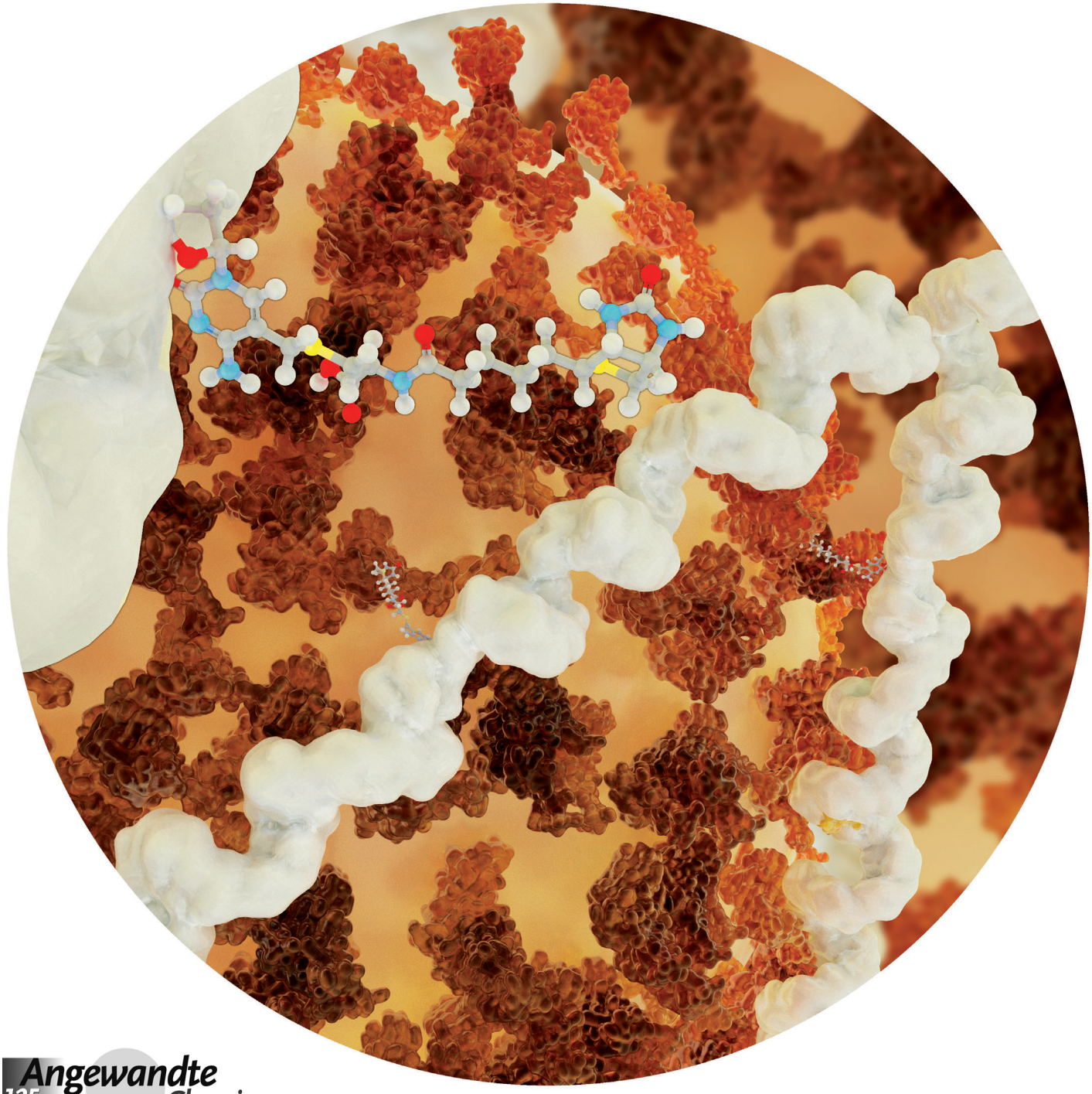
VIP Epigenetic Markers

DOI: 10.1002/ange.201300413

Single-Molecule Detection of 5-Hydroxymethylcytosine in DNA through Chemical Modification and Nanopore Analysis**

*Wen-Wu Li, Lingzhi Gong, and Hagan Bayley**



DNA methylation is an important epigenetic modification of the genome. For example, 5-methylcytosine (5mC) plays a key role in the regulation of gene expression associated with development and tumorigenesis.[1,2] Recently, 5-hydroxymethylcytosine (5hmC),[3,4] 5-formylcytosine (5fC),[5–8] and 5-carboxylcytosine[6,7] have also been found in mammalian DNA. Changes in the level and distribution of 5hmC bases may be associated with the maintenance and differentiation of embryonic stem (ES) cells[8] and transcriptional regulation in those cells.[9] 5hmC has also been shown to play a role in the regulation of chromatin structure and gene expression in central nervous system cells.[10]

It is important, therefore, to develop methods for the detection of epigenetic bases in genomic DNA.[11] A method based on bisulfite treatment, PCR amplification, and nucleotide sequence analysis has become routine for the detection of 5mC in DNA.[12,13] Bisulfite treatment converts cytosine (C) to uracil (U; Figure 1a), whereas 5mC remains intact (Figure 1b).[14–16] However, this method is unable to differentiate 5hmC[17,18] from 5mC, because 5hmC is converted into cytosine 5-methylenesulfonate (CMS) during the process,[19] which behaves as C residues during PCR amplification (Figure 1c). To overcome this problem, an oxidative bisulfite sequencing method has been developed recently to locate 5hmC with single base resolution.[20] Selective chemical oxidation of 5hmC to 5fC by potassium perruthenate ($KRuO_4$) enables bisulfite conversion of 5fC into uracil. Sequences of oxidized/bisulfite-treated DNA are compared with untreated and bisulfite-only sequences to distinguish C, 5mC, and 5hmC. In a second combined enzymatic and chemical approach, 5hmC is converted into C, and C and 5mC into U, before sequencing.[21] The enrichment of 5hmC-containing DNA for targeted sequencing has also been accomplished by enzymatic modification of 5hmC with β-glucosyltransferase followed by chemical modification with a biotin tag.[9,22–24] DNA methyltransferase-directed thiol and selenol substitution of the hydroxy group of 5hmC in DNA has also enabled subsequent modification with a biotin- or fluorescein-containing group.[25] However, the use of expensive enzymes with restricted recognition motifs, the need for the synthesis of analogues of enzyme co-factors or substrates, and the multiple steps required for nucleobase modification are drawbacks of these approaches.

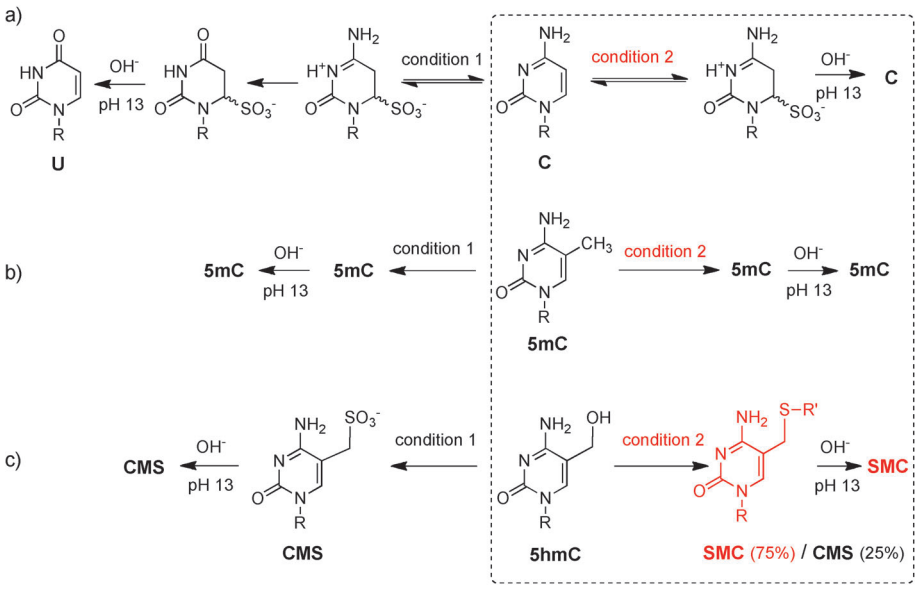


Figure 1. Bisulfite reactions of cytosine derivatives. a) Bisulfite (>4.0 M) at high temperature (condition 1) converts C into U;[12] bisulfite (0.05–0.6 M) in the presence of excess of thiol at lower temperature (condition 2) does not affect C. b) Under either condition 1 or 2, 5mC is unreactive. c) Under condition 1, 5hmC is converted into CMS; under condition 2, 5hmC is converted into CMS as the minor product and SMC as the major product.

Single-molecule sequencing of modified bases in unamplified DNA is an attractive alternative to this method. Single-molecule, real-time differentiation of 5hmC and 5mC has been achieved based on their influence on the rate of a polymerase engaged in sequencing-by-synthesis, but the signals are context-dependent and the error rates are high.[26] Nanopore methods are a promising alternative for single-molecule DNA sequencing[27,28] and sequencing of the four canonical bases (G, A, C, and T) has recently been accomplished.[29–31] In the experiments that demonstrated the feasibility of base identification for strand sequencing, biotinylated DNA strands were immobilized within an engineered protein pore after binding to streptavidin, which prevented translocation through the pore.[32–34] In this way, 5hmC and 5mC could be differentiated from G, A, C, or T.[35] Herein, we report an alternative approach in which 5hmC is chemically modified in situ, permitting the enrichment of rare 5hmC-containing sequences and their identification with a protein nanopore at the single-molecule level.

Bisulfite converts 5hmC into CMS by transformation of the 5-hydroxy group to a sulfonate (Figure 1c; Supporting

[*] Dr. W. W. Li, L. Gong, Prof. H. Bayley
Department of Chemistry, University of Oxford
Mansfield Road, Oxford, OX1 3TA (UK)
E-mail: hagan.bayley@chem.ox.ac.uk
Dr. W. W. Li
Institute for Science and Technology in Medicine, Keele University
Thornburrow Drive, Stoke-on-Trent, ST4 7QB (UK)

[**] We thank Oxford Nanopore Technology for financial support.

Supporting information for this article (including synthesis and characterization of sulfur-substituted 5hmC derivatives, the modification of 5hmC in ssDNA, UPLC-MS analysis, and single-channel electrical recordings) is available on the WWW under http://dx.doi.org/10.1002/anie.201300413.

Information Figure S1 a).[19] We speculated that CMS is produced through formation of an *exo*-methylene intermediate (Figure S1 b), similar to those implicated in the mechanisms of the enzymes T4 deoxycytidylate hydroxymethylase,[36] thiaminase,[37] and thymidylate synthase,[38] and subsequent attack by the nucleophilic sulfite anion (Figure S1 b). If this mechanism is correct, a nucleophilic thiolate should trap the *exo*-methylene intermediate to form the corresponding 5-thiomethyl derivative (Figure 1 c; Figure S1 b). We tested this idea by using two model compounds (5hmC and 5-hydroxymethyl-2′-deoxycytidine (5hmdC)). The reaction of 5hmC with excess of glutathione (GSH) or 1-thio-β-D-glucopyranose in the presence of bisulfite in D_2O at pD 5.0 yielded CMS (25%) and the sulfur-substituted 5hmC adducts (SMC, 75%; Figure S1 a). ^{1}H NMR spectroscopy (Figure S2) showed that the reactions were complete after 18 hours at pD 5.0 and 42 °C. The 5hmC and 5hmdC adducts were isolated by HPLC and characterized by UV/Vis absorption spectroscopy, negative and positive ion mode electrospray ionization MS, and NMR spectroscopy (Figures S3,S4; Tables S1,S2).

When the pD of the reaction was increased from 5.0 to 7.0, the reaction rate decreased (Figure S2). This suggests that protonation of the N-3 atom of 5hmC at pD 5.0, to initiate formation of the *exo*-methylene intermediate, is more important than deprotonation of the thiol ($pK_a = 9.42$ for GSH[39]), which would provide a higher concentration of the nucleophilic thiolate. In another possible pathway, bisulfite reacts with the 5-hydroxy group of 5hmC to form a sulfite ester that subsequently undergoes substitution by either a nucleophilic thiolate or a second sulfite anion (Figure S1 c). Under the same conditions, G, A, T, and 5mC (as nucleosides) were unaltered (data not shown). Remarkably, in the presence of GSH and bisulfite, the nucleobase C in cytidine underwent the initial attack by bisulfite to form 5,6-dihydrocytidine-6-sulfonate,[14,15] but did not undergo hydrolytic deamination to generate U at the relatively low concentration of bisulfite and low temperature that was used.[15,40] Thiol derivatives (e.g. glutathione) do not react with 2′-deoxycytidine and do not affect the formation of 5,6-dihydrocytidine-6-sulfonate (NMR data not shown). Bisulfite is known to participate in the deamination step.[40] During subsequent alkali treatment (adjusted to pH 13 with aqueous 10 N NaOH), the 5,6-dihydrocytosine-6-sulfonate nucleobase was transformed back to C rather than to U (Figure 1 a; Figure S5).

Next, the modification of 5hmC in single-stranded DNA (ssDNA) by bisulfite under conditions 1 and 2 (Figure 1) was carried out. Four 28-mer ssDNAs containing 5mC and 5hmC were used: 5′-ACTGTATCA**X**CTGGTCCTGTAT**Y**TAATA-3′; **1**, **X** = 5hmC and **Y** = 5mC; **2**, **X** = 5hmC and **Y** = C; **3**, **X** = 5hmC and **Y** = 5hmC; **4**, **X** = 5mC and **Y** = C. First, the reaction products of C and 5hmC after bisulfite treatment with a high concentration of bisulfite at a high temperature (condition 1, Figure 1) in these four ssDNAs were characterized by ultra-performance liquid chromatography mass spectrometry (UPLC-MS). Measurement of the masses of the unmodified ssDNAs, the stable intermediate products (after bisulfite treatment), and the final modification products (after alkali treatment; Figure S6) enabled us to determine the number of 5hmC and C bases as CMS and U, respectively (Figure S7, Table S3, and Supporting Information). CMS is stable under both acidic and basic conditions, whereas the 5,6-dihydrouracil-6-sulfonate intermediate is only stable under acidic or neutral conditions (Figure 1 c).[14,15] 5mC is unaltered in the overall reaction (Figure 1 b).

Next, the 5hmC of ssDNA-**1** was modified at pH 5.0 by bisulfite at a lower concentration and temperature in the presence of excess GSH at 42 °C for 36–48 hours (condition 2) followed by alkali treatment (Figure 1 and 2 a). UPLC-MS showed the formation of an adduct of GSH and ssDNA (ssDNA-**1**-SG, yield approximately 70%) and a sulfonated DNA adduct (ssDNA-**1**-CMS, yield approximately 30%; based on the relative intensities of the MS peaks; Figure 3 a,b). At 70 °C for four hours or 42 °C for 15 hours, the reactions were incomplete. No conversion of C into U occurred, and 5mC was also unchanged. In a control experiment with ssDNA-**4**, in which the 5hmC and 5mC of ssDNA-**1** are replaced by 5mC and C, respectively, no modification of any nucleobase (including C into U conversion) was observed, based on UPLC-MS analysis (data not shown). The chemistry of ssDNA-**1** and ssDNA-**4** is therefore consistent with that of the model mononucleotides (Figures S1,S5).

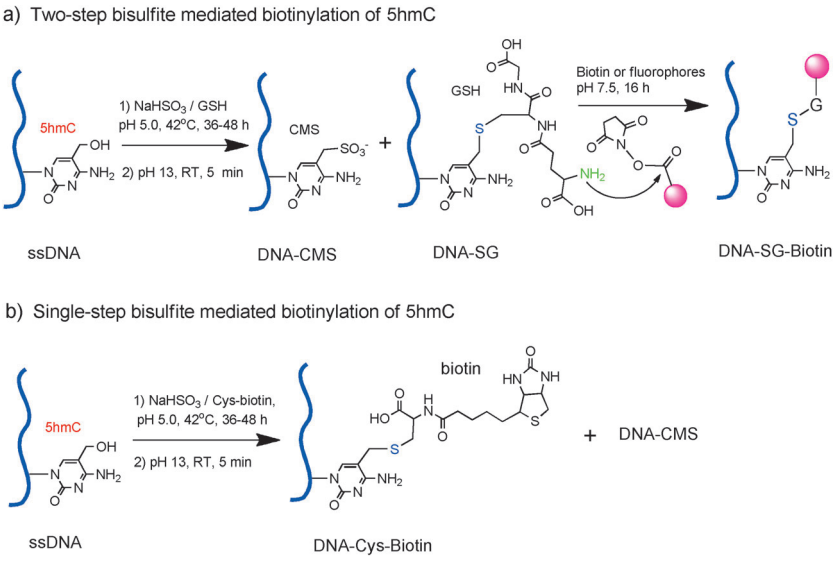


Figure 2. Bisulfite-mediated biotinylation of 5hmC in ssDNA. a) Modification of 5hmC with glutathione (condition 2, Figure 1) and subsequent modification of the primary amino group of the adduct. G, A, T, C and 5mC remain intact, while the 5-hydroxy group of 5hmC is substituted by sulfite to form DNA-CMS (≈30%) and DNA-SG (≈70%). In a second step, the primary amino group of glutathione is modified by biotin- or fluorescein-*N*-hydroxysulfosuccinimidyl ester (shown as spheres). b) Direct bisulfite-mediated biotinylation with *N*-biotinyl-L-cysteine (condition 2, Figure 1).

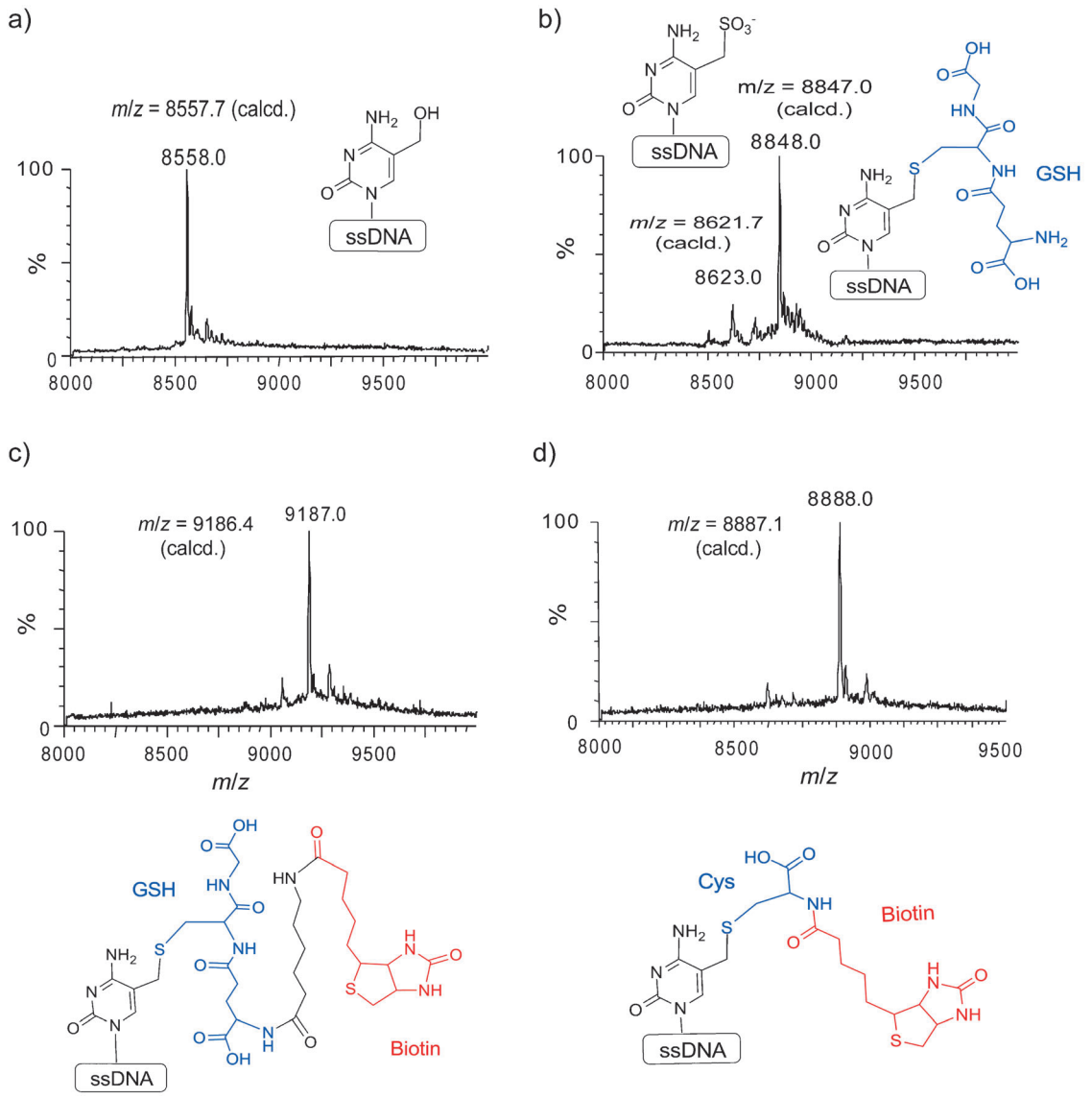


Figure 3. UPLC-MS characterization of ssDNA-**1**, and its modified products. a) Mass spectrum of unmodified ssDNA-**1**. b) Mass spectrum of a single UPLC peak containing a mixture of ssDNA-**1**-CMS and ssDNA-**1**-SG, which were not separated. c) Mass spectrum of a UPLC peak of a biotinamidyl hexanoate glutathione DNA adduct (ssDNA-**1**-SG-Biotin) from the two-step modification approach (Figure 2a). d) Mass spectrum of a biotinamidyl cysteinyl DNA (ssDNA-**1**-Cys-Biotin) adduct from the single-step modification (Figure 2b) of ssDNA-**1**.

Two approaches were explored for the incorporation of biotin into ssDNA at the 5hmC site. First, the N-terminal primary amine of the GSH in ssDNA-**1**-SG was selectively modified with biotin *N*-hydroxysulfosuccinimidyl esters[25,41] with yields ranging from 50% to 95%, depending on the length of the linker in the reagent (Figure 2a and 3c). The overall yield of biotinylation of 5hmC-containing DNA over two steps was 30–65%. Similarly, fluorescein can be incorporated using *N*-hydroxysuccinimidyl esters with yields of around 60% (data not shown). Second, a synthetic biotinyl cysteine derivative (Supporting Information) was used for a single-step incorporation of biotin at 5hmC sites in ssDNA-**1** and a 100-mer ssDNA with yields ranging from 35–55%, as demonstrated by UPLC-MS (Figure 3d; Figure S8). Importantly, the biotinylated ssDNAs can be enriched by using immobilized streptavidin (Figure S9).[25]

The wild-type staphylococcal α-hemolysin (αHL)[42] protein nanopore (which is not optimized for nucleobase detection)[33] was then used to detect modified 5hmC in a ssDNA, using the streptavidin immobilization approach.[33,34] The 40-mer ssDNA (DNA40-hmC, Figure 4a; Figure S10) contained a fragment of the POU5F gene (5′-TATACACAGGCXGATGTGGG-3′) from human embryonic stem cells,[13] in which a 5mC at position 9 with respect to the 3′-end (X) was replaced by 5hmC. Biotinylation at the 3′-end placed the 5hmC in the immobilized DNA40-hmC at the recognition site near the constriction in the lumen of the αHL pore.[33] In the presence of DNA40-hmC and streptavidin, the open pore current (I_O) was decreased to a new level (I_B; Figure 4b). The histogram of $I_{RES\%}$ values ($I_{RES\%} = (I_B/I_O)100$) shows a main peak with $I_{RES\%} = 11.8 \pm 0.1$ (Figure 4b,c), which is set at zero for further comparison with

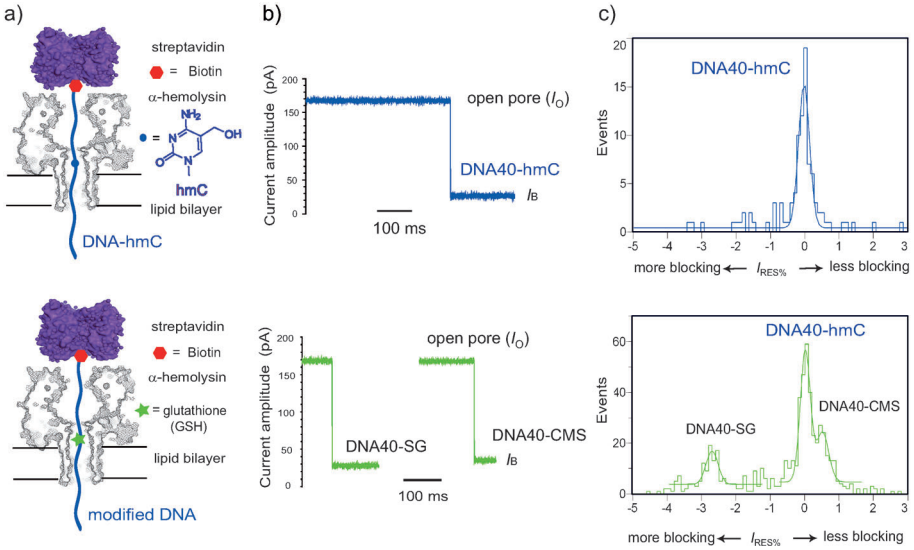


Figure 4. Detection of modified 5hmC in an immobilized ssDNA with the wild-type αHL pore. a) Scheme of the immobilized 40-mer ssDNA (5′-TTTTTTTTTTTTTTTTTTTTTATACACAGGCX-GATGTGGG-TEG-biotin-3′). X = 5hmC (DNA40-hmC, top) or X = glutathione adduct of 5hmC (DNA40-SG, bottom; Figure 2a and 3b). TEG-biotin is a biotin moiety attached to a 15-atom mixed polarity triethylene glycol spacer.[33] b) Current levels for an αHL pore before and after blockade with DNA40-hmC only (top) or DNA40-hmC, DNA40-CMS and DNA40-SG (bottom), after addition of the modified products to the recording chamber (Supporting Information). The recordings were made at +160 mV in 1 M KCl, 20 mM Tris·HCl, containing 0.1 mM EDTA, pH 8.0. c) Histograms of residual current levels for an αHL pore interrogated with DNA40-hmC (top), and an αHL pore interrogated with DNA40-hmC, DNA40-SG and DNA40-CMS (bottom). The value of $I_{RES\%}$ for DNA40-hmC was set equal to 0.

modified ssDNAs. When DNA40-hmC was modified by reaction with glutathione and bisulfite (Figure S10) and added to a recording chamber already containing DNA40-hmC, the histogram showed two additional peaks corresponding to a glutathione-ssDNA conjugate (DNA40-SG) and a sulfonated ssDNA conjugate (DNA40-CMS; Figure 4c). The peak with the larger shift ($\Delta I_{RES\%} = -2.7 \pm 0.1$) relative to DNA40-hmC was tentatively assigned to DNA40-SG, because of the larger size of GSH compared to the sulfonate group. In control experiments with the 40-mer ssDNAs where position 9 was 5mC, G, A, C, or T, histograms of the electrical recordings showed no change (data not shown) after treatment with glutathione and bisulfite and subsequent addition of alkali. Therefore, the combination of chemical modification and nanopore analysis enables us to detect 5hmC within a ssDNA using the wild-type αHL pore.

In summary, we have described a simple non-enzymatic modification of 5hmC, bisulfite-mediated thiol substitution, which allows coupling to peptides, fluorescein, and biotin. This single-step biotinylation procedure should be a useful way to enrich 5hmC-containing DNA fragments, and compares favorably with alternative approaches that use enzymes.[9,24,25] By nanopore analysis, the modified 5hmC residues can easily be distinguished from other nucleobases, which provides a basis for single-molecule sequencing to localize epigenetic markers in unamplified DNA. The approach might be extended to other epigenetic bases such as 5fC, where the aldehyde group could be specifically modified by hydrazine and aminooxy derivatives.[5,8]

Received: January 17, 2013
Published online: April 4, 2013

Keywords: 5-hydroxymethylcytosine · DNA methylation · epigenetics · gene sequencing · nanopores

[1] A. Bird, *Nature* **2007**, *447*, 396.
[2] E. N. Gal-Yam, Y. Saito, G. Egger, P. A. Jones, *Annu. Rev. Med.* **2008**, *59*, 267.
[3] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929.
[4] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930.
[5] T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008.
[6] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303.
[7] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300.
[8] E. A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.
[9] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394.
[10] M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, *Cell* **2012**, *151*, 1417.
[11] M. Münzel, D. Globisch, T. Carell, *Angew. Chem.* **2011**, *123*, 6588; *Angew. Chem. Int. Ed.* **2011**, *50*, 6460.
[12] S. J. Clark, A. Statham, C. Stirzaker, P. L. Molloy, M. Frommer, *Nat. Protoc.* **2006**, *1*, 2353.
[13] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315.
[14] H. Hayatsu, Y. Wataya, K. Kazushige, *J. Am. Chem. Soc.* **1970**, *92*, 724.
[15] R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, *92*, 422.
[16] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827.
[17] Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS One* **2010**, *5*, e8888.
[18] S. G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, *38*, e125.
[19] H. Hayatsu, M. Shiragami, *Biochemistry* **1979**, *18*, 632.

[20] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934.
[21] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368.
[22] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181.
[23] C. X. Song, T. A. Clark, X. Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He, J. Korlach, *Nat. Methods* **2011**, *9*, 75.
[24] C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68.
[25] Z. Liutkevičiūtė, E. Kriukiene, I. Grigaityte, V. Masevicius, S. Klimasauskas, *Angew. Chem.* **2011**, *123*, 2138; *Angew. Chem. Int. Ed.* **2011**, *50*, 2090.
[26] B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nat. Methods* **2010**, *7*, 461.
[27] H. Bayley, *Curr. Opin. Chem. Biol.* **2006**, *10*, 628.
[28] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. A. Schloss, *Nat. Biotechnol.* **2008**, *26*, 1146.
[29] G. M. Cherf, K. R. Lieberman, H. Rashid, C. E. Lam, K. Karplus, M. Akeson, *Nat. Biotechnol.* **2012**, *30*, 344.
[30] E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, J. H. Gundlach, *Nat. Biotechnol.* **2012**, *30*, 349.
[31] E. C. Hayden, **2013**, DOI: 10.1038/nature.2012.10051.
[32] N. Ashkenasy, J. Sanchez-Quesada, H. Bayley, M. R. Ghadiri, *Angew. Chem.* **2005**, *117*, 1425; *Angew. Chem. Int. Ed.* **2005**, *44*, 1401.
[33] D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, H. Bayley, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7702.
[34] E. A. Manrao, I. M. Derrington, M. Pavlenok, M. Niederweis, J. H. Gundlach, *PLoS One* **2011**, *6*, e25723.
[35] E. V. Wallace, D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, H. Bayley, *Chem. Commun.* **2010**, *46*, 8195.
[36] M. M. Butler, K. L. Graves, L. W. Hardy, *Biochemistry* **1994**, *33*, 10521.
[37] R. Nicewonger, C. A. Costello, T. P. Begley, *J. Org. Chem.* **1996**, *61*, 4172.
[38] C. W. Carreras, D. V. Santi, *Annu. Rev. Biochem.* **1995**, *64*, 721.
[39] S. G. Tajc, B. S. Tolbert, R. Basavappa, B. L. Miller, *J. Am. Chem. Soc.* **2004**, *126*, 10508.
[40] M. Sono, Y. Wataya, H. Hayatsu, *J. Am. Chem. Soc.* **1973**, *95*, 4745.
[41] R. P. Viscidi, C. J. Connelly, R. H. Yolken, *J. Clin. Microbiol.* **1986**, *23*, 311.
[42] W. W. Li, T. D. Claridge, Q. Li, M. R. Wormald, B. G. Davis, H. Bayley, *J. Am. Chem. Soc.* **2011**, *133*, 1987.

