Structure and Catalytic Mechanism of a SET Domain Protein Methyltransferase

Raymond C. Trievel,[1] Bridgette M. Beach,[1]
Lynnette M. A. Dirk,[2] Robert L. Houtz,[2]
and James H. Hurley[1,3]
[1]Laboratory of Molecular Biology
National Institute of Diabetes and Digestive and
 Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892
[2]Department of Horticulture
Plant Physiology/Biochemistry/Molecular
 Biology Program
University of Kentucky
Lexington, Kentucky 40546

Summary

Protein lysine methylation by SET domain enzymes regulates chromatin structure, gene silencing, transcriptional activation, plant metabolism, and other processes. The 2.6 Å resolution structure of Rubisco large subunit methyltransferase in a pseudo-bisubstrate complex with S-adenosylhomocysteine and a HEPES ion reveals an all-β architecture for the SET domain embedded within a larger α-helical enzyme fold. Conserved regions of the SET domain bind S-adenosylmethionine and substrate lysine at two sites connected by a pore. We propose that methyl transfer is catalyzed by a conserved Tyr at a narrow pore connecting the sites. The cofactor enters by a "back door" on the opposite side of the enzyme from substrate, promoting highly specific protein recognition and allowing addition of multiple methyl groups.

Introduction

In recent years, protein N-methylation has become an intensively studied regulatory modification of proteins. Two types of protein methylation have been characterized: methylation of the Nη nitrogen atoms of the guanidino side chain of arginines by protein arginine methyltransferases (PRMTs), and methylation of the ϵ-amino groups of lysine residues. Although a number of proteins possess methylated lysines, including histones (Murray, 1964), ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco; Klein and Houtz, 1995), calmodulin (Rowe et al., 1986), and cytochrome c (Durban et al., 1978), until recently, little was known about the enzymes that modify these proteins. The discovery of site-specific histone lysine methyltransferases that regulate chromatin structure and gene silencing (Rea et al., 2000; reviewed by Zhang and Reinberg, 2001; Jenuwein and Allis, 2001; Kouzarides, 2002; Lachner and Jenuwein, 2002) was a breakthrough that led to the current explosion of interest in protein lysine methylation.

The discovery of regulatory histone methylation was made possible by the identification of the SET (Su(var), Enhancer of zeste, Trithorax) domain (Schultz et al., 1998). The SET domain is a ∼110 amino acid motif shared by chromatin-related proteins of previously unknown biochemical function and by plant Rubisco large subunit methyltransferases (LSMTs). Homology between the SET domains of LSMTs and of the SUV39H1 family of position-effect variegation (PEV) modifiers provided the first hint that SUV39H1 proteins were histone methyltransferases (HMTs; Rea et al., 2000). In the two years since the discovery of regulatory histone methylation, the repertoire of histone lysine methylations has rapidly expanded, and, with one exception, with each new class has come the characterization of a new group of SET domain methyltransferases.

The first HMT group to be identified methylates Lys-9 of histone H3 (Rea et al., 2000). The founding member of this group is Su(var)3-9, a mediator of position effect variegation in Drosophila. The mammalian homologs of this protein are Suv39H1 and Suv39H2; these are important for genomic stability (Peters et al., 2001). The Schizosaccharoymes pombe homolog of Su(var)3-9 is clr4, which also methylates histone H3 Lys-9 (Nakayama et al., 2001). Among related histone H3 Lys-9 methyltransferases, G9A methylates both Lys-9 and Lys-27 of histone H3 (Tachibana et al., 2001). The newest group of histone H3 Lys-9 methyltransferases comprises SETDB1/ESET (Schultz et al., 2002, Yang et al., 2002).

Methylation of histone H3 Lys-9 has multiple consequences. It recruits the heterochromatic protein HP1 to chromatin via binding of the HP1 chromodomain to the methylated tail (Lachner et al., 2001, Bannister et al., 2001, Nakayama et al., 2001). This mediates gene silencing in heterochromatin and repression of gene expression in euchromatin by Rb (Nielsen et al., 2001). The structures of the HP1 chromodomain/methyllysine peptide complex have been determined (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002), showing how methyllysine is specifically differentiated from Lys by an aromatic recognition cage. In Neurospora crassa and Arabidopsis thaliana, histone H3 Lys-9 methylation has another function in that it appears to recruit DNA methylases (Tamaru and Selker, 2001; Jackson et al., 2002).

Another major group of HMTs is specific for histone H3 Lys-4. Mammalian SET9 and Saccharomyces cerevisiae SET1 methylate Lys-4 of histone H3 (Roguev et al., 2001; Wang et al., 2001; Nishioka et al., 2002a). This modification antagonizes the gene silencing effect of Lys-9 methylation, and the differences in methylation patterns correlate with the transcriptional activity in different regions of chromatin (Boggs et al., 2001; Litt et al., 2001; Noma et al., 2001). In yeast, histone H3 Lys-4 methylation regulates rDNA silencing (Briggs et al., 2001; Bryk et al., 2002). The consequences of methylation of other Lys residues are not yet as clear. S. cerevisiae SET2 methylates histone H3 Lys-36, which represses transcription (Strahl et al., 2002). PR-SET7 and SET8 methylate Lys-20 of histone H4, which silences genes by interfering with acetylation of adjacent sites of the histone H4 tail (Fang et al., 2002; Nishioka et al., 2002b).

[3]Correspondence: jh8e@nih.gov

There are 368 SET domain proteins identified in the SMART database (Schultz et al., 1998), distributed through all kingdoms of life. To date, the Rubisco large subunit methyltransferases (LSMTs) of plants are the only other class of SET domain enzymes with identified substrates. Indeed, the Rubisco LSMTs were the first SET domain methyltransferases whose enzymology was characterized. Methylation of Rubisco large subunits resembles histone methylation in many respects. Lys-14 is the site of methylation on the Rubisco large subunit, and is located in the flexible N terminus of the large subunit (Schneider et al., 1992). The analogy is clear to the sites of histone methylation in the N-terminal tails of histones H3 and H4, which have been shown to be disordered in the crystal structure of the nucleosome core (Luger et al., 1997). Furthermore, Rubisco, like the nucleosome, is a large assembly, a hexadecamer of eight large and eight small subunits. The activity of Rubisco is not affected by methylation of its N-terminal tail. The precise role of Rubisco tail methylation is unknown at this time, but is thought to involve targeting of other proteins to interact with the tail. LSMTs are highly expressed in leaves during daylight and may be involved in the regulation of Rubisco during photosynthesis (Klein and Houtz, 1995; Ying et al., 1999).

Despite the intense interest in this family of enzymes, thus far, little structural or mechanistic information has been available for them. In order to elucidate the structure, function, substrate specificity, and catalytic mechanism of SET domain methyltransferases, we screened a large series of SET domain-containing proteins for suitability for structural studies. SET domain-containing ORFs from diverse species including yeast, human, *Arabidopsis thaliana*, tobacco, spinach, and pea were cloned and screened by systematically scanning the N- and C-terminal boundaries of their respective SET domains for constructs that were soluble when expressed in *E. coli*. Among the soluble SET domain methyltransferases that were identified using this approach were several Rubisco LSMTs and the human HMT SET9. In screening several different recombinant Rubisco LSMTs, we found that the garden pea (*Pisum sativum*) Rubisco LSMT yielded crystals suitable for determination of the structure of a SET domain-containing enzyme. We determined the structure bound to the cofactor reaction product *S*-adenosylhomocysteine (AdoHcy) and a fortuitously bound molecule of HEPES that mimics lysyl substrate binding. The structure provides the framework for a general understanding of SET domain function, and its implications have been tested by mutational analysis of pea LSMT and human SET9.

Results and Discussion

Overall Structure and Trimerization

The structure of pea Rubisco LSMT bound to AdoHcy was solved by single anomalous dispersion (SAD) from a Xe derivative, and refined at 2.6 Å resolution (Figure 1A). The crystallized enzyme consists of residues 46–482, as compared to the 489 amino acids genetically encoded. Residues 1–36 of pea Rubisco LSMT comprise the chloroplast transit sequence and are removed in the mature protein in vivo (Zheng et al., 1998). The protein

was prepared as a C-terminal hexahistidine fusion, with an intervening TeV protease site. The crystallized protein contains the six residues ENLYFQ from the TeV protease recognition site fused to its C terminus, as well as an additional N-terminal Met from the initiator codon.

The structure of Rubisco LSMT consists of two large lobes and a C-terminal extension (Figure 1B). The N-terminal lobe (residues 50–305) can be further divided into a β sheet domain (residues 69–106 and 222–291) and an α-helical domain (residues 50–68, 107–221, and 292–305). The β sheet domain within the N-terminal lobe corresponds to the SET domain (Figures 1D and 2A). The C-terminal lobe (residues 306–482) comprises a single domain that is entirely helical except for one two-stranded β sheet (Figure 1E). Neither lobe shows similarity to other structures in the protein data bank as judged by searches with CE, DALI, and VAST. The C-terminal portion of the protein (residues 463–482 and the five vector-derived C-terminal residues) protrudes into the interlobe cleft of a neighboring molecule, forming a trimer interface (Figure 1C). Rubisco LSMT is a monomer in solution (Wang et al., 1995; R. Ghirlando, personal communication). We believe the trimer is formed by domain swapping (Liu and Eisenberg, 2002) under the conditions of crystallization.

There are minor differences between the three monomers in the trimer. Two monomers denoted B and C are well ordered throughout, with the exception of the N-terminal three amino acid residues and the C-terminal residues. The A monomer manifests disorder within the N-terminal domain; residues 228–230 and 257–266 of the A monomer are completely absent from electron density. There are modest differences in the angle between the N and C-terminal lobes in the monomers; the interlobe distance is about 30 Å in molecule A as compared to 27–28 Å in molecules B and C. The greatest local structural differences are in residues 252–266. Most of this segment participates in crystal contacts in molecules B and C, whereas all but the first five residues in this segment are disordered in the A molecule.

The SET domain consists of 12 β strands arranged into 5 partially interwoven sheets (Figures 2A and 2B). Sheets I–IV are antiparallel, whereas sheet V is parallel. Sheets I and V are interwoven, in that β12 participates in both sheets. Sheet I contains strands β1, β2, and β12; sheet II strands β3 and β11; sheet III strands β4, β10, and β9; sheet IV strands β5, β7, and β6; and sheet V strands β8 and β12 (Figure 2B). There is a knot-like structure where the C terminus of the SET domain exits underneath the β8-β9 connection. The all-β SET domain bears no resemblance to the classical α/β AdoMet binding fold of other methyltransferases (Schluckebier et al., 1995; Weiss et al., 2000; Zhang et al., 2000).

The SET domain is interrupted between β5 and β6 by a Rubisco LSMT-specific domain (Figures 1D, 2, and 3; we refer to it as the SET-inserted domain, or iSET domain), consisting of 115 residues and six α helices (Figures 1D and 3). The inserted region forms a single folding unit together with the N- and C-terminal helices of the N-terminal lobes, α1 and α8, which we will refer to as the "nSET" and "cSET" regions, respectively. The SET domain and nSET, iSET, and cSET regions of the N-terminal lobe interact extensively with each other. Collectively, these regions form a domain that is shaped like
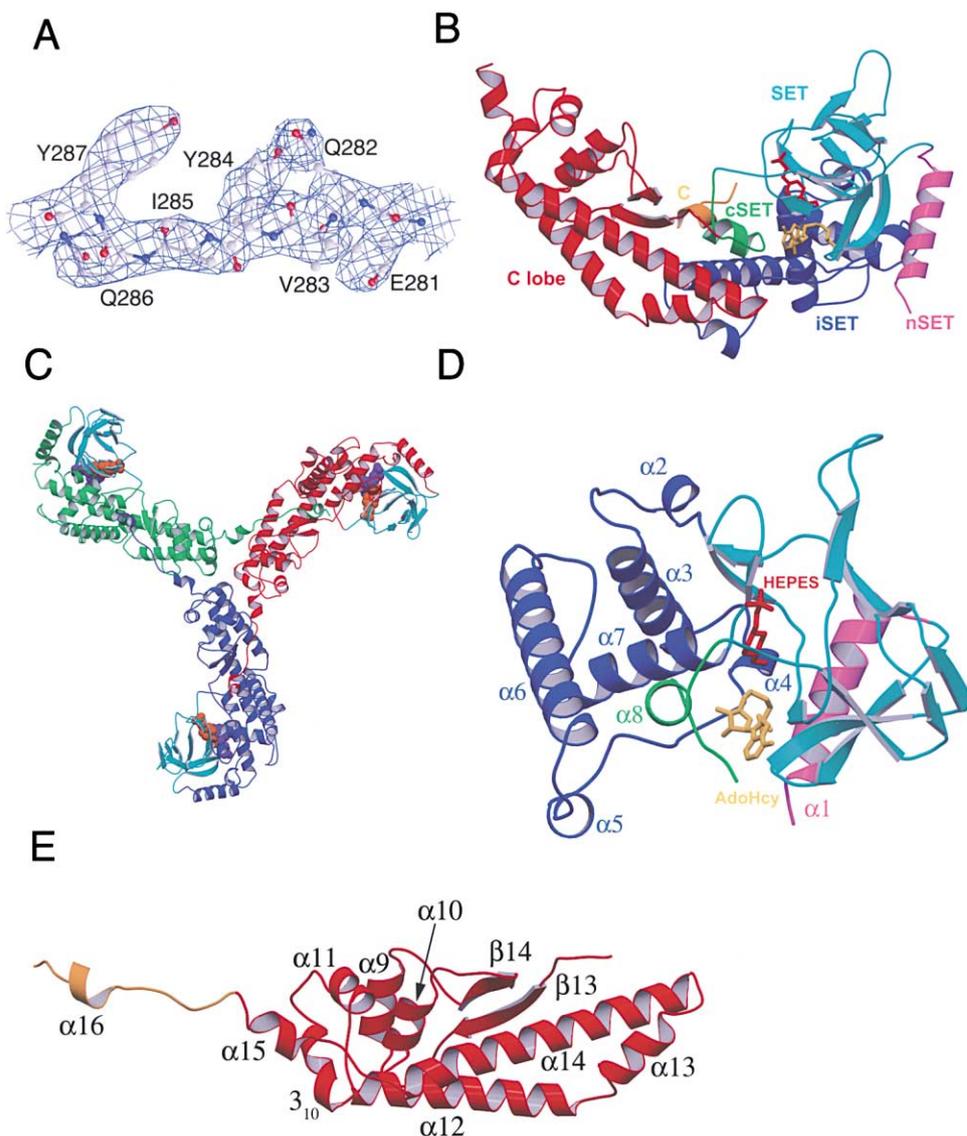
Figure 1. Structure of LSMT and the SET Domain

(A) Experimental electron density taken from the solvent-flattened Xe SAD Fourier synthesis and contoured at 1.0 σ. Density is shown in the vicinity of the putative catalytic Tyr-287 of the B monomer.

(B) Structure of the LSMT monomer together with the putative domain-swapped extension of the neighboring monomer. The nSET, iSET, cSET, and SET regions are colored magenta, blue, green, and cyan, respectively, while the C-terminal lobe is red, and the domain-swapped extension is gold. Stick representations of HEPES and AdoHcy are colored orange and gold, respectively.

(C) Structure of the LSMT trimer observed in the crystal. The SET domain is colored cyan in each monomer, while the remainders of the monomers are denoted in red, green, and blue, respectively. Spacefilling CPK representations of HEPES and AdoHcy are colored orange and violet, respectively.

(D) The N-terminal lobe with secondary structures of the nSET, iSET, and cSET regions and ligands labeled and colored as in (B).

(E) C-terminal lobe with labeled secondary structures, colored as in (B). Figures were generated with Molscript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

a catcher's mitt and grips the SET domain on three sides, burying much of SET domain sheets III and IV.

Two deep clefts are formed at the interface between the SET and the iSET/cSET domains of the N-terminal lobe. The first cleft is formed where SET domain sheet III contacts the cSET region. One molecule of the cofactor AdoHcy is bound in this cleft (Figures 1D, 2A, and 4C). A second cleft is formed where SET domain sheet IV contacts α3 and α7 of the iSET. This narrow cleft opens into a broader gap between the N- and C-terminal lobes of Rubisco LSMT, and contains a molecule of the buffer ion HEPES (Figures 4A, 4B, and 4D). The identities and binding modes of AdoHcy and HEPES were confirmed using Bijvoet difference syntheses (Figure 4B). The syntheses identified the sulfur atoms of the thioether and sulfonic acid moieties of the two ligands. The two clefts meet in the center of the N-terminal lobe at the presumptive active center of the enzyme.
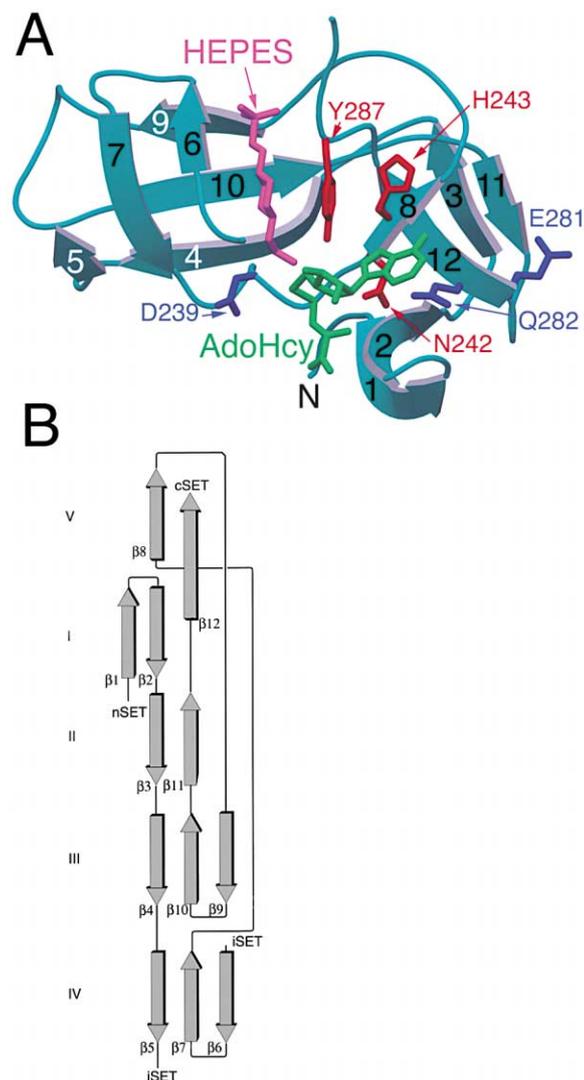
Figure 2. β-Sheet Architecture of the SET Domain

(A) Structure of the SET domain, with β strands numbered. Invariant residues in the SET domain are highlighted in red and conserved are colored blue. HEPES and AdoHcy are shown in magenta and green.

(B) Topology of the SET domain β sheets.

AdoHcy Binding Site

The AdoHcy cofactor binds to a face of the SET domain formed by the β1-β2 turn, the loop preceding β6, and β8 (Figure 2A). While the bulk of the AdoHcy binding site is formed by a cleft in the SET domain, the cSET domain covers the site (Figure 1D). The adenine base binds in a hydrophobic groove between the aliphatic part of the Glu-80 side chain and Phe-302 (Figure 4C). The Ade N6 and N7 are hydrogen bonded to the main chain O and NH of His-243. The ribose O4 approaches the side chain of Asn-242. The positively charged amino nitrogen of the Hcy is hydrogen-bonded to the main chain carbonyls of residues 80 and 82 and the side chain Oδ of Asn-242. The Hcy carboxylate interacts with the side chain of Arg-222 and the main chain amide of Leu-82. The Sδ of the Hcy is situated at the deepest point

in the cleft, poised at the narrow opening to the second cleft in which the HEPES ion is bound.

HEPES and Lysine Binding Site

One molecule of the buffer ion HEPES is bound in the second cleft, where it occupies what we believe is the binding site for the substrate Lys residue. This cleft is formed where β6 and β12 of the SET domain meet the C-terminal section of the iSET domain (Figures 1D and 2A). At the rim of the cleft, Arg-226 and His-252 form hydrogen bonds with the sulfonic acid moiety of the HEPES molecules (Figure 4D). The main chain amide groups of residues 225 and 226 form hydrogen bonds with the sulfonic acid group. Five hydrophobic residues, Phe-224, Ile-241, Ile-285, Tyr-287, and Tyr-300, form the sides of the cleft. The hydroxyethyl moiety and the positively charged piperazine ring of HEPES are bound in this part of the cleft. At the very bottom of the pit is a carbonyl cage formed by the carbonyl groups of residues 221, 222, 239, and 241. The terminal hydroxyl group of the HEPES is closest to the latter two carbonyls in this cage, and donates a hydrogen bond to the D239 carbonyl. The carbonyl cage forms the gate to the first cleft, as described above. The hydroxyl group is poised at the entrance to the cofactor binding cleft, within 3.5 Å of the Sδ of the AdoHcy. The hydroxyl group is 4.8 Å from the hydroxyl of Tyr-287.

The HEPES OH group is within 4 Å of three potential hydrogen bond acceptors, the 239 and 241 carbonyls and the AdoHcy thioether. The HEPES OH is nominally capable of donating only one hydrogen bond, however. The HEPES OH may have been forced into the site by the strength of surrounding interactions, since its direct polar interactions are not all favorable. It appears that this site is designed to accommodate a moiety capable of donating multiple hydrogen bonds, such as a Lys ε-ammonium group. We modeled the Lys side chain of the substrate by overlaying the terminal portion of the side chain on the HEPES piperazine moiety (Figure 5A). With the appropriate torsions in the Lys side chain, the Nζ could donate hydrogen bonds to the Tyr OH and the carbonyls of residues 221 and 222. In the reactive complex, the hydrogen bond-accepting lone pair on the AdoMet S would be occupied by the sulfonium methyl group, and thus would not be available to interact with the Lys. Instead, the methyl group would be positioned in precisely the location required for transfer to the Lys Nζ.

The serendipitous binding of a HEPES ion in the active site provides a plausible model for how the substrate Lys residue interacts with the enzyme. Indeed, HEPES is a weak inhibitor of pea Rubisco LSMT. In addition to the hydrogen bonding interactions described above, the Lys binding site also contains an aromatic cluster like the methyllysine binding pocket of the HP1 chromodomain (Nielsen et al., 2002; Jacobs and Khorasanizadeh, 2002). This is significant because SET domain methyltransferases can generate trimethyllysine, and therefore must be capable of using mono- and dimethyllysine as substrates.

Kinetic Analysis of SET Domain Methyltransferases

In order to verify that the enzymatic function of the protein was not compromised in the construct used for
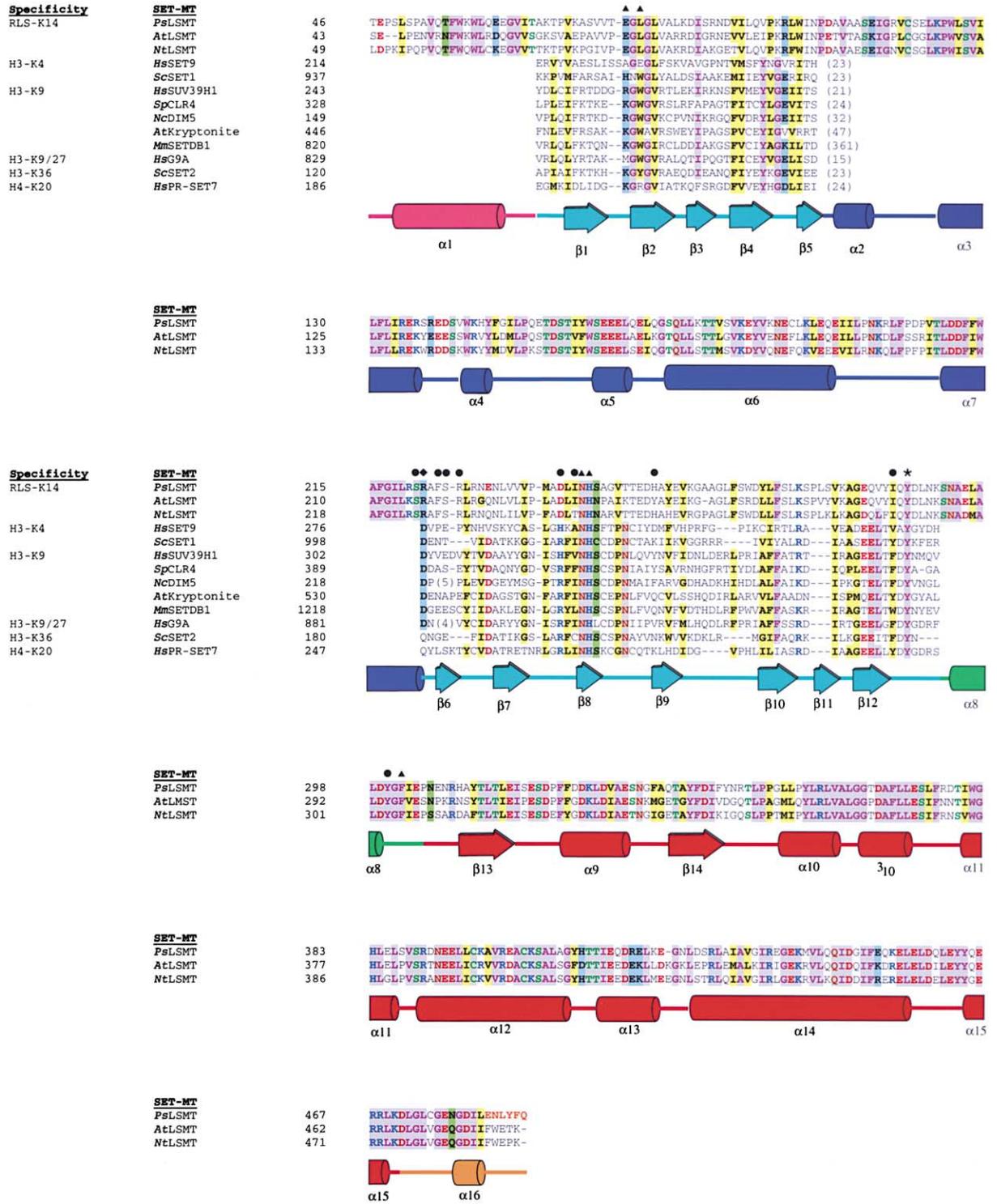
Figure 3. Secondary Structure and Sequence Alignment of the SET Domains of the Plant LSMTs and HMTs

The background text color gray represents invariant residues, while the colors yellow, green, and blue signify conserved hydrophobic, polar, and charged residues, respectively. The symbols ▲ and ● denote residues involved in AdoMet and substrate lysine binding respectively, while ◆ represents residues implicated in both. The putative catalytic Tyr is highlighted by a *. The secondary structure of the pea LSMT is illustrated under the alignment and is colored according to the scheme in Figure 1B. Species abbreviations: *Ps, Pisum sativum*; *At, Arabidopsis thaliana*; *Nt, Nicotiana tabacum*; *Sc, Saccharomyces cerevisiae*; *Hs, Homo sapiens*; *Sp, Schizosaccharomyces pombe*; *Nc, Neurospora crassa*; and *Mm, Mus musculus*. The substrate specificities of the LSMTs for the Rubisco large subunit (RLS) and the HMTs for histones H3 and H4 are listed to the left of the alignment. The C-terminal vector-derived sequence present in the structure is shown in orange letters.
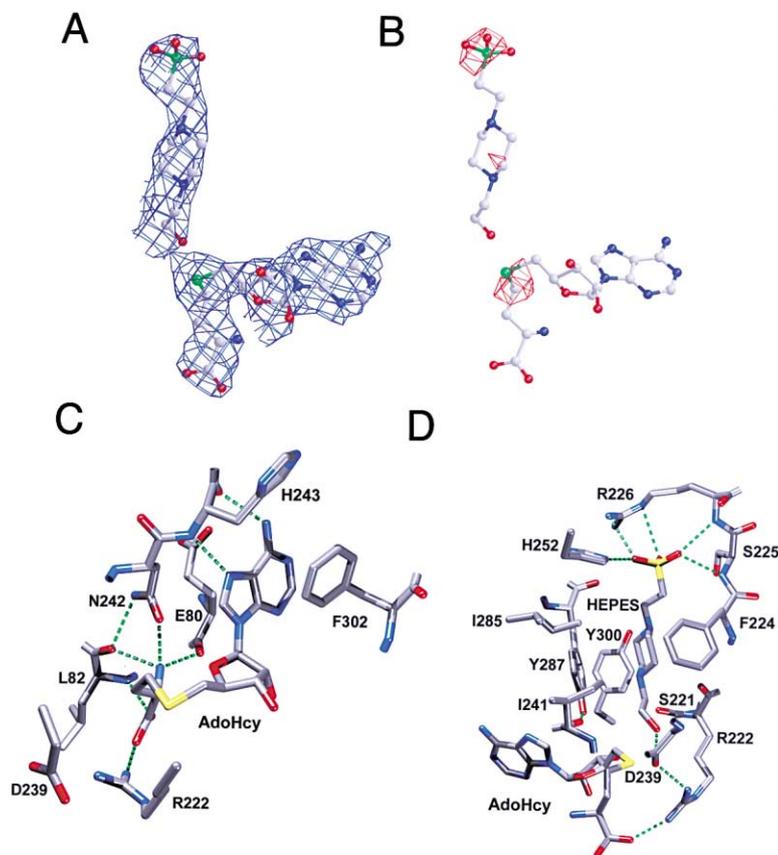
Figure 4. Ligand Binding Sites

(A) Electron density calculated from a ($2F_{obs} - F_c)\alpha_{calc}$ synthesis that was derived from a partially refined model of the LSMT protein prior to inclusion of ligands, contoured at 1.0 σ.

(B) Anomalous difference density from a ($F^+ - F^-)\alpha_{exp}$ synthesis, using Xe derivative structure factors derived from the solvent-flattened experimental map. Essentially the same map is obtained using native structure factors (not shown).

(C) AdoHcy in its binding pocket. Interacting residues are shown, and hydrogen bonds are shown in dashed green lines.

(D) HEPES in its binding pocket, also showing AdoHcy to illustrate their relative position.

crystallization, we determined the kinetic parameters for the enzyme. Recombinant pea Rubisco LSMT 46–482 has a V_{max} = 56 nmol min^{-1} mg protein^{-1}, (k_{cat} = 0.05 s^{-1}, K_m[AdoMet] = 6 μM, and K_m [Rubisco] = 1.4 μM) (Figures 6A and 6E). The V_{max} and K_m(AdoMet) parameters are virtually identical to those reported for a construct spanning the entirety of the mature enzyme (Zheng et al., 1998). We found that pea Rubisco LSMT is marginally stable with respect to mutagenesis, and mutations that have been constructed at the conserved residues Asp-239, Asn-242, His-243, Glu-282, and Tyr-287 in pea Rubisco LSMT have proved insoluble when expressed in *E. coli*. Because the C-terminal GE(E/Q) motif is conserved, we constructed the LSMT variant E281Q. The E281Q mutant has approximately normal K_m values for AdoMet (4.1 μM) and Rubisco (3.0 μM), but its k_{cat} is reduced 16-fold to 0.003 s^{-1} (Figure 6A). Because HEPES was observed bound in the active site, we tested whether it inhibits pea Rubisco LSMT enzyme activity. The presence of 500 mM HEPES (pH 8.0) results in a 36% loss of activity when compared to an identical level of Bicine (pH 8.0) at 1.7 μM Rubisco (data not shown).

We extended the mutational and kinetic analysis to SET9 for two reasons: to determine whether the structure of a SET domain from an LSMT has predictive value with respect to the function of an HMT, and to obtain insight into the catalytic roles of residues for which the pea LSMT mutants were insoluble. Full length recombinant wild-type SET9 has a V_{max} = 5.9 nmol min^{-1} mg protein^{-1} (k_{cat} = 0.004 s^{-1}, K_m[AdoMet] = 6 μM, and K_m [histone H3] = 5.4 μM) (Figure 6B). All SET9 mutants tested were soluble and expressed protein at levels comparable to wild-type (Figure 6D). SET9 mutants N296A, H297A, and Y335F, the counterparts of pea LSMT mutants N242A, H243A, and Y287F, were all enzymatically inactive (Figure 6E). The SET9 H293A mutant, which corresponds to pea LSMT D239A, is active but has lower affinity for AdoMet by about 4-fold (Figures 6B and 6E), consistent with its role in the AdoMet binding site. Surprisingly, it also increases the affinity for histone H3 (Figure 6E). To place constraints on the identity of a possible general base in the reaction, we carried out pH/rate analysis of pea Rubisco LSMT and SET9 (Figure 6C). The rates for the two enzymes have pKa values of 7.1 and 8.3, respectively, as compared to 7.8 for tobacco Rubisco LSMT (Houtz et al., 1991).

Conserved Motifs and Mutational Analysis of the SET Domain

The conserved sequence motifs that define the SET domain can be considered in light of the structure. The first completely conserved residue in the SET domain is Gly-81 (pea Rubisco LSMT numbering), which is part of the β1-β2 turn (Figure 3). This turn sits over the AdoMet binding site, and positions the carbonyls of the adjacent residues 80 and 82 to interact with the AdoHcy NH$_3^+$. The next conserved motif is the (H/R/D)$\Phi\Phi$NHSC sequence, where Φ denotes a large hydrophobic residue. The first residue in this sequence in pea Rubisco LSMT is Asp-239. The carbonyl of this Asp is one of the two
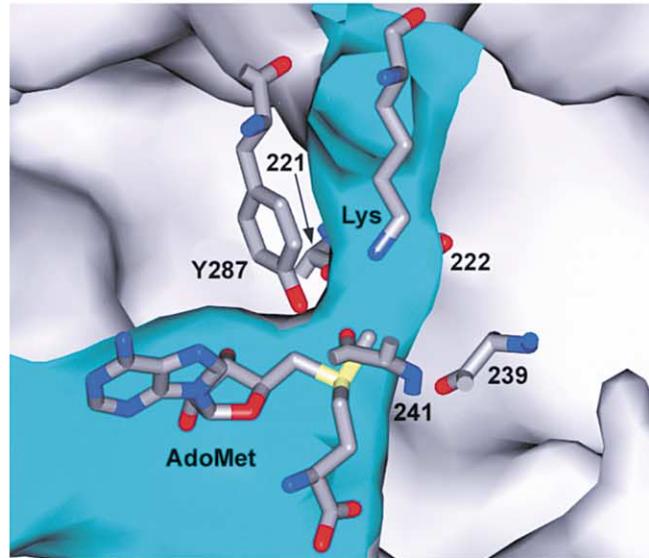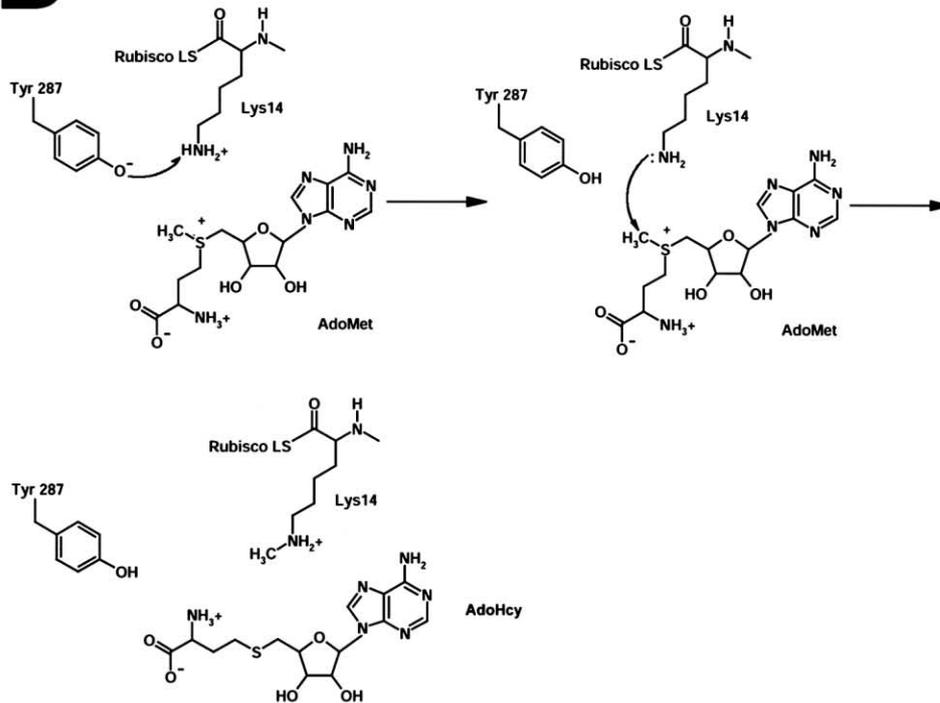
Figure 5. Catalysis by SET Domain Enzymes

(A) Model-built substrate Lysyl residue and AdoMet are shown in a cutaway of a space-filling surface model to illustrate their interaction through a constricted pore (cyan). The lysine was positioned by aligning its carboxylate and Nζ moieties to the HEPES sulfate group and the piperazine nitrogen atom closest to the AdoMet site, respectively.
(B) Proposed reaction mechanism for methyl transfer.

presumed ligands for the Nε of the substrate Lys. The Asp-239 side chain forms a buried salt bridge with Arg-222, which in turn interacts with the AdoHcy COO⁻ moiety. The HMTs have an Arg or His residue at this position instead of an Asp, and they appear, subject to the uncertainty of the sequence alignment in this region, to have an Asp, or in a few cases, a Gln, as a replacement for Arg-222 (Figure 3), maintaining the charged-pair interac-
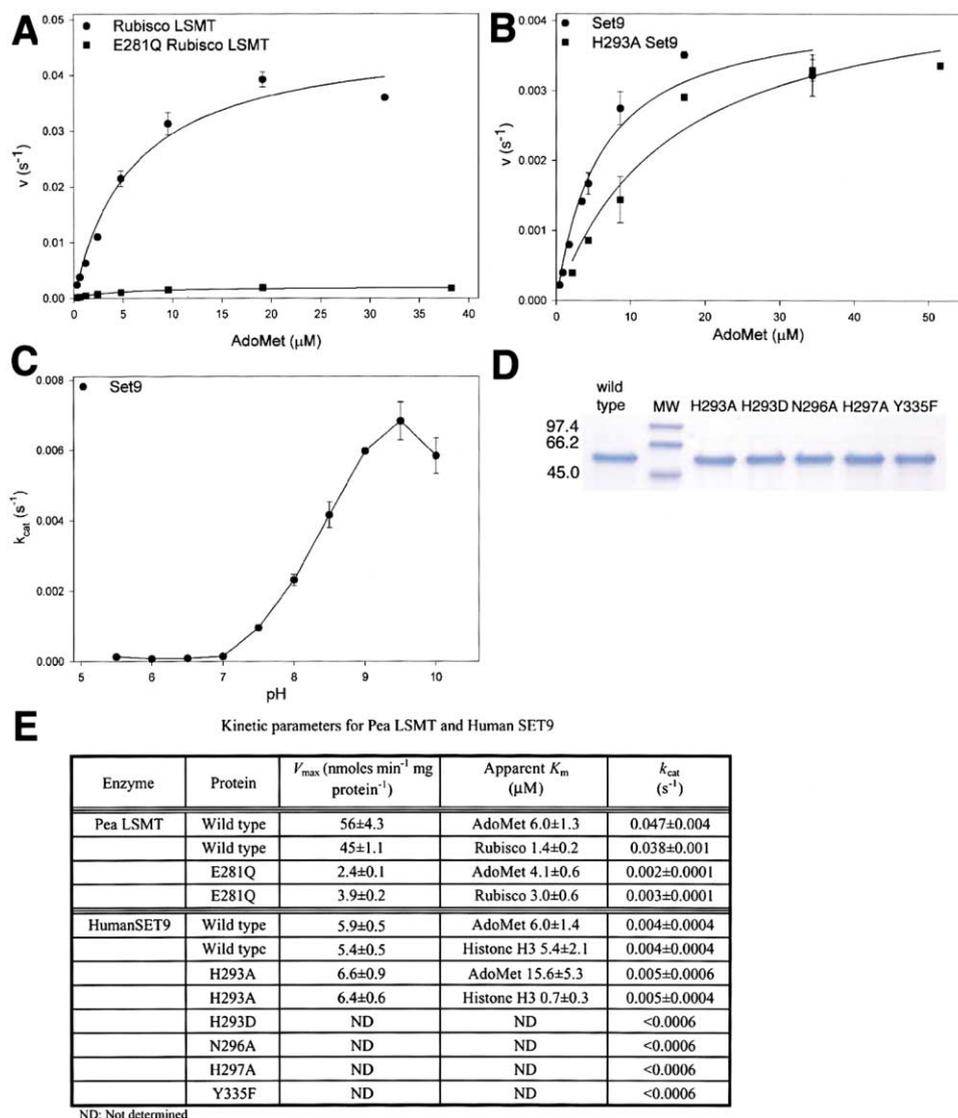
Figure 6. Enzyme Kinetics of Rubisco LSMT and SET9

(A) Rate (in units of turnovers/sec) versus [AdoMet] for wild-type and E281Q LSMT.

(B) Rate versus [AdoMet] for wild-type and H293A SET9.

(C) pH rate profile of k_{cat} versus pH for SET9. Analysis of the plot yields a pK_a value of 8.3 and a pH optimum of 9.5.

(D) Coomasie stained SDS-PAGE gel of the purified wild-type and mutant SET9 constructs used in this analysis.

(E) Summary of the kinetic data for LSMT and SET9 wild-type enzymes and mutants.

tion. In the HMTs, this residue might interact directly with the AdoMet COO^- moiety. In SET9, this position is occupied by His-293. The SET9 H293A mutant is impaired in AdoMet binding, as expected from its role in the structure. Arg at this position favors activity by 20-fold over His in the HMTs (Rea et al., 2000; Tachibana et al., 2001), suggesting that the Arg is better able to reach and bind the AdoMet COO^-. Modeling of the Arg side chain shows that the side chain can be overlaid on that of Arg-222. Mutation of the Arg to Gly in SET2 and PR-SET7 inactivates these enzymes (Strahl et al., 2002, Nishioka et al., 2002b), consistent with a functional role. The two hydrophobic residues are structural anchors.

The conserved Asn-242 and His-243 are linchpins of the active site (Figure 4C). The side chain of the Asn binds the AdoHcy NH_3^+, together with the residue 82

carbonyl, and the side chain also comes into close approach with the ribose O4′. The His side chain is hydrogen bonded to the main chain amide of the putative catalytic base Tyr-287, and also to Glu-304 of the cSET region. These interactions may explain why mutation of this His in Suv39h1, SETDB1, and SET9 (Rea et al., 2000; Schultz et al., 2002; Wang et al., 2001; Figure 6D) inactivates these enzymes. Ser-244 and Ala-245 (the counterpart to the semiconserved Cys in some HMTs) probably owe their conservation and functional importance to their roles in anchoring the key β8 strand.

The C terminus of the SET domain contains the conserved motif GE(E/Q) and an absolutely conserved Tyr. Gly-280 of the GE(E/Q) sequence forms a key part of the β11-β12 turn (Figure 2A). The conservation of the Gly is consistent with its nonstandard ϕ,ψ angles
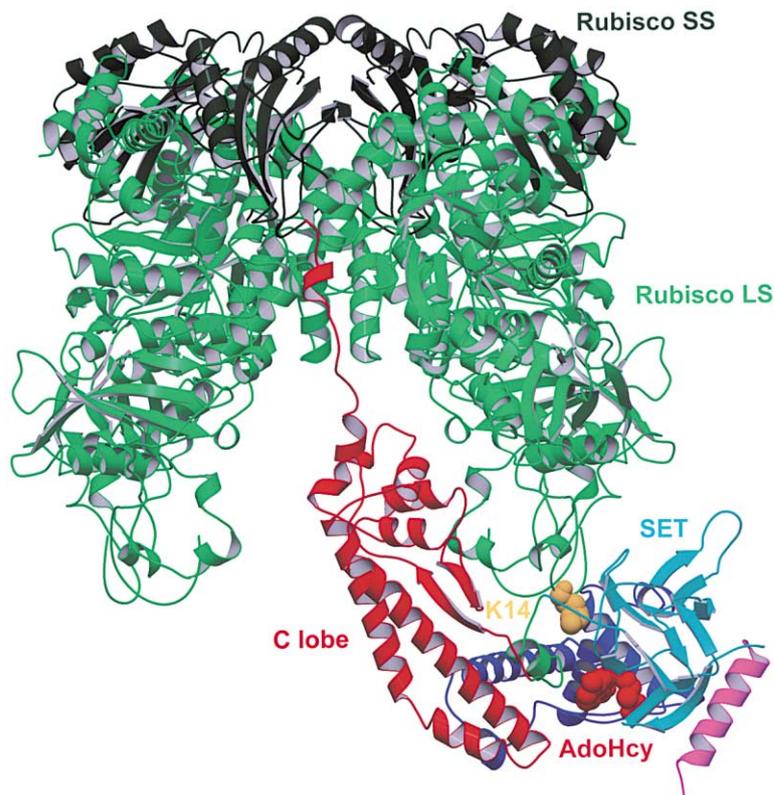
Figure 7. Protein Substrate Binding

The structure of spinach Rubisco (PDB 1RXO) was docked to LSMT to position Lys-14 (gold) as close as possible to the HEPES binding site while avoiding steric collisions between the two proteins and without altering the conformation of the flexible Rubisco N terminus. The Rubisco large and small subunits are colored light and dark green, respectively, while LMST is shown as in Figure 1B, with the exception that AdoHcy is orange.

(101°, −31°), which are energetically disallowed for non-Gly residues. The structure of this turn is important for catalytic activity, since β12 is an important player in the structure of the active site. Glu-281 appears to stabilize the turn by forming a salt bridge with Lys-278. Gln-282 is hydrogen bonded through its Oϵ to the critical Asn-242 and positions it for its role in AdoMet binding. This interaction uses the Gln side chain oxygen, consistent with its replacement by Glu in HMTs. Finally, the absolutely conserved Tyr-287 forms part of the wall of the Lys binding site, and its OH group is close to the expected position of the substrate Lys Nζ. The Y335F mutant of SET9 is completely inactive (Figure 6E).

Catalytic Mechanism

The model for the Lys side chain based on the structure of the HEPES complex puts it in the center of the carbonyl cage, ∼3 Å from the methyl group of a modeled AdoMet (Figure 5A). Because the proximity to the methyl donor and the polarity of the site are appropriate, the Lys Nζ of the substrate probably occupies a similar position in the reactive complex. There are no acidic residues positioned to bind the Nζ of the Lys. However, the carbonyl cage at the base of the pocket could bind the terminal amino group of the substrate Lys (Figure 5A). Histone acetyltransferases (HATs) also use main chain carbonyls to bind the substrate Lys Nζ (Rojas et al., 1999).

The unprotonated Lys side chain is the species that is presumed to make a nucleophilic attack on the AdoMet methyl group. There is no His, Asp, or Glu present with the appropriate geometry to deprotonate the substrate.

However, the OH of the absolutely conserved Tyr-287 is within 4.0 Å of the presumptive location of the substrate Lys Nζ. In the productive Michaelis complex, two positively charged moieties of the bound ligands, the AdoMet sulfonium cation and the ϵ-ammonium cation of the Lys side chain, are juxtaposed in close proximity in a deeply buried site. The pKa of the Tyr could be depressed, activating it as general base (Figure 5B). The pKa values of the reaction rates of SET enzymes tested center around 8, within ∼2 units of the pKa values of Cys, Lys, or Tyr. Only the latter is present in the active site. The optima might be consistent with an alternative model in which the pKa reflects the Lys side chain of the substrate, but a depressed pKa for the substrate Lys would be difficult to reconcile with its high solvent exposure in Rubisco. The structure and pH profiles taken together seem most consistent with a catalytic role for the Tyr.

One of the most remarkable aspects of the Rubisco LSMT active site is that the substrate and cofactor enter separate clefts from separate ends of the enzyme (Figure 5A). The two ligands meet at a pore barely large enough to accommodate the transfer of a methyl group from one cleft to the other. In a more conventional protein fold, binding protein first might sterically obstruct cofactor binding. The two-cleft structure of the SET domain avoids this problem, since it allows the AdoMet a back door entry that is not sterically obstructed by the substrate protein complex. While there are little data on the processivity of methyl group addition, the back door mechanism suggests that multiple methyl groups could be added to a Lys side chain without dissociation of the protein. The back door arrangement is suited for specific

Table 1. Crystallographic Data Collection, Phasing, and Refinement Statistics

(A) Crystallographic data and phasing statistics

Space group:			I222		
Cell dimensions (Å), native:			a = 132.16, b = 156.68, c = 268.44, $\alpha = \beta = \gamma = 90°$		
Cell dimensions (Å), Xe derivative:			a = 132.42, b = 158.14, c = 268.16, $\alpha = \beta = \gamma = 90°$		

	d_{min} (Å)	No. of reflections[a]	Completeness(%)[a]	$<I>/<\sigma>$[a]	R_{sym}(%)[a,b]
Native (λ = 1.5418 Å)	2.6	83,954 (8294)	98.1 (98.1)	31.6 (2.31)	4.7 (50.8)
Xenon (λ = 1.5418 Å)	3.0	56,506 (5580)	99.9 (100.0)	26.4 (4.56)	8.0 (55.1)

Observed Bijvoet ratios[c]

Xe, λ = 1.5418 Å:	5.0%
Figure of merit:	0.252 (0.781)[d]
Phasing power:	1.23

(B) Refinement

Resolution range	35.0–2.6 Å
No. of reflections	74,865
$R_{working}$[e]	23.2%
R_{free}[f]	27.8%
Luzatti coordinate error	0.40
Cross-validated Luzatti coord. error	0.49
Bond-length deviation	0.007 Å
Bond-angle deviation	1.2°
Improper angle deviation	0.79°
Dihedral angle deviation	21.3°
Average B factor	
Overall	69.6 Å²
Protein	69.1 Å²
Ligands	66.4 Å²
Water	63.0 Å²
Bonded mainchain atom B factor rmsd	1.33 Å²
Bonded sidechain atom B factor rmsd	1.72 Å²
Residues in Ramachandran plot ϕ-ψ regions[g]	
Most favored	86.9%
Additionally allowed	12.7%
Generously allowed	0.4%
Disallowed	0%

[a] Values in parentheses are for the highest-resolution bin.
[b] $R_{sym} = \Sigma_h \Sigma_i |I_i(h) - <I(h)>| / \Sigma_h \Sigma_i I_i(h)$
[c] Ratio is calculated as $\Sigma <|\Delta F|>/<|F|>$
[d] Value in parentheses is the figure of merit after solvent flattening.
[e] $R = \Sigma(|F_{obs}| - k|F_{calc}|)/\Sigma|F_{obs}|$
[f] R_{free} is the R value calculated for a test set of reflections, comprising a randomly selected 5% of the data that is not used during refinement.
[g] Ramachandran plot ϕ-ψ regions are defined according to the criteria of the program Procheck (Laskowski et al., 1993).

methylation of large proteins and protein components of very large macromolecular complexes.

Origins of Substrate Specificity

The SET domain packs a spectacular amount of functionality—AdoMet binding, Lys side chain binding, and catalysis—into just 110 amino acid residues. It appears that the SET domain is too small to contain the determinants to recognize the three-dimensional surface of a large protein or multiprotein assembly. We modeled the interaction of Rubisco LSMT and spinach Rubisco (Figure 7). We docked a Rubisco octamer (4 small subunits and 4 large subunits, half of the physiological hexadecamer; PDB file 1RXO) to Rubisco LSMT such that Lys-14 approached the HEPES binding site and collisions were avoided. Rubisco completely fills up the 30 Å wide gap between the N- and C-terminal lobes of Rubisco LSMT. The Rubisco large subunit contacts the SET and iSET regions, but forms much more extensive interac-

tions with the C-terminal lobe. The N-terminal tail of Rubisco is predicted to bind in a pocket formed by the SET and iSET domains and the C-terminal helix of the C-lobe.

Based on the docking, we believe that substrate recognition occurs on two levels. The first level concerns the local sequence context of a particular Lys residue within the tail of a histone or Rubisco. The N and C termini of the SET domain, and the break point at which the iSET region is inserted, are all near the active site. We predict that most of this specificity is encoded by regions outside but immediately adjacent to the core SET domain, including the iSET and cSET regions. All SET domain enzymes contain iSET insertions, although the various iSET regions are not all homologous to each other. The iSET regions range in size from 15 to 361 amino acids (Figure 3). All SET domain enzymes have C-terminal extensions. The structure shows that the region immediately following the SET domain is essential for closing one side of the substrate Lys binding site.

Because of the lack of sequence similarity between these sequences, we predict that the closure of this site may be achieved in different ways in different SET enzymes.

The second level of recognition concerns the presentation of the substrate peptide sequence in the context of an intact protein or macromolecular assembly. In the Rubisco/LSMT model, extensive contacts are formed between the C-terminal lobe of Rubisco LSMT and regions of Rubisco up to 50 Å away from the LSMT active site. Since some SET domain enzymes, such as SET1 (Miller et al., 2001; Roguev et al., 2001), only function as members of very large macromolecular complexes, and in turn recognize substrates that are also incorporated within similar large complexes, these principles are likely to apply more generally.

Conclusions

In summary, the pea LSMT structure provides a three-dimensional framework for understanding specificity and catalysis in the SET domain enzyme family. The structure explains the relationships between conserved sequences and links all members of the family, and explains the universal presence of C-terminal extensions following SET domains. We have been able to propose a catalytic mechanism that facilitates substrate and cofactor binding in the context of a large macromolecular assembly. The availability of a structure that mimics a bisubstrate complex will boost inhibitor design for the future use of chemical biology approaches to probe SET domain function in vivo. Thirteen different SET domains have been implicated so far in either positive or negative roles in cancer (Schneider et al., 2002), and this and subsequent structures of SET domain enzymes bound to substrates and cofactors should be valuable guides for anticancer drug design. Finally, the structure provides a starting point for understanding the interplay between SET and other domains in determining peptide sequence and protein substrate specificity.

Experimental Procedures

Protein Expression, Purification, and Mutagenesis

The DNA sequence encoding residues 46–482 of pea Rubisco LSMT was PCR amplified and cloned into the GATEWAY destination vector pDEST14 (Invitrogen) with a C-terminal hexahistidine-tag and TeV protease cleavage site. The plasmid was transformed into *Escherichia coli* strain BL21(DE3) Codon Plus RIL (Stratagene) and overexpressed by induction with 0.5 mM IPTG and growth at 16°C overnight. Cells were lysed by sonication in 50 mM NaPhosphate (pH 7.0), 500 mM NaCl, and 5 mM β-mercaptoethanol (βME), and LSMT was isolated using Talon Co²⁺ affinity chromatography (Clontech) with a gradient of 0 to 500 mM imidazole (pH 7.0). The peak fractions containing the enzyme were pooled, and the hexahistidine-tag was removed by overnight digestion with TeV protease at 4°C by dialysis in 50 mM TRIS (pH 8.0), 150 mM NaCl, 0.5 mM EDTA, and 5 mM dithiothreitol. LSMT was further purified with a second Talon affinity column and a Superdex S-200 gel filtration column (Amersham-Pharmacia) in 20 mM TRIS (pH 8.0), 150 mM NaCl, and 10 mM βME, and was judged to be at least 99% pure by SDS-PAGE. The protein was then concentrated to approximately 30 mg/ml, flash-frozen in liquid nitrogen, and stored at −80°C. The DNA coding for human SET9 (residues 1–366) was amplified from a human HeLa cell cDNA library (Clontech, 7111-1). It was cloned into the pHIS parallel-2 (Sheffield et al., 1999) vector using EcoRI and BamHI sites. Protein was expressed as above except at 20°C. Mutants were made using the QuikChange site-directed mutagenesis kit (Stratagene). Wild-

type and mutant full-length SET9 constructs were expressed and purified as N-terminal hexahistindine fusions using a similar protocol as for LSMT.

Rubisco LSMT and SET9 Activity Assays and Kinetic Analyses

Assays were performed as previously described (Zheng et al., 1998) using Rubisco purified from spinach leaves according to McCurry et al. (1982) and commercially available calf thymus Histone H3 (Roche Molecular Biochemicals). AdoMet was obtained from Sigma and purified prior to use as described by Chirpich et al. (1970) for Rubisco LSMT assays, or enzymatically synthesized using AdoMet synthetase (Markham et al., 1980) kindly provided by Dr. George Markham. Radiolabeled [³H-*methyl*]AdoMet (~80 Ci/mmol) was obtained from Amersham Pharmacia Biotech. Commercial preparations of AdoMet were corrected for diastereomeric purity (Hoffman, 1986). Rubisco LSMT assays were performed at 30°C for 1–2 min, and SET9 assays at 37°C for 0.08–2 min. Both assays were optimized to limit substrate consumption to 5% or less and were linear with time and enzyme concentration (0.12 to 1.2 μM). Assays were terminated by the addition of 10% TCA following addition of a 100-fold molar excess of unlabeled AdoMet. At low protein substrate levels (≤5 μg Rubisco) and at all levels of Histone H3, 5 μl of 1% Na deoxycholate was added to facilitate protein precipitation. The effect of pH on enzyme activity was determined at saturating substrate levels (3–5 fold above K_m) in a buffer consisting of 25 mM MES, 25 mM CHES, and 50 mM Bis-Tris-Propane adjusted to the appropriate pH. Enzyme activity was plotted versus substrate concentration and the data fitted to the Michaelis-Menten equation using Sigma Plot version 4.0.

Crystallization and Data Collection

Cocrystals of pea Rubisco LSMT complexed to AdoHcy were obtained with a protein concentration of 10 mg/ml in the presence of 400 μM *S*-adenosylhomocysteine (Sigma) at 25°C using the hanging drop method containing a reservoir solution of 100 mM HEPES (pH 6.8) and 1.2 M to 1.35 M NaAcetate. Crystals appeared in 2 to 3 days and grew to average dimensions of 200 μM × 300 μM × 300 μM in 1 to 2 weeks. Crystals were partially dehydrated by equilibration over mother liquor with increasing concentrations of NaAcetate in successive 150 mM steps to a final concentration of 2.4 M NaAcetate. Crystals were cryo-protected in 30% glycerol and flash-frozen in the cryo-stream at 95 K. Xenon-derivatized crystals were prepared in an Xcell pressure chamber (Oxford Cryosystems) using a xenon pressure of 20 bar for 5 min; after depressurization, crystals were immediately flash-frozen in propane (Soltis et al., 1997). Both native and derivative data sets were collected at 95 K using an R-AXIS IV++ image plate detector (Rigaku) using Cu Kα radiation produced by a RU200 X-ray generator (Rigaku) and focused with Osmic confocal mirrors (Rigaku). Data were indexed and reduced using DENZO and scaled using SCALEPACK (HKL Research; Otwinowski and Minor, 1997).

Structural Solution and Refinement

Phases were calculated from the xenon SAD dataset to 3.4 Å in SOLVE (Terwilliger and Berendzen, 1996). Seven xenon sites were found in the asymmetric unit and were input into SHARP for maximum likelihood phase refinement (de La Fortelle and Bricogne, 1997). The resulting map was solvent-flattened using the automated solvent content optimization in SHARP and yielded a traceable electron density map with defined side chain density. An inspection of the maps in the model building program O (Jones et al., 1991) revealed three molecules in the asymmetric unit, which were subsequently built by tracing the chain of the first monomer and using the noncrystallographic symmetry to transpose this model into the density for the remaining two monomers. Once the initial model building was completed, the structure was refined against the Xe derivative data using CNS (Brünger et al., 1998) at 3.0 Å using positional and torsion angle dynamics, and B factor refinement with a maximum likelihood target and noncrystallographic (NCS) restraints set on the iSET region and the C-terminal lobe of the monomers. After the first round of refinement, the resulting structural coordinates were transposed into the native unit cell and refined in CNS

at 2.6 Å. Further rounds of model building revealed density in the active site for AdoHcy and a single molecule of HEPES, which was consistent with the initial maps generated in SHARP. Residues 228–230 and 257–266 of the A molecule were excluded from the model due to their disorder, and 666 water molecules were added. The model has a working R factor of 23.2% and a free R factor of 27.8%. None of the nonglycine residues are in the disallowed region of the Ramachandran plot (Table 1).

References

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Boggs, B.A., Cheung, P., Heard, E., Spector, D.L., Chinault, A.C., and Allis, C.D. (2001). Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat. Genet. *30*, 73–76.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y.R., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by SET1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. Genes Dev. *15*, 3286–3295.

Bryk, M., Briggs, S.D., Strahl, B.D., Curcio, M.J., Allis, C.D., and Winston, F. (2002). Evidence that SET1, a factor required for methylation of histone H3, regulates rDNA silencing in *S. cerevisiae* by a Sir2-independent mechanism. Curr. Biol. *12*, 165–170.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software system for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

Chirpich, T.P., Zappia, V., Costilow, R.N., and Barker, H.A. (1970). Lysine 2, 3 aminomutase. Purification and properties of a pyridoxal phosphate and *S*-adenosylmethionine-activated enzyme. J. Biol. Chem. *245*, 1778–1789.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.

Durban, E., Nochumson, S., Kim, S., Paik, W.K., and Chan, S.K. (1978). Cytochrome c-specific protein-lysine methyltransferase from *Neurospora crassa*. J. Biol. Chem. *253*, 1427–1435.

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.

Hoffman, J.L. (1986). Chromatographic analysis of the chiral and covalent instability of *S*-adenosyl-L-methionine. Biochemistry *25*, 4444–4449.

Houtz, R.L., Royer, M., and Salvucci, M.E. (1991). Partial purification and characterization of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit εN-methyltransferase. Plant Physiol. *97*, 913–920.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556–560.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chro-

modomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080–2083.

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

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Klein, R.R., and Houtz, R.L. (1995). Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase. Plant Mol. Biol. *27*, 249–261.

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

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Lachner, M., and Jenuwein, T. (2002). The many faces of histone lysine methylation. Curr. Opin. Cell Biol. *14*, 286–298.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science *293*, 2453–2455.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: As domains continue to swap. Protein Sci. *11*, 1285–1299.

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 Å resolution. Nature *389*, 251–260.

Markham, G.D., Hafner, E.W., Tabor, C.W., and Tabor, H. (1980). *S*-Adenosylmethionine synthetase from *Escherichia coli*. J. Biol. Chem. *255*, 9082–9092.

McCurry, S.D., Gee, R., and Tolbert, N.E. (1982). Ribulose-1,5-bisphosphate carboxylase/oxygenase from Spinach, Tomato or Tobacco leaves. Methods Enzymol. *90*, 515–521.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D version 2.0: A program for photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, S. (2001). COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. USA *98*, 12902–12907.

Murray, K. (1964). The occurrence of ε-N-methyl lysine in histones. Biochemistry *3*, 10–15.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110–113.

Nielsen, S.J., Schneider, R., Bauer, U.-M., Bannister, A.J., Morrison, A., O'Carroll, D.O., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R.E., and Kouzarides, T. (2001). Rb targets histone H3 methylation and HP1 to promoters. Nature *412*, 561–565.

Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103–107.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. (2002a). SET9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev. *16*, 479–489.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., et al. (2002b). 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.

Noma, K., Allis, C.D., and Grewal, S.I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science *293*, 1150–1155.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Peters, A.H.F.M., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323–337.

Rea, S., Eisenhaber, R., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W.M.P., 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.

Rojas, J.R., Trievel, R.C., Zhou, J.X., Mo, Y., Li, X.M., Berger, S.L., Allis, C.D., and Marmorstein, R. (1999). Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. Nature *401*, 93–98.

Rowe, P.M., Wright, L.S., and Siegel, F.L. (1986). Calmodulin N-methyltransferase. J. Biol. Chem. *261*, 7060–7069.

Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. J. Mol. Biol. *247*, 16–20.

Schneider, G., Lindqvist, Y., and Branden, C.-I. (1992). RUBISCO: structure and mechanism. Annu. Rev. Biophys. Biomol. Struct. *21*, 119–143.

Schneider, R., Bannister, A.J., and Kouzarides, T. (2002). Unsafe SETs: histone lysine methyltransferases and cancer. Trends Biochem. Sci. *27*, 396–402.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. Proc. Natl. Acad. Sci. USA *95*, 5857–5864.

Schultz, D.C., Ayyananthan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., III. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. *16*, 919–932.

Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of parallel expression vectors. Protein Expr. Purif. *15*, 34–39.

Soltis, S.M., Stowell, M.H.B., Wiener, M.C., Phillips, G.N., Jr., and Rees, D.C. (1997). Successful flash-cooling of xenon-derivatized myoglobin crystals. J. Appl. Crystallogr. *30*, 190–194.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.-W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2002). SET2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol. Cell. Biol. *22*, 1298–1306.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). SET domain containing protein, G9A, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J. Biol. Chem. *276*, 25309–25317.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature *414*, 277–283.

Terwilliger, T.C., and Berendzen, J. (1996). Correlated phasing of multiple isomorphous replacement data. Acta Crystallogr. D *52*, 749–757.

Wang, P., Royer, M., and Houtz, R.L. (1995). Affinity purification of Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit ϵN-methyltransferase. Protein Expr. Purif. *6*, 528–536.

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4 specific methyltransferase. Mol. Cell *8*, 1207–1217.

Weiss, V.H., McBride, A.E., Soriano, M.A., Filman, D.J., Silver, P.A., and Hogle, J.M. (2000). The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. Nat. Struct. Biol. *7*, 1165–1171.

Yang, L., Xia, L., Wu, D.Y., Wang, H., Chansky, H.A., Schubach, W.H., Hickstein, D.D., and Zhang, Y. (2002). Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. Oncogene *21*, 148–152.

Ying, Z., Mulligan, R.M., Janney, N., and Houtz, R.L. (1999). Rubisco small and large subunit N-methyltransferases. Bi- and mono-functional methyltransferases that methylate the small and large subunits of Rubisco. J. Biol. Chem. *274*, 36750–36756.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev.. *15*, 2343–2360.

Zhang, X., Zhou, L., and Cheng, X. (2000). Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. EMBO J. *19*, 3509–3519.

Zheng, Q., Simel, E.J., Klein, P.E., Royer, M.T., and Houtz, R.L. (1998). Expression, purification, and characterization of recombinant ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N$^\epsilon$-methyltransferase. Protein Expr. Purif. *14*, 104–112.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank with the accession code 1MLV.