The SET-domain protein superfamily: protein lysine methyltransferases
Shane C Dillon*, Xing Zhang†, Raymond C Trievel‡ and Xiaodong Cheng†

Addresses: *Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. †Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA. ‡Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI 48109-0606, USA.

Correspondence: Shane C Dillon. E-mail: scd@sanger.ac.uk

Published: 2 August 2005

Genome Biology 2005, 6:227 (doi:10.1186/gb-2005-6-8-227)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/8/227

© 2005 BioMed Central Ltd

Abstract

The SET-domain protein methyltransferase superfamily includes all but one of the proteins known to methylate histones on lysine. Histone methylation is important in the regulation of chromatin and gene expression.

Gene organization and evolutionary history

Nucleosomes, the main components of chromatin, consist of histones, and histone proteins have positively charged amino-terminal tails that are exposed on the outside of nucleosomes. These tails are subject to several post-translational covalent modifications, including acetylation, phosphorylation, ubiquitination, sumoylation and methylation (reviewed in [1]). Methylation been found on a range of lysine residues in various histones: K4 (using the single-letter amino-acid code for lysine), K9, K27, K36 and K79 in histone H3, K20 in histone H4, K59 in the globular domain of histone H4 [2] and K26 of histone H1B [3]. Several proteins responsible for the methylation of specific residues have been characterized, and all but one of these contains a SET domain; they make up the SET-domain protein methyltransferase family (Table 1). The exception to the rule is the DOT1 family, members of which methylate K79 in the globular region of histone H3 and which are structurally not related to SET-domain proteins [4-6]. Recent work suggests that SET-domain-containing proteins methylate a few proteins in addition to histones (see later); they should therefore be named protein lysine methyltransferases rather than histone lysine methyltransferases. The function of SET-domain proteins is to transfer a methyl group from S-adenosyl-L-methionine (AdoMet) to the amino group of a lysine residue on the histone or other protein, leaving a methylated lysine residue and the cofactor byproduct S-adenosyl-L-homocysteine (AdoHcy). Methylation of specific histone lysine residues serves as a post-translational epigenetic modification that controls the expression of genes by serving as ‘markers’ for the recruitment of particular complexes that direct the organization of chromatin.

The SET domain (Figure 1) was first recognized as a conserved sequence in three Drosophila melanogaster proteins: a modifier of position-effect variegation, Suppressor of variegation 3-9 (Su(var)3-9) [7], the Polycomb-group chromatin regulator Enhancer of zeste (E(z)) [8], and the trithorax-group chromatin regulator trithorax (Trx) [9]. The domain, which is approximately 130 amino acids long, was characterized in 1998 [10] and SET-domain proteins have now been found in all eukaryotic organisms studied. There are currently 157 entries for human SET-domain proteins in the SMART database [11] and 93 entries in the Pfam database [12], although both databases contain duplicate entries. Seven main families of SET-domain proteins are known - the SUV39, SET1, SET2, EZ, RIZ, SMYD, and SUV4-20 families - as well as a few orphan members (see Table 2 for a list of the members of each family in humans and their properties). Proteins within each family have similar sequence motifs surrounding the SET domain, and they often also share a higher level of similarity in the SET domain.
The SUV39 family has been characterized in most detail. Members of this family - human SUV39H1, murine SUV39H2, and *Schizosaccharomyces pombe* Cryptic loci regulator 4 (CLR4) - were the first SET-domain protein lysine methyltransferases to be characterized, following the discovery of sequence homology between their SET domains [13]. These proteins, with other members of the family such as D. melanogaster Su(var)3-9, specifically methylate lysine 9 of histone H3 (H3 K9) [13]. Human SUV39H1 and its closely related paralog, SUV39H2, are 55% identical at the amino-acid level. The structures of the genes encoding the two proteins are shown in Figure 2a,b: both have six exons, and they share identical intron-exon junctions. It appears that the members of the family are involved in both euchromatin and heterochromatin, but another member of the same family, G9a, is the predominant histone H3 K9 methyltransferase in mammalian euchromatin [14]. There are two isoforms of G9a in the mouse: the short form (GenBank accession number NP_671493) corresponds to human G9a and the long form (NP_665829), which lacks intron one, has additional Arg-Gly repeats at the amino terminus. No human expressed sequence tag (EST) corresponding to the long form of G9a has yet been isolated, although the sequence is present in the genome. Similar to the situation with SUV39H1, G9a also has a closely related paralog in mammals, G9a-like-protein-1 (GLP1). The human G9a gene has 28 exons and is

| Histone lysine | Function(s) | Histone lysine methyltransferases* |
|---------------|-------------|-----------------------------------|
| H1 K26        | Transcriptional silencing | *Hs EZH2* (catalytic subunit of Polycomb repressive complex 3) [3] |
| H3 K4         | Transcriptional activation | *Dm Trx; Hs MLL1 (ALL-1, HRX), MLL2 (ALR-1), and MLL3 (HALR)* |
|               | Transcriptional activation and elongation | *Hs SET1; Sc SET1* |
|               | Transcriptional activation | *Hs SET7/9* |
|               | Transcriptional activation (in conjunction with ASH1-mediated methylation of H3 K9 and H4 K20) | *Dm ASH1* |
| H3 K9         | Heterochromatic and euchromatic silencing; DNA methylation | *Dm Su(var)3-9; Hs and Mm SUVAR39H1 and UVAR39H2; Sp CLR4* |
|               | Euchromatic silencing; DNA methylation | *Hs and Mm G9a; Hs GLP1 (EuHMT1)* |
|               | Euchromatic silencing | *Hs and Mm ESET (SETDB1)* |
|               | Heterochromatic silencing; DNA methylation | *Nc DIM-5* |
|               | Heterochromatic silencing; DNA methylation | *At KRYPTONITE* |
|               | Transcriptional activation (in conjunction with ASH1-mediated methylation of H3 K4 and H4 K20) | *Dm ASH1* |
| H3 K27        | Euchromatic silencing | *Dm E(z); Hs EZH1 and EZH2 (catalytic subunit of Polycomb repressive complex 2)* |
| H3 K36        | Transcriptional elongation and silencing | *Sc SET2* |
|               | Transcriptional regulation | *Mm NSD1* |
| H3 K79        | Demarcation of euchromatin | *Sc and Hs DOT1 (a non-SET domain histone lysine methyltransferase)* |
| H4 K20        | Cell cycle-dependent silencing, mitosis and cytokinesis [52,53] | *Hs and Dm SET8* |
|               | Heterochromatic silencing | *Hs and Mm SET8* |
|               | Transcriptional regulation | *Mm NSD1* |
|               | Transcriptional activation (in conjunction with ASH1-mediated methylation of H3 K4 and H3 K9) | *Dm ASH1* |
|               | Recruitment of checkpoint protein Crb2 to sites of DNA damage | *Sp SET9* [49] |

Histone lysine methylation sites, functions, and associated histone lysine methyltransferases, which are listed according to the lysine that they methylate.

*Species abbreviations: At, Arabidopsis thaliana; Dm, Drosophila melanogaster; Hs, Homo sapiens; Mm, Mus musculus; Nc, Neurospora crassa; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe. Adapted from [54-56]; additional references listed in the table are those not cited in these reviews.*
yet been annotated as such. The zebrafish GLP ortholog amino-terminal third of the protein. The interacting salt bridge, and a F/Y switch controlling whether the product is a mono-, di- or tri-methylated histone [57] are indicated.

The involvement of invariant residues in binding to AdoMet and the target lysine, catalysis, the structural pseudoknot (see Figure 3), an intra-molecular interacting salt bridge, and a F/Y switch controlling whether the product is a mono-, di- or tri-methylated histone [57] are indicated.

Figure 1

A protein sequence alignment of the SET domains of several representative histone lysine methyltransferases (HKMT) grouped according to their histone-lysine specificity. All sequences are human with the exceptions of Saccharomices cerevisiae SET1 and SET2, Schizosaccharomices pombe CLR4, and Neurospora DIM-5. See Table 2 for the family designations of each human protein shown. The alignment between SET7/9 and DIM-5 is based on their structures [16]. The white text on a black background denotes invariant residues; black text on a gray background indicates conserved residues.

About 17.3 kilobases (kb) long (Figure 2c). GLP1 is 45% identical to G9a and most of the divergence is in the amino-terminal third of the protein. The GLP1 gene has 25 exons - it lacks homologs of the first three introns of G9a - and the 20 exons from the 3’ end have identical junctions to those found in G9a. The GLP gene is quite large, 120 kb in human and 92 kb in mouse, with introns as long as 16 kb (Figure 2d). No obvious orthologs of G9a or GLP can be found in the worm, frog or yeast genomes; in the D. melanogaster genome there is one gene (CAE49087) is 45% identical to human GLP, and the gene shares all but three of its 23 intron-exon junctions with human GLP1.

G9a and SUV39H1 both belong to the same family of SET-domain proteins and both have pre-SET and post-SET domains surrounding the SET domain, but they do not share any intron-exon junctions, even though a number of these junctions occur within the highly conserved SET domain. The other two SUV39 family proteins, ESET (also called SETDB1) and CLLL8 (SETDB2) also have significant similarities in their genomic structures with each other but not with G9a or SUV39H1 (data not shown). Several proteins in other SET families are also found in closely related pairs: EZH1 and EZH2 (members of the EZ family), MLL1 (also called HRX1) and MLL2 (HRX2, both members of the SET1 family), SET1 and SET1L (SET1 family), NSD2 (WHSC1) and NSD3 (WHSC1L; SET2 family), and SUV4-20H1 and SUV4-20H2 (SUV4-20 family).
## Table 2

Properties of some human SET-domain proteins

| Chromosomal location | Gene size (kb) | Number of coding exons | Protein size (amino acids) | Pre-SET (9 Cys, 3 Zn), post-SET (CXCX\(_2\)C) | Domains common to the family in addition to the SET domain | Domains unique to particular members | GenBank accession number |
|----------------------|---------------|-------------------------|---------------------------|---------------------------------------------|-----------------------------------------------|-------------------------------------|-------------------------|
| **SUV39 family**     |               |                         |                           |                                             |                                               |                                     |                         |
| SUV39H1              | Xp11.23       | 12.4                    | 6                         | 412                                         | 4 Cys, chromo                                 | 4507321                             |                         |
| SUV39H2              | 10p13         | 24                      | 6                         | 477 (Mm)*                                   | 4 Cys, chromo                                 | 9956936                             |                         |
| G9a                  | 6p21.33       | 17.3                    | 28                        | 1,210                                       | E/KR-rich, NRSF-binding, ankyrin repeats       | 18375637                            |                         |
| GLP1 (EuHMT1)        | 9q34.3        | 120                     | 25                        | 1,267                                       | Same as G9a                                   | 20372683                            |                         |
| ESET (SETDB1)        | 1q21.2        | 37                      | 21                        | 1,291                                       | Tudor, MBD                                    | 505110                              |                         |
| CLLL8 (SETDB2)       | 13q14.2       | 40                      | 14                        | 719                                         | MBD                                           | 13994282                            |                         |
| **SET1 family**      |               |                         |                           |                                             |                                               |                                     |                         |
| MLL1 (HRX, ALL1)     | 11q23.3       | 86                      | 36                        | 3,969                                       | AT hook, Bromo                                 | 1170364                             |                         |
| HRX2 (MLL4)          | 19q13.12      | 20                      | 37                        | 2,715                                       | Same as above                                 | 12643900                            |                         |
| ALR (MLL2)           | 12q13.12      | 34                      | 54                        | 5,262                                       | PHD, ring finger                              | 2358285                             |                         |
| MLL3                 | 7q36.1        | 299                     | 58                        | 4,911                                       | PHD, ring finger                              | 21427632                            |                         |
| SET1 (ASH2)          | 16p11.2       | 26                      | 18                        | 1,707                                       | RRM, poly-S/E/P                               | 6683126                             |                         |
| SET1L                | 12q24.31      | 14                      | 11                        | 1,092 (Mm)*                                 | RRM, poly-S/E/P                               | 23468263                            |                         |
| **SET2 family**      |               |                         |                           |                                             |                                               |                                     |                         |
| WHSC1 (NSD2)         | 4q16.3        | 79                      | 21                        | 1,365                                       | PWWP, PHD, HMG, ring finger                   | 6683809                             |                         |
| WHSC1 (NSD3)         | 8p12          | 73                      | 23                        | 1,437                                       | PWWP, PHD, ring finger                        | 13699811                            |                         |
| NSD1                 | 5q35.3        | 160                     | 23                        | 2,696                                       | PWWP, PHD, ring finger                        | 19923586                            |                         |
| HIF1 (HYPB)          | 3p21.31       | 106                     | 19                        | 2,061                                       | WWV                                            | 12697196                            |                         |
| ASH1                 | 1q22          | 184                     | 27                        | 2,969                                       | AT hook, bromo, BAH, PHD                      | 7739725                             |                         |
| **RIZ family**       |               |                         |                           |                                             |                                               |                                     |                         |
| RIZ (PRDM2)          | 1p36.21       | 86                      | 9                         | 1,719                                       | C2H2 zinc finger                              | 9955379                             |                         |
| BLIMP1 (PRDM1)       | 6q21          | 19                      | 6                         | 789                                         | C2H2 zinc finger                              | 3493158                             |                         |
| **SMYD family**      |               |                         |                           |                                             |                                               |                                     |                         |
| SMYD3                | 1q44          | 758                     | 12                        | 428                                         | ZI-MYND                                        | 30913569                            |                         |
| SMYD1                | 2p11.2        | 43                      | 9                         | 490                                         | ZI-MYND                                        | 38093643                            |                         |
| **EZ family**        |               |                         |                           |                                             |                                               |                                     |                         |
| EZH1                 | 17q21.2       | 26                      | 19                        | 747                                         | Pre-SET (~15 Cys)                              | 3334182                             |                         |
| EZH2                 | 7q36.1        | 40                      | 19                        | 746                                         | 2 SANT                                         | 3334180                             |                         |
| **SUV4-20 family**   |               |                         |                           |                                             |                                               |                                     |                         |
| SUV4-20H1            | 11q13.2       | 57                      | 9                         | 876                                         | Post-SET (CXCX\(_2\)C)                        | 50659081                            |                         |
| SUV4-20H2            | 19q13.42      | 8                       | 8                         | 462                                         | Post-SET (CXCX\(_2\)C)                        | 31543168                            |                         |
| **Others**           |               |                         |                           |                                             |                                               |                                     |                         |
| SET7/9               | 4q31.1        | 45                      | 8                         | 366                                         | MORN                                          | 25091213                             |                         |
| SET8 (PR-SET7)       | 12q24.31      | 26                      | 8                         | 393                                         |                                               | 25091219                             |                         |

The seven families of SET-domain proteins are classified according to the sequences surrounding their SET domain. *Complete human SUV39H2 and SET1L cDNAs are not available in current databases, but partial cDNA and genomic sequences corresponding to the mouse sequences (Mm) are present. For the pre-SET and post-SET domains, the number and (if known) the arrangement of cysteines in the domain is given. Domain abbreviations and definitions: ankyrin repeats, tandemly repeated modules of about 33 amino acids; AT hook, DNA binding motif with a preference for A/T-rich regions; BAH, Bromo adjacent homology domain; bromo, bromodomain, which can interact specifically with acetylated lysines; chromo, chromatin organization modifier domain; CXXC, domain with two cysteines separated by two amino acids; E/KR-rich, glutamine- or lysine/arginine-rich domains; HMG, high mobility group domain; MBD, methyl-binding domain; MORN, membrane occupation and recognition nexus repeat; NRSF-binding, binds neuron-repressive silencing factor/repressor element 1 silencing transcription factor; PHD, folds into an interleaved type of Zn-finger chelating two Zn ions; poly-S/E/P, runs of serine, glutamate or proline; PWWP, domain including a conserved Pro-Trp-Pro motif; RRM, RNA recognition motif; SANT, DNA-binding domain that specifically recognizes the sequence YAC(G/T)G; Tudor, domain of unknown function present in several RNA-binding proteins; WW, contains two highly conserved tryptophans and binds proline-rich peptide motifs; ZI-MYND, ‘myeloid, Nervy, DEAF-1’ domain consisting of a cluster of cysteine and histidine residues.
**Characteristic structural features**

The structures of SET-domain proteins that are currently known include: the crystal structures of two members of the SUV39 family, *Neurospora crassa* Decrease in DNA methylation 5 (DIM-5) [15,16] and *S. pombe* CLR4 [17]; four structures of human SET7/9 in various configurations [18-21]; a nuclear magnetic resonance (NMR) structure of a viral protein that contains only the SET domain [22]; and

---

**Figure 2**

Schematic representations of the gene and primary protein structures of two pairs of related SET-domain histone methyltransferases in the SUV39 family. (a) Human SUV39H1 (gene, mRNA and protein); (b) human SUV39H2 (gene and mRNA for comparison with SUV39H1); (c) human G9a (gene, mRNA and protein); (d) human GLP1 (EuHMT1; mRNA for comparison with G9a; the gene structure is not shown because of the large size of the intron). Black boxes in the genes and white boxes in the mRNAs denote exons; numbers above each gene are exon numbers; numbers within exons indicates their size in nucleotides; thin lines in the genes indicate the introns and untranslated regions of the first and the last exons (these are shown to scale with the length of the exons except where lines are broken). (a,c) Protein structures are shown on the same scale as the coding region of the corresponding mRNA, so that the corresponding exons for each protein domain can be directly aligned. Domains are indicated above protein structures, and the number of conserved cysteines (Cys) in each domain is also shown. Abbreviations: Chromo, chromodomain; E-rich, glutamine-rich domain; KR-rich, domain rich in lysine and arginine; NRSF-binding, a domain involved in binding neuron-restrictive silencing factor/repressor element 1 silencing transcription factor. (c,d) The intron-exon junctions indicated with asterisks are those that differ between G9a and GLP1.
a structure of the non-histone protein methyltransferase Rubisco LSMT, an unclassified member of the superfamily [23,24]. These structures revealed that the SET domain forms a novel β fold not seen in any other previously characterized AdoMet-dependent methyltransferases (reviewed in [25]). The fold has a series of curved β strands forming several small sheets, packed together with pre-SET (or N-SET) and post-SET (or C-SET) domains or regions (Figure 3). The pre-SET domain of SUV39-family proteins (see Table 2) contains nine invariant cysteine residues that are grouped into two segments of five and four cysteines separated by various numbers of amino acids (CXCC_CX4_CXC_X5-CX5_CXCC_C, where N is 46 in DIM-5 and 28 in CLR4). The nine cysteines of the pre-SET domain of DIM-5 coordinate three zinc ions to form an equilateral triangular cluster, Zn2,Cys9 (Figure 3a). The SET domain, which may have evolved through the duplication of a three-stranded unit [26], is folded in all the solved structures into several small β sheets surrounding a knot-like structure by threading of the carboxyl terminus through an opening of a short loop formed by a preceding stretch of the sequence (Figure 3). This remarkable ‘pseudoknot’ fold brings together the two most-conserved sequence motifs of the SET domain (RFIN_HXCP and ELXFDY; see Figure 1) to form an active site in a location immediately next to the pocket where the methyl donor binds and to the peptide-binding cleft. The post-SET region of DIM-5 contains three conserved cysteine residues, arranged CXCC_C, that are essential for its histone lysine methyltransferase activity [15]. The structure of DIM-5 in a ternary complex with an H3 K9 peptide and AdoHcy [16] reveals that, as expected from their arrangement, these three post-SET-domain cysteines coordinate a zinc ion tetrahedrally together with cysteine 244 of the SET-domain signature motif RFINHXCPN in the pseudoknot near the active site (Figure 3a). Consequently, a narrow channel is formed to accommodate the side chain of the target lysine. Three ternary structures - SET7/9 in complex with a peptide containing histone H3 K4 [21], DIM-5 in complex with a histone H3 K9 peptide [16], and Rubisco LSMT in complex with a free lysine [24] - reveal that the target lysine is inserted into a narrow channel so that the target nitrogen would be in close proximity to the methyl donor AdoMet at the opposite end of the channel. Close examination of the region carboxy-terminal to the SET domain in many proteins, including members of the SUV39, SET1, and SET2 families, suggests that the post-SET-domain metal center observed in DIM-5 is universal among all those members of the superfamily that have the cysteine-rich post-SET domain. For almost all SET-domain proteins, there appears to be an absolute correlation between the presence of the post-SET domain and a cysteine corresponding to Cys244 of DIM-5 near the active site. Comparison of DIM-5 with SET7/9 [19,21] and the Rubisco LSMT [23,24], two SET-domain proteins that do not have Cys-rich pre-SET and post-SET domains, reveals a remarkable example of convergent evolution. In particular, as in DIM-5, these two enzymes rely on residues carboxy-terminal to the SET domain for the formation of lysine channel, but they do so by packing of an α helix, rather than a metal center, onto the active site.

Localisation and function

The lysine ε-amino groups of histones can be mono-, di-, or tri-methylated and, depending on the specific residue(s) modified, this methylation is associated with the formation of repressive heterochromatin, with transcriptional activation and elongation by RNA polymerase II or with the transcriptional silencing of euchromatic genes (Table 1). In heterochromatin, the H3 K9 methyl mark is recognized by heterochromatin protein 1 (HP1) in mammals and its homolog Swi6 in S. pombe; these proteins bind to the modified residue via their chromodomains, a domain shared by many regulators of chromatin structure [27,28], resulting in the formation of transcriptionally silent heterochromatin. Methylation of H3 K9 and H3 K27 is also associated with transcriptional silencing in euchromatin [29]. In contrast, di- and tri-methylation of Lys4 on the same histone (H3 K4) is associated with active transcription [30]. The Saccharomyces cerevisiae SET1 and SET2 complexes are involved in transcriptional elongation as part of the RNA polymerase II holoenzyme [31,32]. Tri-methylation of H3 K4 by SET1 is associated with regions of each gene that are transcribed early, in contrast to the SET2-mediated methylation of H3 K36, which is associated with downstream regions that are transcribed in the later stages of transcriptional elongation. The mammalian nuclear-receptor-binding SET-domain-containing protein (NSD1, a member of the SET2 family) has been found to play a crucial role in post-implantation development, methylating H3 K36 and H4 K20 [33]. ESET (also called SETDB1), which predominantly methylates H3 K9 in transcriptionally silent euchromatin, is also required for peri-implantation development [34]. ESET has been reported to bind the co-repressor KAP-1, which acts as a molecular scaffold, targeting ESET and HP1 to euchromatic genes silenced by KRAB-domain zinc-finger proteins [35]. Human SET7/9 mono-methylates H3 K4 [16,21], whereas S. cerevisiae SET1 di- or tri-methylates the same residue [36]. The strict lysine specificity of these enzymes is in distinct contrast to Drosophila ASH1 (a member of the SET1 family), mammalian G9a, human EZH1 and EZH2 and mouse NSD1, enzymes that can methylate two or more different lysine residues (Table 1). In some cases, the functions of SET-domain enzymes are not confined to histone methylation. For instance, human SET7/9 has recently been reported to methylate Lys189 in the general transcription factor TAF10, resulting in an increased affinity for RNA polymerase II and transcriptional activation of certain TAF10-dependent genes [37]. SET7/9 has also been reported.
to methylate p53, increasing the stability of this short lived tumor-suppressor protein [38]. These observations suggest that we should not narrowly define the SET-domain proteins as histone lysine methyltransferases but instead call them protein lysine methyltransferases.

DIM-5 tri-methylates H3 K9 [16], and this marks chromatin regions for DNA methylation [39]. Other members of the SUV39 family - KRYPTONITE of Arabidopsis thaliana [40,41], Suv39h1 of mouse [42], and mammalian G9a [43] - have been implicated in DNA methylation. In contrast to the tri-methylation of H3 K9 by which DIM-5 marks regions for DNA methylation [39], the critical mark for DNA methylation by KRYPTONITE is di-methylation of H3 K9 [40].

**Mechanism**

SET-domain proteins differ from other AdoMet-dependent methyltransferases in several respects; most notably, the binding sites for the histone substrate and AdoMet are located on opposite faces of the SET domain (Figure 4). The substrate-binding clefts are connected through a deep channel that runs through the core of the SET domain and permits transfer of the methyl group from AdoMet to the amino group of the lysine substrate. This unusual arrangement of the substrate binding sites was originally proposed to permit multiple rounds of lysine methylation without dissociation of the protein substrate from the SET domain [23]. Subsequent biochemical studies of SET-domain methyltransferases such as the histone H3 K9 tri-methylase DIM-5 [16] have supported this model of processive lysine methylation by members of the superfamily.

The crystal structures of N. crassa DIM-5 [15], human SET7/9 [18,19,21], and pea Rubisco LSMT [23] have revealed that the AdoMet-binding pocket is structurally conserved among SET-domain methyltransferases. The conserved G-X-G motif (in which X is generally a bulky hydrophobic residue) and the asparagine and histidine of the RFINHXCXPN motif (Figure 1) engage in hydrogen-bond and van der Waals interactions with the cofactor (Figure 4a). In addition, a positively charged residue that is structurally conserved in the cofactor binding site but is not conserved within the SET domain sequence forms a salt bridge with the carboxylate of AdoMet, as illustrated by the interactions between the cofactor and the side chain of Lys294 in SET7/9 (Figure 4a). Finally, the side chain of an aromatic residue from the C-SET or post-SET region forms an interaction involving stacking of aromatic π rings with the adenine moiety of AdoMet. The cumulative effect of these interactions causes AdoMet to adopt a horseshoe-shaped conformation, positioning the labile methyl group into the methyltransfer pore that links the cofactor-binding and lysine-binding sites.

The crystal structures of the ternary complexes of SET7/9 [21] and DIM-5 [16], respectively bound to AdoHcy and histone H3 peptides have yielded insights into the catalytic mechanism of the SET-domain methyltransferases. In both proteins, the walls of the lysine-binding channel are formed by hydrophobic residues that engage in van der Waals interactions with the lysine side chain (Figure 4b,c). At the base of the channel is the methyltransfer pore, which connects the pocket to the AdoMet-binding cleft. This pore is rimmed...
with several structurally conserved carbonyl oxygens as well as the hydroxyl group of the invariant tyrosine from the carboxyl terminus of the SET domain (Figure 1). These carbonyl and hydroxyl oxygens have been proposed to facilitate the transfer of the methyl group during catalysis [16,21,24]. In addition, tyrosine residues in the lysine-binding clefts of SET7/9 (Tyr245 and Tyr305) and DIM-5 (Tyr178) hydrogen-bond to the lysine ε-amino group, aligning it for a methyltransfer with AdoMet (Figure 4b,c). Mutation of Tyr245 or Tyr305 in SET7/9 (Figure 4b) alters its specificity from an H3 K4 mono-methylase to a tri- and di-methylase, respectively [16,21], whereas an Phe281Tyr mutation in the lysine-binding pocket of DIM-5 (Figure 4c) converts this protein to an H3 K9 mono- or di-methylase [16]. These mutations exemplify the F/Y switch (Figure 1) that establishes SET-domain product specificities. Taken together, these results...
have yielded insights into the catalytic mechanism and methyltransfer specificity of SET-domain methyltransferases.

The crystal structures of SET7/9 [21] and DIM-5 [16] bound to peptide fragments of histone H3 have also revealed the determinants for methylation of K4 and K9, respectively. The two enzymes bind to their cognate histone methylation sites in a structurally analogous orientation. The histone substrate binds in an extended conformation in a groove formed by the β6 strand and the loop exiting the thread-loop motif in the carboxyl terminus of the SET domain (Figure 4d,e). The backbone of the histone peptide is anchored in this site by a short parallel β sheet with the β6 strand. Specificity for methylation of Lys4 and Lys9 in histone H3 is achieved through recognition of the residues flanking each lysine residue. Residues Arg2, Thr3, and Gin5 in histone H3 are recognized through hydrogen bonds in the substrate binding cleft of SET7/9 [20] (Figure 4e). To compensate for the lack of side-chain interactions, the substrate-binding site of DIM-5 engages in a more extensive β-sheet hydrogen bonding with the backbone of histone H3 than does the substrate-binding site of SET7/9. Collectively, the crystallographic studies of SET7/9 and DIM-5 bound to histone H3 peptides have provided a framework for understanding the histone lysine specificity of different SET-domain enzymes.

Frontiers
As the function of SET-domain proteins becomes clearer, it is apparent that they can also be perturbed in disease. The recent recognition of the role played by the SET-domain protein SMYD3 in the proliferation of colorectal and hepatocellular carcinomas [44] may pave the way for the development of specific inhibitors of SMYD3 activity in cancer treatment. SMYD3 expression is upregulated in these cancers, and its histone H3 K4 methyltransferase activity activates oncopgenes and other genes associated with the cell cycle. MLL1, the human homolog of Drosophila Trx and a member of the SET1 family, is often implicated in leukemia as a result of aberrant Hox gene activation mediated by histone H3 K4 methylation [45]. Moreover, EZH2 is involved in metastatic prostate and breast cancer [46,47], suggesting that identification and targeted inhibition of SET-domain proteins involved in cancer may be useful for the treatment of patients in the future. The recent identification and characterization of two new SET-domain methyltransferases, SUV4-20H1 and SUV4-20H2, which function in H4 K20 tri-methylation [48], suggests that more SET-domain methyltransferases await characterization. In addition, the previously uncharacterized S. pombe SET9 protein has recently been shown to be able to methylate H4 K20 [49]. This modification does not seem to play a role in controlling gene expression or heterochromatin formation, but rather appears to be responsible for the recruitment of the checkpoint protein Crb2 to sites of DNA damage, unveiling yet another role for SET-domain proteins. The recent identification of the first histone demethylase, LSD1, which is conserved from S. pombe to humans, reveals that regulation of histone methylation is even more dynamic than was thought [50]. More recently, evidence has been provided that a cytosolic EZH2-associated methyltransferase complex regulates actin polymerization in various cell types, suggesting that SET-domain proteins may have many different roles in the cell [51].

Acknowledgements
S.C.D. is funded by the Wellcome Trust. X.D.C. is a Georgia Research Alliance Eminent Scholar and is funded by NIH grants GM62945 and GM61355.

References
1. Fischle W, Wang Y, Allis CD: Histone and chromatin cross-talk. Curr Opin Cell Biol 2003, 15:172-183.
2. Zhang L, Eugeni EE, Parthim MR, Freitas MA: Identification of novel histone post-translational modifications by peptide mass fingerprinting. Chromosoma 2003, 112:77-86.
3. Kuzmichev A, Jenewein T, Tempst P, Reinberg D: Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell 2004, 14:183-193.
4. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y: Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol 2002, 12:1052-1058.
5. Van Leeuwen F, Gafken PR, Gottschling DE: Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell 2002, 109:745-756.
6. Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K: Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev 2002, 16:1518-1527.
7. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G: 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 1994, 13:3822-3831.
8. Jones RS, Gelbart WM: The Drosophila Polycistron-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol Cell Biol 1993, 13:6357-6366.
9. Stassen MJ, Bailey D, Nelson S, Chinwalla V, Harte PJ: 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 1995, 52:209-223.
10. Jenuwein T, Laibl G, Dorn R, Reuter G: SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci 1998, 54:80-93.
11. Letunic I, Copley RR, Schmidt S, Ciacchelli FD, Doerks T, Schultz J, Ponting CP, Bork P: SMART 4.0: towards genomic data integration. Nucleic Acids Res 2004, 32:D142-D144.
12. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonhammer EL, et al.: The Pfam protein families database. Nucleic Acids Res 2004, 32:D138-D141.
13. Rea S, Eisenhaber F, O’Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T: Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 2000, 406:593-599.
14. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, Shinkai Y: G9a histone methyltransferase plays a dominant role in euchromatin histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 2002, 16:1779-1791.
15. Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Selker EU, Cheng X. Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. Cell 2002, 111:17-127.

16. Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Cheng X. Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell 2003, 12:177-185.

17. Min J, Zhang X, Cheng X, Grewal SI, Xu RM. Structure of the SET domain histone lysine methyltransferase Cerv4. Nat Struct Biol 2002, 9:82-93.

18. Jacobs SR, Hap JM, Devarakonda S, Kim Y, Rastinejad F, Khosravizadeh S. The active site of the SET domain is constructed on a knot. Nat Struct Biol 2002, 9:833-838.

19. Kwon T, Chang JH, Kwik E, Lee CW, Joachimak A, Kim YC, Lee J, Cho Y. Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. EMBO J 2003, 22:292-303.

20. Wilson JR, Jing C, Walker PA, Martin SR, Howell SA, Blackburn GM, Gamblin SJ, Xiao B. Crystal structure and functional analysis of the histone methyltransferase SET7/9. Cell 2002, 111:105-115.

21. Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell S, Taylor IA, Blackburn GM, Gamblin SJ: Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature 2003, 421:652-656.

22. Manzur KL, Farooq A, Zeng L, Plotnikova O, Koch AW, Sachidanand, Zhou M: A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. Nat Struct Biol 2003, 10:187-196.

23. Trievel RC, Flynn EM, Hourt RL, Hurley JH: Structure and catalytic mechanism of a SET domain protein methyltransferase. Cell 2002, 111:91-103.

24. Trievel RG, Flynn EM, Hourt RL, Hurley JH: Mechanism of multiple lysine methyltransferase by the SET domain enzyme Rubisco LSmT. Nat Struct Biol 2003, 10:545-552.

25. Schubert H, Blumenthal RM, Cheng X: Many paths to methyltranser: a chronicle of convergence. Trends Biochem Sci 2003, 28:329-335.

26. Aravind L, Iyer LM: Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit. Cell Cycle 2003, 2:369-376.

27. 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 2001, 410:116-120.

28. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T: Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 2001, 410:120-124.

29. Hwang KK, Eisenberg JC, Worman HJ: Transcriptional repression of euchromatic genes by Drosophila heterochromatin protein 1 and histone modifiers. Proc Natl Acad Sci USA 2001, 98:11423-11427.

30. Schreiber M, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T: Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Curr Opin Cell Biol 2003, 11:709-719.

31. Ng HH, Robert F, Young RA, Struhl K: The Paf1 complex group protein ezh2 controls actin polymerization and cell signaling. Cell 2005, 121:425-436.

32. Lachner M, Reinberg D: The many faces of histone lysine methyltransferase: a signature pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev 2004, 18:1251-1262.

33. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T: Active genes are tri-methylated at K4 of histone H3. Nature 2002, 419:407-411.

34. Kouskouti A, Scheer E, Staub A, Talianidis I: Gene-specific modulation of TAF10 function by SET9-mediated methylation. Mol Cell 2004, 14:175-182.

35. Schuch S, Kurash JK, Walker JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ, et al.: Regulation of p3 activity through lysine methylation. Nature 2004, 432:353-360.

36. Tamaru H, Zhang X, McMillen D, Singh PB, Nakanjara Y, Grewal SI, Allis CD, Cheng X, Selker EU: Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neospora caninum. Nat Genet 2003, 34:75-79.

37. Jackson JP, Johnson L, Yasencakova Z, Zhang X, Perez-Burgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE: Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 2004, 112:308-315.

38. Jackson JP, Lindroth AM, Cao X, Jacobsen SE: Control of Cpg DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 2002, 416:556-560.

39. Lehnerth B, Ueda Y, Dierick DA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH: Suv39-mediated histone H3 lysine 9 methylation directly DNA methylation to major satellite repeats at pericentric heterochromatin.Curr Biol 2003, 13:129-130.

40. Xiao B, Tashibana M, Giese M, Heard E, Shinkai Y, Waga S: Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. J Biol Chem 2003, 278:14996-15000.

41. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakanura Y: SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol 2004, 6:731-740.

42. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, Hess JL: MLL targets SET domain methyletransferase activity to the Prad-Willi gene promoter. Mol Cell 2002, 10:1107-1117.

43. Ramballay-Sl, Dhanasekaran SM, Zhou M, Barrett TR, Kumar-Sinha C, Sanga MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, et al.: The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 2002, 419:624-629.

44. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RG, Otte AP, Hayes DF, et al.: EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci USA 2003, 100:11606-11611.

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

46. Sanders SL, Porosito M, Matta J, Baehler J, Allshire RC, Kouzarides T: Methylation of histone H4 lysine 20 controls recruitment of Cen2 to sites of DNA replication. Cell 2004, 119:633-643.

47. Shi Y, Lan F, Matson C, Mulligan P, Whetstone JR, Cole PA, Casero RA, Shi Y: Histone demethylation mediated by the nuclear amino oxidase homolog LSD1. Cell 2004, 119:941-953.

48. Su IH, Dobruckecker MW, Dickenson E, Oser M, Basavaraj A, Marquenar R, Viale A, Reinberg D, Wuiling C, Tarakhovsky A: Polycomb group protein ezh2 controls actin polymerization and cell signaling. Cell 2005, 121:425-436.

49. Julien E, Herr W: A switch in mitotic histone H4 lysine 20 methylation status is linked to M phase defects upon loss of HCF-1. Mol Cell 2004, 14:713-725.

50. Karachentsev D, Sarma K, Reinberg D, Steward R: PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. Genes Dev 2005, 19:431-435.

51. Lachner M, Jenuwein T: The many faces of histone lysine methylation. Curr Opin Cell Biol 2002, 14:286-298.

52. Sims RJ 3rd, Nischio K, Reinberg D: Histone lysine methylation: a signature for chromatin function. Trends Genet 2003, 19:629-639.

53. Scheja S, Tachibana M, Tamaru H, Smith KM, Jia D, Zhang X, Selker EU, Shinkai Y, Cheng X: In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem 2005, 280:5563-5570.