

Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT

Raymond C Trievel¹, E Megan Flynn², Robert L Houtz² & James H Hurley¹

SET domain protein methyltransferases catalyze the transfer of methyl groups from the cofactor S-adenosylmethionine (AdoMet) to specific lysine residues of protein substrates, such as the N-terminal tails of histones H3 and H4 and the large subunit of the Rubisco holoenzyme complex. The crystal structures of pea Rubisco large subunit methyltransferase (LSMT) in ternary complexes with either lysine or ϵ -N-methyllysine (MeLys) and the product S-adenosylhomocysteine (AdoHcy) were determined to resolutions of 2.65 and 2.55 Å, respectively. The ζ -methyl group of MeLys is bound to the enzyme via carbon–oxygen hydrogen bonds that play a key role in catalysis. The methyl donor and acceptor are aligned in a linear geometry for S_N2 nucleophilic transfer of the methyl group during catalysis. Differences in hydrogen bonding between the MeLys ϵ -amino group and Rubisco LSMT and SET7/9 explain why Rubisco LSMT generates multiply methylated Lys, whereas SET7/9 generates only MeLys.

The methylation of the ϵ -amino group of lysine residues in proteins regulates a number of critical biological events. These mechanisms include the regulation of chromatin structure via methylation of specific lysines in the N-terminal tails of histones H3 and H4 throughout eukaryotes^{1–5} and regulation of Rubisco function by methylation of Lys14 in the N-terminal tail of the Rubisco large subunit in several plant species^{6,7}. With one exception, all known lysine methylation of proteins is catalyzed by specific S-adenosylmethionine-dependent methyltransferases that contain a conserved 110-amino acid motif known as the SET (Su(var) 3-9, Enhancer-of-zeste, Trithorax) domain^{8,9}. Mono-, di- and tri-methylation of lysine residues can promote the recruitment of proteins containing specific methyllysine-binding domains, such as the chromodomain^{10,11}, and can inhibit alternative post-translational modifications, such as phosphorylation and acetylation^{9,12}. Recently, much attention has focused on differential regulation by different levels of methylation at a given lysine and the effect the methylation level has on transcriptional regulation^{13,14}. Thus, the ability or inability of a SET domain enzyme to mono-, di- and tri-methylate lysine residues is likely to be critical to its regulatory function.

The central importance of SET domain methyltransferases in chromatin structure and other regulatory mechanisms has spurred intense interest in determining the structural basis for the specificity, regulation and catalytic activity of these enzymes. In recent months, the structures of the histone methyltransferases DIM-5 from *Neurospora crassa*¹⁵, human SET7/9 (refs. 16–19), *Schizosaccharomyces pombe* CLR4 (ref. 20) and a viral enzyme (vSET of *Paramecium bursaria* chlorella virus 1) have been determined²¹, as well as that of the Rubisco large subunit methyltransferase (LSMT) from the garden pea²². All of the structures reveal the presence of an all β -sheet catalytic

core flanked by a variety of divergent N- and C-terminal extensions, and insertions between the first and second conserved sequence motifs in the SET domain^{23–25}. These structures revealed an invariant tyrosine as a key catalytic residue. Four of the six structures published to date contain either a well-ordered AdoMet substrate or AdoHcy product cofactor bound in the AdoMet-binding site. A consensus model for AdoMet binding has been derived on the basis of the strong agreement between geometries observed for three distinct SET7/9 cofactor complex structures^{16,17,19} and the LSMT–AdoHcy complex²². The cofactor is bound in such a way that the methyl group of AdoMet projects into a narrow, deeply buried pocket that is the putative lysine substrate-binding cleft. Gamblin and co-workers¹⁹ have recently determined the structure of SET7/9 bound to AdoHcy and a histone H3 MeLys4-containing peptide. This structure is of a product–product complex because SET7/9 only catalyzes monomethylation of Lys4 in histone H3 (refs. 18,19).

Despite the interest in understanding the structural basis for substrate specificity and catalytic mechanism of the SET domain enzymes, no structure of a ternary SET domain enzyme–substrate–cofactor complex has been reported so far. We observed the fortuitous binding of a HEPES buffer ion in the putative lysine-binding site of LSMT²². We subsequently confirmed that HEPES is a weak competitive inhibitor of the enzyme²². This finding suggested to us that at sufficiently high concentrations, free lysine might be a substrate of the enzyme, which may provide the basis for determining the structure of a catalytically competent substrate bound to a SET domain enzyme. In addition, we reasoned that if free lysine were a substrate of LSMT, free ϵ -N-methyllysine (MeLys) might be one as well because LSMT is capable of mono-, di- or tri-methylating Lys14 in the large subunit of Rubisco^{26,27}. Indeed, we found that LSMT has enzyme activity for free

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA. ²Department of Horticulture, Plant Physiology/Biochemistry/Molecular Biology Program, University of Kentucky, Lexington, Kentucky 40546-0091, USA. Correspondence should be addressed to J.H.H. (jh8e@nih.gov)



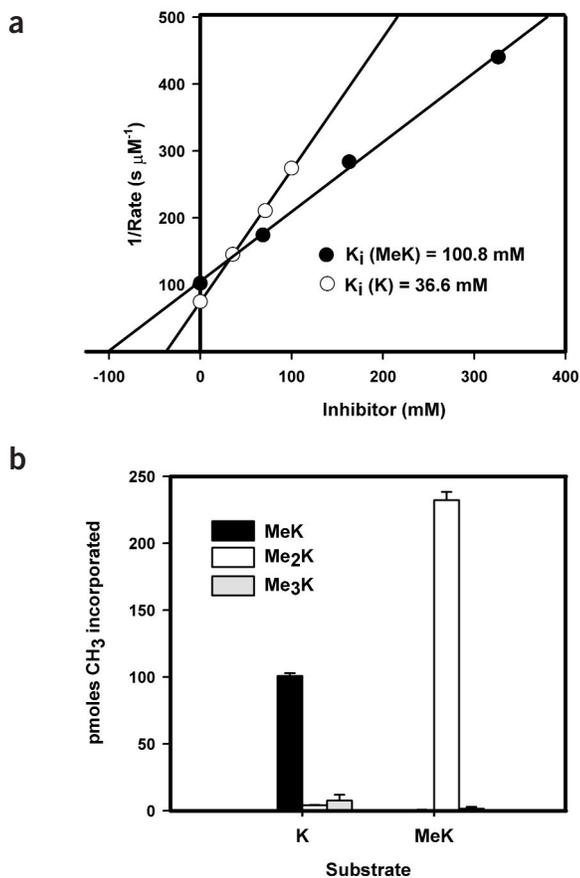


Figure 1 Lys and MeLys are alternative substrates for Rubisco LSMT. **(a)** Plot of the concentration of Lys and MeLys versus the reciprocal of the enzymatic rate. The x -intercept is equal to $-K_i$. **(b)** Incorporation of ^3H -methyl groups into Lys and MeLys. Radiolabeled AdoMet was incubated with LSMT and either Lys or MeLys as substrates. The methylated products were then separated by TLC, and the bands corresponding to each Lys compound were excised to determine the amount of ^3H -methyl incorporation.

concentrations of each Lys compound to quantify the inhibition of Rubisco methylation. Secondary replots of the reciprocal rate versus concentration of lysine and MeLys were used to determine the K_i values for Lys and MeLys, respectively (Fig. 1a). The K_i values were 37 mM for Lys and 101 mM for MeLys compared with ~700 mM estimated for HEPES (data not shown), indicating that the binding of Lys and MeLys to Rubisco LSMT is substantially stronger than that of HEPES. Because we were able to observe HEPES bound to crystals of Rubisco LSMT, we hypothesized that it would also be possible to observe Lys and MeLys in crystalline complex. The pattern of inhibition with Lys and MeLys is competitive with respect to Rubisco; therefore, we conclude that they bind to the same site.

We sought to determine whether Lys and MeLys are alternative substrates for Rubisco LSMT. Product analysis via thin layer chromatography (TLC) demonstrates that LSMT catalyzes the transfer of ^3H -methyl from radiolabeled AdoMet to Lys to generate MeLys, Me₂Lys and Me₃Lys, with MeLys having the highest level of ^3H -methyl incorporation (Fig. 1b). LSMT also catalyzed the methylation of MeLys to form Me₂Lys and Me₃Lys. MeLys was converted to Me₂Lys to a larger extent (about two-fold) than Lys to MeLys, even though the apparent K_m for MeLys is nearly three-fold higher than that for Lys. Me₂Lys was also examined as a potential substrate and as an inhibitor in the presence of Rubisco, but virtually no activity as a substrate or inhibitor was observed even at 300 mM (data not shown). The k_{cat} values for Lys ($6.2 \times 10^{-5} \text{ s}^{-1}$) and MeLys ($2.5 \times 10^{-4} \text{ s}^{-1}$) are small compared to Rubisco (0.047 s^{-1})²², which is consistent with the expectation that Rubisco would be preferred to these non-physiological substrates. For either random or ordered reaction mechanisms, an alternate substrate acts as a competitive inhibitor with respect to the appearance of product of the normal substrate²⁸. Therefore, the K_i values derived from the x -intercepts of these plots are equivalent to the K_m for Lys or MeLys

MeLys and determined the structure of this complex as well. These structures of substrate complexes reveal snapshots of the methyl transfer reaction in a SET domain enzyme.

RESULTS

Lys and MeLys are LSMT substrates

To determine if Lys and MeLys compete for binding to the same site as Rubisco, enzyme assays were performed in the presence of increasing

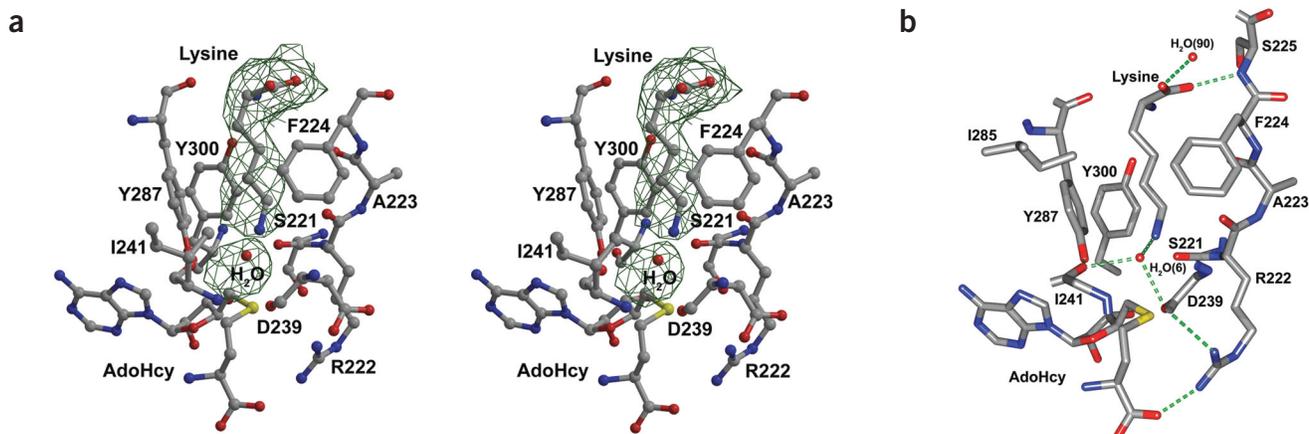


Figure 2 Binding of Lys in the lysine-binding pocket of LSMT. **(a)** Stereo view of the simulated-annealing $F_o - F_c$ electron density omit map of the lysine-binding cleft calculated in the absence of Lys and water molecules. Several residues constituting the binding cleft are shown, as well as the product AdoHcy. The map is contoured at 2.2σ . Electron density map figures were rendered in BobScript⁴⁴⁻⁴⁶ and Raster3D⁴⁴. **(b)** Lys bound in the lysine-binding pocket of LMST. Residues and water molecules interacting with Lys are illustrated, and hydrogen bonds are denoted with dashed green lines. Active site models were rendered using Swiss-Pdb Viewer⁴⁷ and POV-Ray (http://www.povray.org).

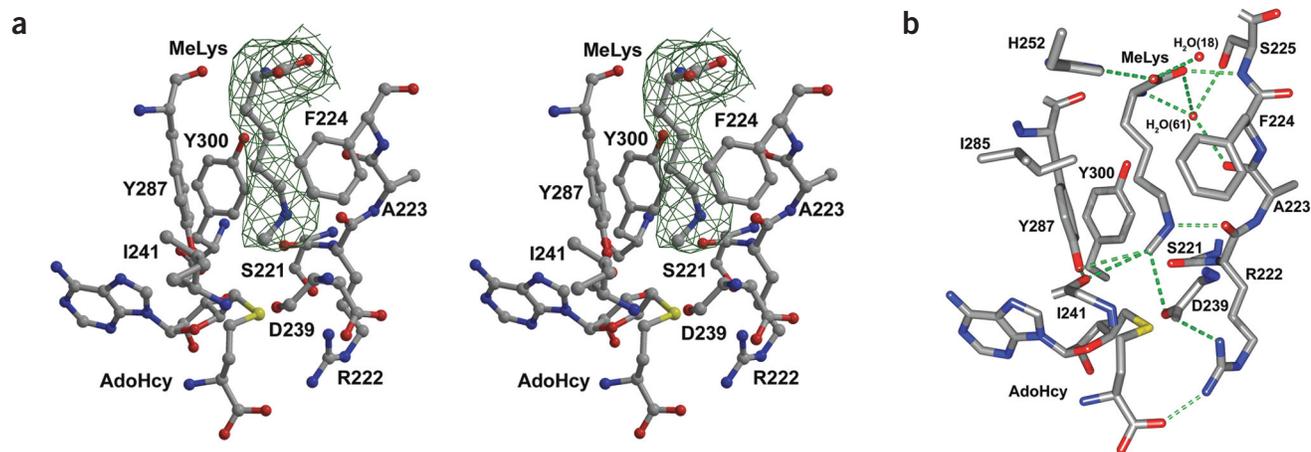


Figure 3 Binding of MeLys in the lysine-binding pocket of LSMT. (a) Stereo view of the simulated-annealing $F_o - F_c$ electron density omit map of the lysine-binding cleft calculated in the absence of MeLys and water molecules. Several residues constituting the binding cleft are shown, as well as the product AdoHcy. The map is contoured at 2.2σ . (b) MeLys bound in the lysine-binding pocket of LMST. Residues and water molecules interacting with MeLys are illustrated, and hydrogen bonds are rendered as dashed green lines.

as a substrate for LSMT²⁸. The kinetic characterization of LSMT with respect to Lys and MeLys demonstrates that they are the minimal competent substrates recognized by LSMT.

Structure of the lysine complex

Crystals of the lysine-AdoHcy-LSMT and MeLys-AdoHcy-LSMT complexes were grown using crystallization solutions buffered with 100 mM of either Lys acetate or MeLys acetate in place of the HEPES buffer used in the crystallization of the AdoHcy-LSMT binary complex²². The structure of the lysine-AdoHcy-LSMT ternary complex was determined using Fourier difference analysis based on the previously determined substrate-free Rubisco LSMT structure with the HEPES and water molecules removed. There are three molecules in the asymmetric unit in these crystals. In one of the molecules (A), the Lys is relatively mobile. In the other two molecules, the Lys side chain is in well-defined electron density (Fig. 2a). The average B -factor of the bound Lys is 58 \AA^2 , which is substantially lower than that observed for HEPES (average B -factor 77 \AA^2) bound in the active site.

The free lysine molecule is bound in a deep pocket composed of residues $\beta 6$ – $\beta 12$ of the SET domain, the C-terminal section of the iSET region (residues 111–220) and the cSET region (residues 290–305). The Lys-binding site corresponds closely to the locus of HEPES binding²². The carboxylate of Lys, which would correspond to the C-terminal portion of the lysyl residue in a polypeptide chain, interacts with the main chain amide of Ser225 and a water molecule (S90), whereas the α -amino group hydrogen bonds to the carbonyl oxygen of Tyr287 (Fig. 2b). The side chain is in a completely extended conformation, with all dihedral angles in *trans* conformations. The aliphatic portion of the Lys side chain interacts with the hydrophobic residues Phe224, Ile285, Tyr287 and Tyr300. The side chain $N\zeta$ forms a short hydrogen bond (2.5 \AA) with a water molecule (S6); this water molecule in turn hydrogen bonds with the main chain carbonyl

groups of Asp239 and Ile241 (Fig. 2b). The water molecule is bound tightly to the ϵ -amino and carbonyl groups, with a B -factor of 37 \AA^2 . These carbonyl groups, along with the carbonyls of Ser221 and Arg222, ring the center of the active site, and we refer to them as the ‘carbonyl cage’²².

Structure of the MeLys complex

The overall structure of the MeLys-AdoHcy-LSMT ternary complex is similar to the Lys ternary complex, with an r.m.s. deviation of 0.86 \AA for all protein atoms. The conformations of the residues in the active site are essentially identical to the complex containing Lys (Fig. 3a), and the MeLys has an average B -factor of 54 \AA^2 , which is similar to the value for the bound Lys. The entire MeLys molecule, including the amino acid moiety, is shifted by $\sim 1 \text{ \AA}$ away from the center of the active site. The outward shift accommodates the bulk of the methyl group and permits the α -carboxylate to make a hydrogen bond with the side chain of His252 (Fig. 3b). The α -carboxylate group also makes hydrogen bond contacts with the amide of Ser225 and a water molecule (S18), while the α -amino group hydrogen bonds to the carbonyl

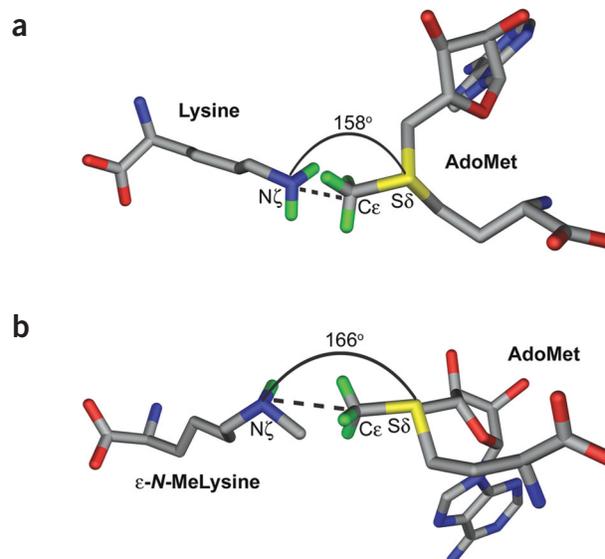


Figure 4 Stereochemical mechanism of methyl group transfer. (a) Model of the lysine-AdoMet substrate complex on the basis of the LSMT SET-Lys-AdoHcy complex. The formation of the nascent bond between the deprotonated $N\zeta$ of the lysine and the $C\epsilon$ methyl group of AdoMet is denoted with a dashed line, and the geometry of S_N2 reaction is shown with a black arc. (b) Model of the MeLys-AdoMet substrate complex based on the LSMT SET-MeLys-AdoHcy complex. The figure is labeled as in a.

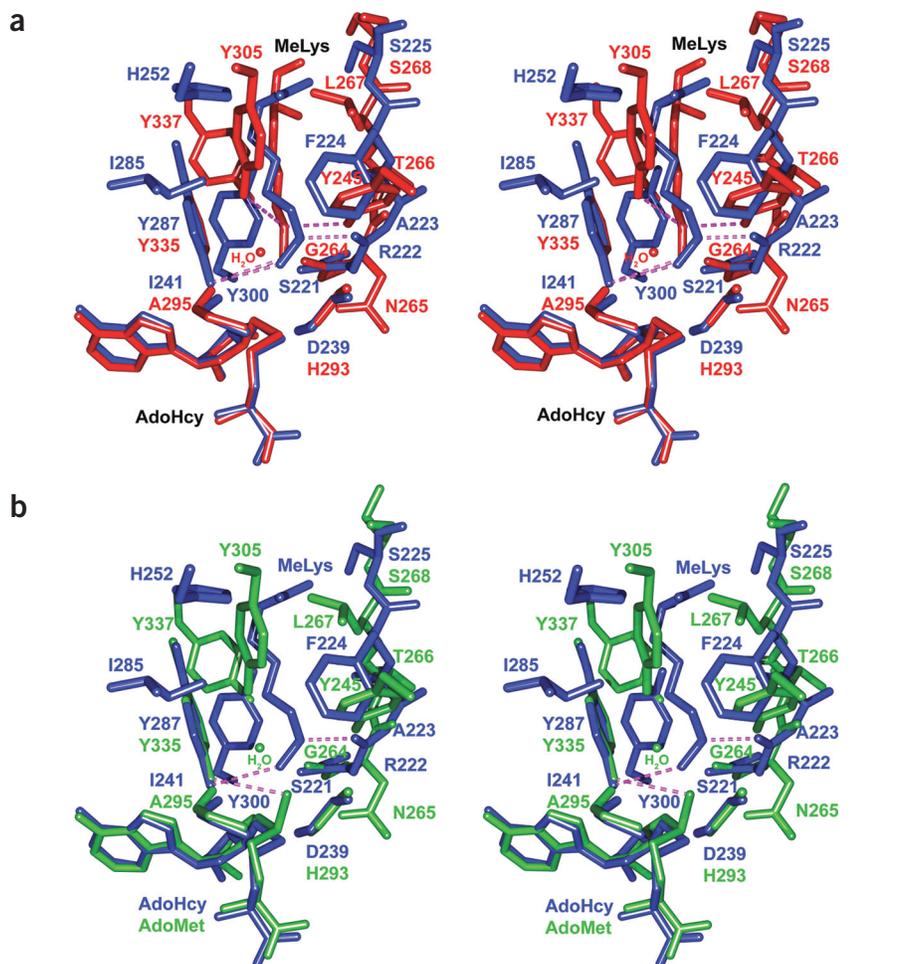


Figure 5 Comparison of hydrogen bonding in the active sites of LSMT and SET7/9. **(a)** Stereo view of the superimposition of the lysine-binding clefts of LSMT (blue) and the SET7/9-histone H3 MeLys4-AdoHcy ternary complex (PDB entry 1O9S) (red). Hydrogen bonds between the protein and ϵ -amino groups, and carbon-oxygen hydrogen bonds between the invariant Tyr and methyl groups are illustrated with dashed magenta lines. **(b)** Stereo view of the superimposition of the lysine-binding clefts of LSMT (blue) and the SET7/9-AdoMet complex (PDB entry 1N6A) (green). Hydrogen bonds are illustrated as in **a**.

oxygen of Tyr287. In addition, both the α -amino and carboxylate groups make water-mediated (S61) hydrogen bonds to the carbonyl of Ala223 and the side chain hydroxyl of Ser225. The most significant difference from the Lys complex is that the ϵ -amino group does not make water-mediated hydrogen bonds with the carbonyls of Asp239 and Ile241, as observed in the Lys complex; instead, N ζ makes a direct 2.5 Å hydrogen bond with the carbonyl of Arg222.

DISCUSSION

Carbon-oxygen hydrogen bonds

The MeLys ζ -methyl moiety displaces the S6 water molecule in the lysine complex and interacts with the same surrounding polar groups (Fig. 5a,b). It is remarkable that this methyl group makes no hydrophobic interactions. The methyl moiety of MeLys is positioned within 3.3 Å of the invariant Tyr287 hydroxyl group and 3.6 Å and 3.4 Å of the carbonyl oxygens of Asp239 and Ile241, respectively. These C \cdots O bond distances are shorter than typical van der Waals interactions, which are generally ≥ 3.7 Å between aliphatic carbon and oxygen atoms when the distance is measured along the C-H bond²⁹.

Derewenda and colleagues²⁹ have noted the ability of methyl groups to serve as hydrogen bond donors in proteins, and subsequent analyses have revealed many instances where carbon-oxygen hydrogen bonds are important for protein stability and function^{30–35}. In addition, carbon-oxygen hydrogen bonds can play important roles in enzyme catalysis, such as in the catalytic triad of serine hydrolases³⁶ and in the catalytic glutamate of acyl-CoA dehydrogenases³⁷. The observed short distances and favorable angles between the ζ -methyl and neighboring oxygen atoms, the pattern in which three oxygens interact with the methyl in tetrahedral geometry, the lack of hydrophobic interactions around the methyl group and the electropositive character of the ζ -methyl suggest that carbon-oxygen hydrogen bonds coordinate the methyl group in the carbonyl cage. The electronegativity of the ϵ -amino group of MeLys gives the ζ -methyl group a partial positive charge, which is consistent with the acidification of the hydrogens of the methyl groups and their engagement in weak hydrogen bonding with the oxygens of the carbonyl cage and the invariant Tyr287. Quantum mechanical modeling suggests that carbon-oxygen hydrogen bonds in Lys residues can be especially strong and have interaction energies comparable to classical hydrogen bonds because of the polarization effect of the ϵ -amino moiety³⁵.

These interactions also occur in other SET domains, although they have not previously been discussed as such in the literature. In the SET7/9 ternary complex (PDB entry 1O9S), there are carbon-oxygen hydrogen bonds between the MeLys methyl group and the hydroxyl of the invariant Tyr335 (3.4 Å), the carbonyl group of Gly264 (3.3 Å) and the Asn265 side chain amide oxygen (3.6 Å)¹⁹.

Furthermore, carbon-oxygen hydrogen bonds are also present between the methyl group of AdoMet and the hydroxyl of Tyr335 (3.5 Å), the carbonyl of His293 (3.0 Å) and the side chain amide of Asn265 (3.1 Å) in the 1.7-Å resolution crystal structure of SET7/9 bound to AdoMet (PDB entry 1N6A)¹⁷. The latter structure shows that carbon-oxygen hydrogen bonds coordinate the methyl group before transfer. We propose that a major, and perhaps the primary, function of the invariant Tyr and the carbonyl cage in the SET domain is to form carbon-oxygen hydrogen bonds with the methyl group, activating it for transfer.

Mechanism of methyl transfer

Both Lys and MeLys are substrates of LSMT. Here we consider the implications for the catalytic mechanism of methyl group transfer to both a Lys and MeLys substrate. The ϵ -amino of the Lys points toward the AdoHcy thioether S atom located 3.8 Å away (Fig. 4a). This geometry allows sufficient room to model the methyl group on AdoMet by adding it to the structure of AdoHcy with appropriate stereochemistry using the known structure of AdoMet. The ϵ -amino group of Lys

points toward the AdoMet methyl group, rather than Tyr287. The structure of Lys bound in the complex is congruent with the requirements for a direct in-line S_N2 nucleophilic attack. The angle between the $N\zeta$, $C\epsilon$, and $S\delta$ atoms is 157° in the Lys complex, whereas the angle is 166° between these atoms in the MeLys complex (Fig. 4a,b). Given an experimental uncertainty of ~ 10 – 20° , these bond angles correlate well with the value of 180° expected theoretically for a linear S_N2 transfer mechanism³⁸. The Lys complex structure presents the most direct evidence so far that SET domain enzymes act through a direct in-line S_N2 nucleophilic attack.

MeLys is both a substrate and product of the LSMT reaction. The observed MeLys complex reflects features of both substrate- and product-like complexes. The MeLys ϵ -amino group is 4.8 \AA away from the AdoHcy thioether S atom, which is 1 \AA farther than in the Lys complex. The angle between the $N\zeta$, $C\epsilon$ and $S\delta$ atoms (Fig. 4b) is 166° in the MeLys complex, which is appropriate for methyl transfer. However, the ζ -methyl group partially occludes the pathway for in-line attack, and a small shift of this group is predicted to occur in the reactive complex with AdoMet.

Only the deprotonated Lys substrate has a free lone pair capable of nucleophilic attack on the AdoMet methyl group. We and other groups had suggested that the invariant Tyr could be the base responsible for deprotonating the ϵ -ammonium group prior to methyltransfer^{15–17,22}. However, a close examination of the hydrogen bonding patterns in the active site does not support this concept. In the Lys-bound complex, the ϵ -amino group of the substrate Lys and Tyr287 hydroxyl are separated by $>3.3 \text{ \AA}$, indicating weak or no hydrogen bonding between the two moieties. In the MeLys complex, the ϵ -amino is $>4.3 \text{ \AA}$ from the Tyr hydroxyl and instead is tightly hydrogen bound to the carbonyl of Arg222. The lack of hydrogen bonding between Lys and MeLys and Tyr287 is inconsistent with the function of this residue in base catalysis in SET domain methyltransferases. The absence of a general base implies that a deprotonated Lys residue is the true substrate of SET domain enzymes. In this model, the Lys is deprotonated in solution before binding the enzyme. Because deprotonated Lys is a minor species in solution at $\text{pH} < 10$, this model would be consistent with the exceptionally low turnover numbers and high pH optima for SET domain enzymes.

Methylation multiplicity

LSMT is capable of trimethylating Lys14 of the Rubisco large subunit and can catalyze the methylations of free Lys and MeLys. However, not all SET domain enzymes are capable of catalyzing multiple methylations. The human histone Lys methyltransferase (HKMT) SET7/9 catalyzes only the monomethylation of Lys4 in histone H3 (refs. 18,19). We superimpose the SET7/9–MeLys peptide complex with the present structures in order to ascertain why some SET domain enzymes are capable of multiple methylations while others are not. The most informative comparison is between the LSMT complex with MeLys substrate complex and the ternary product complex of SET7/9 bound to AdoHcy and a histone H3 MeLys4 peptide.

A superimposition of the LSMT and SET7/9 lysine-binding clefts reveals several structural differences that can account for the different methylation specificities of each enzyme (Fig. 5). Many of the structural elements in the LSMT lysine-binding pocket, such as the carbonyl cage, β_6 strand and invariant Tyr287, are conserved in SET7/9. However, there are several significant differences between the lysine-binding pockets. The most notable is the presence of Tyr245 and Tyr305 in SET7/9, which hydrogen bond to the ϵ -amino group of the MeLys residue in the ternary complex. Phe224, His252 and Ile285 occupy comparable positions in the lysine-binding cleft of LSMT. In

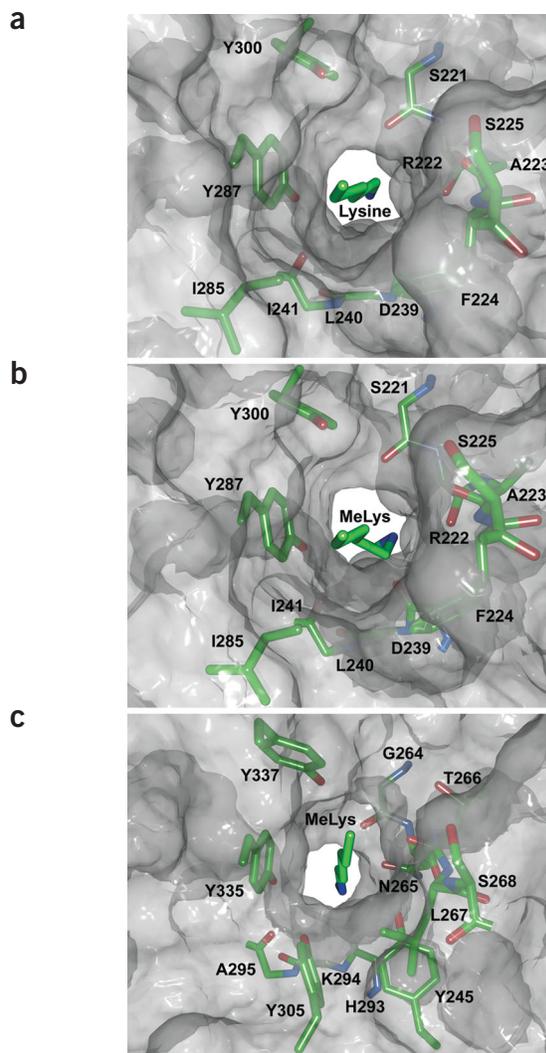


Figure 6 Space filling comparison of the lysine-binding sites of LSMT and SET7/9. (a) Molecular surface of the lysine-binding cleft in the LSMT SET–Lys–AdoHcy complex. The side chain of Arg 226 and the backbone atoms of the lysine substrate were removed for clarity. (b) Molecular surface of the lysine-binding pocket in the LSMT SET–MeLys–AdoHcy complex, depicted as in a. (c) Molecular surface of the lysine-binding cleft in the SET7/9 ternary complex, illustrated as in a.

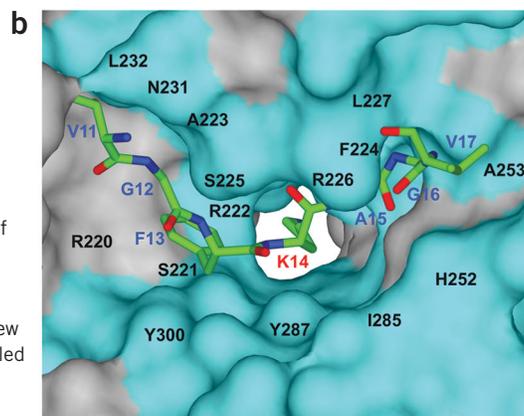
SET7/9, the side chains of Tyr245, Tyr305 and Leu267 protrude farther into the lysine-binding pocket than the comparable residues in LSMT. This intrusion narrows the diameter of the lysine-binding pocket in SET7/9 in comparison to LSMT (Fig. 6). As a consequence, MeLys4 binds in an all-*trans* extended conformation to SET7/9, whereas the MeLys binds to LSMT in a kinked conformation.

The ϵ -amino groups of the MeLys side chains bound in the LSMT and SET7/9 ternary complexes have completely different hydrogen bonding interactions. After methyltransfer occurs in SET7/9, hydrogen bonds between ϵ -amino group and the hydroxyl moieties of Tyr245 and Tyr305 orient the $N\zeta$ and $C\eta$ atoms of MeLys and the $S\delta$ thioether atom of AdoHcy in a nearly linear geometry of 177° . This conformation directly blocks both the binding of AdoMet and the accessibility of the ϵ -amino group for a second round of methyl transfer, as noted by Gamblin and co-workers¹⁹. The strong hydrogen bonds with Tyr245 and Tyr305 lock the MeLys side chain χ_2 angle into

a Rubisco Large Subunit N-Terminal Sequences

Arabidopsis thaliana: Ac-PQTETKASVGF_{K14}AGVKEYKLTYY
 Pea (*Pisum sativum*): Ac-PQTETKAKVGF_{K14}AGVKDYKLTYY
 Tobacco (*Nicotiana tabacum*): Ac-PQTETKASVGF_{K14}AGVKEYKLTYY
 Spinach (*Spinacia oleracea*): Ac-PQTETKASVGF_{K14}AGVKDYKLTYY
 Consensus Sequence: VGF_{K14}AGV

Figure 7 Substrate specificity of LSMT. (a) Sequence alignment of the N-terminal residues of the large subunits of Rubisco from several plant species. The N termini begin at Pro3, which are N α acetylated. The consensus recognition sequence for methylation of Lys14 by LSMT is shown in blue. (b) Modeling of the consensus sequence peptide into the protein substrate-binding cleft in LSMT. The side chain of Arg 226 was removed to provide an unobstructed view of the binding site. Residues that are within van der Waals contact of the peptide model labeled on the molecular surface of LSMT. Molecular surfaces associated with residues that are identical in LSMT sequences are colored cyan.



a *trans* geometry. This stabilizes the product complex in a non-productive orientation. The structural comparison presented here helps explain why the Y245A mutation permits SET7/9 to catalyze the di- and trimethylation of Lys4 in histone H3 (ref. 19).

In LSMT, the ϵ -amino group of MeLys hydrogen bonds to the carbonyl oxygen of Arg222, which positions it 1 Å closer to the reactive site than the ϵ -amino of SET7/9-bound MeLys. This configuration orients the substrate MeLys N ζ and C η atoms and the AdoHcy S δ atom into a non-linear geometry with an angle of 121° (Fig. 4b), rotating the methyl away from the line of nucleophilic attack as compared with the nearly straight-line geometry of the SET7/9 complex. In summary, a combination of steric constriction in the SET7/9 active site as compared with LSMT and hydrogen bonds between the ϵ -amino group and Tyr245 and Tyr305 prevent SET7/9 from carrying out multiple methylations. LSMT is capable of carrying out multiple Lys methylations because its active site is less constricted and because hydrogen-bonding groups are available to position the ϵ -amino group in a reactive conformation.

Methylation site specificity

The basis for the methylation site specificity in the SET domain histone methyltransferases has been grounds for intense interest because site specificity is at the heart of the 'histone code'. Similar to most of the characterized HKMTs, LSMT is exquisitely specific for Lys14 in the large subunit of the Rubisco holoenzyme^{26,27}. A sequence alignment of several Rubisco large subunit homologs reveals a consensus sequence VGF_{K14}AGV (in which K₁₄ is the methylation site) (Fig. 7a), which is relatively solvent exposed and lacks defined secondary structure in the Rubisco holoenzyme³⁹. The Lys and MeLys α -amino and -carboxylate groups hydrogen bond to residues in the lysine-binding cleft and neighboring water molecules, providing a basis for modeling a consensus sequence peptide into the LSMT active site (Fig. 7b).

The peptide was manually docked in an extended conformation and makes numerous backbone interactions in the LSMT protein substrate-binding site. The N- to C-terminal direction of the peptide is the same as in the SET7/9 complex (not shown). Three out of seven positions in the consensus methylation sequence of Rubisco are Gly or Ala. Consistent with the preponderance of small residues in the substrate sequence, the peptide-binding groove contains few pockets capable of binding a large side chain. Thus, one aspect of specificity is the steric exclusion of residues with medium or large side chains at these positions. The exception to this observation is Phe13, which can bind in a deep hydrophobic pocket immediately adjacent to the lysine-binding cleft in LSMT. LSMT binds to the large subunit with

high affinity ($K_d = \sim 0.1$ nM) and requires the presence of the both its N- and C-terminal lobe domains to achieve this tight binding (R.L.H., unpub. data). Structural studies of the LSMT–Rubisco holoenzyme complex will be necessary to appreciate the interactions that provide specificity between this SET domain methyltransferase and its substrate.

METHODS

LSMT methyltransferase assays. Rubisco LSMT was assayed as described²² with several modifications. Assays with Rubisco as the substrate and Lys–HCl (Sigma) or MeLys–HCl (Bachem) as inhibitors contained 200 mM bicine, pH 8.2, and were incubated for 1 min at 30 °C. Assays for product analysis with Lys (150 μ M) or MeLys (150 μ M) as substrates contained higher levels of LSMT (33 μ M) and ³H-labeled AdoMet (82 μ M) and were carefully adjusted to a pH of ~ 8.8 with NaOH. The reactions (20 μ l) were incubated for 40 min at 30 °C, and terminated by the addition of 30 μ l of water. A 5- μ l sample of the reaction was immediately spotted onto a 2-cm wide band onto a silica gel TLC plate (20 cm \times 20 cm, 250 μ m thickness, 2–25 μ m particle size, 60 Å pore size) (Sigma), and the remainder was stored at -80 °C. The TLC plate was developed and visualized as reported⁴⁰. R_f values corresponding to known standards for Lys (0.53), MeLys (0.44), Me₂Lys (0.76) and Me₃Lys (0.21) were examined for radioactivity by removing the appropriate area with a razor blade and placing the silica gel in a 5.0 ml vial with 30% (v/v) methanol followed by 4.5 ml of liquid scintillation cocktail. Radioactivity was determined by liquid scintillation spectroscopy. Control reactions consisted of assays minus LSMT or assays minus substrate. This TLC separation technique results in complete separation of Lys and all methylated derivatives, as well as AdoMet, which remains restricted to an area 1 cm above and including the origin ($R_f \leq 0.06$). All data were graphed and analyzed using Sigma Plot version 8.0.

Crystallization and data collection. Pea LSMT (residues 46–482) was expressed and purified as described²². LSMT ternary complex crystals were grown in hanging drops at 25 °C with 10 mg ml⁻¹ protein in 0.95–1.10 M sodium acetate, 400 μ M AdoHcy, 1 mM tris (2-carboxyethyl)phosphine hydrochloride (TCEP) and in either 100 mM Lys or MeLys acetate buffer, pH 6.8. Crystals grew to a size of 300–400 μ m in ~ 3 weeks. Drops containing the crystals were then dehydrated by successively transferring the cover slips over higher concentrations of sodium acetate to a final concentration of 2.4 M. Crystals were then cryoprotected in 30% (v/v) glycerol and frozen in the cryo-stream for data collection. Data set was collected at 95 K on an R-Axis IV++ detector (Rigaku) using Cu K α radiation from a Rigaku RU200 generator (Rigaku) and focused with Osmic confocal mirrors (Rigaku). Data were then indexed, reduced and scaled using DENZO and SCALEPACK⁴¹.

Structure solution and refinement. The crystal structure of LSMT bound to AdoHcy and a HEPES buffer molecule (PDB entry 1MLV) was used as a starting model for refinement of the Lys and MeLys complexes. The HEPES and water molecules were removed, and the Lys and MeLys structures were refined

Table 1 Crystallographic data collection and refinement statistics

Crystallographic data		
Space group	I222	
Cell dimensions	Lys	MeLys
<i>a</i> (Å)	131.96	130.41
<i>b</i> (Å)	156.90	153.19
<i>c</i> (Å)	267.55	266.32
λ (Å)	1.5418	1.5418
<i>d</i> _{min} (Å)	2.65	2.55
Number of reflections ^a	79,741 (7,902)	83,829 (7,768)
Completeness (%) ^a	98.9 (99.0)	96.5 (93.6)
$\langle I \rangle / \langle \sigma \rangle$ ^a	25.9 (2.51)	30.2 (2.45)
<i>R</i> _{sym} (%) ^{a,b}	5.6 (49.4)	4.4 (48.5)
Refinement		
Resolution range (Å)	30.0–2.65	30.0–2.55
Number of reflections	72,448	74,004
<i>R</i> _{work} (%) ^c	22.7	22.8
<i>R</i> _{free} (%) ^d	26.6	26.9
Luzatti coordinate error	0.38	0.41
Cross-validated Luzatti coordinate error	0.44	0.49
R.m.s. deviation		
Bond length (Å)	0.007	0.007
Bond angle (°)	1.2	1.3
Improper angle (°)	0.79	0.78
Dihedral angle (°)	21.5	21.3
Average <i>B</i> -factor (Å ²)		
Overall	66.5	70.4
Protein	66.8	70.9
Ligands	53.8	59.9
Water	64.0	62.8
<i>B</i> -factor r.m.s. deviation (Å ²)		
Bonded main chain atom	1.40	1.32
Bonded side chain atom	1.82	1.65
Residues in Ramachandran plot ϕ - ψ regions (%) ^e		
Most favored	87.6	87.8
Additionally allowed	12.1	11.9
Generously allowed	0.3	0.3
Disallowed	0	0

^aValues in parentheses are for the highest-resolution bin. ^b $R_{\text{sym}} = \sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)$. ^c*R*-factor = $\sum (|F_o| - |F_c|) / \sum |F_o|$. ^d*R*_{free} is the *R*-factor value calculated for a test set of reflections, comprising a randomly selected 5% of the data that is not used during refinement. ^eRamachandran plot ϕ - ψ regions are defined according to the criteria of Procheck⁴⁸.

against their datasets using CNS⁴². After several rounds of simulated-annealing torsion angle molecular dynamics refinement, $2F_o - F_c$ and $F_o - F_c$ omit maps revealed density for either a well-ordered Lys or MeLys in each of the three LSMT molecules in the asymmetric unit for each respective complex. Lys and MeLys molecules were subsequently built into the omit map density for the respective structures using O⁴³. The structures were then subsequently refined with several cycles of manual model building and molecular dynamic, positional and *B*-factor refinements using NCS restraints on the domains and regions flanking the SET domain. Residues 258–266 in the A molecule were omitted from the Lys complex structure because of their disorder in the crystal, whereas residues 258–265 were omitted from the A molecule in the MeLys complex. During later rounds of refinement, 644 and 652 molecules of water were added to the Lys and MeLys complexes, respectively. The final structure of the Lys complex has a *R*_{work} of 22.7% and a *R*_{free} of 26.6%, whereas the MeLys

complex has *R*_{work} and *R*_{free} of 22.8% and 26.9%, respectively. The refined models have none of their non-glycine residues present in the disallowed regions of the Ramachandran plot.

Coordinates. The coordinates have been deposited in the Protein Data Bank (accession codes 1OZV (Lys) and 1P0Y (MeLys)).

ACKNOWLEDGMENTS

We would like to thank S. Gamblin for SET7/9–AdoHcy–histone H3 ternary complex coordinates, and D. Engelman and G. Felsenfeld for insightful discussions. Research by R.L.H. and E.M.F. was supported by the DOE, Division of Energy Biosciences.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 10 March; accepted 28 May 2003

Published online 22 June 2003; doi:10.1038/nsb946

- Kouzarides, T. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **12**, 198–209 (2002).
- Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).
- Rice, J.C. & Allis, C.D. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* **13**, 263–273 (2001).
- Strahl, B.D., Ohba, R., Cook, R.G. & Allis, C.D. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl. Acad. Sci. USA* **96**, 14967–14972 (1999).
- Noma, K., Allis, C.D. & Grewal, S.I. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**, 1150–1155 (2001).
- Houtz, R.L., Stults, J.T., Mulligan, R.M. & Tolbert, N.E. Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase. *Proc. Natl. Acad. Sci. USA* **86**, 1855–1859 (1989).
- Klein, R.R. & Houtz, R.L. Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit *N*-methyltransferase. *Plant Mol. Biol.* **27**, 249–261 (1995).
- Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**, 5857–5864 (1998).
- Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
- Bannister, A.J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
- Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D. & Grewal, S.I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
- Bernstein, B.E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. USA* **99**, 8695–8700 (2002).
- Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
- Zhang, X. *et al.* Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127 (2002).
- Jacobs, S.A. *et al.* The active site of the SET domain is constructed on a knot. *Nat. Struct. Biol.* **9**, 833–838 (2002).
- Kwon, T. *et al.* Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J.* **22**, 292–303 (2003).
- Wilson, J.R. *et al.* Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111**, 105–115 (2002).
- Xiao, B. *et al.* Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652–656 (2003).
- Min, J., Zhang, X., Cheng, X., Grewal, S.I. & Xu, R.M. Structure of the SET domain histone lysine methyltransferase Clr4. *Nat. Struct. Biol.* **9**, 828–832 (2002).
- Manzur, K.L. *et al.* A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nat. Struct. Biol.* **10**, 187–196 (2003).
- Triebel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L. & Hurley, J.H. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111**, 91–103 (2002).
- Marmorstein, R. Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem. Sci.* **28**, 59–62 (2003).
- Yeates, T.O. Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell* **111**, 5–7 (2002).
- Dutnall, R.N. & Denu, J.M. Methyl magic and HAT tricks. *Nat. Struct. Biol.* **9**, 888–891 (2002).
- Houtz, R.L., Poneleit, L., Jones, S.B., Royer, M. & Stults, J.T. Posttranslational modifications in the amino-terminal region of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from several plant species. *Plant Physiol.* **98**, 1170–1174 (1992).

27. Houtz, R.L., Royer, M. & Salvucci, M.E. Partial purification and characterization of ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit *N*-methyltransferase. *Plant Physiol.* **97**, 913–920 (1991).
28. Segel, I.H. *Enzyme Kinetics* (John Wiley & Sons, New York; 1975).
29. Derewenda, Z.S., Lee, L. & Derewenda, U. The occurrence of C-H \cdots O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248–262 (1995).
30. Bella, J. & Berman, H.M. Crystallographic evidence for C α -H \cdots O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264**, 734–742 (1996).
31. Fabiola, G.F., Krishnaswamy, S., Nagarajan, V. & Patabhi, V. C-H \cdots O hydrogen bonds in β -sheets. *Acta Crystallogr. D* **53**, 316–320 (1997).
32. Wahl, M.C. & Sundaralingam, M. C-H \cdots O hydrogen bonding in biology. *Trends Biochem. Sci.* **22**, 97–101 (1997).
33. Senes, A., Ubarretxena-Belandia, I. & Engelman, D.M. The C α -H \cdots O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Pro. Natl. Acad. Sci. USA* **98**, 9056–9061 (2001).
34. Weiss, M.S., Brandl, M., Suhnel, J., Pal, D. & Hilgenfeld, R. More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* **26**, 521–523 (2001).
35. Scheiner, S., Kar, T. & Gu, Y. Strength of the C α -H \cdots O hydrogen bond of amino acid residues. *J. Biol. Chem.* **276**, 9832–9837 (2001).
36. Derewenda, Z.S., Derewenda, U. & Kobos, P.M. (His)C ϵ -H \cdots O=C < hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.* **241**, 83–93 (1994).
37. Bach, R.D., Thorpe, C., & Dmitrenko, O. C-H \cdots carboxylate oxygen hydrogen bonding in substrate activation by acyl-CoA dehydrogenases: synergy between the H-bonds. *J. Phys. Chem.* **106**, 4325–4335 (2002).
38. Coward, J.K. Chemical mechanisms of methyl transfer reactions: comparison of methylases with nonenzymatic 'model reactions' in *The Biochemistry of Adenosylmethionine* (eds. Salvatore, F., Borek, E., Zappia, V., William-Ashman, H.G. & Schlenk, F.) 127–144 (Columbia University Press, New York; 1977).
39. Duff, A.P., Andrews, T.J. & Curmi, P.M. The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J. Mol. Biol.* **298**, 903–916 (2000).
40. Rebouche, C.J. & Broquist, H.P. Carnitine biosynthesis in *Neurospora crassa*: enzymatic conversion of lysine to ϵ -*N*-trimethyllysine. *J. Bacteriol.* **126**, 1207–1214 (1976).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
42. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
43. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
44. Esnouf, R.M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134 (1997).
45. Esnouf, R.M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
46. Merritt, E.A. & Bacon, D.J. Raster3D version 2.0: a program for photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
47. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
48. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1993).

