Supporting Information

for

Simultaneous Detection of 5-Methylcytosine and 5-Hydroxymethylcytosine at Specific Genomic Loci by Engineered Deaminase-Assisted Sequencing

Neng-Bin Xie,[a,b,c,‡] Min Wang,[d,e,‡] Tong-Tong Ji,[d] Xia Guo,[d] Fang-Yin Gang,[a] Ying Hao,[d] Li Zeng,[d] Ya-Fen Wang,[a] Yu-Qi Feng,[a] Bi-Feng Yuan [a,b,c,d,*]

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

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

[c] Cancer Precision Diagnosis and Treatment and Translational Medicine Hubei Engineering Research Center, Zhongnan Hospital of Wuhan University, Wuhan Research Center for Infectious Diseases and Cancer, Chinese Academy of Medical Sciences, Wuhan 430071, China.

[d] College of Chemistry and Molecular Sciences, Wuhan University, Wuhan 430072, China.

[e] College of Chemical Engineering and Environmental Chemistry, Weifang University, Weifang, 261061, China.

‡ These authors contributed equally to this work.

* Corresponding author:

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

Table of Contents

Preparation of 5mC spike-in DNA

A series of double-stranded 5mC spike-in DNA were synthesized for the calculation of the deamination rates of 5mC in different sequence contexts. Detailed sequences are provided in Tables S2.

For the preparation of DNA-C/5mC spike-in, 0.5 ng of synthetic DNA from Takara was used as a template for PCR amplification. The PCR was performed in a 50-μL mixture containing 1 unit of Q5 High-Fidelity DNA polymerase, 0.2 mM of dNTP, 5 μL of 10× reaction buffer, 2 μL of 10 μM forward primer, and 2 μL of 10 μM reverse primer (Table S4). The PCR reaction consisted of 95°C for 5 min, 30 cycles of 95°C for 1 min, 61°C for 1 min, 68°C for 1 min, followed by an elongation at 68°C for 10 min. The PCR products were separated using agarose gel electrophoresis and extracted with a Gel Extraction kit (Omega Bio-Tek Inc., Norcross, GA, USA).

For the DNA-5mC2 spike-in, 0.5 ng of synthetic DNA from Takara was used as a template for PCR amplification. The PCR was performed in a 50-μL mixture containing 1 unit of Q5 High-Fidelity DNA polymerase, 0.2 mM of dATP, dTTP, and dGTP, 0.2 mM of 5mdCTP, 5 μL of 10× reaction buffer, 2 μL of 10 μM forward primer, and 2 μL of 10 μM reverse primer (Table S4). The PCR reaction consisted of 95°C for 5 min, 30 cycles of 95°C for 1 min, 53°C for 1 min, 68°C for 1 min, followed by an elongation at 68°C for 10 min. Similar protocols with different primers (Table S4) were used for the preparation of *THRA* 5mC spike-in DNA and *ALS2CL* 5mC spike-in DNA. For the DNA-5mC2 spike-in, *THRA* 5mC spike-in DNA, and *ALS2CL* 5mC spike-in DNA, all cytosines were replaced with 5mC, excluding the cytosines in the PCR primers.

Table S1. Sequences of DNA substrates with different modifications.

Name	Sequence (5' to 3')
DNA-C1	GAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGA TTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAG GATGAAGATGATTATATGATCGATGGTCCGTATGCGTAG AATACGTTGTTGTAGTGATTATAATGGAGTGAGAATGTA GATGAGTGGAGTAGGTAGTAAGATGTAGTGGTGAAGAG AGTAATTGTTAGTGGAATGTTGGAGAGGAT
DNA-5mC1	GAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGA TTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAG GATGAAGATGATTATATGAT5mCGATGGT5mC5mCGTAT G5mCGTAGAATA5mCGTTGTTGTAGTGATTATAATGGAG TGAGAATGTAGATGAGTGGAGTAGGTAGTAAGATGTAG TGGTGAAGAGAGTAATTGTTAGTGGAATGTTGGAGAGG AT
DNA-5hmC1	GAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGA TTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAG GATGAAGATGATTATATGAT5hmCGATGGT5hmC5hmCGT ATG5hmCGTAGAATA5hmCGTTGTTGTAGTGATTATAATG GAGTGAGAATGTAGATGAGTGGAGTAGGTAGTAAGATG TAGTGGTGAAGAGAGTAATTGTTAGTGGAATGTTGGAG AGGAT
DNA-C2	GATGAGAGGTGATTATGAGTCCGTATAGTGTGGTAGGAT ATGCGTATGTAGTATGATAGTGTACGTGTAGGTTATATG ATTATTCGAGTATGATTATATGATATGGTGAGTATGTGT AGAATCAGAGAGTAATTGTTAGTGGAATGTTGG
DNA-5mC2	GATGAGAGGTGATTATGAGTC5mCGTATAGTGTGGTAGG ATATG5mCGTATGTAGTATGATAGTGTA5mCGTGTAGGT TATATGATTATT5mCGAGTATGATTATATGATATGGTGA GTATGTGTAGAATCAGAGAGTAATTGTTAGTGGAATGTT GG
DNA-5hmC2	GATGAGAGGTGATTATGAGTC5hmCGTATAGTGTGGTAG GATATG5hmCGTATGTAGTATGATAGTGTA5hmCGTGTAG GTTATATGATTATT5hmCGAGTATGATTATATGATATGGT GAGTATGTGTAGAATCAGAGAGTAATTGTTAGTGGAAT GTTGG
DNA-C/5mC	AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGAT TATGAGTA5mCGTATAGTGTTAGGATAGAGTTCGTGTAG TAATAGGATGATAGTATGATTATATGAATGGTGAGTATG TGTAGAATCAGAGAGTAATTGTTAGTGGAATGTTGG

Table S2. Sequences of spike-ins.

Name	Sequence (5' to 3')
DNA-C/5mC spike-in	AGGAGTCAGCTGGGCTCACGCGCACTCTGAGAGGTGA TTATGAGTA5mCGTATAGTGTTAGGATAGAGTTCGTGT AGTAATAGGATGATAGTATGATTATATGAATGGTGAG TAAGTTATGTGTGAGGTGATGATGA
DNA-5mC2 spike-in	AGGAGTCAGCTGGGCTCACGCGCACTCTAGTGATATG TGATGAGAGGTGATTATGAGTC5mCGTATAGTGTGGT AGGATATG5mCGTATGTAGTATGATAGTGTA5mCGTGT AGGTTATATGATTATT5mCGAGTATGATTATATGATAT GGTGAGTATGTGTAGAATAGTTAGTAGTAAGTTATGT GTGAGGTGATGATGA
THRA 5mC spike-in DNA	AGGAGTTAGTTGGGTTCACGTGCATTTGAGGGTATAG TTGTTTAGATTAAGTATGTATTAAG5mCGAGGTTTTTT TATTTGTGAGTGTAGGTTTTAGTTTAGAAGGGGAAGTT ATGTGTGAGGTGATGATGA
ALS2CL 5mC spike-in DNA	AGTGACGCTGAGCTTGACGTCGCGCTGTTTAGTTTTGA GTGTTTATGT5mCGATGGTTTTTAAATTTAAGGGTTTA GTGAAATATAGGTTTTTGGGTTTATTTTTTGTATTTTTG AGAGAGTAATTGTTAGTGGAATGTTGG

Table S3. Sequences of primers used in the synthesis of different DNA.

Synthetic DNA	Sequence (5' to 3')
DNA-C1, DNA-5mC1 and DNA-5hmC1	Forward: GAGTGACGCTGAGCTTGACGTCGCGC
	Reverse: ATCCTCTCCAACATTCCACTAACAATTA
DNA-C2, DNA-5mC2 and DNA-5hmC2	Forward: GATGAGAGGTGATTATGAGTC
	Reverse: CCAACATTCCACTAACAATTACTC
DNA-C/5mC	Forward:AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGATTATGAGTA5mC
	Reverse: CCAACATTCCACTAACAATTACTCTCTG
DNA-C/5mC spike-in	Forward:AGGAGTCAGCTGGGCTCACGCGCACTCTGAGAGGTGATTATGAGTA5mC
	Reverse: TCATCATCACCTCACACATAACTTACTC
DNA-5mC2 spike-in	Forward: AGGAGTCAGCTGGGCTCACGCGCACTCT
	Reverse: TCATCATCACCTCACACATAACTTACT
THRA 5mC spike-in DNA	Forward: AGGAGTTAGTTGGGTTCACGTGC
	Reverse: TCATCATCACCTCACACATAACTTC
ALS2CL 5mC spike-in DNA	Forward: AGTGACGCTGAGCTTGACGTCGCGC
	Reverse: CCAACATTCCACTAACAATTACTC

Table S4. Sequences of primers used in the amplification of deaminated DNA.

Deaminated DNA	Sequence (5' to 3')
DNA-C1, DNA-5mC1 and DNA-5hmC1	Forward: GAGTGATGTTGAGTTTGATGTTGTGT
	Reverse: ATCCTCTCCAACATTCCACTAACAATTA
DNA-C2, DNA-5mC2 and DNA-5hmC2	Forward: GATGAGAGGTGATTATGAGTT
	Reverse: CCAACATTCCACTAACAATTACTC
DNA-C/5mC	Forward: AGTGATGTTGAGTTTGATGTTGT
	Reverse: CCAACATTCCACTAACAATTACTCTCTA
DNA-C/5mC spike-in	Forward: AGGAGTTAGTTGGGTTTATGTGTATTT
	Reverse: TCATCATCACCTCACACATAACTTACTC
DNA-5mC2 spike-in	Forward: AGGAGTTAGTTGGGTTTATGTGTATTT
	Reverse: TCATCATCACCTCACACATAACTTACT
5mC sites in *THRA* gene body	Forward: TTGTTTAGTTTTAGGGTTTGGTTTT
	Reverse: CAAATTCCCTCTCCAAAAAAAATA
5mC sites in *ALS2CL* gene body	Forward: ATAATAGGTATTGGTTATTGGGTGT
	Reverse: ACAAATTCTTAAAACCACCTCAAAC
THRA 5mC spike-in DNA	Forward: AGGAGTTAGTTGGGTTTATGTGT
	Reverse: TCATCATCACCTCACACATAACTTC
ALS2CL 5mC spike-in DNA	Forward: AGTGATGTTGAGTTTGATGTTGTGT
	Reverse: CCAACATTCCACTAACAATTACTC

Table S5. Information of the individual cytosine sites detected in genome.

Genome location	Gene	Gene element
chr17.38222379, CRCh37	*THRA*	Gene body
chr3.46713993, CRCh37	*ALS2CL*	Gene body

Figure S1. The schematic illustration of plasmid for the expression of wtA3A or eA3A proteins.

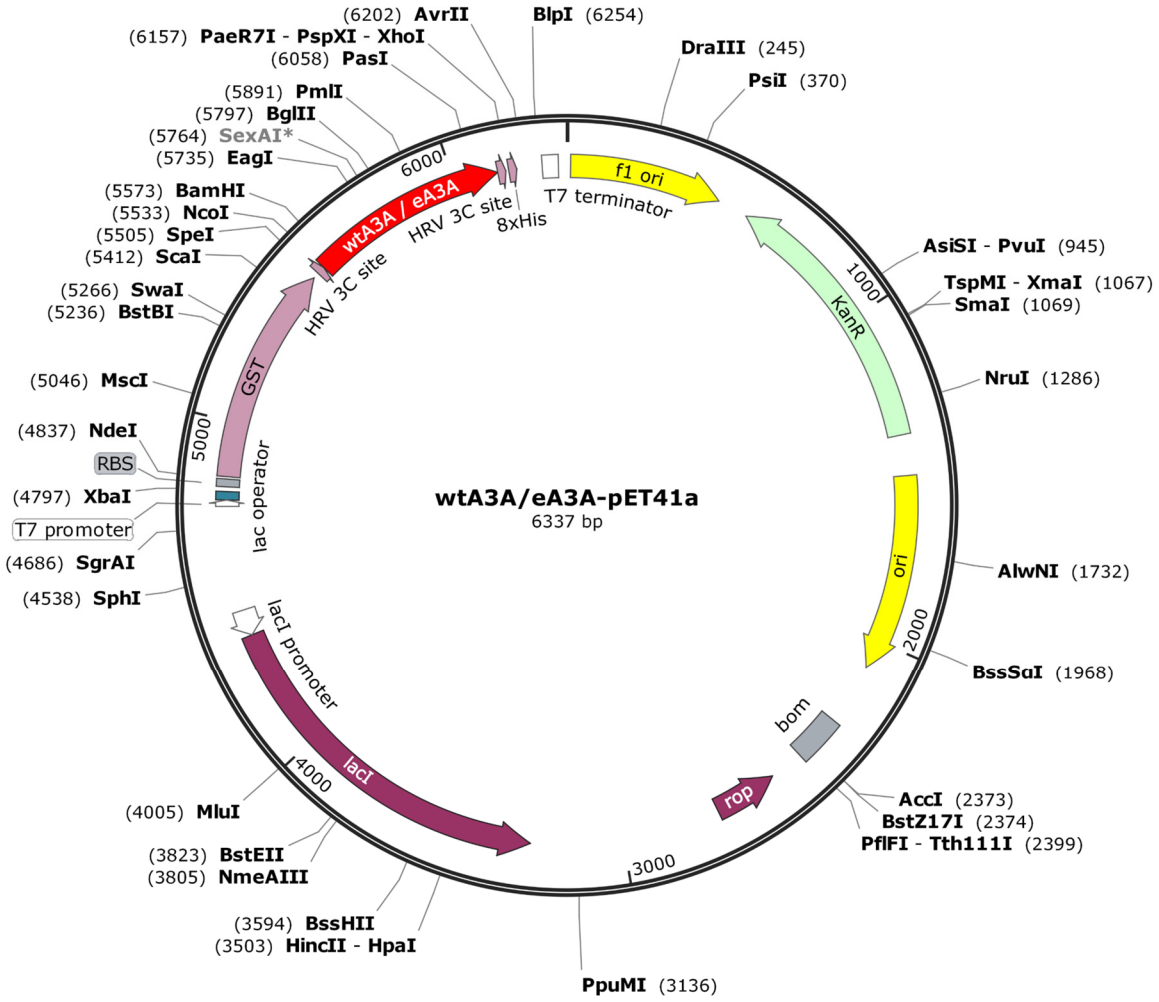

Figure S2. SDS-PAGE analysis of the purified wtA3A and eA3A-M5. "M" represents protein marker. (A) SDS-PAGE analysis of the purified wtA3A. (B) SDS-PAGE analysis of the purified eA3A-M5.

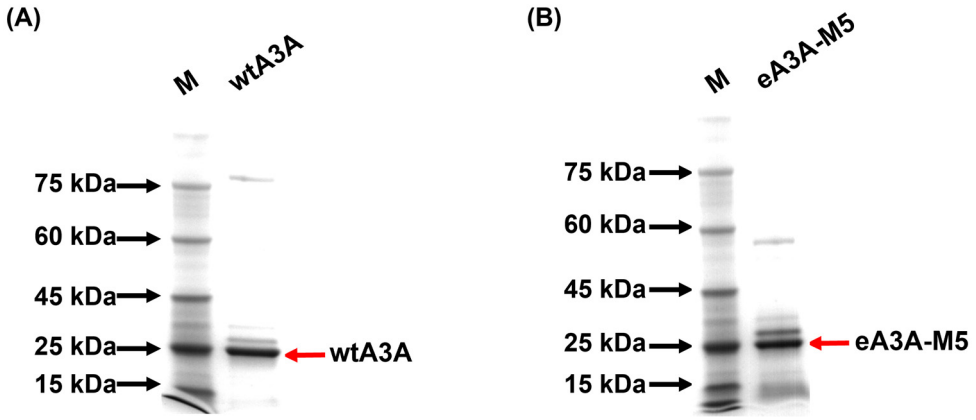

Figure S3. The schematic illustration of analytical strategies for mapping 5mC and 5hmC. (A) The schematic illustration of bisulfite treatment-based mapping methods. (B) The schematic illustration of deaminase treatment-based mapping methods. (C) The schematic illustration of pyridine borane treatment-based mapping methods.

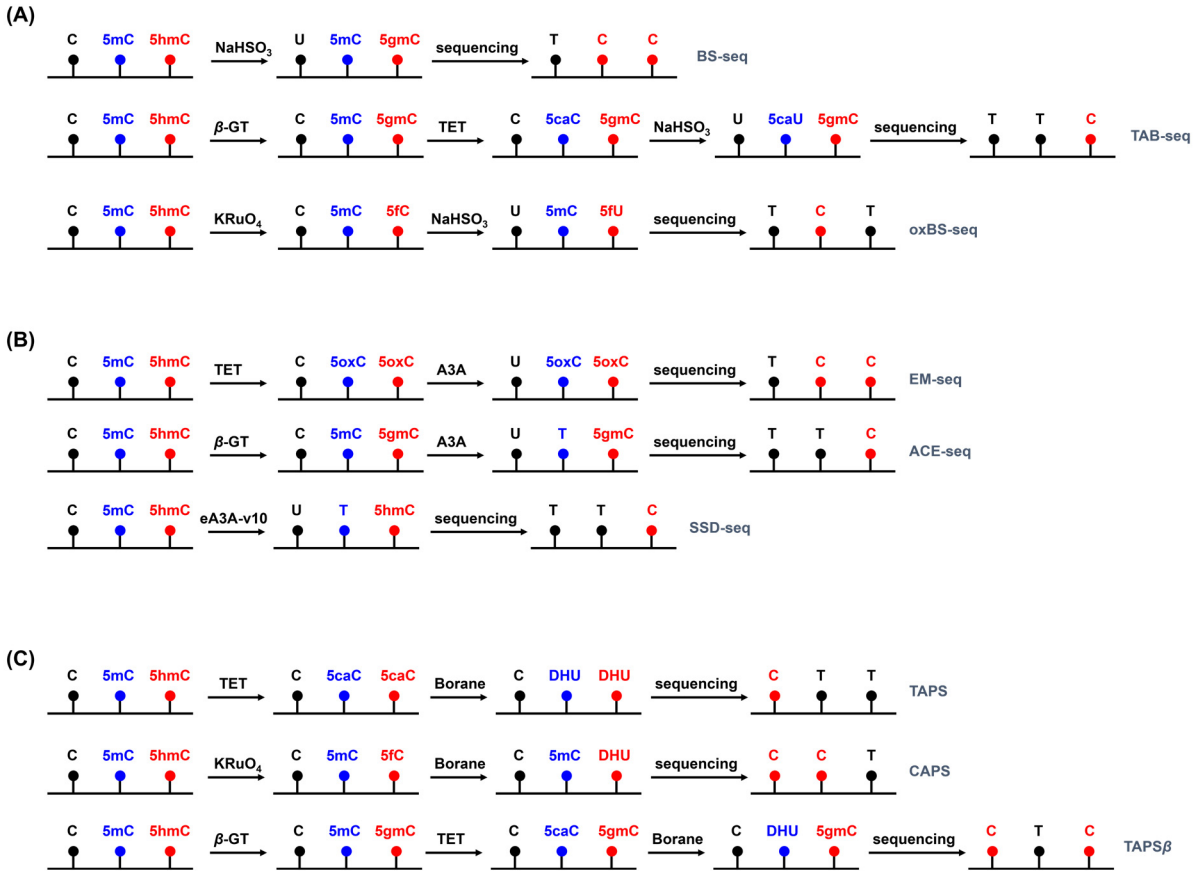

Figure S4. Assessment of wtA3A specificity for C, 5mC, and 5hmC across various sequence contexts. (A) The amino acid compositions of wtA3A and engineered A3A mutants (eA3A-v1 to eA3A-M5). (B) Sanger sequencing results for DNA substrates containing C, 5mC, and 5hmC (DNA-C1, DNA-5mC1, and DNA-5hmC1, respectively) as treated by wtA3A. wtA3A converted all the C and 5mC sites to U and T, respectively, which were subsequently read as T during sequencing; 5hmC sites underwent partial deamination, resulting in a mixed read of C and T.

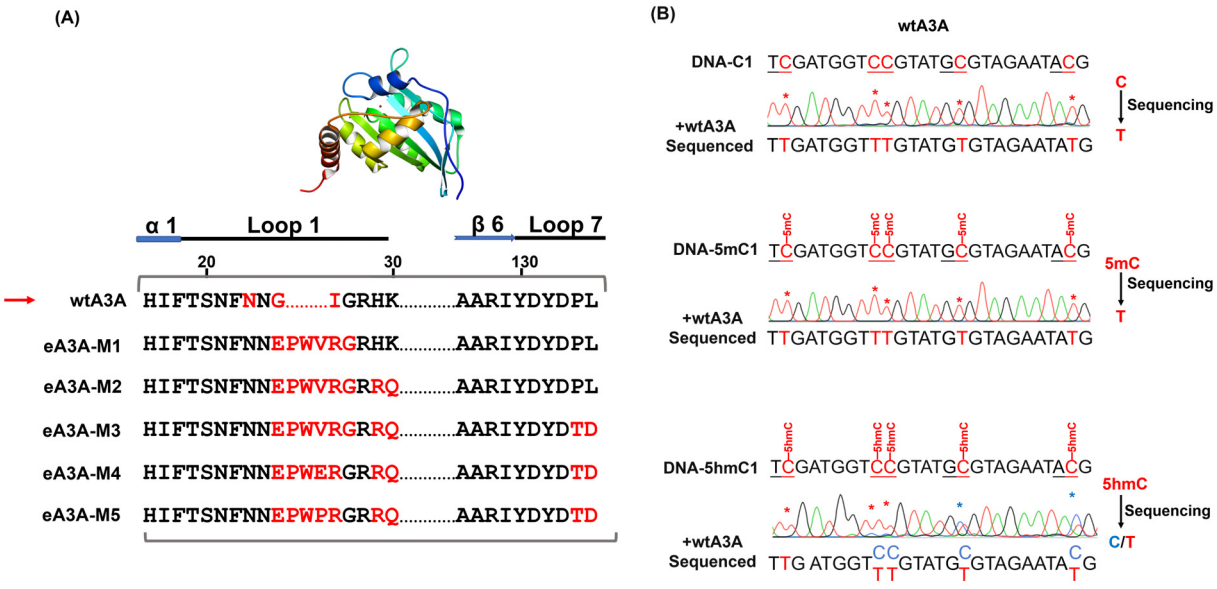

Figure S5. Assessment of eA3A-M1 specificity for C, 5mC, and 5hmC across various sequence contexts. (A) The amino acid compositions of wtA3A and engineered A3A mutants (eA3A-v1 to eA3A-M5). (B) Sanger sequencing results for DNA substrates containing C, 5mC, and 5hmC (DNA-C1, DNA-5mC1, and DNA-5hmC1, respectively) as treated by eA3A-M1. eA3A-M1 converted all the C and 5mC sites to U and T, respectively, which were subsequently read as T during sequencing; 5hmC sites underwent partial deamination, resulting in a mixed read of T and C.

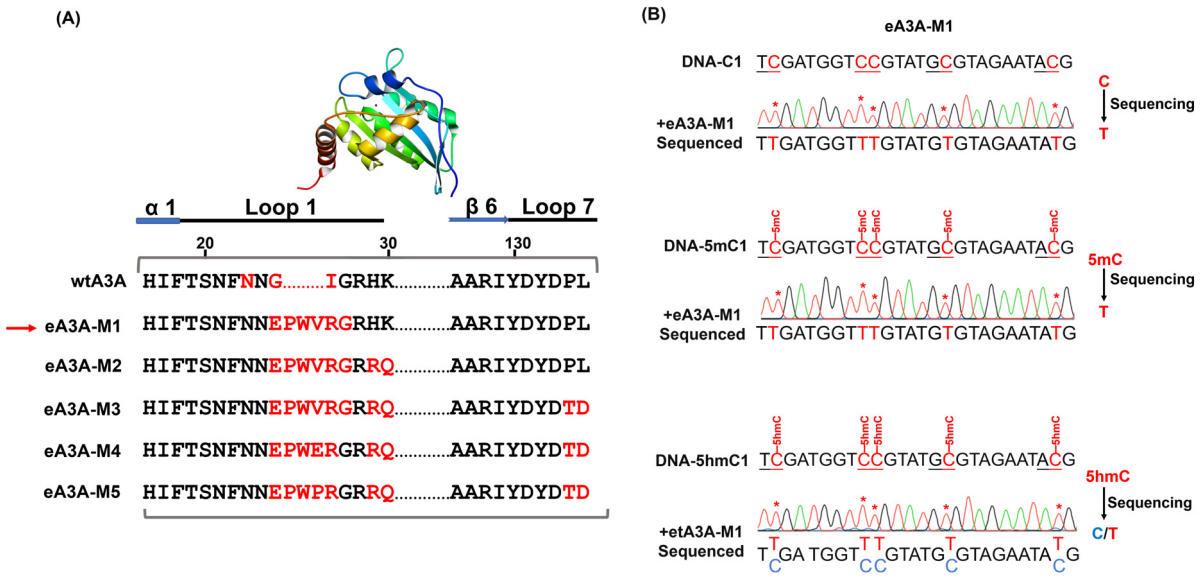

Figure S6. Assessment of eA3A-M2 specificity for C, 5mC, and 5hmC across various sequence contexts. (A) The amino acid compositions of wtA3A and engineered A3A mutants (eA3A-v1 to eA3A-M5). (B) Sanger sequencing results for DNA substrates containing C, 5mC, and 5hmC (DNA-C1, DNA-5mC1, and DNA-5hmC1, respectively) as treated by eA3A-M2. eA3A-M2 converted all the C and 5mC sites to U and T, respectively, which were subsequently read as T during sequencing; 5hmC sites underwent partial deamination, resulting in a mixed read of C and T.

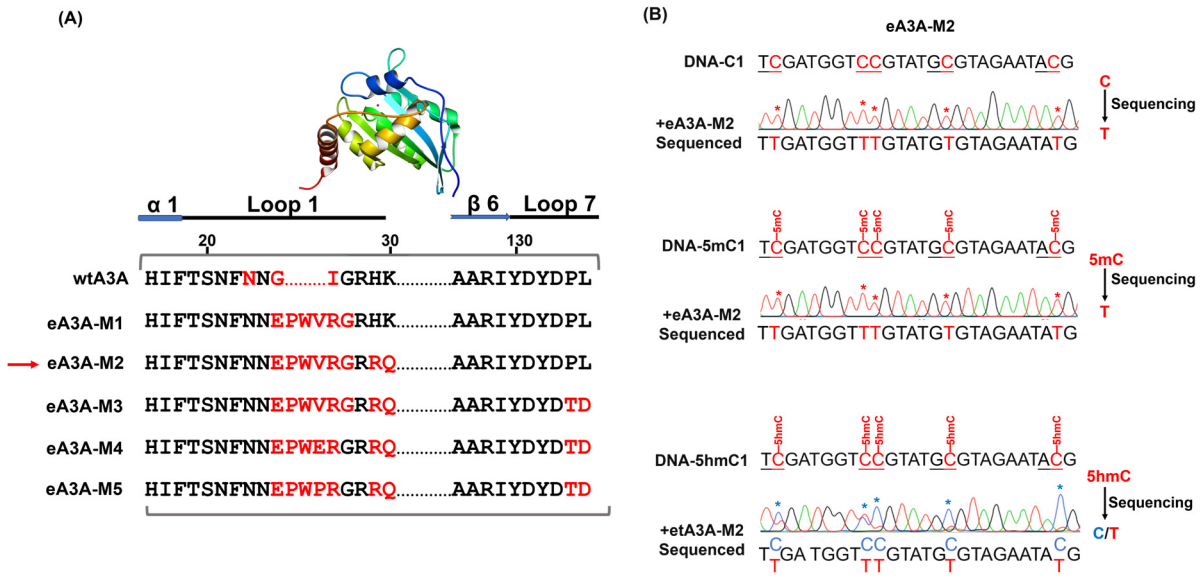

Figure S7. Assessment of eA3A-M3 specificity for C, 5mC, and 5hmC across various sequence contexts. (A) The amino acid compositions of wtA3A and engineered A3A mutants (eA3A-v1 to eA3A-M5). (B) Sanger sequencing results for DNA substrates containing C, 5mC, and 5hmC (DNA-C1, DNA-5mC1, and DNA-5hmC1, respectively) as treated by eA3A-M3. eA3A-M3 converted all the C and 5mC sites to U and T, respectively, which were subsequently read as T during sequencing; 5hmC sites remained unaltered and continued to be read as C.

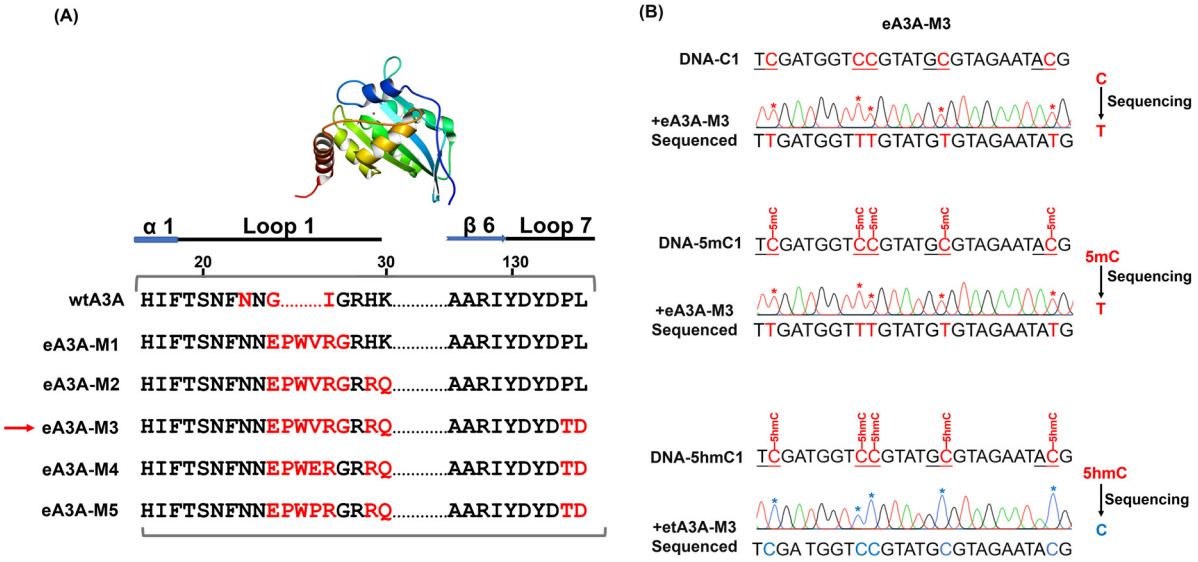

Figure S8. Assessment of eA3A-M4 specificity for C, 5mC, and 5hmC across various sequence contexts. (A) The amino acid compositions of wtA3A and engineered A3A mutants (eA3A-v1 to eA3A-M5). (B) Sanger sequencing results for DNA substrates containing C, 5mC, and 5hmC (DNA-C1, DNA-5mC1, and DNA-5hmC1, respectively) as treated by eA3A-M4. After the treatment of eA3A-M4, C and 5mC sites underwent partial deamination, resulting in a mixed read of C and T; 5hmC sites remained unaltered and continued to be read as C.

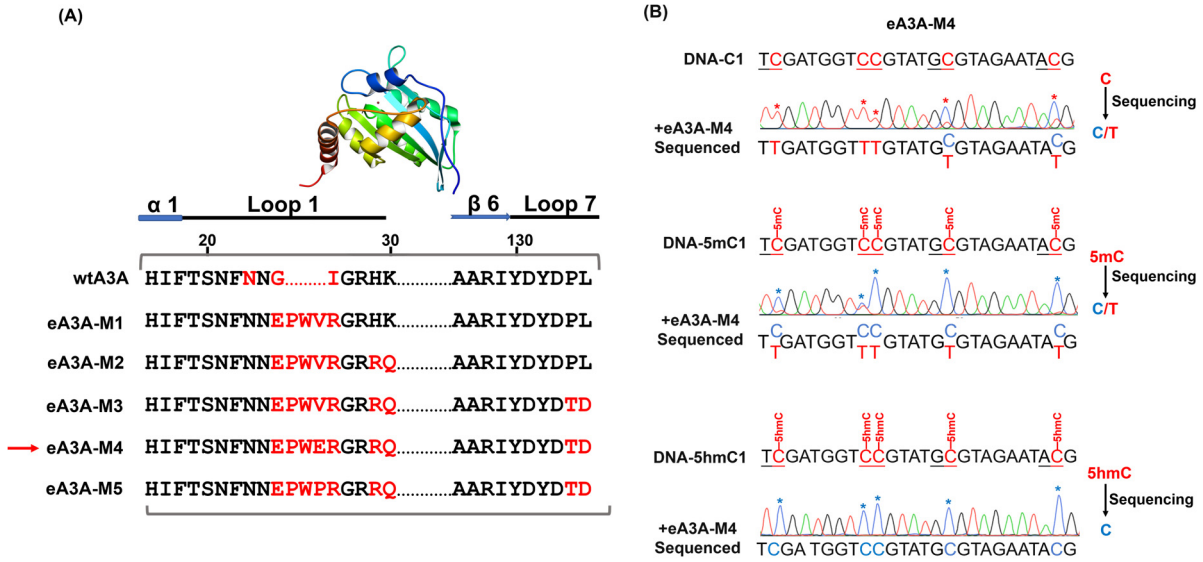

Figure S9. Evaluation of the deamination rates of C and 5mC in the same sequence context from different DNA under eA3A-M5 treatment. (A) DNA-C/5mC and DNA-C/5mC spike-in were mixed up with the content of DNA-C/5mC spike-in being 0.1% in the mixture followed by eA3A-M5 treatment and sequencing. (B) The deamination rate of 5mC from DNA-C/5mC and DNA-C/5mC spike-in treated by eA3A-M5 for different time. (C) The deamination rate of C from DNA-C/5mC and DNA-C/5mC spike-in treated by eA3A-M5 for different time.

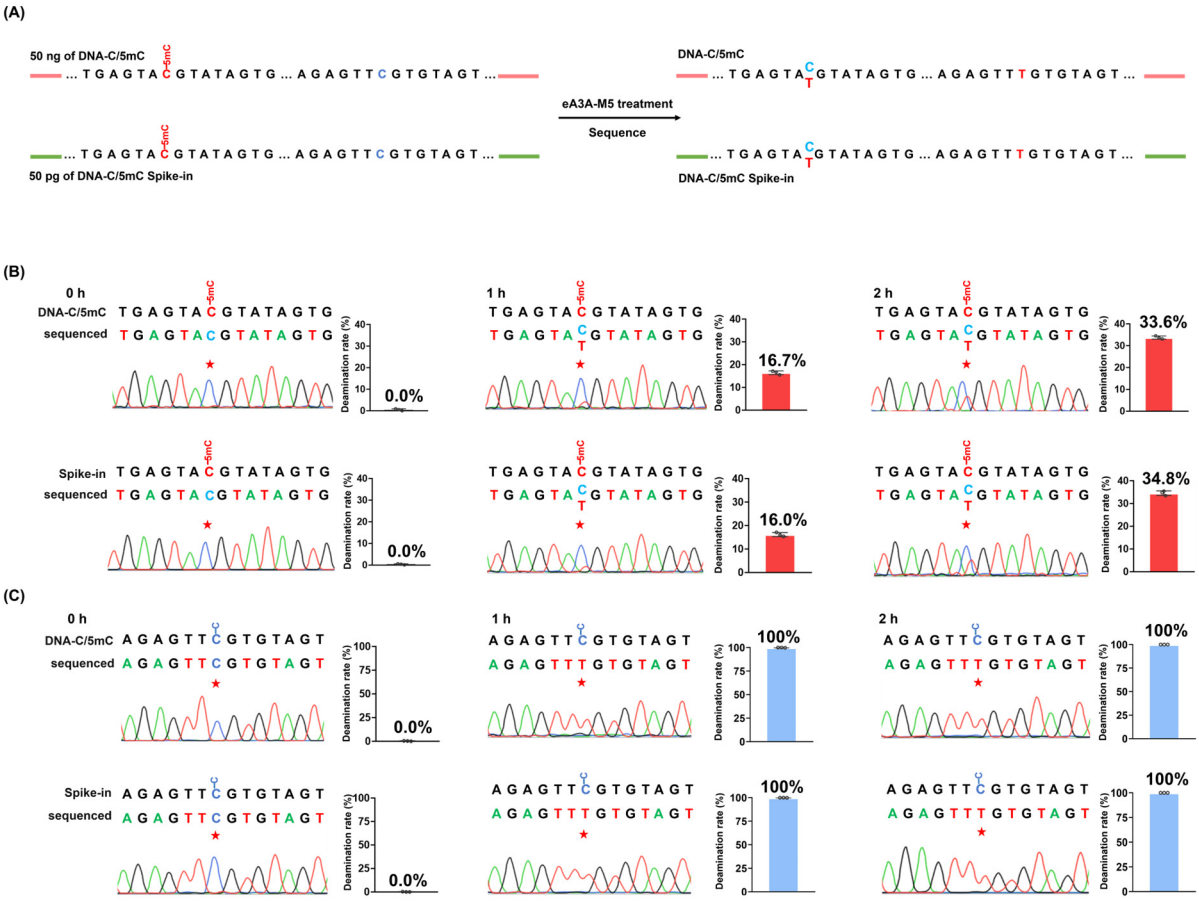

Figure S10. Quantitative assessment of 5mC level at GC sites containing C and 5mC by EDA-seq. (A) DNA-C2 and DNA-5mC2 were mixed at varying ratios, with DNA-5mC2 ranging from 0% to 100%. A 0.1% DNA-5mC2 spike-in was introduced to determine the deamination rate of 5mC. (B) Sanger sequencing results of GC sites in the DNA-C2 and DNA-5mC2 mixture, along with the DNA-5mC2 spike-in.

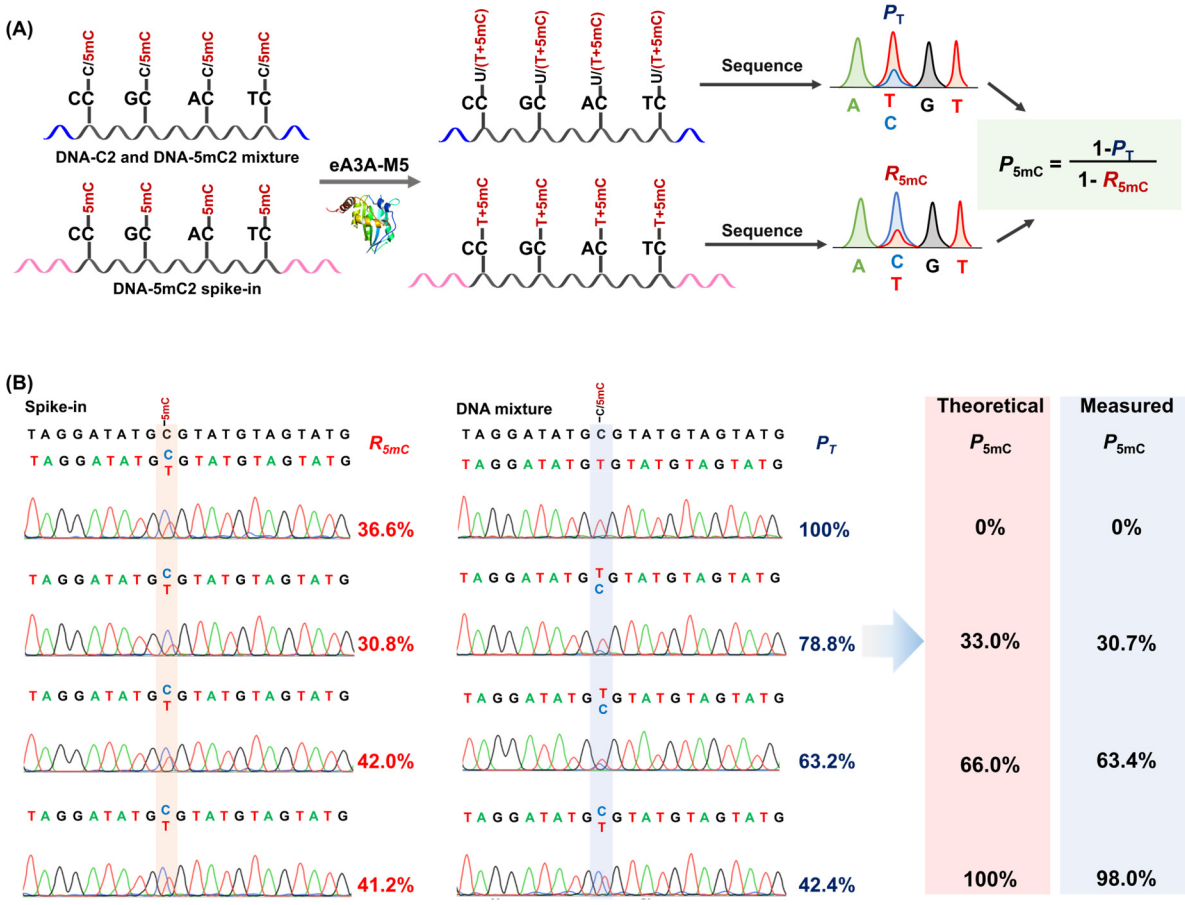

Figure S11. Quantitative assessment of 5mC level at AC sites containing C and 5mC by EDA-seq. (A) DNA-C2 and DNA-5mC2 were mixed at varying ratios, with DNA-5mC2 ranging from 0% to 100%. A 0.1% DNA-5mC2 spike-in was introduced to determine the deamination rate of 5mC. (B) Sanger sequencing results of AC sites in the DNA-C2 and DNA-5mC2 mixture, along with the DNA-5mC2 spike-in.

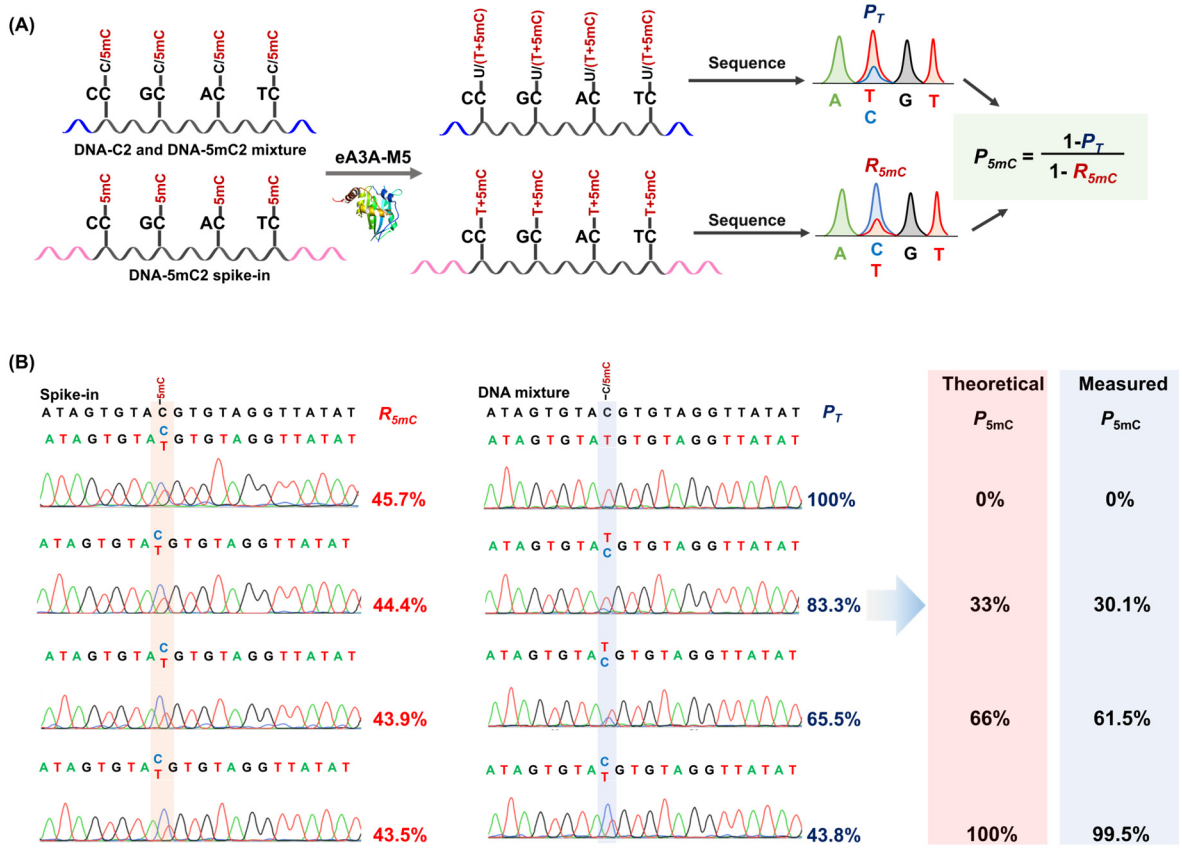

Figure S12. Quantitative assessment of 5mC level at CC sites containing C and 5mC by EDA-seq. (A) DNA-C2 and DNA-5mC2 were mixed at varying ratios, with DNA-5mC2 ranging from 0% to 100%. A 0.1% DNA-5mC2 spike-in was introduced to determine the deamination rate of 5mC. (B) Sanger sequencing results of CC sites in the DNA-C2 and DNA-5mC2 mixture, along with the DNA-5mC2 spike-in.

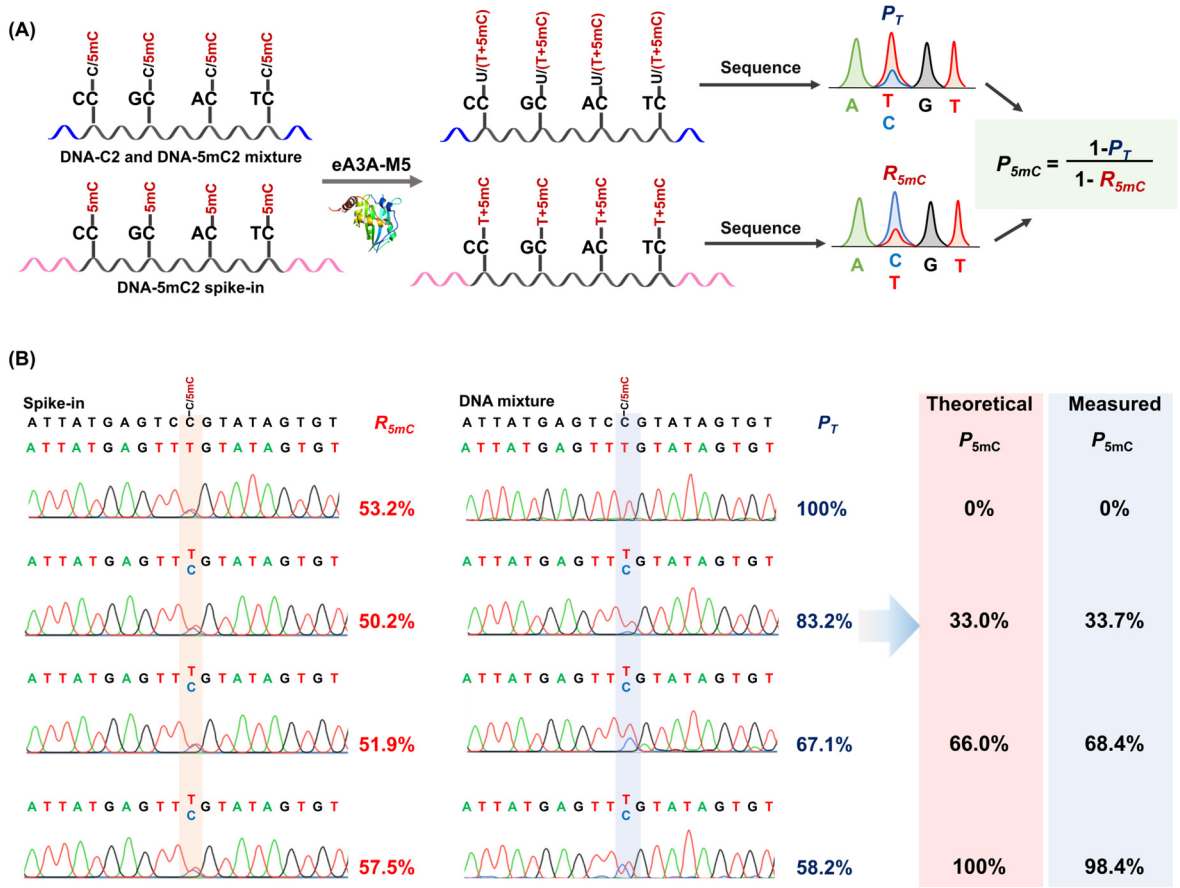

Figure S13. Quantitative assessment of C, 5mC, and 5hmC levels at GC sites containing C, 5mC, and 5hmC. (A) DNA-C2, DNA-5mC2, and DNA-5hmC2 were mixed at proportions of 30%, 30%, and 40%, respectively. A 0.1% DNA-5mC2 spike-in was included in the mixture to determine the deamination rate of 5mC. (B) Sanger sequencing results of GC sites in the DNA-C2, DNA-5mC2, and DNA-5hmC2 mixture, as well as the DNA-5mC2 spike-in, after treatment with eA3A-M5 for varying durations.

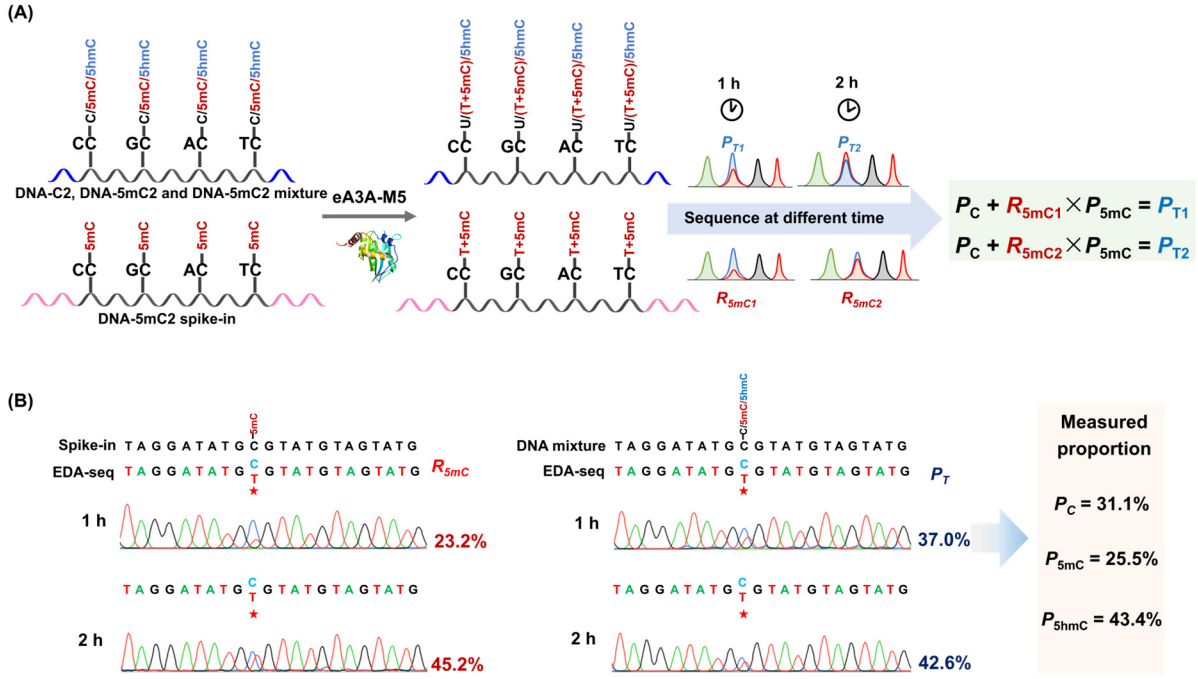

Figure S14. Quantitative assessment of C, 5mC, and 5hmC levels at AC sites containing C, 5mC, and 5hmC. (A) DNA-C2, DNA-5mC2, and DNA-5hmC2 were mixed at proportions of 30%, 30%, and 40%, respectively. A 0.1% DNA-5mC2 spike-in was included in the mixture to determine the deamination rate of 5mC. (B) Sanger sequencing results of AC sites in the DNA-C2, DNA-5mC2, and DNA-5hmC2 mixture, as well as the DNA-5mC2 spike-in, after treatment with eA3A-M5 for varying durations.

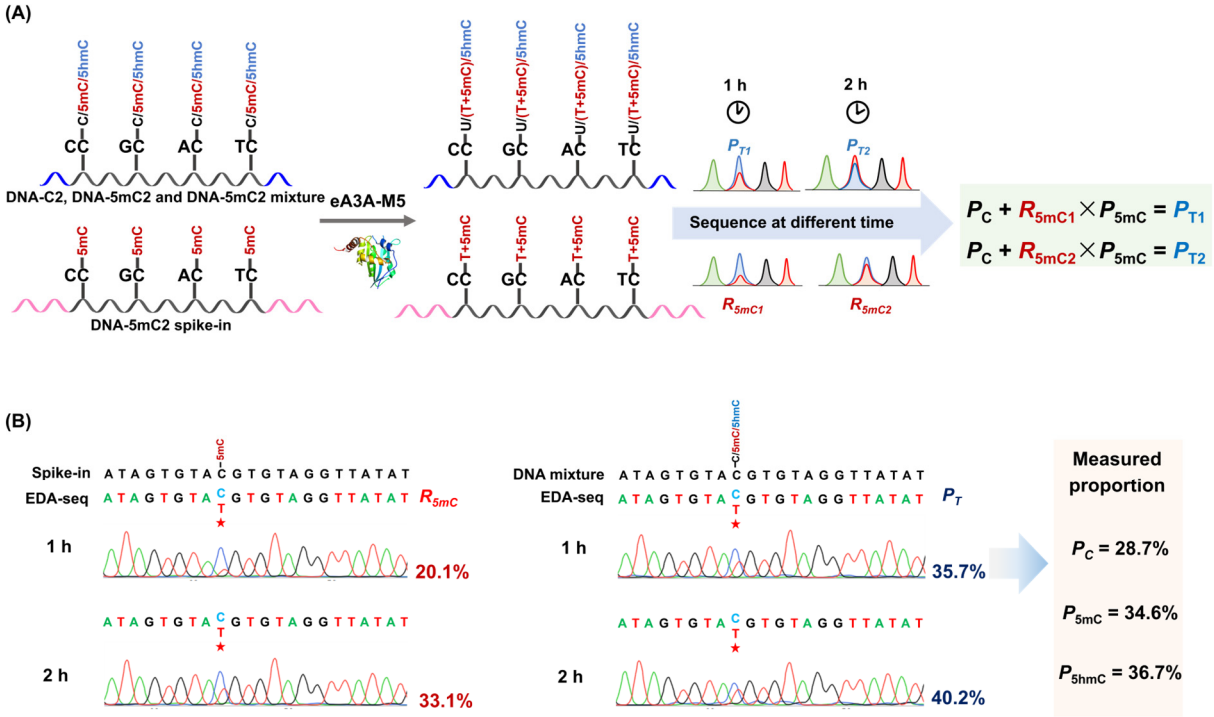

Figure S15. Quantitative assessment of C, 5mC, and 5hmC levels at CC sites containing C, 5mC, and 5hmC. (A) DNA-C2, DNA-5mC2, and DNA-5hmC2 were mixed at proportions of 30%, 30%, and 40%, respectively. A 0.1% DNA-5mC2 spike-in was included in the mixture to determine the deamination rate of 5mC. (B) Sanger sequencing results of CC sites in the DNA-C2, DNA-5mC2, and DNA-5hmC2 mixture, as well as the DNA-5mC2 spike-in, after treatment with eA3A-M5 for varying durations.

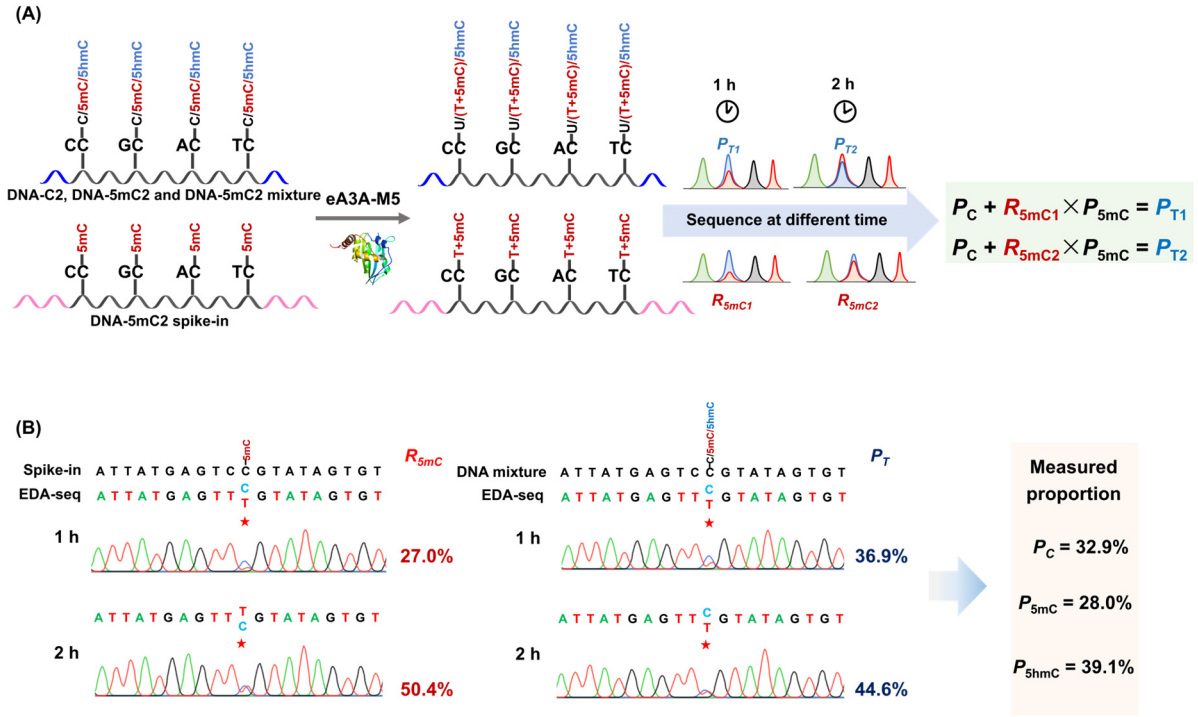

Figure S16. The schematic illustration of BS-seq and ACE-seq. (A) In BS-seq, all the C sites are deaminated and read as T; 5mC and 5hmC are not deaminated and read as C. The sequence results of BS-seq provides the total proportion of 5mC and 5hmC in the individual cytosine sites. (B) In ACE-seq, all the 5hmC sites are first glycosylated by *β*-glucosyltransferase. After wtA3A treatment, all the C and 5mC sites are deaminated and read as T; the glycosylated 5hmC (5gmC) are not deaminated and read as C. The sequence results of ACE-seq provides the proportion of 5hmC in the individual cytosine sites.