†Electronic Supporting Information（ESI）

Highly sensitive quantification of site-specific 5-hydroxymethylcytosine at single-base resolution by HpaII-mediated ligation PCR

Zhenhao Zhang,[a] Xiao Shan,[a] Pengbo Zhang,[b] Weiliang Liu,[a] Jingli Yan,*[a] Zhengping Li*[a]

[a] *Z. Zhang, W. Liu, H. Pian, Prof. Dr. J. Yan, Prof. Dr. Z. Li,*
Key laboratory of Applied Surface and Colloid Chemistry Ministry of Education, Key Laboratory of Analytical Chemistry for Life Science of Shaanxi Province, School of Chemistry and Chemical Engineering, Shaanxi Normal University, Xi'an 710062, Shaanxi Province (P. R. China).
E-mail: lzpbd@snnu.edu.cn
yanjingli@snnu.edu.cn
[b] *Dr. Pengbo Zhang*
School of Chemistry and Biology Engineering, University of Science and Technology Beijing, Beijing, 100083 (P. R. China).

Table of Contents

Experimental Procedures

1. Materials and reagents

HpaII Restriction Endonucleases, Exonuclease I (Exo I) and Exonuclease III (Exo III) were purchased from New England Biolabs (USA). Ampligase (Thermostable DNA ligase) was obtained from Epicentre Biotechnologies (USA). JumpStart™ *Taq* DNA polymerase and Potassium perruthenate ($KRuO_4$) were ordered from Sigma-Aldrich (USA). Dr.GenTLE® precipitation carrier, nuclease-free water and dNTP mixture (2.5 mM each) were purchased from TaKaRa Biotechnology Co., Ltd (Dalian, China). 20× SYBR Green I (20 ng/μL stock solution in DMSO) was brought from Xiamen Bio-Vision Biotechnology (Xiamen, China). EpiTect® Plus DNA Bisulfite kit was purchased from Qiagen (Germany). UNIQ-10 Spin Column Oligo DNA Purification Kit was obtained from Shanghai Sangon Biological Engineering Technology & Services Co., Ltd (Shanghai, China). Mouse brain genomic DNA was purchased from Celprogen Inc (USA). All oligonucleotides sequences (Table S1) were synthesized and purified by TaKaRa Biotechnology Co., Ltd (Dalian, China). All other reagents were analytical reagent grade and used as purchased without further purification.

2. Table S1 the sequences of the nucleic acids used in this work

Name	Sequence
Target C	5′-CCAGGTCCCACAGATCTATCACCCGGGGCTCTTCAAACTCTGCAGG -3′
Target 5mC	5′-CCAGGTCCCACAGATCTATCACC^{5m}CGGGGCTCTTCAAACTCTGCAGG -3′
Target 5hmC	5′-CCAGGTCCCACAGATCTATCACC^{5hm}CGGGGCTCTTCAAACTCTGCAGG -3′
Target 5fC	5′-CCAGGTCCCACAGATCTATCACC^{5f}CGGGGCTCTTCAAACTCTGCAGG -3′
Probe B	5′-GGGATACTGGAACCTGATGATGACTAATACCTGGCTGATAAAATTTAAAAAACCCCA -3′
Probe A	5′-phosphate-AATAATAAATCTATAAAACCTTTACTACTCTCACACACTCTATGCTTGCTACCG*T*C*G-phosphorothioates-3′
Forward primer	5′- GGGATACTGGAACCTGATGATGAC -3′
Reverse primer	5′- CGACGGTAGCAAGCATAGAGTGTG -3′

Note: target C、target 5mC and target 5hmC were synthetic DNA according to Lrpap1 gene in chromosome 5 of Mus musculus strain C57BL/6J. The underlined bases were the detected sites in the sequences. The green part of probe A and probe B was the complementary sequences with targets. The probe A was modified with a phosphate group at its 5′ end. The * labeled red oligonucleotides were modified with phosphorothioates at 3′ end of probe A which could prevent the ligated templates from digestion by exonuclease. Forward primer and reverse primer were used for the PCR reaction.

3. Cleavage with HpaII restriction endonucleases

16 μL (10 μM) each of single-stranded target (containing target C, target 5mC or target 5hmC, respectively) and its complementary DNA were respectively mixed in 4 μL of CutSmart® Buffer (50 mM KAc, 20 mM Tris-Ac, 10 mM Mg $(Ac)_2$, 100 μg/ml BSA, pH 7.9 @ 25°C) and incubated for 5 min at 95 °C followed by incubating for 30 min at room temperature to anneal and yield dsDNA targets. The dsDNA products or genomic DNA samples were then supplemented with 80 U of HpaII to cleave the dsDNA target within the HpaII site. The reaction was carried out at 37 °C for 12 h followed by inactivation of the enzyme at 80 °C for 20 min. A 2720 thermal cycler (Bio-Rad, CA, USA) was employed to control the reaction temperature. Subsequently, the products were precipitated with ethanol and Dr.GenTLE® Precipitation carrier in a microfuge tube and then were centrifuged at 12000 rpm for 30 min. Finally, the supernatant fluids were removed and the precipitations were dissolved with 44 μL of nuclease-free water.

4. Potassium perruthenate (KRuO4) oxidation and bisulfite treatment

The above reaction products were respectively supplemented with 32 μL of NaOH (0.12 M) solution and incubated at 37 °C for 30 min for denaturation. The products were rapidly cooled on ice for 5 min and then 4 μL of $KRuO_4$ (15 mM in 0.05 M NaOH) solution was added. The reaction was held on ice for 1 hour with three times vortexing. Subsequently, each of reaction product was supplemented with 16 μL of NaOH (0.05 M) to re-denature for 30 min at 37 °C and was cooled again on ice. Then 4 μL of $KRuO_4$ (15 mM in 0.05 M NaOH) was added in and the reoxidation was carried out in the same oxidation process as described above. The method of double oxidation was applied to synthetic dsDNAs to reach a higher oxidation efficiency. Finally, the products were purified with UNIQ-10 Spin Column Oligo DNA Purification Kit according to the manufacturer's instructions. The purified products were subjected to bisulfate conversion using the EpiTect® Plus DNA Bisulfite kit according to the manufacturer's instructions.

5. Ligation reaction and enzyme digestion reaction

The mixture (9 μL) of ligation reaction consists of 2 nM Probe A, 2 nM Probe B, 1 U of Ampligase, Ampligase buffer (20 mM Tris-HCl, 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.1% Triton X-100, pH 8.3). The 1 μL of appropriate amounts of treated synthetic dsDNAs or genomic DNA samples were added to the mixture. This solution was heated at 94 °C for 3 min, followed by incubating at 52 °C for 20 min to ligate probe A and probe B relying on treated target 5hmC. After the ligation reaction, the products were immediately put on ice. Subsequently, the reaction mixture was digested in the presence of Exo I (20 U) and Exo III (60 U) in a final volume of 11.6 μL. The reaction of enzyme digestion was carried out through the following program: 37 °C for 4 h and followed by inactivation of Exo I and Exo III at 85 °C for 20 min.

6. PCR amplification

The PCR reaction mixture contained 0.5 U of JumpStart™ *Taq* DNA polymerase, reaction buffer (20 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, pH 8.3), 200 nM forward primer, 200 nM reverse primer, 200 μM dNTPs, 0.4× SYBR Green I and nuclease-free water. 2 μL of the ligation products were added to PCR reaction mixture to give the final volume of 10 μL. The mixture was mixed thoroughly and subjected to PCR reaction according to the following thermal cycle: 94 °C for 2 min followed by 50 cycles of 94 °C for 20 s and 60 °C for 30 s. PCR reaction was carried out in StepOne Real-Time PCR System (Applied Biosystems, USA), in which the real-time fluorescence intensity was monitored.

7. Influence of the amount of the HpaII restriction endonuclease

The methylation-sensitive restriction endonuclease (MSRE) HpaII that could specifically recognize the sequence CCGG and cleave the dsDNA containing unmethylated C was used to discriminate C from 5mC and 5hmC. The target C of incomplete digestion in the digestion reaction would interfere the detection of target 5hmC. To investigate the amount of enzyme for completely digesting target C, each of 4 μM dsDNA of target C, target 5mC or target 5hmC was respectively subjected to the digestion reaction with HpaII at different concentrations. The digestion products were separated by the non-denaturing polyacrylamide gel electrophoresis (PAGE).

A 16% PAGE was used to separate the digestion products in Tris-Borate-EDTA (TBE) electrophoresis buffer (9mM Tris-HCl, 9mM boric acid, 0.2mM EDTA, pH 7.9) at a 110 V constant voltage for 80 min at room temperature. The gel was stained by 4s Red plus Nucleic acid stain and imaged with a Gel Doc Ez imager (BioRad, USA).

As shown in Fig. S1a, the target 5mC and target 5hmc had not been digested at all (lane 1, 2 and lane 4, 5). Meanwhile, the target C had been cleaved to form short strands, showing the high specificity of HpaII digestion. In addition, one can see from lane 6 and lane 3 that 40 U HpaII was insufficient to completely digest 4 μM of target C, but the complete digestion could be observed with 80 U HpaII. Therefore, 80 U HpaII was employed in the experiment of the ligation PCR assay.

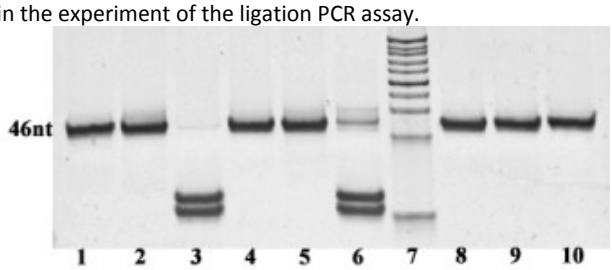

Fig. S1 Influence of the amount of HpaII restriction endonuclease on the digestion reaction. Lane 7: DNA markers (20 bp DNA ladder); Lane 1, 2, 3: digestion reaction with 80 U of HpaII restriction endonuclease; lane 4, 5, 6: digestion reaction with 40 U of HpaII restriction endonuclease; lane 8, 9, 10: digestion reaction with 0 U of HpaII restriction endonuclease. Lane 1, 4, 8: double strand target 5hmC. Lane 2, 5, 9: double strand target DNA 5mC. Lane 3, 6, 10: double strand target C.

8. Optimization of the amount of Ampligase

The ligation reaction relied on the catalysis of the Ampligase with high DNA-dependent ligation activity. In this assay, the ligation efficiency would be affected by the amount of Ampligase. The effect of amount of Ampligase ranging from 0.5 U to 2 U on the proposed ligation PCR assay were investigated.

As shown in Fig. S2a and S2b, when the amount of Ampligase were increased from 0.5 U to 1 U, the PCR amplification signals produced by target C, target 5mC and blank were always straight lines that meant the PCR amplification signals could not be detected within the reaction time of 50 cycles, but the PCR amplification reaction produced by target 5hmC was gradually accelerated and the C_T value produced by target 5hmC was reduced with the increase of Ampligase dosage. Based on this evidence, the more amount of Ampligase would achieve higher ligation efficiency templated by target 5hmC. When the amount of Ampligase was 2 U (Fig. S2c) the PCR amplification reaction was further accelerated. Nonetheless, target C and target 5mC showed obvious PCR amplification signals. These results indicated that the nonspecific ligation had already existed, which had disturbed the detection of target 5hmC when the amount of Ampligase was increased to 2 U. In conclusion, 1 U was enough to complete the ligation reaction and could meet the specific requirement of detection. In this regard, the amount of 1 U Ampligase was chosen as the optimal amount in ligation PCR assay.

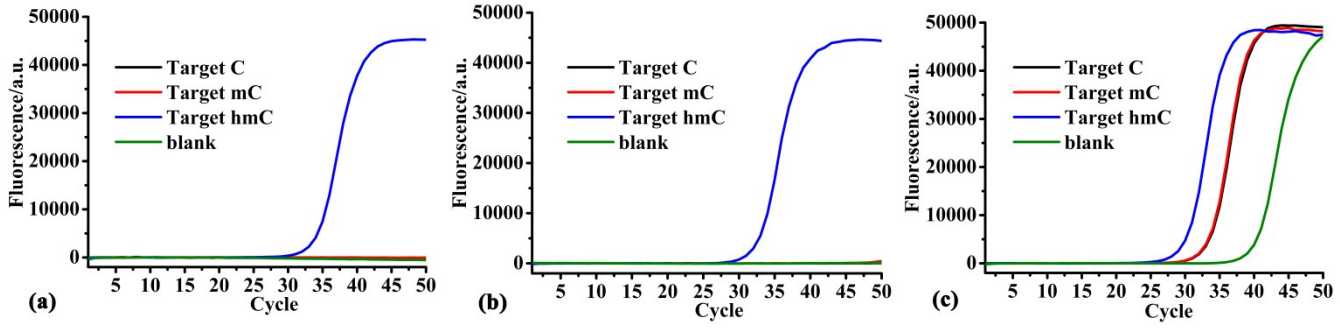

Fig. S2 Optimization of the amount of Ampligase in the ligation PCR assay. The real-time fluorescence curves produced by 35 fM of target C, target 5mC and target 5hmC were recorded respectively under the catalysis of (a) 0.5 U, (b) 1 U and (c) 2 U Ampligase. Other experimental conditions were the same as described in the experimental procedure except the amount of Ampligase.

9. Optimization of the concentration of DNA probes

The non-specific amplification was an inevitable problem for PCR amplification and it was the important reason to limit the sensitivity and specificity for the detection of targets. The appropriate concentrations of ligation probes would make a contribution to reduce the generation of non-specific amplification. In the other hand, the concentrations of probe A and probe B were also a critical factor for the ligation efficiency. So the amount of probe A and probe B (kept the same concentration in this study) were optimized with the different concentrations in the range of 400 pM to 10 nM.

As shown in Fig. S3a and S3b, when the concentration of the probes were increased from 400 pM to 2 nM, only the target 5hmC produced the real-time PCR fluorescence signals and the reaction produced by target 5hmC was gradually accelerated with the increase of probes dosage, while the PCR amplification signals produced by target C, target 5mC could not be observed within the reaction time. These results demonstrated that when the concentration of probes was lower than 2 nM, target C and 5mC did not interfere with the analysis of target 5hmC, but the greater concentration of probes could result in higher ligation efficiency at the low levels of the target 5hmC, which could lead to higher sensitivity for detection of target 5hmC. When the concentration of probes continuously increased to 10 nM (Fig. S3c), the target C, target 5mC and blank produced the non-specific amplification signals. The reason might be that the 10 nM probes were too excessive to detect the targets of low concentration, which leaded to the non-specific ligation of the probe A and probe B independent on the template. Thus 2 nM probes could make a remarkable distinction of the real-time fluorescence signals among target C, target 5mC, target 5hmC and blank and achieved higher sensitivity. Hence, 2 nM of probe A and probe B was employed for the assay.

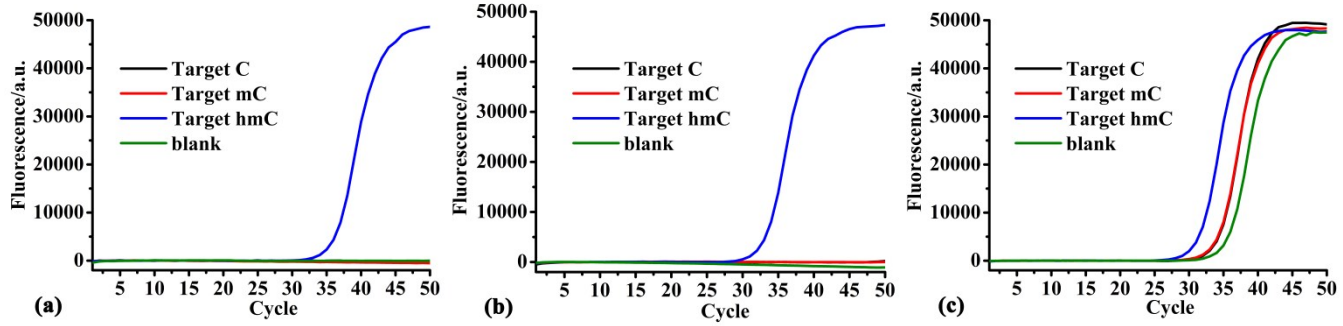

Fig. S3 Affection of the concentration of probes on the ligation PCR assay. The real-time fluorescence curves produced by 35 fM of target C, target 5mC and target 5hmC were recorded respectively under the different probe concentrations of (a) 400 pM, (b) 2 nM and (c) 10 nM. Other experimental conditions were the same as described in the experimental procedure except the concentration of probe A and probe B.

10. Optimization of the temperature of ligation reaction.

The sequences of the treated targets 5hmC differs from the target 5mC by only one base. The temperature of ligation reaction played an important role to differentiate target 5hmC from target 5mC. Therefore, the investigation of the influence of the ligation temperature on this assay was carried out. The ligation reaction was performed at the temperature of 48 °C, 52 °C and 56 °C, respectively.

As shown in Fig. S4, when the ligation reaction was performed at 48 °C (Fig. S4a), although the target 5hmC can be distinguished from target C and 5mC, both target C and 5mC can product obvious non-specific amplification signals, which can interfere with the detection of target 5hmC. When the ligation temperature was elevated to 52 °C (Fig. S4b), the PCR amplification signals produced by target C, target 5mC and blank were always straight lines that meant the non-specific amplification signals were eliminated. When the temperature continued to rise to 56 °C, there was still no non-specific amplification produced by target C, target 5mC and blank, but the C_T value of target 5hmC increased. On the other hand, the experimental results indicated that as the temperature increases from 48 °C to 56 °C, the C_T value increases gradually, but the efficiency of ligation gradually decreases. When the probe B were fully complementary to the target 5hmC, Tm was 53.5 °C. When the probe B were one-base mismatched complementary to the target C and 5mC, Tm was 51.5 °C. The 48 °C is lower than the 51.5 °C of Tm which could lead to the non-specificity ligation and result in non-specific PCR amplification produced by target C and 5mC. The 56 °C is higher than the 53.5 °C of Tm resulting in low ligation efficiency. Therefore, considering ligation efficiency and nonspecific ligation, 52 °C was selected as the ligation temperature in this assay.

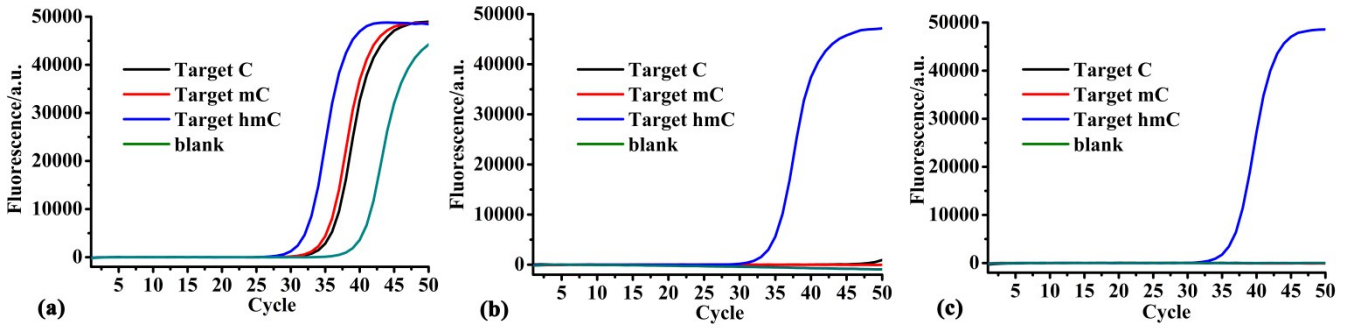

Fig. S4 The influence of ligation temperature on ligation-based PCR assay. Real-time fluorescence signals produced by 35 fM of target C， target 5mC and target 5hmC were recorded with different ligation temperatures of (a) 48 °C, (b) 52 °C, (c) 56 °C. Other experimental conditions were the same as described in the experimental procedure except the temperature of ligation reaction.

11. The detection of 5hmC in mouse brain genomic DNA

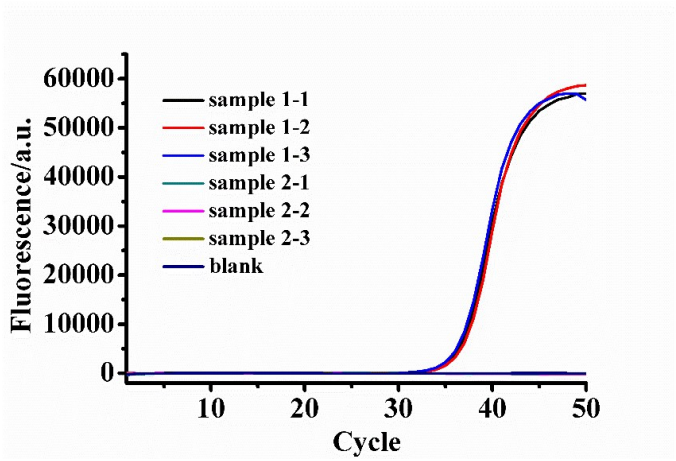

Fig. S5 Determination of 5hmC in genomic DNA from Lrpap1 gene of mouse brain. Real-time fluorescence curves were produced by 140 ng genomic DNA of mouse brain with oxidation and without oxidation. The sample 1 was mouse brain genomic DNA with oxidation step. The sample 2 was mouse brain genomic DNA without oxidation step. Three parallel experiments were carried out on both samples respectively, denoted by" -1", "-2" and "-3". The fluorescence curves produced by the three sample 2 without oxidation step could not be detected and their traces are coincident with the blank. The blank was detected with the same procedures but without genomic DNA sample. The measurand was according to the experimental protocol described above.

12. Table S2: Relevant MSREs that recognizes different XCGX sequence

The relevant MSREs that recognize different XCGX sequence have been listed in Table S2. The methylation sensitivity of these MSRE is cited. The recognition sequences of these enzymes contain six kinds of XCGX sequence which are ACGT, TCGC, CCGC, CCGG, GCGC, GCGG, respectively. "/" in sequence is the digestive site for MSRE. By comparing the properties of these enzymes and HpaII, we believe that these enzymes excepted Msp I may be applicable to this method.

Table S2: Relevant MSREs that recognizes different XCGX sequence

Name	Recognition sequence (5'-3')	Dcm	CpG Methylation
HpaII	C/CGG	Not sensitive	blocked
Msp I	C/CGG	Not sensitive	Not sensitive
AciI	C/CGC	Not sensitive	blocked
HinP1I	GC/GC	Not sensitive	blocked
FspI	TGC/GCA	Not sensitive	blocked
HhaI	GCG/C	Not sensitive	blocked
HpyCH4IV	A/CGT	Not sensitive	blocked
NotI	GC/GGCCGC	Not sensitive	blocked
PmlI	CAC/GTG	Not sensitive	blocked
SacII	CCGC/GG	Not sensitive	blocked
AclI	AA/CGTT	Not sensitive	blocked
AsiSI	GCGAT/CGC	Not sensitive	blocked
SnaBI	TAC/GTA	Not sensitive	blocked
ZraI	GAC/GTC	Not sensitive	blocked