# Supporting Information

## One-pot intramolecular cyclization of 5-hydroxymethylcytosine for sequencing DNA hydroxymethylation at single-base resolution

Yangde Ma[a,b], Ning Zhang[b], Shaokun Chen[a,b], Jingwei Sun[a], Yan Liu[a,b], Xiangjun Li [a,*], Hailin Wang [a,b,c]

a. University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing, 100049, P.R.China.

b. State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, No. 18 Shuangqing Road, Beijing, 100085, P.R.China.

c. Institute of Environment and Health, Jianghan University, No. 8, Sanjiaohu Road, Wuhan, Hubei, 430056, P.R.China.

*To whom correspondence should be addressed. E-mail: lixiangj@ucas.ac.cn.

### Contents

## Experimental Section

### Chemicals and reagents

All of 5hmC-ODN were synthesized at Takara Biotechnology (Dalian, China) and other primers were purchased from Sangon Biotech (Shanghai, China). Organic chemical reagents, such as acetophenone, 3-oxobutyronitrile, phenylacetylene, benzoylacetonitrile, 2-acetyfuran and dimethyl sulfoxide (DMSO), 1,4-dioxane, 2-methoxyethanol were from Innochem (Beijing, China). Ammonium bicarbonate (for mass spectrometry, ~99%) and formic acid (for mass spectrometry, ~98%) used for UHPLC-MS/MS were purchased from Fluka Analytical (Buchs, Switzerland). 5-(Hydroxymethyl)-2'-deoxycytidine (5hmC), 5-(formyl)-2'-deoxycytidine (5fC), 2'-deoxycytidine (dC), LC-MS grade methanol (Fluka, Shanghai, China), were ordered from Sigma-Aldrich (St. Louis, USA). [$D_3$]-5-(hydroxymethyl)-2'-deoxycytidine ([$D_3$]-5hmC) and [$^{15}N_3$] -2'-deoxycytidine ([$^{15}N_3$]-dC) were purchased from Toronto Research Chemicals (Toronto, Canada). $KRuO_4$ was obtained from Alfa Aesar (Beijing, China). Amicon Ultra-0.5 mL centrifugal filters (M.W. cutoff: 3 kDa and 10 kDa ) were bought from Merck Millipore(Billerica, MA, U.S.A.). Snake venom phosphodiesterase (SVP), and calf intestine alkaline phosphatase (CIP) were purchased from Sigma-Aldrich (St. Louis, USA). Deoxyribonuclease I (DNase I) was purchased as lyophilized powder from Sangon Biotech (Shanghai, China).

### General chemical labeling of 5hmC

Typically, coupling agent (8 μmol), t-BuOK (1 mg, 0.01 mmol), benzophenone (1 mg, 5.5 μmol) and 1,4-dioxane (200 μL) were all added to a centrifuge tube (1.5 mL). After being well homogenized, the contents were transferred into a new centrifuge tube (1.5 mL) containing 1 μg of lyophilized DNA (76mer 5hmC-ODN, 11mer 5hmC-ODN or 5hmC nucleoside). The reaction was performed at 90 °C for 2 h (0.3 h or 6 h, depending on the experimental design) under continuous agitation. The DNA product was separated by mixing with 600 μL of deionized water and subsequent centrifugation. The supernatant was purified at least three times by centrifugal filters (M.W. cutoff: 10 kDa ) to eliminate the organic solvent and inorganic salts.

**Evaluating conversion rate of 5xC-to-T (5xC= 5mC, 5hmC, 5fC)**

By using 5hmC as an example, first, the labelled 5hmC-ODN was used as model DNA in the 50 μL PCR. The reaction mixture included the use of Model-F and Model-R as primers, dNTP mixture, and polymerase (Taq DNA polymerase or Q5 High-Fidelity DNA polymerase). After purification with the PCR Clean-Up System, the obtained DNA products were directly ligated with T4 vector by using the pClone007 Blunt Simple Vector Kit (TSINGKE). Furthermore, we sequenced a certain number of monoclonal E.coli and calculated the C-to-T conversion rate by counting the proportion of T-clone (the monoclonal *E.coli* that has a mutation of 5hmC to T) among the results.

**Agarose gel electrophoresis**

To supervise the durability of the reaction, the obtained DNA samples were loaded into a 2% agarose gel containing 0.1 μg/mL Gel-red for electrophoresis. This electrophoresis run at 160 V/cm for 15 min in 1×TAE buffer. After electrophoresis, the gel was visualized and photographed on a G:BOX gel documentation system.

**DNA digestion and UHPLC-MS/MS analysis of dC, 5mC, 5hmC, 5fC**

Prior to LC-MS/MS analysis, the DNA samples (5 μg) were digested into 2'-deoxynucleosides by three enzymes, namely, DNase I (1.0 U), snake venom phosphodiesterase (SVP, 0.02 U), and calf intestine alkaline phosphatase (CIP, 5.0 U). After incubated at 37 °C for 6 h, the mixture was purified by ultrafiltration device. Subsequently, the device was centrifuged at 12000 r/min for 20 min at 4 °C. Finally, the filtrate was transferred into sample vial together with [$^{15}N_3$]-dC (final concentration: 100 nM) and [$D_3$]-5hmC (final concentration: 100 pM) that severed as stable isotope for calibrating internal standards.

UHPLC-MS/MS analysis of 5mC and dC was performed on an Agilent 1290 UHPLC System coupled with the triple quadrupole mass spectrometer (6470B, Agilent). A reversed-phase Agilent Zorbax Eclipse Plus C18 column (2.1 × 50 mm I.D., 1.8 μm particle size) was employed for the separation of monodeoxynucleosides. In detail, dC was separated with the mobile phase containing 5.0 % component A (0.1 % formic) and 95.0 % component B (pure methonal) at a flow rate of 0.25 mL/min. A

multiple reaction monitoring (MRM) mode with a jet stream electrospray ionization source (Santa Clara, CA, U.S.A.) was adopted to quantify the tested deoxynucleosides: m/z 242→126 for 5mC, m/z 228→112 for dC, and m/z 231→115 for [$^{15}N_3$]-dC. The mass spectrometer detection adopted positive ion mode. Capillary voltage was set at +3500 V. Fragmentation voltage was set at 90 V and the collision energy was 5 eV. All samples were analyzed for three times with an injection volume of 5.0 μL.

Similarly, UHPLC-MS/MS analysis of 5hmC and 5fC was performed on an Agilent 1290 UHPLC System coupled with the triple quadrupole mass spectrometer (6470B, Agilent). Nevertheless, the mobile phase contained component A (2 mM $NH_4HCO_3$) and component B (pure methonal) at a flow rate of 0.25 mL/min, a gradient elution was used: 0−3.0 min, 5.0% B; 3.0−5.0 min, 20.0% B; 5.0−8.0 min, 5.0% B. The mass spectrometer detection adopted positive ion mode. A multiple reaction monitoring (MRM) mode was adopted to quantify the tested deoxynucleosides: m/z 256→140 for 5fC; m/z 258→142 for 5hmC; m/z 261→145 for [$^{15}N_3$]-5hmC. Capillary voltage was set at +3500 V. Fragmentation voltage was set at 90 V and the collision energy was 5 eV. All samples were analyzed three times with an injection volume of 5.0 μL.

**MALDI-TOF-MS analysis**

Initially, the matrix was obtained by mixing nicotinic acid (NA), anthranilic acid (AA) and diammonium hydrogen citrate (DHC) in a molar ratio of 2:1:0.006. Then samples were pretreated by mixing with the matrix (1:1, v/v) and deposited on a 384 Anchor-Chip MALDI target plate (Bruker Daltonics, Billerica, MA, USA) for drying and crystallization.

MALDI-TOF-MS analysis was performed on a Bruker Autoflex III Smartbeam mass spectrometer (Bruker Daltonics, Bremen, Germany). To get high sensitivity and best mass resolution, the apparatus operates in the conditions: a mass range of m/z 100 to 4000 in the linear positive ion mode with broadband, 90%−100% laser energy, $2\times10^{-6}$ mbar vacuum, 200 Hz trigger frequency and 200 s extraction delay time was set. Meanwhile, the mass spectra were recorded from an accumulation of 20 laser scans and each scan was accumulated from 500 laser shots. The m/z range of analyte was calibrated with the synthetic DNA of known molecular weight.

## Supplementary Figures and Tables

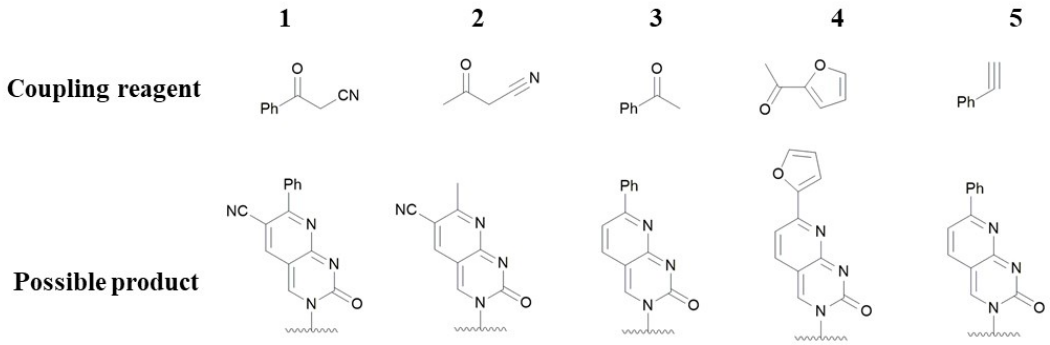


**Figure S1.** Five poteintial ketones were screened for 5hmC labeling reaction, with the possible coupling product. **1**, benzoylacetonitrile. **2**, 3-oxobutyronitrile. **3**, acetophenone(Ace). **4**, 2-acetylfuran. **5**, phenylacetylene(Phe).

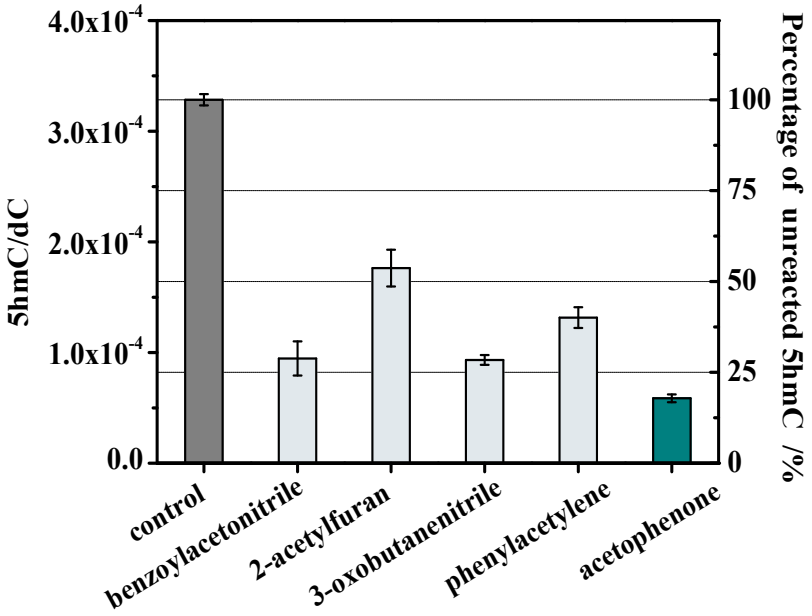


**Figure S2.** The variation of 5hmC/dC in mES genomic DNA before and after coupling reaction under general chemical labeling reaction.

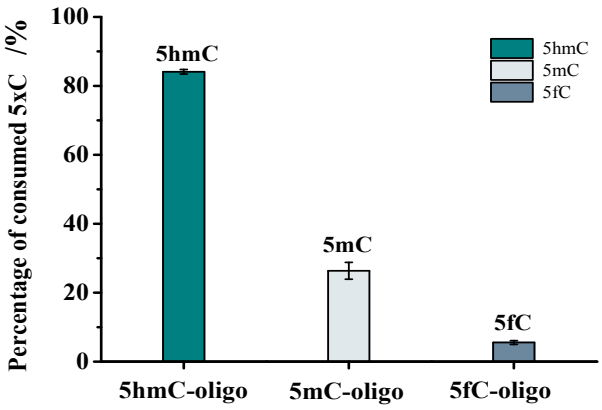


**Figure S3.** The percentage of consumed 5xC in 5xC-ODN (5xC=5mC, 5hmC or 5fC) after coupled with acetophenone for 2 h.

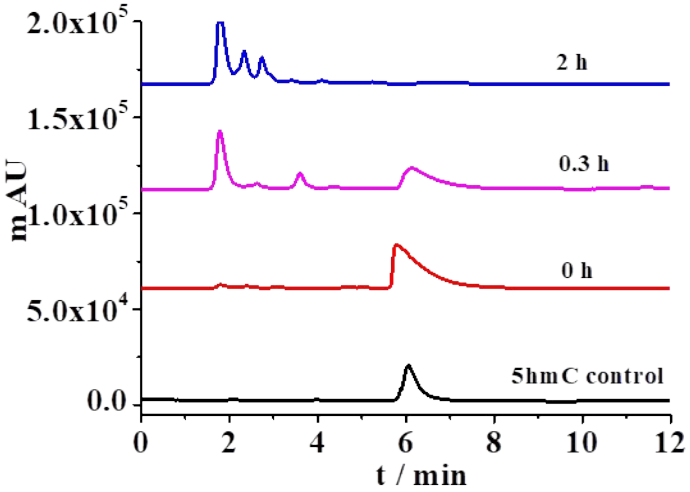


**Figure S4.** The liquid chromatography analysis of 5hmC nucleoside before and after coupled with acetophenone for 2 h.

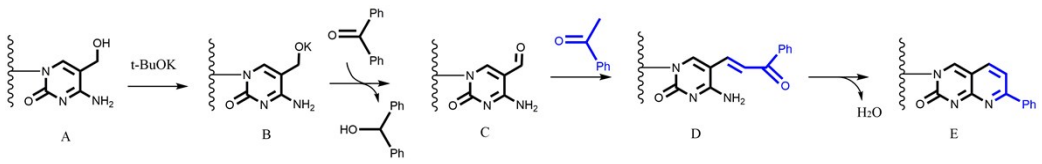


**Figure S5.** A possible coupling reaction pathway between 5hmC and acetophenone.

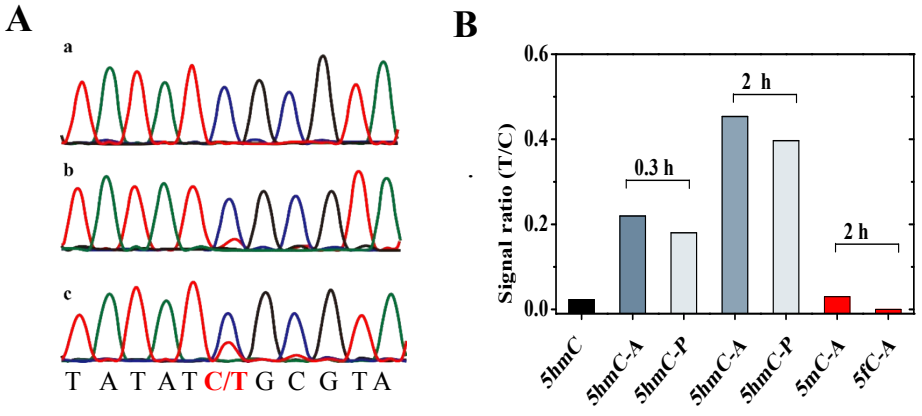


**Figure S6.** (A) Direct Sanger sequencing analysis of 76mer 5hmC-ODN before and after coupled with acetophenone, a, b, c respectively correspond to reaction time 0 h, 0.2 h and 2 h. (B) The T/C signal ratio of PCR-amplified samples derived from 76mer 5xC-ODN (5xC=5mC, 5hmC or 5fC) that treated with acetophenone (Ace) and phenylacetylene (Phe) in different reaction time.

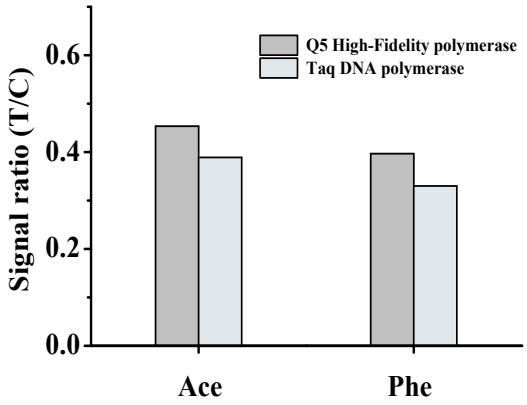


**Figure S7.** The T/C signal ratio of PCR-amplified samples derived from 76mer 5hmC-ODN that treated with acetophenone (Ace) and phenylacetylene (Phe) using Taq DNA polymerase and Q5 High-Fidelity polymerase.

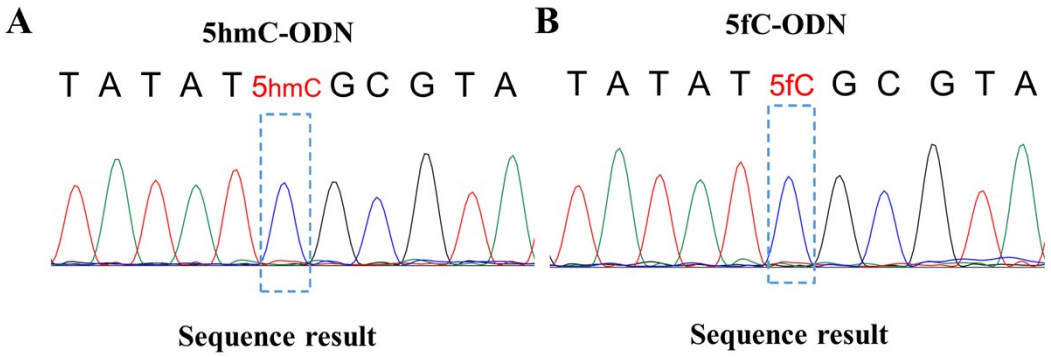


**Figure S8**. Direct Sanger sequencing analysis of 76mer 5mC-ODN (A) and 5fC-ODN (B) after labelled with acetophenone.

**Table S1.** All DNA sequence mentioned in the work

| ODN name | Sequence (5'-3') |
|---|---|
| **76mer 5hmC-ODN** | CCTCACCATCTCAACCAATATTATATTACGCGTATAT/**5hmC**/GCGTATTTCGCGTTATAATATTGAGGGAGAAGTGGTGA |
| **76mer 5mC-ODN** | CCTCACCATCTCAACCAATATTATATTACGCGTATAT/**5mC**/GCGTATTTCGCGTTATAATATTGAGGGAGAAGTGGTGA |
| **76mer 5fC-ODN** | CCTCACCATCTCAACCAATATTATATTACGCGTATAT/**5fC**/GCGTATTTCGCGTTATAATATTGAGGGAGAAGTGGTGA |
| **T-ODN** | CCTCACCATCTCAACCAATATTATATTACGCGTATAT/**T**/GCGTATTTCGCGTTATAATATTGAGGGAGAAGTGGTGA |
| **C-ODN** | CCTCACCATCTCAACCAATATTATATTACGCGTATAT/**C**/GCGTATTTCGCGTTATAATATTGAGGGAGAAGTGGTGA |
| **11mer 5hmC-ODN** | GAGT/**5hmC**/TGGAGA |
| **Model-F primer** | CCTCACCATCTCAACCAAT |
| **Model-R primer** | CTCCGACATTATCACTACCATCAACCACCCATCCTACCTGGACTACATTCTTATTCAGTATTCACCACTTCTCCCTCAAT |
| **76-seq** | CTCCGACATTATCACTACCA |

**Table S2**. Screening of optimized reaction conditions[a]

| Entry | coupling agent | temperature /°C | solvent | time/h | Percentage of 5hmC consumed |
|---|---|---|---|---|---|
| 1 | benzoylacetonitrile | 90 | 1,4-dioxane | 2 | 54.5 |
| 2 | 2-acetylfuran | 90 | 1,4-dioxane | 2 | 48.1 |
| 3 | 3-oxobutyronitrile | 90 | 1,4-dioxane | 2 | 66.4 |
| 4 | phenylacetylene | 90 | 1,4-dioxane | 2 | 79.8 |
| 5 | acetophenone | 90 | 1,4-dioxane | 2 | 84.3 |
| 6 | acetophenone | 90 | 1,4-dioxane | 0.3 | 29.3 |
| 7 | acetophenone | 90 | 1,4-dioxane | 3 | 74.9 |
| 8 | acetophenone | 90 | 1,4-dioxane | 4 | 67.4 |
| 9 | acetophenone | 90 | 1,4-dioxane | 5 | 67.0 |
| 10 | acetophenone | 90 | 1,4-dioxane | 6 | 64.3 |
| 11 | acetophenone | 90 | DMF | 2 | 58.0 |
| 12 | acetophenone | 90 | DMSO | 2 | 52.3 |
| 13 | acetophenone | 90 | Toluene | 2 | 65.1 |
| 14 | acetophenone | 90 | $H_2O$ | 2 | 5.6 |
| 15 | acetophenone | 80 | 1,4-dioxane | 2 | 12.2 |
| 16 | acetophenone | 90 | 1,4-dioxane/$H_2O$ (v/v=99:1) | 2 | 10.1 |
| 17 | phenethyl alcohol | 90 | 1,4-dioxane | 2 | 24.2 |
| 18[b] | acetophenone | 90 | 1,4-dioxane | 2 | 69.2 |
| 19[c] | acetophenone | 90 | 1,4-dioxane | 2 | 38.1 |
| 20[d] | acetophenone | 90 | 1,4-dioxane | 2 | 44.5 |
| 21[e] | acetophenone | 80 | Toluene/$H_2O$ (v/v=1:1) | 6 | 35.4 |

[a] Reaction condition:coupling agent (8 μmol), t-BuOK (1.1 mg, 10 μmol), benzophenone (1mg, 5.5 μmol) and 1,4-dioxane (200 μL), continuous nitrogen. [b] Reaction with continuous oxygen. [c] Reaction with the extra addition of tris (triphenylphosphine) ruthenium(II) chloride (0.5 mg). [d] Reaction with the extra addition of Benzylidene [1,3-bis(2,4,6-trimethylphenyl)-2-imidazolidinylidene] dichloro (tricyclohexyl-phosphine) ruthenium (0.5 mg). [e] Reaction with the extra addition of silver trifluoromethanesulfonate (0.5 mg)