Specificity and mechanism of the histone methyltransferase Pr-Set7

Bing Xiao,[1,4] Chun Jing,[1,4] Geoff Kelly,[2,4] Philip A. Walker,[1] Frederick W. Muskett,[2] Thomas A. Frenkiel,[2] Stephen R. Martin,[1] Kavitha Sarma,[3] Danny Reinberg,[3] Steven J. Gamblin,[1,5] and Jonathan R. Wilson[1]

[1]Structural Biology Group, [2]MRC Biomedical NMR Centre, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, United Kingdom; [3]Howard Hughes Medical Institute, Division of Nucleic Acids Enzymology, Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

Methylation of lysine residues of histones is an important epigenetic mark that correlates with functionally distinct regions of chromatin. We present here the crystal structure of a ternary complex of the enzyme Pr-Set7 (also known as Set8) that methylates Lys 20 of histone H4 (H4-K20). We show that the enzyme is exclusively a mono-methylase and is therefore responsible for a signaling role quite distinct from that established by other enzymes that target this histone residue. We provide evidence from NMR for the C-flanking domains of SET proteins becoming ordered upon addition of AdoMet cofactor and develop a model for the catalytic cycle of these enzymes. The crystal structure reveals the basis of the specificity of the enzyme for H4-K20 because a histidine residue within the substrate, close to the target lysine, is required for completion of the active site. We also show how a highly variable component of the SET domain is responsible for many of the enzymes' interactions with its target histone peptide and probably also how this part of the structure ensures that Pr-Set7 is nucleosome specific.

In eukaryotic cells DNA is stored in the form of chromatin, which is repressive to many of the activities that require access to the genetic information. The building unit of chromatin, the nucleosome, consists of an octamer of histone proteins that form a core large enough to accommodate two turns (147 base pairs) of double stranded DNA (Luger et al. 1997; Davey et al. 2002). The N-terminal portions of the four different histone proteins (H2A, H2B, H3, and H4) that make up the nucleosome project from the core structure and interact with specific binding partners. These histone tails are thought to make contacts with the linker DNA between nucleosomes and between nucleosomes themselves, thus participating in the formation of higher-order chromatin structure (Wolffe and Hayes 1999). Post-translational modification of the histone tails appears important in determining whether chromatin adopts a compacted structure and is associated with silenced DNA—heterochromatin—or if it appears as an extended structure and is associated with transcriptionally active DNA—euchromatin (Grewal and Moazed 2003; Vermaak et al.

2003). Post-translational modifications include acetylation, phosphorylation, methylation, and ubiquitination (Jaskelioff and Peterson 2003; Mellone et al. 2003). It has been proposed that different combinations of these modifications represent a code that is interpreted by the binding of various chromatin associated proteins (Strahl and Allis 2000; Jenuwein and Allis 2001). The net effect of the readout of this epigenetic code is at the level of gene transcription (Fischle et al. 2003; Grewal and Moazed 2003; Rice et al. 2003). For example, it has been established in eukaryotes that tri-methylation of histone H3 Lys 9 (H3-K9), by the Suv3.9 enzymes, leads to recruitment of the chromatin-binding protein HP1, which is a marker for condensation into heterochromatin and subsequent gene silencing (Lehnertz et al. 2003). Conversely, methylation at H3-K4 is associated with gene activation (Santos-Rosa et al. 2003).

A family of enzymes containing a conserved domain, called the SET domain, was initially implicated in gene silencing and position effect variegation (Jenuwein et al. 1998) and has subsequently been shown to be responsible for methyl transfer from *S*-adenosylmethionine (AdoMet) to the histone lysine side-chain nitrogen (Nε) (Rea et al. 2000; Roguev et al. 2001). It seems that all but one of the methylated lysine residues present on histones is subject to modification by SET family enzymes (Sawada et al. 2004). The structural and functional prop-

erties of this class of enzymes have been recently reviewed (Breiling and Orlando 2002; Kouzarides 2002; Lachner et al. 2003), but a key biological question remains how members of this family of enzymes achieve specificity for their particular target lysine residue. A second important issue is establishing whether a particular lysine residue is mono-, di-, or tri-methylated and whether these distinct modifications give rise to different physiological consequences.

Mono-methylation of lysine H4-K20 has been shown to be cell-cycle regulated, and its methylation status contributes to chromosome behavior during mitosis and proper cytokinesis (Rice et al. 2002; Karachentsev et al. 2005). Moreover, using antibodies able to discriminate between H4-K20 mono-, di-, and tri-methylation states (Schotta et al. 2004), staining of mouse fibroblasts revealed that tri-methylated H4-K20 was associated with transcriptionally inactive pericentric heterochromatin whereas mono-methylated and di-methylated H4-K20 were found to be broadly distributed but generally associated with euchromatic regions. A number of different SET proteins have been reported as being able to methylate H4-K20 including Pr-Set7 (Nishioka et al. 2002), also separately reported as Set8 (Fang et al. 2002), Suv4-20h1 and Suv4-20h2 (Schotta et al. 2004), and NSD1 (Rayasam et al. 2003). Both Suv4-20h enzymes are capable of tri-methylating H4-K20. Pr-Set7 is highly selective for H4-K20 and has a marked preference for nucleosomal substrate over either histone octamer or H4 polypeptide (Fang et al. 2002; Nishioka et al. 2002). Gene knockout techniques involving deletion of a homolog of the Pr-Set7 (Set8) gene in *Drosophila* have shown that the gene is essential for development (Nishioka et al. 2002). Also in *Drosophila* it has been shown that K20 methylation is associated with transcriptionally silent dense chromatin regions (Rice et al. 2002). We have therefore undertaken the structural and functional characterization of Pr-Set7 to better understand its biological role. We report here the crystal structure of a ternary complex of the enzyme and demonstrate that it adds a single methyl group to H4-K20. We also show that a residue from the peptide substrate contributes to the active site of the enzyme and thus plays an important part in determining the enzyme's specificity, and we suggest how the enzyme interacts with its nucleosome substrate. Finally, we consider how the C-flanking domains of SET proteins are likely to become ordered upon the addition of cofactor and how this influences the catalytic cycle of these enzymes.

Results and Discussion

The overall structure of Pr-Set7

The structure of the ternary complex of Pr-Set7 with a 10-residue peptide based on histone H4 (mono-methylated on the target lysine) and the cofactor product *S*-adenosylhomocysteine (AdoHcy) is shown in Figure 1 in both Ribbons (A) and space-filling representation (B,C).
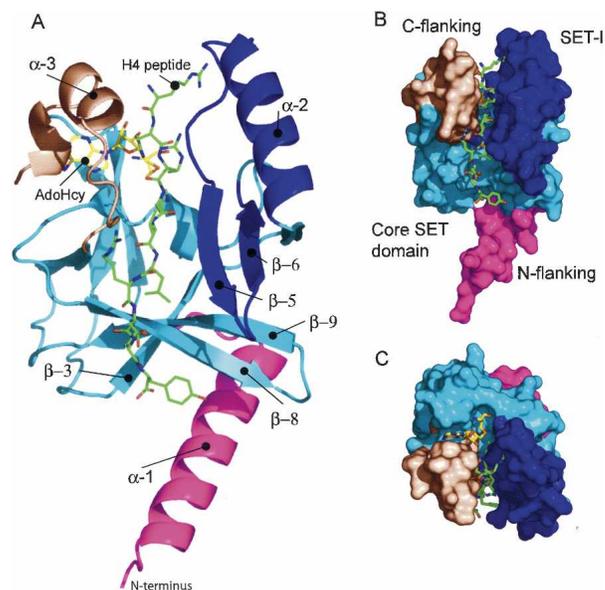

Figure 1. Ternary structure of the histone 4 K20 methyltransferase Pr-Set7. (*A*) A ribbons representation of Pr-Set7 colored according to its subdomains: N-flanking (magenta); core SET domain (cyan); SET-I (blue); C-flanking (beige). The H4 peptide and AdoHcy cofactor products are shown as stick representations with the peptide carbons in green and AdoHcy carbons in yellow. Key secondary structure elements have been labeled. (*B*) Surface representation of Pr-Set7 in the same orientation as *A* showing the relationship of the four subdomains and the peptide-binding groove. (*C*) Surface representation rotated 90° about the *X*-axis to show the AdoHcy-binding pocket.

The X-ray structure was solved using phases derived from a mercury-based multiwavelength anomalous diffraction experiment and refined at 1.5-Å Bragg spacing. Relevant crystallographic statistics are presented in Table 1. The protein construct that proved most suitable for structural studies was N-terminally truncated (and thus contains residues 175–352), but its enzymological properties are essentially the same as full-length construct (Fang et al. 2002).

The overall architecture of the SET domain is largely similar to other family members, but there are considerable variations in the size and conformations of many of the loops. A region of marked structural variation between different family members, and of particular interest, is the insert or Set-I region located approximately in the middle of the SET domain (colored blue in Fig. 1). There is considerable sequence variation within the Set-I region and this part of the domain is responsible for mediating many of the enzyme's interactions with its substrate (Xiao et al. 2003b).

The sequences of the N- and C-flanking domains of Pr-Set7 do not belong to any of the recognized classes of these types of domains found in other members of the SET family. The truncated construct contains just 38 residues from the N-flanking domain (colored magenta in Fig. 1) and these are organized as an α-helix packing against the SET domain. This structural feature has not

Table 1. *Data and refinement statistics for the ternary complex of Pr-Set7, H4 peptide, and AdoHcy*

Space group	P1
Unit cell a = 42.2 Å, b = 46.3 Å, c = 52.0 Å, α = 64.7°, β = 86.7°, γ = 90.6°	
Resolution range (Å)	20.0–1.5
R_{symm} (%)	6.6 (51.5)[a]
I/σ	9.2 (1.8)[a]
Completeness (%)	94.0
Final model	
No. of atoms Pr-Set7, H4 peptide, AdoHcy	1290, 97, 25
Water molecules	612
$R_{working}$ (%)[b]	18.7
R_{free} (%)[c]	20.6
rms deviation bond lengths (Å), angles (°)	0.011, 1.579

[a]The average value across the resolution range; value in parentheses is that for the highest resolution bin (1.55–150 Å).
[b]$R_{working} = \Sigma \, ||F_o| - |F_c||/\Sigma|F_o|$.
[c]$R_{free} = \Sigma_T \, ||F_o| - |F_c||/\Sigma_T|F_o|$, where T is a test data set of 5% of the total reflections randomly chosen and set aside before refinement.

been seen as such before but in all known SET domain structures, packing interactions between the core of the SET domain and elements of the N-flanking domain have been observed (Xiao et al. 2003b). Although the C-flanking domain of Pr-Set7 (colored beige in Fig. 1) is shorter, and bears no obvious sequence similarity to that of Set7/9, it is similarly organized. In both enzymes this region consists of a loop (occupying a similar region of space relative to the core SET domain) terminating in a helical structure. In Pr-Set7 the helix is interrupted by a proline residue and then terminates with a short span of 3_{10} helix. It seems that there are corresponding parallels in the functional roles of this C-flanking region between Pr-Set7 and Set7/9. In particular, both contribute to the completion of the lysine-access channel and both contain a tryptophan residue that makes a packing interaction with the cofactor. Although the C-flanking domain of Dim-5 is structurally divergent from both Set7/9 and PrSet-7, it too appears to have an analogous role in completing the active site (Xiao et al. 2003b; Zhang et al. 2003).

Pr-Set7 is a monomethylase

An intricate network of hydrogen bonds position the side chain of the methylated lysine residue such that the methyl group is directly in line with Nε, to which it is attached, and the sulphur atom of the AdoHcy (Fig. 2A, left panel). Key residues in this network are Tyr 334 and Tyr 245. This arrangement is very reminiscent of the ternary complex of Set7/9 (Fig. 2A, upper right panel). Comparison of the active sites of these two enzymes shows that the size of the pocket, containing the amino group of the lysine, is quite similar in both cases. In contrast the active site of Dim-5, a SET domain enzyme that tri-methylates its target lysine residue (Zhang et al.

2003), is more open and able to accommodate mono-, di-, or tri-methylated lysine (Fig. 2A, lower right panel). This suggested that Pr-Set7, as with Set7/9, is only able to accommodate mono-methylated lysine, and may therefore only be capable of adding a single methyl group to its target lysine. In order to address this notion experimentally two different approaches were undertaken.

First, unmodified H4 30-mer peptide was incubated with Pr-Set7 and the reaction product was isolated by reverse phase HPLC and analyzed by MALDI time-of-flight mass spectrometry (Fig. 2B). The increase in the mass of peptide, after the HKMT reaction had been driven to completion, was consistent with the addition of a single methyl group. Second, recombinant nucleosomes were methylated by Pr-Set7 and the products of the reaction were analyzed with antibodies specific to mono-, di-, and tri-methyl H4-K20. Native nucleosomes served as a positive control for the presence of the modification (Fig. 2C). Although all three modified forms can be detected in the native nucleosome preparation, only the mono-methyl K20 specific antibody cross-reacts with recombinant nucleosome incubated with Pr-Set7. Taken together these results confirm that Pr-Set7 is, indeed, responsible for the addition of a single methyl group to H4-K20 presented either as a peptide or nucleosomal substrate.

Recent studies have demonstrated the existence of additional HKMT enzymes with specificity toward H4-K20: Suv4-20h1 and Suv4-20h2 (Schotta et al. 2004), which have been shown to catalyze the tri-methylation of this residue, and NSD1 (Rayasam et al. 2003) although there are no data as to how many methyl groups this enzyme adds. An alignment of these sequences with Pr-Set7 and other SET domain HMKTs of known structure is shown in Figure 2D. Modeling of the Suv4-h20 and NSD1 sequences onto various SET domain structures (data not shown) suggests that the active sites of these enzymes at the base of the lysine access channel is likely to be more open than that observed for Pr-Set7 or for Set7/9. The sequences of NSD1 and Dim-5 are closely related, they share 50% sequence similarity over the SET domain, and both Tyr 178 and Phe 281 of Dim-5 in particular (Tyr 1702 and Phe 1787, respectively) are conserved in NSD1 (see Fig. 2D). This analysis strongly predicts that NSD1, like Dim-5, will be a tri-methylase. The sequence of the Suv4-20h enzymes are less closely related to a SET protein of known structure, the closest match is again with Dim-5, but in this case there is just 34% sequence similarity (there is 31% similarity between Suv4-20h1 and Pr-Set7). Nonetheless, it appears that Tyr 245 of Pr-Set7, which is conserved in all other SET proteins whose structure is known, is replaced by a leucine residue (Leu 219) in Suv4-20h1. Substitution of Phe $\rightarrow$ Leu would be expected to lead to a more open active site that would correlate with these enzymes being tri-methylases. Similarly, Tyr 334 of Pr-Set7 is replaced by a substantially smaller cysteine residue in Suv4-20h1 (Cys-297), although care must be taken in interpreting the effect of this change because in Set7/9, where the equivalent residue is Val-333 (on the Set7/9
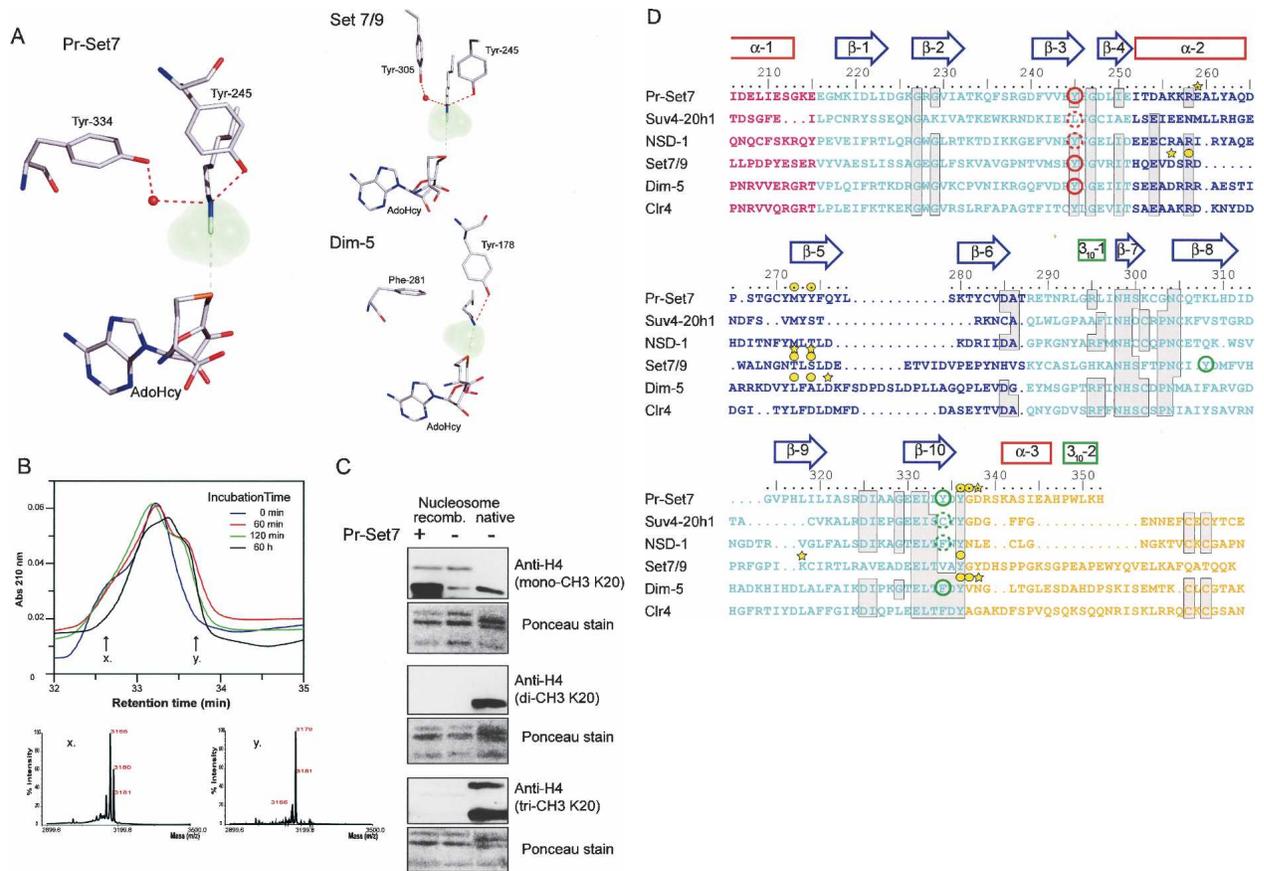
Figure 2. Pr-Set7 is a monomethylase. (*A*) Key residues in the active site of Pr-Set7 in stick representation. The hydrogen bonds that coordinate the lysine Nε atom are represented by dashed red lines. A dashed white line represents a theoretical line between the lysine Nε and the AdoHcy S atom. The transferred methyl group is also shown in a space-filling representation (in green) to give an indication of size and why it is not possible to accommodate extra bulky methyl groups at the position of the hydrogens currently forming H-bonds either with Tyr 245 or the water molecule. The *upper right* panel shows the equivalent view in Set7/9 (PDB 1o9s), also a mono-methylase, and the *lower right* panel shows the equivalent view in Dim-5 (PDB 1peg), a tri-methylase. (*B*) Analytical analysis of the reaction products obtained following incubation of Pr-Set7 with unmodified H4 30-mer peptide. The products were separated by HPLC and fractions analyzed by MALDI time of flight mass spectrometry, two example traces of which are shown *beneath*. (*C*) Western blots obtained after incubating Pr-Set7 with nucleosome substrates and AdoMet. The substrates are reconstituted recombinant nucleosomes (*Xenopus* histone) or derived from Hela cells (human histone) as indicated. In each the *upper* panel has been stained with the indicated antibody with a Ponceau stained control shown *beneath*. (*D*) Sequence alignment of the Pr-Set7 SET domain with two other histone 4 K20 methyltransferase proteins (Suv-20h1 [accession no. AAT00539] and NSD-1 [NP_758859]) and with SET domains of known structure (Set7/9 [NP_085151], Dim-5 [AAL35215], and Clr4 [T43745]). The sequence is colored according to domain. The secondary structure elements of Pr-Set7 are indicated *above*. Green and red circles indicate the Pr-Set7 active site Y-245 and Y-334 equivalent residues (dashed circles for intuited equivalent residues). Yellow circles indicate protein main-chain–peptide interactions; yellow stars indicate protein side-chain–peptide interactions.

β-22 strand), it is a tyrosine coming from the Set7/9 β-20 strand (equivalent to Pr-Set7 β-8) that occupies the equivalent position in the active site.

Substrate binding, specificity, and SET-I

There is well-defined electron density for all 10 residues of the H4 peptide in the ternary complex structure. As can be seen from the surface representation of Pr-Set7 (Fig. 1), the peptide sits in a groove on the surface of the SET domain. The groove is largely defined by the β-5 strand and α-2 helix (both from Set-I) on one side and by

the loop connecting β-10 and α-3 (from the C-flanking domain) on the other. The gross structural features of this binding groove are conserved in Set7/9 and Dim-5. An equally well-defined and conserved pocket, accommodating the AdoHcy cofactor, is also evident on the opposite face of the SET domain (Fig. 1C).

The network of polar interactions that stabilize the peptide substrate complex and that appear to contribute largely to substrate specificity is shown schematically in Figure 3. Many of these interactions are made with the Set-I domain (blue) and somewhat fewer with the C-flanking domain (beige). There are no such interac-
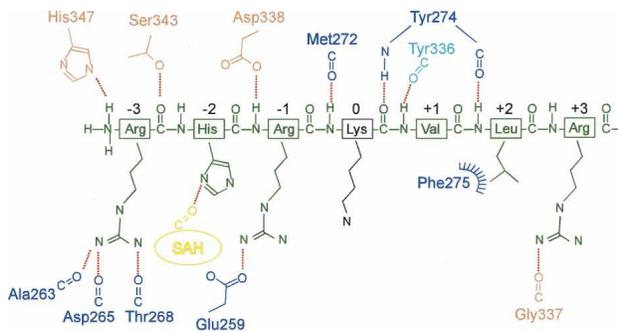
Figure 3. H4 peptide-binding interactions in Pr-Set7. A schematic representation of the interactions observed between Pr-Set7 and its H4 peptide substrate. Hydrogen bonds are shown as a red dotted line. The peptide main-chain interactions are shown *above* the peptide trace (green) and side-chain interactions are shown *below*. The target lysine is shown in black but its interactions are omitted in this figure for clarity. The Pr-Set7 residues involved in interactions are colored according to the scheme used in Figure 1 and protein side-chain or main-chain interactions are indicated accordingly.

tions between the peptide and the core of the SET domain. The structure of Pr-Set7 therefore highlights the role played by these two variable regions of the enzyme in mediating substrate specificity. Although the Set-I is a region of high sequence variability within the SET family (see Fig. 2D; Xiao et al. 2003b), its structural architecture is conserved. It consists of a helix(α-2)–loop-strand(β-5)–loop-strand(β-6), although the lengths of the various elements vary across family members. The β-5 strand that runs approximately anti-parallel to the peptide makes main-chain-to-main-chain hydrogen bonds with it (Pr-Set7 residues 272 and 274) that are conserved in all three ternary complexes now available (Fig. 2D). Another conserved main-chain-to-main-chain hydrogen bond is made from the (+1) position on the peptide with a tyrosine residue located at the boundary of the conserved SET domain and the C-flanking domain (Tyr 336 in Pr-Set7, Tyr 335 in Set7/9, and Tyr 283 in Dim-5). This tyrosine side chain forms one side of the target lysine-access channel. Notably, there are no polar interactions between Pr-Set7 and peptide residues located C-terminal to Arg(+3). This suggests that the determinants of specificity, at least in terms of the residues flanking the target lysine residue, are relatively limited. Similarly we observed that only a few substrate residues, in close proximity to the target lysine, were responsible for the discrimination between lysine residues in ternary complexes of Set7/9 (Xiao et al. 2003a; Chuikov et al. 2004).

Although the lysine-access channel connecting the substrate- and cofactor-binding sites is a common feature of SET enzymes in general, a remarkable feature of the Pr-Set7 structure is that the peptide substrate itself contributes to the structure of the channel (Fig. 4A). Thus the histidine side chain at the (–2) position completes the channel and occupies a position approximately similar to that of protein residues in other SET domain structures (for example, Tyr 337 in Set7/9 and

Trp 308 in Dim-5). The conformation of the side chain of His(–2) is stabilized by a hydrogen bond made with the carboxyl group on the AdoHcy cofactor (Fig. 3).

Given the remarkable structural role of this histidine residue we were interested to assess its role in determining substrate specificity. First, using "wild-type" peptide consisting of the first 29 residues of human histone H4 (plus a C-terminal tyrosine residue added to aid determination of peptide concentration) for activity measurements we obtained Km = 440 µM and Vmax = 260 nmol CH3/min/mg enzyme (Fig. 4B). We then carried out further HKMT assays using peptides with different amino acids substituted at the (–2) position; His → Ala, His → Gln, and His → Tyr (Fig. 4B). The rational for using His → Gln was that this residue could, in principle, make a similar hydrogen bond with the cofactor as histidine. The His → Tyr mutant was tried because there is a tyrosine, supplied by the enzyme, in the Set7/9 ternary complex. For the Ala(–2) and Gln(–2) peptides no HKMT activity was detectable. In the case of Tyr(–2), activity was observed at the higher peptide concentrations but the behavior was not Michaelian. Therefore, in order to determine whether the histidine residue at the (–2) position was exerting its effect through binding, catalysis, or perhaps both, we measured the binding constants for wild-type and mutant peptide. We first tried using isothermal titration calorimetry and fluorescence titrations to monitor binding, but neither system gave suitable signal changes. Subsequently, we used biotin-tagged peptides bound to streptavidin sensor chips to monitor binding of Pr-Set7 using surface plasmon resonance (SPR). Typical sensorgrams (or response curves) for wild-type and mutant peptide at different Pr-Set7 concentrations are shown in Figure 4C. The binding process is clearly concentration dependent, indicative of a bimolecular reaction. A plot of k_{obs} against enzyme concentration is also shown in Figure 4C (inset) for the wild-type peptide that suggests an on rate of ~60 M^{-1} sec^{-1} and, from the intercept, an off rate of ~0.001 sec^{-1}, giving Kd ~ 20 µM. There are no substantial differences in the on or off rates for wild-type and His → Ala mutant peptides. We conclude therefore that the principal effect of the histidine residue at (–2) is on providing a side chain to facilitate the completion of a catalytically competent active site. This conclusion seems compatible with our X-ray structure and probably accounts for why the His → Tyr peptide has some activity. The fact that this histidine residue is required for efficient methyltransferase activity provides an important contribution to the substrate specificity of Pr-Set7.

Role of the N-flanking domain

Although the structure of the part of the N-flanking domain present in our construct is quite different from that seen in other SET proteins, it may well play a similar role. The N-flanking helix of Pr-Set7 packs against β-strands 3, 8, and 9 of the core of the SET domain (Fig. 1A), which is broadly equivalent to the packing of a smaller helix E against β-strands 8, 14, and 15 in Dim-5

A

Substrate Histidine; H(-2)
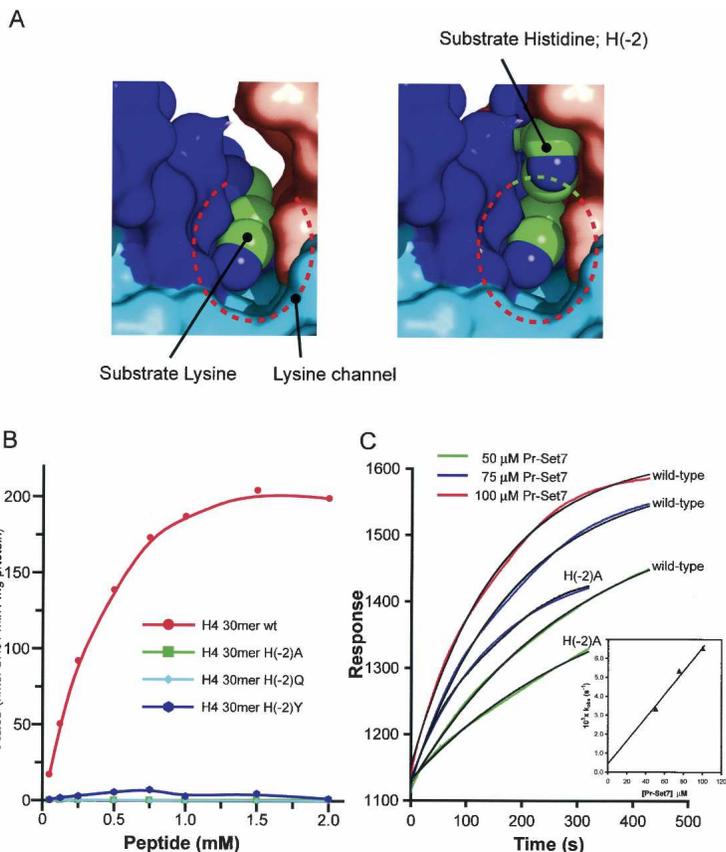
Substrate Lysine   Lysine channel

Figure 4. The H4 peptide substrate contributes to formation of the lysine access channel in Pr-Set7. (*A*) A surface representation of the lysine access channel in Pr-Set7 colored according to the scheme in Figure 1. The substrate H4-K20 side chain is shown as a space-filling representation. The *left* panel is a view without the His(−2) residue and to the *right* the histidine has been added completing the access channel shown by the dashed line. (*B*) Histone methyltransferase assay of Pr-Set7 incubated with increasing concentration of H4 30-mer peptide substrate. Traces are shown for wild-type substrate and peptides with substituted residues at the −2 position relative to the target lysine as indicated. Rates are derived from the incorporation of [^{3}H CH$_3$] after 10 min as measured by liquid scintillation counting of peptide separated from free AdoMet by reverse phase chromatography. (*C*) Plots showing the binding phase from surface plasmon resonance experiments for Pr-Set7 protein injections (as indicated) onto surfaces containing either wild-type or H(−2)A peptide. The *inset* shows a plot of the observed on rate versus Pr-Set7 concentration.

B

C

(Zhang et al. 2003). In Set7/9 the equivalent strands from the core of the SET domain define the center of the contact between it and its N-flanking domain. Certainly in the cases of Pr-Set7 and Set7/9 further truncation of the N-flanking domain results in unstable protein. Thus it seems likely that at least one role of this domain is to provide structural stabilization for the SET domain. Of course this does not preclude there being additional functions of this part of the protein that have not yet been identified.

Cofactor binding and dynamic behavior of the C-flanking domain

The C-flanking domain of Pr-Set7 consists of a loop followed by two short segments of helix (Figs. 1, 5A). Although these structural elements are located differently from the equivalent ones in Set7/9 (Fig. 5A), in both cases the C-flanking domain is responsible for providing a tryptophan residue that makes crucial interactions with the adenine ring of the AdoHcy cofactor. In fact, if the structures of Pr-Set7 and Set7/9 are aligned on the AdoHcy moieties then the tryptophan benzyl rings coincide (Trp 349 of Pr-Set7 and Trp 352 of Set7/9), as shown in Figure 5A. Interestingly, in the nonhistone SET domain methyltransferase Rubisco LSMT an equivalent role is played by residue Phe 302 (Trievel et al. 2003), and in Dim-5 there is related hydrophobic packing of residue

Leu 317 (Zhang et al. 2003). The Pr-Set7 and Set7/9 structures suggest that cofactor binding stabilizes the conformation of their C-flanking domains. To date, ordered C-flanking domains have only been observed for cofactor or ternary complexes of SET proteins (Kwon et al. 2003; Trievel et al. 2003; Xiao et al. 2003a; Zhang et al. 2003). This is not to suggest that binding of cofactor requires interaction with the C-flanking region, because several studies have shown that C-terminally truncated constructs still interact with cofactor (Jacobs et al. 2002; Wilson et al. 2002). We note that two different structural studies of Set7/9 with cofactor gave rise to an ordered C-flanking domain in one case (Kwon et al. 2003) but disordered in another (Jacobs et al. 2002). Interestingly, docking of the C-flanking domain of Set7/9 from our ternary complex onto the latter binary complex shows that this conformation of the C-flanking domain is prohibited by the lattice packing arrangement. It is therefore formally possible that cofactor binding favors the ordering of the C-flanking domain but that this conformation is not so energetically favorable that it cannot be offset by favorable lattice energy (i.e., the lattice energy is more favorable than the ordering of the C-flanking domain). Although NMR has been used previously to monitor cofactor binding to Set7/9 (Jacobs et al. 2002), these studies used a truncated construct that lacked the C-flanking domain; we were therefore interested in using NMR to investigate whether there were changes in the structure of the C-flanking domain induced by cofactor binding.
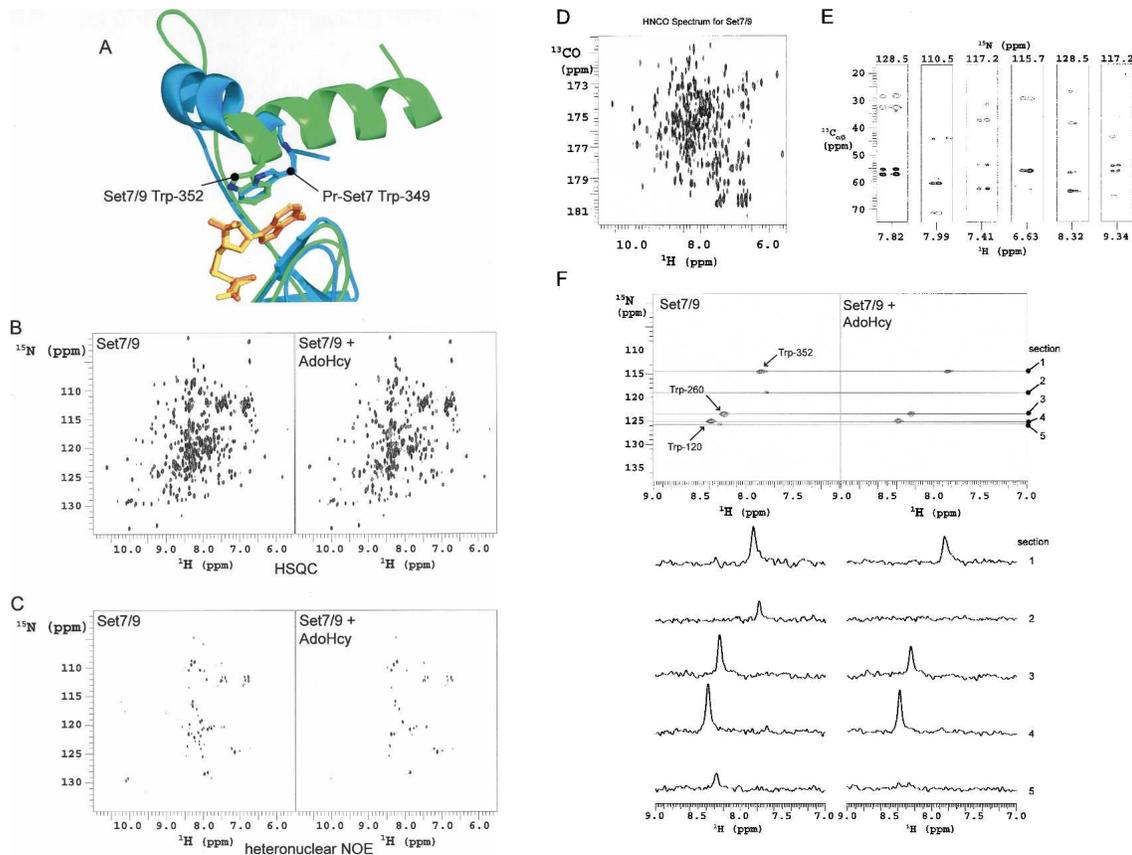
Figure 5. Flexibility of the C-flanking domain. (*A*) A ribbons representation of the structures of the C-flanking domains of Pr-Set7 (blue) and Set7/9 (green) aligned on the AdoHcy cofactor (yellow and orange, respectively). The figure shows that the benzyl rings of Trp 349 of Pr-Set7 and Trp 352 of Set7/9 occupy equivalent positions and make the same interaction with the adenine ring of AdoHcy. (*B*) ^{15}N-HSQC spectra of apo-Set7/9 (*left*) and Set7/9 with 1.2 M equivalents of AdoHcy (*right*). Note there are fewer resonances in the binary spectrum. (*C*) {^{1}H} ^{15}N-heteronuclear NOE data. The plots represent the difference spectra (saturated – 0.8× control) for apo-Set7/9 (*left*) and Set7/9 + AdoHcy (*right*). (*D*) ^{1}H-^{13}CO projection of the 3D trosy-HNCO spectrum used for accurate peak counting. (*E*) Selected sections from the 3D trosyHNCO projection demonstrating the characteristic "doubling" of resonances for some residues. (*F*) ^{15}N-HSQC spectra of selectively Trp-^{15}N-labeled apo-Set7/9 (*left*) and Set7/9 with 1.2 M equivalents of AdoHcy (*right*). The assigned Trp resonances are indicated. *Below* are cross-sections through the spectra at the positions indicated.

We initially prepared isotopically labeled samples of both Pr-Set7 and Set7/9 for this purpose but found that the Set7/9 sample could be brought to a much higher concentration without precipitation and was stable for a much longer time. Subsequent studies were therefore carried out using Set7/9, although the observations are likely to be applicable to the C-flanking domains of SET proteins in general.

The solution behavior of apo-Set7/9 was investigated using ^{1}H{^{15}N} transverse relaxation optimized (TROSY) NMR spectroscopy (Pervushin et al. 1997). Initial inspection of ^{1}H-^{15}N correlation spectra of the Set7/9 construct (108–366 → 264 residues) indicated the presence of a larger number of ^{1}H-^{15}N correlations than the 245 [264 – (18 prolines + 1 N terminus)] expected (Fig. 5B, left panel). Careful peak counting on a 3D trosy-HNCO spectrum revealed at least 270 resonances corresponding to backbone amide protons, 25 more than predicted. Addition of 1.2 M equivalents of AdoHcy (Fig. 5A, right panel) eliminated the discrepancy in the peak count (reducing the number of ^{1}H-^{15}N by at least 35). These observations indicate structural heterogeneity in the protein in the absence of cofactor. Importantly, almost every peak in the spectrum with excess AdoHcy overlays (to within the line-width) a peak in the spectrum of apo-Set7/9. This suggests that some parts of the apo-Set7/9 are able to adopt two distinct conformations and that one of these closely resembles that found in the binary complex of the enzyme with AdoHcy. Heteronuclear ^{1}H{^{15}N}-NOE measurements (Fig. 5C) indicate significant picosecond–nanosecond time scale internal motion for a number of the resonances that disappear on cofactor binding. These data indicate that there is a high degree of internal flexibility in the nonbinary-like component of the apo-Set7/9 structure. As we will go on to describe, this largely unstructured population is probably predominately associated with the C-flanking domain.

These observations provide a structural explanation for previous observations we have reported for two distinct states of histone H3 peptide bound to Set7/9 (Xiao et al. 2003a). At least in terms of the target lysine residue we showed, by specific ^{13}C/^{15}N labeling of this residue

and ^{1}H{^{13}C}-HSQC spectra, that in the Set7/9/peptide complex this residue was largely disordered. However, addition of stoichiometric AdoMet lead to 90% of the population attaining HSQC signals associated with a bound, or restricted motion, form. At that time we interpreted this observation as indicating that the methylated side chain was better ordered than the unmethylated form. However, it now seems more likely that the lysine side chain becomes ordered upon addition of cofactor because the latter promotes ordering of the C-flanking domain and consequently the completion of the lysine access channel that prevents flexibility in the lysine side chain. To further test this hypothesis we undertook the sequential assignment of apo-Set7/9 using ^{2}H,^{13}C,^{15}N-labeled material and TROSY-based triple resonance methods (Fig. 5D). The backbone assignment of the "binary-like" conformation of apo-SET7/9 has progressed satisfactorily and is >70% complete. However, the extent of the assignments around the active site, including the C-flanking region, of the nonbinary-like conformation of the apo-Set7/9 has been restricted by its limited resonance dispersion. Nonetheless, during the sequential assignment procedure, characteristic "doubling" of resonances was observed in this region and was eliminated by AdoHcy binding (Fig. 5E).

As an additional approach toward understanding flexibility in the C-flanking domain we pursued a residue-specific labeling strategy. The Set7/9 construct used for these studies contains just three tryptophan residues, two of which, Trp 260 in the SET domain and Trp 352 at the beginning of the α-helix of the C-flanking domain, are close to the cofactor-binding site. The spectra of the Trp-^{15}N-labeled protein in the absence and presence of excess AdoHcy are shown in Figure 5F. The observation of five correlations in the spectrum of the apo form (left panel), decreasing to three in the AdoHcy complex (right panel) provides further evidence for the presence of two conformations of the enzyme in the vicinity of the active site in the apo form. Moreover, the close similarity between the positions of the peaks in the spectrum of the complex with three of the peaks in the apo spectrum further supports the conclusion that one of the conformations adopted by the apo protein is "binary-like." Attempts to demonstrate direct exchange between these conformations using magnetization transfer techniques (Yamazaki et al. 1995) were unsuccessful, but estimates based on the signal-to-noise ratio of the spectra indicate an upper limit of ~2 sec^{-1} for the interconversion rate. We conclude that the origin of the heterogeneity lies in a slow equilibrium between a species in which the C-flanking domain, and probably other parts of the active site, adopts a conformation analogous to that observed in the cofactor-bound complex, and a second species in which this region is likely to be unstructured (as evinced by its limited resonance dispersion). Heteronuclear ^{15}N{^{1}H}-NOE (Fig. 5C) measurements indeed indicate significant picosecond–nanosecond time scale internal motion for a number of the resonances that disappear on cofactor binding.

Given that cofactor binding is necessary for proper or-

dering of the C-flanking domain, it may then be the case that the target lysine residue does not insert itself through the preformed access channel but, instead, lies in the groove formed by the SET domain and is then fixed in place by the C-flanking domain folding over it once cofactor is bound (Fig. 6). This model seems plausible and has implications for the actions of other SET enzymes that either di- or tri-methylate their target lysine residues and are often referred to as being processive. For the initial methylation to take place the lysine side chain must be deprotonated. As discussed elsewhere (Trievel et al. 2003; Xiao et al. 2003b), it is not apparent from SET structures that there is an active site residue that carries out this role. In the absence of such a residue it has to be assumed that the enzyme relies upon the small, but rapidly exchanging, population of deprotonated target lysines. This idea is consistent with the observed pH dependence of HKMT reactions catalyzed by several different SET enzymes. After the transfer of the first methyl group onto the lysine, this side chain again carries a net positive charge and must be further deprotonated before a subsequent methylation step can occur. Of course, a new AdoMet molecule must also replace the cofactor product AdoHcy as well. Upon dissociation of the AdoHcy we would therefore expect the C-flanking domain to become largely disordered and for the lysine access channel to be partially broken down. In this condition the enzyme complex, with the mono-methylated substrate perhaps still bound (although there are no data to support an ordered binding mechanism), would be able to deprotonate the target lysine to bulk solvent. Binding of the next AdoMet molecule would then drive the reordering of the C-flanking domain to generate a catalytically competent form of the enzyme ready for the next methyl transfer step. Clearly further experimentation will be needed to test this model and see how closely the dynamic features of Set7/9 apply to other SET enzymes in general.
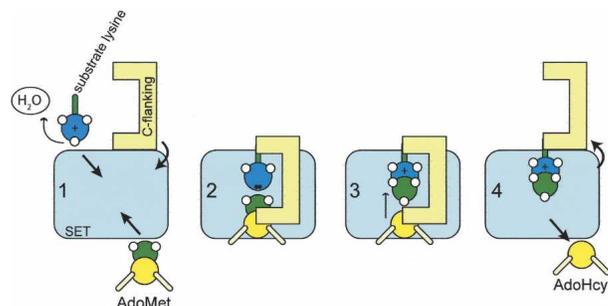

Figure 6. Processivity of SET domain enzymes. Our observations fit the following scheme: (*1*) The lysine loses a proton to the bulk solvent and binds to the methyltranferase. (*2*) Upon binding of the cofactor (AdoMet) the C-flanking domain completes the active site. (*3*) Methyl transfer occurs via a simple nucleophilic substitution reaction. (*4*) The C-flanking domain dissociates from the SET domain allowing release of the cofactor product (AdoHcy). In the case of a monomethylase, for example Pr-Set7, the methylated lysine can now dissociate. For higher-order methyltransferase another cofactor molecule could bind and the cycle repeated.

Pr-Set7 is nucleosome specific: molecular basis for this specificity?

An unusual, and somewhat disturbing, feature of the SET enzymes that have been biochemically characterized is that they show maximal enzyme activity with low or zero salt concentration and limited activity at approximately physiological concentrations of salt (Trievel et al. 2002; Zhang et al. 2002). It is perhaps noteworthy that these reported studies were performed using peptide substrates and not the full-length target molecules that must be the physiologically relevant substrates. Therefore, we first confirmed earlier reports that the preferred substrate for Pr-Set7 is an assembled nucleosome rather than histone H4 or histone octamers, and found this to be the case (Fig. 7A). We then compared the HKMT activity of Pr-Set7 with respect to histone peptide and nucleosome substrates at increasing NaCl concentrations (Fig. 7B). These experiments show that while Pr-Set7 activity against peptide substrate is substantially reduced at approximately physiological salt concentration, under the same conditions nucleosomes are still an effective substrate. We take this as further
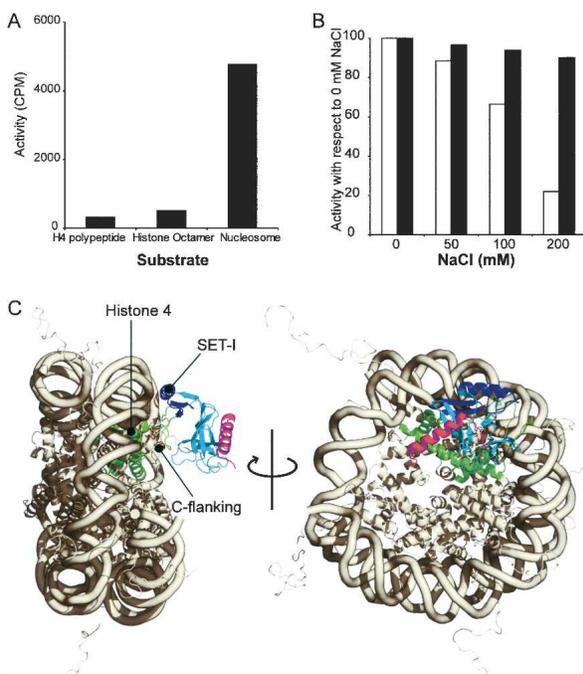

Figure 7. Interaction of Pr-Set7 with the nucleosome. (*A*) Histone methyltransferase assay showing incorporation of [^{3}H CH$_3$] into H4-K20 as peptide, histone octamer, or nucleosome substrate. (*B*) Histone methyltransferase assay following incorporation of [^{3}H CH$_3$] into H4 peptide (white) or nucleosome (black) substrate in the presence of increasing concentrations of NaCl. Each set of data has been normalized to the activity at 0 mM NaCl for that substrate to aid comparison. (*C*) A model showing Pr-Set7 (colored according to the scheme in Fig. 1) docked to the nucleosome (PDB entry 1KX5) and binding to one of the H4 histones (green). To the *right* is an orthogonal view of the model showing how Pr-Set7 may be orientated with respect to the nucleosome.

support for the idea that nucleosomes are the pertinent substrate for Pr-Set7 and that the anomalous behavior of SET enzyme activity in general, with respect to salt concentration, may be a reflection of using peptide mimics of the enzymes' substrates that only partially reflect their behavior. The implication of this notion is that there must be more extensive contacts between the SET enzyme and its physiological substrate than those observed in peptide complexes. We were intrigued therefore to consider how Pr-Set7 might interact with nucleosomes.

In this task we were very fortunate in the fact that K20 of histone H4, being much closer to the core of the nucleosome than other characterized sites of modification, is well ordered in the crystal structures of the nucleosome (Davey et al. 2002). Therefore, most of the H4 based peptide in our ternary complex is represented in the structure of the nucleosome. It is possible then to dock our ternary complex onto the nucleosome structure by superposing the common parts of the H4 peptide. If we make some adjustments to this docked model to avoid steric clashes, the final model shown in two orthogonal views in Figure 7C is obtained. It is a particularly intriguing feature of this model that the principal contact between the enzyme and the nucleosome are mediated by the variable Set-I region (and to a lesser extent the equally variable C-flanking domain) which, as discussed earlier, seems to play the role of a specificity cassette for different SET enzymes. It would certainly seem reasonable that the extended recognition surface of SET enzymes for their physiological substrate would also be mediated by this variable region. Although attempts to test this model directly by generating site-directed mutants in the Set-I region have been thwarted by the instability of the resultant mutated proteins, the restraints on the docking procedure from the crystal structures of Pr-Set7 and the nucleosome mean that the final model ought to be a close approximation to the actual complex formed in vivo.

Materials and methods

Substrate peptides and cofactors

Synthetic peptides were prepared by the Peptide Synthesis Facility, University of Bristol. A 10-mer peptide RHRK(monoMe)VLRDNY (residues 17–25 of human histone H4 with a C-terminal Tyr added for quantification) was used in crystallization of the complex. A 30-mer peptide (derived from residues 1–29 of histone H4 with a C-terminal Tyr added for quantification) SGRGKGGKGLGKGGAKRHRKVLRDNIQGIY or a peptide with either a Y, Q, or A residue substituted for the histidine indicated (underlined) was used to investigate the importance of histidine at the –2 position relative to the target lysine. Unlabeled AdoMet and AdoHcy were obtained from Fluka. Radiolabeled [methyl-^{3}H] AdoMet was obtained from Amersham Biosciences.

Protein crystallization and data collection

Crystals of the ternary complex of Pr-Set7 (175–352) with 10-mer peptide and AdoHcy (ratio 1:2:2) were grown by the vapor

diffusion method at 4°C–10°C. The protein concentration was 12 mg/mL and the crystallization solution was 100 mM PIPES (pH 7.0), 100 mM KSCN with 30% MePEG 2000 as precipitant. The crystals had space group P1, with cell dimensions a = 42.2 Å, b = 46.3 Å, c = 52.0 Å, α = 64.5° β = 86.7° γ = 90.6°, and two molecules per asymmetric unit. The crystals were cryoprotected by stepwise addition of PEG 400 to a final concentration of 5%. The Hg-MAD data were collected at three wavelengths on BM14, ESRF using MAR CCD area detectors at resolution 2.9 Å and processed using DENZO and SCALEPACK (Otwinowski and Minor 1997). The initial phase was calculated from the MAD data using SOLVE and RESOLVE (Terwilliger and Berendzen 1999). The complete model was built on the map calculated using wARP (Collaborative Computational Project, Number 4, 1994) with a high resolution data set (1.5 Å) collected on ID14EH2, ESRF. Refinement used ARP and REFMAC (Collaborative Computational Project, Number 4, 1994) and crystallographic statistic information are shown in Table 1.

NMR spectroscopy

^{2}H,^{13}C,^{15}N-labeled Set7/9 was prepared in 20 mM Tris (pH 7.0), 100 mM NaCl, 0.5 mM TCEP as described previously (Xiao et al. 2003a). Residue-specific labeling of tryptophan residues was achieved using M9 minimal medium supplemented with ^{15}N-labeled tryptophan and 19 unlabeled amino acids. Sequential assignments were accomplished using TROSY-based HNCA, HN(CA)CB, and HN(CA)CO experiments (Yang and Kay 1999) performed on a Varian Inova spectrometer operating at 800 MHz. ^{1}H{^{15}N}-HSQC spectra of specific ^{15}N-Trp labeled samples were recorded on a Bruker Avance 600 MHz spectrometer.

Histone methyltransferase assays

For methyltransferase assays with 30-mer peptides (Fig. 4B); Pr-Set7 (full length) at 1 µM concentration was incubated with peptide in a buffer of 50 mM Tris (pH 8.5), 1 mM EDTA, 1 mM DTT with 250 µM AdoMet and 0.625 µM [methyl-^{3}H] AdoMet for 10 min at 37°C. The reaction was stopped by the addition of 5% acetonitrile and 0.05% trifluoro-acetic acid and acidified with glacial acetic acid. The peptide was separated from unincorporated AdoMet by reverse phase chromatography on Sep-Pak C18 cartridges (Waters). The peptide fraction was eluted in 40% acetonitrile and incorporation of [methyl-^{3}H] measured by scintillation counting. For analytical analysis of reactions products the [methyl-^{3}H] AdoMet was omitted and the reaction products were treated with 8 M urea and separated by reversed phase chromatography using a Zorbax 300SB C18 column (Rockland Technologies) with a gradient of 0%–40% acetonitrile (0.05% TFA) at 55°C. Selected fractions were analyzed by MALDI time-of-flight mass spectrometry (Peptide Synthesis Facility, University of Bristol).

For methyltransferase assays with nucleosome substrate (Fig. 7A,B) Pr-Set7 was incubated with 1 µg of each substrate at 30°C for 5 min in a reaction buffer containing 50 mM Tris-HCl (pH 8.5), 5 mM MgCl$_2$, 4 mM DTT, 1 µM [methyl-^{3}H] AdoMet, and 200 µM AdoMet (Sigma). The total reaction volume was adjusted to 25 µL and the reaction was stopped by the addition of SDS sample buffer. The reaction products were fractionated on a 15% SDS-PAGE and transferred to PVDF membrane. After transfer, the membrane was stained with Coomassie brilliant blue, sprayed with EN^{3}HANCE (NEN), and exposed for ~12 h to Kodak XAR film. Recombinant *Xenopus laevis* histone octamers and nucleosomes were prepared as described previously (Luger et al. 1999). Native nucleosomes were prepared from Hela cells. Western analysis of nucleosomes were carried out as described by Rice et al. (2002). Anti-H4-K20 polyclonal antibodies were purchased from Upstate Biotech.

Surface plasmon resonance

The affinity of Pr-Set7 for "wild-type" H4 and "H(−2)A" peptide was determined using surface plasmon resonance. Two 30-mer peptides (sequence as described above) were synthesized incorporating an additional N-terminal biotin group. These peptides were used to modify two cells of a sensor chip with a streptavidin surface (sensor chip SA Biacore AB). The chip was equilibrated in EP-HBS buffer (at 20 µL/min) and Pr-Set7 was injected at the indicated concentrations and sensograms collected using a Biacore 2000 instrument (Biacore AB).

Coordinates of the X-ray structure of Pr-Set7 have been published in the Protein Data Bank under accession code ****, and NMR assignments for Set7/9 have been published in the BioMagResBank under accession code 2bgz.

References

Breiling, A. and Orlando, V. 2002. SET domain proteins reSET gene expression. *Nat. Struct. Biol.* **9:** 894–896.

Chuikov, S., Kurash, J.K., Wilson, J.R., Xiao, B., Justin, N., Ivanov, G.S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J., et al. 2004. Regulation of p53 activity through lysine methylation. *Nature* **432:** 353–360.

Collaborative Computational Project, Number 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50:** 760–763.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.* **319:** 1097–1113.

Fang, J., Feng, Q., Ketel, C.S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J.A., and Zhang, Y. 2002. Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr. Biol.* **12:** 1086–1099.

Fischle, W., Wang, Y., and Allis, C.D. 2003. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425:** 475–479.

Grewal, S.I., and Moazed, D. 2003. Heterochromatin and epigenetic control of gene expression. *Science* **301:** 798–802.

Jacobs, S.A., Harp, J.M., Devarakonda, S., Kim, Y., Rastinejad, F., and Khorasanizadeh, S. 2002. The active site of the SET domain is constructed on a knot. *Nat. Struct. Biol.* **9:** 833–838.

Jaskelioff, M., and Peterson, C.L. 2003. Chromatin and transcription: Histones continue to make their marks. *Nat. Cell Biol.* **5:** 395–399.

Jenuwein, T., and Allis, C.D. 2001. Translating the histone code. *Science* **293:** 1074–1080.

Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. 1998. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol. Life Sci.* **54:** 80–93.

Karachentsev, D., Sarma, K., Reinberg, D., and Steward, R.

2005. PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes & Dev.* **19:** 431–435.

Kouzarides, T. 2002. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **12:** 198–209.

Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J., and Cho, Y. 2003. Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J.* **22:** 292–303.

Lachner, M., O'Sullivan, R.J., and Jenuwein, T. 2003. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116:** 2117–2124.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13:** 1192–1200.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* **389:** 251–260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. 1999. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304:** 3–19.

Mellone, B.G., Ball, L., Suka, N., Grunstein, M.R., Partridge, J.F., and Allshire, R.C. 2003. Centromere silencing and function in fission yeast is governed by the amino terminus of histone H3. *Curr. Biol.* **13:** 1748–1757.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., et al. 2002. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9:** 1201–1213.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. In *Methods in enzymology* (ed. J.C.W. Carter and R.M. Sweet), pp. 307–326. Academic Press, New York.

Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94:** 12366–12371.

Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. 2003. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22:** 3153–3163.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406:** 593–599.

Rice, J.C., Nishioka, K., Sarma, K., Steward, R., Reinberg, D., and Allis, C.D. 2002. Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes & Dev.* **16:** 2225–2230.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. 2003. Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol. Cell* **12:** 1591–1598.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. 2001. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* **20:** 7137–7148.

Santos-Rosa, H., Schneider, R., Bernstein, B.E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S.L., Mellor, J., and Kouzarides, T. 2003. Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol. Cell* **12:** 1325–1332.

Sawada, K., Yang, Z., Horton, J.R., Collins, R.E., Zhang, X., and Cheng, X. 2004. Structure of the conserved core of the yeast Dot1p, a nucleosomal histone H3 lysine79 methyltransferase. *J. Biol. Chem.* **279:** 43296–43306.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Dev.* **18:** 1251–1262.

Strahl, B.D., and Allis, C.D. 2000. The language of covalent histone modifications. *Nature* **403:** 41–45.

Terwilliger, T.C., and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55 (Pt 4):** 849–861.

Trievel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L., and Hurley, J.H. 2002. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111:** 91–103.

Trievel, R.C., Flynn, E.M., Houtz, R.L., and Hurley, J.H. 2003. Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. *Nat. Struct. Biol.* **10:** 545–552.

Vermaak, D., Ahmad, K., and Henikoff, S. 2003. Maintenance of chromatin states: An open-and-shut case. *Curr. Opin. Cell Biol.* **15:** 266–274.

Wilson, J.R., Jing, C., Walker, P.A., Martin, S.R., Howell, S.A., Blackburn, G.M., Gamblin, S.J., and Xiao, B. 2002. Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111:** 105–115.

Wolffe, A.P., and Hayes, J.J. 1999. Chromatin disruption and modification. *Nucleic Acids Res.* **27:** 711–720.

Xiao, B., Jing, C., Wilson, J.R., Walker, P.A., Vasisht, N., Kelly, G., Howell, S., Taylor, I.A., Blackburn, G.M., and Gamblin, S.J. 2003a. Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421:** 652–656.

Xiao, B., Wilson, J.R., and Gamblin, S.J. 2003b. SET domains and histone methylation. *Curr. Opin. Struct. Biol.* **13:** 699–705.

Yamazaki, T., Pascal, S.M., Singer, A.U., Formankay, J.D., and Kay, L.E. 1995. Nmr pulse schemes for the sequence-specific assignment of arginine guanidino N-15 and H-1 chemical-shifts in proteins. *J. Am. Chem. Soc.* **117:** 3556–3564.

Yang, D.W., and Kay, L.E. 1999. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *J. Am. Chem. Soc.* **121:** 2571–2575.

Zhang, X., Tamaru, H., Khan, S.I., Horton, J.R., Keefe, L.J., Selker, E.U., and Cheng, X. 2002. Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111:** 117–127.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. 2003. Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell* **12:** 177–185.