Electronic Supplementary Material (ESI) for Chemical Communications

Optical detection of epigenetic marks: Sensitive quantification and direct imaging of individual hydroxymethylcytosine bases

Yael Michaeli*, Tamar Shahal*, Dmitry Torchinsky, Assaf Grunwald, Rony Hoch, and Yuval Ebenstein.

Supporting information

DNA samples: 5hmC- saturated DNA fragments of 1 kb and 70 bp were prepared by PCR amplification of lambda DNA (New England Biolabs; (NEB), Ipswich MA, USA), using the following primers: forward primer: 5-CTCATGCTGAAAACGTGGTG-3, reverse primer: 5-GGACAGGACCAGCATACGAT-3 and forward primer: 5-/5Alex488N/TAAATTAGTTACACAGGAAA-3 reverse primer: 5-AAGCCACAA CTCTAATTTT-3 for 1 kb and 70 bp DNA fragments, respectively (Integrated DNA Technologies Inc, Coralville, IA USA). A typical reaction was performed in a volume of 50µl, and contained 200 ng of template DNA, 2 units of Vent (exo-) (NEB), 200µM of dAGT (Sigma-Aldrich Israel Ltd. Rehovot, Israel) and 5hmC (Bioline Reagents Ltd., London, UK) nucleotides in NEB thermopol buffer. Reaction mixtures were incubated at 95°C for 2 minutes as an initial step, followed by 30- 35 cycles of 30 seconds at 95°C, 30 seconds at 55°C and 2 minutes at 72°C or 30 seconds at 95°C, 30 seconds at 42°C and 30 seconds at 72°C for the 1kb and 70 bp products, respectively, and finally 5 minutes at 72°C.

For the analysis of the labeling efficiency of 5hmC, the 70 bp lambda DNA fragments were prepared with Alexa Fluor 647- dCTP (Molecular Probes, Eugene, OR, *USA*) instead of 5hmC. The reaction was performed in a volume of 50µl, and contained 200 ng of template DNA, 2 units of Vent (exo-) , 50µM of dATP dGTP and dTTP and 50 µM Alexa Fluor 647- dCTP nucleotides in NEB thermopol buffer. Reaction mixtures were incubated at 95°C for 2 minutes as an initial step, followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 42°C and 30 minutes at 50°C, and finally 10 minutes at 72°C. This control Alexa Fluor-647- PCR product contained three cytosine sites and represents 100% labeling efficiency. All PCR products were cleaned of free nucleotides and primers using QIAquick PCR purification kit (QIAGEN GmbH, Hilden, Germany).

For extraction of DNA from mouse tissues, 5prime ArchivePure DNA cell/tissue kit was used according to manufacturer's instructions.

Preparation of hydroxymethylated lambda DNA by Sequence specific labeling: Lambda phage intact genomes (48.5 kb) were labeled with 5hmC nucleotides by nick translation[1] . 10µg of DNA was incubated with 10u/µg Nt.BspQI (NEB) nicking enzyme in 100µl NEB buffer 3 for 2h at 50°C, followed by heat inactivation for 20 min at 80°C. For labeling of the DNA, nicked DNA was incubated for 2h at 72°C in 200µl NEB thermopol buffer with 2u/µg Vent (exo-) and 250nM dNTPs. For incorporation of hydroxymethyl, DNA was incubated with 250nM dATP dGTP and dTTP and 250nM of 5-hydroxymethyl-labeled cytosine (Bioline).

Fluorescent-labeling of 5hmC by click chemistry: In the case of lambda DNA for single molecule optical mapping, 10 µg of Nt.BspQI site Hydroxymethylated DNA in 50mM Hepes (Sigma- Aldrich

Israel Ltd.) was incubated with 20 units of T4-beta-glucosyltransferase (NEB) for glucosylation of 5hmC, in the presence of NEB buffer 4 and 150µM UDP-azide glucose (Active Motif, Carlsbad, CA, USA), for 2h at 37°C. As a control reaction, UDP-azide glucose was replaced by UDP-glucose at the same molar concentration. The click chemistry reaction was performed by the addition of 250µM Alexa Fluor 555 DIBO alkyne (Molecular Probes) for 1h at 37°C. Buffer was then exchange to 50mM Hepes, and sample was stained with 1µM YOYO-1.

For labeling 5hmC- saturated PCR products, 1-3 µg was first glucosylated by incubation with UDP-glucose-azide (Active Motif) at a molar ratio of 1:30 (5hmC: UDP- glucose-azide) and 50 units of T4-beta-glucosyltransferase (NEB), in the presence of NEB buffer 4, overnight at 37°C. Two types of click reactions were used: First, a copper- free reaction, with Alexa Fluor 647- DIBO alkyne (Molecular Probes) (for the 1kb PCR product) at a molar ratio of 1:100 (5hmC: DIBO) in 10 mM PBS. Second, a copper- dependent reaction with Alexa Fluor 647- alkyne at a molar ration of 1:100 (5hmC: alkyne) in the presence of 200 mM triethylammonium acetate buffer, 50% DMSO, freshly prepared 0.5 mM ascorbic acid and 0.5 mM Cu-TBTA (Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine complex in 55% DMSO. The copper in the second reaction serve as a catalyzer which is not required when using cyclooctynes as in the DIBO alkyne. Both reactions were incubated at 22°C overnight. The reaction sample was degassed by nitrogen before addition of Cu-TBTA complex and flashed before incubation. As a control reaction, UDP- glucose- azide was replaced by UDP-glucose at the same molar concentration. Each incubation step preceded a purification step with Qiagen PCR purification columns (QIAGEN).

5hmC quantification by UV-Vis Spectroscopy: In order to determine the percentage of 5hmC in an examined DNA sample, a calibration curve was obtained with a DNA sample that contains a known 5hmC percentage. For this purpose, Alexa Fluor 647- 5hmC labeled- 1 kb PCR product was mixed with increasing concentrations of non- labeled XL1 bacteria plasmids and the absorption ratio at 260nm for DNA and 647nm for labeled 5hmC was plotted. Increasing amounts of plasmids (600, 1200 and 1750, 3500 ng) were added to fixed amounts of 5hmC- saturated 1 kb lambda fragments (in the order of 4 and 2 ng). The 1kb fragment contained 29% 5hmC and was fluorescently labeled by the click reaction. Plasmids were extracted from a XL1 PE bacteria by a DNA purification system (Promega, Madison WI, USA). Absorption measurements were conducted on a NanoPhotometer® P 300 (IMPLEN, Munich, Germany).

Analysis of labeling efficiency: For this purpose, we used Alexa Fluor 488-labeled-70bp PCR products that were prepared with Alexa Fluor 647- dCTP nucletides, or with 5hmC nucleotides. The 5hmC product was subjected to click reaction and the two sets of DNA samples were analyzed by electrophoresis and run side by side through a 3% agarose gel, in TBE buffer, at 80 volts. The gel was imaged on a multicolour laser gel scanner, GE Healthcare FLA5000. A normalization of the DNA amount loaded on the gel was achieved by comparing the fluorescence intensity of the 70 bp bands at 510 nm, resulting from the Alexa Fluor 488 molecule bound to the forward primer of both control and click reaction products, following excitation with 473 nm laser. The efficiency of the click labeling

procedure was calculated by comparing the bands fluorescence intensity at 665 nm, resulting from the Alexa Fluor- labeled dC nucleotides or 5hmC subjected to a click reaction, following excitation with 635 nm laser. Fluorescence intensity measurements were analyzed by ImageJ: http://rsbweb.nih.gov/ij/.

DNA extension: Surfaces for DNA extensions were prepared according to Sidorova *et al*.[2] with minor modifications. In short, 24 × 24 glass cover slips were cleaned by 7 hours to overnight incubation in a freshly made 2:1 (v/v) mixture of nitric acid (70%) and hydrochloric acid (37%). The incubation proceeded in a chemical fume hood and was followed by an extensive wash with ultrapure water (18 MΩ), ethanol and dried under a stream of nitrogen. Dry slides were immersed in a premixed solution containing 595µl N-trimethoxysilylpropyl-N,N,N-trimethylammonium chloride and 216µl of vinyltrimethoxysilane in 300ml ultrapure water and incubated overnight at 65ºC. After incubation, slides were thoroughly washed with ultrapure water and ethanol and stored at 4 ºC in ethanol. The silane solution was freshly made and thoroughly mixed before the slides were introduced into the mixture. Stored slides were normally used within 2 weeks. Lambda DNA molecules were extended on silanized glass slides by placing a 5µl drop of pre-labeled Lambda DNA in 50mM HEPES buffer and 200mM dithiothreitol (DTT) in- between a dry silanized glass slide and a non-treated microscope glass slide (lambda DNA concentration was as appropriate for single molecule imaging).

Data acquisition and analysis: Extended DNA molecules were imaged on a MORE imaging system (TILL photonics GmbH, Munich ,Germany) with an Olympus UPlanApo 60X 1.35 NA oil immersion objective. A 150 W Xenon lamp with galvanometer driven filter switching was used as an excitation source. The filter sets used to image YOYO-1 stained DNA and the Alexa Fluor 555 labels were 482/18 and 561/14 bandpass excitation filters, 405/488/561/640 quadband beamsplitter and a 446/523/600/677 quadband emission filter (all from Semrock, Rochester, NY, USA). Images were acquired by a DU888 EMCCD (Andor, Belfast, Ireland) with an EM gain setting of 200 and integration times of 200 ms and 1300 ms for YOYO-1 and Alexa Fluor 555 respectively.

We mapped the positions of the Alexa-Fluor 555 fluorescence spots along the DNA in order to verify that they are located in the expected positions of 5hmC sites. In order to determine the genomic position of the tags we measure the distance (in pixels) between the Tag signal position and the far end of the DNA. This value was divided by the measured full length of the DNA to give a normalized position value (ranging from 0 to 100). Fluorescence spots were mapped manually using a Matlab program written for this purpose. The spectrally separated images of the stretched DNA and of the Alexa Fluor 555 fluorescent spots were overlaid to visualize fluorescent tags bound on the hydroxymethylated DNA bases. Using the Matlab function Improfile, a line was manually drawn along each long DNA strand (> 80 pixels) that showed at least four fluorescent spots along its contour and had orientation that was clearly evident from the spot pattern. The Improfile function projects the intensity values of each pixel along the drawn line on a pixel vs. intensity plot. The DNA lengths were measured by subtracting the two y-intercepts on the DNA channel that represent DNA ends. Fluorescence spots generated by labeled 5hmC bases may be accurately localized by 2D Gaussian fitting[3,4]. Images were analyzed using a custom Matlab program that extracts the position coordinates

of fluorescent spots. After localizing each spot, we measured its distance to one end of the DNA in order to assign a genomic position. In order to account for the large variation in stretching factor between different DNA molecules we normalized the measured locations to units of percentage of the whole genome (100% representing the far end of the phage genome at 48500 base pairs). By dividing the distance of the tag location from the DNA end point by the total length of the template DNA molecule, a normalized value for all detected tags is calculated, allowing statistical analysis of the pooled data as shown in the main text.

References:

[1] Das, S. K.; Austin, M. D.; Akana, M. C.; Deshpande, P.; Cao, H.; Xiao, M. *Nucleic acids research* **2010**, *38*, e177.

[2] Sidorova, J. M.; Li, N.; Schwartz, D. C.; Folch, A.; Monnat, R. J. *Nature protocols* **2009**, *4*, 849–61.

[3] Ebenstein, Y.; Gassman, N.; Kim, S.; Antelman, J.; Kim, Y.; Ho, S.; Samuel, R.; Michalet, X.; Weiss, S. *Nano letters* **2009**, *9*, 1598–603.

[4] Kim, S.; Gottfried, A.; Lin, R. R.; Dertinger, T.; Kim, A. S.; Chung, S.; Colyer, R. A.; Weinhold, E.; Weiss, S.; Ebenstein, Y. *Angewandte Chemie (International ed. in English)* **2012**, *51*, 3578–81.