Supplementary Materials for:

Detection of 5-Methylcytosine and 5-Hydroxymethylcytosine in DNA via Host-Guest Interactions inside α-Hemolysin Nanopore

Tao Zeng,[ab]‡ Lei Liu,[a]‡ Ting Li,[a] Yuru Li,[a] Juan Gao,[a] Yuliang Zhao,[b]* and Hai-Chen Wu[a]*

[a]Key Laboratory for Biomedical Effects of Nanomaterials & Nanosafety, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

[b]National Center for Nanoscience and Technology of China, Beijing 100190, China.

*To whom correspondence should be addressed. Email: haichenwu@ihep.ac.cn; Tel: +8610-88235745; Fax: +8610-88235745; or Email: zhaoyl@nanoctr.cn.

‡These authors contributed equally to this work.

Materials and Methods

Materials and characterisation. 1,2-Diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPc) was purchased from Avanti Polar Lipids (Beijing distributor). DNA6 and DNA8 were purchased from Takara Bio Inc. (Dalian) and all the other DNA oligomers were purchased from Sangon Biotech (Shanghai). Micro Bio-Spin P6 gel columns (Tris buffer) were purchased from Bio-Rad (Hercules, CA). All the columns were pre-equilibrated three times with 80 μL deionised water prior to use. All chemicals were purchased from Sigma-Aldrich, Alfa Aesar and J&K, and used without further purification. All samples and buffers were prepared in deionised water (Millipore, MA). DNA mass spectrometry was analysed on Thermo-Finnigan LCQ Deca XP Plus. HRMS data of new compounds were obtained using Bruker ultrafleXtreme MALDI-TOF/TOF. NMR spectra were recorded on Bruker Avance III 500 spectrometers. FT-IR measurements were performed on a Bruker Tensor27 spectrometer. DNA concentration was measured on Malcom e-spect ES-2.

Protein preparation. Wild type αHL-D8H6 and mutant monomers (K8L, K147N, wildtype background with D8H6 tail) were produced by expression in BL21 (DE3) pLysS *Escherichia coli* cells. The monomers were then assembled into homoheptamers on rabbit red blood cell membranes followed by purification with 8% SDS-PAGE as described earlier.[53] The purified

heptamer protein was conserved in buffer (10 mM Tris-HCl, pH 7.9, 50 mM NaCl) and stored at -70°C.

Bisulfite conversion. The bisulfite treatment of ssDNA was conducted according to an established protocol with a few minor modifications.[31] Briefly, ssDNA (100 μL, 100 μM) was incubated with sodium metabisulfite (400 μL, 2.5 M, pH 5.0, containing 10 mM hydroquinone) at 70 °C for 1 hour. Then, the mixture was cooled to room temperature and subjected to ultrafiltration (molecular weight cut-off 3000, rpm 14000 *g*, and 1 hour). The concentrated sample was then purified with Micro Bio-Spin P6 columns. Next, bisulfite-treated ssDNA (40 μL) was desulfonated in 0.5 M NaOH (10 μL) at room temperature for 5 minutes, followed by neutralization with 5 μL 1.0 M HCl and desaltation with Micro Bio-Spin P6 columns. The purified sample was then lyophilised and stored for ESI-MASS analysis and further modifications. The recovery yield of this step is about 10% due to the DNA degradation, but the conversion efficiency is quantitative.

5mC-DNA modification.

Condensation with O-(pent-4-yn-1-yl)hydroxylamine. Typically, 10 μL bisulfite-treated oligonucleotide (100 μM) was heated with 100 μL *O*-(pent-4-yn-1-yl)hydroxylamine hydrochloride (~1 M, pH 5.2 titrated with diethylamine) and 20 μL phosphate buffer (0.2 M, pH 5.2) at 60 ºC for 12 h. Afterwards, the sample was purified with Micro Bio-Spin P6 columns. The filtrate (DNA-Alkyne) was lyophilized and re-dissolved in 10 uL deionised water for the following coupling reaction and mass spectrometry analysis. The coupling efficiency of this step is about 90%.

Conjugation with ferrocene azide. The conjugation between DNA-Alkyne and ferrocene azides was carried out using Cu(I)-catalyzed click chemistry. 3.33 μL DNA-Alkyne, 2 μL HEPES buffer (100 mM, pH 7.4), 1.17 μL deionised water, 2 μL ferrocene azide (N_3-PEG_4-Fc: 100 mM in dimethyl sulfoxide), 1 μL sodium ascorbate (20 mM), 0.5 μL copper(II) nitrate (20 mM) were mixed together (total volume: 10 μL) and incubated at 25 ºC for 2 h. After quenched by EDTA (2 μL, 100 mM), the sample was purified with Micro Bio-Spin P6 columns. The filtrate (DNA-Fc) was analysed by ESI-MS to determine the conjugation efficiency and stored at -20 ºC. The coupling efficiency of this step is about 100%.

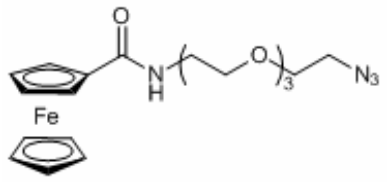
5hmC-DNA Modification.

$KRuO_4$ oxidation. The procedure for the oxidation of 5hmC-DNA was adapted from literature[21]. Typically, NaOH (12 μL, 0.1 M) and $KRuO_4$ (0.5 μL, 15 mM in 0.1 M NaOH) were sequentially added to the starting 5hmC-DNA solution (12.5 μL, 80 μM), and the mixture was incubated on an ice bath for an hour. The reaction was quenched by the addition of MES buffer (25 μL, 200 mM 2-(*N*-morpholino)ethanesulfonic acid, pH 6.5), and the sample was purified with Micro Bio-Spin P6 columns. The filtrate (DNA-Aldehyde) was immediately subjected to the following coupling reaction with *O*-(pent-4-yn-1-yl)hydroxylamine. The recovery yield of this step is about 27% due to the DNA damage by $KRuO_4$; the oxidation conversion is about 100%.

Coupling with O-(pent-4-yn-1-yl)hydroxylamine. To 20 μL DNA-Aldehyde sample were added 10 μL NH_4OAc buffer (100 mM NH_4OAc, 100 mM *p*-anisidine, pH 4.5) and 10 μL *O*-(pent-4-yn-1-yl)hydroxylamine (50 mM), and the mixture were incubated at 25 °C for 12 hours. Afterwards, the solution was ultrafiltrated (molecular weight cut-off 3000, rpm 14000 *g*, 30 min × 3) and purified with Micro Bio-Spin P6 columns. The filtrate was lyophilised and re-dissolved in 20 μL deionised water for ESI-mass spectrometry analysis and further conjugation. The coupling efficiency of this step is about 98.7%.

Conjugation with N_3-PEG_4-Fc. This procedure is similar to that of the conjugation of 5mC-containing DNA-Alkyne with N_3-PEG_4-Fc, except that N_3-PEG_4-Fc was dissolved in acetonitrile (2 μL, 100 mM) and the incubation was performed in an ice bath for 2 hours. The coupling efficiency of this step is about 94.3%.

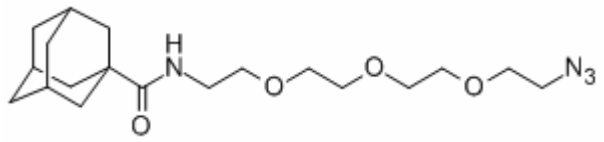
Syntheses and characterisation of new compounds.



N_3-PEG_4-Fc

1-Amino-11-azido-3,6,9-trioxaundecane (218 mg, 1.68 mmol) and *N*-hydroxysuccinimide ester of ferrocenecarboxylic acid (328 mg, 1.12 mmol) were stirred in

an aqueous solution of sodium bicarbonate (84 mg, 1.68 mmol) at room temperature for 12 h. After extraction of the aqueous layer with chloroform (20 mL × 3), the organic phase was combined, dried over Na_2SO_4, and evaporated under reduced pressure. The obtained crude product was purified on a silica gel column to afford an orange solid (345 mg, 72%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.34 (s, 1H), 4.67 (s, 2H), 4.30 (s, 2H), 4.17 (s, 5H), 3.68-3.59 (m, 12H), 3.55 (m, 2H), 3.34 (t, J = 4.9 Hz, 2H). ^{13}C NMR of N_3-PEG_4-Fc. ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.26 (s), 76.17 (s), 70.57 (s), 70.55 (s), 70.50 (s), 70.30 (s), 70.22 (s), 69.94 (s), 69.92 (s), 69.70 (s), 68.18 (s), 50.57 (s), 39.23 (s). IR (KBr): 3323, 2868, 2103, 1632, 1106 cm^{-1}. HRMS (MALDI-TOF/TOF) Exact mass calcd. for $C_{19}H_{26}FeN_4O_4$: 430.1303; found 430.1299.



N_3-PEG_4-Ad

1-Amino-11-azido-3,6,9-trioxaundecane (203 mg, 0.93 mmol) and *N*-hydroxysuccinimide ester of 1-adamantanecarboxylic acid (257 mg, 0.93 mmol) were stirred in an aqueous solution of sodium bicarbonate (78 mg, 0.93 mmol) at room temperature for 24 h. After extraction of the aqueous layer with chloroform (15 mL × 3), the organic phase was combined, dried over Na_2SO_4, and evaporated under reduced pressure. The obtained crude was purified on a silica gel column to afford colorless oil (60 mg, 17%). ^{1}H NMR (500 MHz, $CDCl_3$) 3.68-3.58 (m, 10H), 3.52 (t, J = 4.9 Hz, 2H), 3.42 (dd, J = 10.0, 5.2 Hz, 2H), 3.37 (t, 4.6 Hz, 2H), 2.01 (s, 3H), 1.82 (s, 6H), 1.69 (q, J = 12.2 Hz, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 178.07 (s), 70.69 (s), 70.56 (s), 70.19 (s), 70.10 (s), 69.90 (s), 50.59 (s), 40.52 (s), 39.09 (s), 38.87 (s), 36.47 (s), 28.06 (s). IR (KBr): 3342, 2905, 2103, 1644, 1122 cm^{-1}. HRMS Exact mass calcd. for $C_{19}H_{32}N_4O_4$ [M+H]: 381.2502; found 381.2492.

Single-channel current recording. A bilayer of DPhPc was formed across an aperture 100-150 μm in diameter in a 20-μm thick polytetrafluoroethylene film (Goodfellow, Malvern, PA) that divided a planar bilayer chamber into two compartments, *cis* and *trans*. Both compartments contained 1 mL of buffer. DNA samples were added to the *cis* compartment, which was connected to ground. The *trans* compartment was connected to the head-stage of the

amplifier. All experiments were carried out in 3 M KCl, 10 mM Tris, pH 8.0, at 22.5 ± 2 °C, unless otherwise stated. Currents were recorded by a patch clamp amplifier (Axopatch 200B; Axon instruments, Foster City, CA) with an internal low-pass Bessel filter operating at 10 kHz, and sampled at 100 kHz by a computer equipped with a Digidata 1440 A/D converter (Axon instruments).

Data analysis. Current traces were analysed with Clampfit 10.2 software (Axon Instruments). Events were detected using the Event Detection feature, and used to construct amplitude and dwell time histograms. Origin 8.5 (Microcal, Northampton, MA) and Clampfit 10.2 were used for histogram construction, curve fitting and graph presentation. Adobe Illustrator CS2 was used for making figures.

Electroosmotic flow inside αHL.

According to ref. 46, the net water flux (J_w) caused by the electroosmotic flow in KCl solution can be expressed as:

$$J_w = N_w \cdot \frac{gV}{e} \cdot \left(\frac{P_{K^+}/P_{Cl^-} - 1}{P_{K^+}/P_{Cl^-} + 1} \right) \qquad (1)$$

where N_w is the number of water molecules transported per ion; g is the unitary conductance of the pore in S; V is applied potential; e is the charge of an electron in C; P_{K^+}/P_{Cl^-} is the ion permeability ratios.

By assuming a very simple model without any free energy barrier, the ionic conductance of electrolyte inside a nanopore should be given by:

$$g = 6.02 \times 10^{26} (\mu_K + \mu_{Cl}) c_{KCl} e \pi D_{pore}^2 / 4L_{pore} \qquad (2)$$

where $\mu_K = 7.62 \times 10^{-8}$ m^2/Vs and $\mu_{Cl} = 7.91 \times 10^{-8}$ m^2/Vs are the electrophoretic mobilities of potassium and chloride ions, respectively; c_{KCl} is the KCl concentration in M; D_{pore} is the nanopore diameter; and L_{pore} is the length of the nanopore.

From equation 1 and equation 2, we can obtain $J_w \propto V c_{KCl}$.

Supplementary figures

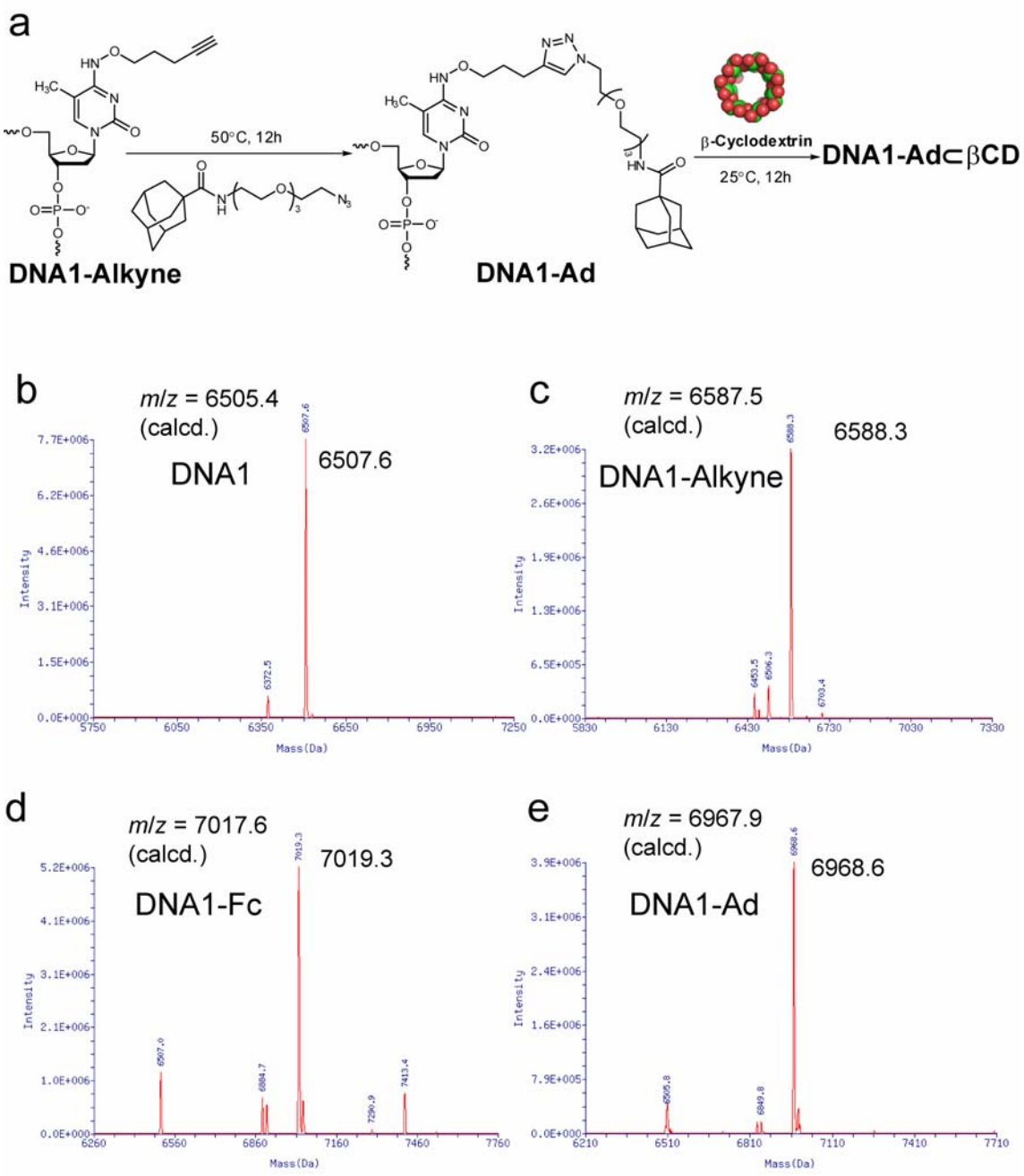


Figure S1. Modification of DNA1 with Ad-βCD and mass spectrometry characterization. (a) Chemical reactions of the modification of DNA1 with Ad-βCD complex. (b) Mass spectrometry characterization of DNA1. (c) Mass spectrometry characterization of the product DNA1-Alkyne. (d) Mass spectrometry characterization of the product DNA1-Fc. (e) Mass spectrometry characterization of the product DNA1-Ad.

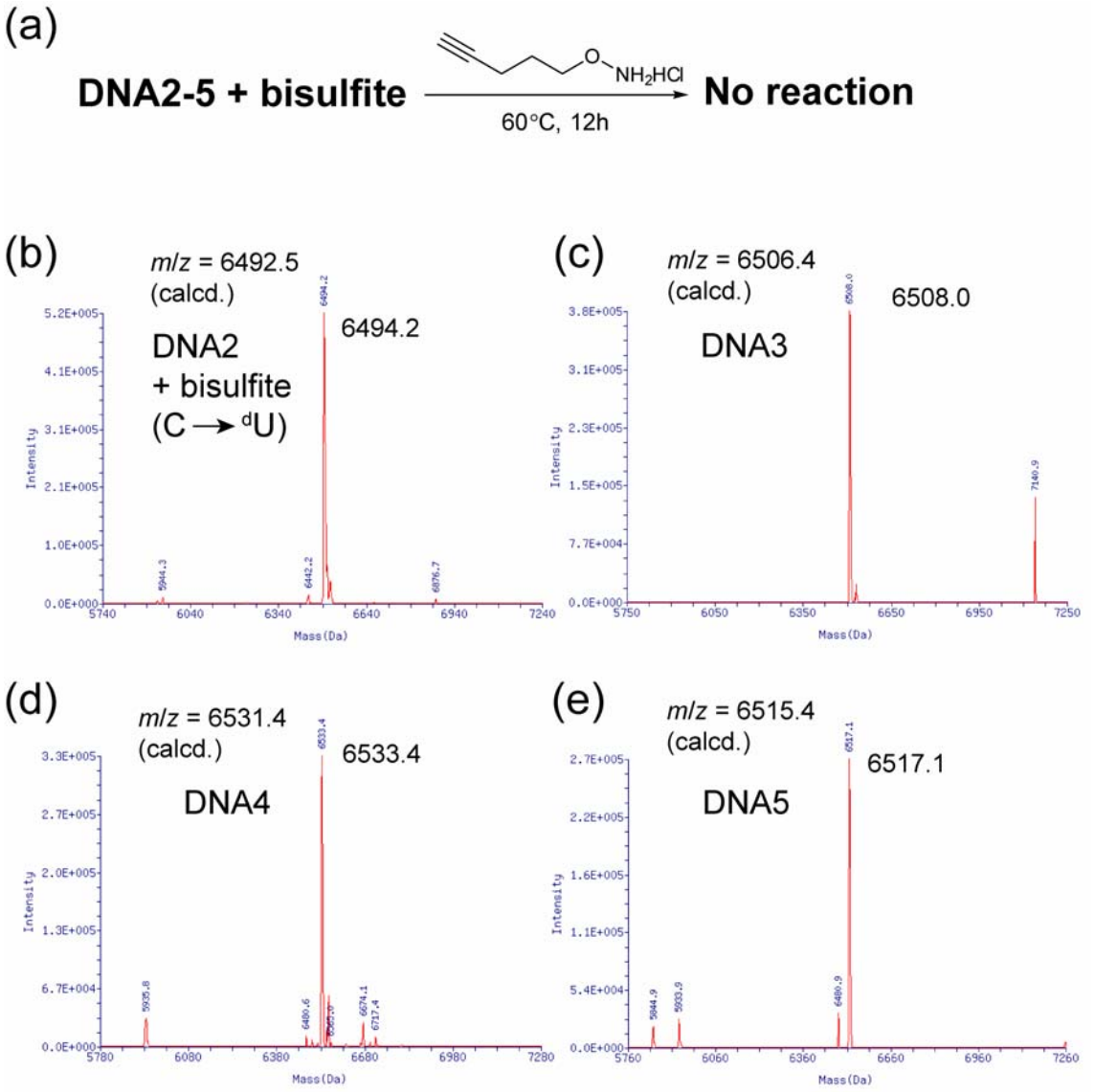


Fig. S2 Modification reactions of the control DNAs (DNA2-5). (a) Reactions of DNA2-5 with *O*-(pent-4-yn-1-yl)hydroxylamine after bisulfite treatment. (b) Mass spectrometry characterisation of the product DNA2-^{d}U. The cytosine in DNA has been converted to uracil after bisulfite treatment. (c) Mass spectrometry characterisation of DNA3 after the reaction. (d) Mass spectrometry characterisation of DNA4 after the reaction. (e) Mass spectrometry characterisation of DNA5 after the reaction.

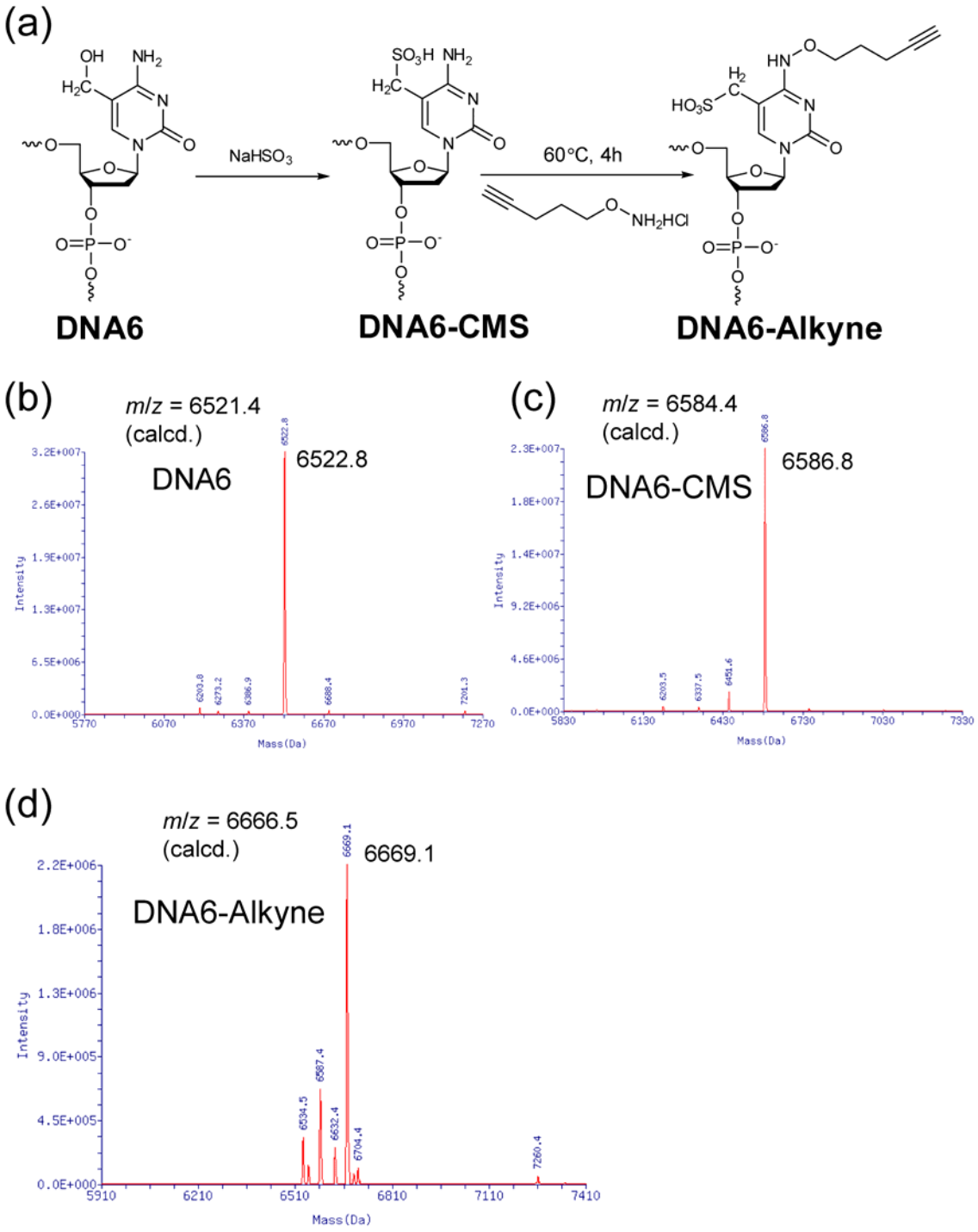


Fig. S3 Reaction of DNA6 with *O*-(pent-4-yn-1-yl)hydroxylamine after bisulfite treatment. (a) Bisulfite treatment of 5hmC-containing DNA6 and subsequent reaction with *O*-(pent-4-yn-1-yl)hydroxylamine. (b) Mass spectrometry characterization of DNA6. (c) Mass spectrometry characterisation of DNA6-CMS. (d) Mass spectrometry characterisation of DNA6-Alkyne.

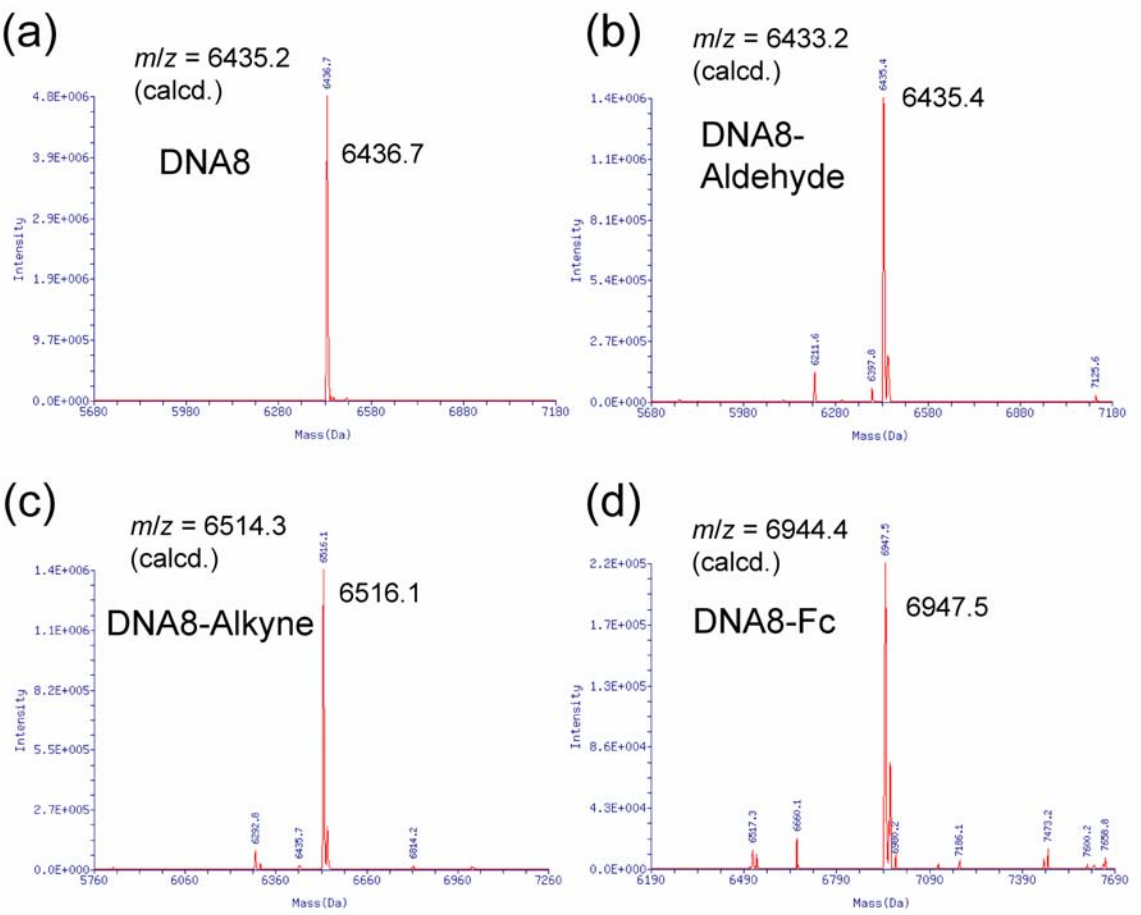


Fig. S4 Selective modifications of 5hmC-containing DNA8. (a) Mass spectrometry characterisation of DNA8. (b) Mass spectrometry characterisation of DNA8-Aldehyde. (c) Mass spectrometry characterisation of DNA8-Alkyne. (d) Mass spectrometry characterisation of DNA8-Fc.

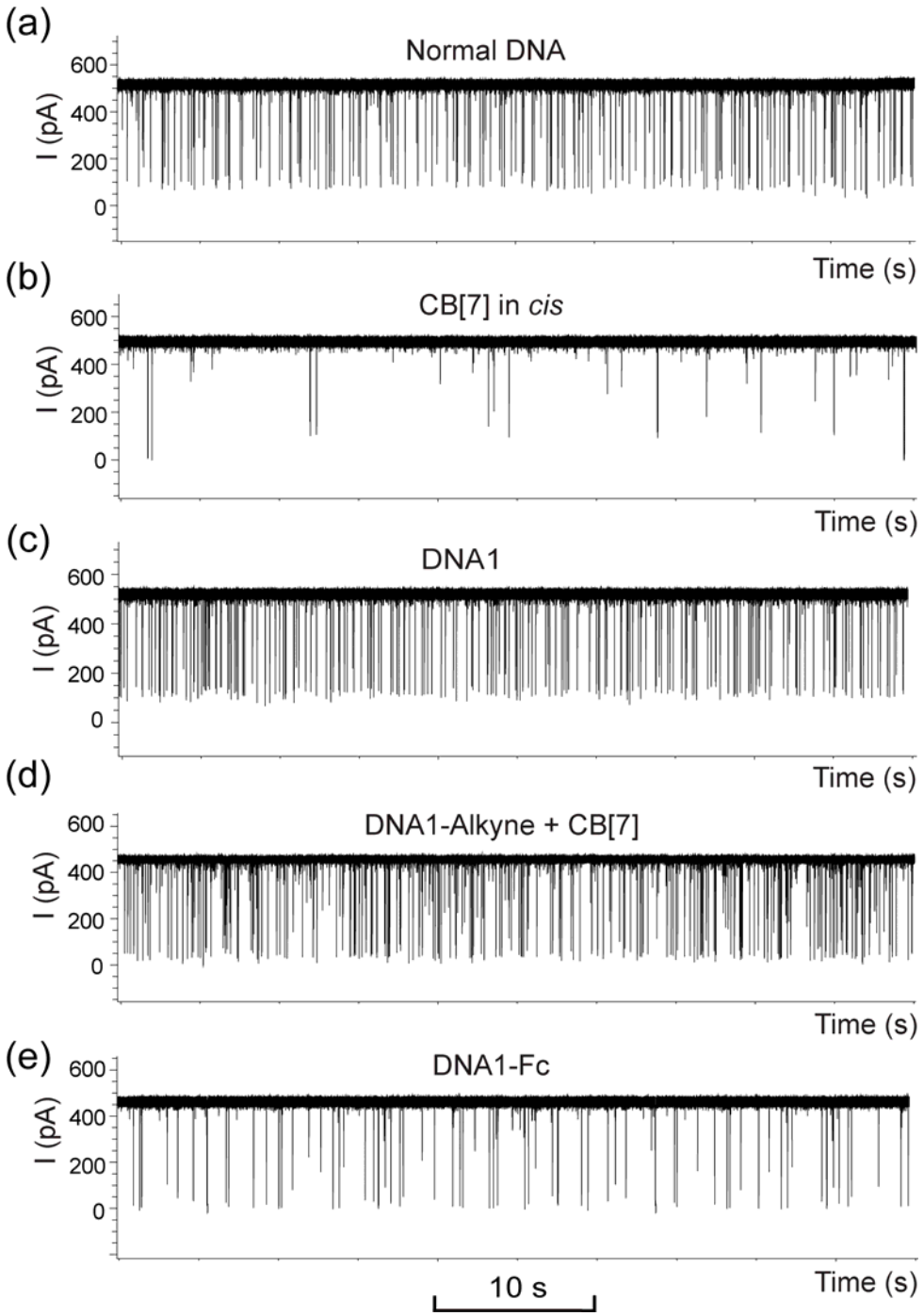


Fig. S5 Translocation current traces of control experiments. (a) DNA9 with random sequence. (b) CB[7] in *cis* (concentration: 50 μM). (c) DNA1. (d) DNA1-Alkyne and CB[7] (concentration: 50 μM). (e) DNA1-Fc. All the DNA concentrations were 1.0 μM except for DNA1-Fc (0.25 μM). DNA and CB[7] were all placed in *cis*. Data were acquired in the buffer of 3 M KCl and 10 mM Tris, pH 8.0, with the transmembrane potential held at +160 mV.

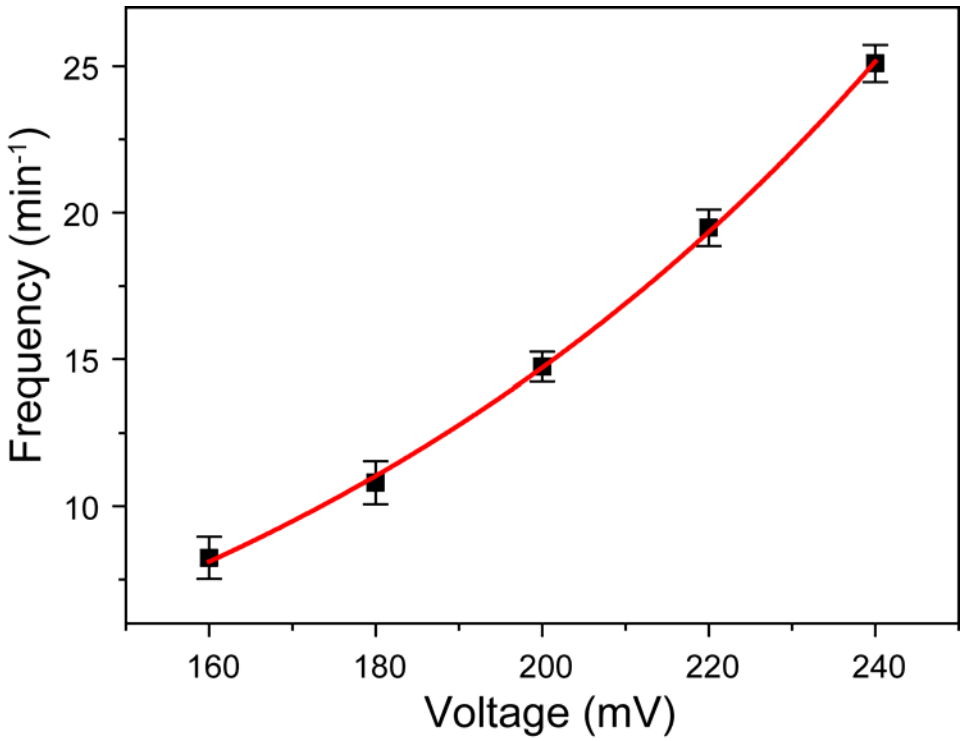


Fig. S6 Voltage-dependence of the frequency of signature events. Traces were recorded in 3 M KCl buffered with 10 mM Tris, pH 8.0. DNA1-Fc (final concentration: 0.25 μM) were incubated with CB[7] (final concentration 50 μM) at room temperature for 2 hours before measurement.

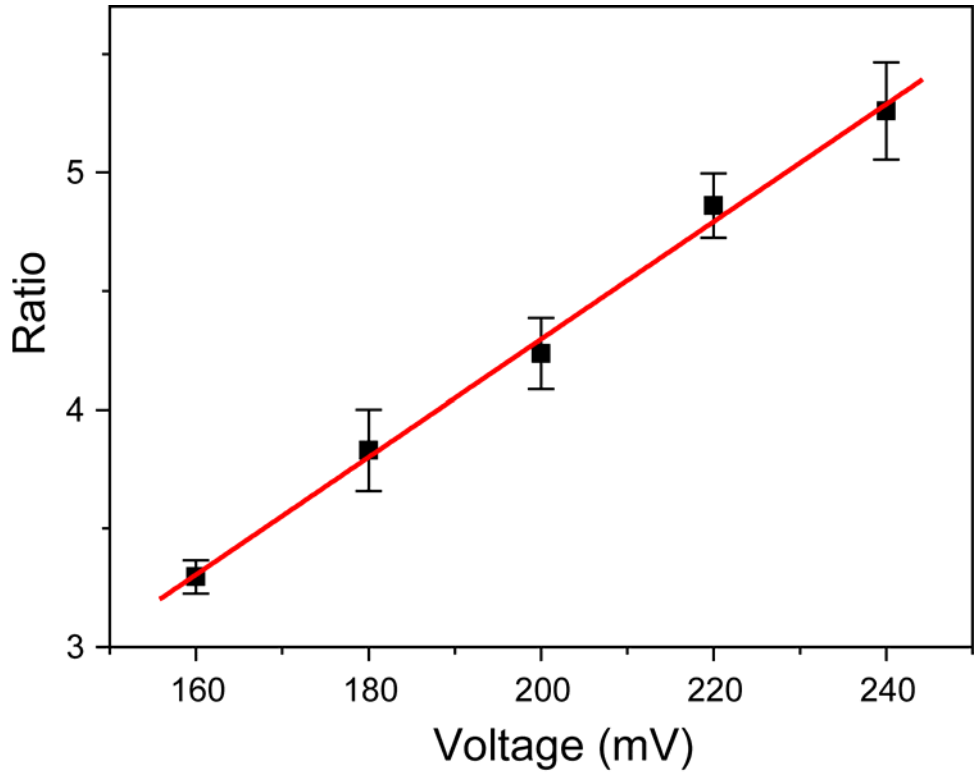


Fig. S7 Voltage-dependence of the ratio of multi-level signature events to single-level events. In the range of 160 mV to 240 mV, the ratio of multi-level signature events to single-level events increased monotonically with the applied potential. The data could be fitted by a linear regression. Traces were recorded in 3 M KCl buffered with 10 mM Tris, pH 8.0. DNA1-Fc (final concentration: 0.25 μM) were incubated with CB[7] (final concentration 50 μM) at room temperature for 2 hours before measurement.

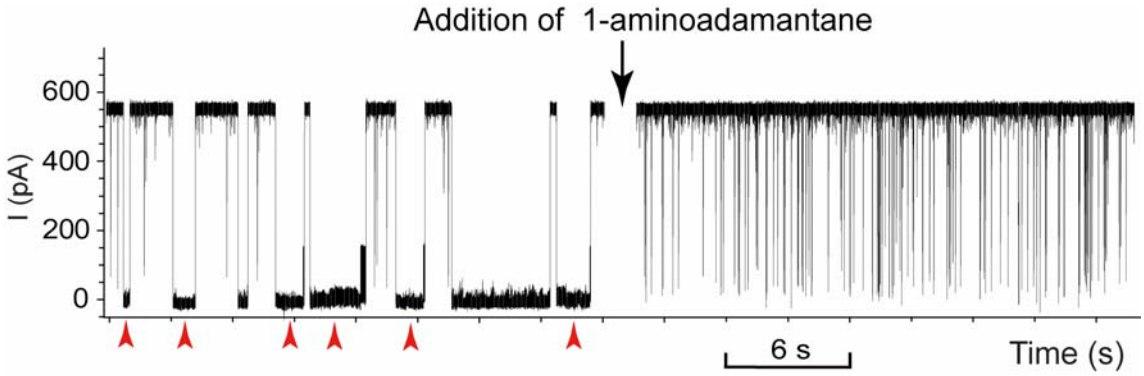


Fig. S8 Influence of a competing "guest" molecule on the generation of signature events. When 1-aminoadamantane (final concentration 1.0 μM) was added in *cis* chamber and incubated with DNA1-Fc⊂CB[7] for 30 min, all the signature events disappeared and only ssDNA translocation events were observed. Traces were recorded in 3 M KCl buffered with 10 mM Tris, pH 8.0. DNA1-Fc (final concentration 0.25 μM) were incubated with CB[7] (final concentration 50 μM) at room temperature for 2 hours before measurement. The transmembrane potential was held at +200 mV.

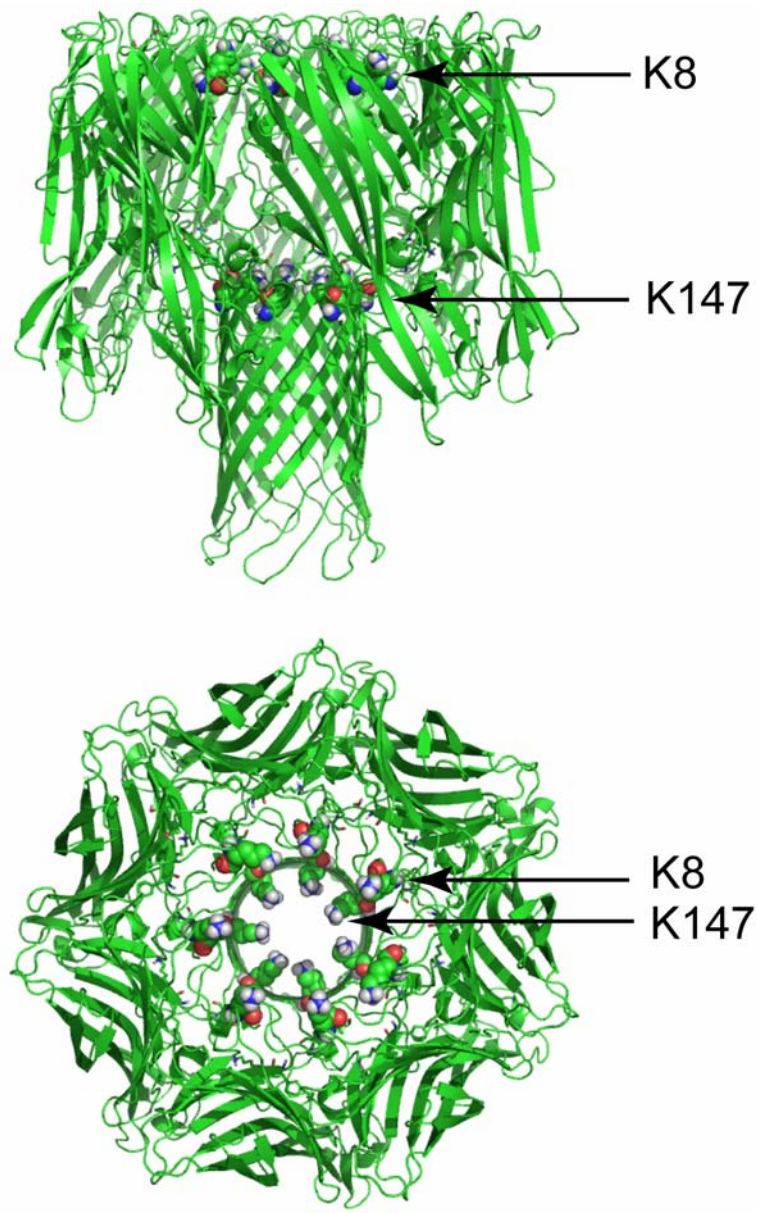


Fig. S9 Positions of amino acids K8 and K147 inside αHL. The images were generated using PyMOL software based on the crystal structure of αHL.[54]

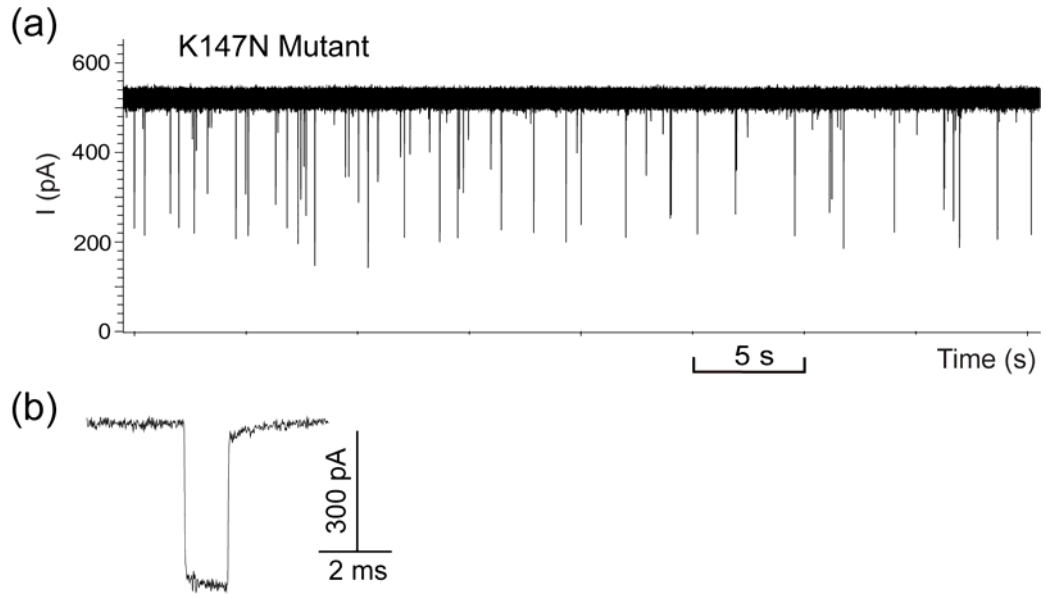


Fig. S10 Translocation of DNA1-Fc⊂CB[7] through engineered αHL mutant K147N. (a) Representative current trace of the translocation of DNA1-Fc⊂CB[7] through mutant K147N. (b) Expanded view of a current event. Traces were recorded in 3 M KCl buffered with 10 mM Tris, pH 8.0. DNA1-Fc (final concentration: 0.25 μM) were incubated with CB[7] (final concentration 50 μM) at room temperature for 2 hours before measurement. The transmembrane potential was held at +200 mV.

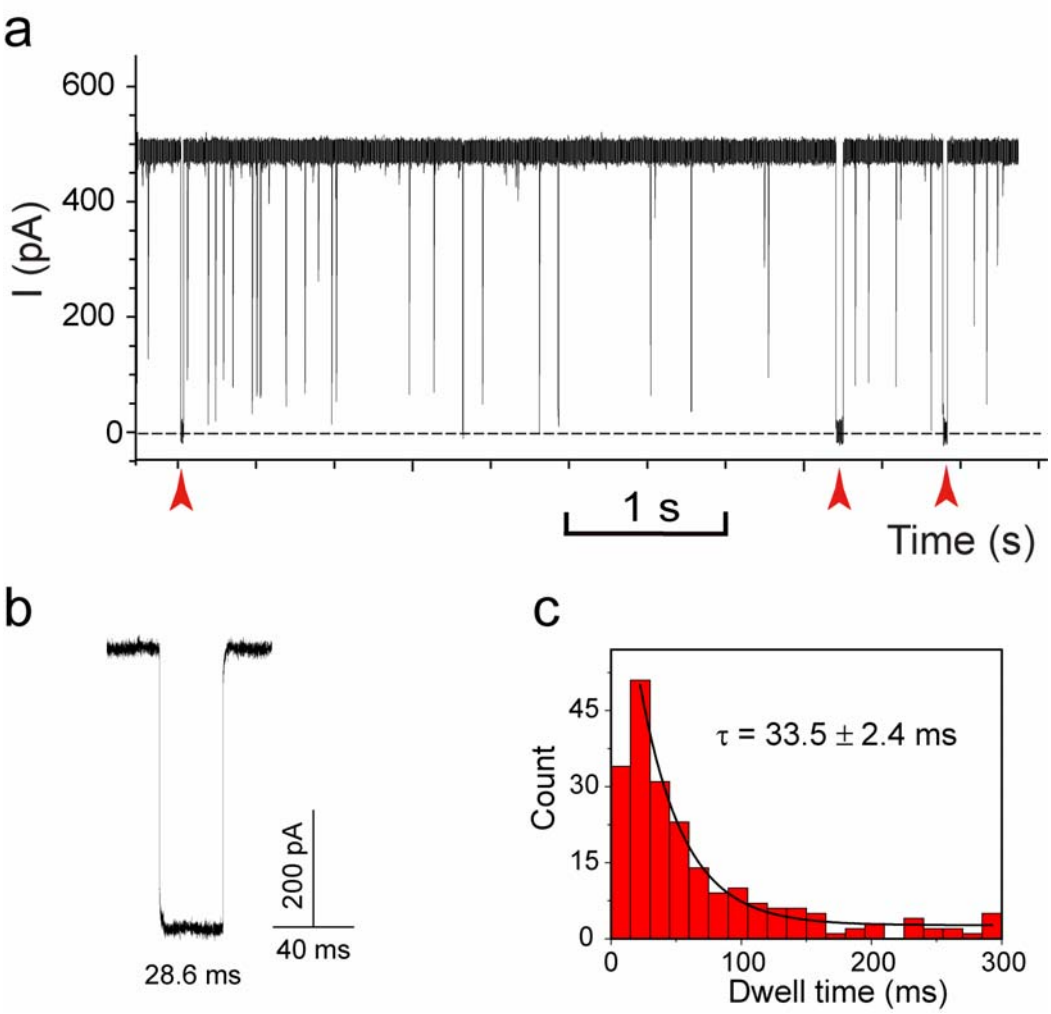


Fig. S11 Translocation of DNA1-Ad⊂βCD through αHL. (a) A representative current trace of the translocation of DNA1-Ad⊂βCD (final concentration 0.25 μM) through αHL. Data acquired in the buffer of 3 M KCl and 10 mM Tris, pH 8.0, with the transmembrane potential held at +160 mV. Red arrows indicate the long current events. (b) Expanded view of one long current event. (c) The dwell time histogram of the long events.

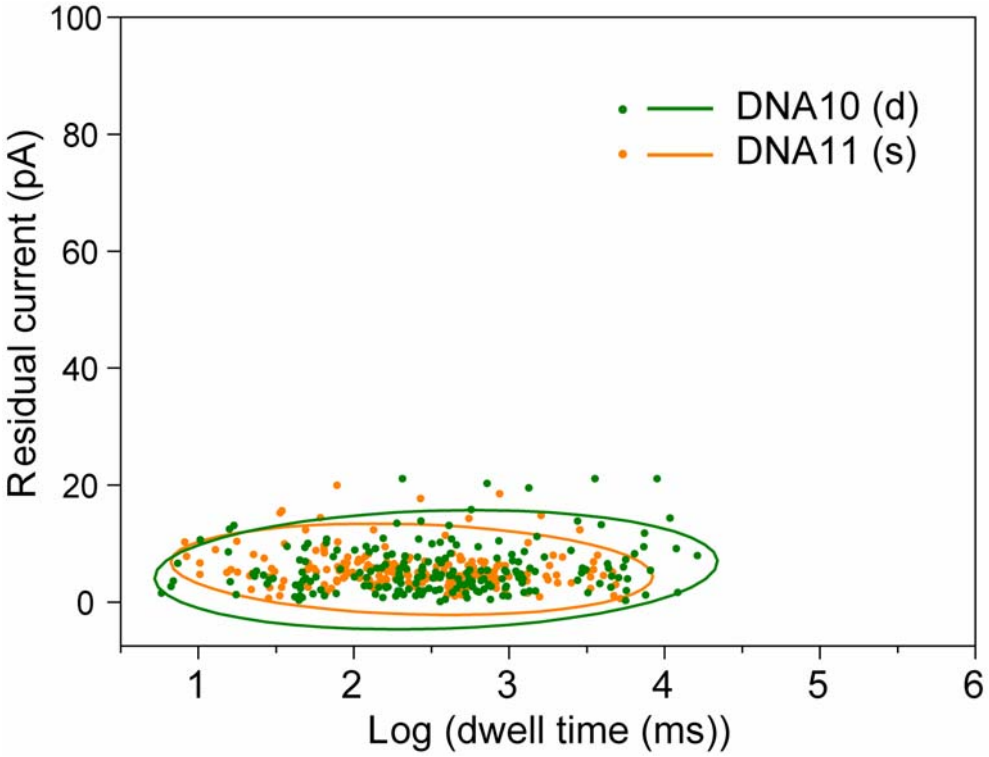


Fig. S12 Comparison of the translocation of modified singly-methylated DNA and doubly-methylated DNA. Scatter plots of the events show residual current *versus* event durations of Level 1 in the signature events. The green data points are from the translocation of doubly-methylated DNA10-Fc⊂CB[7] and the orange data points are from the translocation of singly-methylated control DNA11-Fc⊂CB[7]. The ovals contain >95% of the events of each DNA. Traces were recorded in 3 M KCl buffered with 10 mM Tris, pH 8.0. The transmembrane potential was held at +200 mV.

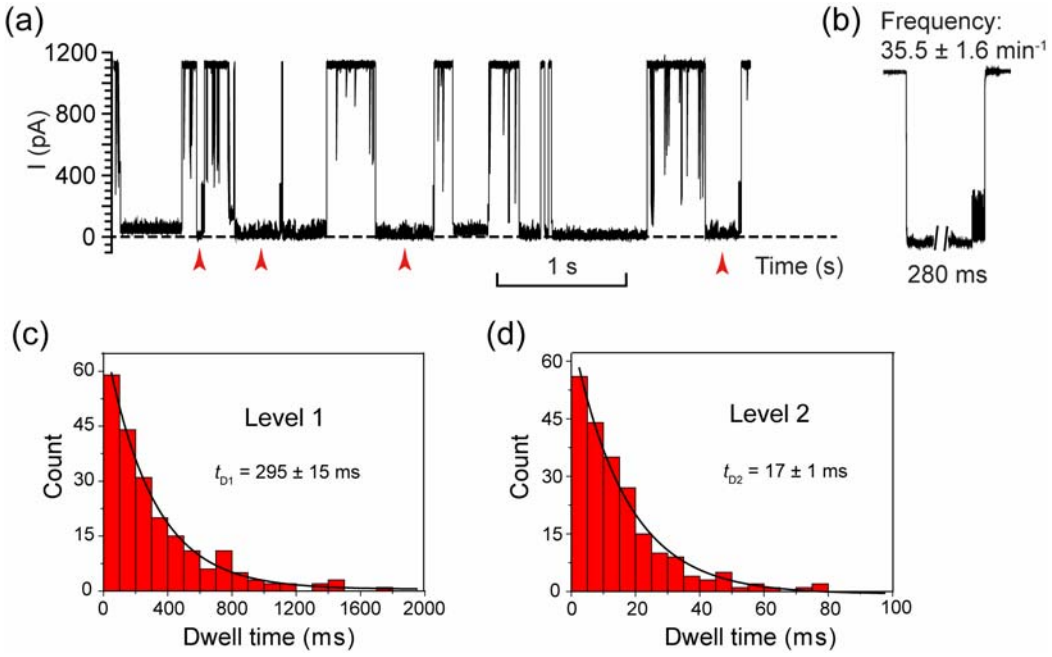


Fig. S13 Translocation of longer modified methylated DNA12 under elevated temperature. (a) A representative current trace of the translocation of DNA12-Fc⊂CB[7] (final concentration 0.25 μM) through αHL. Data were acquired in the buffer of 3 M KCl and 10 mM Tris, pH 8.0, at 46 ± 2 °C with the transmembrane potential held at +240 mV. Red arrows indicate the multi-level current events. (b) Expanded view of a typical current signature. (c) Dwell-time histogram for Level 1. The solid line in the histogram is a single exponential fit to the data. For Level 1, the mean duration t_{D1} = 295 ± 15 ms. (d) Dwell-time histogram for Level 2. The solid line in the histogram is a single exponential fit to the data. For Level 2, the mean duration t_{D2} = 17 ± 1 ms.

Supplementary tables

Table S1 Sequences of studied DNA oligomers

Name	Sequence
DNA1	5’-AAAAAAAAAA**(5mC)**AAAAAAAAAA-3’
DNA2	5’-AAAAAAAAAA**C**AAAAAAAAAA-3’
DNA3	5’-AAAAAAAAAA**T**AAAAAAAAAA-3’
DNA4	5’-AAAAAAAAAA**G**AAAAAAAAAA-3’
DNA5	5’-AAAAAAAAAA**A**AAAAAAAAAA-3’
DNA6	5’-AAAAAAAAAA**(5hmC)**AAAAAAAAAA-3’
DNA7	5’-TATGA**C**CTGA**(5mC)**TAGATA**C**G**C**T-3’
DNA8	5’-TATGA**C**CTGA**(5hmC)**TAGATA**C**G**C**T-3’
DNA9	5’-TATGA**C**CTGA**C**TAGATA**C**G**C**T-3’
DNA10	5’-AAAAAAAAAA**(5mC)**AAAAAAAAAA**(5mC)**AAAAAAAAAA-3’
DNA11	5’-AAAAAAAAAA**(5mC)**AAAAAAAAAAAAAAAAAAAA-3’
DNA12	5’-AAAAAAAAAAAAAAAAAAAA**(5mC)**AAAAAAAAAAAAAAAAAAAA-3’

(5hmC)—5-hydroxymethylcytosine; **(5mC)**—5-methylcytosine

Table S2 Mean dwell time of Level 1 and Level 2 in different concentrations of electrolyte (transmembrane potential +200 mV)

		3 M KCl	2 M KCl	1 M KCl
Event duration (ms)	Level 1	609.2 ± 29.6	448.1 ± 19.8	423.8 ± 25.1
	Level 2	**68.4 ± 3.4**	**22.9 ± 1.1**	**0.81 ± 0.04**

Supplementary references

53. S. Cheley, G. Braha, X. F. Lu, S. Conlan and H. Bayley, *Protein Sci.*, 1999, **8**, 1257-1267.

54. L. Z. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley and J. E. Gouaux, *Science*, 1996, **274**, 1859-1866.