HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2019 June 17.

Published in final edited form as:

Nat Chem Biol. 2019 February ; 15(2): 132–140. doi:10.1038/s41589-018-0188-z.

Structural basis of 7SK RNA 5′ γ-phosphate methylation and retention by MePCE

Yuan Yang[1,2], **Catherine D. Eichhorn**[1,2], **Yaqiang Wang**[1], **Duilio Cascio**[1], and **Juli Feigon**[1,*]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA

[2]These authors contributed equally: Yuan Yang, Catherine D. Eichhorn

Abstract

Among RNA 5′-cap structures, γ-phosphate monomethylation is unique to a small subset of noncoding RNAs, 7SK and U6 in humans. 7SK is capped by methylphosphate capping enzyme (MePCE), which has a second non-enzymatic role as a core component of the 7SK RNP that is an essential regulator of RNA transcription. We report 2.0 and 2.1 Å X-ray crystal structures of human MePCE methyltransferase domain bound to S-adenosylhomocysteine (SAH) and uncapped or capped 7SK substrates, respectively. 7SK recognition is achieved by protein contacts to a 5′ hairpin-single-stranded RNA region, explaining MePCE specificity for 7SK and U6. The structures reveal SAH and product RNA in a near-transition state geometry. Surprisingly, binding experiments show that MePCE has higher affinity for capped vs uncapped 7SK, with kinetic data supporting a slow product release model. This work reveals the molecular mechanism of methyl transfer and 7SK retention by MePCE for subsequent assembly of 7SK RNP.

Keywords

RNA modification; X-ray crystallography; U6 snRNA; RNA polymerase III; NMR; 7SK RNP; methyltransferase; RNA capping; METTL16; BCDIN3D

Introduction

5′ capping is a tightly regulated step in RNA processing that is required for RNA stability, localization, and function[1,2]. All caps, which vary in an RNA- and organism dependent manner, feature one or more methyl groups added by an S-adenosylmethionine (SAM, Fig.

*Corresponding author feigon@mbi.ucla.edu.

Author Contributions

Y.Y. and C.D.E. designed and performed experiments, analyzed data, and wrote the paper, Y.W. prepared RNA samples, D.C. helped with X-ray data collection and processing, J.F. supervised all aspects of the work, analyzed data, and wrote the paper.

Declaration of Interests

The authors declare no competing interests.

Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank with the following accession codes: 6DCB (MePCE-SAH-7SK) and 6DCC (MePCE-SAH-me7SK). All other data generated or analyzed in this study are included in the published article (and its supplementary information files) or are available from the corresponding author upon reasonable request.

1a) dependent methyltransferase (MTase). A unique monomethyl γ-phosphate 5′ cap (mpppG) (Fig. 1b)[3] has been identified in a small subset of RNA polymerase III transcripts, i.e. human 7SK long noncoding RNA (lncRNA) and U6 spliceosomal RNA, rodent B2 short interspersed element RNA, and plant U3 small nucleolar RNA[3–5]. 7SK and U6 associate with the methylphosphate capping enzyme MePCE (also called BCDIN3 in mammals and Bin3 in Drosophila)[6,7], but only 7SK has been conclusively shown to be capped by MePCE[6,8].

MePCE also has a non-enzymatic function as a component of the core 7SK RNP, and knock-down significantly decreases 7SK levels in vivo[6,9]. 7SK regulates RNA polymerase II transcription elongation in higher eukaryotes by sequestering and inactivating positive elongation factor b (P-TEFb)[10,11]. P-TEFb phosphorylates RNA polymerase II C-terminal domain and negative transcription elongation factors to transition RNA polymerase II from a promoter-proximal paused state to productive transcription elongation[12]. 7SK constitutively assembles with MePCE at the 5′-end and La-related protein group 7 (Larp7) at the 3′-end[13–15]. On binding 7SK, Larp7 and MePCE interact to form a core RNP where MePCE is inactivated[7–9]. Hexim1/2 and P-TEFb subsequently assemble onto 7SK RNP to sequester P-TEFb in an inactive form[10,11,16,17].

Little is known about 7SK biogenesis, and two alternative secondary structures for 7SK have been proposed[18,19], a “linear 7SK model” with four stem-loops (Fig. 1c) [8,18], supported by chemical mapping and enzymatic footprinting experiments, and a “circular 7SK model” based on sequence conservation with up to eight stem-loops and where the first ten nucleotides at the 5′-end basepair with nucleotides near the 3′-end (Supplementary Fig. 1)[19,20]. Both RNA secondary structure models have an identical terminal stem-loop 4 (SL4), where Larp7 binds[7,21–23]. However, the secondary structures differ at the 5′-end, where MePCE binds and caps (Supplementary Fig. 1). The consensus motif for the determinants of U6 capping comprises a 5′ triphosphate, hairpin beginning at the 5′-end, and adjacent 3′ single-stranded RNA (ssRNA) with sequence AUAUAC[24]. In 7SK, the sequence and secondary structure at the 5′-end corresponding to the linear 7SK model are important for MePCE assembly in vivo[7]. Whether one or both of the proposed secondary structures participates in MePCE capping and/or core 7SK RNP assembly remains to be determined.

MePCE contains a C-terminal MTase domain that is highly conserved among eukaryotes and a highly variable N-terminal region of low sequence complexity[25]. The structure of the MePCE MTase domain bound to S-adenosylhomocysteine (SAH) (MePCE–SAH) was determined by the Structural Genomics Consortium (PDB ID 5UNA). Here, we report structures of MePCE MTase domain crystallized in the presence of 7SK substrate RNA and either SAH or SAM. Surprisingly, despite transfer of the methyl group from SAM to the 5′ γ-phosphate of 7SK substrate during crystallization, the RNA remains bound to MePCE, providing a unique example of a product-bound RNA MTase structure. Together with analysis of enzyme kinetics and reactant and product binding affinities, the crystal structures of MePCE–SAH in complex with uncapped (MePCE–SAH–7SK) and capped (MePCE–SAH–me7SK) RNAs reported here reveal the mechanism of methyl transfer to a phosphate group and the determinants of 7SK recognition and retention and define the first steps of assembly of 7SK core RNP.

Results

Crystal structures of the MePCE MTase domain bound to 7SK

Based on studies indicating that MePCE binds the proximal end of the 7SK SL1 and adjacent ssRNA in vivo (Fig. 1c)[7], we designed and synthesized an RNA hairpin-ssRNA construct comprising the 5′-end of 7SK SL1 with a UUCG tetraloop and an eight nucleotide ssRNA 3′ overhang (SL1p) (Fig. 1c, inset). SAM or SAH do not co-purify with MePCE MTase domain (residues 400–689, $MePCE_{MT}$), as determined by LC-MS (Supplementary Fig. 2). To verify that SL1p is capped by $MePCE_{MT}$ we incubated [*methyl*-^{13}C]-SAM, SL1p, and $MePCE_{MT}$ under single turnover conditions to completion and analyzed the subsequently purified SL1p by NMR spectroscopy (Fig. 1d). A unique resonance attributed to a methyl group on SL1p ([*methyl*-^{13}C]-SL1p) was observed, with a significant chemical shift difference compared to [*methyl*-^{13}C]-SAM (Fig. 1d). High-resolution electrospray ionization mass spectrometry confirmed that SL1p was fully monomethylated (meSL1p) (Supplementary Fig. 3). These assays were performed in the absence of divalent cations (Fig. 1d; Supplementary Fig. 3), indicating that MePCE does not require divalent cations for catalysis.

$MePCE_{MT}$ was crystallized in complex with SL1p in the presence of either SAH or SAM, and structures were solved from crystals that diffracted to 2.0 Å and 2.1 Å resolution, respectively (Supplementary Fig. 4; Supplementary Table 1). The crystal structure of the sample with added SAM lacked the electron density expected for the SAM methyl moiety, whereas clear additional density was observed at an RNA 5′ γ-phosphate oxygen, indicating that methyl transfer occurred during crystallization (Supplementary Fig. 4). Thus, crystal structures of $MePCE_{MT}$ in complex with SAH and SL1p in uncapped (MePCE–SAH–7SK) and capped (MePCE–SAH–me7SK) forms were determined (Supplementary Fig. 4). The structures are nearly identical (all-atom RMSD 0.21Å), with significant differences only at the SAH and 5′ triphosphate (Supplementary Figs. 4 and 5). In both structures, $MePCE_{MT}$ has the expected Rossman-like fold common to RNA methyltransferases[1,2] with a αβαβαβαβαβαββ topology (Fig. 2a,b), also seen in RNA-free MePCE–SAH (Fig. 2c). Helices α1, α2, and α3 lie underneath the central β-sheet (↑3↑2↑1↑4↑5↓7↑6) and α4, α5, and α6 lie above the β-sheet. Noncanonical helices α5′ and α6′ are located adjacent to α5 and α6, respectively, at the β-sheet edge (Fig. 2b,c). In MePCE–SAH–me7SK (Fig. 2b) and MePCE–SAH–7SK (Supplementary Fig. 4a), $MePCE_{MT}$ has two additional helices, α0 and α7, located in the structure between α3 and α5′. SAH and the RNA 5′ triphosphate are buried in the active site (Fig. 2b inset, d). Notably, no divalent or trivalent cations are observed in the active site, even for crystals soaked with Mg^{2+} or Sm^{3+} (see Online Methods). SL1p forms a ten basepair hairpin with an asymmetric (1–3) internal loop separating the bottom three and top seven basepairs and a UUCG tetraloop, and eight single-stranded nucleotides that stack on each other below the terminal basepair (Figs. 1c and 2b; Supplementary Fig. 4). The internal loop, which is outside of the $MePCE_{MT}$ binding site, has crystal contacts with nucleotides Gua113-Ade116 from another molecule in the asymmetric unit that form non-native interactions (Supplementary Fig. 6).

Comparison of MePCE–SAH–me7SK (Fig. 2b) and MePCE–SAH (PDB ID 5UNA) (Fig. 2c) highlights the conformational changes in $MePCE_{MT}$ that occur on RNA binding. The structures of MePCE–SAH in the absence and presence of SL1p are globally similar (all atom RMSD 0.29 Å), but there are major differences around the active site and RNA-protein interface (Fig. 2e,f). In the absence of RNA, the N-terminus (residues 418–431) and C-terminal β6-β7 loop (residues 665–676) are partially disordered (Fig. 2c,f), while in the presence of meSL1p, a helix α0 forms near the N-terminus, loop residues that connect to α1 become ordered, and the β6-β7 loop adopts an ordered structure with a short helix α7 in the middle (Fig. 2b,e). Together with helices α5′ and α6′, which do not change conformation significantly when RNA binds except for K625 and R626 sidechains, these elements form a tunnel that encloses the RNA triphosphate (triphosphate binding tunnel) to position it near the cofactor in the active site (Fig. 2d,g,h). These new structured elements are stabilized by hydrogen bonds (H-bonds) and stacking interactions to both RNA substrate and each other, and sequester SAH in the cofactor binding pocket with only its sulfur group exposed to the γ-phosphate in the tunnel (Fig. 2d,g,h). In contrast, in the absence of RNA, SAH homocysteine is solvent accessible (Fig. 2i). In summary, SL1p binding induces conformational changes in $MePCE_{MT}$ that enclose SAH in the active site.

Extensive interaction network between $MePCE_{MT}$ and 7SK

In MePCE–SAH–me7SK, there is an extensive interaction network between a highly basic surface on MePCE and the SL1p 5′ triphosphate, two terminal basepairs and the following six ssRNA nucleotides (Figs. 2b, 3; Supplementary Fig. 4). Gua1 and Gua2 in the first two basepairs are recognized by $MePCE_{MT}$ helices α5′, α6′ and α7, Gua1–Cyt108 and first two ssRNA nucleotides straddle helix α0, and the following four ssRNA nucleotides interact primarily with helix α3 (Figs. 2b, 3a). The interior of the triphosphate-binding tunnel (Fig. 2d; Supplementary Fig. 5) is accessible to solvent at the guanosine and α-phosphate and narrows at the β- and γ-phosphates, excluding solvent. The oxygens on the 5′ triphosphate have multiple direct H-bonds (to all three phosphates) and indirect H-bonds (to α-phosphate) from residues on helices α7, α6′, α5′, and α0 and adjoining loops (Fig. 3a–e; Supplementary Table 2) that form the tunnel. Additional H-bonds from α6′ and α7 recognize the attached Gua1 base and ribose, respectively (Fig. 3e; Supplementary Table 2). These H-bonds fix the otherwise highly flexible 5′ triphosphate to position the γ-phosphate oxygen for methyl transfer.

SL1p terminal two basepairs and following four ssRNA nucleotides (Ade109-Ura112) are sequence specifically recognized by $MePCE_{MT}$, and there are backbone and ribose interactions to Gua113 and Cyt114 (Fig. 3a). $MePCE_{MT}$ interacts in the major groove at the first two basepairs, forming H-bonds to Gua1 and Gua2, as well as the phosphates (Fig. 3e–h). F674 (α7) stacks between the Gua1 ribose and the R425 (α0) guanidinium plane, which then stacks on R433 (α1) to encompass the “right wall” of the triphosphate-binding tunnel (Figs. 2e and 3e,g). The terminal Gua1-Cyt108 basepair and ssRNA nucleotides Ade109 and Ura110 straddle helix α0, which is enriched in tyrosines whose aromatic rings splay out around α0 (Fig. 3f). This “Tyr wheel” has intricate interactions with ssRNA nucleotides outside the active site, the 5′ triphosphate within the active site, and helix α7 residues (Fig. 3a–c, f, g, i–k). Specifically, Y424 has π-π base-stacking and H-bonding interactions with

Ade109 base and ribose (Fig. 3f,i). Cyt108, Ade109 and Ura110 have continuous base stacking (Fig. 3f,j). K420 side-chain on the Tyr wheel stacks between Ura110 and Gua111 and has H-bonds to Ura110 (Fig. 3f,j,k). Gua111 has an H-bond to S482 (α3) (Fig. 3k) and stacks on Gua113 (Fig. 3f), with Ura112 base flipped 120° out about the Gua111 ribose. Ura112 base has stacking and H-bonding interactions with four residues from α2 and α3 (Fig. 3l). Ura112 is the last RNA residue with $MePCE_{MT}$ base recognition; Gua113 and Cyt114 have only backbone interactions with helix α3 residues (Fig. 3m). The residues that participate in triphosphate and basepair recognition are highly conserved from fission yeast to human, whereas helix α3 is highly conserved among vertebrates but poorly conserved for fission yeast and plants (Supplementary Fig. 7). The large MePCE-RNA interaction surface explains the in vivo mutagenesis data that a 5′ hairpin structure followed by ssRNA nucleotides is required for the mpppG cap[7,24].

Cofactor and triphosphate are in a near-transition state

The positions of SAH in MePCE–SAH–7SK and MePCE–SAH–me7SK are nearly identical with only minor rotations of γ and β phosphates and shifting of the SAH sulfur (Fig. 4a,b; Supplementary Fig. 5). The SAH adenosine and homocysteine moieties are anchored in a bidentate binding cleft at the back wall of the active site and the sulfur atom points away from the β-sheet toward the Gua1 γ-phosphate oxygen atom that peeks out of the triphosphate-binding tunnel (Fig. 2b, 4a-c). The SAH adenosine is recognized by multiple H-bonds and stacking interactions primarily from motifs I and Ia residues (Fig. 4d; Supplementary Fig. 7; Supplementary Table 2). The SAH homocysteine amide donates a H-bond to G451 (β1-α2 loop, motif I) and L581 (β4, motif II) backbone carbonyl groups, and the SAH carboxylate group has three H-bonds to the R433 guanidinium (α1) and Y422 hydroxyl (α0) (Fig. 4e; Supplementary Table 2). Notably, in RNA-free MePCE−SAH, Y422 and R433 are partially disordered (Supplementary Fig. 8). In contrast, they are highly ordered in the RNA-bound structures (Supplementary Fig. 5) indicating that RNA binding to $MePCE_{MT}$ further stabilizes SAH in its binding cleft.

In MePCE–SAH–me7SK, the methyl carbon has been transferred from SAM to the Gua1 γ-phosphate oxygen (Figs. 1b and 4g; Supplementary Fig. 4) with an O-C bond distance of 1.40 Å (Supplementary Fig. 4). Superposition of the MePCE–SAH–7SK and MePCE–SAH–me7SK reveals a local shift in the SAH sulfur by 0.89 Å, which would allow room for the methyl group (Fig. 4b). For MePCE–SAH–7SK, the distance between the SAH sulfur and Gua1 γ-phosphate oxygen (3.37 Å) equals the sum of their van der Waals (vdW) radii (3.32 Å) (Fig. 4f) (http://www.rsc.org/periodic-table), while for MePCE–SAH–me7SK the distance between the SAH sulfur and the Gua1 γ-phosphate O-methyl carbon (3.22 Å) is slightly smaller than the sum of the vdW radii (3.50 Å) (Fig. 4g). These distances and geometries show a remarkable resemblance to a transition state, where the methyl group would be located at an intermediate distance between the sulfur and γ-phosphate for in-line transfer. This close proximity of reacting atoms at the active site is achieved by both the bidentate cofactor binding cleft (Fig. 4a) that fixes the sulfur atom in place and the triphosphate-binding tunnel (Fig. 4c) through which the γ-phosphate is anchored and positioned towards the cofactor sulfur atom.

Although we do not have structural information for the starting state of the methyl transfer reaction (SAM and uncapped RNA), the fact that MePCE–SAH–7SK and MePCE–SAH–me7SK structures both have near-transition state geometry suggests a similar geometry would be present in a MePCE–SAM–7SK complex, modeled in Fig. 4h. Taken together, we infer the methyl transfer reaction mechanism to be the following: SAM cofactor binds first to $MePCE_{MT}$, and subsequent RNA substrate binding further stabilizes and sequesters SAM in its binding cleft. The negatively charged Gua1 γ-phosphate oxygen is positioned in close proximity to the SAM methyl group to promote nucleophilic attack at the positively charged methyl sulfonium moiety (Figs. 1a,b and 4h). Notably, the γ-phosphate oxygens are fully deprotonated and negatively charged at neutral pH[26], unlike protonated RNA substrates such as 2′-OH or $N2H_2$, and the SAM positively-charged methyl sulfonium group is less stable than the neutral SAH sulfur, making the MePCE methyl transfer reaction extremely electrostatically favorable and limiting the reverse reaction. After transfer of the methyl group to the Gua1 γ-phosphate oxygen, both SAH byproduct and capped RNA product remain bound in the active site (Figs. 2b and 4b,g), facilitated by the extensive protein-RNA interactions discussed above.

To investigate the functional importance of interactions in the triphosphate-binding tunnel, we made individual alanine substitutions at three highly conserved residues ranging from closest to the methyl transfer reaction to the solvent exposed Gua1 ribose, Y421A (α0), K585A (α5′), and F674 (α7) (Fig. 4c) and compared their methyltransferase activity to wild-type (wt) $MePCE_{MT}$ within a single turnover (Fig. 4i). Y421, located at the floor of the triphosphate-binding tunnel (Fig. 2b,d), donates an H-bond to and buttresses the γ-phosphate (Fig. 3a,b,f) and is the closest side-chain to the methyl group prior to transfer (Fig. 4h). Y421A is nearly catalytically dead (0.6% activity) (Fig. 4i), indicating the critical importance of Y421 for formation of the active site and stabilization of the γ-phosphate. Alanine substitution of K585, which has H-bonds to the γ- and β-phosphate oxygens (Fig. 3a–e), decreases activity to 32% (Fig. 4i) indicating K585 has a significant role within the $MePCE_{MT}$ triphosphate recognition network. Partial activity is likely maintained by the remaining nine of eleven direct H-bonds to the triphosphate and retention of the majority of the left side (helix α5′) of the tunnel. Alanine substitution of F674, which would expose the right side (helix α7) and destabilize the ceiling of the tunnel by disrupting the stacking interactions with R425 and Gua1 ribose, decreases activity to 2.5% (Fig. 4i). These results verify the structural and functional importance of the triphosphate-binding tunnel residues and highlight the significance of residues in the regions that become ordered on RNA binding.

MePCE binds tightly to capped RNA

Under single turnover conditions SL1p is fully monomethylated in less than 15 min (Fig. 1d; Supplementary Fig. 3). In time-dependent activity assays at different SL1p concentrations, we similarly observe quantitative methylation under single turnover conditions ($MePCE_{MT}$ in molar excess to RNA); however, nonlinear behavior is observed under multiple turnover conditions (Fig. 5a). These data reflect an initial fast phase corresponding to a quantitative single turnover, followed by a slow phase corresponding to a second turnover event that is limited by product release[27]. To explain these results, we first investigated the equilibrium

dissociation constants (K_D) of cofactor SAM and byproduct SAH to $MePCE_{MT}$ by isothermal titration calorimetry (ITC) (Supplementary Fig. 9a,b). $MePCE_{MT}$ binds SAM with 13-fold lower affinity compared to SAH, suggesting that SAH may be a product inhibitor to MePCE, as seen in some SAM-dependent Rossman-fold MTases[28–32].

We next investigated by ITC the binding of $MePCE_{MT}$ to various RNA substrates including SL1pΔss (no ssRNA), SL1pΔ1–3 (no internal loop), and three pairs of unmethylated and methylated constructs–SL1p (Fig. 5b,c), S1alt from the proposed 5′ stem of the alternative circular 7SK model, and SL1 of the U6 snRNA (U6-SL1 and meU6-SL1), the only other predicted substrate of MePCE in humans (Fig. 5d; Supplementary Fig. 9; Table 1; Supplementary Table 3). To ensure that only binding energies were measured, the experiments were performed in the presence of SAH rather than SAM. SL1pΔss had no detectable binding by ITC (K_D >> 1 μM) indicating that the ssRNA 3′ overhang is required for high affinity binding. In contrast, no difference in binding affinity was observed between SL1p and SL1pΔ1–3, consistent with a lack of interactions observed between $MePCE_{MT}$ and the internal asymmetric loop in the crystal structures.

U6-SL1 has a 4-fold higher binding affinity compared to 7SK SL1p (Table 1; Supplementary Table 3; Supplementary Fig. 9). Like SL1p, U6-SL1 has a terminal Gua-Cyt basepair and although the penultimate basepair is a Ura-Ade instead of a Gua-Cyt, equivalent H-bonds could be made between protein and base. The first 6 nts of the 3′ ssRNA sequence are alternating purine and pyrimidine nucleotides (AUAUACU), versus AUGUGCG in 7SK SL1p (Fig. 5d). The two Ade are expected to stack similarly to SL1p Gua111 and Gua113[33], and the top Ade, like Gua111, could accommodate an H-bond to S482 (Fig. 3f,k). In contrast, the binding affinity of $MePCE_{MT}$ to S1alt is 8-fold lower than to SL1p. S1alt and SL1p both have two Gua-Cyt basepairs, but in S1alt the 3′ ssRNA region is not alternating purine-pyrimidine and is three nucleotides shorter (AAAUG) (Fig. 5d). Given that SL1p binds $MePCE_{MT}$ with higher affinity and that MePCE likely associates with 7SK as it is being transcribed before the 5′-end and the region near the 3′-ends could basepair[9], these data suggest that the linear 7SK model is the substrate for MePCE, consistent with biochemical experiments[8].

$MePCE_{MT}$ binds to meSL1p with ~2-fold higher affinity than SL1p ($\Delta\Delta G_{cap}$ 0.6 ± 0.3 kcal/mol) (Table 1; Supplementary Table 3; Supplementary Fig. 9). The unexpected finding that the product binds with higher affinity than the substrate explains how 7SK is retained by MePCE after capping. MePCE binds to S1alt almost an order of magnitude more weakly than to SL1p, but binds meS1alt with comparable affinity to uncapped SL1p, with an ~5-fold increase in affinity for meS1alt vs S1alt ($\Delta\Delta G_{cap}$ 1.0 ± 0.2 kcal/mol). These results do not exclude an assembly model in which MePCE binds circular 7SK after capping of linear 7SK.

Although $MePCE_{MT}$ binds tightly to both capped and uncapped U6-SL1 ($\Delta\Delta G_{cap}$ 0.0 ± 0.2 kcal/mol) with comparable affinities, the affinity is about the same as for meSL1p, suggesting that MePCE remains bound to U6 snRNA after capping, consistent with MePCE pull-down assays[6,7].

Catalytic turnover is inhibited by product and byproduct binding

The above results indicate that both capped RNA and SAH byproduct can act as inhibitors. To validate the proposed product inhibition mechanism, we performed methyltransferase activity assays to determine if binding affinities are correlated with turnover (Fig. 5e and Supplementary Fig. 10). As discussed above, $MePCE_{MT}$ exhibits fast kinetics within one turnover and a slow phase for the second turnover and beyond that is likely limited by product release (Fig. 5a). As expected, addition of Mg^{2+} to the methyltransferase assay had no statistically significant effect on activity. Under multiple turnover conditions, S1alt has 19% increase in *apparent* activity relative to SL1p. As meS1alt has reduced binding affinity compared to meSL1p, these results suggest that meS1alt has weaker product inhibition, allowing $MePCE_{MT}$ to proceed to multiple turnovers. In contrast, the *apparent* activity on U6-SL1 is 36% less than on SL1p. We note that under multiple turnover conditions with U6-SL1, one stoichiometric methylation is completed at around 15 min, indicating the decreased activity with U6-SL1 is not due to an abortive methyl transfer reaction, but rather to a decreased rate of product release, consistent with higher binding affinity of MePCE for meU6-SL1 than meSL1p (Supplementary Table 3). SL1pΔss has no detectable binding to $MePCE_{MT}$ by ITC but is capped, although with 39% lower activity. This moderately reduced activity probably reflects both weaker protein-RNA binding and drastically reduced product inhibition. As further support for the product inhibition model, we doped the MTase assays with SAH or meSL1p, which each decreased activity by 23% and 34%, respectively.

In summary, the enzymatic activity of MePCE is an extreme case of product inhibition wherein both byproduct and product have higher affinity toward the MTase domain than the corresponding cofactor and RNA substrate. The retention of a monomethylated RNA product also explains why MePCE only methylates a single γ-phosphate oxygen. Together, the structural, binding, and enzymatic data demonstrate that MePCE generally functions as a single-turnover enzyme and explains the dual role of MePCE in 7SK RNP biogenesis in capping the 5′ γ-phosphate to protect RNA from degradation and as a constitutive component of the 7SK core RNP.

Comparison to other MTases highlights unique features of MePCE

We compared substrate and cofactor binding sites of MePCE to structures of two other RNA capping MTases bound to substrate, VP39 (PDB ID 1AV6), a well-studied mRNA ribose O2′-cap1 MTase, and Tgs1, an mRNA TMG-capping MTase (PDB ID 3GDH) (Supplementary Fig. 11). In contrast to the large RNA binding surface on MePCE, VP39 and Tgs1 have base-specific contacts to only two or one nucleotides, respectively. MePCE has a large positively-charged surface area and buried active site with an enclosed cofactor, whereas VP39 and Tgs1 have much smaller positively-charged surface areas and relatively shallow active sites with solvent accessible cofactor. In the MePCE active site, methyl donor and acceptor atoms are within vdW radius (Fig. 4f,g; Supplementary Fig. 11g). The donor and acceptor atoms in VP39 and Tgs1 are farther apart, likely because both the ribose 2′-OH group (VP39) and the N2 amino group (Tgs1) require a deprotonation step prior to methyl transfer (Supplementary Fig. 11h-i).

BCDIN3D[34,35], a recently discovered MePCE paralog that methylates the 5′ α-phosphate of $tRNA^{His}$ and pre-miRNA, has significant sequence homology to MePCE including helix α0 Tyr wheel residues and residues involved in 5′-end and cofactor recognition (Supplementary Fig. 7), suggesting it may recognize its 5′ hairpin-single-stranded substrates in a similar way to MePCE. In contrast, MTase METTL16 that also modifies 7SK and U6, lacks helix α0 and conserved residues involved in 5′-end recognition by MePCE, consistent with its internal m6A modification function[36,37]. This comparison of RNA-capping MTases highlights the unique features of MePCE: a large substrate recognition surface; enclosed active site; and substrate and cofactor poised in near-transition state geometry.

Discussion

RNA 5′ caps are required for RNA biogenesis, stability, splicing, cellular localization and protein interactions[1,2]. However, the molecular mechanisms of the MTases responsible for capping have not been well characterized, primarily due to a lack of structural information on substrate recognition. The MePCE–SAH–me7SK crystal structure is to our knowledge the first structure of an RNA MTase in complex with its capped RNA product at the active site. Together with the MePCE–SAH–7SK crystal structure, a reaction mechanism where the reactants and products remain in a near-transition state geometry in the active site for the entire reaction process is revealed. Comparison to the substrate-free MePCE–SAH structure shows a disorder-to-order conformational change for the N- and C-terminal regions upon binding RNA that encapsulates and blocks diffusional exit of the cofactor and positions the substrate for reaction. The structures also reveal atomic details of an extensive interaction network that recognizes 5′ triphosphate and basepaired and single-stranded nucleotides by both unique and canonical Rossman-fold methyltransferase α-helices. Significantly, the binding affinity of capped 7SK SL1p to $MePCE_{MT}$ with SAH is higher than that of the uncapped RNA. The higher affinity of products over reactants to MePCE explains how MePCE becomes a stable component of the 7SK RNP after catalysis. Structural and mechanistic comparisons of MePCE to other RNA MTases described above explain further why MePCE is inhibited by both product and byproduct, and how it retains product in the active site. Although there are apparently some protein MTases that retain their products, e.g. protein L-isoaspartyl methyltransferase[30] and protein arginine methyltransferase 3 (PRMT3) through its N-terminal Zn-finger domain[38], MePCE is the only known RNA MTase that retains product after catalysis.

MePCE concentrates at the 7SK and U6 promoter regions[9], and it is likely that 5′ capping occurs co-transcriptionally once the specific helix-single-stranded RNA region is transcribed. Our binding analysis and activity assays indicate that U6-SL1 interacts with MePCE with higher affinity than 7SK and is capped by MePCE in a similar product-inhibited manner. These observations validate U6 as a MePCE substrate and explain the association of MePCE with U6 in vivo (based on pull-down assays), at least prior to formation of the U4/U5/U6 tri-snRNP[6,7,39]. An RNA construct corresponding to the 5′-end in the circular 7SK model binds to $MePCE_{MT}$ with ~8-fold lower affinity than a construct corresponding to the linear 7SK model, and in vivo the circular 7SK model could not form until transcription of full-length 7SK is nearly complete, suggesting that the linear 7SK

model is the substrate for MePCE capping. However, Larp7, which has strand melting and annealing properties[40] common to LARPs and La protein[41,42] may remodel 7SK on assembly, altering the MePCE-RNA interface and promoting the stable association between MePCE and Larp7. The retention of MePCE on 7SK as a core component of the 7SK RNP highlights the unique nature of 7SK – it is capped at the 5′-end by a specialized MTase that remains bound after catalysis and the same MTase is then joined with a 3′-end-associating La-related protein to form a stable ternary RNP. This work provides a structural and mechanistic basis for the role of MePCE in the first steps of 7SK RNP assembly.

Online Methods

Protein expression and purification

The pET28a-LIC recombinant plasmid with gene coding for His_6-tagged human MePCE (FLJ20257) MTase domain (residues 400–689) was a gift from Cheryl Arrowsmith (Addgene plasmid 25304) and was transformed into *Escherichia coli* BL21-Gold(DE3) (Agilent Technologies) for protein expression. Protein point substitutions were carried out with Q5 site-directed mutagenesis kit (NEB) and verified by DNA sequencing. Bacterial cultures were grown in minimal media at 37 °C to $O.D._{600}$ 0.6, transferred to 18 °C for an hour prior to induction with 0.5 mM IPTG for 18 hours. Cells were pelleted, resuspended with Resuspension buffer (Buffer R, 20 mM HEPES pH 8, 1 M NaCl, 1 mM TCEP, 15 mM imidazole, 5% glycerol, 1mM PMSF) supplemented with lysozyme and sonicated. Cell lysate was clarified by centrifugation and filtration, and the His_6-tagged proteins were purified by Ni sepharose affinity column (HisTrap HP; GE Healthcare) and further purified by size exclusion chromatography (SEC; HiLoad 26/600 Superdex 75; GE Healthcare) in Crystallization buffer (Buffer C, 20 mM HEPES pH 7.5, 250 mM KCl, 1 mM TCEP) or Binding buffer (Buffer B, 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM TCEP). The concentration of protein peak fractions was measured and the protein was stabilized by adding 2–3-fold molar-excess of S-(5′-Adenosyl)-L-homocysteine (SAH) (Sigma A9384) or S-(5′-Adenosyl)-L-methionine p-toluenesulfonate salt (SAM) (Sigma A2408; ≥80% purity) cofactors.

In vitro transcription of RNA

RNA samples were prepared by in vitro transcription as described previously[22] using chemically synthesized DNA templates (Integrated DNA Technologies). Briefly, the transcription reaction was incubated for 4–6 hours at 37 °C, followed by purification by 15–20% denaturing polyacrylamide gel electrophoresis, excision of RNA and elution using an Elutrap device (GE Waters). Collected RNA eluents were buffer exchanged into sterilized nanopure water, supplied with counter ions by addition of high salt buffer (1.5 M KCl), and briefly buffer exchanged into nanopure water before being annealed by heating to 95 °C for 5 minutes and snap cooling on ice for 1 hour. RNAs were then desalted into sterilized nanopure water or Buffer C/B and concentrated with Amicon ultra centrifugal filters (3 kDa MWCO, Millipore). Stock solutions of RNA were stored in −20 °C freezer.

Methylated RNA sample preparation

MePCE$_{MT}$, RNA and SAM were mixed in molar ratios of 1:1:3 or 0.5:1:3 and incubated at 37 °C for >24 hours in buffer B. RNA was purified by phenol chloroform extraction, followed by ethanol precipitation and buffer exchanged into nanopure water. Sample homogeneity was verified by mass spectrometry (Supplementary Figs. 3 and 12–14).

Crystallization and data collection

All crystallization trials were performed using hanging drop vapor diffusion method at 4°C with MePCE$_{MT}$ in complex with 7SK SL1p RNA and SAM or SAH at 8–10 mg/mL concentration. The molar ratio of cofactor-bound MePCE$_{MT}$ and SL1p RNA was determined by electrophoretic mobility shift array (EMSA) where the RNA reached stoichiometric saturation by protein. In our hands, the SL1p RNA was saturated by 1.5 molar access of MePCE$_{MT}$ in the presence of SAM or SAH. To assemble the MePCE MTase−SL1p complex, a dilute protein solution from the SEC elution in buffer C with added SAM or SAH was combined with RNA in buffer C and concentrated with 3 kDa MWCO Amicon ultra centrifugal filters (Millipore) at 4°C. The final molar concentrations of MePCE$_{MT}$ (ε = 51715 L·mole^{-1}·cm^{-1} at A_{280}) is 0.19–0.24 mM and SL1p RNA (ε = 317600 L·mole^{-1}·cm^{-1} at A_{260}) is 0.13–0.16 mM.

Prior to crystallization, the protein–RNA solutions were mixed with the well solution buffers at a 2:1 v/v ratio. For the MePCE–SAH−7SK structure, single crystals grew in about one week with a well solution of 0.1 M phosphate/citrate, pH 4.2, 0.214 M lithium sulfate and 17.25% w/v PEG 1,000. The crystals were transferred in a cryoprotectant containing 50% glycerol and flash frozen in liquid nitrogen. The structure was solved by molecular replacement using Phaser[43] with the MePCE−SAH structure (PDB ID: 5UNA, chain E) as a search model. RNA density for the 5′-end triphosphate and the first three basepairs was apparent in the electron density map. RNA was built into electron density with Coot[44] and iteratively refined to 2.0 Å with Phenix[45].

For the MePCE–SAH−me7SK structure, the sample was prepared as described above using SAM instead of SAH. Single crystals grew in approximately one week and continued to grow over two months in a well solution of 0.1 M phosphate/citrate, pH 4.2, 0.2 M lithium sulfate and 25% w/v PEG 1,000. The crystals were transferred to an undiluted drop of well solution supplemented with 1.2 mM $MgCl_2$, soaked for 2 hours and flash frozen in liquid nitrogen. The structure was solved by molecular replacement using Phaser with the refined MePCE–SAH−7SK model. The additional methyl group on the γ-phosphate was apparent from both $2F_o$-F_c and F_o-F_c maps (Supplementary Fig. 4). The γ-phosphate monomethylated GTP geometry was built with phenix.elbow[46], and the final structure was refined to 2.1 Å with Phenix. Independent soaking experiments were also performed with both MePCE–SAH−me7SK and MePCE–SAH−7SK crystals in the following conditions: 1.2 mM $MgCl_2$, 1.2 mM $SmCl_3$, 2.4 mM $MgCl_2$ and 2.4 mM $SmCl_3$. Two soaking times – overnight or 2 hrs – were used for all conditions. MePCE–SAH−me7SK crystal from the 2 hr 1.2 mM $MgCl_2$ condition was used for final refinement and deposition.

All data were collected on beamline 24-ID-C (NE-CAT) at the Advanced Photon Source at Argonne National Laboratory (APS-ANL, Argonne, IL, USA) at a wavelength of 0.9792 Å and temperature of 100 K with a Dectris Pilatus 6M detector. Data were indexed and integrated using XDS and scaled using XSCALE[47]. Final iterative rounds of model building and refinement were performed using Coot and Phenix with TLS refinement[48]. Composite refine omit maps[49] were generated with Phenix and PyMOL 2.0.6 was used to generate structural figures. Crystallization data collection and refinement statistics are presented in Supplementary Table 1. Electrostatic potential of MePCE–SAH–7SK, VP39, and Tgs1 structures was calculated with Adaptive Poisson-Boltzmann Solver (APBS) through PDB2PQR[50] and APBS[51] webservers.

NMR spectroscopy

Capped RNA samples for ^{13}C NMR experiments were generated by incubating freshly purified $MePCE_{MT}$, S-methyl-^{13}C SAM (99 atom % ^{13}C; Sigma Aldrich/ISOTEC), and RNA samples in a molar ratio of 1:3:1 in Buffer C at 37 °C for one hour. The reactions were quenched and RNA separated by phenol-chloroform extraction. The RNA samples were further purified by ethanol precipitation, air dried and dissolved in NMR buffer. The uncapped and capped RNA NMR samples contained 50–100 μM RNA, in a buffer of 10 mM sodium phosphate at pH 6.2, 60 mM KCl, 5% (v/v) D_2O and 0.6 mM NaN_3. All NMR spectra were collected at 15 °C using Bruker DRX 500 MHz and Avance 800 MHz spectrometers equipped with cryogenically cooled probes. Data was collected with TopSpin (Bruker), processed with NMRPipe[52] and analyzed in NMRFAM-Sparky[53].

1D ^1H flipback NMR spectra[54] was used to quantify (S,S)- and (R,S)-diastereomers of SAM. SAM salts were dissolved in the RNA NMR buffer. The peaks at 3.01 ppm and 2.97 ppm were assigned to methyl protons from (S,S)- and (R,S)-diastereomers, respectively, based on the decrease in relative intensity of (S,S)-peak after a 7-day incubation and previously reported chemical shift values[55]. SAM p-toluenesulfonate salt was found to contain 76.63% and SAM iodide salt (Sigma A4377) contained 73.79% (S,S)-diastereomer. All SAM concentrations in the methyltransferase and ITC experiments refer to the total concentration of SAM as measured by A_{260}. However, the SAM concentrations were adjusted to the active (S,S)-diastereomer when calculating molar methyl transfer.

Mass spectrometry analysis of RNA and purified $MePCE_{MT}$ protein

The uncapped and capped RNA samples were desalted with ammonium acetate precipitation to replace the Na^+ on the RNA backbone[56]. The desalted RNA samples were dissolved in ultrapure water to 20 μM and analyzed with direct infusion method on a Q exactive plus orbitrap mass spectrometer (ThermoFisher Scientific) on negative mode with 70,000 resolving power at 200 m/z. 100 scans were acquired to improve the S/N.

SAM/SAH standard mixture and freshly purified $MePCE_{MT}$ protein were analyzed on ZORBAX C3 column 3.0×150 mm with bead size 3.5 μm (Agilent Technologies) using Agilent 6530 Q-TOF ESI with a 1260 Infinity LC. Buffer A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. A linear gradient from 1–99% B

was applied for a duration of 8 min and elutions were analyzed on positive mode of ESI-TOF mass spectrometer.

Methyltransferase assay

In vitro RNA methyltransferase assays were adapted from established protocols[57,58]. All experiments were performed in a reaction buffer of 50 mM HEPES, pH 7.5, 150 mM KCl, 1 mM TCEP at 37 °C with 1 μM $MePCE_{MT}$ and 20 μM SAM. SAM is a mixture of 20:1 molar ratio of S-(5′-Adenosyl)-L-methionine p-toluenesulfonate salt (SAM) (Sigma A2408; ≥80% purity) to S-(5′-Adenosyl)-L-[methyl-^{3}H]-methionine ([me^{3}H]-SAM) (stock solution of 6.7 μM (81.9 Ci/mmol) in 10 mM H_2SO_4/EtOH (9:1, v/v); PerkinElmer Life Sciences). Reaction mixtures of 25 μL without RNA substrates in thin-walled PCR tubes were incubated at 37 °C for 5 min, and reactions were initiated by adding 5 μL of RNA stock solutions (6X concentrations in the range of 3–60 μM). Reactions were quenched by prompt heating to 98 °C and maintained for 2 min in a pre-heated thermocycler at different time points. 25 μL out of total 30 μL reaction mixtures were spotted on Amersham Hybond-N+ membrane (GE Healthcare), air-dried for 30 min, washed with 100 mL of 50 mM $NaHCO_3$ pH 9 solution for 45 min twice and air-dried for 90 min. The membranes were placed in vials and soaked with 5 mL of Safety-Solve scintillation solution (Research Products International 111177) overnight and counted for three cycles of 5 min using a Beckman scintillation counter. Product RNA concentration is quantitated by converting cpm counts to molar concentrations with 58% counting efficiency for the scintillation counter, SAM-to-[me^{3}H]-SAM molar ratio, and (S,S)-diastereomer purity of 76.63% as determined from NMR.

The progress curves with varying RNA concentrations were fit to a two-step kinetics equation as follows:

$$[\text{meSL1p}] = A[1 - e^{-kt}] + Vt \quad \text{(equation 1)}$$

where A is the burst amplitude, k is the burst rate, and V is the reaction speed for the linear slow phase. Non-linear regression analysis was performed with Igor Pro 8 (WaveMetrics, Inc.). For comparing protein substitutions with wild-type $MePCE_{MT}$, 10 μM of SL1p was used and relative activities within one turnover were calculated from the 2 min time points. For comparing different RNA substrates and the effect of the product inhibitors under multiple turnover conditions, all RNA concentrations are 10 μM and the SAH/meSL1p inhibitor concentrations are 5 μM. 30 min time points were used for calculating multiple turnover relative activities. To compute p-values between different groups of datasets, the single factor ANOVA1 test associated with Igor Pro 8 (WaveMetrics, Inc.) was used. The between-groups degrees of freedom is n-1, where n is the number of groups.

Isothermal titration calorimetry

The binding dissociation coefficients (K_D) for binding of $MePCE_{MT}$ to cofactors SAM and SAH and substrate RNA constructs were determined using a MicroCal 200 ITC instrument (GE Healthsciences). All samples were in buffer B and experiments were performed at 298

K. buffer B was used for the final size exclusion chromatography purification step and $MePCE_{MT}$ was used for ITC experiments within 48 hours of purification without further concentration or buffer exchange. $MePCE_{MT}$ was added to the cell in concentrations between 8–9 μM and titrant (SAM, SAH, or RNA) was 100–250 μM. SAM and SAH were dissolved directly into buffer B and the concentration was determined by UV spectroscopy using the extinction coefficient ε=15400 L/(mole·cm) at A_{260}. RNA was extensively buffer exchanged into buffer B. Calorimetric data was fit using ORIGIN 7 (MicroCal). The binding parameters stoichiometry (N), entropy (ΔS), enthalpy (ΔH), and association constant (K) were kept as floating variables during each fit. Experiments were performed 2–6 times with each experiment fit individually and binding parameters averaged. Thermodynamic parameters and number of replicates are listed in Supplementary Table 3.

Multiple sequence alignment

Sequences from the MTase domains of human (Hs), mouse (Mm), Danio rario (Dr), Drosophila melanogaster (Dm), Arabadopsis thaliana (At), Pea (Pea), and Saccharomyces pombe (Sp) MePCE; and human BCDIN3 and Mettl16 were aligned using the NCBI constraint-based multiple alignment tool (COBALT)[59] and analyzed using ESPript 3.0[60].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grant GM107567 to J.F. and American Cancer Society Postdoctoral Fellowship 126777-PF-14-179-01-DMC to C.D.E. We acknowledge NMR equipment grant NIH S10OD016336 and DOE grant DE-FC0302ER63421 for partial support of NMR and X-ray core facilities. The authors thank M. Capel, K. Rajashankar, N. Sukumar, F. Murphy, I. Kourinov and J. Schuermann of Northeastern Collaborative Access Team (NE-CAT) beamline ID-24 at the Advanced Photon Source (APS) of Argonne National Laboratory, which are supported by NIH grants P41 RR015301 and P41 GM103403. Use of the APS is supported by DOE under Contract DE-AC02-06CH11357. We thank Y. Chen for the help with mass spectrometry data collection and analysis in the UCLA MIC mass spectrometry facility, supported in part by the NIH instrumentation grant 1S10OD016387, M. Collazo for help with crystal optimization, K. Jain, J. Lowenson and S. Clarke for helpful discussions.

References

1. Byszewska M, Śmietański M, Purta E & Bujnicki JM RNA methyltransferases involved in 5’ cap biosynthesis. RNA Biol 11, 1597–1607 (2014). [PubMed: 25626080]
2. Schapira M Structural chemistry of human RNA methyltransferases. ACS Chem Biol 11, 575–582 (2016). [PubMed: 26566070]
3. Singh R & Reddy R Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. Proc Natl Acad Sci USA 86, 8280–8283 (1989). [PubMed: 2813391]
4. Shumyatsky GP, Tillib SV & Kramerov DA B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5′ end. Nucleic Acids Res 18, 6347–6351 (1990). [PubMed: 1700854]
5. Shimba S, Buckley B, Reddy R, Kiss T & Filipowicz W Cap structure of U3 small nucleolar RNA in animal and plant cells is different. gamma-Monomethyl phosphate cap structure in plant RNA. J. Biol. Chem 267, 13772–13777 (1992). [PubMed: 1618872]
6. Jeronimo C et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell 27, 262–274 (2007). [PubMed: 17643375]

7. Muniz L, Egloff S & Kiss T RNA elements directing in vivo assembly of the 7SK/MePCE/Larp7 transcriptional regulatory snRNP. Nucleic Acids Res 41, 4686–4698 (2013). [PubMed: 23471002]
8. Brogie JE & Price DH Reconstitution of a functional 7SK snRNP. Nucleic Acids Res 45, 6864–6880 (2017). [PubMed: 28431135]
9. Xue Y, Yang Z, Chen R & Zhou Q A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP. Nucleic Acids Res 38, 360–369 (2010). [PubMed: 19906723]
10. Yang Z, Zhu Q, Luo K & Zhou Q The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature 414, 317 (2001). [PubMed: 11713532]
11. Nguyen VT, Kiss T, Michels AA & Bensaude O 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature 414, 322 (2001). [PubMed: 11713533]
12. Kohoutek J P-TEFb- the final frontier. Cell Division 4, 19 (2009). [PubMed: 19723344]
13. Krueger BJ et al. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res 36, 2219–2229 (2008). [PubMed: 18281698]
14. Markert A et al. The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Reports 9, 569–575 (2008). [PubMed: 18483487]
15. He N et al. A La-Related Protein Modulates 7SK snRNP Integrity to Suppress P-TEFb-Dependent Transcriptional Elongation and Tumorigenesis. Mol Cell 29, 588–599 (2008). [PubMed: 18249148]
16. Michels AA et al. Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. The EMBO Journal 23, 2608–2619 (2004). [PubMed: 15201869]
17. Muniz L, Egloff S, Ughy B, Jády BE & Kiss T Controlling Cellular P-TEFb Activity by the HIV-1 Transcriptional Transactivator Tat. PLOS Pathogens 6, e1001152 (2010). [PubMed: 20976203]
18. Wassarman DA & Steitz JA Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. Mol. Cell. Biol 11, 3432–3445 (1991). [PubMed: 1646389]
19. Marz M et al. Evolution of 7SK RNA and Its Protein Partners in Metazoa. Mol. Biol. Evol 26, 2821–2830 (2009). [PubMed: 19734296]
20. Yazbeck AM, Tout KR & Stadler PF Detailed secondary structure models of invertebrate 7SK RNAs. RNA Biol 15, 158–164 (2018). [PubMed: 29219696]
21. Uchikawa E et al. Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. Nucleic Acids Res 43, 3373–3388 (2015). [PubMed: 25753663]
22. Eichhorn CD, Chug R & Feigon J hLARP7 C-terminal domain contains an xRRM that binds the 3' hairpin of 7SK RNA. Nucleic Acids Res 44, 9977–9989 (2016). [PubMed: 27679474]
23. Eichhorn CD, Yang Y, Repeta L & Feigon J Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7. Proc Natl Acad Sci USA (2018).
24. Singh R, Gupta S & Reddy R Capping of mammalian U6 small nuclear RNA in vitro is directed by a conserved stem-loop and AUAUAC sequence: conversion of a noncapped RNA into a capped RNA. Mol. Cell. Biol 10, 939–946 (1990). [PubMed: 2304469]
25. Cosgrove MS, Ding Y, Rennie WA, Lane MJ & Hanes SD The Bin3 RNA methyltransferase targets 7SK RNA to control transcription and translation. Wiley Interdiscip Rev RNA 3, 633–647 (2012). [PubMed: 22740346]
26. Cheng H, Sukal S, Callender R & Leyh TS γ-Phosphate Protonation and pH-dependent Unfolding of the Ras·GTP·Mg2+ Complex: A VIBRATIONAL SPECTROSCOPY STUDY. J. Biol. Chem 276, 9931–9935 (2001). [PubMed: 11124953]
27. Copeland RA in Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis 188–265 (John Wiley & Sons, New York, 2000).
28. Husain N et al. Structural basis for the methylation of A1408 in 16S rRNA by a panaminoglycoside resistance methyltransferase NpmA from a clinical isolate and analysis of the NpmA interactions with the 30S ribosomal subunit. Nucleic Acids Res 39, 1903–1918 (2011). [PubMed: 21062819]

29. Shi YQ & Rando RR Kinetic mechanism of isoprenylated protein methyltransferase. J. Biol. Chem 267, 9547–9551 (1992). [PubMed: 1577795]
30. Johnson BA & Aswad DW Kinetic properties of bovine brain proteinl-isoaspartyl methyltransferase determined using a synthetic isoaspartyl peptide substrate. Neurochemical Res 18, 87–94 (1993).
31. Swiercz R, Person, Maria D & Bedford, Mark T Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). Biochemical Journal 386, 85–91 (2005). [PubMed: 15473865]
32. Clarke S & Banfield K in Homocysteine in Health and Disease (eds Ralph Carmel & Donald W. Jacobsen) 63–78 (Cambridge University Press, Cambridge, 2001).
33. Turner DH & Mathews DH NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res 38, D280–282 (2010). [PubMed: 19880381]
34. Martinez A et al. Human BCDIN3D monomethylates cytoplasmic histidine transfer RNA. Nucleic Acids Res 45, 5423–5436 (2017). [PubMed: 28119416]
35. Xhemalce B, Robson, Samuel C. & Kouzarides T. Human RNA Methyltransferase BCDIN3D Regulates MicroRNA Processing. Cell 151, 278–288 (2012). [PubMed: 23063121]
36. Warda AS et al. Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Reports 18, 2004–2014 (2017). [PubMed: 29051200]
37. Pendleton KE et al. The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell 169, 824–835.e814 (2017). [PubMed: 28525753]
38. Choi S, Jung C-R, Kim J-Y & Im D-S PRMT3 inhibits ubiquitination of ribosomal protein S2 and together forms an active enzyme complex. Biochim Biophys Acta (BBA) - General Subjects 1780, 1062–1069 (2008). [PubMed: 18573314]
39. Didychuk AL, Butcher SE & Brow DA The life of U6 small nuclear RNA, from cradle to grave. RNA 24, 437–460 (2018). [PubMed: 29367453]
40. Hussain RH, Zawawi M & Bayfield MA Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. Nucleic Acids Res 41, 8715–8725 (2013). [PubMed: 23887937]
41. Bayfield MA, Yang R & Maraia RJ Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). Biochim Biophys Acta 1799, 365–378 (2010). [PubMed: 20138158]
42. Maraia RJ, Mattijssen S, Cruz-Gallardo I & Conte MR The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. Wiley Interdiscip Rev RNA 8 (2017).
43. McCoy AJ et al. Phaser crystallographic software. J Appl Crystallogr 40, 658–674 (2007). [PubMed: 19461840]
44. Emsley P & Cowtan K Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132 (2004). [PubMed: 15572765]
45. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221 (2010). [PubMed: 20124702]
46. Moriarty NW, Grosse-Kunstleve RW & Adams PD electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr D Biol Crystallogr 65, 1074–1080 (2009). [PubMed: 19770504]
47. Kabsch W Xds. Acta Crystallogr D Biol Crystallogr 66, 125–132 (2010). [PubMed: 20124692]
48. Painter J & Merritt EA Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62, 439–450 (2006). [PubMed: 16552146]
49. Hodel A, Kim S-H & Brünger AT Model bias in macromolecular crystal structures. Acta Crystallogr A 48, 851–858 (1992).
50. Dolinsky TJ et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35, W522–W525 (2007). [PubMed: 17488841]

51. Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA 98, 10037 (2001). [PubMed: 11517324]
52. Delaglio F et al. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J. Bio. NMR 6, 277–293 (1995).
53. Lee W, Tonelli M & Markley JL NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31, 1325–1327 (2015). [PubMed: 25505092]
54. Grzesiek S & Bax A The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. J. Am. Chem. Soc 115, 12593–12594 (1993).
55. Vinci CR & Clarke SG Recognition of Age-damaged (R,S)-Adenosyl-L-methionine by Two Methyltransferases in the Yeast Saccharomyces cerevisiae. J. Biol. Chem 282, 8604–8612 (2007). [PubMed: 17264075]
56. Shah S & Friedman SH An ESI-MS method for characterization of native and modified oligonucleotides used for RNA interference and other biological applications. Nature Protocols 3, 351 (2008). [PubMed: 18323805]
57. Jain K, Jin CY & Clarke SG Epigenetic control via allosteric regulation of mammalian protein arginine methyltransferases. Proc Natl Acad Sci USA 114, 10101–10106 (2017). [PubMed: 28874563]
58. Smietanski M et al. Structural analysis of human 2′-O-ribose methyltransferases involved in mRNA cap structure formation. Nature Comm 5, 3004 (2014).
59. Papadopoulos JS & Agarwala R COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23, 1073–1079 (2007). [PubMed: 17332019]
60. Robert X & Gouet P Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42, W320–324 (2014). [PubMed: 24753421]

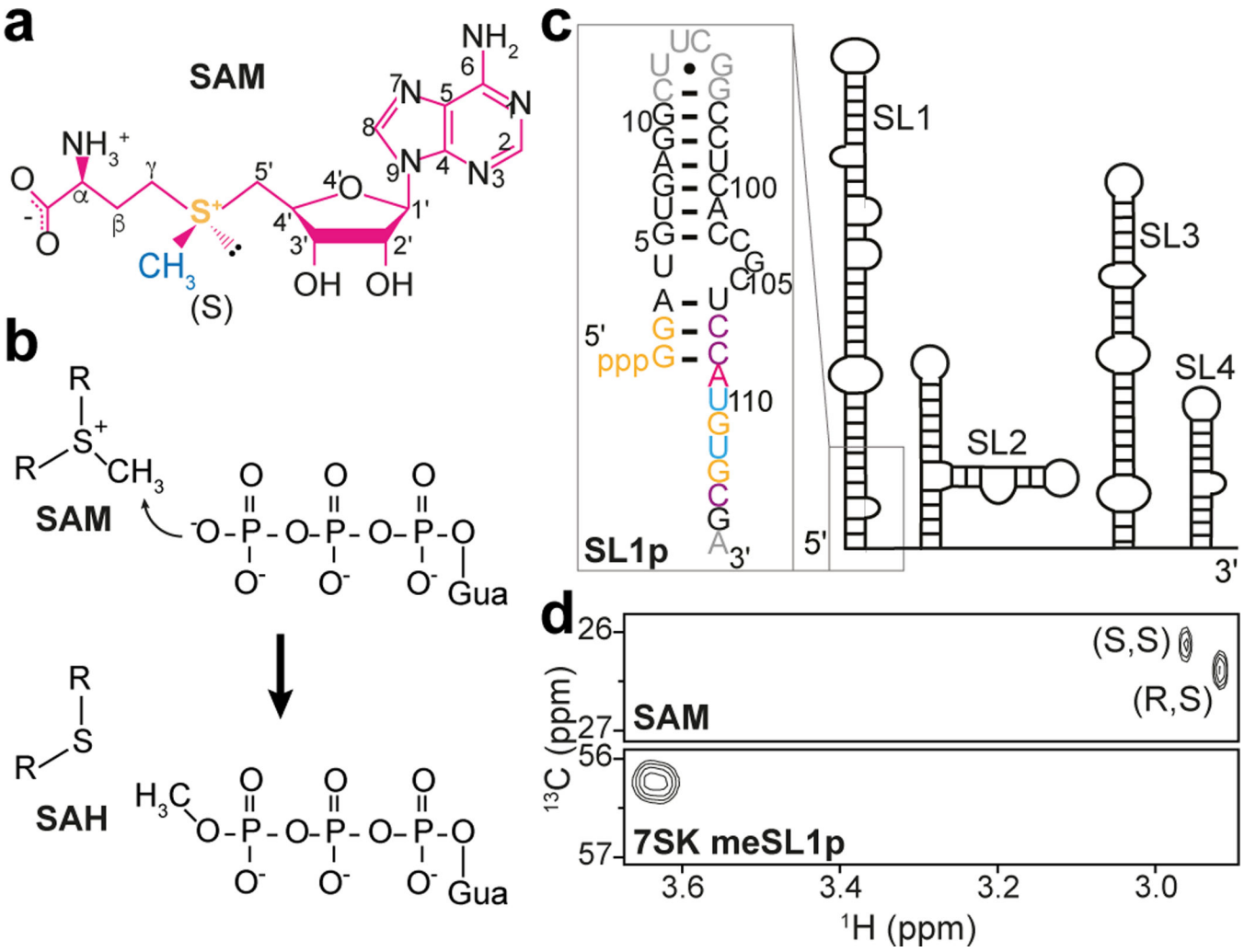

Fig. 1.

MePCE binds and caps 7SK. **a.** S-adenosylmethionine (SAM) cofactor, shown with methyl sulfonium group as (S) stereoisomer. Chemical bonds colored pink, sulfur colored yellow, and methyl group colored blue. **b.** Proposed transfer reaction mechanism of SAM methyl to RNA 5′ γ-phosphate oxygen by MePCE. **c.** Cartoon schematic of proposed secondary structure of 7SK, linear model. *inset*: sequence and secondary structure of SL1p used in this study. Residues involved in MePCE interactions are colored by nucleotide (adenine is pink, guanine is gold, uridine is cyan, cytidine is purple). Native residues not involved in protein binding are colored black and non-native residues are colored gray. **d.** ^{1}H-^{13}C HMQC spectra of free [*methyl*-^{13}C]-SAM (top) and the quantitatively capped 7SK [*methyl*-^{13}C]-SL1p. The two peaks for free [*methyl*-^{13}C]-SAM are from the (S,S) and (R,S) diastereomers. The unique resonance of meSL1p methyl has chemical shift values of 3.65 ppm (^{1}H) and 56.0 ppm (^{13}C).

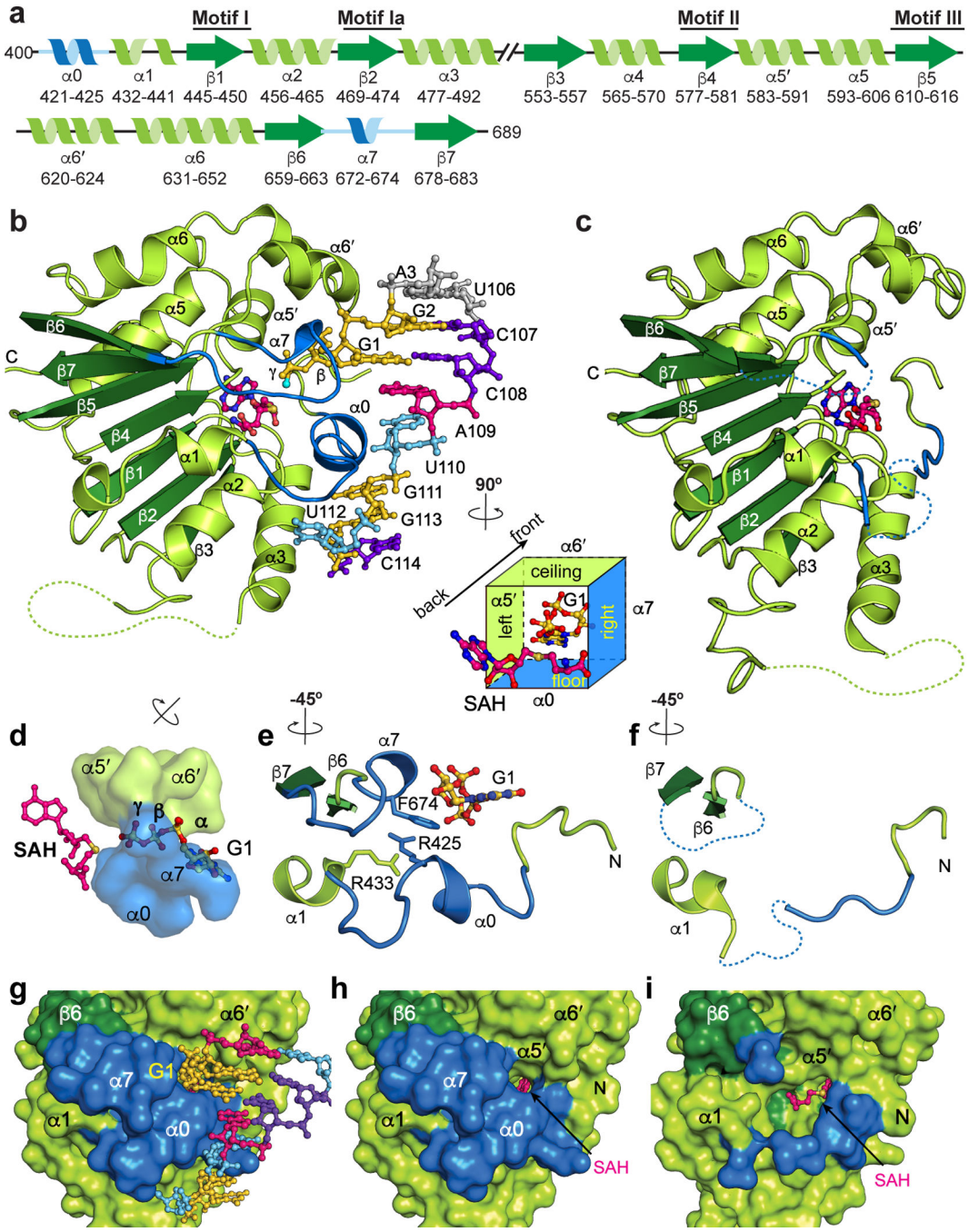

Fig. 2.

Structural changes in $MePCE_{MT}$ on binding RNA. **a.** Secondary structure boundaries of $MePCE_{MT}$. β-strands are colored dark green, α-helices and loops are colored lime, and loops that become ordered on binding RNA are colored blue. **b.** Crystal structure of MePCE–SAH–me7SK, with SAH colored pink and meSL1p colored as in Fig. 1a. The methyl carbon is colored cyan. meSL1p and SAH are shown as ball and stick and $MePCE_{MT}$ is shown as cartoon. *Inset*: Cube representation of the active site topology, defined by helix α0 at the floor, helix α5′ to the left, helix α6′ at the ceiling, the β-sheet edge at the back,

helix α7 to the right, and SL1p at the front, as seen from the back wall. **c.** Crystal structure of MePCE–SAH (PDB ID 5UNA). **d.** The triphosphate-binding tunnel formed by helix α0 and α7 (blue) and helix α5′ and α6′ (green) in MePCE–SAH–7SK structure. Protein residues are shown in transparent surface representation. RNA residue Gua1 and SAH are shown as ball and stick. **e–f.** $MePCE_{MT}$ N- and C-terminal loops in the presence (**e**) of SL1p h are structured and form α0 and α7 helices (colored blue), while in the absence (**f**) of SL1p are disordered (represented as dashed blue lines). **g–h.** Surface representations of MePCE–SAH–7SK. **(g)** SAH is buried by $MePCE_{MT}$ N- and C-terminal regions and SL1p. (**h**) Removal of SL1p RNA visually reveals the sulfur atom of SAH at the base of the triphosphate-binding tunnel. (**i**) SAH is solvent accessible in the absence of bound RNA for MePCE–SAH (PDB ID 5UNA).

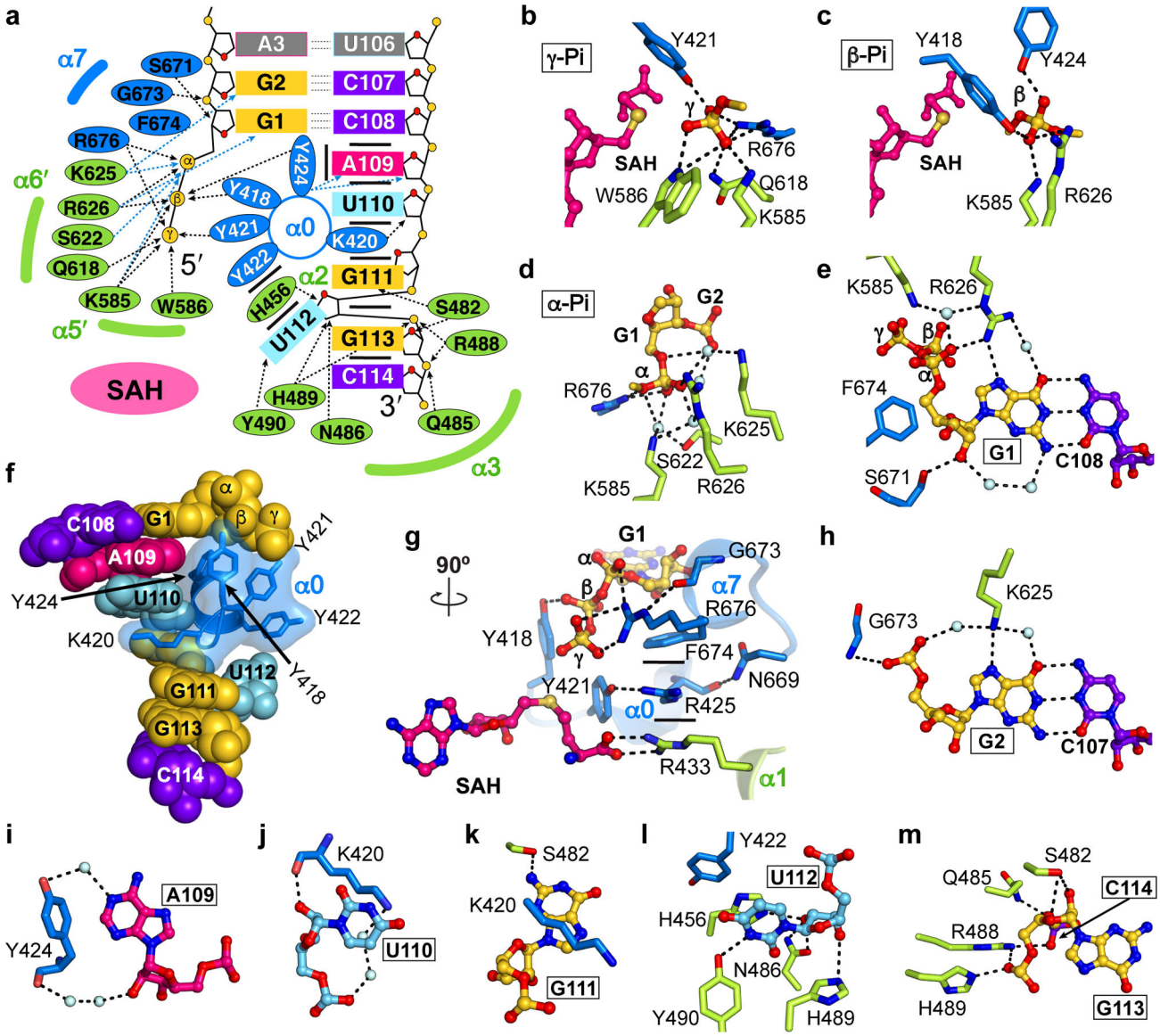

Fig. 3.

Sequence- and structure-specific recognition of 7SK by $MePCE_{MT}$ **a.** Schematic showing the specific interactions between $MePCE_{MT}$ and 7SK SL1p. α0 is indicated as blue circle with RNA-interacting residues as spokes around the helix (Tyr wheel), locations of other helices are shown with curved lines, and SAH is highlighted with a magenta oval. Dashed lines represent H-bonds and blue dashed lines represent water-mediated H-bonds. Bold lines represent stacking interactions. **b–e.** Interactions between $MePCE_{MT}$ and pppGua1 for (**b**) γ-phosphate, (**c**) β-phosphate, (**d**) α-phosphate, and (**e**) Gua1 nucleotide. **f.** Tyr wheel in helix α0 and adjacent loop recognizes the helix-single-stranded junction of SL1p. Surface, stick and cartoon are shown for the Tyr residues and K420, and RNA residues are shown in ball and stick. **g.** Active site residues that participate in specific stacking and H-bonding interactions that stabilize helices α0 and α7. **h.** Protein residues that interact with Gua2

nucleotide and backbone. **i–m.** Interactions between $MePCE_{MT}$ and the single-stranded RNA nucleotides Ade109 (**i**); Ura110 (**j**); Gua111 (**k**); Ura112 (**l**); Gua113 and Cyt114 (**m**).

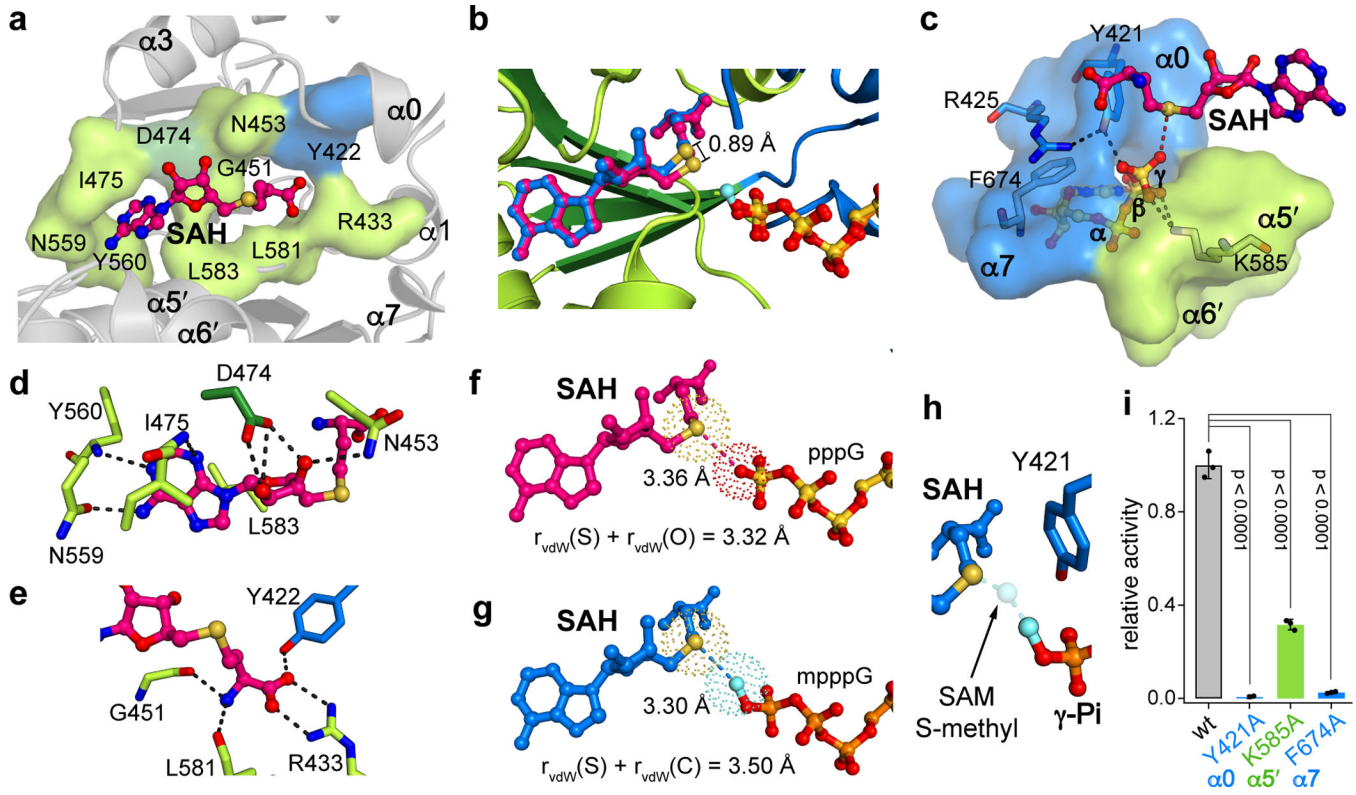

Fig. 4.

Active site of MePCE$_{MT}$ organizes cofactor and triphosphate in near transition state. **a.** Surface representation of the cofactor binding cleft in MePCE–SAH–7SK. MePCE$_{MT}$ is shown in gray cartoon. Surface coloring is identical to that in Fig. 2. **b.** Overlay of MePCE–SAH–7SK and MePCE–SAH–me7SK structures highlighting the positions of SAH and γ-phosphate in the active site. In MePCE–SAH–7SK, SAH is magenta and phosphorus atoms are gold; in MePCE–SAH–me7SK, SAH is blue and phosphorus atoms are orange with the methyl group in cyan. **c.** Triphosphate binding residues position the γ-phosphate oxygen in line with SAH sulfur in MePCE–SAH–7SK. The triphosphate-binding tunnel is shown as surface representation colored as in Fig. 2. **d.** Specific H-bonding and stacking interactions between the SAH adenosyl group and MePCE$_{MT}$. **e.** H-bonds between SAH homocysteine group and MePCE$_{MT}$. **f.** vdW radii of sulfur atom and γ-phosphate oxygen atom in MePCE–SAH–7SK showing the direct contact between the two. **g.** vdW radii of sulfur atom and methyl carbon atom in MePCE–SAH–me7SK showing the slight overlap between the two atoms. **h.** Positions of SAH sulfur, γ-mePi, and Y421 in MePCE–SAH–me7SK, with hypothetical position of SAM methyl group shown as transparent ball. Y421 stacks above path of methyl transfer. Atoms colored as in (**g**). **i.** Methyltransferase activities of wild type (wt) MePCE$_{MT}$ and point substitutions K585A (α0), F674A (α5′) and Y421A (α7) using SL1p as RNA substrate. All activities are within one turnover and are scaled relative to wt. Dots indicate values for each of the three independent reactions; bars indicate mean values; error bars are s.d. of three independent reactions performed for each protein construct. The single factor ANOVA1 test was used to compute the p-values, and the exact p-values are: 0.000041 (Y421A), 0.0000077 (K585A) and 0.0000072 (F674A).

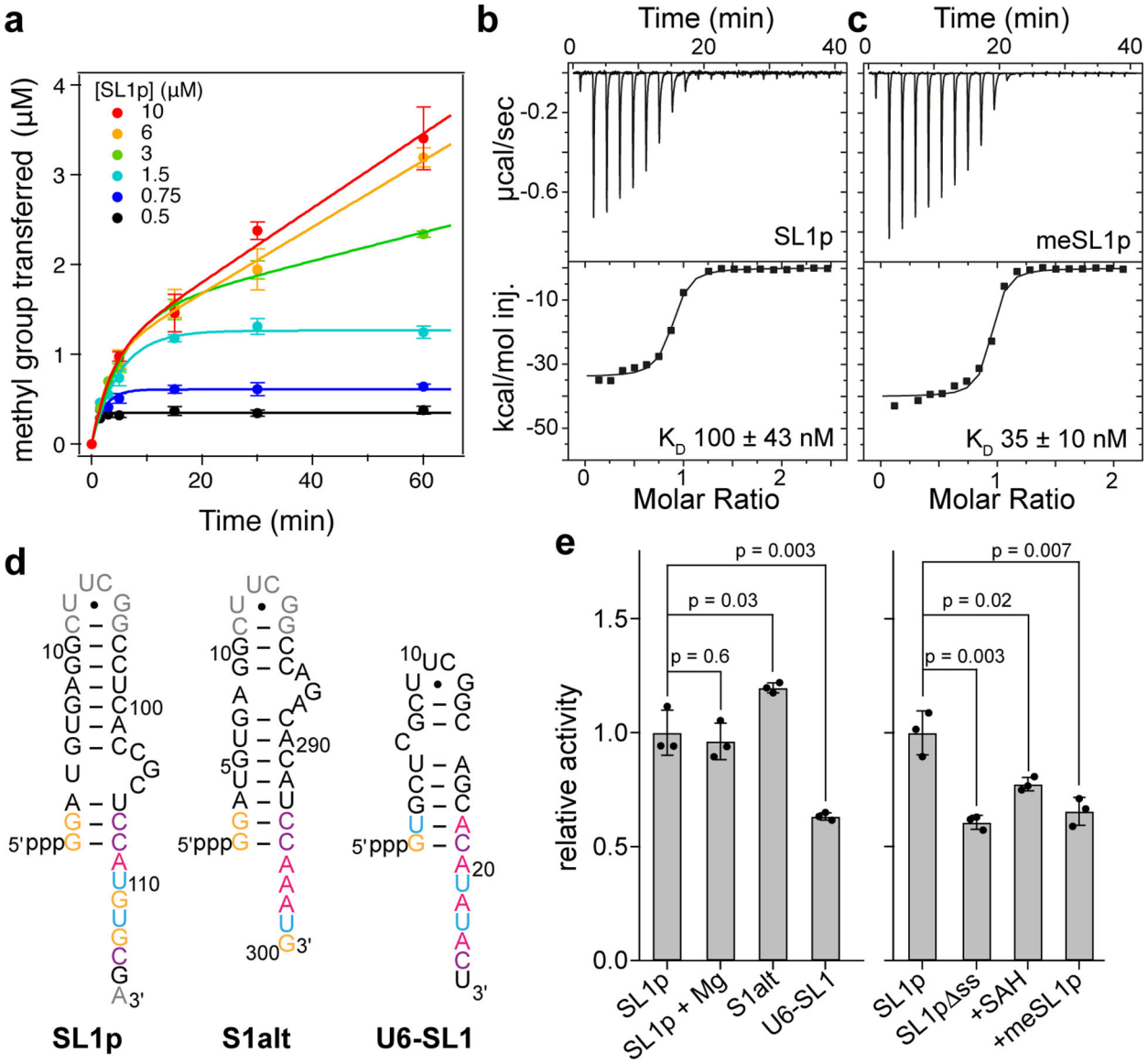

Fig. 5.

MePCE$_{MT}$ multiple turnover kinetics and binding experiments reveal product and byproduct inhibition and retention. **a.** MePCE$_{MT}$ activity assay show nonlinear progress curves for varying concentrations of SL1p. Error bars are standard deviations for three independent reactions for time points ≥ 5 min. Time points taken at 1.5 min and 3 min were performed once. Data points reflect mean values for time points ≥ 5 min. The solid lines are best fits to equation 1 (see Online Methods). **b–c.** ITC data and plots of MePCE$_{MT}$ binding to RNA substrates SL1p (**b**) and methylated SL1p (**c**). 3 equivalents of SAH were added to MePCE$_{MT}$ in the ITC experiments, and the numbers of replicates are included in Supplementary Table 3d. Sequence and secondary structures of three RNA constructs assayed for MePCE$_{MT}$ binding. **e.** Methyltransferase *apparent* activity (observed sum of single-turnover and product-inhibited activity) of MePCE$_{MT}$ with different RNA substrates or with inhibitors added under multiple turnover conditions. The two sets of plots indicate

two sets of experiments, each performed with one enzyme stock with a SL1p control. All activities are measured in three independent reactions and the averages are scaled to the SL1p in the absence of Mg^{2+}. Dots indicate values for each of the three independent reactions; bars indicate mean values; error bars are s.d. from the three reactions. The single factor ANOVA1 test was used to compute the p-values, and the exact p-values are shown on the graph.

Table 1.

Thermodynamic parameters of MePCE binding to cofactor and RNA substrates

	K_D (nM)	N
Ligand		
SAM	1580 ± 180	0.8 ± 0.3
SAH	122 ± 35	0.8 ± 0.1
7SK *		
SL1p	100 ± 43	0.9 ± 0.1
meSL1p	35 ± 10	0.9 ± 0.1
S1alt	800 ± 99	0.6 ± 0.1
meS1alt	146 ± 42	0.7 ± 0.1
SL1pΔss	n.b. **	
SL1pΔ1–3	100 ± 50	0.8 ± 0.1
U6 *		
U6-SL1	30 ± 11	0.8 ± 0.1
meU6-SL1	27 ± 4	0.8 ± 0.1

* MePCE is in the presence of 3 equivalents of SAH

** n.b. is no binding

Individual replicates were fit using Origin 7 using a one-site fitting equation. The parameters determined from n fits were averaged to determine the mean and standard deviation, where n is as follows: SAM (n=2); SAH (n=4); SL1p (n=6); meSL1p (n=3); S1alt (n=2); meS1alt (n=3); SL1pΔss (n=2); SL1pΔ1–3 (n=4); U6-SL1 (n=4); meU6-SL1 (n=2)