Supporting Information

for

Direct Single-Nucleotide Resolution Sequencing of DNA 5-Methylcytosine Using Engineered DNA Methyltransferase-Mediated CMD-seq

Wei Liu,[1,2,#] Zhao-Cheng Ma[1,#] Shan Zhang,[1] Fang-Yin Gang,[1] Tong-Tong Ji,[1] Yao-Hua Gu,[1,3] Neng-Bin Xie,[1] Shu-Yi Gu,[1] Xia Guo,[1] Tian Feng,[1] Yu Liu,[1,4,*] Jun Xiong,[1,*] Bi-Feng Yuan[1,2,5,6,*]

[1] Department of Occupational and Environmental Health, School of Public Health, Wuhan University, Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, China.

[2] Research Center of Public Health, Renmin Hospital of Wuhan University, Wuhan University, Wuhan 430060, China.

[3] School of Nursing, Wuhan University, Wuhan 430071, China.

[4] Hubei Key Laboratory of Tumor Biological Behaviors, Cancer Clinical Study Center, Zhongnan Hospital of Wuhan University, Wuhan 430071, China.

[5] State Key Laboratory of Metabolism and Regulation in Complex Organisms, College of Life Sciences, Wuhan University, Wuhan 430072, China.

[6] Hubei Key Laboratory of Biomass Resource Chemistry and Environmental Biotechnology, Wuhan University, Wuhan 430071, China.

[#] These authors contributed equally to this work.

* Corresponding authors:

Bi-Feng Yuan. E-mail: bfyuan@whu.edu.cn

Jun Xiong. Email: jxiong@whu.edu.cn

Yu Liu. Email: liuyu97@whu.edu.cn

This file contains:

Page S3-S5	Synthesis of carboxy-S-adenosyl-*L*-methionine (caSAM)
	Expression and purification of A3A protein
	Enzymatic digestion of DNA
	LC-MS/MS analysis
	Preparation of double-strand DNA with C and 5mC
	Evaluation of the deamination activity of A3A by sequencing
Page S6	Table S1. Amino acid sequences of the proteins used in this study.
Page S7	Table S2. Sequences of oligonucleotides.
Page S8	Table S3. The mass spectrometry parameters for the analysis of nucleosides.
Page S9	Table S4. The conditions for steady-state kinetics study.
Page S10	Table S5. Sequences of PCR primers.
Page S11	Figure S1. Comparison of the amino acid sequences of the four DNA methyltransferases.
Page S12	Figure S2. Expression of M.MedI-WT protein.
Page S13	Figure S3. The schematic illustration of plasmid for the expression of M.MedI-N377K protein.
Page S14	Figure S4. Expression of M.MpeI-N374K protein.
Page S15	Figure S5. Expression of M.MpeI-E45D/N374K protein.
Page S16	Figure S6. Rate versus caSAM concentration curves in the steady-state kinetics study.
Page S17	Figure S7. Expression of A3A protein.
Page S18	Figure S8. Sequencing results of 216-bp dsDNA after A3A treatment.
Page S19	References

Synthesis of carboxy-S-adenosyl-*L*-methionine (caSAM)

caSAM was synthesized, purified, and characterized using a previously established protocol.[1] Briefly, 10 mg of *S*-adenosyl-*L*-homocysteine (SAH, obtained from Aladdin Industrial Inc.) was dissolved in 1.7 mL of 150 mM aqueous ammonium bicarbonate solution. The resulting mixture was incubated at 37°C for 24 h with shaking at 1500 rpm. Subsequently, 12 mL of chilled methanol was added, and the mixture was incubated at -80°C for 12 h. The precipitate was isolated via centrifugation at 2000 rpm for 30 min at 4°C, followed by two washes with chilled methanol. caSAM was then dissolved in deionized water (ddH_2O) and further purified using high-performance liquid chromatography (HPLC). The purified caSAM was collected, and its concentration was determined using a B-500 spectrophotometer (Metash Instruments Co., Ltd., Shanghai, China) with an extinction coefficient of $\varepsilon_{260} = 15.4\ cm^{-1}\ mM^{-1}$. The aliquots were stored at -20°C.

Expression and purification of A3A protein

The A3A protein was expressed and purified according to previously established protocols.[2, 3] The A3A coding sequence was cloned into the pET-41a plasmid, which encodes a glutathione S-transferase (GST) tag and a recognition sequence for the human rhinovirus 3C protease (HRV 3C) (Figure S7). The plasmid was transformed into *E. coli* BL21 (DE3) pLysS competent cells (Sangon), and protein expression was induced by the addition of 0.5 mM isopropyl-*β*-D-thiogalactoside (IPTG, Sangon) when the cell density (OD600) reached 0.4-0.6. The cells were then cultured at 16°C for 16 h and harvested by centrifugation at 10,000 g for 30 min. Cell lysis was performed by sonication in PBS buffer supplemented with 2 mM DTT and 1 mM phenylmethylsulfonyl fluoride (PMSF). The supernatant was separated by centrifugation and filtered through a 0.22-μm membrane. Following incubation with Glutathione Sepharose 4B beads (Sangon) and cleavage with HRV 3C protease (Sangon), the protein was eluted and concentrated using a size-exclusion column (Millipore, Germany) equilibrated with 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.01 mM EDTA, 0.5 mM DTT, and 0.01% Tween-20. Prior to storage at -80°C, an equal volume of 50% (v/v) glycerol was added. The purified A3A protein was verified by SDS-PAGE (Figure S7B).

Enzymatic digestion of DNA

Enzymatic digestion of DNA was performed under neutral conditions, as previously described.[4] In brief, DNA samples that had been treated with DNA methyltransferases were denatured at 95°C for 10 min, followed by rapid cooling on ice water. The enzymatic digestion reaction was carried out in a 30-μL solution containing 5 units of DNase I, 15 units of alkaline phosphatase, 180 units of S1 nuclease, 0.25 units of phosphodiesterase I, and 3 μL of digestion buffer (500 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM $ZnSO_4$, pH 7.0) at 37°C for 4 h. The digested samples were then extracted with chloroform on three separate occasions. The aqueous layer was collected, lyophilized to dryness, reconstituted in water, and subsequently analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

LC-MS/MS analysis

The nucleoside analysis was performed using a liquid chromatography-tandem mass spectrometry (LC-MS/MS) system, consisting of a Shimadzu 8045 mass spectrometer (Kyoto, Japan) equipped with an electrospray ionization source (Turbo Ionspray) and a Shimadzu LC-30AD UPLC system. The liquid chromatography (LC) separation was carried out on a Shimadzu Shim-pack GIST C18 column (2.1 mm i.d. × 100 mm, 2.0 μm) at a flow rate of 0.3 mL/min and a temperature of 40°C. The mobile phases used were water with 0.05% formic acid (solvent A) and methanol (solvent B). A gradient elution program was employed, consisting of 0-1.5 min at 5% B, 1.5-3 min at 5-40% B, 3-5 min at 40% B, and 5-12 min at 5% B. The mass spectrometry (MS) detection was performed in positive ion mode, with an interface temperature of 300°C. The collision-induced dissociation gas was optimized to 230 kPa, and the electrospray capillary voltage was set at 4.0 kV.

Preparation of double-strand DNA with C and 5mC

The 216-bp double-stranded DNA molecules (216-bp C-DNA and 216-bp 5mC-DNA) were synthesized via PCR amplification (Table S2). A 228-bp DNA fragment with the sequence 5′-ACTAGT/DNA-C/CTCGAG-3′ was designed and cloned into the pET-41a plasmid (pET-41a-4CG) using SpeI and XhoI restriction sites. This cloned plasmid served as the template for PCR amplification. The reaction was performed in a 50 μL reaction solution consisting of 1×

reaction buffer, 4 ng of pET-41a-4CG plasmid, 1 U of Taq DNA polymerase, 0.2 mM dNTP (with dCTP replaced by 5mdCTP for the preparation of 216-bp 5mC-DNA), 0.4 μM forward primer, and 0.4 μM reverse primer (Table S5). The PCR protocol involved an initial denaturation at 95°C for 2 min, followed by 35 cycles of 95°C for 15 s, 52°C for 20 s, and 72°C for 30 s, with a final extension at 72°C for 5 min. The 400-bp DNA was produced under identical conditions, except using the pUC19 plasmid as the PCR template. PCR products were verified by agarose gel electrophoresis and purified using the MicroElute DNA Clean-Up kit (Omega Bio-Tek Inc., Norcross, Georgia). The concentrations of the purified C-DNA and 5mC-DNA were quantified using a B-500 spectrophotometer (Metash Instruments Co., Ltd., Shanghai, China). The detailed sequences of these double-stranded DNA molecules can be found in Table S2.

Evaluation of the deamination activity of A3A by sequencing

The 216-bp dsDNA molecules (216-bp C-DNA and 216-bp 5mC-DNA, Table S2) were utilized to assess the deamination activity of A3A. Initially, 20 ng of dsDNA was denatured at 95°C for 10 min in a 20% DMSO (v/v) solution, followed by rapid cooling on ice water for 5 min. The deamination reaction was conducted in a 20 μL solution containing 10 μM A3A, 20 mM 2-morpholinoethanesulfonate (MES) (pH 6.5), and 0.1% Triton X-100, under the following temperature gradient: 4°C for 5 min, linear ramping temperature from 4-37°C at + 0.1°C/13 s, 37°C for 2 h, and linear ramping temperature from 37-50°C at + 0.1°C/13 s. The reaction was terminated by heating at 95°C for 10 min. Subsequently, PCR amplification was performed in a 50 μL solution comprising 10 μL of 5× reaction buffer, 0.625 U of EpiMark hot start Taq DNA polymerase, 0.2 mM dNTP, 0.4 μM forward primer, and 0.4 μM reverse primer (Table S5). The PCR protocol consisted of 30 cycles of initial denaturation at 95°C for 5 min, 95°C for 30 s, 50°C for 30 s, and 68°C for 1 min, followed by an additional 10 min of elongation at 68°C. The resulting PCR products were subjected to Sanger sequencing (Figure S8).

Table S1. Amino acid sequences of the proteins used in this study.

Name	Amino acid sequence (N terminal to C terminal)
M.MpeI-N374K	MNSNKDKIKVIKVFEAFAGIGSQFKALKNIARSKNWEIQHSGMVEWFVDA IVSYVAIHSKNFNPKIEQLDKDILSISNDSKMPISEYGIKKINNTIKASYLNYA KKHFNNLFDIKKVNKDNFPKNIDIFTYSFPCQDLSVQGLQKGIDKELNTRS GLLWEIERILEEIKNSFSKEEMPKYLLMENVKNLLSHKNKKNYNTWLKQL EKFGYKSKTYLLNSKNFDNCQNRERVFCLSIRDDYLEKTGFKFKELEKVK NPPKKIKDILVDSSNYKYLNLNKYETTTFRETKSNIISRSLKNYTTFNSENY VYNINGIGPTLTASGANSRIKIETQQGVRYLTPLECFKYMQFDVNDFKKVQ STNLISENKMIYIAGKSIPVKILEAIFNTLEFVNNEE
M.MpeI- E45D/N374K	MNSNKDKIKVIKVFEAFAGIGSQFKALKNIARSKNWEIQHSGMVDWFVD AIVSYVAIHSKNFNPKIEQLDKDILSISNDSKMPISEYGIKKINNTIKASYLN YAKKHFNNLFDIKKVNKDNFPKNIDIFTYSFPCQDLSVQGLQKGIDKELNT RSGLLWEIERILEEIKNSFSKEEMPKYLLMENVKNLLSHKNKKNYNTWLK QLEKFGYKSKTYLLNSKNFDNCQNRERVFCLSIRDDYLEKTGFKFKELEK VKNPPKKIKDILVDSSNYKYLNLNKYETTTFRETKSNIISRSLKNYTTFNSE NYVYNINGIGPTLTASGANSRIKIETQQGVRYLTPLECFKYMQFDVNDFKK VQSTNLISENKMIYIAGKSIPVKILEAIFNTLEFVNNEE
M.MedI-WT	MKEIKIFEFFSGIGSQMKALKNLEKSLNFTTKSVGACDFYIDAIVSYMCIHH GNLEPENDFTKEEMISILDKFKFSNNSKDIVSENYFKKINEQKLRQLFPYLF AFINNDYFNKKYDKNILQYERLNATDIRDFDTLPDNIDILTYSFPCQDLSQQ GKQKGIEKNTRSGLLYEIERILKLNLNNLPKVLILENVKALVSKKFINQFNA WINVLSELGYKSSWKIMNASDFGSAQNRERVFMVSVLSDENFEFPKINEN NSKNVSNIWEYDGEHKIIQLDSNQKMNDFKITKNKIQKAFINNYSNFNSEN YIYSINSKGATLTASGANSRLKFWVNNEIQIMNSLEALLYMGFERDDYEKI KSSNLLNENKIIFTAGNSISVEVLETLFKKIIKEVITDEQ
M.MedI-N377K	MKEIKIFEFFSGIGSQMKALKNLEKSLNFTTKSVGACDFYIDAIVSYMCIHH GNLEPENDFTKEEMISILDKFKFSNNSKDIVSENYFKKINEQKLRQLFPYLF AFINNDYFNKKYDKNILQYERLNATDIRDFDTLPDNIDILTYSFPCQDLSQQ GKQKGIEKNTRSGLLYEIERILKLNLNNLPKVLILENVKALVSKKFINQFNA WINVLSELGYKSSWKIMNASDFGSAQNRERVFMVSVLSDENFEFPKINEN NSKNVSNIWEYDGEHKIIQLDSNQKMNDFKITKNKIQKAFINNYSNFNSEN YIYSINSKGATLTASGANSRLKFWVNNEIQIMNSLEALLYMGFERDDYEKI KSSNLLNENKIIFTAGKSISVEVLETLFKKIIKEVITDEQ
A3A	MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQ HRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSP CFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQ VSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQG N

Table S2. Sequences of oligonucleotides.

Name	Sequence (5'-3')
28-CG-F	TATTTATACGTAACCGATCGTATTTTAT
28-CG-R	ATAAAATACGATCGGTTACGTATAAATA
49-CG-FAM-F	AATTATTAAAAATATATAAAACCGGATTAAATATAATATAAATATAATT-FAM
49-CG-R	AATTATATTTATATTATATTTAATCCGGTTTTATATATTTTTAATAATT
216-bp C-DNA	AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGATTATGAGTAT GTATAGTGTTAGGAAGAGTGTAGTAATAGGATGAAGATGATTATAT GATCGATGGTCCGTATGCGTAGAATACGTTGTTGTAGTGATTATAAT GGAGTGAGAATGTAGATGAGTGGAGTAGGTAGTAAGATGTAGTGGT GACAGAGAGTAATTGTTAGTGGAATGTTGG
216-bp 5mC-DNA	AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGATTATGAGTAT GTATAGTGTTAGGAAGAGTGTAGTAATAGGATGAAGATGATTATAT GAT5mCGATGGT5mC5mCGTATG5mCGTAGAATA5mCGTTGTTGTAGT GATTATAATGGAGTGAGAATGTAGATGAGTGGAGTAGGTAGTAAGA TGTAGTGGTGA5mCAGAGAGTAATTGTTAGTGGAATGTTGG
400-bp C-DNA	GTGAGTGAGTGAGATTGATGTGAGTGATTTTTCTCTGGTCCCGCCGC ATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGC ATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCAT CGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCAT CAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATC AGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACG CGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGAT GAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAA GTAATTGTTAGTGGAATGT
400-bp 5mC-DNA	GTGAGTGAGTGAGATTGATGTGAGTGATTTTT5mCT5mCTGGT5mC 5mC5mCG5mC5mCG5mCAT5mC5mCATA5mC5mCG5mC5mCAGTTG TTTA5mC5mC5mCT5mCA5mCAA5mCGTT5mC5mCAGTAA5mC5mCGG G5mCATGTT5mCAT5mCAT5mCAGTAA5mC5mC5mCGTAT5mCGTGAG 5mCAT5mC5mCT5mCT5mCT5mCGTTT5mCAT5mCGGTAT5mCATTA 5mC5mC5mC5mC5mC5mCATGAA5mCAGAAAT5mC5mC5mC5mC5mCTT A5mCA5mCGGAGG5mCAT5mCAGTGA5mC5mCAAA5mCAGGAAAAAA 5mC5mCG5mC5mC5mCTTAA5mCATGG5mC5mC5mCG5mCTTTAT5mCA GAAG5mC5mCAGA5mCATTAA5mCG5mCTT5mCTGGAGAAA5mCT5mC AA5mCGAG5mCTGGA5mCG5mCGGATGAA5mCAGG5mCAGA5mCAT 5mCTGTGAAT5mCG5mCTT5mCA5mCGA5mC5mCA5mCG5mCTGATGA G5mCTTTA5mC5mCG5mCAG5mCTG5mC5mCT5mCG5mCG5mCGTTT 5mCGGTGATGACGGTGAAAAGTAATTGTTAGTGGAATGT

Table S3. The mass spectrometry parameters for the analysis of nucleosides.

Analyte	Precursor ion	Product ion	Q1 Prerod/V	CE / V	Q3 Prerod/ V
G	268.2	152.1	22	10	29
A	252.2	136.1	20	15	20
C	228.2	112.1	11	10	20
T	243.2	127.0	12	10	22
5mC	242.2	126.1	20	11	24
5camC	286.2	170.1	20	11	30

Table S4. The conditions for steady-state kinetics study.

DNA methyltransferase	Concentration (nM)	Reaction time (min)
M.MpeI-N374K	640	30
M.MpeI-E45D/N374K	75	15
M.MedI-WT	750	40
M.MedI-N377K	75	2

Table S5. Sequences of PCR primers.

Name	Sequence (5'-3')
216-up	AGTGACGCTGAGCTTGACGTCGCGC
216-down	CCAACATTCCACTAACAATTACTCTCT
216-A3A-up	AGTGATGCTGAGTTTGATGTTGTGT
216-A3A-down	CCAACATTCCACTAACAATTACTCTCTA
216-MF	AGTGACGTTGAGTTTGACGTC
216-EXT	TAACTTCACGAACGTACATCGCCGGCCAACATTCCACTAAC AATTACTCTCT
216-Edown	CGTAACTTCACAAACATACATCACCAA
400-up	GTGAGTGAGTGAGATTGATGTGAGTGATTTTTC
400-down	ACATTCCACTAACAATTACTTTTCACCGTCATCAC
400-MF	TGAGTGATTTTTTTTTGGTTTCGTCGTA
400-MR	GAAACGCGCGAAACAACTACGA
RASSF1A-F	GGTTAGGGATTAGTTGTCGTGT
RASSF1A-R	GCCGCCCAATCTAAATCCTAAA
SHOX2-F	GTTGTATTTGTTTGTTGAAATGGTATTTTT
SHOX2-R	CTTAATAATCTTCAAAAACTATTAATAAAATCC

Figure S1. Comparison of the amino acid sequences of the four DNA methyltransferases.

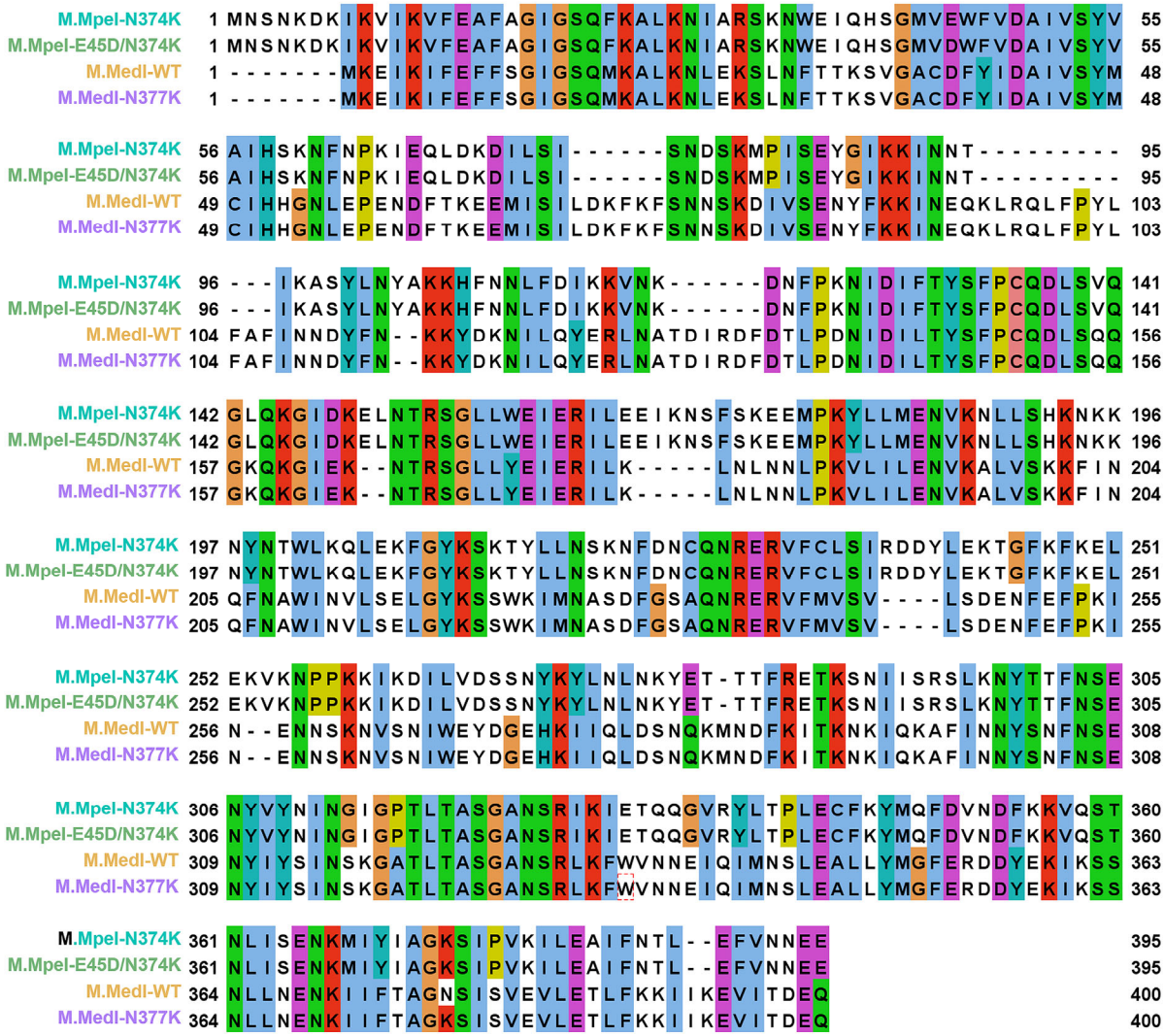


Figure S2. Expression of M.MedI-WT protein. (A) The schematic illustration of plasmid for the expression of M.MedI-WT protein. (B) SDS-PAGE analysis of the purified M.MedI-WT protein.

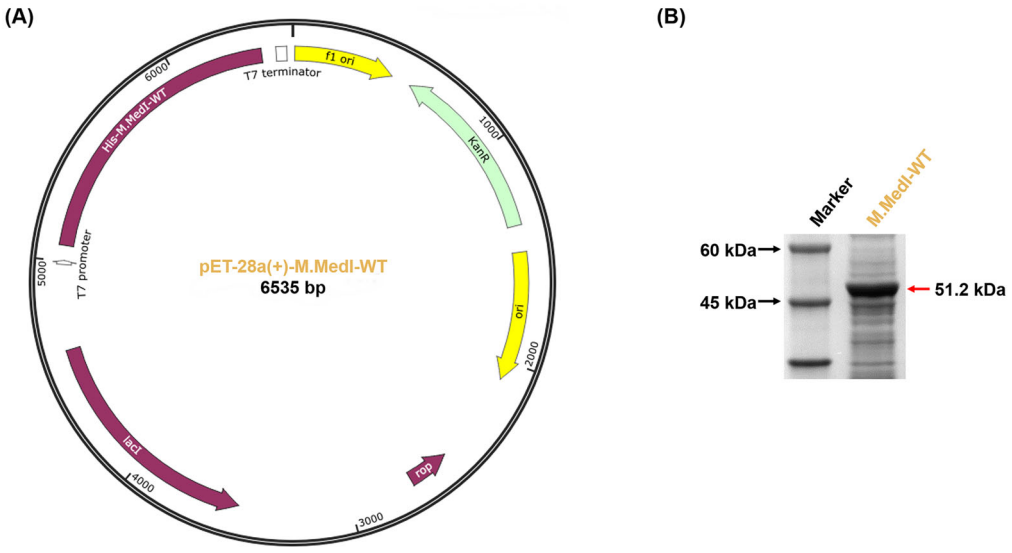


Figure S3. The schematic illustration of plasmid for the expression of M.MedI-N377K protein.

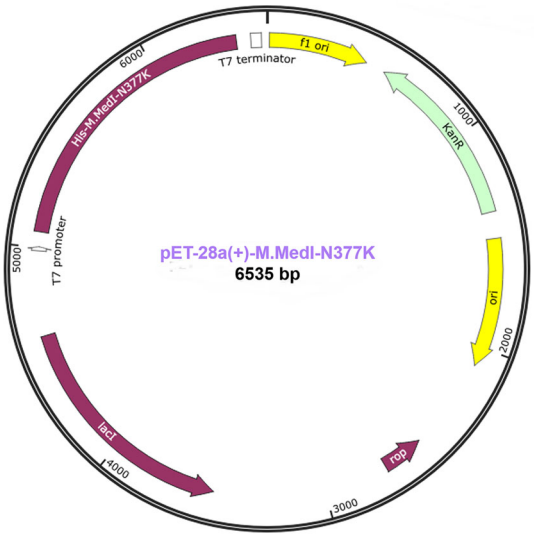


Figure S4. Expression of M.MpeI-N374K protein. (A) The schematic illustration of plasmid for the expression of M.MpeI-N374K protein. (B) SDS-PAGE analysis of the purified M.MpeI-N374K protein.

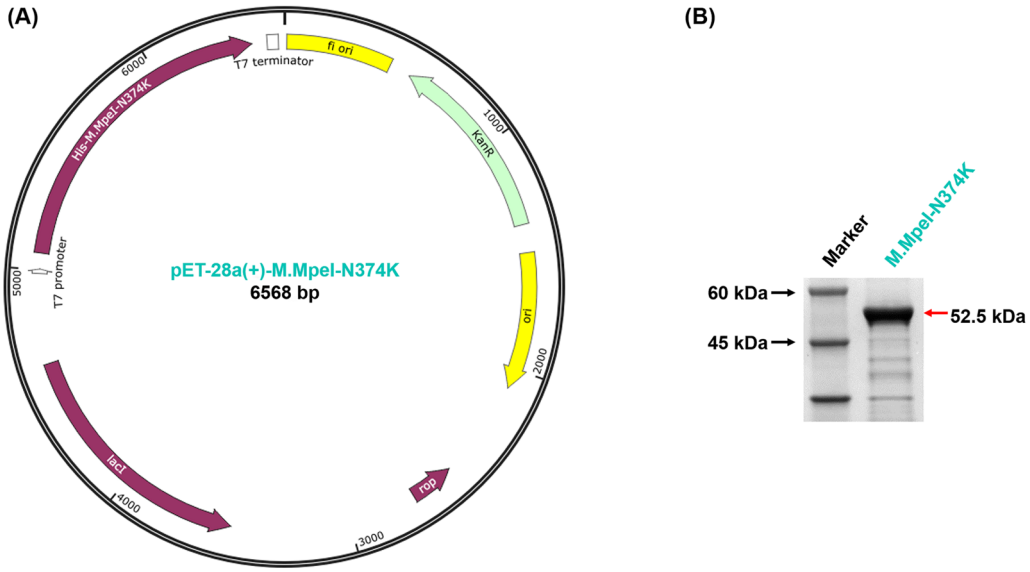


Figure S5. Expression of M.MpeI-E45D/N374K protein. (A) The schematic illustration of plasmid for the expression of M.MpeI-E45D/N374K protein. (B) SDS-PAGE analysis of the purified M.MpeI-E45D/N374K protein.

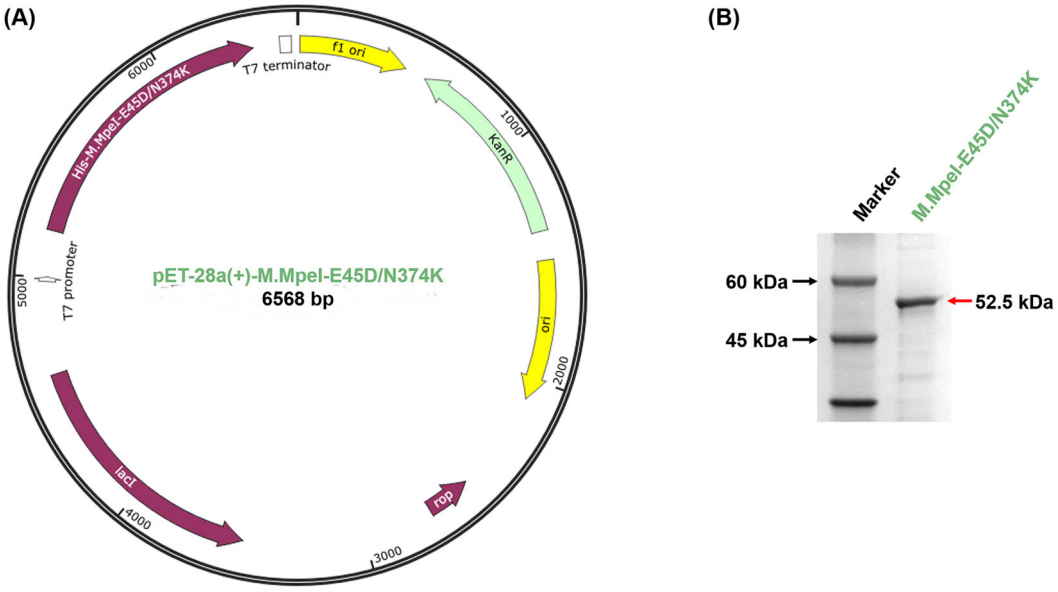


Figure S6. Rate versus caSAM concentration curves in the steady-state kinetics study. The data were fitted to the Michaelis-Menten equation. (A) DNA treated with M.MpeI-N374K. (B) DNA treated with M.MpeI-E45D/N374K. (C) DNA treated with M.MedI-WT. (D) DNA treated with M.MedI-N377K. A 49-bp duplex DNA (49-CG-FAM) was used as the substrate in this study.

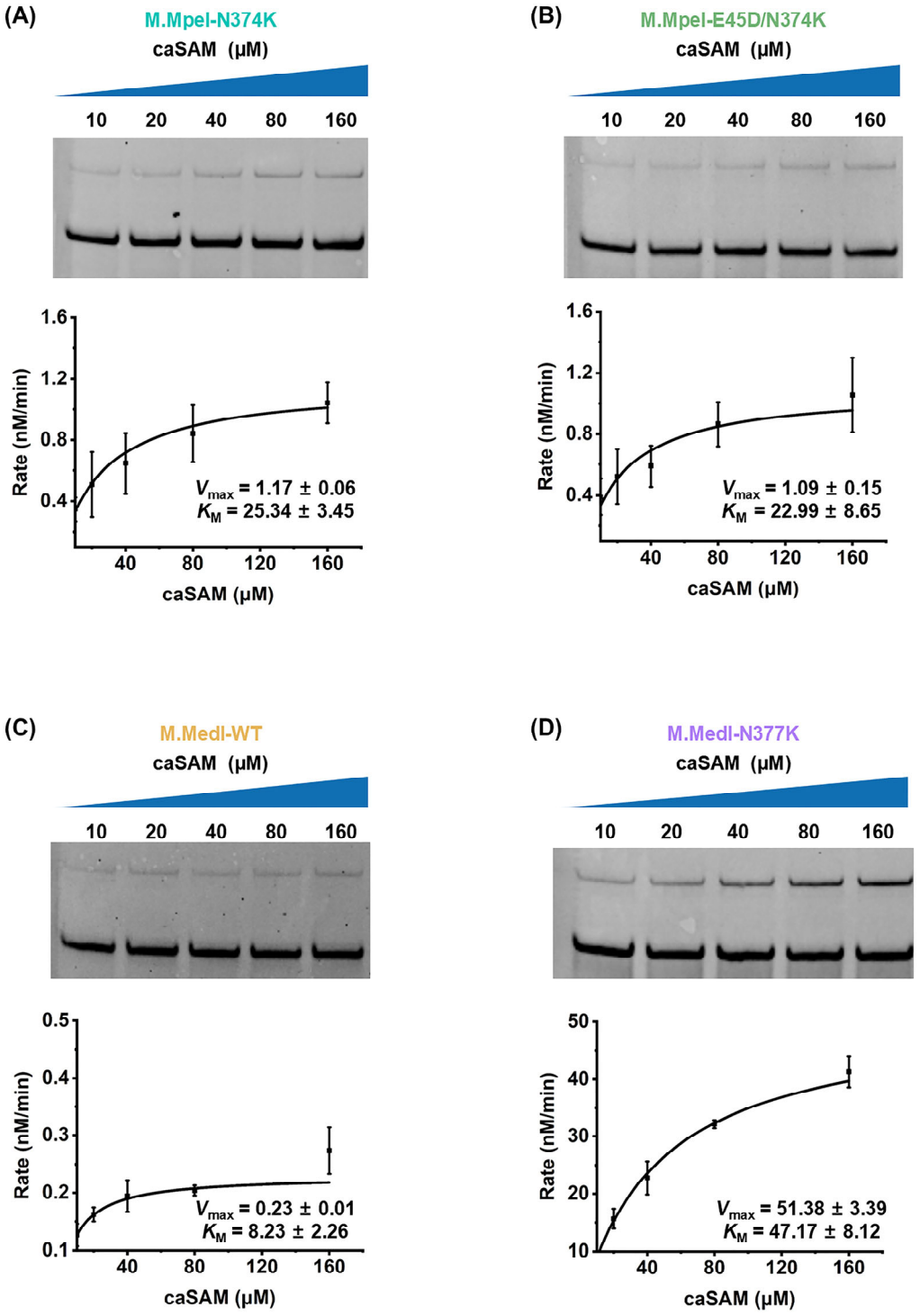


Figure S7. Expression of A3A protein. (A) The schematic illustration of plasmid for the expression of A3A protein. (B) SDS-PAGE analysis of the purified A3A protein.

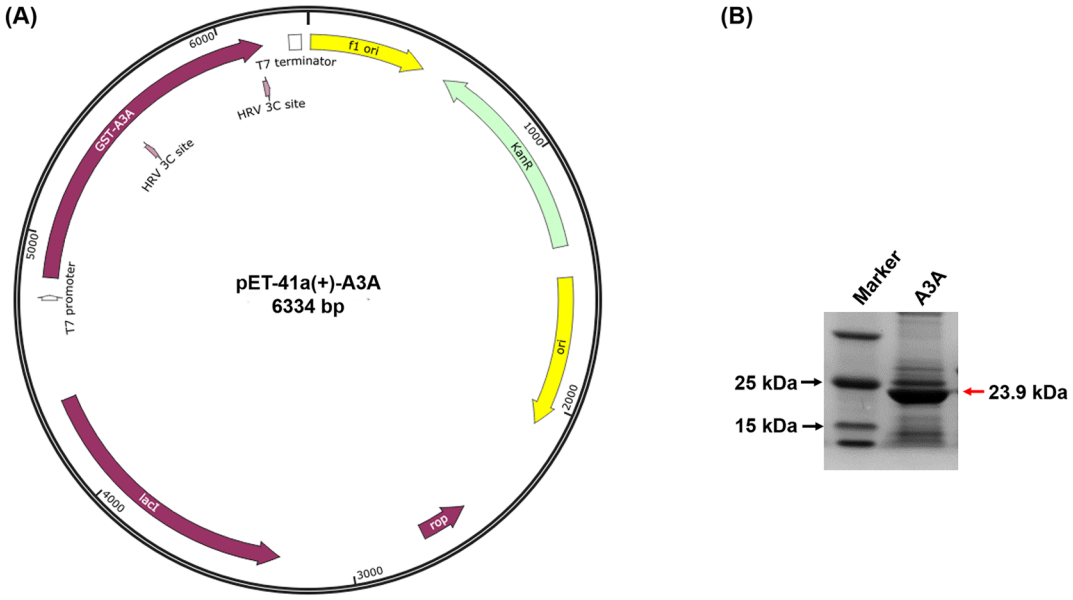
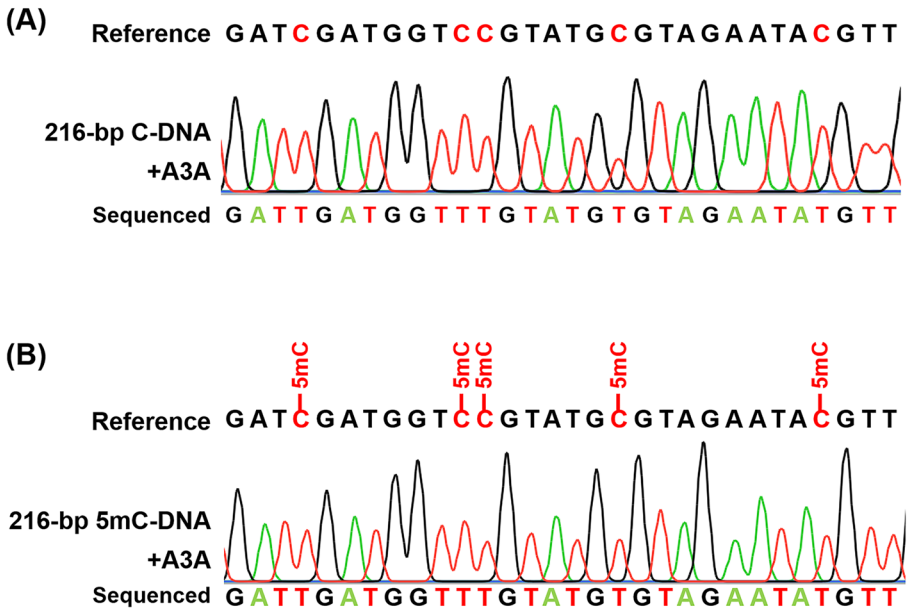


Figure S8. Sequencing results of 216-bp dsDNA after A3A treatment. (A) 216-bp C-DNA. (B) 216-bp 5mC-DNA.



References

1. J. Xiong, K. K. Chen, N. B. Xie, T. T. Ji, S. Y. Yu, F. Tang, C. Xie, Y. Q. Feng and B. F. Yuan, *Anal Chem*, 2022, **94**, 15489-15498.
2. N. B. Xie, M. Wang, T. T. Ji, X. Guo, J. H. Ding, B. F. Yuan and Y. Q. Feng, *Chem Sci*, 2022, **13**, 7046-7056.
3. N. B. Xie, M. Wang, T. T. Ji, X. Guo, F. Y. Gang, Y. Hao, L. Zeng, Y. F. Wang, Y. Q. Feng and B. F. Yuan, *Chem Sci*, 2024, **15**, 10073-10083.
4. C. J. Ma, G. Li, W. X. Shao, Y. H. Min, P. Wang, J. H. Ding, N. B. Xie, M. Wang, F. Tang, Y. Q. Feng, W. Ci, Y. Wang and B. F. Yuan, *ACS Cent Sci*, 2023, **9**, 1799-1809.