ChemComm

COMMUNICATION

View Article Online
View Journal

Structural insights into how 5-hydroxymethylation influences transcription factor binding†

Cite this: DOI: 10.1039/c3cc48151d

Received 24th October 2013,
Accepted 19th November 2013

DOI: 10.1039/c3cc48151d

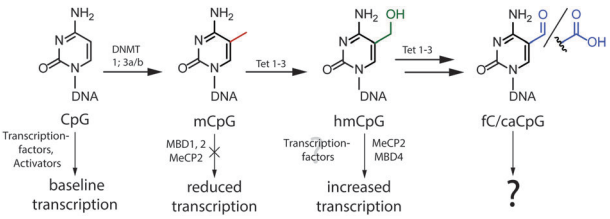
www.rsc.org/chemcomm

Lukas Lercher,‡[a] Michael A. McDonough,‡[a] Afaf H. El-Sagheer,[bc] Armin Thalhammer,[a] Skirmantas Kriaucionis,[d] Tom Brown*[b] and Christopher J. Schofield*[a]

Transcription factor binding and high resolution crystallographic studies (1.3 Å) of Dickerson–Drew duplexes with cytosine, methylcytosine and hydroxymethylcytosine bases provide evidence that C-5 cytosine modifications could regulate transcription by context dependent effects on DNA transcription factor interactions.

Cytosine-5-methylation at CpG islands, as catalysed by DNA methyltransferases (DNMTs), is an established mechanism for transcriptional regulation,[1] normally resulting in reduced gene expression. 5-Methylcytosine (5mC) can be oxidized by the ten-eleven-translocation (TET) 2-oxoglutarate (2OG) dependent oxygenases,[2] to yield 5-hydroxymethylcytosine (5hmC) and further oxidation products (5-formyl- and 5-carboxycytosine) (Scheme 1). 5hmC is enriched in embryonic stem and neuronal cells[3] and 5mC modification is proposed to play roles in maintenance of pluripotency,[4] development[5] and gene expression.[6] Despite evidence that 5mC oxidation products are common, and that abnormal TET activity is linked to disease, especially cancer, its biological roles are unclear. There is evidence that 5mC oxidation regulates transcription by attenuating DNA–protein interactions.[7–11] 5hmC formation could contribute to transcriptional regulation in part by attenuating DNA stability.[12] 5mC increases dsDNA stability,[13,14] consistent with the role of 5mC in CpG islands in transcriptional repression. We found that in some sequence contexts, 5hmC can, in part, reverse the stabilizing effect of 5mC.[12] We report crystallographic studies on 5hmC structures in dsDNA and investigate its effect on transcription factor binding.

Pioneering studies on DNA explored the self-complementary "Dickerson sequence",[15] which enables high resolution analysis.[16]



Scheme 1 Proposed general effects of redox-mediated cytosine modifications on transcription due to changes in dsDNA stability. The proposed effects may be linked to transcription factor binding. MBD, methyl-CpG-binding domain protein; MeCP2, methyl-CpG-binding protein 2.

Various DNA modifications have been structurally characterized in dsDNA and in complex with partners (see additional background information in ESI†). Recently, a structure of a B-DNA dodecamer with 5hmC was reported;[17] this study did not reveal effects of 5hmC modification on B-DNA structure. For direct comparison between different modifications, all parameters during crystallization, harvesting and X-ray data collection should be as close as possible.[18] We thus chose the Dickerson B-DNA sequence containing a CpG (CGCGAATTXGCG where X = C, 5mC or 5hmC, at positions 9 and 21 with continuous numbering), to investigate influences of cytosine modification on DNA structure.

Initially, we investigated the effect of 5mC and 5hmC on the thermal stability of the Dickerson sequences. UV melting temperature (T_m) analysis agree with previous studies[12,14,19,20] showing a higher T_m for 5mC than for C or 5hmC (5hmC [60.1 °C] < C [63.04 °C] < 5mC [64.01 °C]). We then crystallised and determined structures of Dickerson dsDNA dodecamers under the same crystallisation conditions (Table S1, ESI†). All crystals were isomorphous with the same space group and near equivalent unit cell dimensions (a ± 1.7%, b ± 0.5%, c ± 0.5%) and provided data to 1.3 Å resolution.

The structures of the B-DNA duplexes are very similar with RMSDs of all atoms: 0.18 Å (C-mC), 0.27 Å (C-hmC) and 0.35 Å (mC-hmC; Fig. S1, ESI†). Helix and bp step parameters were calculated using 3DNA[21] and were similar for all structures. Introduction of 5hmC results in a 0.8 Å (4.5%) widening of the major groove at the site of modification and a 1.1 Å (6.5%) narrower

[a] Department of Chemistry and the Oxford Centre for Integrative Systems Biology, Chemistry Research Laboratory, Oxford, UK. E-mail: Christopher.schofield@chem.ox.ac.uk
[b] Department of Chemistry, Chemistry Research Laboratory, Oxford, UK. E-mail: tom.brown@chem.ox.ac.uk
[c] Chemistry Branch, Department of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez University, Suez, Egypt
[d] Laboratory of Epigenetic Mechanisms, Ludwig Institute for Cancer Research, University of Oxford, ORCRB, Oxford, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3cc48151d

‡ These authors contributed equally to this work.

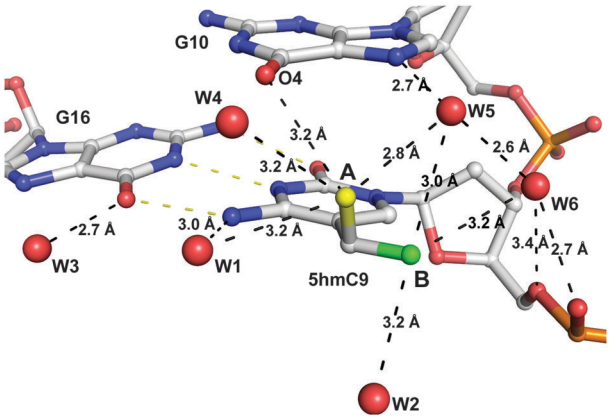


Fig. 1 Hydrogen bonding of the 2 refined 5hmC hydroxyl conformations (A, yellow, 70% occupancy; B, green, 30% occupancy) to observed surrounding waters and O4 of the 3′-adjacent G for conformation A.

major groove in the central AT tract (Table S2, ESI†) when compared to unmodified C. mC is reported to increase major groove width in a CpT context,[22] but we did not observe significant change in groove size on methylation in the CpG context.

The electron density for the hydroxymethyl group of 5hmC is well resolved and was refined in 2 conformations (Fig. 1). The major conformation (70% occupancy; conformation A) has the alcohol positioned to make a weak hydrogen bond to O4 of the 3′-adjacent G (3.4 Å with good geometry). Note analogous interactions are not possible with other bases at the +1 position. For pyrimidine bases, the distance between N4 (with C) or O4 (T) would be too great to form a hydrogen bond; while for A, geometric constraints likely hinder formation of a H-bond to N4. The alternative conformation (30%, conformation B) has the hydroxyl group orientated towards, and interacting with, the backbone phosphate group oxygen OP2 *via* a bridging water molecule (W6, Fig. 1). The different conformations observed for the hmC hydroxyl may impart sequence context dependent roles, because a CpG context can influence positioning of the hydroxyl.

A magnesium–water cluster induced likely by crystal packing in the immediate region of C21 (continuous numbering; C9 on the antisense strand; Fig. S1, ESI†) that can influence the water structure at this site was observed. C9 however, is free from direct crystal packing effects and was used for analysis of modification dependent solvation effects. The Solvation Web Service for nucleic acids (SWS)[23] was used to generate a "water density map", summarizing the hydration of all G–C base pairs in the nucleic acid database (NDB) with base pair parameters similar to our structures (Fig. S2D, ESI†). In each of our structures the water positions (W1–W6) around the modified cytosines are similar with small changes to the position of waters (W4 and W2) in 5hmC likely due to the introduction of the alcohol (Fig. S2, ESI†). The solvent structure for both hydroxyl conformations is likely to be very similar since all surrounding waters are held in place by an extensive hydrogen bonding network involving neighbouring bases and the phosphate backbone.

It has been suggested that 5hmC and further oxidation products might attenuate DNA–protein interactions, namely transcription factor (TF) binding. We investigated this possibility; focusing on basic-Helix-Loop-Helix (bHLH) TFs, because of interest in O_2 sensing regulated by the Hypoxia-Inducible Factor (HIF). HIF1α/β, a heterodimeric bHLH per arnt sim (PAS) TF, is a key regulator of the mammalian hypoxic response[24] and binds to hypoxic response elements (HRE; ACGTG) containing a CpG.[25] HIF levels/activity are regulated by oxygenases, which, like the TETs, are 2-OG, Fe^{2+}, and O_2 dependent. Thus, TET mediated 5mC oxidation has the potential to regulate transcription in a pO_2 dependent manner. Since structural data is not available for the bHLH region of HIF, we also investigated a related class of bHLH leucine zipper (ZIP) TFs binding to the HRE related E-Box sequence (CACGTG). The E-Box CpG is subject to methylation and most bHLH-ZIP TFs are not observed to bind to the E-Box sequence on CpG methylation.[26] Sequence alignments indicate high similarity in the bHLH regions of PAS and ZIP TFs (Fig. S3, ESI†), suggesting studies using the structurally characterised bHLH-ZIP TFs may provide useful information on bHLH-PAS binding.

Initially, 2 bHLH-ZIP TFs, MAX (produced in *E. coli*) and USF (produced by *in vitro* transcription translation; IVTT) were used to investigate the effect of different methylation/hydroxymethylation states at the central CpG on interaction with their cognate E-Box sequence. With electrophoretic mobility shift assays (EMSA) we observed strong binding of both MAX and USF to the unmodified E-Box sequences (Fig. 2A and B). USF and MAX showed weaker binding to all the hemi-modified oligonucleotides. Symmetric modification of 5mC or 5hmC almost completely abolishes binding of both transcription factors. The modification type, 5mC or 5hmC, was not observed to influence binding affinity to USF or MAX within our limits of detection. USF binding appears to be less affected by hemi-cytosine-modification than MAX binding.

We measured titration curves for USF and MAX and differently modified oligonucleotides (Fig. S4, ESI†). For MAX, the Kd_{app} for the hemi-modified E-Box sequence bearing 5hmC or 5mC increases 6-fold over the unmodified E-Box sequence. A similar trend was observed for USF. Precise quantification of this interaction was not

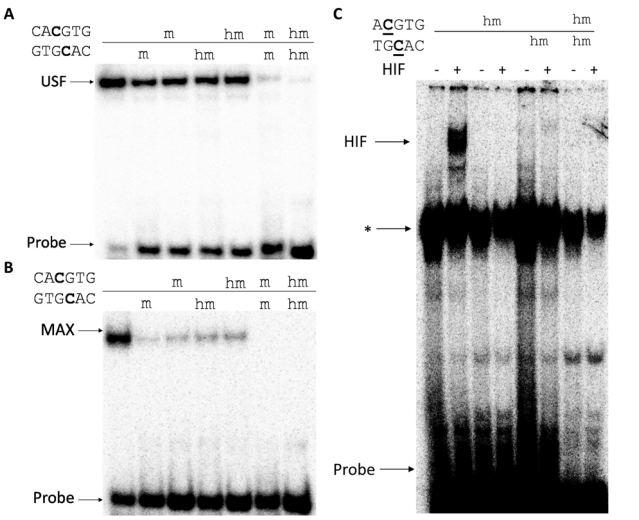


Fig. 2 Electrophoretic mobility shift assays with radiolabeled dsDNA probes bearing different modifications to cytosine C-5 at the central CpG of E-Box's or HRE's. The modified C is highlighted with the modification above the lane. Proteins used are IVTT produced USF (A) or purified MAX (B) with the E-Box probes and IVTT produced HIF (C) with the HRE probe (for sequences see ESI†). Non-specific and non-modification-sensitive binding can be observed in (C) (denoted with *).

possible with the IVTT produced USF, since other DNA and protein components in the mixture might influence binding. To confirm this trend for USF, we tested the competition between an unmodified labelled oligonucleotide and increasing concentrations differentially modified cold competitor oligonucleotides for USF binding (Fig. S4D, ESI†). The unmodified oligonucleotide efficiently competes for USF binding. The hemi-modified oligonucleotides also compete, albeit at higher concentrations, symmetrically modified oligonucleotides do not compete for USF binding under the tested conditions.

To investigate the different effect of 5hmC on MAX and USF binding, we overlaid our 5hmC dsDNA structure with those of the DNA binding bHLH domain of two bHLH-ZIP TFs, MAX[27] and USF[28] (Fig. S5, ESI†). A conserved Arg (Arg35 in MAX, Arg211 in USF) that makes direct contacts with the G in the central CpG in the MAX and USF dsDNA complex adopts a different conformation in the two structures. It is predicted that the Arg35 Cδ atom of MAX would clash with a C-5 modified cytosine, while in USF Arg211 is in a position that may tolerate the presence of C-5 modification without steric hindrance; this proposal is supported by the observation that USF-E-box binding is less influenced by C-5 modification (Fig. 2).

To investigate HIF binding, we used PCR to prepare an HRE containing oligonucleotide (see ESI†) globally substituting cytosines for 5mC[25] or 5hmC by PCR. We performed EMSAs with full length HIF1α/β (IVTT) and observed no HIF binding on substitution of C with mC or hmC (Fig. S6, ESI†).

To verify that this is due to modification of the CpG in the HRE, we performed EMSAs using chemically synthesized HRE oligonucleotides containing only one modification at the CpG. We observed clear HIF binding to the unmodified HRE, but the presence of hmC in the central CpG, even in the hemi-modified form, abolishes HIF binding within the limit of detection (Fig. 2 and Fig. S6, ESI†).

To investigate this difference we analysed a crystal structure of the basic helix-loop-helix domain of the related PAS-bHLH heterodimer CLOCK–BMAL1[29] bound to DNA. Superimposition of the CLOCK–BMAL1 complex with our 5hmC structure shows a conserved arginine (Arg47 in CLOCK and Arg85 in BMAL1) in position to clash with the cytosine-5 modification (Fig. S7, ESI†) as was observed for the bHLH-ZIP TFs structures, this arginine is also conserved in HIF1. We propose this clash with the conserved arginine to be responsible for the reduced affinity of HIF towards C-5 modifications.

We found that both 5mC and 5hmC influence bHLH TF binding to the same extent within our limits of detection. Hemi-modification is tolerated by 2 of the 3 TFs tested, while symmetric-modification abolishes detectable binding in all cases. 5mC is usually found in symmetrically modified CpGs, as maintained by DNMT1. Notably, hemi-hydroxylated CpGs are not reported substrates for DNMT1.[30] A mechanism whereby 5mC hydroxylation marks methylated CpG islands for activation after cell division, by providing some TFs an opportunity for occupancy of new sites in the genome, can be envisioned. Inhibition of HIF binding by the presence of 5hmC, an oxygen dependent modification, may constitute a new mechanism of gene-specific transcriptional control in the hypoxic response.

Overall, our results support the proposal that cytosine modifications regulate transcription in a sequence and context dependent manner; they thus have the potential to play distinct roles in development and pathophysiology. We propose the effects of cytosine modifications are mediated by a combination of effects on DNA itself (*i.e.* modulation of its stability) and on its binding interactions. Both our structures and those of Renciuk *et al.*[17] also indicate that static DNA structures in isolation are insufficient to explain the observed effects of cytosine modifications on DNA chemistry and biology; however the available structures and future ones of DNA-transcription factor complexes will help in understanding of how post-oligomerisation modifications regulate gene expression.

Notes and references

1 M. Münzel, D. Globisch and T. Carell, *Angew. Chem., Int. Ed.*, 2011, **50**, 6460–6468.
2 M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930–935.
3 S. Kriaucionis and N. Heintz, *Science*, 2009, **324**, 929–930.
4 J. A. Hackett, R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down and M. A. Surani, *Science*, 2013, **339**, 448–452.
5 T. Wang, Q. Pan, L. Lin, K. E. Szulwach, C.-X. Song, C. He, H. Wu, S. T. Warren, P. Jin, R. Duan and X. Li, *Hum. Mol. Genet.*, 2012, **21**, 5500–5510.
6 M. R. Branco, G. Ficz and W. Reik, *Nat. Rev. Genet.*, 2012, **13**, 7–13.
7 C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes and H. Leonhardt, *PLoS One*, 2011, **6**, e21306.
8 M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis and N. Heintz, *Cell*, 2012, **151**, 1417–1430.
9 O. Yildirim, R. Li, J.-H. Hung, P. B. Chen, X. Dong, L.-S. Ee, Z. Weng, O. J. Rando and T. G. Fazzio, *Cell*, 2011, **147**, 1498–1510.
10 C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell and M. Vermeulen, *Cell*, 2013, **152**, 1146–1159.
11 M. Iurlaro, G. Ficz, D. Oxley, E.-A. Raiber, M. Bachman, M. Booth, S. Andrews, S. Balasubramanian and W. Reik, *Genome Biol.*, 2013, **14**, R119.
12 A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown and C. J. Schofield, *Chem. Commun.*, 2011, **47**, 5325–5327.
13 C. M. Rodríguez López, B. Guzmán Asenjo, A. J. Lloyd and M. J. Wilkinson, *Anal. Chem.*, 2010, **82**, 9100–9108.
14 M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein and M. Drndic, *J. Am. Chem. Soc.*, 2010, **133**, 486–492.
15 H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 2179–2183.
16 T. K. Chiu and R. E. Dickerson, *J. Mol. Biol.*, 2000, **301**, 915–945.
17 D. Renciuk, O. Blacque, M. Vorlickova and B. Spingler, *Nucleic Acids Res.*, 2013, **41**, 9891–9900.
18 X. Shui, L. McFail-Isom, G. G. Hu and L. D. Williams, *Biochemistry*, 1998, **37**, 8341–8355.
19 C. M. Rodríguez López, A. J. Lloyd, K. Leonard and M. J. Wilkinson, *Anal. Chem.*, 2012, **84**, 7336–7342.
20 M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff and T. Carell, *Chem.–Eur. J.*, 2011, **17**, 13782–13788.
21 X. J. Lu and W. K. Olson, *Nucleic Acids Res.*, 2003, **31**, 5108–5121.
22 U. Heinemann and M. Hahn, *J. Biol. Chem.*, 1992, **267**, 7332–7341.
23 P. Auffinger and Y. Hashem, *Bioinformatics*, 2007, **23**, 1035–1037.
24 R. Chowdhury, A. Hardy and C. J. Schofield, *Chem. Soc. Rev.*, 2008, **37**, 1308–1319.
25 R. H. Wenger, I. Kvietikova, A. Rolfs, G. Camenisch and M. Gassmann, *Eur. J. Biochem.*, 1998, **253**, 771–777.
26 G. Prendergast and E. Ziff, *Science*, 1991, **251**, 186–189.
27 A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff and S. K. Burley, *Nature*, 1993, **363**, 38–45.
28 A. R. Ferre-D'Amare, P. Pognonec, R. G. Roeder and S. K. Burley, *EMBO J.*, 1994, **13**, 180–189.
29 Z. Wang, Y. Wu, L. Li and X.-D. Su, *Cell Res.*, 2013, **23**, 213–224.
30 V. Valinluck and L. C. Sowers, *Cancer Res.*, 2007, **67**, 946–950.