggests that the two putative active sites in the dimer are positioned on the opposite sides of the dimer.

To verify the location of the enzyme active site, we performed NMR titration of vSET with the cofactor SAM and the H3-K27 peptide (residues 15–30) by recording a series of 2D ^{1}H-^{15}N HSQC spectra of the protein as a function of ligand concentration. The NMR data collection was optimized to minimize the breakdown of SAM in aqueous solution and/or enzymatic conversion of the H3-peptide substrate. The residues that showed the most significant spectral changes, which include chemical shift perturbations and line broadening of backbone amides upon binding to SAM, are clustered at the knot-like structure, where the SET-domain signature sequence resides, and in the C-terminal tail (Fig. 5*b*). The residues most perturbed by binding to the H3 peptide are located at the dimer interface, as well as in the C-terminal tail (Fig. 5*c*). These results, which are consistent with spectral changes of vSET upon binding to both SAM and H3 peptide (Fig. 5*d*), confirm the location of the enzyme active site near the signature sequence and reveal that the subunit–subunit interactions are attuned to the enzyme catalysis. These results also argue that the structurally flexible C-terminal segment of vSET is involved in SAM binding and/or substrate recognition as an integral part of the enzyme active site.

Mutagenesis of vSET

To probe the molecular nature of the active site, we performed site-directed mutagenesis of vSET on residues in the enzyme signature sequence, the SAM-binding motif and the C-terminal segment (Fig. 6*a*). To minimize potential structural alterations, the residues selected for mutational analysis are either solvent-exposed or located in the structurally flexible regions of the protein. All of the selected residues were individually mutated to Ala except for Ala66, which was changed to Arg because of high conservation of an Arg at this position in SET domains (Fig. 1). Assessment of mutational effects on the vSET structure by gel filtration chromatography reveals that all these individual mutants, except A66R, behave as a dimer similar to the wild type enzyme (data not shown). The latter mutant seems to exist in both the monomeric (elution volume = 15.2 ml) and dimeric (elution volume = 14.2 ml) states (data not shown), possibly

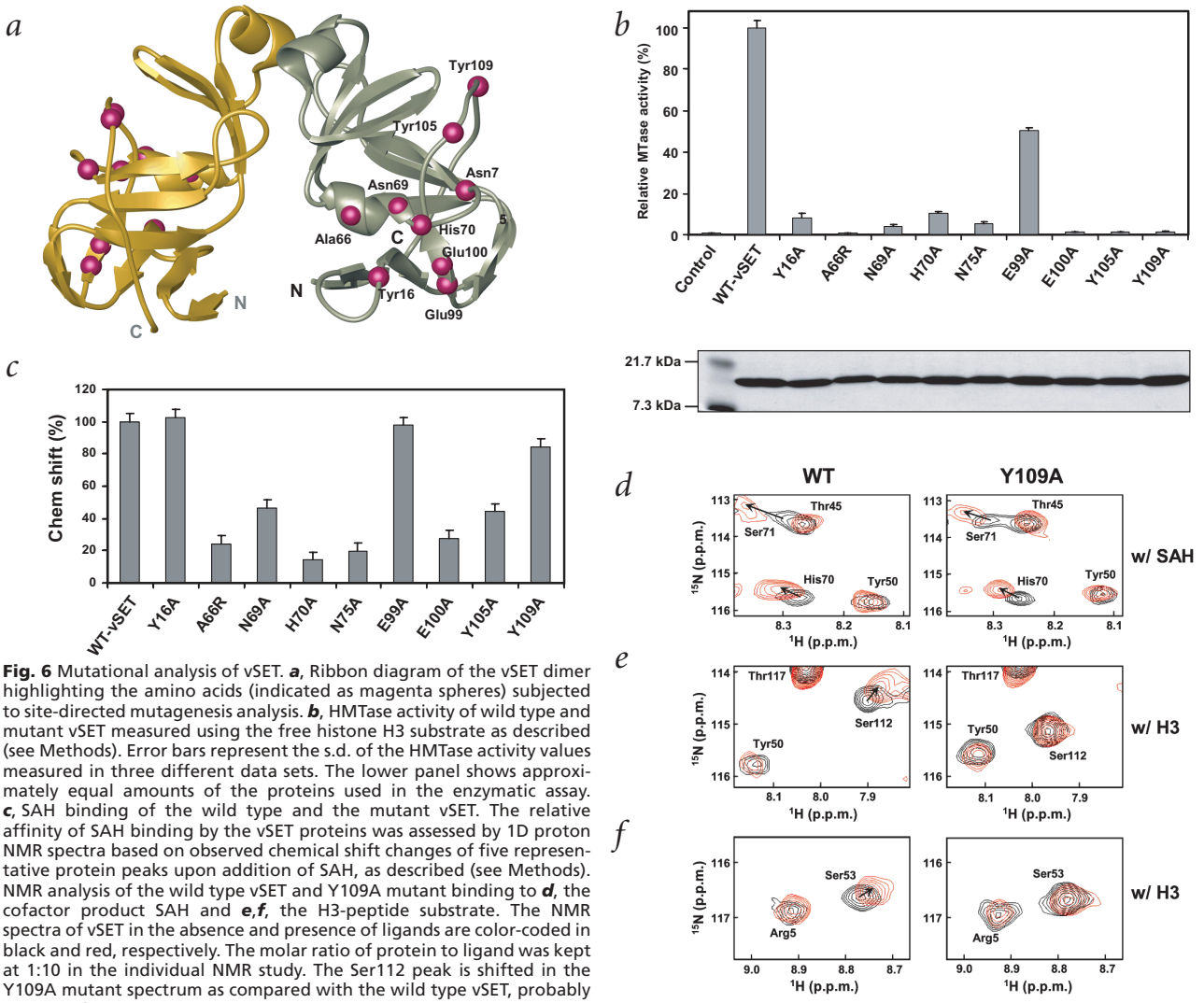

Fig. 6 Mutational analysis of vSET. ***a***, Ribbon diagram of the vSET dimer highlighting the amino acids (indicated as magenta spheres) subjected to site-directed mutagenesis analysis. ***b***, HMTase activity of wild type and mutant vSET measured using the free histone H3 substrate as described (see Methods). Error bars represent the s.d. of the HMTase activity values measured in three different data sets. The lower panel shows approximately equal amounts of the proteins used in the enzymatic assay. ***c***, SAH binding of the wild type and the mutant vSET. The relative affinity of SAH binding by the vSET proteins was assessed by 1D proton NMR spectra based on observed chemical shift changes of five representative protein peaks upon addition of SAH, as described (see Methods). NMR analysis of the wild type vSET and Y109A mutant binding to ***d***, the cofactor product SAH and ***e***,***f***, the H3-peptide substrate. The NMR spectra of vSET in the absence and presence of ligands are color-coded in black and red, respectively. The molar ratio of protein to ligand was kept at 1:10 in the individual NMR study. The Ser112 peak is shifted in the Y109A mutant spectrum as compared with the wild type vSET, probably resulting from the neighboring point mutation.

because this point mutation is a rather nonconservative amino acid substitution and its location is close to the dimer interface.

To examine functional effects of these mutations, we measured the activity of the wild type enzyme and mutants in an *in vitro* MTase assay (see Methods). Although the E99A mutation caused a marked decrease in the MTase activity, mutations of Y16A, A66R, N69A, H70A, N75A, E100A, Y105A or Y109A resulted in an >90% reduction to nearly complete loss of the enzyme activity as compared with the wild type vSET (Fig. 6*b*). To understand the molecular basis of the mutational effects on the enzyme activity, we performed NMR titration to assess binding of these mutant proteins to *S*-adenosyl-L-homocysteine (SAH), a product of SAM after methyl group donation (Fig. 6*c*). The results showed that individual mutations of N69A, H70A, N75A, E100A or Y105A caused two- to five-fold reductions in SAH binding, suggesting that these absolutely conserved residues are indeed located at the active site and play a direct role in coordinating SAM binding and/or catalysis. The A66R mutant also showed >70% reduction in SAH binding, although it is not clear whether this resulted from disruption of vSET dimerization triggered by the point mutation, which would cause nearly complete loss of the enzymatic activity (Fig. 6*b*).

The Y16A mutant, which lost ~90% of the MTase activity, retained full SAH-binding affinity as compared with the wild type enzyme. Because of its proximity to the SAM-binding site, these data suggest that Tyr16 is possibly involved in substrate binding or in catalysis. It is surprising that Y109A mutation effectively abrogated the MTase activity, because Tyr109 is not conserved and is located in the flexible C-terminal tail, and its mutation did not greatly reduce SAH binding (Fig. 6*b*,*c*). To understand the molecular basis of the effect of this mutation on vSET function, we performed more detailed NMR analysis. NMR spectral comparison between the mutant and the wild type proteins shows that this point mutation did not perturb vSET structure (Fig. 6*d*), consistent with the gel filtration chromatography analysis. This mutant protein retained cofactor-binding capability (Fig. 6*d*) but lost the ability to interact with the H3 peptide substrate (Fig. 6*e*,*f*), implying that the loss of its enzymatic activity is due to the loss of its ability to bind substrate. This conclusion agrees with a previous observation that a C-terminal truncation of SET8 (343-IEAHPWL-352) caused nearly complete loss of its HMTase function[26], suggesting that residues in the flexible C-terminal sequence of vSET are responsible for substrate recognition.

Table 2 NMR structural statistics for the PBCV-1 SET domain

Total experimental restraints	2,059	
Total NOE distance restraints	1,919	
Intramolecular	1,888	
Ambiguous	108	
Unambiguous		
Manually assigned	1,398	
ARIA assigned[1]	382	
Intraresidue	852	
Inter-residue		
Sequential $\|i-j\| = 1$	330	
Medium $1 < \|i-j\| \leq 4$	130	
Long Range $\|i-j\| > 4$	468	
Intermolecular		
Manually assigned	31	
Hydrogen bond restraints	78	
Dihedral angle restraints	62	
	Monomer	**Dimer**
Final energies (kcal mol^{-1})[2]		
E_{TOT}	289.4 ± 29.5	578.5 ± 28.8
E_{NOE}	47.2 ± 14.2	86.2 ± 13.4
E_{DIH}	2.0 ± 0.6	3.8 ± 1.2
E_{LJ}	–505.3 ± 23.7	–1,002.6 ± 31.4
Ramachandran plot (%)		
Full molecule[3]		
Most favorable regions	55.8 ± 2.9	53.3 ± 4.4
Additionally allowed regions	37.9 ± 3.7	38.1 ± 3.9
Generously allowed regions	5.2 ± 1.8	6.1 ± 2.1
Disallowed regions	1.1 ± 0.9	2.5 ± 1.3
Secondary structure[4]		
Most favorable regions	89.6 ± 2.8	84.5 ± 3.0
Additionally allowed regions	10.4 ± 2.2	15.5 ± 2.4
Generously allowed regions	0.0 ± 0.0	0.0 ± 0.0
Disallowed regions	0.0 ± 0.0	0.0 ± 0.0
R.m.s. deviations of atomic coordinates (Å)		
Full molecule[2]		
Backbone	0.89 ± 0.13	1.25 ± 0.14
Heavy atoms	1.45 ± 0.13	1.85 ± 0.15
Secondary structure[4]		
Backbone	0.53 ± 0.07	0.96 ± 0.11
Heavy atoms	1.06 ± 0.09	1.51 ± 0.10

[1]See ref. 51.
[2]Based on 20 lowest energy-minimized structures. The Lennard-Jones Potential was not used during any refinement stage. None of these final structures show NOE-derived distance restraint violations >0.5 Å or dihedral angle restraint violations >5°.
[3]Residues 1–101.
[4]Residues 2–3, 6–9, 23–25, 28–31, 42–47, 50–54, 57–61, 64–69, 76–79, 88–91, 94–96 and 99–101.

Interestingly, although the C-terminal residues are generally not conserved in SET domains (Fig. 1), the C-terminal sequence composition between vSET and human G9a[25], which both show selective HMTase activity for Lys27 of H3, is remarkably similar in two ways. (i) Both have two consecutive aromatic residues at positions corresponding to Tyr109 and Trp110 in vSET; and (ii) both have a glycine (Gly106 in vSET) connecting the knot-like structure and the flexible C-terminal sequence (Fig. 3*e*). Notably, the similar pattern of amino acid conservation in the C-terminal region is also seen in the SET domains of human EZH2 and *Drosophila* E(z) (Fig. 1), which have been recently shown to methylate specifically Lys27 in histone H3 in Polycomb group-mediated gene silencing[28,31,32]. These results strongly suggest that the C-terminal sequence including Tyr109 contributes directly to substrate specificity of vSET for Lys27 in histone H3.

Structural comparison to other protein MTases

Sequence alignment of regions of SET domains shows that all SET domains have a two-domain structural architecture (Fig. 1). Although the antiparallel β-barrel domain I of vSET is structurally conserved in the SET domain family, the domain II structure probably varies significantly in different SET domains because of high sequence variations. Moreover, the residues mediating the dimer interface in vSET are not all conserved in the SET domain family, implying that the dimer structure of vSET is possibly present only in a selected number of SET domains. The SET domains from human ALL-1/MLL (a Trithorax homolog) and *Drosophila* Ash1 and Trithorax, as demonstrated by *in vitro* GST pull-down assay, have been reported to homodimerize[39,40]. Interestingly, the majority of the residues in vSET involved in subunit–subunit interactions in the dimer are found to be homologous to those present in the corresponding positions in Ash1 (Fig. 1), suggesting that the structural mechanism used for Ash1 dimerization may be similar to that in vSET. However, because of lack of obvious sequence similarity to vSET, the structural basis for SET domain dimerization of the other two proteins has not been determined.

These structural predictions from our vSET structure study are confirmed by the recently reported three-dimensional structures of SET domains from proteins SET7/9 (refs. 41,42), DIM-5 (ref. 43), Clr4 (refs. 44) and Rubisco LMST[45], which have been published in the literature during the review of this work for publication. Comparison of these newly available SET domain structures reveals the following salient points that further highlight the distinctive features of the vSET structure, which represents a full-length SET domain protein. First, the core SET domain structure from these proteins contains a common two-domain architecture, consisting of a conserved antiparallel β-barrel structure (domain I) and a structurally variable insert (domain II). Second, vSET exists in a unique dimeric structure in solution, whereas all the other SET domain structures are monomeric, agreeing with the divergent sequences in domain II. Third, although vSET forms a dimer, other SET domain proteins contain the pre-SET and post-SET domain motifs, which form an integral part of SET domain structures. Finally, the cofactor-binding site and the catalytic center of these SET domains seem to be constructed on the unusual but conserved knot-like structure.

The vSET structure is characteristically different from the structures of the protein arginine methyltransferases PRMT3 (ref. 37) and yeast HMT1 (ref. 38), which also have a two-domain structure. However, the two domains of the PRMTases consist of different structural folds — the N-terminal domain is composed of a classical α/β Rossmann fold and the C-terminal domain contains a barrel-like structure — which are connected in tandem rather than tethered in a domain insertion fashion as seen in vSET. In addition, the SAM-binding pocket in the PRMTases is formed on an edge of a Rossmann fold capped by two N-terminal α-helices, in sharp contrast to the location of the cofactor-binding site that is constructed on an antiparallel β-barrel fold in vSET. Furthermore, PRMTases form a noncrystallography-related symmetric dimer or possibly a hexamer in solution[37,38], but the dimer is arranged in a head-to-tail orientation that creates a ring-like structure with two-fold symmetry. In this dimer, the SAM-binding site of the enzyme is positioned at the periphery of a central cavity of the ring structure[38].

Conclusions

The three-dimensional structure and substrate specificity of the SET domain from PBCV-1 reported here provides new insights into the structure and molecular basis of catalysis of the SET domain HMTase family. SET domain HMTases differ drastically from protein arginine methyltransferases both in structures of cofactor SAM-binding domains and in configuration of the methyltransferase active sites. The overall two-domain architecture of the core SET domain structure, consisting of a conserved antiparallel β-barrel (domain I) and a structurally variable insert (domain II), is conserved. The structurally flexible C-terminal tail of SET domains is likely directly involved in substrate recognition and constitutes an integral part of the conserved HMTase active site. Sequence variations in different SET domains, particularly in the C-terminal tail and the domain II region, may confer the different substrate specificities at distinct lysine methylation sites in histones. Oligomerization of SET domain structures through domain II–domain II tethering, as seen in vSET, could contribute to proficient lysine methylation of both copies of histone H3 or H4 in nucleosomes. Finally, the structural conservation and highly specific HMTase activity shown by this viral SET domain protein extends our realization of the evolutionary conservation of histone lysine methylation, from virus to human, as a fundamentally important epigenetic control mechanism for gene transcription and silencing.

Methods

Sample preparation. The A612L gene (Swiss-PROT accession number AAC96946) encoding the full-length 119-residue SET domain protein from *P. bursaria* chlorella virus 1 (vSET)[30] was subcloned into pET22b(+) expression vector (Novagen). The construct was expressed in *E. coli* BL21(DE3) cells (Novagen) at 37 °C. To rule out the possibility that the small size of the PBCV-1 vSET resulted from a DNA sequencing error, a 760-nucleotide region encompassing the A612L gene was amplified from PBCV-1 DNA by PCR and re-sequenced (J. Van Etten, pers. comm.), which confirmed that the nucleotide sequence was indeed identical to that originally reported[30]. Uniformly $^{15}N/^{13}C$ and ^{15}N-labeled proteins were prepared by growing bacteria in a minimal medium containing $^{15}NH_4Cl$ with or without $^{13}C_6$-glucose, respectively. A uniformly $^{15}N/^{13}C$-labeled and fractionally deuterated protein sample was prepared by growing the cells in 75% D_2O. The vSET protein was isolated from inclusion bodies and denatured with 6 M guanidine-HCl. Protein refolding was accomplished by step-wise dialysis using a 50 mM HEPES buffer, pH 7.5, containing 300 mM NaCl, 2–10% (v/v) glycerol, 0.1 mM EDTA and 5 mM β-mercapto-ethanol (β-ME). The refolded protein was purified by Source 15S cation exchange chromatography (Amersham) followed by Superose 12/20 gel filtration chromatography (Amersham). NMR samples contained ~0.6 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-ME in H_2O/D_2O (9:1) or D_2O.

NMR spectroscopy. All NMR spectra were acquired at 37 °C on a 600 MHz or 500 MHz Bruker DRX NMR spectrometer. ^1H, ^{13}C and ^{15}N resonances of the protein were assigned with standard 3D deuterium-decoupled triple-resonance spectra of HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB and (H)C(CO)NH-TOCSY[34,46] recorded on a uniformly $^{15}N/^{13}C$-labeled and fractionally deuterated protein. The side chain assignments were completed with 3D HCCH-TOCSY[35] data collected from a uniformly $^{15}N/^{13}C$-labeled protein. NOE-derived distance restraints were obtained from ^{15}N- or ^{13}C-edited 3D NOESY spectra[35]. The intermolecular NOEs used in defining the dimer structure of vSET were detected in ^{13}C-edited (F_1), $^{13}C/^{15}N$-filtered (F_3) 3D NOESY spectra using a $^{13}C/^{15}N$-labeled/nonlabeled protein sample (1:1). φ-angle restraints were determined from $^3J_{HN,H\alpha}$ coupling constants measured in a 3D HNHA-*J* spectrum[35]. Slowly exchanging amide protons were identified from a series of 2D ^{15}N-HSQC spectra recorded after the H_2O buffer was changed to D_2O buffer. The steady-state $\{^1H\}$-^{15}N heteronuclear NOE experiments were performed with 4.5 s recycle delays between scans using water flip-back pulses for minimum water saturation and gradients for coherence selection[47]. All NMR spectra were processed and analyzed by using NMRPipe[48] and NMRView[49].

Structure calculations. Structures of vSET were calculated with a distance geometry-simulated annealing protocol with X-PLOR[50]. Initial structure calculations of the vSET subunit were done with manually assigned NOE-derived distance restraints. Hydrogen-bond distance restraints were added at the late stage of structure calculations for residues with characteristic NOE patterns. The converged structures were then used for the iterative automated assignment of the NOE spectra by ARIA[51] that integrates with X-PLOR for the structure refinement. For the final dimer structure calculation, only one subunit was subjected to experimental distance and angular restraints. Both subunits in the dimer were subjected to symmetry restraints, which were applied in the form of the X-PLOR non-crystallography-related symmetry restraint and distance symmetry restraints as described[52]. The NOE-derived restraints were categorized on the basis of the observed NOE peak intensities. Additionally, 78 hydrogen-bond distance restraints for 36 hydrogen bonds and 62 φ-angle restraints determined from the 3D HNHA-*J* experiment were also used in the calculations. For the final 20 lowest energy NMR structures, no distance or torsional angle restraint was violated by >0.5 Å or >5°, respectively (Table 2).

***In vitro* histone methyltransferase assay.** The *in vitro* histone methyltransferase assay was carried out using a procedure similar to that reported[21]. Briefly, the enzymatic reaction was carried out in the optimized buffer conditions as follows: a 50 μl of 20 mM Tris-HCl buffer, pH 8.0, containing 20 mM KCl, 10 mM $MgCl_2$, 10 mM β-ME, 10 μg of free histone protein substrates (Roche) and 300 nCi of ^{14}C-labeled SAM (Amersham) for 0.5–1 h at 25 °C or 37 °C. Resulting proteins from ^{14}C-labeled SAM the reactions were separated by SDS-PAGE and visualized by Coomassie staining and fluorography. Methylation reaction using histone-peptide substrates was assayed as described[12]. The reversed-phase HPLC analysis of the peptide reaction products was conducted with a HP 1050 instrument (Hewlett-Packard) equipped with a diode array detector. The HPLC analysis was conducted with an ODS-Hypersil column (Keystone Scientific) of 250 × 4.6 mm using a gradient of 5–60% (v/v) acetonitrile/water in the presence of 0.1% trifluoroacetic acid. The flow rate was 1.00 ml min^{-1}, and the signals were detected with λ at 214 nm. The histone peptides were obtained from Bio-Synthesis. Mass spectrometry analysis was performed at the Protein Core Facility at Columbia University.

Site-directed mutagenesis. Mutant proteins were prepared using the QuikChange site-directed mutagenesis kit (Stratagene). The presence of appropriate mutations was confirmed by DNA sequencing.

Ligand titration. Ligand titration experiments were performed by recording a series of 2D ^{15}N-HSQC spectra on uniformly ^{15}N-labeled vSET protein (~0.25 mM) in the presence of different amounts of ligand concentration in the 0–2.0 mM range. The protein sample and the stock solutions of the ligands were all prepared in the same 50 mM phosphate buffer containing 700 mM NaCl, 300 mM urea, 0.1 mM EDTA and 5 mM β-ME, pH 6.5.

Analytical ultracentrifugation. Sedimentation velocity and sedimentation equilibrium runs were performed in a Beckmann XLA analytical ultracentrifuge equipped with absorption optics. Sedimentation coefficients were calculated from sedimentation velocity experiments conducted at rotor speed of 56,000*g* and corrected to standard conditions with water blanks at 20 °C. Sedimentation equilibrium runs were performed at 22,000*g* or 24,000*g*, and the molecular mass of the protein was evaluated using the in-house written program SEGAL. All runs were performed at 20 °C, and the partial specific volume of 0.73 $cm^3\ g^{-1}$ for the protein samples was used. Protein samples were in the NMR buffer (50 mM

sodium phosphate, pH 6.5, containing 700 mM NaCl, 300 mM urea, 0.1 mM EDTA and 5 mM β-ME) or in similar buffers that contain no urea and 300 mM or 150 mM NaCl.

Coordinates. The atomic coordinates of the vSET structure have been deposited in the Protein Data Bank (accession code 1N3J).

Acknowledgments

We thank J. Van Etten for providing the cDNA clone of PBCV-1 SET domain protein, and L. Chen and S. Mutjaba for technical advice. We are also grateful to A. Lustig (Biozentrum, University of Basel, Switzerland) for performing the analytic ultracentrifugation experiments. A.F. is a recipient of a Wellcome Trust Postdoctoral Fellowship. The work was supported by a grant from the National Institutes of Health to M.-M.Z.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 24 September, 2002; accepted 30 December, 2002.

1. Tschiersch, B. *et al*. 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 (1994).
2. Jones, R.S. & Gelbart, W.M. The *Drosophila* Polycomb-group gene *Enhancer of Zeste* contains a region with sequence similarity to Trithorax. *Mol. Cell. Biol.* **13**, 6357–6366 (1993).
3. Stassen, M.J., Bailey, D., Nelson, S., Chinwalla, V. & Harte, P.J. 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 (1995).
4. Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**, 5857–5864 (1998).
5. Jenuwein, T. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266–273 (2001).
6. Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).
7. Kouzarides, T. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **12**, 198–209 (2002).
8. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
9. Turner, B.M. Cellular memory and the histone code. *Cell* **111**, 285–291 (2002).
10. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
11. Bannister, A.J. *et al*. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
12. Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D. & Grewal, S.I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
13. Noma, K.-I., Allis, C.D. & Grewal, S.I. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**, 1150–1155 (2001).
14. Schotta, G. *et al*. Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* **21**, 1121–1131 (2002).
15. Nielsen, S.J. *et al*. Rb targets histone H3 methylation and HP1 to promoters. *Nature* **412**, 561–565 (2001).
16. Strahl, B.D. *et al*. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell Biol.* **22**, 1298–1306 (2002).
17. Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D. & Felsenfeld, G. Correlation between histone lysine methylation and developmental changes at the chicken β-globin locus. *Science* **293**, 2453–2455 (2001).
18. Heard, E. *et al*. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**, 727–738 (2001).
19. Tamaru, H. & Selker, E.U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283 (2001).
20. Taverna, S.D., Coyne, R.C. & Allis, C.D. Methylation of histone H3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell* **110**, 701–711 (2002).
21. Rea, S. *et al*. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
22. Briggs, S.D. *et al*. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* **15**, 3286–3295 (2001).
23. Wang, H. *et al*. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol. Cell* **8**, 1207–1217 (2001).
24. Nishioka, K. *et al*. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev.* **16**, 479–489 (2001).
25. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.* **276**, 25309–25317 (2001).
26. Fang, J. *et al*. Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr. Biol.* **12**, 1086–1099 (2002).
27. Santos-Rosa, H. *et al*. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
28. Czermin, B. *et al*. *Drosophila* Enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
29. Van Etten, J.L. & Meints, R.H. Giant viruses infecting algae. *Annu. Rev. Microbiol.* **53**, 447–494 (1999).
30. Li, Y. *et al*. Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology* **237**, 360–377 (1997).
31. Cao, R. *et al*. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
32. Muller, J. *et al*. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
33. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
34. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158 (1999).
35. Clore, G.M. & Gronenborn, A.M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 249–363 (1994).
36. Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339 (1981).
37. Zhang, X., Zhou, L. & Cheng, X. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J.* **19**, 3509–3519 (2000).
38. Weiss, V.H. *et al*. The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. *Nat. Struct. Biol.* **7**, 1165–1171 (2000).
39. Rozovskaia, T. *et al*. Trithorax and ASH1 interact directly and associate with the trithorax group-responsive bxd region of the *Ultrabithorax* promoter. *Mol. Cell Biol.* **19**, 6441–6447 (1999).
40. Rozovskaia, T. *et al*. Self-association of the SET domains of human ALL-1 and of *Drosophila* TRITHORAX and ASH1 proteins. *Oncogene* **19**, 351–357 (2000).
41. Wilson, J. *et al*. Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111**, 105–115 (2002).
42. Jacobs, S.A. *et al*. The active site of the SET domain is constructed on a knot. *Nat. Struct. Biol.* **9**, 833–838 (2002).
43. Zhang, X. *et al*. Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127 (2002).
44. Min, J., Zhang, X., Cheng, X., Grewal, S.I. & Xu, R.M. Structure of the SET domain histone lysine methyltransferase Clr4. *Nat. Struct. Biol.* **9**, 828–832 (2002).
45. Trievel, R., Beach, B., Dirk, L., Houtz, R. & Hurley, J. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111**, 91–103 (2002).
46. Yamazaki, T., Lee, W., Arrowsmith, C.H., Mahandiram, D.R. & Kay, L.E. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655–11666 (1994).
47. Farrow, N.A. *et al*. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
48. Delaglio, R. *et al*. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
49. Johnson, B.A. & Blevins, R.A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
50. Brunger, A.T. *X-PLOR version 3.1: a system for X-ray crystallography and NMR* (Yale University Press, New Haven; 1993).
51. Nilges, M. & O'Donoghue, S. Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spectrosc.* **32**, 107–139 (1998).
52. Nilges, M. A Calculation strategy for the structure determination of symmetry dimers by ^{1}H NMR. *Proteins* **17**, 297–309 (1993).