Supplementary Information

Identification of Epigenetic DNA Modifications with a Protein Nanopore

Emma V. B. Wallace, David Stoddart, Andrew J. Heron, Ellina Mikhailova, Giovanni Maglia, Timothy J. Donohoe, Hagan Bayley*

Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, OX1 3TA, Oxford, UK

Email : hagan.bayley@chem.ox.ac.uk

Methods

Protein Preparation. Heptameric αHL E111N/K147N (NN) and E111N/K147N/M113Y (NNY) were produced as described.[1] The mutant genes were prepared by using a kit for site-directed mutagenesis (QuikChange II XL, no. 200533-5, Stratagene), and the DNA sequences were verified.

Planar Bilayer Recordings. Electrical recordings were carried out in a planar lipid bilayer apparatus.[2] Bilayers of 1,2-diphytanoyl-*sn*-glycerol-3-phosphocholine (DPhPC, Avanti Polar Lipids) were formed across an aperture in a 25 μm-thick polytetrafluoroethylene film (Teflon) (Goodfellow Cambridge, no. FP301200) that separated the *cis* and the *trans* compartments of the apparatus. The aperture was first pre-treated with hexadecane in *n*-pentane (10 mg mL^{-1}). Electrolyte solution (0.5 mL of 1 M KCl, 25 mM Tris.HCl, pH 8.0, containing 0.1 mM EDTA) was then added to each compartment. DPhPC dissolved in *n*-pentane (10 mg mL^{-1}) was added to both surfaces and the solvent was allowed to evaporate. The bilayer was formed by the lowering and raising of the electrolyte level past the aperture.

Current recordings were performed by using a patch clamp amplifier (Axopatch 200B, Axon Instruments) with the ground connected to the *cis* compartment. The NN or NNY protein was added to the *cis* compartment and after a single insertion event had occurred, DNA was added to the same side at a final concentration of 200 nM. Biotinylated

oligonucleotides were obtained as lyophilized pellets from ATDBio Ltd. Solutions of the biotinylated oligonucleotides (100 μM in 10 mM Tris.HCl, pH 8.0, 0.1 mM EDTA) were mixed with equal volumes of streptavidin (25 μM) (Sigma-Aldrich) in the same buffer and incubated for at least 5 min before use. Using the Episodic Stimulation mode in the pClamp software (version 10.1, Molecular Devices), a Digidata 1440A digitizer (Molecular Devices) was used to apply repeated voltage steps to drive a DNA•streptavidin complex into the pore (positive potential) and then eject the DNA•streptavidin complex (negative potential). Initially, +110 mV or +160 mV was applied to the *trans* side for 900 ms to drive the negatively charged DNA into the pore. We eventually settled with +110 mV, as we found nucleobase discrimination at R_1 to be optimal under these conditions. The capture of a DNA strand by the αHL pore was observed as a stepwise decrease in the open pore current level (I_O) to a lower current level (I_B). A voltage of –140 mV was then applied for 50 ms to eject the immobilized DNA from the pore. The applied potential was then stepped to 0 mV for 50 ms. The sequence was repeated for at least 100 cycles for each DNA species added. The signal was low-pass filtered at 1 kHz and sampled at 5 kHz with a computer equipped with a Digidata 1440A digitizer (Molecular Devices).

Data Analysis. Data were analyzed and presented by using pClamp (version 10.1, Molecular Devices). Single-channel searches were performed to find the average current level for each ssDNA blockade event (I_B). The mean I_B value for each oligonucleotide was determined by performing a Gaussian fit to a histogram of the I_B values. The current blockade for each oligonucleotide was expressed as a percentage of the open pore current (I_O): $I_{RES} = (I_B/I_O) \times 100$. When comparing several oligonucleotides, a single oligonucleotide species was first added to the *cis* chamber and the current trace required to determine I_B was recorded. A second (and if required, a third, a fourth, a fifth and a sixth) oligonucleotide species was added and additional currents recorded. When these experiments were repeated the oligonucleotides were added to the chamber in a different order.

Notes and references

1. S. Cheley, O. Braha, X. Lu, S. Conlan and H. Bayley, *Protein Sci.*, 1999, **8**, 1257-1267.
2. M. Montal and P. Mueller, *Proc. Natl. Acad. Sci. USA*, 1972, **69**, 3561-3566.

Exp.	Io, pA	I_{RES} CpG, %	I_{RES} GpG, %	ΔI_{RES} GpG, %	I_{RES} ApG, %	ΔI_{RES} ApG, %	I_{RES} hmCpG, %	ΔI_{RES} hmCpG, %	I_{RES} 5mCpG, %	ΔI_{RES} 5mCpG, %	I_{RES} TpG,%	ΔI_{RES} TpG, %
1	113	30.0	27.8	−2.2	27.4	−2.6	23.6	−6.4	21.3	−8.7	17.1	−12.9
2	121	29.4	27.3	−2.1	26.8	−2.6	23.4	−6.0	21.3	−8.1	17.3	−12.1
3	113	29.7	27.5	−2.2	27.0	−2.7	23.5	−6.2	21.4	−8.3	17.8	−11.9
4	114	29.4	27.1	−2.3	26.7	−2.7	23.1	−6.3	20.8	−8.6	16.8	−12.6
Mean	115	29.6	27.4	−2.2	27.0	−2.7	23.4	−6.2	21.2	−8.4	17.3	−12.4
SD	4	0.3	0.3	0.1	0.3	0.1	0.2	0.2	0.3	0.3	0.4	0.5

Table S1 The open pore (I_O) and residual current levels (I_{RES}) produced when NNY pores are blocked with the XpG_{R1} oligonucleotides, where X represents G, A, T, C, 5mC or hmC at position 9 (Fig. 2a, main text). ΔI_{RES} is defined as $I_{RES}^{XpGR1} - I_{RES}^{CpGR1}$.

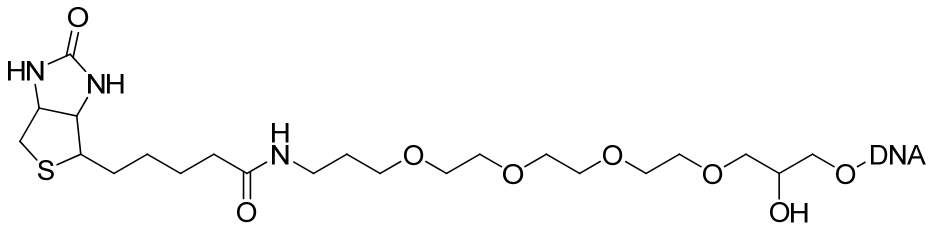

Fig. S1 The chemical structure of the biotin-TEG linker used to biotinylate the 3' terminus of the DNA oligonucleotides. The structure was produced with ChemBioDraw Ultra 11.

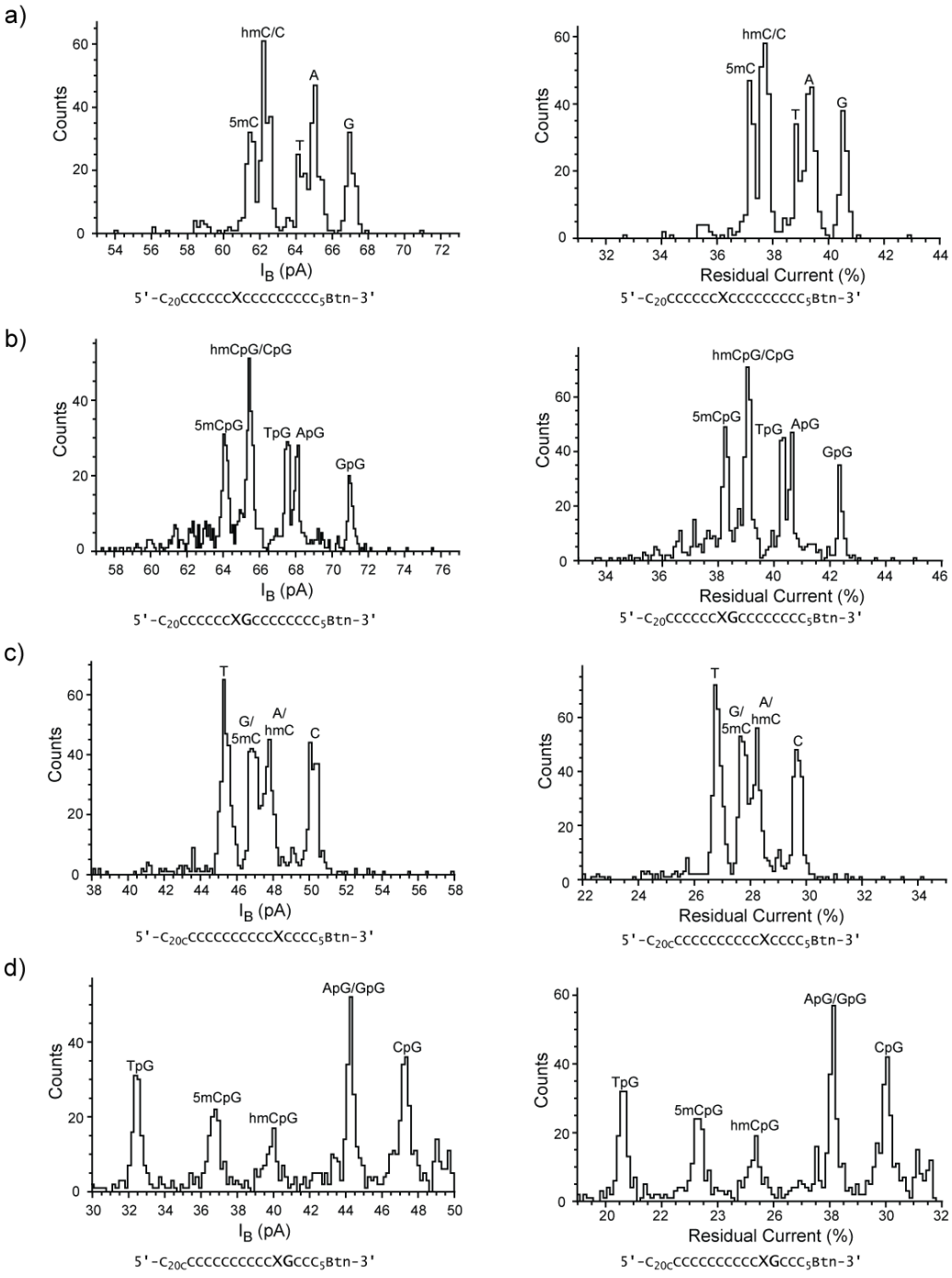

Fig. S2 Initial investigations to determine the ability of the αHL pore to discriminate between modified bases in immobilized oligonucleotides. The histograms shown are from typical experiments (each conducted at least 3 times), where the DNA was captured at +160 mV. The left-hand column displays I_B values and the right-hand column I_{RES} values. (a) NN pores probed with X_{R2} oligonucleotides. (b) NN pores probed with XpG_{R2} oligonucleotides. (c) NNY pores probed with X_{R1} oligonucleotides. (d) NNY pores probed with XpG_{R1} oligonucleotides.

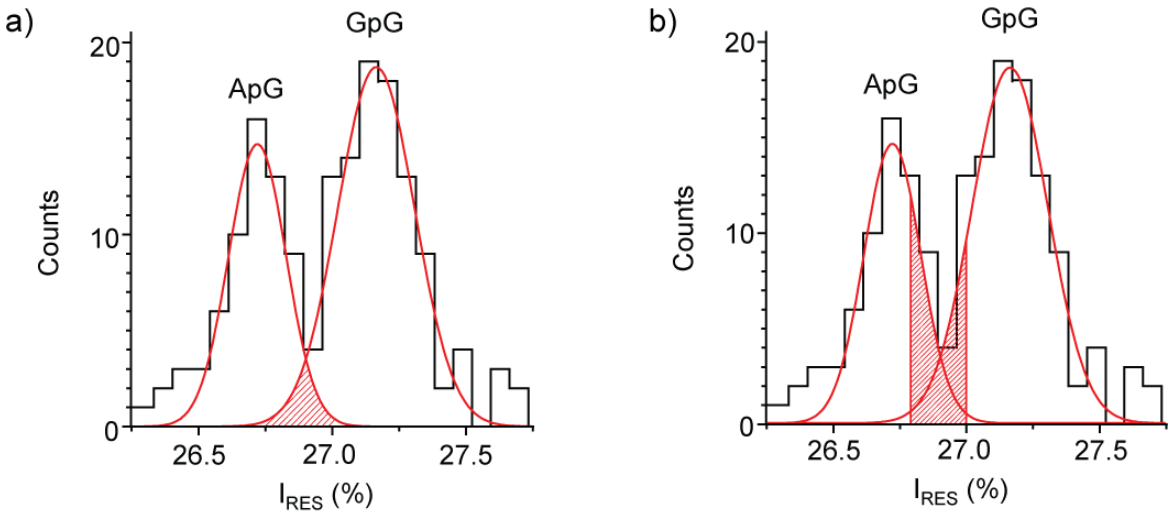

Fig. S3 Gaussian fits to the ApG and GpG histogram populations from an experiment with the NNY pore and oligonucleotides XpG_{R1} at +110 mV. We determined the overlap between the ApG_{R1} and GpG_{R1} populations from Gaussian fits to the histogram peaks. The data are binned at 0.07% I_{RES}, and a double Gaussian function is fitted to the data, with a constraint fixing the y-offset to zero. (a) From the Gaussian fits, we calculated that the two distributions share an overlapping area (shaded) of 7.6% of the total area of the two peaks. (b) For a more meaningful measure of the ability to accurately call nucleobases, we also calculated that the region where nucleotides cannot be called with more than 95% confidence (shaded) contains ~17% of the total events.