POLYPEPTIDE SUBSTRATE SPECIFICITY OF \( P_s \) LSMT: A SET-DOMAIN PROTEIN METHYLTRANSFERASE

Roberta Magnani, Nihar R. Nayak, Mitra Mazarei#, Lynnette M.A. Dirk, and Robert L. Houtz

From the Department of Horticulture, Plant Physiology/Biochemistry/Molecular Biology Program, 401D Plant Science Building, University of Kentucky, Lexington, KY 40546-0312 and the #Department of Plant Sciences, 252 Ellington Plant Sciences Building, University of Tennessee, Knoxville, TN 37996-4561

Address correspondence to: Robert L. Houtz, Department of Horticulture, Plant Physiology/Biochemistry/Molecular Biology Program, 401D Plant Science Building, University of Kentucky, Lexington, Kentucky 40546-0312; Tel. 859 257-1982; Fax. 859 257-7874; E-Mail: rhoutz@uky.edu

Rubisco large subunit methyltransferase is a SET-domain protein responsible for the trimethylation of Lys-14 in the large subunit of Rubisco. The polypeptide substrate specificity determinants for pea Rubisco large subunit methyltransferase were investigated using a fusion protein construct between the first 23 amino acids from the large subunit of Rubisco and human carbonic anhydrase II. A total of 40 conservative and non-conservative amino acid substitutions flanking the target Lys-14 methylation site (positions \( P_{-3} \) to \( P_{+3} \)) were engineered in the fusion protein. The catalytic efficiency (\( k_{cat}/K_m \)) of \( P_s \) LSMT was determined using each of the substitutions and a polypeptide consensus recognition sequence deduced from the results. The consensus sequence, represented by X - Gly/Ser - Phe/Tyr - Lys - Ala/Lys/Arg - Gly/Ser - \( \pi \), where X is any residue, Lys is the methylation site, and \( \pi \) is any aromatic or hydrophobic residue was used to predict potential alternative substrates for \( P_s \) LSMT. Four chloroplast-localized proteins were identified including \( \gamma \)-tocopherol methyltransferase (\( \gamma \)-TMT). In vitro methylation assays using \( P_s \) LSMT and a bacterially-expressed form of \( \gamma \)-TMT from \textit{Perilla frutescens} confirmed recognition and methylation of \( \gamma \)-TMT by \( P_s \) LSMT in vitro.

Four chloroplast-localized proteins were identified including \( \gamma \)-tocopherol methyltransferase (\( \gamma \)-TMT). In vitro methylation assays using \( P_s \) LSMT and a bacterially-expressed form of \( \gamma \)-TMT from \textit{Perilla frutescens} confirmed recognition and methylation of \( \gamma \)-TMT by \( P_s \) LSMT in vitro.

RNAi-mediated knockdown of the \( P_s \) LSMT homologue (\( N_t \) LSMT) in transgenic tobacco plants resulted in a 2-fold decrease of \( \alpha \)-tocopherol, the product of \( \gamma \)-TMT. The results demonstrate the efficacy of consensus sequence driven identification of alternative substrates for \( P_s \) LSMT as well as identification of functional attributes of protein methylation catalyzed by LSMT.

Besides being the world’s most abundant enzyme (1) and essentially the only significant route for carbon entry into the earth’s biosphere, Rubisco (EC 4.1.1.39) is also extensively modified with a plethora of co- and post-translational modifications (2;3). Two of these modifications are methylations, one on the \( \varepsilon \)-amine of Lys-14 in the large subunit (LS, (2)), and another on the \( \alpha \)-amine of Met-1 in the small subunit (SS, (4)). Methylation at Lys-14 is catalyzed by LSMT (EC. 2.1.1.127); a highly conserved SET-domain protein lysine methyltransferase (PKMT) found in all plant species as well as orthologues in representative organisms across all eukaryotic species (5). LSMT transcripts are coordinately expressed with the LS of Rubisco through several light-regulated promoter elements, and are predominantly found in leaf tissue (6). Despite much effort, an exact role for LSMT and Lys-14 methylation in the LS of Rubisco remains obscure. Although there is a partial correlation of LSMT activity with plants species known to exhibit light/dark regulation of Rubisco activity through...
There are other plant species without Lys-14 methylation that nonetheless have LSMT homologues, some of which are alternatively spliced (7). Thus, there may be several different roles played by Lys-14 methylation as well as LSMT that vary in a species-specific manner (3).

SET-domain PKMTs are well known for their role in the methylation of histones and subsequent influence on epigenetic gene expression (8-10). While PKMTs have in the past been regarded as having exceptionally high polypeptide substrate specificity, recent results clearly demonstrate that PKMTs can have more than one polypeptide substrate. Human Smyd2 (11), specific for Lys-36 in H3 histones, also methylates Lys-370 in p53 (12) and SET1 from *Saccharomyces cerevisiae*, specific for Lys-4 in H3 histones, methylates, as well, lysyl residues in DAM1 (13), a kinetochore protein. Human SET 7/9 methylates Lys-4 in H3 histones (14;15), but also methylates Lys-189 in TAF10 (16) as well as Lys-372 in p53 (17). Examination of the structural basis for substrate flexibility in SET 7/9 revealed a consensus recognition sequence that was used to tentatively identify other polypeptide substrates with database search engines (18). Identification of alternative substrates for PKMTs and consensus sequence requirements for catalytic activity broadens the repertoire of target polypeptide substrates, but, more importantly, can also lead to an increased understanding of previously unknown functional aspects of protein methylation (19). PKMT protein specificity studies frequently rely on synthetic polypeptide substrates and/or site-directed mutagenesis of the target polypeptide substrate. With LSMT however, polypeptide mimics of the N-terminal region of the LS, specifically acetylated-Pro-3 to Tyr-25, do not bind with any appreciable affinity to LSMT and furthermore, the hexadecameric plant forms of Rubisco cannot be expressed and assembled in bacteria. To explore the determinants of polypeptide substrate specificity for *Pisum sativum* (Pea) LSMT (PsLSMT) and potentially identify new alternative substrates, an artificial polypeptide substrate was created by fusing the C-terminus of the first 23 amino acids from the LS of spinach Rubisco (N-terminal residue directly determined as N-acetyl-Proline; from Pro-3 to Tyr-25) to the N-termini of human carbonic anhydrase II (HCA II). This fusion protein (LS:HCA II) was acceptable as a substrate for PsLSMT and allowed an examination of the amino acid sequence requirements for recognition of Rubisco by PsLSMT. In the present work, we report the amino acid consensus sequence required for the trimethylation of Lys-14 in the LS of Rubisco by examining the effect of amino acid replacements in the region encompassing residues Val-11 to Val-17. The consensus sequence was used to screen protein databases and identify putative alternate substrates. Chloroplastic γ-tocopherol methyltransferase (γ-TMT) was identified and confirmed as a potential alternative substrate for PsLSMT in vitro. Analysis of the amounts of α-tocopherol (the product of γ-TMT) in LSMT-knockdown tobacco plants suggested that methylation of γ-TMT results in increased activity in vivo.

**Experimental Procedures**

**Enzymes, Substrates, and Chemicals** – PsLSMT (residues 38 to 485, 51.7 kDa) was expressed and purified as previously described (20-22) and was judged to be greater than 95% homogeneous as determined by SDS-PAGE. Des(methyl) spinach Rubisco was purified from spinach leaves (23). S-Adenosyl-L-methionine (AdoMet) from Sigma was purified prior to use (24). HiTrap™ chelating columns, and [3H-methyl]AdoMet (~70-80 Ci mmol⁻¹) were from GE Healthcare (Piscataway, NJ). AdoMet was diluted to a specific activity of 1-4 µCi nmol⁻¹ for enzyme assays. Immobilon-P (PVDF) membrane and MF-Millipore membrane filters (0.025 µm) were from Millipore Corporation (Billerica, MA). His-Mag Agarose Beads, Benzonase® nuclease, pET 23b and 23d expression plasmids were from Novagen (EMD Biosciences, Merck KGaA, Darmstadt, Germany). pGEM-T Easy Vector System I was from Promega Corporation (Madison, WI). His-Mag Agarose Beads, Benzonase® nuclease, pET 23b and 23d expression plasmids were from Novagen (EMD Biosciences, Merck KGaA, Darmstadt, Germany). pGEM-T Easy Vector System I was from Promega Corporation (Madison, WI). BL21(DE3)pLysS competent cells, PfuUltra™ High-Fidelity DNA polymerase and the StrataScript® First-Strand cDNA Synthesis Kit were from Stratagene (La Jolla, CA). T4 DNA ligase and restriction enzymes were products of Invitrogen (Carlsbad, CA) and acrylamide/bisacrylamide solution was from Bio-Rad Laboratories (Hercules, CA). QIAquick® gel extraction kit and RNeasy Plant
I and Easy, the full length fusion was digested with further 20 cycles. After TA cloning to pGEM-T and OL34.4 (0.15 fusion DNA was amplified by including OL34.1 PCR reaction and after 10 cycles, full length gel-purified products were mixed together in a purified with a QIAquick® gel extraction kit. The (Metaphor agarose for the 96-bp amplicon) and separate PCR reactions created the respective primer set and a 3'-end I site. After two Xho 31-bp overlap with the 3'-end of the previous synthesized for amplification of the HCA II with a additional primers (OL34.3, 34.4; Table S1) were Table S1 with an intervening I site. Two Nco vector (25) with 5' FL) in a modified pET 31 bacterial expression of Medicine, University of Florida, Gainesville, FL) in a modified pET 31 bacterial expression vector (25) with 5' Nco I and 3' Xho I restriction sites. Twenty-two base pair overlapping sense and antisense oligonucleotides (OL34.1, 34.2; Table S1), encoding the N-terminal amino acids of the LS of Rubisco from Pro-3 to Tyr-25 (with an initiating Met) and the first 5 amino acids of HCA II, were synthesized as indicated in Supplementary S1, encoding the N-terminal amino acids of the LS of Rubisco from Pro-3 to Tyr-25 (with an initiating Met) and the first 5 amino acids of HCA II, were synthesized as indicated in Supplementary Table S1 with an intervening I site. Two additional primers (OL34.3, 34.4; Table S1) were synthesized for amplification of the HCA II with a 31-bp overlap with the 3'-end of the previous primer set and a 3'-end Xho I site. After two separate PCR reactions created the respective fragments, the products were separated on gels (Metaphor agarose for the 96-bp amplicon) and purified with a QIAquick® gel extraction kit. The gel-purified products were mixed together in a PCR reaction and after 10 cycles, full length fusion DNA was amplified by including OL34.1 and OL34.4 (0.15 µM of each; Table S1) for a further 20 cycles. After TA cloning to pGEM-T Easy, the full length fusion was digested with Nde I and Xho I, ligated with T4 DNA ligase into predigested pET 23b expression plasmid. The ligated fusion DNA in the LS:HCA II construct as template DNA and the resultant amplicon was digested with Nde I and Xho I, and ligated into pET 23b expression vector. All recombinant clones were sequenced for validation of insert and desired mutation.

**LS:HCA II Fusion Protein Constructs** – The DNA encoding human carbonic anhydrase (HCA II) was obtained as a kind gift from Prof. P. Laipis (Dept. of Biochemistry and Molecular Biology, College of Medicine, University of Florida, Gainesville, FL) in a modified pET 31 bacterial expression vector (25) with 5' Nco I and 3' Xho I restriction sites. Twenty-two base pair overlapping sense and antisense oligonucleotides (OL34.1, 34.2; Table S1), encoding the N-terminal amino acids of the LS of Rubisco from Pro-3 to Tyr-25 (with an initiating Met) and the first 5 amino acids of HCA II, were synthesized as indicated in Supplementary Table S1 with an intervening I site. Two additional primers (OL34.3, 34.4; Table S1) were synthesized for amplification of the HCA II with a 31-bp overlap with the 3'-end of the previous primer set and a 3'-end Xho I site. After two separate PCR reactions created the respective fragments, the products were separated on gels (Metaphor agarose for the 96-bp amplicon) and purified with a QIAquick® gel extraction kit. The gel-purified products were mixed together in a PCR reaction and after 10 cycles, full length fusion DNA was amplified by including OL34.1 and OL34.4 (0.15 µM of each; Table S1) for a further 20 cycles. After TA cloning to pGEM-T Easy, the full length fusion was digested with Nde I and Xho I, ligated with T4 DNA ligase into predigested pET 23b expression plasmid. The ligated fusion DNA in the LS:HCA II construct was verified by DNA sequencing and transformed to BL21(DE3)pLysS cells for expression.

**Sequence Mutations** – Mutations were generated in the LS:HCA II construct using mutated oligonucleotides as listed in Supplemental Table S1. Equal amounts of mutated (42 single point substitutions and 3 multiple point; Table S1) and reverse primer (OL34.4; Table S1) were mixed with the LS:HCA II construct as template DNA and the resultant amplicon was digested with Nde I and Xho I, and ligated into pET 23b expression vector. All recombinant clones were sequenced for validation of insert and desired mutation.

**LS:HCA II Protein Purification** – Purification of LS:HCA II proteins was performed as described by Banerjee et al. (26), with the following modifications. An induction time of 5 h at 25 °C with 0.4 mM IPTG was used once the culture in Luria Bertani medium (LB) containing ampicillin (100 µg ml⁻¹) and chloramphenicol (34 µg ml⁻¹) reached an A₆₀₀ of 0.5. Bacteria were harvested by centrifugation at 7000 × g for 10 min and lysed using 4 cycles of freeze-thaw in the presence of 0.01% (v/v) Benzonase® nuclease with 1 mM phenylmethanesulfonyl fluoride. After centrifugation at 10,000 × g, the supernatant was applied to a HiTrap™ chelating column charged with 0.3 M ZnSO₄ and LS:HCA II was eluted with a linear gradient of imidazole from 0 to 125 mM. Fractions containing LS:HCA II were pooled, concentrated and dialyzed for 1 h using MF- Millipore membrane filters against 50 mM Tris-HCl, 5 mM MgCl₂ and 1 mM EDTA pH 8.

**Western blot and phosphor-images** – Western blot and phosphor-image analyses were performed on proteins electrophoretically transferred from SDS-PAGE gels (12.5 % acrylamide) to PVDF membranes. Western blots were performed using an antibody specific for the N-terminus of the LS of spinach Rubisco (residues 3-25). Phosphor-image analyses of radiolabel incorporation from [³H-methyl]AdoMet into proteins transferred to PVDF membranes utilized exposure times from 24 to 48 h depending on the level of radioactivity. PsLSMT Methyltransferase Assays – PsLSMT activity was determined as previously described (20;21). Briefly, assays contained 100 mM bicine, 20 mM MgCl₂ pH 8, [³H-methyl]AdoMet (7-9 x 10⁶ dpm nmol⁻¹ at various concentrations), 4 µg of PsLSMT, and indicated amounts of LS:HCA II fusion protein in a final volume of 20 µl. The reactions were incubated at 30 °C for two min, and terminated by addition of 25 volumes of 10% (v/v) TCA. The precipitated protein pellet was dissolved with 150 µl of 0.1 N NaOH and precipitated again prior to dissolution in 50 µl of formic acid, addition of scintillation cocktail and determination of radioactivity. Kinetic parameters (Kₘ and kₗ) were obtained from fitted Michaelis-Menten
equations using SigmaPlot (version 10), except when mutations resulted in large increases in the $K_m$. Under these conditions the slope of the velocity versus substrate concentration plot was used as an estimate of $k_{cat}/K_m$ (27). All assays were optimized for linearity with time and enzyme concentration, and substrate consumption was limited to 20% or less.

$\gamma$-TMT Cloning, Expression and Purification – Seeds of *Perilla frutescens* were cultured in the greenhouse. Total RNA from immature embryos was isolated with the RNaseasy Plant Mini Kit following the manufacturer’s directions. First-strand cDNA was synthesized using an oligo-dT primer and the reagents provided with the StrataScript® First-Strand cDNA Synthesis Kit. The full-length cDNA of the $\gamma$-TMT gene (GenBank accession number AF213481) was then amplified from the first-strand cDNA using the oligo 5’-GTCAAAATCACCAATTACCTATTT-3’, oligo-dT, and PfuUltra™ High-Fidelity DNA polymerase. PCR products were electrophoresed on a 1.0 % agarose gel and the expected 1369 bp amplicon excised from the gel. The DNA fragment was purified using QIAquick® gel Extraction Kit, ligated into pGEM-T Easy® cloning vector and sequenced. The full-length $\gamma$-TMT cDNA was truncated at the N-terminus according to the predicted chloroplast processing site from ChloroP (5) to obtain a clone encoding the mature protein. Accordingly, a forward primer (5’-NcoI restriction site), a reverse primer (5’-I restriction site), and PfuUltra™ High-Fidelity DNA polymerase were used to amplify the partial cDNA. The amplicon was ligated into pET 23d in frame with the histidine tag. The vector was then transformed to BL21(DE3)pLysS cells for expression. Transformed colonies were induced when $A_{600}$ was 0.6 with 0.4 mM IPTG and vigorously shaken at 25 °C overnight. Harvested cells were disrupted with 4 cycles of freeze-thaw and one quick cycle of sonication, and the tagged protein was purified using His-Mag Agarose Beads following manufacturer’s instructions.

**Construction of NiLSMT-Knockdown Tobacco Plants** - The full length cDNA for tobacco Rubisco LSMT (29) was cloned in the sense or antisense orientation into the multiple cloning site of the binary vector pKYLX (30) under the control of the CaMV 35S promoter. *Agrobacterium tumefaciens* strain GV3850 was transformed with the sense or antisense construct by triparental mating (31) with *Escherichia coli* HB101 containing the pRK2013 helper plasmid. The sense or antisense construct was introduced into tobacco (cv Petite Havana) using *A. tumefaciens*-mediated transformation following the standard leaf disk method (32). PCR was used to confirm the presence of the sense or antisense 35S-Rubisco LSMT transgene in the genome of the tobacco plants. Plants (T2) identified by PCR as positively transformed were crossed sense with antisense to generate RNAi for Rubisco LSMT (33). The progeny (F1) were screened for *in vitro* substrate susceptibility of tobacco Rubisco to incorporate methyl groups from [3H-methyl]AdoMet in the presence of PsLSMT.

$\gamma$-TMT Activity Assays and Product Analyses – A tobacco leaf disk (12 mm diameter) was incubated in 150 µl of buffer (50 mM Bisine pH 8, 5 mM MgCl$_2$), containing 3 mM $\gamma$-tocopherol and 10 µl of [3H-methyl]AdoMet (77 Ci mmol$^{-1}$) for 3 h under constant illumination (360 µmol m$^{-2}$ sec$^{-1}$) at room temperature (~25 °C). Total tocopherols were then extracted as previously described (34) by grinding the leaf disk in 5 ml of 0.5 M KOH in methanol after the addition of 20 µg of tocot. To determine the incorporation of [3H-methyl]AdoMet in the presence of PsLSMT, the incorporation was stopped by the addition of 1 ml of nanopure H$_2$O and 5 ml of hexane. The samples were vortexed and centrifuged for 2 min at 600 x g for phase separation. The hexane was collected, evaporated under nitrogen, and the residue resuspended in acetonitrile:methanol 95:5 (v/v). Tocopherols were separated on a Waters Nova-Pak® C18 4 mm reverse phase column (300 mm x 3.9 mm) with acetonitrile:methanol 95:5 (v/v) at 1 ml min$^{-1}$ for 35 min using detection at 214 nm. The peak corresponding to $\alpha$-tocopherol was collected for determination of radiolabel incorporation.

Determination of $\alpha$-tocopherol amounts were essentially identical to the aforementioned procedure except 20 disks were used for each
determination. All analyses were replicated at least four times and % recovery calculated from the tocol internal standard.

RESULTS

The LS:HCA II protein fusion was evaluated as a polypeptide substrate for PsLSMT. The fusion protein exhibited a slightly higher molecular mass than HCA II (~ 2.6 kDa, corresponding to the N-terminal sequence from the LS of Rubisco), immuno-reactivity with an antibody specific for the N-terminal amino acid sequence (residues 3-25) of the LS of Rubisco, and, more importantly, radiolabel incorporation from [3H-methyl]AdoMet in the presence of PsLSMT (Fig. 1A). The fusion protein was thus considered as a potentially effective tool for mapping the polypeptide substrate specificity of PsLSMT. There are three additional lysyl residues (other than the Lys-14 methylation site) within the LS N-terminal sequence located at positions 8, 18, and 21. Mutagenesis of these residues to either Arg or Ala did not affect methylation of Lys-14 by PsLSMT, whereas mutagenesis of the Lys-14 methylation site abolished methylation (Fig. 1B). The methylation site specificity was, thus, not altered in the LS:HCA II protein. Additionally, product analyses confirmed that the predominant methylation product was Me3Lys (Fig. 1C), identical to previous results with Rubisco.

Kinetic analyses of PsLSMT using the LS:HCA II protein as a polypeptide substrate revealed an increase in the apparent $K_m$ compared with Rubisco from 1.4 $\mu$M to 1.2 mM, a decrease in $k_{cat}$ from 2.4 min$^{-1}$ to 0.6 min$^{-1}$, and a change in catalytic efficiency from 1.7 $\mu$M$^{-1}$ min$^{-1}$ to 0.5 mM$^{-1}$ min$^{-1}$ (Fig. 2A). Although reduced, the catalytic rate constant was still within the range of activities reported for several other SET-domain PKMTs (27;35;36); however, the weaker binding affinity suggests that other regions of PsLSMT interact with Rubisco outside the immediate target methylation site and that those regions contribute to the affinity of polypeptide substrate binding. Nevertheless, this is unlikely to compromise the determination of substrate specificity but likely, instead, to provide a more conservative estimate of substrate specificity. An examination of AdoMet kinetics revealed a change in the $K_m$ from 6 $\mu$M to 19 $\mu$M, and a similar change in $k_{cat}$ as described above.

Each of the flanking three residues surrounding the Lys-14 methylation site (those before Lys-14 denoted as P$_{-3}$, P$_{-2}$ and P$_{-1}$ and those after Lys-14 denoted P$_{+1}$, P$_{+2}$ and P$_{+3}$) was individually mutagenized using amino acid substitutions reflecting the entire range of general side chain chemistries, and the LS:HCA II polypeptide evaluated as a substrate for PsLSMT (Fig. 3). Kinetic plots of PsLSMT activity revealed changes in both $k_{cat}$ and apparent $K_m$; therefore, catalytic efficiency ($k_{cat}/K_m$) was considered as a better measure of polypeptide substrate specificity. As an example, PsLSMT activity using mutated LS:HCA II Val-11 to Lys and Val-17 to Leu had similar apparent $K_m$ values and, yet, due to differences in $k_{cat}$, PsLSMT processed the Val-11 to Lys substrate more efficiently than unmodified substrate whereas PsLSMT processed the Val-17 to Leu substrate with similar efficiency. Furthermore, some mutations led to increases in the $K_m$ of PsLSMT to the extent that saturated velocities could not be obtained. Under these conditions however the $k_{cat}/K_m$ can be estimated from the initial slope of the enzyme velocity versus substrate concentration plot (27), as denoted in Fig. 3 by L following the efficiency estimate.

Out of forty single-point substitutions (excluding substitutions at Lys-14), thirteen were acceptable as substrates (Fig. 3). The remaining substitutions either resulted in protein which was insoluble, or supported activity too low to be considered a substrate for PsLSMT. There was limited symmetry of the tolerated substitutions around the methylation site in that the majority of replacements at the P$_{-3}$ and P$_{+3}$ positions, regardless of degree of conservation, were suitable as PsLSMT substrates. All substitutions at position 11 (P$_{-3}$) were tolerated, suggesting considerable flexibility in the position normally occupied by Val-11; whereas, negative and polar substitutions at P$_{+3}$ of the hydrophobic Val-17 were not considered substrates. Generally, substitutions were not well tolerated at P$_{-2}$ and P$_{-1}$; only Ser was acceptable as a replacement for either Gly-12 or Gly-16. At the P$_{+1}$ site, Tyr, but not Trp, partly substituted for Phe-13, suggesting a preference at this position for residues with an aromatic side chain but with a particular side chain

[5]

Downloaded from http://www.jbc.org/ by guest on October 30, 2017
volume: 189.0 Å³ for Phe and 193.6 Å³ for Tyr were suitable but neither 227.8 Å³ for Trp nor 153.2 Å³ for His were acceptable (37). At the P1 position several substitutions supported low but detectable PsLSMT activity and both Arg and Lys were well tolerated substitutions.

Three non-conservative substitutions resulted in increased substrate acceptability, namely, Lys and Ser at position P3 (~3-fold and 2-fold, respectively), and Lys at position P1 (~4-fold). The relatively large increase in PsLSMT catalytic efficiency observed when using the Ala-15 to Lys fusion protein was examined further by evaluating the kinetic constants for PsLSMT using this mutated substrate (Fig. 4). The nearly 4-fold increase in PsLSMT catalytic efficiency with this substrate was a consequence of an increase in $k_{cat}$ (~2.5-fold) and a decrease in $K_m$ (~1.75-fold, Fig. 4A). Because the observed changes in $k_{cat}$ could at least in part be due to methylation at Lys-15 as well as Lys-14, we replaced Lys-14 with Arg while maintaining the Ala-15 to Lys replacement. Surprisingly, the double replacement, Lys-14 to Arg and Ala-15 to Lys, was nearly as acceptable a substrate as the unmodified fusion protein (Fig. 4B), although the $K_m$ of PsLSMT using it as a substrate was 2-fold higher. The kinetics of AdoMet binding for PsLSMT was also examined for the Ala-15 to Lys replacement (Fig. 4C) with results similar to those reported in Fig. 2. In this case, the $K_m$ of PsLSMT for AdoMet using the mutated LS:HCA II was almost half of that of the unmodified fusion protein, while the $k_{cat}$ was just slightly higher. We attempted to discern if the Ala-15 to Lys protein was methylated at both Lys-14 and Lys-15 by conducting end-point enzymatic assays. However, the maximum level of $[^3H]$-methyl incorporation never exceeded 2-3 moles per mole of LS:HCA II, suggesting that methylation was restricted to either Lys-14 or Lys-15, but the results do not exclude the possibility that Lys-15, when present, may also be methylated.

A consensus sequence for polypeptide substrate recognition by PsLSMT (represented by $[X\text{ - G]y/Ser\text{- Phe/Tyr\text{- Lys\text{- Ala/Lys/Arg\text{- Gly/Ser - π \text{ where X is any residue, Lys is the methylation site, and π is any aromatic or hydrophobic residue}}}$ was deduced from the data in Figure 3. Scansite (38) was used to search the Swiss-Prot protein database for potential alternative protein substrates for PsLSMT using the consensus sequence along with discriminators for chloroplast-localized plant proteins. A total of 372 proteins from different plant species were identified. The majority represented the LS of Rubisco (367) but five were chloroplast-localized proteins other than the LS of Rubisco (Fig. 5), including γ-tocopherol methyltransferase (γ-TMT), a methyltransferase involved in the synthesis of vitamin E (39). The γ-TMT was considered particularly interesting among the possible alternative substrate candidates because the consensus motif, localized to the C-terminus, was conserved among different species. We tested if γ-TMT could be recognized as a substrate for PsLSMT using purified γ-TMT from bacterial expression of a partial cDNA, and subsequent incubation with PsLSMT in methyltransferase assays. The incorporation of radiolabel from $[^3H\text{-methyl}]$AdoMet demonstrated that γ-TMT was a substrate for PsLSMT (Fig. 6B), and that the predominant product was Me₂K (Fig. 6C). We next attempted to determine if methylation affected the activity of γ-TMT by stoichiometrically methylating γ-TMT in vitro with PsLSMT and determining γ-TMT enzyme activity. Unfortunately, the conditions necessary for Lys methylation (extended incubation with AdoMet and PsLSMT) resulted in γ-TMT enzyme inactivation, an observation which was not surprising given the documented lability of γ-TMT (40-42). However, Lys methylation could affect other in vivo aspects of γ-TMT such as stability, membrane localization, or interactions with other proteins, and these could have significant indirect effects on activity, which might not be manifested by in vitro Lys methylation and γ-TMT activity studies. Therefore, we evaluated in vivo γ-TMT activity as well as amounts of the γ-TMT product, α-tocopherol, in tobacco LSMT (NtLSMT)-knockdown tobacco plants, where the level of NtLSMT was significantly reduced (data not shown). An analysis of α-tocopherol amounts revealed a 46% reduction in the NtLSMT-knockdown plants compared with untransformed plants (Fig. 7). However, despite a number of attempts to confirm the reduction of γ-TMT activity using enzymatic assays, none were successful. Assays using crude leaf lysates or whole chloroplast suspensions as sources of γ-TMT were always subject to rapid losses in
activity that could be explained by the aforementioned instability of the enzyme. Other attempts using measurements of radiolabel incorporation from [3H-methyl]AdoMet into α-tocopherol using leaf disks from NtLSMT-knockdown plants did not show any significant difference in γ-TMT activity between NtLSMT-knockdown and untransformed plants (data not shown). This could be due to the kinetic parameters of the enzyme, reported to have a very low V\textsubscript{max}, on the order of fkat mg\textsuperscript{-1} of enzyme (spinach = 7.61 nmol h\textsuperscript{-1} mg\textsuperscript{-1} protein; pepper = 0.18 nmol h\textsuperscript{-1} mg\textsuperscript{-1} protein, (40;41)). Therefore, measurements of product under physiological conditions were taken as a good indirect indication of a change of the in vivo activity of γ-TMT. We also noted a chloroplast-membrane bound polypeptide in extracts from knockdown plants that could be methylated by PsLSMT (data not shown), but in the absence of specific antibodies we could not unambiguously identify this polypeptide as γ-TMT.

**DISCUSSION**

LSMT catalyzes the trimethylation of lysyl residue 14 in the LS of Rubisco by a processive mechanism (35) and like other SET-domain PKMTs contains separate binding domains for the polypeptide substrate and AdoMet co-factor (20). Until recently, PKMTs were believed to exhibit high polypeptide substrate specificity catalyzing site-specific methylation of a single lysyl residue in a single polypeptide substrate. However, the discovery of multiple polypeptide substrates for SET7/9 (16-18), SET1 (13), and Smyd2 (12) confirm a broader role for PKMTs and flexibility in the polypeptide substrate binding cleft. At least one study, capitalizing on high-resolution ternary complexes formed between SET 7/9 and polypeptide substrates, used a consensus recognition sequence to predict other potential polypeptide substrates for SET7/9 and verified the substrate in vitro (18).

Although LSMT has provided important structural and molecular information regarding the SET-domain and the catalytic mechanism for protein lysyl methylation (20;21;35;43), ternary complexes between LSMT and polypeptide substrates has not proven feasible as with other SET-domain PKMTs. Small synthetic polypeptides do not bind with appreciable affinity to LSMT and co-crystallization with Rubisco has been unsuccessful. However, molecular modeling studies have provided structural suggestions of how LSMT may recognize and bind the highly conserved and disordered N-terminal region of the large subunit of Rubisco, and what characteristics may be influential in establishing specificity (21). The model suggests that, in general, only amino acid residues with small side chains are tolerated in the polypeptide binding site with the exception of the P\textsubscript{-1} position, where there may be a requirement for a residue with a relatively large aromatic side chain accommodated by a relatively deep hydrophobic pocket next to the Lys-14 binding site. For the most part, the results presented here confirm these predictions with a few notable exceptions. Substitutions at the P\textsubscript{-1} position confirm a requirement for a residue with an aromatic side chain of precise dimensions since the only substitution catalytically tolerated by PsLSMT was a tyrosine residue. Tyrosine has a side chain volume comparable with phenylalanine, while other aromatic substitutions such as tryptophan or histidine were not recognized by PsLSMT as substrates and have side chains with volumes either smaller or larger than phenylalanine. PsLSMT barely tolerated substitutions with Ser at the two flanking P\textsubscript{-2} and P\textsubscript{+2} positions and was completely intolerant of other substitutions. Thus, there could be steric constraints for small side chain residues although Ala was also not tolerated. If rotational freedom is required to assist in polypeptide substrate binding, by allowing a kinked polypeptide conformation, glycine residues would be necessary at these positions. A kinked peptide conformation has been observed for ternary complexes between H3 peptides and SET7/9 (44). The three other positions investigated as possible determinants in the polypeptide substrate specificity of PsLSMT for Rubisco, P\textsubscript{3}, P\textsubscript{+1}, and P\textsubscript{+3} yielded some results which were considerably less predictable from the molecular docking model. Substitutions which resulted in equal or larger PsLSMT catalytic efficiency than the unmodified fusion protein were all located at these positions. PsLSMT tolerated all replacements at the P\textsubscript{3} position; Val-11 to Lys and Val-11 to Ser were substrates which supported
PsLSMT catalytic efficiencies approximately three and two fold above the original fusion substrate respectively, and the Val-11 to Ala substitution nearly equal to the unchanged fusion protein. Val-11 to Phe and Val-11 to Glu were still acceptable as substrates for PsLSMT albeit supporting lower enzyme efficiencies, making the P_{i-3} site the most tolerant of changes and likely able to accept any substitution. Surface residues in PsLSMT surrounding the polypeptide binding cleft within van der Waals distance of the Val-11 position include Leu-232, Asn-231, and Arg-220. Thus, potential interactions between these residues and the polypeptide substrate do not readily explain the increased catalytic efficiency supported by use of the Val-11 to Lys- and Val-11 to Ser-modified polypeptide substrates. However, at the P_{i-3} position normally occupied by Val-17, adjacent surface residues of PsLSMT are mostly aromatic and hydrophobic including Phe-224, Leu-227, and Ala-253. The higher efficiencies of PsLSMT noted using the Val-17 to Leu substitution as substrate and the lower but measurable efficiency using either the Val-17 to Ala or Val-17 to Phe substitutions are all compatible with potential interactions with these residues. The increased tolerance by PsLSMT of the flanking P_{i-1} (Ala-15) position for replacement by Lys or Arg suggest a preference at this site for a positively charged residue. However, this site is near surface residues in PsLSMT that are also positively charged including His-252 and Arg-226. It is possible that the higher efficiencies with PsLSMT observed using the Ala-15 to Lys as substrate is a consequence of methylation at Lys-14 as well as Lys-15 since we were unable to independently discern the methylation status at these two sites. It is interesting to note that a consensus sequence derived for the Lys-72 methylation site in cytochrome c has been described as X-K-K-X where the first K is the methylation site and X is any residue (45).

The identification of several other plastid-localized proteins as potential substrates for PsLSMT and the confirmation of γ-TMT as shown here, demonstrate the utility of this approach for the potential identification of alternative substrates for PsLSMT. However for a number of reasons including protein folding and membrane localization, these polypeptides may not necessarily be methylated in vivo. In this case the use of NtLSMT-knockdown plants provides a convenient tool for the verification and examination of the functional aspects of PsLSMT-catalyzed protein methylation. The identification of multiple protein substrates for SET-domain PKMTs adds an increased level of complexity to a class of enzymes previously thought to have a rather high degree of polypeptide substrate specificity. Knowledge of the diversity of polypeptide substrates used by SET-domain PKMTs is important in furthering our understanding of the biological role of protein methylation and additionally supports the idea that PKMTs may play more of a global role in regulating cellular processes. Moreover, in the absence of knowledge identifying the diversity of polypeptide substrates for any particular PKMT, efforts to dissect in vivo functional significance through insertion mutagenesis or gene silencing will be potentially confounded by the pleiotropic effects of (des) methyl forms of multiple substrates.

REFERENCES


8
FOOTNOTES

ψThis work was supported by the Department of Energy grant (#DE-FG02-92ER20075) to RLH.

The abbreviations used are: SET, founding members of the domain SU-(VAR)3-9, E(Z), and TRX; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; PKMT, protein lysine methyltransferase; AdoMet, S-adenosyl-L-methionine; MeLys, ε-N-monomethyllysine; Me2Lys, ε-N-dimethyllysine; Me3Lys, ε-N-trimethyllysine; PsLSMT, Pisum sativum large subunit of Rubisco methyltransferase; NtLSMT, Nicotiana tabacum large subunit of Rubisco methyltransferase; LS, large subunit; PVDF, polyvinylidene difluoride; γ-TMT, γ-tocopherol methyltransferase; TLC, thin layer chromatography; SDS-PAGE, polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate; TCA, trichloroacetic acid; IPTG, isopropyl β-D-1-thiogalactopyranoside; EDTA, ethylenediaminetetraacetic acid; PCR, polymerase chain reaction.

FIGURE LEGENDS

Fig. 1. Characterization of the LS:HCA II fusion protein. A, In groups of three from left, Rubisco (a), HCA II (b), and LS:HCA II (c) as visualized by Coomassie Blue-stained SDS-PAGE, immunoblots with antibody against the N-terminus from the LS of Rubisco, and phosphor-image of radiolabel incorporation from [3H-methyl]AdoMet catalyzed by PsLSMT. B, Specificity of PsLSMT for Lys-14 in the LS:HCA II polypeptide substrate as demonstrated by incorporation of radiolabel from [3H-methyl]AdoMet after changing Lys-8, Lys-18, and Lys-21 to Arg or Ala and the absence of radiolabel incorporation from [3H-methyl]AdoMet after changing Lys-14 to Arg or Ala. The top panel is a Coomassie Blue-stained SDS-PAGE gel and the lower panel corresponds to a phosphor-image of PVDF blot from a similar gel. HCA II is shown as a control. C, TLC product analyses of radiolabeled methylated lysyl residues of protein hydrolysate from LS:HCA II after incubation with [3H-methyl]AdoMet and PsLSMT.

Fig. 2. Kinetic analyses of PsLSMT with LS:HCA II as a polypeptide substrate. A, PsLSMT velocity plots with varying concentrations of LS:HCA II at saturated levels of AdoMet, and B, varying concentrations of AdoMet at saturating levels of LS:HCA II. Kinetic constants were derived from fitting the data to the Michaelis-Menten equation with SigmaPlot (version 10).

Fig. 3. Analysis of PsLSMT polypeptide substrate specificity. The catalytic efficiency (kcat/Km) of PsLSMT was evaluated as a consequence of conservative and non-conservative single amino changes in LS:HCA II. The top row shows the amino acid sequence surrounding the Lys-14 target methylation site in the LS of Rubisco (and LS:HCA II) and the substitutions introduced in the sequence are reported in the left column. Data are reported as mM⁻¹min⁻¹ and changes relative to unmodified LS:HCA II (0.52 mM⁻¹min⁻¹) are indicated in color. Green indicates substitutions that increased substrate acceptability, yellow for substitutions with values comparable to unmodified LS:HCA II, red for substitutions with lower substrate
acceptability, and grey for substitutions that resulted in \( PsLSMT k_{cat}/K_m \) values too low for those fusion proteins to be considered substrates. Black indicates LS:HCA II proteins with either expression levels too low for purification or expression only as insoluble inclusion bodies. L indicates that the estimation of \( k_{cat}/K_m \) was made from a linear plot as indicated in Experimental Procedures. Dashes indicate mutations which were not tested.

**Fig. 4.** Kinetic analyses of \( PsLSMT \) with Ala-15 to Lys LS:HCA II and the double mutant Lys-14 to Arg and Ala-15 to Lys LS:HCA II polypeptide substrates. A, \( PsLSMT \) velocity plots with Ala-15 to Lys LS:HCA II and B, Lys-14 to Arg and Ala-15 to Lys LS:HCA II at saturating levels of AdoMet. C, \( PsLSMT \) velocity plots with AdoMet and saturating levels of Ala-15 to Lys LS:HCA II. Kinetic constants were derived from fitting the data to the Michaelis-Menten equation with SigmaPlot (version 10).

**Fig. 5.** Identification of potential alternative \( PsLSMT \) polypeptide substrates. Using the consensus recognition motif derived from the data in Fig. 3 (indicated at top), chloroplast-localized candidate polypeptides were identified in Swiss-Prot databases by ScanSite. The data were culled to remove the 367 Rubisco LS sequences which were also identified. Each of the four candidate polypeptides and the corresponding target methylation sequence (the middle lysyl residue) are shown.

**Fig. 6.** Demonstration of \( \gamma \)-TMT as a novel substrate for \( PsLSMT \). \( \gamma \)-TMT from *Perilla frutescens* was expressed in *E. coli* and purified as indicated in Experimental Procedures. The purified enzyme was tested as a polypeptide substrate for \( PsLSMT \) by incubation with [\(^3\text{H-methyl}]\text{AdoMet} \) and after SDS-PAGE, proteins were electroblotted to a PVDF membrane and the membrane imaged for radiolabel. A, lane 1, molecular mass markers (in kDa), lane 2, purified \( \gamma \)-TMT incubated with [\(^3\text{H-methyl}]\text{AdoMet} \) alone, lane 3, purified \( \gamma \)-TMT incubated with [\(^3\text{H-methyl}]\text{AdoMet} \) and \( PsLSMT \). B, The phosphor-image of lanes 2 and 3 from A. C, Phosphor-image of TLC product analyses of radiolabeled methylated lysyl residues from the protein hydrolysate of \( \gamma \)-TMT after its incubation with [\(^3\text{H-methyl}]\text{AdoMet} \) and \( PsLSMT \).

**Fig. 7.** Determination of \( \alpha \)-tocopherol amounts in \( NtLSMT \)-knockdown tobacco plants. Leaf disks from \( NtLSMT \)-knockdown tobacco plants were analyzed as described in Experimental Procedures for the \( \gamma \)-TMT product, \( \alpha \)-tocopherol. Data presented are the means from 4 replications \( \pm \) SD. The asterick (*) represents the Anova analyses (randomized complete block) which had an F-ratio significant at the 5% level for the \( \alpha \)-tocopherol amounts of \( NtLSMT \)-knockdown and untransformed plants.
Figure 2

**Panel A**

- **Turnover (min⁻¹)** vs **LS: HCA II (µM)**
- $k_{cat} = 0.65$ min⁻¹
- $K_m = 1.2$ mM

**Panel B**

- **Turnover (min⁻¹)** vs **AdoMet (µM)**
- $k_{cat} = 1.11$ min⁻¹
- $K_m = 19.23$ µM
### Figure 3

<table>
<thead>
<tr>
<th></th>
<th>V11</th>
<th>G12</th>
<th>F13</th>
<th>K14</th>
<th>A15</th>
<th>G16</th>
<th>V17</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1.50</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>H</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>0.33</td>
<td>0.06</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.32</td>
<td>-</td>
</tr>
<tr>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>0.25</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>W</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>E</td>
<td>0.30</td>
<td>-</td>
<td>-</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>0.90</td>
<td>0.27</td>
<td>-</td>
<td>-</td>
<td>0.04</td>
<td>0.14</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.04</td>
</tr>
<tr>
<td>A</td>
<td>0.55</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.19</td>
</tr>
<tr>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.65</td>
</tr>
<tr>
<td>E</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>V</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.08</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$k_{m}/K_m$ (mM$^4$/min$^3$)
Figure 4

A

Turnover (min\(^{-1}\))

\[ k_{\text{cat}} = 1.65 \text{ min}^{-1} \]
\[ K_m = 0.71 \text{ mM} \]
\[ k_{\text{cat}}/K_m = 2.32 \text{ min}^{-1}\text{mM}^{-1} \]

LS: HCA II (K\(_{14}K_{15}\)) (μM)

B

Turnover (min\(^{-1}\))

\[ k_{\text{cat}} = 1.16 \text{ min}^{-1} \]
\[ K_m = 2.15 \text{ mM} \]
\[ k_{\text{cat}}/K_m = 0.54 \text{ min}^{-1}\text{mM}^{-1} \]

LS: HCA II (R\(_{14}K_{15}\)) (μM)

C

Turnover (min\(^{-1}\))

\[ k_{\text{cat}} = 1.45 \text{ min}^{-1} \]
\[ K_m = 12.09 \text{ μM} \]

AdoMet (μM)
**Figure 5**

<table>
<thead>
<tr>
<th>Consensus sequence</th>
<th>P_3</th>
<th>P_2</th>
<th>P_1</th>
<th>P_0</th>
<th>P_1</th>
<th>P_2</th>
<th>P_3</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>G/S</td>
<td>F/Y</td>
<td>K</td>
<td>A/K/R</td>
<td>G/S</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Potential targets for \( \text{PsLSMT} \)**

- **Gamma-tocopherol methyltransferase.** *Arabidopsis thaliana and other species*  
  EGYKKGV
- **NAD(P)H-quinone oxidoreductase chain 5.** *Marchantia polymorpha*  
  GSYKAGL
- **Granule-bound starch synthase 1, precursor.** *Ipomoea batatas and Manihot esculenta*  
  HSYKRGV
- **Acyl carrier protein 1, precursor.** *Cassava glauca*  
  ASPKASL
Figure 6
Figure 7

[Bar chart showing ng α-tocopherol (mg fresh weight)^{-1} for NiLSMT knockdown plants vs. Untransformed plants. The chart indicates a significant difference (*) between the two groups.]
Polypeptide substrate specificity of psLSMT: a SET-domain protein methyltransferase

Roberta Magnani, Nihar R. Nayak, Mitra Mazarei, Lynnette M. A. Dirk and Robert L. Houtz

J. Biol. Chem. published online July 17, 2007

Access the most updated version of this article at doi: 10.1074/jbc.M702069200

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2007/07/18/M702069200.DC1