aKMT catalyzes extensive protein lysine methylation in the hyperthermophilic archaeon *Sulfolobus islandicus* but is dispensable for the growth of the organism

Yindi Chu[a‡], Yanping Zhu[b‡], Yuling Chen[c], Wei Li[d], Zhenfeng Zhang[a], Di Liu[d], Tongkun Wang[a], Juncai Ma[a, d], Haiteng Deng[c], Zhi-Jie Liu[b,e], Songying Ouyang[b*] and Li Huang[a*]

Supplemental Experimental Procedures

Southern hybridization

Genomic DNAs from the parental strain and ΔaKMT were extracted and digested with *Eco*RV. The cleavage products were subjected to electrophoresis in 1% agarose gel. Southern hybridization was preformed using a radiolabeled probe (Table S1), which annealed to a sequence shown in Fig. S1A.

Immunoblotting

Cells from the wild type, the deletion mutant or the complementary strain grown exponentially at 75^{o}C were harvested and resuspended in a calculated volume of 1× sample buffer for SDS–PAGE to the same cell density. Equal volumes of each sample were loaded onto two 15% SDS polyacrylamide gels in parallel. After electrophoresis, one of the gels was stained with Commassie brilliant blue R-250. The other gel was processed for immunoblotting. Proteins were transferred electrophoretically to a PVDF membrane (Merck Millipore). The membrane was incubated with rabbit anti-aKMT antibodies. An anti-rabbit IgG-HRP conjugate (Promega) was used as the secondary antibody and was detected by the chemiluminescent method.

Supplemental Results

Key residues for the methyltransferase activity of aKMT

DALI searches suggest that the *T. thermophilus* PrmA is the closest homologue of aKMT in structure among protein lysine methyltransferases whose structures have been determined. Taking

advantage of the availability of the crystal structure of PrmA in complex with ribosomal protein L11, the substrate of the methyltransferase (1,2), we stacked the methylation site K39 of L11 with aKMT (Fig. S6A). K39 was found to be in the close vicinity of the N-terminus and residues 102-109 of aKMT as well as the methyl group of SAM. We then conducted an alanine substitution scan on amino acid residues in 6-HVPYVP-11 and 102-FLLTNVNE-109 (Fig. S6B). V7A and V10A were about 40-50% as active as wild type aKMT. By comparison, alanine substitution at Y9, P11, F102, L103, L104, or N108 resulted in ~90% loss of the activity. These results should be discussed with caution since the possibility can't be ruled out that these substitutions affected the activity of aKMT by disrupting the conformation of the protein. V10, P11, and F102 are involved in hydrophobic interactions with SAM. The three residues, along with the other two hydrophobic residues L103 and L104, constitute a hydrophobic cave, which is large enough to accommodate an inserted lysine residue. The methyl group of SAM is located at the opening of the cave and oriented toward the benzene ring of F102. The methyl transfer reaction may take place spontaneously in this hydrophobic cave in a manner similar to that for PrmA (2,3). The alanine substitution of these residues would reduce the hydrophobic property of the SAM binding pocket, especially in the area surrounding the active methyl group, decreasing the activity of the protein. Notably, Y9 is oriented towards outside of the SAM binding pocket, and yet Y9A was completely inactive. Replacement of Tyr with Phe, Leu, or Trp at this position occurs in homologous proteins. Intriguingly, in the case of leucine replacement, a tyrosine residue is found adjacent to the Leu residue. In our further mutagenesis assays, as shown in Fig. S6B, Y9F and Y9W were as active as wild type aKMT, whereas Y9L was barely active, suggesting that a benzene ring at this position serves an important role in methyl transfer by aKMT. It is noticed that the main chain between Y9 and P8 is greatly altered in aKMT-SAH, as compared to that in aKMT-SAM, but little translocation of the Y9 side chain is detected. Taken together, these results suggest that Y9 serves as a hinge in the conformational switch of the N-terminal loop of aKMT and that the benzene ring is responsible for the proper positioning of Y9 relative to the SAM pocket.

REFERENCES

1. Demirci, H., Gregory, S. T., Dahlberg, A. E., and Jogl, G. (2007) Recognition of ribosomal protein

L11 by the protein trimethyltransferase PrmA. *Embo J* **26**, 567-577

2. Demirci, H., Gregory, S. T., Dahlberg, A. E., and Jogl, G. (2008) Multiple-site trimethylation of ribosomal protein L11 by the PrmA methyltransferase. *Structure* **16**, 1059-1066

3. Cameron, D. M., Gregory, S. T., Thompson, J., Suh, M. J., Limbach, P. A., and Dahlberg, A. E. (2004) Thermus thermophilus L11 methyltransferase, PrmA, is dispensable for growth and preferentially modifies free ribosomal protein L11 prior to ribosome assembly. *J Bacteriol* **186**, 5819-5825

Supplemental Tables

Table S1. Oligomucleotides used in this study.

Use	Designation[1]	Sequence[2]
Construction of the expression vector for aKMT	aKMT(f)	5'-GGAATTCCATATGTTGAGTTACGTACCGCATG TTCC
	aKMT(r)	5'ATAAGAATGCGGCCGCTTTATGTTCACCAATTA CATAG
Construction of the N-terminal deletion mutants of aKMT	ND6(f)	5'-GCCGCATATGGTTCCTTATGTACCAACGCCAG
	ND6(r)	5'-ATAAGAATGCGGCCGCTTTATGTTCACCAATT ACATAG
	ND12(f)	5'-GCCGCATATGCCAGAAAAAGTTGTAAGAAG
	ND12(r)	5'-ATAAGAATGCGGCCGCTTTATGTTCACCAATT ACATAG
Construction of the single point mutants of aKMT	H6A(f)	5'-TTGAGTTACGTACCGGCTGTTCCTTATGTAC
	H6A(r)	5'-GCCGGTACGTAACTCAACATATGTATATC
	V7A(f)	5'-AGTTACGTACCGCATGCTCCTTATGTACC
	V7A(r)	5'-GCATGCGGTACGTAACTCAACATATGTATATC
	P8A(f)	5'-TACGTACCGCATGTTGCTTATGTACCAACG
	P8A(r)	5'-CAACATGCGGTACGTAACTCAACATATG
	Y9A(f)	5'-GTACCGCATGTTCCTGCTGTACCAACGCCAG
	Y9A(r)	5'-GCAGGAACATGCGGTACGTAACTCAACATAT G
	V10A(f)	5'-CCGCATGTTCCTTATGCACCAACGCCAG
	V10A(r)	5'-GCATAAGGAACATGCGGTACGTAACTCAAc
	P11A(f)	5'-CATGTTCCTTATGTAGCAACGCCAGAAAAAG
	P11A(r)	5'-CTACATAAGGAACATGCGGTACGTAAC

T12(f)	5’-GTTCCTTATGTACCAGCGCCAGAAAAAG
T12(r)	5’-CTGGTACATAAGGAACATGCGGTACGTAAC
Y9F(f)	5’-TACCGCATGTTCCTTTTGTACCAACGCC
Y9F(r)	5’-AAAGGAACATGCGGTACGTAACTCAA
Y9W(f)	5’-TACCGCATGTTCCTTGGGTACCAACGCCAG
Y9W(r)	5’-CCAAGGAACATGCGGTACGTAACTCAA
Y9L(f)	5’-TACCGCATGTTCCTTTGGTACCAACGCCAG
Y9L(r)	5’-CAAAGGAACATGCGGTACGTAACTCAA
E59A(f)	5’-AGGCTGTAGGTGTAGCAATTAACGATGAG
E59A(r)	5’-GCTACACCTACAGCCTTCTTTACGTT
N87A(f)	5’-TCAATAGTAAAAGGTGCTTTCTTCGAGGTC
N87A(r)	5’-GCACCTTTTACTATTGAAGCTCTTCCC
F88A(f)	5’-ATAGTAAAAGGTAATGCCTTCGAGGTCGA
F88A(r)	5’-GCATTACCTTTTACTATTGAAGCTCTT
F102A(f)	5’-ACTGTAGTTACAATGGCCCTTTTAACAAA
F102A(r)	5’-GCCATTGTAACTACAGTAGCCTCCGAA
L103A(f)	5’-GTAGTTACAATGTTCGCTTTAACAAATGT
L103A(r)	5’-GC GAACATTGTAACTACAGTAGCCTCC
L104A(f)	5’-GTTACAATGTTCCTTGCAACAAATGTTAA
L104A(r)	5’-GCAAGGAACATTGTAACTACAGTAGCC
T105A(f)	5’-ACAATGTTCCTTTTAGCAAATGTTAATG
T105A(r)	5’-CTAAAAGGAACATTGTAACTACAGTA
N106A(f)	5’-ATGTTCCTTTTAACAGCTGTTAATGAGATG
N106A(r)	5’-GCTGTTAAAAGGAACATTGTAACTACA
V107A(f)	5’-TCCTTTTAACAAATGCTAATGAGATGCT
V107A(r)	5’-GCATTTGTTAAAAGGAACATTGTAAC
N108A(f)	5’-CTTTTAACAAATGTTGCTGAGATGCTAAAACC

	N108A(r)	5’-GCAACATTTGTTAAAAGGAACATTGTA
	E109A(f)	5’-TAACAAATGTTAATGCGATGCTAAAACCAA
	E109A(r)	5’-GCATTAACATTTGTTAAAAGGAACAT
Probes for Southern hybridization	Probe_wt	5’-GCTTCAATAGTAAAAGGTAATTTC
	Probe_mt	5’-GAGAAGTTCTCGAGAGTGTATC
Construction of the aKMT deletion mutant	Tg-arm(f)	5’-GTCGACTTGAGTTACGTACCGCATGTTCCTT
	Tg-arm(r)	5’-ACGCGTTCATTTATGTTCACCAATTACATAG
	L-arm(f)	5’-CCATGGGTTATCGCAACAAGATACGAGGTAA
	L-arm(r)	5’-CTCGAGAACTTCTCACTTCAAAATAATCCTA
	R-arm(f)	5’-CTCGAGATGAAAATTTTGATAGTAAAAAGTG
	R-arm(r)	5’-CTCGAGAGTGTATCTCTATGTAATTGGTGAA
Detection of the *kmtA* allele	Ext(f)	5’-GATTCAGCCTTAGCTCTAATATATG
	Ext(r)	5’-CCTTTTGGAGTTAACGTCCTACCC
	Int(f)	5’-TTGAGTTACGTACCGCATGTTCCTT
	Int(r)	5’-GGTTTAAGCTCTTTCTCGAGCTTTG
Construction of a vector for the expression of aKMT in *S. islandicus*	aKMT pSeSD(f)	5’-CATATGTTGAGTTACGTACCGCATGTTC
	aKMT pSeSD(r)	5’-GTCGACTTTATGTTCACCAATTACATAG
Primers for quantitative RT-PCR	SiRe_0696(f)	TGAAGTTTGTTAGGCACG
	SiRe_0696(r)	TTACATTGTCTGGCACTA
	SiRe_0587(f)	GGAATGGCTTATCAAACA
	SiRe_0587(r)	CAATACCGCAAATACTTACT
	SiRe_0431(f)	TAACCGAACCAACAGACG

SiRe_0431(r)	TAATGCCATTTCCACAGC
SiRe_0591(f)	AACAAACCAAGCCCGTAT
SiRe_0591(r)	AAGCGTCAACTACTTTCTCACT
SiRe_2427(f)	AAAGCTCCAGAATCAGCG
SiRe_2427(r)	GCCCAGCAAATAATACATC
SiRe_0125(f)	TAACAGTAAATTCCCTAATC
SiRe_0125(r)	TATCATACCGTCCCACCA
SiRe_0453(f)	GCTGAGGACGCTAGATTA
SiRe_0453(r)	GCTGATAGGTCGGCAAGT
SiRe_2463(f)	CTGGAATCTTACCGTTGC
SiRe_2463(r)	GCTTCCTCTTTCCCACTA
SiRe_0706(f)	TGAAGACAGTTCCCAGAG
SiRe_0706(r)	TTGCCATGTACTTATCCAAT
SiRe_0707(f)	CCGAAGGAATCCCAATAC
SiRe_0707(r)	CGTGGGCTAACCTGACTG
sis 16S(f)	TCAACGCCTGGAATCTTACCA
sis 16S(r)	TCGCTCGTTGCCTGACTTAA

[1] Forward and reverse primes are abbreviated as f and r, respectively, in parentheses.

[2] Restriction sites are underlined.

Table S2. Identification of proteins in the parental strain (S2a) and ΔaKMT（S2b）by mass spectrometry (Extra files). Valid identifications are highlighted in cyan in the 'identified proteins' worksheet. The 'Best Site Probabilities' values (shown in brackets) of the peptides with a methylated lysine residue at the C-terminus were manually added under 'modifications' in the 'methylated peptides' worksheet. The assignment of the modification is considered reliable when the value is greater than 75. The exponentially modified protein abundance index (emPAI) was determined to allow the estimation of protein abundance, and the emPAI values of proteins identified in both strains and methylated in ΔaKMT were manually added to S2b. ΔCn shows the normalized score difference between the currently selected PSM and the highest-scoring PSM for that spectrum. The #PSMs are the total number of identified peptide sequences (PSMs) for the protein, including those redundantly identified. The Protein Score shows how an identified peptide matches a protein. The q-value is used to quantify the FDR, and PEP is posterior error probability. These two parameters are calculated from the distribution of search results against a normal database and a false decoy database.

Table S3. List of proteins methylated only in ΔaKMT.

Proteins	Unmethylated form identified in parent	Proteins	Unmethylated form identified in parent
SiRe_0027	+	SiRe_1336	-
SiRe_0057	+	SiRe_1346	-
SiRe_0078	-	SiRe_1369	+
SiRe_0120	+	SiRe_1383	+
SiRe_0124	+	SiRe_1402	+
SiRe_0128	+	SiRe_1420	+
SiRe_0185	+	SiRe_1421	+
SiRe_0198	+	SiRe_1461	+
SiRe_0214	+	SiRe_1527	-
SiRe_0218	+	SiRe_1578	+
SiRe_0245	+	SiRe_1587	-
SiRe_0265	+	SiRe_1601	+
SiRe_0331	-	SiRe_1611	+
SiRe_0381	-	SiRe_1639	-
SiRe_0414	+	SiRe_1679	+
SiRe_0417	-	SiRe_1703	+
SiRe_0434	+	SiRe_1712	+
SiRe_0538	-	SiRe_1746	-
SiRe_0566	+	SiRe_1763	+
SiRe_0590	+	SiRe_1773	+
SiRe_0601	+	SiRe_1785	+
SiRe_0618	+	SiRe_1787	-
SiRe_0654	+	SiRe_1802	+
SiRe_0668	-	SiRe_1902	+
SiRe_0680	-	SiRe_1917	+
SiRe_0701	+	SiRe_1927	+
SiRe_0742	+	SiRe_2030	-
SiRe_0904	+	SiRe_2148	+
SiRe_0930	-	SiRe_2155	+
SiRe_0944	-	SiRe_2244	-
SiRe_0959	+	SiRe_2261	+
SiRe_0965	+	SiRe_2276	+
SiRe_0971	+	SiRe_2281	+
SiRe_0991	+	SiRe_2309	-
SiRe_0992	+	SiRe_2341	-
SiRe_1014	+	SiRe_2373	+
SiRe_1061	-	SiRe_2377	+
SiRe_1093	+	SiRe_2390	-
SiRe_1104	+	SiRe_2401	+
SiRe_1124	-	SiRe_2425	+

SiRe_1141	-	SiRe_2454	+
SiRe_1145	+	SiRe_2512	+
SiRe_1177	+	SiRe_2553	-
SiRe_1182	+	SiRe_2556	+
SiRe_1184	+	SiRe_2612	+
SiRe_1255	-	SiRe_2630	+
SiRe_1331	-	SiRe_2632	-

Table S4. ArCOG category analysis of proteins methylated by aKMT. (Extra file)

Table S5 Data collection and refinement statistics[1].

Parameter	aKMT(Se-Met)-SAM	aKMT(native)-SAH
PDB ID		
X-ray source	SSRF 17U1	SSRF 17U1
Crystal to detector distance (mm)	170	300
Number of images	720	180
Oscillation width (º)	0.5	1
Wavelength (Å)	0.9790	0.9792
Space group	C2	P2
Cell dimensions		
a, b, c (Å)	48.30, 57.99, 55.73	38.23, 51.90, 40.71
α, β, γ (°)	90.00, 100.93, 90.00	90.00, 117.84, 90.00
Mosaicity (º)	0.45	1.94
No. of protein molecules/ASU	1	1
No. of ligand/ASU	1	1
No. of phasing sites (selenium) /ASU	4	-
Resolution range (Å)	50.00-1.30 (1.33-1.30)	50.00-1.84 (2.28-1.84)
Rsym (%) cif	10.4(7.4)	40.4(32.8)
Mean I/σ(I) cif	51.5(2.1)	33.35(2.10)
Completeness (%) cif	95.5(67.0)	93.7(68.3)
Redundancy	7.0(4.4)	3.3(2.1)
Refinement		
Resolution (Å)	28.60-1.40	35.90-1.84
No. of reflections	69363	11289
Rwork/Rfree (%)	18.31/21.24	18.35/21.69
No. of atoms	1476	1416
No. of protein atoms	1233	1268
No. of ligand atoms	27	26
No. of waters	215	122
B-factors	25.03	34.84
Protein	22.69	34.69
Ligand	16.57	27.95
Water	39.60	37.96
R.M.S. deviations		
Bond lengths (Å)	0.006	0.007
Bond angles (º)	1.222	1.192
Ramachandran analysis		
Favored region (%)	98.70	98.10
Allowed region (%)	1.30	1.90
Outliers (%)	0.00	0.00
Parameter	aKMT(Se-Met)-SAM	aKMT(native)-SAH
PDB ID		

X-ray source	SSRF 17U1	SSRF 17U1
Crystal to detector distance (mm)	170	300
Number of images	720	180
Oscillation width (º)	0.5	1
Wavelength (Å)	0.9790	0.9792
Space group	C2	P2
Cell dimensions		
a, b, c (Å)	48.30, 57.99, 55.73	38.23, 51.90, 40.71
α, β, γ (°)	90.00, 100.93, 90.00	90.00, 117.84, 90.00
Mosaicity (º)	0.45	1.94
No. of protein molecules/ASU	1	1
No. of ligand/ASU	1	1
No. of phasing sites (selenium) /ASU	4	-
Resolution range (Å)	50.00-1.30 (1.33-1.30)	50.00-1.84 (2.28-1.84)
Rsym (%) cif	10.4(7.4)	40.4(32.8)
Mean I/σ(I) cif	51.5(2.1)	33.35(2.10)
Completeness (%) cif	95.5(67.0)	93.7(68.3)
Redundancy	7.0(4.4)	3.3(2.1)
Refinement		
Resolution (Å)	28.60-1.40	35.90-1.84
No. of reflections	69363	11289
Rwork/Rfree (%)	18.31/21.24	18.35/21.69
No. of atoms	1476	1416
No. of protein atoms	1233	1268
No. of ligand atoms	27	26
No. of waters	215	122
B-factors	25.03	34.84
Protein	22.69	34.69
Ligand	16.57	27.95
Water	39.60	37.96
R.M.S. deviations		
Bond lengths (Å)	0.006	0.007
Bond angles (º)	1.222	1.192
Ramachandran analysis		
Favored region (%)	98.70	98.10
Allowed region (%)	1.30	1.90
Outliers (%)	0.00	0.00

[1]The numbers in parentheses represent values for the highest resolution shell.

Table S6. Summary of transcriptomic analysis.

	Parent	**△aKMT**
Raw Reads	10,040,707	10,500,000
Clean reads	9,976,077	10,202,562
Average length (bp)	2 x 100	2 x 100
Q20 base	94.98%	94.14%
Q30 base	84.05%	85.24%
GC%	60%	60%

Table S7. Differentially expressed genes in the parental strain and ΔaKMT. (Extra file)

Supplemental Figures

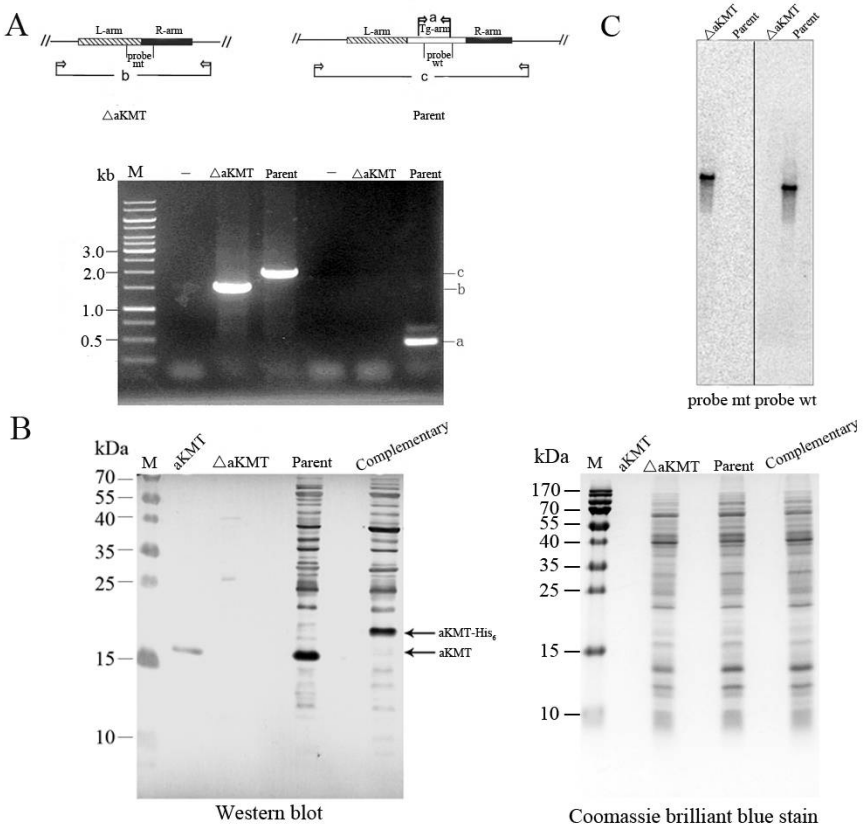

Figure S1. Verification of the aKMT mutant construct. (A) PCR. PCR reactions were carried out on genomic DNAs extracted from the parental strain and ΔaKMT. PCR products were subjected to electrophoresis in 1% agarose gel. -, no template DNA was added. The positions of the PCR primers and the probes used in Southern hybridization (Fig. S1C) are show in the diagram on top. PCR with primer pair Ext(f)/Ext(r) would produce fragment b (ΔaKMT) or c (Parent), whereas PCR with primer pair Int(f)/Int(r) would yield product a (Parent) or nothing (ΔaKMT). (B) Analysis of the synthesis of aKMT in the parental strain, ΔaKMT and the complementary strain. Samples from the three strains with the same number of cells were subjected to SDS-PAGE. The proteins were transferred electrophoretically to a PVDF membrane for immunoblotting with anti-aKMT antibodies (*left*). A separate gel identical to that for immunoblotting was stained with Coomassie brilliant blue R-250 (*right*). (C) Southern hybridization. Genomic DNAs from the parental strain and ΔaKMT were digested with *Eco*RV, and the cleavage products were subjected to Southern hybridization using a radiolabeled probe.

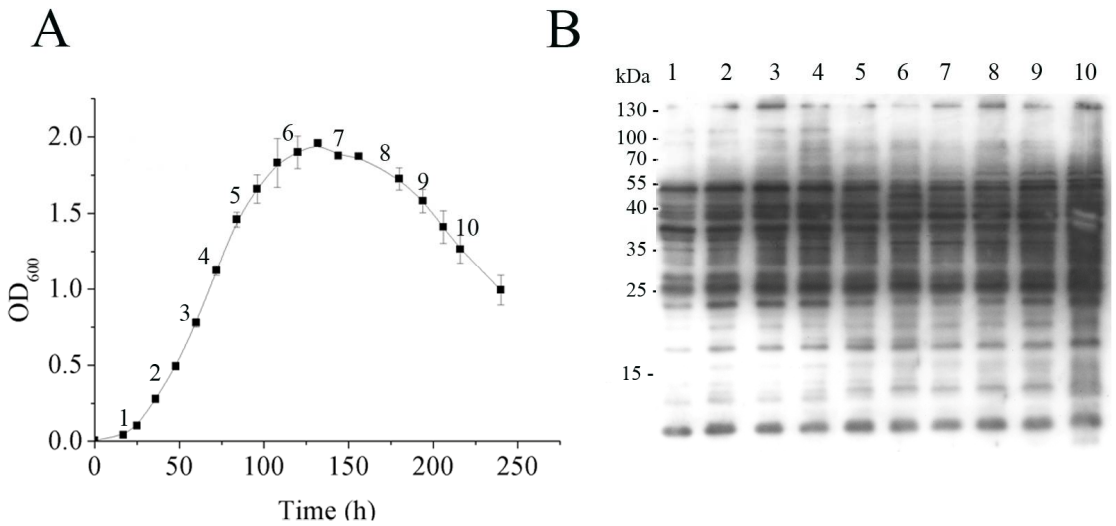

Figure S2. Protein lysine methylation in cells taken at various growth stages. The parental strain was grown at 75°C, and samples were taken at indicated time points (A). Cells were subjected to 15% SDS-PAGE, and proteins were transferred electrophoretically to a PVDF membrane. Immunoblotting (B) was performed with anti-mono/dimethyllysine antibodies.

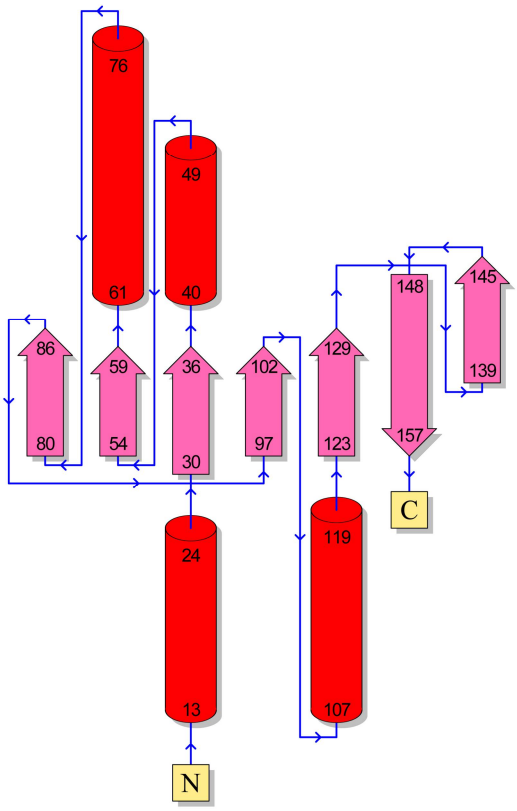

Figure S3. A topological diagram of the aKMT-SAM structure. α helices and β sheets are shown in red and pink, respectively. The secondary structure elements are organized as indicated by arrows.

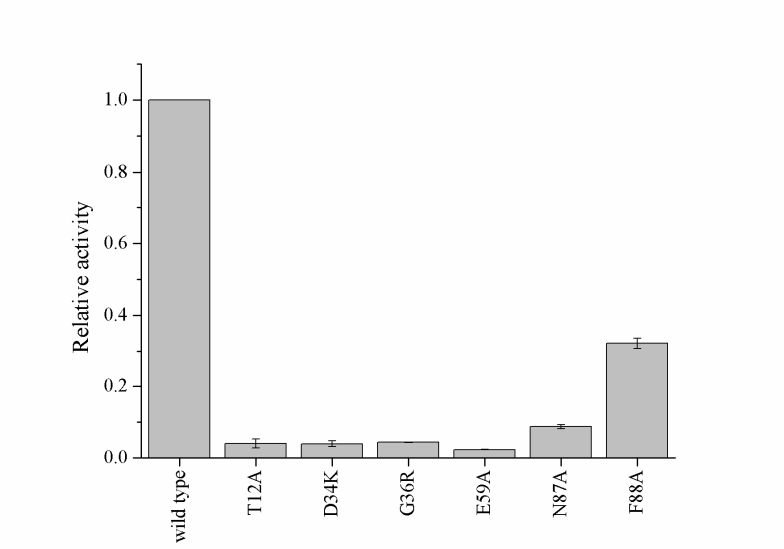

Figure S4. Methyltransferase activity assays of mutant aKMT proteins. Wild-type or mutant aKMT was incubated with recombinant Cren7 and [^{3}H]SAM. The samples were treated for liquid scintillation counting. All numbers are an average of three independent measurements.

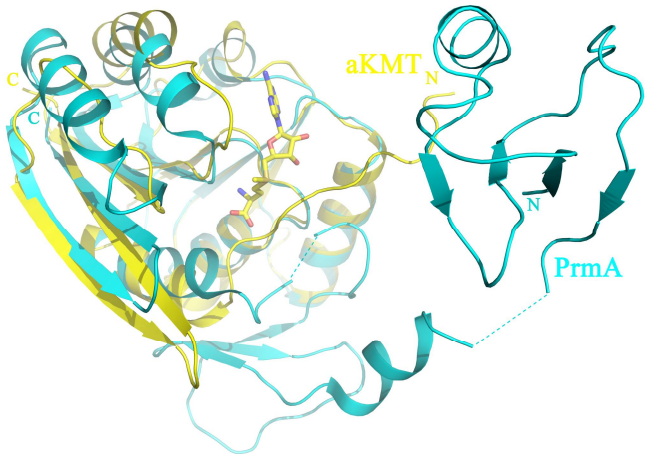

Figure S5. Structure comparison between aKMT (yellow) and *Thermus thermophilus* PrmA (cyan).

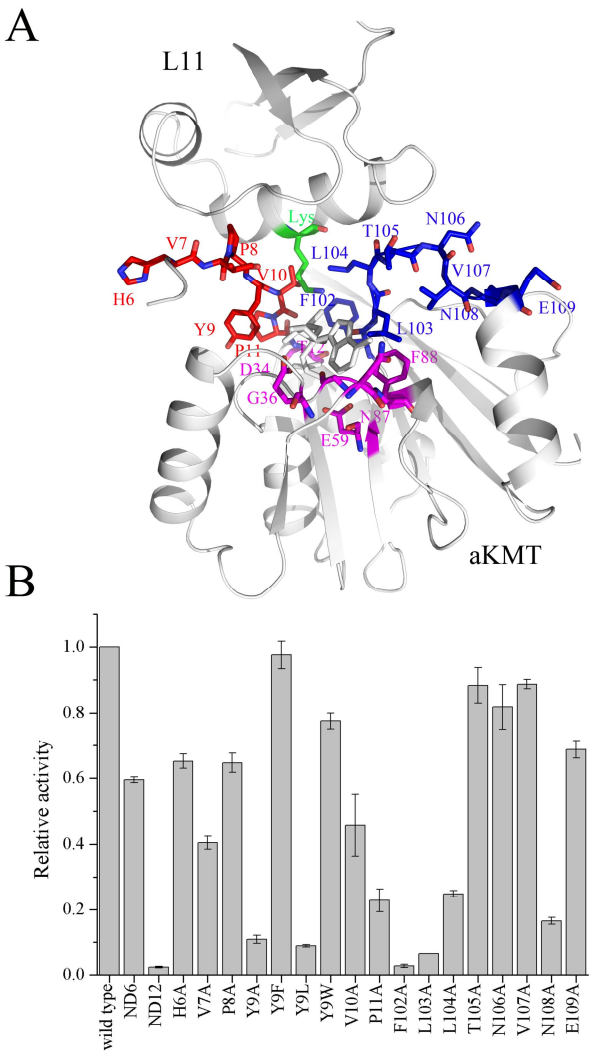

Figure S6. Identification of amino acid residues important for the activity of aKMT. (A) A model for the insertion of Lys39 of ribosomal protein L11 (PDB code: 2NXN) from *Thermus thermophilus* into the SAM-binding pocket of aKMT. SAM is shown as silk sticks and Lys39 of L11 is in green. Amino acid residues speculated to be involved in the methyl transfer reaction are shown as colored sticks. N-terminal H6-P11 and F102-E109, and residues forming hydrogen bonds with SAM are shown in red, blue and purple, respectively. (B) Methyltransferase activity of mutant aKMT proteins. Wild-type or mutant aKMT was incubated with recombinant Cren7 and [^{3}H]SAM. The samples were processed for liquid scintillation counting. All numbers are an average of three independent measurements.

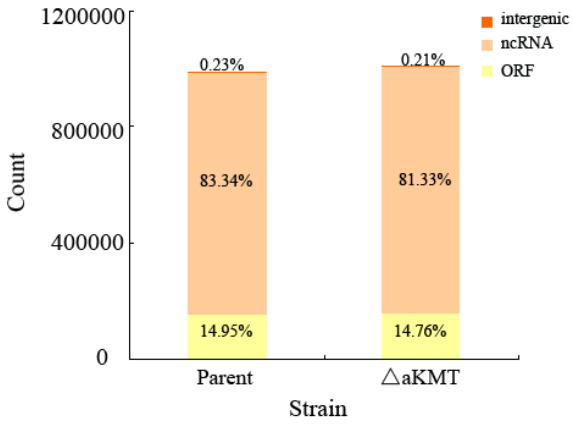

Figure S7. Numbers of RNA-seq reads and proportions of reads mapped to ORF, ncRNA and intergenic regions from the parental strain and ΔaKMT.

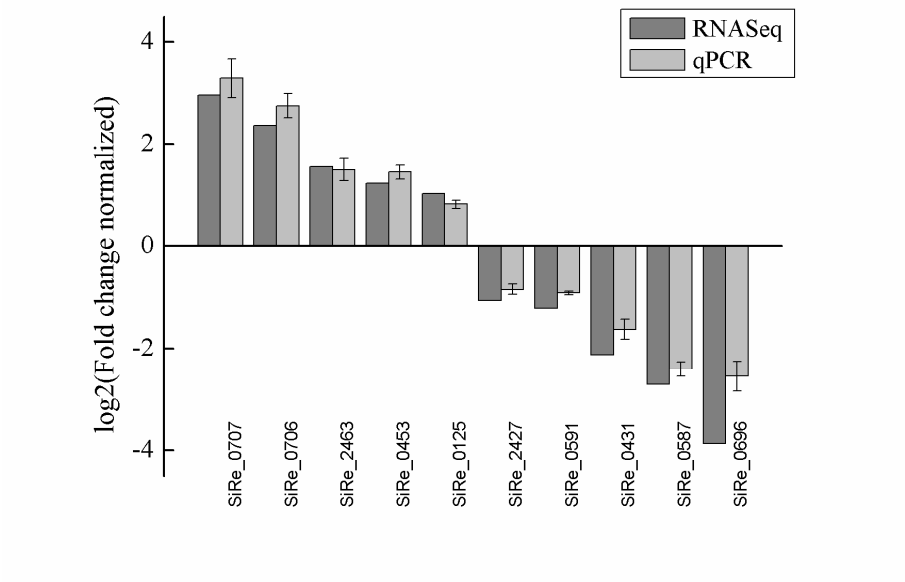

Figure S8. Validation of the RNA-seq results using qRT-PCR. Total RNA isolated from the parental strain and ΔaKMT cultures were used for cDNA synthesis. 10 genes were selected for qRT-PCR assays. Three technical replicates were set for each qPCR reaction.

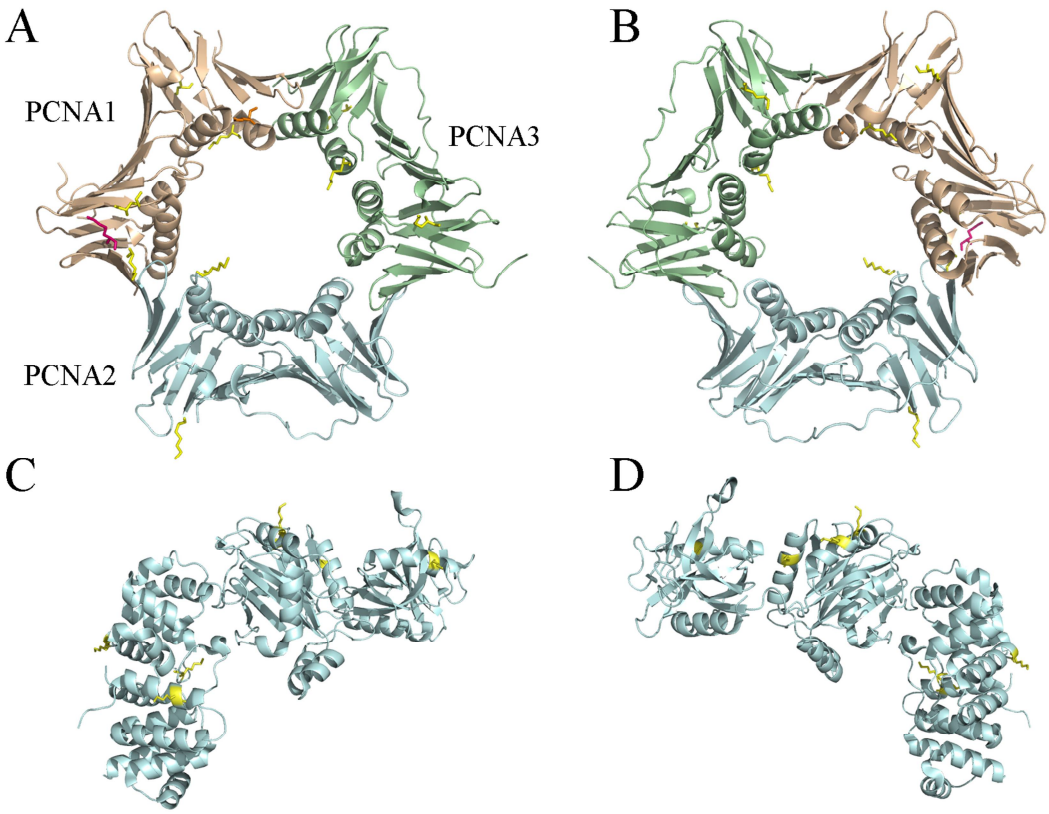

Figure S9. Location of methylated lysine residues in the structures of PCNA (A and B) and Lig1 (C and D). The structures of the heterotrimic PCNA and Lig 1 from *S. islandicus* were generated by SWISS-MODEL using *S. solfataricus* PCNA (PDB: 2IJX) and Lig1 (PDB: 2HIV), respectively, as the model, and shown in cartoon representation. The back view of the structures in (A) and (C) is shown in (B) and (D), respectively. Lysine residues that were monomethylated, dimethylated or modified in both forms are shown by a yellow, red or orange stick, respectively.