Published in final edited form as:

Chem Commun (Camb). 2010 November 21; 46(43): 8195–8197. doi:10.1039/c0cc02864a.

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, and Hagan Bayley*

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

Abstract

Two DNA bases, 5-methyl cytosine (5mC) and 5-hydroxymethylcytosine (hmC), marks of epigenetic modification, are recognized in immobilized DNA strands and distinguished from G, A, T and C by nanopore current recording. Therefore, if further aspects of nanopore sequencing can be addressed, the approach will provide a means to locate epigenetic modifications in unamplified genomic DNA.

5-Methylcytosine (5mC) is the most common epigenetic modification in eukaryotic genomes. Roughly 5% of cytosines in the human genome are methylated, mainly at CpG sites,¹⁻³ and there is considerable variation in the pattern of methylation with cell type and state.^{4, 5} Changes in methylation patterns are required for development^{5, 6} and are associated with tumorigenesis and other disease states.² A second rarer epigenetic modification was reported recently: 5-hydroxymethylcytosine (hmC),⁷⁻⁹ which arises from the oxidation of 5mC by the enzyme TET1.⁸ The role of hmC is under investigation. It has already been found to be enriched in the hippocampus and cortex of mouse brain,¹⁰ and hmC turnover has been implicated in embryonic stem cell maintenance and differentiation.¹¹

Techniques have been developed to determine the sites of DNA methylation in genomic DNA, but recent developments that combine bisulfite chemistry with second generation sequencing is quickly supplanting them.⁴ Bisulfite treatment results in the conversion of C to U by deamination, while 5mC remains unchanged. In the new genome-wide approaches, genomic DNA is fragmented, treated with bisulfite, amplified and subjected to deep sequencing, revealing both the sites and extent of methylation. The complexity of the sample may be reduced by the enrichment of methylated sequences (e.g. with anti-5mC antibodies) or by the selective amplification of target sequences. The major disadvantage of this technique is incomplete conversion of C to U, which results in false positives. As yet, there are no reliable methods for the localization of hmC in a genome sequences.¹²⁻¹⁴

Over the last 5 years, second generation sequencing has revolutionized genomics by dramatically increasing sequencing speeds and reducing costs.¹⁵⁻¹⁷ Nevertheless, a third generation of sequencers based on single-molecule approaches is under development, which will have further advantages including greatly reduced cycle times for each base and the ability to read unamplified DNA.^{18, 19} Such devices based on fluorescence-detected sequencing-by-synthesis are now reaching the market.^{20, 21} 5mC and hmC have been detected indirectly by this means, by relying on changes in polymerase kinetics near a

†Electronic supplementary information (ESI) available: Methods and Figs. S1 and S2. See DOI: xxxxxxxxxxxxxxxxx

*hagan.bayley@chem.ox.ac.uk.

modified base, but the error rate is appreciable and multiple passes over a circular DNA template are required.²²

Nanopore sequencing is another promising third generation approach^{18, 23} with the potential to determine modified bases in unamplified genomic DNA. Two primary means for nanopore sequencing have been proposed.^{18, 24} In one of these, exo sequencing, a DNA strand is digested by a processive exonuclease to release nucleoside monophosphates that are identified by the nanopore.^{25, 26} Towards this end, nucleotides, including the 5mC monophosphate,¹⁶ have been identified with an α -hemolysin (α HL) protein nanopore equipped with a cyclodextrin adapter.^{25, 26} In a second proposed approach, strand sequencing, which is addressed in this paper, an intact DNA strand is fed through a nanopore and the bases are read off directly.^{23, 24}

At present, the speed at which DNA is translocated through protein nanopores is too fast for direct strand sequencing (around 1-3 μ s per nucleobase)²⁷ and means to slow the DNA are under intense investigation.²⁸⁻³⁰ In the interim, differences in ionic current arising from single nucleobase substitutions in immobilized ssDNA are being used to probe base identification.³¹⁻³⁴ Upon immobilization of the DNA within the α HL pore, the open pore current (I_O) is reduced to a new level (I_B). The residual current (I_{RES}) is expressed as a percentage of the open-pore current: $I_{RES} = (I_B/I_O) \times 100$. When DNA is immobilized through a streptavidin-biotin complex, three recognition points within the β barrel of α HL, termed R_1 , R_2 and R_3 , are capable of distinguishing individual unmodified nucleobases in a DNA strand (Fig. 1).³³ The α HL E111N/K147N (NN) and E111N/K147N/M113Y (NNY) mutants have previously been shown to give superior discrimination between the standard bases (G, A, T and C) at R_2 and R_1 , respectively.^{33, 34} The aim of the present study was to establish whether nanopore technology is able to discriminate epigenetic modifications within DNA strands. We now report conditions under which NNY can distinguish between the standard bases and 5mC and hmC, at R_1 , which required the difficult and unprecedented separation of six distinct I_{RES} levels.

First, recognition point R_2 within the NN pore was probed with six poly(dC) oligonucleotides (X_{R2}) that contained 5mC, hmC, G, A, T and C substitutions at position 14 (relative to the 3' biotin tag, Fig. S1a), a position known to occupy R_2 . A second set of six oligonucleotides was obtained that contained in addition a G substitution at position 13 (Xp_{R2}), to emulate the CpG context in which 5mC, and presumably hmC, are most often found (Fig. S1b). Unfortunately, neither set of six oligonucleotides yielded the desired six distinct current levels, when elicited at +160 mV.

Similar sets of oligonucleotides were then obtained to probe recognition point R_1 . The oligonucleotides in the first of these sets contained nucleobase substitutions at position 9 (Fig. 1) and were termed the 5mC $_{R1}$, hmC $_{R1}$, G $_{R1}$, A $_{R1}$, T $_{R1}$ and C $_{R1}$ (X_{R1}) oligonucleotides, while the oligonucleotides in the second set contained in addition a G nucleobase at position 8 and were termed the 5mCpG $_{R1}$, hmCpG $_{R1}$, GpG $_{R1}$, ApG $_{R1}$, TpG $_{R1}$ and CpG $_{R1}$ (Xp_{R1}) oligonucleotides (Fig. 2a). The NNY pore was probed with each set of oligonucleotides. Again, neither set produced six distinct current levels at +160 mV. However, the dispersion between the two most widely separated oligonucleotide blockade levels in the Xp_{R1} set (Fig. S2d) was larger than for the X_{R1} set (Fig. S2c) and only the GpG $_{R1}$ and ApG $_{R1}$ oligonucleotides gave overlapping levels (Fig. S2d). In an attempt to enhance the separation between these two sequences, applied potentials from +100 mV to +180 mV were examined and it was discovered that discrimination was improved at +110 mV. Indeed, at this potential, each Xp_{R1} yielded a distinct current level (Fig. 2). Under these conditions, when immobilized within NNY, the difference in residual current levels (ΔI_{RES}) between the 5mCpG $_{R1}$ oligonucleotide and the CpG $_{R1}$ oligonucleotide is $-8.3 \pm$

0.1% ($n = 4$) ($\Delta I_{\text{RES}} = I_{\text{RES}}^{5\text{mCpGR1}} - I_{\text{RES}}^{\text{CpGR1}}$). It is therefore possible to distinguish two bases in a DNA strand that differ only in a methyl group at the C5 position (Fig. 2). The hmCpG oligonucleotide also produced a distinct current block. The difference in the residual current levels between hmCpGR1 and CpGR1 is $\Delta I_{\text{RES}} = -6.1 \pm 0.1\%$ ($n = 4$). While the molecular basis for the observed differences in I_{RES} levels is unclear, the current levels for hmCpG and 5mCpG are distinct from all other bases.

Present techniques for mapping 5mC are tedious and usually yield a consensus methylation pattern, while the actual pattern may vary from cell to cell. By using nanopore technology, we have demonstrated that it is possible to distinguish directly all four standard bases, 5mC and hmC in individual ssDNAs. Therefore, if this technology can be integrated into a practicable nanopore device, it will benefit fundamental studies in epigenetics and medical diagnostics. In particular, the absence of bisulfite chemistry and DNA amplification will reduce errors, and there is the potential for long sequence reads that will facilitate methylation haplotyping.³⁵

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from the NIH, the European Commission's seventh Framework Programme (FP7) READNA Consortium and Oxford Nanopore Technologies. E.V.B.W. and D.S. are supported by BBSRC Doctoral Training Grants.

Notes and references

1. Herman JG, Baylin SB. *N Engl J Med.* 2003; 349:2042–2054. [PubMed: 14627790]
2. Brena RM, Huang TH, Plass C. *J Mol Med.* 2006; 84:365–377. [PubMed: 16416310]
3. Beck S, Rakyan VK. *Trends Genet.* 2008; 24:231–237. [PubMed: 18325624]
4. Lister R, Ecker JR. *Genome Res.* 2009; 19:959–966. [PubMed: 19273618]
5. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. *Nature.* 2009; 462:315–322. [PubMed: 19829295]
6. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. *Nature.* 2008; 454:766–770. [PubMed: 18600261]
7. Kriaucionis S, Heintz N. *Science.* 2009; 324:929–930. [PubMed: 19372393]
8. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. *Science.* 2009; 324:930–935. [PubMed: 19372391]
9. Loenarz C, Schofield CJ. *Chem Biol.* 2009; 16:580–583. [PubMed: 19549596]
10. Munzel M, Globisch D, Bruckl T, Wagner M, Welzmler V, Michalakakis S, Muller M, Biel M, Carell T. *Angew Chem Int Ed Engl.* 2010; 49:5375–5377. [PubMed: 20583021]
11. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. *Nature.* 2010; 466:1129–1133. [PubMed: 20639862]
12. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. *PLoS One.* 5:e8888. [PubMed: 20126651]
13. Jin SG, Kadam S, Pfeifer GP. *Nucleic Acids Res.* 2010; 38:125–131.
14. Nestor C, Ruzov A, Meehan R, Dunican D. *Biotechniques.* 2010; 48:317–319. [PubMed: 20569209]
15. Schloss JA. *Nature Biotechnology.* 2008; 26:1113–1115.
16. Mardis ER. *Genome Med.* 2009; 1:40. [PubMed: 19435481]

17. Venter JC. *Nature*. 2010; 464:676–677. [PubMed: 20360717]
18. Bayley H. *Curr Opin Chem Biol*. 2006; 10:628–637. [PubMed: 17113816]
19. Gupta PK. *Trends Biotechnol*. 2008; 26:602–611. [PubMed: 18722683]
20. Pushkarev D, Neff NF, Quake SR. *Nat Biotechnol*. 2009; 27:847–852. [PubMed: 19668243]
21. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. *Science*. 2009; 323:133–138. [PubMed: 19023044]
22. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. *Nat Methods*. 2010; 7:461–465. [PubMed: 20453866]
23. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wigginton M, Schloss JA. *Nature Biotechnology*. 2008; 26:1146–1153.
24. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. *Proc.Natl.Acad.Sci.USA*. 1996; 93:13770–13773. [PubMed: 8943010]
25. Astier Y, Braha O, Bayley H. *J Am Chem Soc*. 2006; 128:1705–1710. [PubMed: 16448145]
26. Clarke J, Wu H, Jayasinghe L, Patel A, Reid S, Bayley H. *Nature Nanotechnology*. 2009; 4:265–270.
27. Meller A, Nivon L, Brandin E, Golovchenko J, Branton D. *Proc.Natl.Acad.Sci.USA*. 2000; 97:1079–1084. [PubMed: 10655487]
28. Cockroft SL, Chu J, Amarin M, Ghadiri MR. *J Am Chem Soc*. 2008; 130:818–820. [PubMed: 18166054]
29. de Zoysa RS, Jayawardhana DA, Zhao Q, Wang D, Armstrong DW, Guan X. *J Phys Chem B*. 2009; 113:13332–13336. [PubMed: 19736966]
30. Kawano R, Schibel AE, Cauley C, White HS. *Langmuir*. 2009; 25:1233–1237. [PubMed: 19138164]
31. Ashkenasy N, Sánchez-Quesada J, Bayley H, Ghadiri MR. *Angew.Chem.Int.Ed.Engl*. 2005; 44:1401–1404. [PubMed: 15666419]
32. Purnell RF, Schmidt JJ. *ACS Nano*. 2009; 3:2533–2538. [PubMed: 19694456]
33. Stoddart D, Heron A, Mikhailova E, Maglia G, Bayley H. *Proc. Natl. Acad. Sci. USA*. 2009; 106:7702–7707. [PubMed: 19380741]
34. Stoddart D, Maglia G, Mikhailova E, Heron A, Bayley H. *Angew Chem Int Ed*. 2010; 49:556–559.
35. Shoemaker R, Deng J, Wang W, Zhang K. *Genome Res*. 2010; 20:883–889. [PubMed: 20418490]

Fig. 1.

Oligonucleotides immobilized inside the α HL protein nanopore. Schematic representation of a DNA oligonucleotide (blue circles) threaded from the 5' end and immobilized inside an α HL pore (grey, cross-section) through the use of a biotin (yellow)-streptavidin (green) linkage. The lumen of the α HL pore can be divided into two halves, each approximately 5 nm in length; an upper vestibule located between the *cis* entrance and the central constriction, and a fourteen-stranded, transmembrane, antiparallel β barrel, located between the central constriction and the *trans* entrance. The three nucleotide recognition points, R_1 , R_2 and R_3 are indicated.³³ R_1 is located near the central constriction and is capable of recognizing bases at positions 8 to 11 (relative to the 3'-biotin tag, Fig. S1). R_2 is located near the center of the barrel and recognizes bases 13 to 16. The single nucleotide substitutions investigated in this study were made at position 9 (red circle) and position 14 (pink circle). Cytosine in genomic sequences can be epigenetically modified to form 5mC and hmC (right).

Fig. 2.

The NNY pore discriminates between modified nucleotides at R_1 . (a) Histogram (top) of residual current levels for NNY pores interrogated with six DNA strands (middle) that differ at only one position X (bold), where X represents G, A, T, C, 5mC or hmC. Gaussian fits were performed for each peak, and the mean value of the residual current for each oligonucleotide (with the standard deviation) is tabulated (bottom). The result displayed is from a typical experiment carried out at an applied potential of +110 mV. See Table S1 for full list of experimental results. (b) Typical current levels for the NNY pore when blocked at +110 mV with ApGR₁ and GpGR₁ oligonucleotides (left, filtered at 50 Hz for display). Event histogram (right) displaying the subset of residual current levels produced by ApGR₁ and GpGR₁ oligonucleotide blockades. The shaded area is the region where the fitted Gaussians overlap and comprises 7.6% of the area of the two Gaussians (see also Fig. S3).