Chemical Science

ROYAL SOCIETY OF CHEMISTRY

REVIEW

Check for updates

Cite this: *Chem. Sci.*, 2026, **17**, 2045

The role of cytosine modification symmetry in mammalian epigenome regulation

Zeyneb Vildan Cakil, Lena Engelhard and Daniel Summerer *

Received 18th November 2025
Accepted 17th December 2025

DOI: 10.1039/d5sc09022a

rsc.li/chemical-science

5-Methylcytosine (mC) is a key regulatory element of mammalian genomes, and plays important roles in development and disease. mC is predominantly written onto CpG dyads by DNA methyltransferases, and can be further oxidized by ten-eleven translocation dioxygenases (TETs) to 5-hydroxymethyl-, 5-formyl-, and 5-carboxylcytosine. This process results in different symmetric and asymmetric combinations of cytosine forms across the two strands of CpGs, each of which represents a unique physicochemical signature in the major groove of DNA. A comprehensive understanding of the individual functions of oxidized mC modifications can therefore only be achieved by considering both strands of CpG dyads. Here, we provide a brief overview of the current state of knowledge on the sequencing and mapping of individual CpG dyad states, their influence on the intrinsic properties of DNA, and their interactions with chromatin proteins.

1 Introduction

1.1 TET-generated cytosine modifications and their existence as symmetric and asymmetric CpG dyad marks

5-Methylcytosine (mC, Fig. 1a) is the central epigenetic mark of mammalian DNA and acts as a key regulator of transcription, with important roles in developmental and (patho)physiological processes, including genomic imprinting, X-chromosome-inactivation, and cancer development.[1] mC is predominantly written into CpG dyads by *S*-adenosylmethionine (SAM)-dependent DNA methyltransferases (DNMT),[2] and is typically associated with transcriptional silencing.[3,4] mC can thereby be written *de novo* by DNMTa/3b, but can also be maintained over cell divisions by the maintenance enzyme DNMT1 that selectively methylates hemi-methylated CpG dyads generated during replication (Fig. 1a). This makes mC a rather stable, inheritable nucleobase. However, mC can also be passively reversed to C by DNA replication and an absence of maintenance methylation (passive dilution, PD).[3] Moreover, α-ketoglutarate- and Fe(II)-dependent ten-eleven-translocation dioxygenases (TETs) can trigger an active demethylation pathway by oxidizing mC to the

Faculty of Chemistry and Chemical Biology, TU Dortmund University, Otto-Hahn-Str. 4a, 44227 Dortmund, Germany. E-mail: daniel.summerer@tu-dortmund.de



Zeyneb Vildan Cakil

Zeyneb Vildan Cakil obtained her BSc degree in Molecular Biology and Genetics at Middle East Technical University (METU) and her MSc degree in Chemistry with a Chemistry and Life Sciences track co-accredited by Sorbonne University and Paris Science and Letters (PSL) University. Her master's thesis with A. Gautier focused on the development of a chemically induced proximity technology with intrinsic fluorescence imaging and sensing capabilities named CATCH-FIRE. She is currently pursuing her PhD degree in the group of Daniel Summerer, where she works on non-canonical cytosine modifications and identification of their unique readers.



Lena Engelhard

Lena Engelhard received her BSc degree and MSc degree in Biochemistry from Ruhr University Bochum (RUB). She completed her master's thesis at the Max Planck Institute of Molecular Physiology in Dortmund in the laboratory of Stefan Raunser, focusing on the structural elucidation of proteins through high-resolution cryo-electron microscopy. She is currently pursuing her PhD in the laboratory of Daniel Summerer, where she investigates epigenetic DNA modifications with an emphasis on proteomics.

This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

(cc) BY

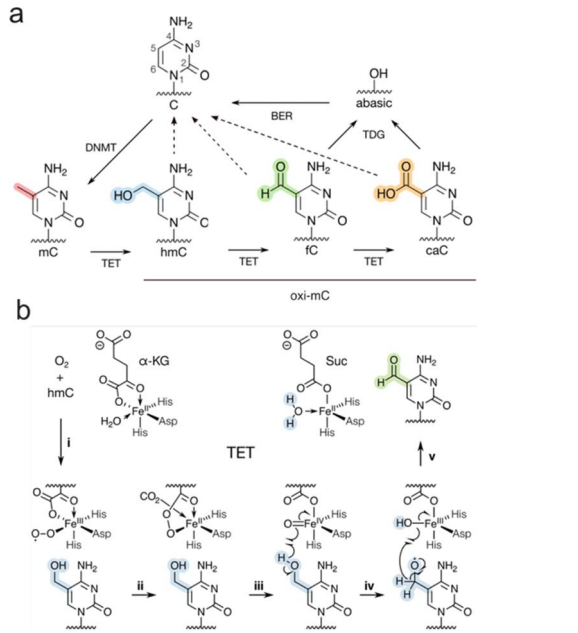


Fig. 1 (a) Cytosine methylation and active demethylation in mammals. DNMT: DNA methyltransferase, TET: ten-eleven translocation dioxygenase, TDG: thymine DNA glycosylase, BER: base excision repair. Dotted arrows: passive dilution.[14] (b) Proposed mechanism of TET-catalyzed oxidation (example shown for oxidation of hmC to fC). α-KG: α-ketoglutarate, Suc: succinate.[18,19]

"oxi-mCs" 5-hydroxymethylcytosine (hmC),[5,6] 5-formylcytosine (fC), and 5-carboxylcytosine (caC)[7–9] (Fig. 1b). Base excision repair (BER) can restore C *via* the excision of fC and caC by thymine DNA glycosylase (TDG), and repair of the generated abasic site (active modification – active removal).[7,10] In addition to this pathway, hemi-modified CpGs containing oxi-mCs seem to compromise DNMT1-catalyzed maintenance methylation compared to hemi-methylated CpG. Consequently, oxi-mC can promote replication-dependent passive dilution of mC *via* active modification and passive dilution (with effects increasing with oxidation state[11–15]). Notably, for reviews discussing these pathways, see ref. 16 and 17.

Cytosine modifications can occur in a strand-symmetric and -asymmetric fashion in the double-stranded genome. Whereas non-CpG (CpH) methylation (and partially hydroxymethylation) can occur at certain levels in a tissue-specific manner and is inherently asymmetric,[20,21] the palindromic nature of the CpG dyad itself allows many different combinations of cytosine modifications to be presented in the DNA major groove, and thus provides rich symmetry information. Indeed, whereas the aforementioned maintenance methylation of hemi-methylated dyads ("mC/C", for the top and bottom strands, respectively) represents an evolved mechanism to maintain them in a symmetric mC/mC state, the oxidation of mC to hmC, fC and caC by TETs occurs in a step-wise and non-processive manner,[22–25] and theoretically can give rise to fifteen (mostly asymmetric) modification combinations (Fig. 2a). Whereas TETs exhibit substrate preferences in view of pre-existing CpG modifications in dsDNA,[24] they can also act on ssDNA, a process that would be inherently independent of modifications of the opposite strand.[26] Whereas part of these dyads may not occur in appreciable numbers in genomes owing to the low levels of fC and particularly of caC (see below), this nevertheless creates



Daniel Summerer

Daniel Summerer earned his PhD degree in the group of A. Marx in nucleic acid chemistry and DNA polymerase design. He did post-doctoral research with P. G. Schultz at the Scripps Research Institute on the expansion of the genetic code with photo-responsive (e.g., fluorescent and photocaged) amino acids, enabling studies of dynamic processes in gene regulation, e.g., via light control of transcription factors. After research in the field of (epi)genomics in the biotech sector, he started his independent career at the University of Konstanz in 2011 and was appointed as Professor for Chemical Biology at TU Dortmund University in 2015. In his research, he employs nucleic acid and protein chemistry as well as directed molecular evolution to design tools that provide new insights into epigenetic mechanisms of chromatin regulation.

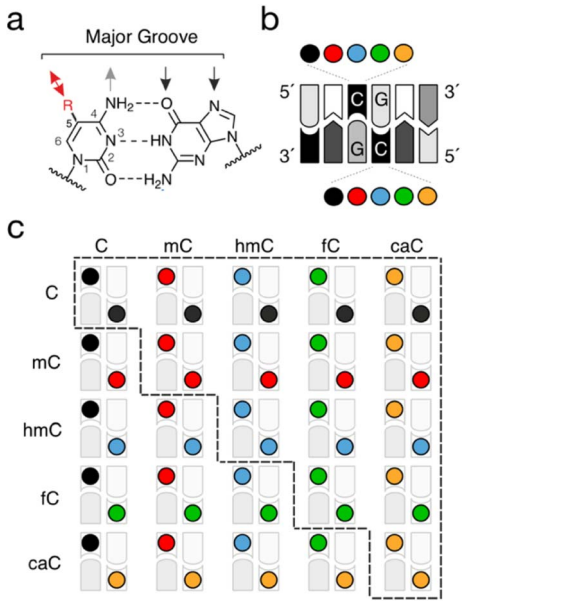


Fig. 2 Combinations of cytosine modifications in the two strands of CpG dyads. (a) Chemical information displayed in the major groove of dsDNA. Black and grey arrows denote hydrogen bond acceptors and donors, respectively. Red arrow shows alternative interactions of cytosine 5-substituents. (b) Schematic view of a CpG dyad with possible combinations of cytosine nucleobases. (c) All theoretically possible combinations from (b). Note that 15 different combinations occur within the context of the CpG itself (marked by a dashed line), whereas 25 combinations can occur overall when considering the CpG's sequence context. Color code as in Fig. 1a.

© 2026 The Author(s). Published by the Royal Society of Chemistry

a complex landscape of combinatorial CpG marks across the genome. In particular, dyads involving the most frequent cytosine nucleobases C, mC and hmC are expected to occur at significant levels in stem cells and neurons,[16] and unique genomic distributions have been observed in first pilot studies (see below).[27–29] Each individual dyad thereby presents a unique combination of 5-substituents in the major groove, and may uniquely affect the physicochemical properties of dsDNA, as well as interactions with nuclear proteins. The dyads thus represent unique signals with distinct regulatory effects that may arise from dedicated, dyad-specific reader protein interactions, or simply from a unique modulation of pathways relying on a selective recognition of symmetric C/C or mC/mC dyads.

A comprehensive understanding of the individual CpG dyad's functions requires a broad knowledge of their properties as a basis, including their direct effects on the structure and physico-chemical properties of DNA, their genomic levels and locations (including tissue dependence and dynamics), and their individual interaction profiles with the nuclear proteome. We here review the most recent developments in these fields that have helped to shed light on this question and that will guide future studies.

2 Simultaneous sequencing of C, mC, and hmC for mapping symmetric and asymmetric CpG dyad states in mammalian genomes

Understanding individual CpG dyad state functions depends on the ability to map their genomic locations. The overall genomic levels of oxi-mCs differ significantly between each other, are tissue-dependent, and can be highly dynamic (reviewed in ref. 30). For example, whereas mC is distributed rather evenly among somatic cells, oxi-mC levels are particularly high in embryonic stem cells (ESC) and neurons (but low in other somatic and particularly many cancer cells), with hmC showing the highest levels in neurons (>10–20% of the levels of mC[30]). In addition, hmC can also show high stability.[31] In comparison, fC and caC show far lower levels (~2 and another 1–2 orders of magnitude lower than hmC in mouse ESC (mESC), respectively[30]). These general differences will also translate into different levels of the individual CpG dyad states, which should be considered when judging the physiological relevance of a particular state. A large number of methods have been introduced for sequencing and mapping of oxi-mCs in general (reviewed in ref. 32 and 33), and have enabled detailed maps and correlations with particular regulatory regions. Briefly, hmC is typically found enriched in active enhancers and gene bodies, and in the latter correlates with active transcription (fC and caC are also enriched in gene bodies of actively transcribed genes). In contrast, low levels of all oxi-mCs are typically found in regions surrounding transcriptional start sites of active promoters.[34–38]

However, most aforementioned methods provide selectivity only for single or grouped, but not for each individual cytosine form in a given sequencing run. This limits their analytical value to statistic assumptions rather than accurate determinations of the actual CpG dyad states. Nevertheless, such maps generally indicated that hmC is typically asymmetric.[37–40] Moreover, an early single molecule imaging study identified hmC/mC as a frequent modification in mouse cerebellum DNA.[41]

Recently, the first methods for the selective, simultaneous detection of C, mC and hmC in one experiment run have emerged that provide potential for refined maps with CpG dyad state resolution. These approaches include restriction enzyme-based techniques like DARESOME[42] and Dyad-seq[27] that however are limited in coverage and resolution by their dependence on specific restriction sites. Other approaches use complex multistep protocols based on chemical nucleobase conversions to achieve nucleotide resolution with whole genome coverage. For example, the EnIGMA protocol[43] and its further developed version "SCoTCH-Seq"[29,44] employ hairpin adapters to store the original DNA sequence and its mC pattern as complement *via* a primer extension and maintenance methylation step, after which hmC and mC are revealed by deamination protocols based either on bisulfite or on multiple enzymatic conversions involving A3A deaminase (Fig. 3a). Another conversion-based approach is SIMPLE-Seq[45] that employs a K_2RuO_4 treatment to oxidize hmC to fC followed by a malononitrile labeling, leading to an adduct that reads as U and is recorded in a complement strand by a primer extension step. Then, the sample is subjected to a TET oxidation step (converting all modified Cs into caC) and subsequent borane reduction, ultimately converting mC in the original template strand to DHU, which results in another C to U transition that can be sequenced. Whether a C to U transition is a result of hmC or mC conversion is decoded by the use of a caC-modified primer in the first primer extension step.[45] It should be noted that none of the aforementioned methods is designed to sequence the less abundant fC and caC modifications that will read as C or U, as well.

Finally, in addition to restriction- and conversion-based strategies, direct sequencing methods promise particularly simple mapping of modified cytosines.[46] Here, a very recent study employed nanopore sequencing with duplex-paired reads to map hmC in mouse cerebellum (Fig. 3b) [28]

So far, only three studies have harnessed one of the aforementioned techniques for establishing genomic maps of individual CpG dyad states. Duplex-paired nanopore sequencing provided highly valuable insights into the locations and overall frequencies of all hmC CpG dyad states in the mouse cerebellum genome that generally exhibits high hmC levels. Symmetric mC/mC was found to be the most abundant state overall, accounting for 53% of all duplex base-calls, whereas symmetric unmodified C/C dyads accounted for 23% (Fig. 3c). Strikingly, hmC-containing dyads occurred predominantly in the form of asymmetric hmC/mCs (72% of all hmC dyads, accounting for 13% of all dyads), which are the first products that TETs generate from their initial mC/mC substrate[28] (this data roughly aligns with Dyad-Seq data for mESC DNA[27]). In contrast, symmetric hmC/hmC dyads (a direct subsequent TET

© 2026 The Author(s). Published by the Royal Society of Chemistry

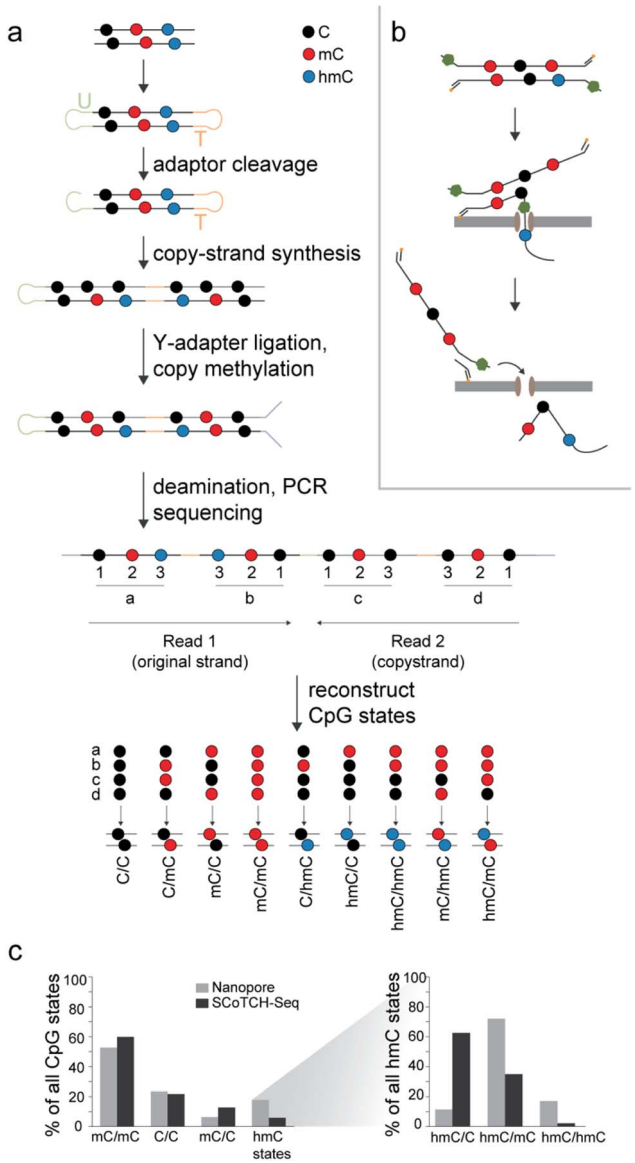


Fig. 3 Strategies for simultaneous sequencing of C, mC and hmC enable genomic mapping of individual CpG dyad states. (a) Scheme of the SCoTCH-Seq approach. (b) Scheme of nanopore sequencing with duplex-paired reads. (c) Genomic levels of individual CpG states as reported for nanopore sequencing of mouse cerebellum DNA[28] and SCoTCH-Seq for mESC DNA.[29]

product) and asymmetric hmC/C dyads (that can theoretically arise from an existing hmC dyad by passive dilution or repair, or from direct off-target oxidation of mC/C dyads by TETs[16]) were found at much lower levels (2–3%, Fig. 3c).

Finally, SCoTCH-Seq has been used for dyad-resolved mapping in mESC genomes, and showed related distributions: mC/mC dyads accounted for 60%, C/C for 22%, mC/C for 13% and all hmC dyads for 6% (Fig. 3c).[29] Similar to what was observed in the nanopore study, only a very small fraction of hmC dyads were symmetric (2%), whereas 35% existed as hmC/mC and 63% as hmC/C dyads. The orientation of asymmetric dyad modifications was thereby generally random, *i.e.*, did not show strand-bias in respect to gene orientation. An overall conclusion from these studies is that TETs indeed oxidize the two strands of mC/mC dyads independently from each other, corroborating previous *in vitro* studies that indicated a non-processive activity.[22–25] The studies also refined previous insights into the genomic distribution of hmC. For example, metagene profiles established by SCoTCH-Seq showed an enrichment of hmC/mC and hmC/C in the bodies of actively transcribed genes, and a depletion at transcription start sites, both being significantly more pronounced for hmC/C. Similarly, primed enhancers showed far higher hmC levels than poised or active enhancers that both showed elevated levels only of hmC/C in their flanking regions. This indicates that dyad-resolved maps are extremely helpful for unraveling the function of individual CpG dyad states. With several methods now available at least for parallel C, mC, and hmC sequencing, it is expected that a growing number of dyad-resolved oxi-mC maps will be reported in the near future.

A limitation in mapping dyad modifications is however the inability of current sequencing approaches to extend simultaneous sequencing of oxi-mCs beyond C, mC and hmC. Achieving this maximal chemical resolution will require significant further efforts, but – given the unique properties of fC and caC, and their ability to control interactions with central chromatin proteins – would be highly valuable.[47–54]

Another bottleneck for broader studies is the low levels of many dyad states, resulting in a requirement for costly whole genome deep sequencing. Besides the generally low levels of fC and caC modifications, hmC levels are also low in somatic and particularly cancer tissues.[30,31] Given the importance of hmC as a cancer biomarker[55,56] and the still poorly understood functions of fC and caC dyads, future studies would thus greatly benefit from enrichment methods that are applicable to undenatured dsDNA fragments, and offer general selectivity for single cytosine modifications or even for specific dyad modification combinations prior to sequencing.

Current hmC-enrichment strategies typically rely on anti-hmC antibodies[57–60] that however typically require DNA denaturation and thus destruction of CpG dyad information. Alternatively, T4-β-glucosyltransferase is widely used for enrichment of hmC,[60–62] whereas engineered MBD proteins have been employed for the enrichment of hmC/mC dyads[63] and provide a direct and simple access to lower resolution maps.[64]

3 The impact of oxidized mC modifications on the physicochemical properties of DNA

DNA exists in double-stranded form in the mammalian nucleus, and it is complexed with histones and other chromatin proteins. Different CpG dyad states may influence protein interactions in diverse ways, for example, by direct recognition or clash with the substituents themselves, by indirect effects on the electronic properties of the nucleobase that affect hydrogen bonding and stacking, by changes in duplex shape readout *via* altered groove geometries, or by altered duplex flexibility. The 5-

© 2026 The Author(s). Published by the Royal Society of Chemistry

substituents differ in hydrogen bonding properties, steric demand and conformational flexibility. The hydroxyl group of hmC is found in crystal structures in a main orientation pointing towards the 3′-nucleobase, and a second conformation undergoing a water-mediated phosphate interaction. In contrast, the fC-formyl and caC-carboxyl groups are fixed in the plane of the cytosine nucleobase, and hydrogen bond to the N4 amino group (Fig. 4a and b).[65–67] Both fC and caC substantially alter the electronic and chemical properties of the nucleobase, such as charge distribution, polarity, and pK_a (with caC being negatively charged).[10,68]

Effects of mC itself on local DNA structure have been extensively studied. Among the main findings are an increase of the local curvature of DNA, as well as effects on the groove geometries (slightly widened major *versus* narrower minor groove, respectively). Similarly to mC, hmC has been shown to result in a slight local widening of the major groove[65] compared to fC- and caC-modified DNA, with the latter exhibiting an enlarged minor groove compared to other modified duplexes in the same study.[70] However, a thorough solution NMR study with dsDNAs bearing a central CpG in two different sequence contexts afforded structures for C, mC and hmC-modified duplexes that overall showed only modest local effects of the modifications that were smaller than the differences induced by the sequence contexts themselves.[71] Similar findings were made for fC- and caC-modified DNA.[72,73]

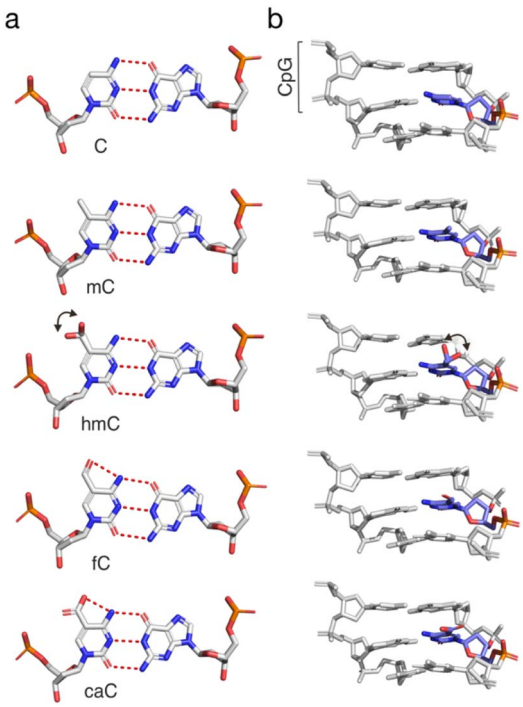


Fig. 4 Structures of (a) CG base pairs and (b) trinucleotide duplexes from the Dickerson–Drew dodecamer containing C, mC, hmC, fC, or caC (PDB entries 436D, 4C63, 4I9V, 4QC7, and 4PWM, respectively).[65–67,69] Hydrogen bonds shown as dotted red lines. Note the conformational freedom of the hmC hydroxyl group indicated by two main orientations in the crystal.

Many aspects of the physicochemical effects of mC and oxi-mCs have been studied as well. Influences on the duplex stability analyzed by thermal melting analyses are highly sequence-dependent. For example, whereas increased TMs observed for a number of mC-modified sequences indicate a generally stabilizing effect,[74–77] a slight destabilization has been observed for the Dickerson–Drew dodecamer (DDD).[67] In contrast, hmC has been reported to have either a slightly stabilizing effect (less than mC[71,74,75,77]), a slightly destabilizing effect,[76,78] or no effect.[67] Strikingly, either a slightly destabilizing[65] or no effect[67] has thereby been reported for the same sequence context (DDD with two different hmC modification settings). Further oxidation of hmC to fC or caC leads to sequence-dependent effects as well. Whereas fC seems to have little impact,[67,77] caC has been reported to have either a weak[77] or a strongly stabilizing effect.[67] In-solution NMR studies with fC- or caC-modified duplexes afforded destabilizing effects in both cases.[72,73]

Effects of the modifications on base stacking have been reported for mC (increased stacking[79–81]), whereas hmC, fC, and caC show similar stacking patterns in crystal structures.[67]

Finally, molecular dynamics simulations and circularization experiments suggest that mC and even more hmC tend to increase DNA stiffness, though again with sequence-dependence.[71] Single molecule DNA looping studies with varying numbers of modifications were in agreement with a stiffening effect of mC. In contrast, hmC and particularly fC increased the flexibility, whereas caC showed little impact.[82]

Taken together, oxi-mC modifications – while having little effect on the overall conformation of B-DNA in modification patterns studied to date – do partially impact the local duplex structure and can affect the stability and flexibility of dsDNA in a sequence-dependent manner. In light of the observed sequence dependencies and the employment of dsDNA oligonucleotides with single or only a few modification sites in the majority of the aforementioned studies, it is still poorly understood how dense modifications – such as those observed in CpG islands that have high physiological relevance – may lead to more pronounced or even alternative effects. Similarly, the aforementioned studies predominantly focused on DNA containing hemi-modified CpGs. It is therefore not understood, how CpG dyad states may uniquely affect DNA properties, which calls for systematic, comparative studies.

4 Proteins reading CpG modification symmetry

Many proteins engage with CpG dyads during their turnover in the cell cycle, and can exhibit selectivity for specific oxi-mC dyad states. Most importantly, this applies to DNMTs,[11,12,83] TET dioxygenases,[8,18,84] and TDG[7,10,85] as factors responsible for the writing and erasing of mC (Fig. 1). In addition, methyl-CpG-binding domain (MBD) proteins, the canonical readers of mC/mC dyads, mediate communication with heterochromatin-associated factors for transcriptional silencing, and show specific dyad state preferences (see below).[3,4] Whereas MBDs

© 2026 The Author(s). Published by the Royal Society of Chemistry

preferentially read symmetric mC/mC CpGs, CXXC and SET- and RING-associated (SRA) domains can recognize the dyad in the non- or hemimethylated state, respectively, and are contained in a variety of chromatin factors. Here, the CXXC of TET3 has been shown to read caC,[51] whereas the SRA domain of UHRF2 selectively reads hmC.[86]

In addition, fC has been shown to form imine cross-links with lysine/arginine residues of histones and other factors *in vitro*,[49,87] with potential roles for nucleosome positioning *in vivo*.[48] As another general factor, RNA polymerase II has been shown to be stalled by fC and caC *in vitro*,[53] and a specific interaction with caC has been identified in a crystal structure that may account for this effect.[54] Importantly, beyond such general chromatin factors, oxi-mC dyad states can also modulate the affinity of specific transcription factors (TFs). In the following, we exclusively discuss canonical MBD readers and transcription factors, because of their immediate relevance for the two main regulatory pathways of (oxi)-mC. We will thereby focus on comparative studies involving different oxi-mC dyad states, since only these enable a judgement of actual CpG symmetry functions. Concerning the interactions of other protein factors with oxi-mCs, we refer readers to previous reviews.[88,89]

4.1 MBD proteins

The prototypical and best-characterized family of reader proteins for canonical mC/mC CpGs comprises the methyl-CpG-binding domain (MBD) proteins.[90] The members of the core MBD family in mammals are MeCP2, MBD1, MBD2, MBD3, and MBD4.

Among these, MeCP2, MBD1, and MBD2 exhibit high selectivity for symmetrically methylated CpGs over unmethylated CpGs, and frequently associate with chromatin-modifying complexes such as histone methyltransferases and ATP-dependent remodelers, thereby linking DNA methylation to histone modifications and transcriptional repression.[91]

In contrast, the core family member MBD3 as well as MBD5 and MBD6 diverge in key residues and secondary structure elements, resulting in loss of high-affinity mC recognition.[91] Structural analyses of MBD–DNA complexes revealed a conserved mode of mCpG recognition. MBD proteins share a small, asymmetric fold that engages the mCpG through two arginine residues hydrogen-bonding to the Hoogsteen face of the CpG guanines (Fig. 5a shows the representative MeCP2-mCpG complex[92,94]). Adjacent residues, including a conserved aspartate and tyrosine, help stabilize these interactions, and form part of hydrophobic pockets that accommodate the two cytosine 5-methyl groups (Fig. 5b). Early reports suggested that hmC is recognized by MBD3 and MeCP2 (ref. 95 and 96). However, subsequent proteomic analyses[97,98] and comprehensive binding studies failed to confirm these observations. In fact, hmC is bound with markedly reduced affinity by MBD family proteins compared to mC.[11,99,100] These findings imply that hmC may serve as an intrinsic modulator capable of alleviating mC- and MBD-dependent silencing even without active demethylation. Consistent with this, genome-wide maps

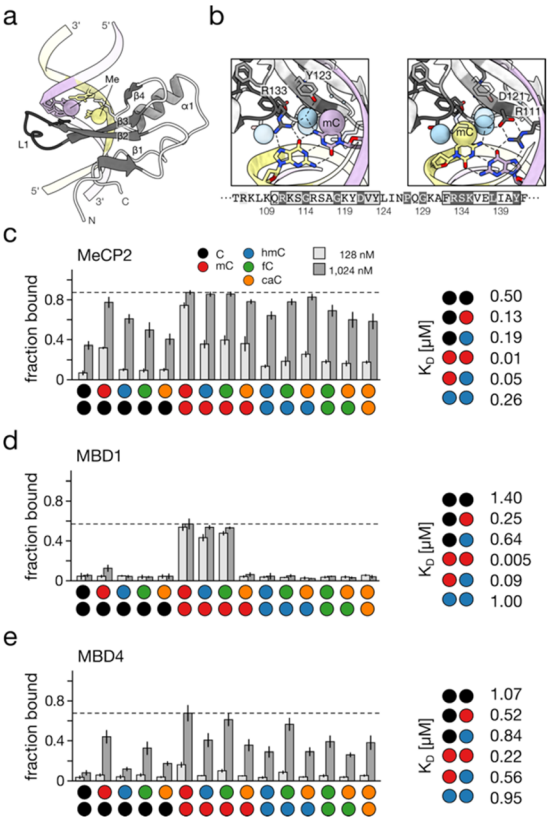


Fig. 5 Read-out of CpG dyad states by MBD domains. (a) Overview of crystal structure of the MBD domain of human MECP2 (PDB 3c2i[92]) with methyl groups shown as spheres. (b) Details of G and mC recognition by MECP2 in the two CpG dyad strands. Note the recognition of dyad Gs by conserved Arg residues. Strand coloring as in (a), methyl groups shown as spheres in color according to strand, water shown as blue spheres. Hydrogen bonds shown as dotted black lines. MECP2 protein sequence of shown area indicated below. (c–e) Exemplary binding profiles of MBD domains of human MeCP2, MBD1 and MBD4 to the 15 CpG dyad states (left[93]) and K_D values for frequent CpG dyad states containing C, mC or hmC (right, MBD4 from mouse).[11] All data from electromobility shift assays (EMSA). Figure adapted from Buchmuller *et al.*,[63] in accordance with its creative commons attribution 4.0 international license: http://creativecommons.Org/licenses/by/4.0/.

correlate hmC with transcriptionally active regions.[62] Several studies reported broader evaluations of the dyad state preferences of individual MBDs (Fig. 5c–e show exemplary data for three of the five core family MBDs). An overall finding is that all MBDs recognize mC/mC dyads with highest affinity, but with varying degrees of selectivity.[11] Whereas MBD1, MBD2 and MeCP2 show low nanomolar affinity for mC/mC and comparably high selectivity, MBD4 and particularly MBD3 show lower affinity and selectivity. For the former three MBDs, hmC/mC is the most important off-target (5–18-fold lower affinity than that for mC/mC (Fig. 5c–e[11]). A general tendency among dyads is that mC contributes to high affinity, and that hmC reduces affinity less than non-modified C. fC and caC tend to generally cause similar affinity reductions to hmC, though for MBD1, caC causes significant reductions (Fig. 5d). As visible in Fig. 5c–e, the MBDs overall feature different specificity profiles.[93] This combined data provides biochemical support for the hypothesis

© 2026 The Author(s). Published by the Royal Society of Chemistry

that different oxi-mC dyad states may act as individual regulatory signals, since they may individually modulate a main pathway of mC-dependent chromatin regulation. The advent of dyad-resolved sequencing techniques will help in studying this question in the cellular context.[27–29]

4.2 Transcription factors

Many transcription factors recognize target sequences containing a CpG, and they can either be repelled or attracted by mC.[101] Oxi-mCs equip the dyad with additional information that can be (anti)read by TFs. Examples for TFs that prefer oxi-mCs and for which detailed biochemical and partially structural information is available are SALL1 and 4, which both possess a C_2H_2 zinc finger domain with a preference for hmC, and are involved in the recruitment of TET dioxygenases to promote hmC oxidation.[102] Wilms tumor protein 1 (WT1), another zinc finger TF that is found mutated in nephroblastoms, has been shown to prefer C, mC or caC over hmC or fC.[103] The presence of hmC also increases the binding of the basic helix–loop–helix (bHLH) TF TCF4 to E-box motif sequences.[104] Finally, MAX, another bHLH TF and dimerization partner of the master regulator MYC (Fig. 6a), has been shown to bind its E-box target sequence in the presence of a caC or C rather than mC, hmC or fC. Both symmetric and hemi-modified states were characterized, revealing individual affinities for each dyad. A crystal structure revealed a conserved arginine involved in recognition of the caC-paired G residue, and a second arginine in a ~6 A distance to the caC carboxyl group that may be involved in water-mediated interactions (Fig. 6b).[52]

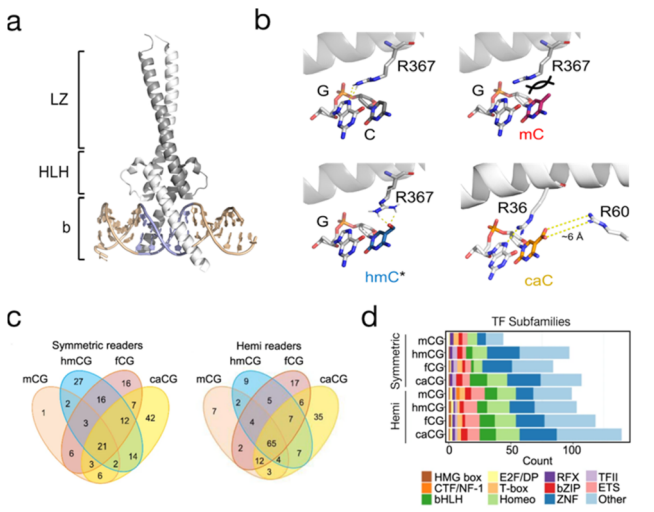


Fig. 6 Oxi-mC (anti)reading by transcription factors. (a) Crystal structure of the MYC/MAX heterodimer (PDB 1NKP) bound to E-box DNA. Grey: MAX, white: MYC, blue: E-box. (b) Interaction of MYC R367 in the MYC/MAX dimer with E-box guanine 4 paired with C (PDB 1NKP), models of MYC/MAX bound to mC/mC or hmC/hmC dyads, respectively,[107] and interaction of MAX R36 and R60 in the MAX_2 dimer with the caC/caC dyad (PDB 5EYO; *interaction with hmC has not been observed in the E-box sequence). (c) Venn diagrams show readers for mC, hmC, fC, and caC modifications in symmetric and hemi-modified CpGs. (d) The readers of symmetric and hemi-modified CpGs are observed for many TF subfamilies. Figure adapted from Song *et al.*,[108] in accordance with its creative commons attribution 4.0 international license: http://creativecommons.Org/licenses/by/4.0/.

Besides studies dedicated to specific TFs, pull-down/proteomics screens have been particularly powerful in discovering TFs with general oxi-mC preferences (for overviews of (anti)reader candidates from such studies, see ref. 97, 102, 105, and 106). An overall observation in these experiments has been that fC and caC attract a higher number of readers than hmC. Importantly, the potential of proteomics in this field is still largely untapped, since previous studies employed probes exclusively containing symmetric CpG states. Besides the higher number of theoretically existing asymmetric dyad states (Fig. 2c), symmetric states may also be underrepresented (*e.g.*, in mESC, only ~2% of hmC is symmetric, whereas 98% resides in hmC/C and hmC/mC dyads; Fig. 3c[27–29]). A recent study reported proteomics screens with DNA promoter probes containing the symmetric and asymmetric CpG states C/C, mC/mC, hmC/hmC, hmC/mC and hmC/C.[107] Each probe version attracted a different set of readers, and a significant overlap between hmC/C *vs.* C/C and hmC/mC *vs.* mC/mC states hinted at the importance of the canonical C and mC bases for affinity. TFs with verified dyad state preferences included RFX5, MYC and MAX. Interestingly, the two latter proteins showed a preference for hmC in the form of MAX_2 homo- or MYC/MAX heterodimers, and a possible hmC–Arg interaction was proposed by a model (Fig. 6b). Nevertheless, the specific sequence in which this hmC reading occurs is still unknown (it did not occur in the E-box itself).[107] Moreover, a previous detailed study covering symmetric and hemi-modified CpG dyad states and the MAX_2 homodimer found differential read-out of different dyad states. Most importantly, caC was recognized within the E-box sequence with high affinity, and a crystal structure revealed involvement of the structurally conserved arginine (in addition to another arginine, Fig. 6b).[52]

Another caveat of proteomics screens is that pull-down probes typically cover a short, specific DNA sequence (*e.g.*, a promoter fragment), and thus do not contain all possible CpG contexts that may actually be bound by TFs. An improvement in this regard is advanced probe designs with a high density of TF target sequences.[106] In addition, a complementary, TF-centric approach named digital affinity profiling *via* proximity ligation (DAPPL) has been reported which allows for sampling all possible CpG sequence contexts in a high throughput assay involving >1000 recombinant human TFs.[108] Here, TF-GST fusion constructs are immobilized on GSH beads, and the GST is tethered to a dsDNA oligonucleotide. Then, a library of dsDNAs containing a central N_8CGN_8 sequence is incubated with the pooled TF-GST-bead collection, allowing for a ligation of the two dsDNAs in case of a binding event. Ligation products contain a TF-specific and a CpG dyad state-specific barcode, and can be amplified and pool-sequenced.

With this approach, consensus sequences and CpG state preferences for symmetric and hemi-modified CpGs containing mC and all oxi-mCs have been established for all involved TFs. In general, a higher number of readers was identified for hemi-modified as compared to symmetric CpGs (Fig. 6c), and readers spanned all major TF subfamilies (Fig. 6d). Interestingly, some

© 2026 The Author(s). Published by the Royal Society of Chemistry

CpG states could either increase or decrease the affinity of TFs dependent on the sequence context, with examples being PKNOX2, ETS1, or MLX. Taken together, these studies illustrate how specific oxi-mC dyad states can modulate important TF-DNA interactions (including their sequence preferences), which may serve as a basis for regulatory functions *in vivo*. An extension of proteomics screens to additional sequences and of the DAPPL approach to other dyad states would enable a more complete picture of this aspect.

5 Conclusions

Cytosine methylation is an essential regulator of mammalian chromatin, with important functions in cell differentiation, development and cancer. Oxi-mCs add another layer of complexity to this regulation which extends beyond their role as active demethylation intermediates. Each oxi-mC decorates the DNA major groove with a sterically unique and polar (in case of fC even electrophilic) functional group that provides unique regulatory potential. The functions of oxi-mCs reside on their influence on DNA's physicochemical properties, their interactions with chromatin proteins, and their genomic locations. A wealth of information is now available about these aspects in general. However, much less is known about the individual properties of different CpG dyad states, owing to a lack of broad, comparative studies. The ability to simultaneously sequence and map C, mC, and hmC CpG states will greatly contribute to a better understanding of their individual functions. In contrast, the impact of different CpG states on the stability, structure, and dynamics of the DNA duplex itself (particularly in the context of physiologically relevant sequences, such as densely modified CpG islands) is still poorly understood, and will require broader studies. First interactomes of individual CpG states are now available from proteomics and *in vitro* selection studies, but these are still far from complete in view of the sampled dyad states and sequence contexts. Most importantly, studying the actual physiological relevance of any of the above findings remains difficult. Whereas dyad-resolved sequencing/mapping methods represent a major step towards meaningful correlation studies, the mixed local occurrence of multiple CpG states and the comparatively poor resolution of complementary mapping methods (*e.g.*, ChIP-Seq) complicate the situation. Moreover, no perturbation methods are available to selectively induce particular dyad states *in vivo*. Given that an increasing number of studies indicate that different oxi-mC CpG dyad states are not functionally equivalent, and that they represent unique regulatory information, broader studies (particularly *in vivo*) are urgently required to fully unravel their individual functions in normal and disease states.

Author contributions

Z. V. C., L. E. and D. S wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (Su 726/9-1 and Su 726/10-1). We acknowledge support by the TU Dortmund and the International Max-Planck Research School for Living Matter (IMPRS-LM). Fig. 4–6 have been prepared with PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC. Fig. 6a and b was adapted from indicated reference under permission from the authors.

References

1 Z. D. Smith and A. Meissner, DNA methylation: roles in mammalian development, *Nat. Rev. Genet.*, 2013, **14**, 204–220.
2 A. Bird, DNA methylation patterns and epigenetic memory, *Genes Dev.*, 2002, **16**, 6–21.
3 J. A. Law and S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, *Nat. Rev. Genet.*, 2010, **11**, 204–220.
4 M. Rodriguez-Paredes and M. Esteller, Cancer epigenetics reaches mainstream oncology, *Nat. Med.*, 2011, **17**, 330–339.
5 S. Kriaucionis and N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, *Science*, 2009, **324**, 929–930.
6 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, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, *Science*, 2009, **324**, 930–935.
7 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 and G. L. Xu, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, *Science*, 2011, **333**, 1303–1307.
8 S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science*, 2011, **333**, 1300–1303.
9 T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier and T. Carell, The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA, *Angew. Chem., Int. Ed.*, 2011, **50**, 7008–7012.
10 A. Maiti and A. C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites, *J. Biol. Chem.*, 2011, **286**, 35334–35338.
11 H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang and X. Cheng, Recognition and potential mechanisms for replication

© 2026 The Author(s). Published by the Royal Society of Chemistry

and erasure of cytosine hydroxymethylation, *Nucleic Acids Res.*, 2012, **40**, 4841–4849.
12 J. Otani, H. Kimura, J. Sharif, T. A. Endo, Y. Mishima, T. Kawakami, H. Koseki, M. Shirakawa, I. Suetake and S. Tajima, Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells, *PLoS One*, 2013, **8**, e82961.
13 D. Ji, C. You, P. Wang and Y. Wang, Effects of tet-induced oxidation products of 5-methylcytosine on DNA replication in mammalian cells, *Chem. Res. Toxicol.*, 2014, **27**, 1304–1309.
14 C. L. Seiler, J. Fernandez, Z. Koerperich, M. P. Andersen, D. Kotandeniya, M. E. Nguyen, Y. Y. Sham and N. Y. Tretyakova, Maintenance DNA Methyltransferase Activity in the Presence of Oxidized Forms of 5-Methylcytosine: Structural Basis for Ten Eleven Translocation-Mediated DNA Demethylation, *Biochemistry*, 2018, **57**, 6061–6069.
15 A. Wei, H. Zhang, Q. Qiu, E. B. Fabyanic, P. Hu and H. Wu, 5-hydroxymethylcytosines regulate gene expression as a passive DNA demethylation resisting epigenetic mark in proliferative somatic cells, *bioRxiv*, 2023, preprint, DOI: 10.1101/2023.09.26.559662.
16 X. Wu and Y. Zhang, TET-mediated active DNA demethylation: mechanism, function and beyond, *Nat. Rev. Genet.*, 2017, **18**, 517–534.
17 F. R. Traube and T. Carell, The chemistries and consequences of DNA and RNA methylation and demethylation, *RNA Biol.*, 2017, 1–9, DOI: 10.1080/15476286.2017.1318241.
18 L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo and Y. Xu, Structural insight into substrate preference for TET-mediated oxidation, *Nature*, 2015, **527**, 118–122.
19 H. Tarhonskaya, A. M. Rydzik, I. K. H. Leung, N. D. Loik, M. C. Chan, A. Kawamura, J. S. O. McCullagh, T. D. W. Claridge, E. Flashman and C. J. Schofield, Non-enzymatic chemistry enables 2-hydroxyglutarate-mediated activation of 2-oxoglutarate oxygenases, *Nat. Commun.*, 2014, **5**, 1–10.
20 E. K. Schutsky, J. E. DeNizio, P. Hu, M. Y. Liu, C. S. Nabel, E. B. Fabyanic, Y. Hwang, F. D. Bushman, H. Wu and R. M. Kohli, Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase, *Nat. Biotechnol.*, 2018, **36**, 1083.
21 B. He, C. Zhang, X. Zhang, Y. Fan, H. Zeng, J. Liu, H. Meng, D. Bai, J. Peng, Q. Zhang, W. Tao and C. Yi, Tissue-specific 5-hydroxymethylcytosine landscape of the human genome, *Nat. Commun.*, 2021, **12**, 4249.
22 E. Tamanaha, S. Guan, K. Marks and L. Saleh, Distributive Processing by the Iron(II)/alpha-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases, *J. Am. Chem. Soc.*, 2016, **138**, 9345–9348.
23 C. B. Mulholland, F. R. Traube, E. Ugur, E. Parsa, E. M. Eckl, M. Schonung, M. Modic, M. D. Bartoschek, P. Stolz, J. Ryan, T. Carell, H. Leonhardt and S. Bultmann, Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency, *Sci. Rep.*, 2020, **10**, 12066.
24 D. J. Crawford, M. Y. Liu, C. S. Nabel, X. J. Cao, B. A. Garcia and R. M. Kohli, Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism, *J. Am. Chem. Soc.*, 2016, **138**, 730–733.
25 M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X. J. Cao, B. A. Garcia, G. A. Cisneros and R. M. Kohli, Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine, *Nat. Chem. Biol.*, 2017, **13**, 181–187.
26 J. E. DeNizio, M. Y. Liu, E. M. Leddin, G. A. Cisneros and R. M. Kohli, Selectivity and Promiscuity in TET-Mediated Oxidation of 5-Methylcytosine in DNA and RNA, *Biochemistry*, 2019, **58**, 411–421.
27 A. Chialastri, S. Sarkar, E. E. Schauer, S. Lamba and S. S. Dey, Combinatorial quantification of 5mC and 5hmC at individual CpG dyads and the transcriptome in single cells reveals modulators of DNA methylation maintenance fidelity, *Nat. Struct. Mol. Biol.*, 2024, **31**, 1296–1308.
28 D. O. Halliwell, F. Honig, S. Bagby, S. Roy and A. Murrell, Double and single stranded detection of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore sequencing, *Commun. Biol.*, 2025, **8**, 243.
29 J. S. Hardwick, S. Dhir, A. Kirchner, A. Simeone, S. M. Flynn, J. M. Edgerton, R. de Cesaris Araujo Tavares, I. Esain-Garcia, D. Tannahill, P. Golder, J. M. Monahan, W. S. Gosal and S. Balasubramanian, SCoTCH-seq reveals that 5-hydroxymethylcytosine encodes regulatory information across DNA strands, *Proc. Natl. Acad. Sci. U. S. A.*, 2025, **122**, e2512204122.
30 T. Carell, M. Q. Kurz, M. Müller, M. Rossa and F. Spada, Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA, *Angew Chem. Int. Ed. Engl.*, 2018, **57**, 4296–4312.
31 M. Bachman, S. Uribe-Lewis, X. P. Yang, M. Williams, A. Murrell and S. Balasubramanian, 5-Hydroxymethylcytosine is a predominantly stable DNA modification, *Nat. Chem.*, 2014, **6**, 1049–1055.
32 M. J. Booth, E. A. Raiber and S. Balasubramanian, Chemical methods for decoding cytosine modifications in DNA, *Chem. Rev.*, 2015, **115**, 2240–2254.
33 M. K. Bilyard, S. Becker and S. Balasubramanian, Natural, modified DNA bases, *Curr. Opin. Chem. Biol.*, 2020, **57**, 1–7.
34 C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin and C. He, Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming, *Cell*, 2013, **153**, 678–691.
35 L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang and Y. Zhang, Genome-wide Analysis Reveals TET- and TDG-Dependent 5-

© 2026 The Author(s). Published by the Royal Society of Chemistry

Methylcytosine Oxidation Dynamics, *Cell*, 2013, **153**, 692–706.
36 S. G. Jin, X. Wu, A. X. Li and G. P. Pfeifer, Genomic mapping of 5-hydroxymethylcytosine in the human brain, *Nucleic Acids Res.*, 2011, **39**, 5015–5024.
37 M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik and S. Balasubramanian, Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution, *Science*, 2012, **336**, 934–937.
38 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 and C. He, Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome, *Cell*, 2012, **149**, 1368–1380.
39 P. Giehr, C. Kyriakopoulos, K. Lepikhov, S. Wallner, V. Wolf and J. Walter, Two are better than one: HPoxBS - hairpin oxidative bisulfite sequencing, *Nucleic Acids Res.*, 2018, **46**, e88.
40 C. Kyriakopoulos, P. Giehr and V. Wolf, H(O)TA: estimation of DNA methylation and hydroxylation levels and efficiencies from time course data, *Bioinformatics*, 2017, **33**, 1733–1734.
41 C. X. Song, J. Diao, A. T. Brunger and S. R. Quake, Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 4338–4343.
42 R. Viswanathan, E. Cheruba, P. M. Wong, Y. Yi, S. Ngang, D. Q. Chong, Y. H. Loh, I. B. Tan and L. F. Cheow, DARESOME enables concurrent profiling of multiple DNA modifications with restriction enzymes in single cells and cell-free DNA, *Sci. Adv.*, 2023, **9**, eadi0197.
43 Y. Kawasaki, Y. Kuroda, I. Suetake, S. Tajima, F. Ishino and T. Kohda, A Novel method for the simultaneous identification of methylcytosine and hydroxymethylcytosine at a single base resolution, *Nucleic Acids Res.*, 2017, **45**, e24.
44 J. Fullgrabe, W. S. Gosal, P. Creed, S. Liu, C. K. Lumby, D. J. Morley, T. W. B. Ost, A. J. Vilella, S. Yu, H. Bignell, P. Burns, T. Charlesworth, B. Fu, H. Fordham, N. J. Harding, O. Gandelman, P. Golder, C. Hodson, M. Li, M. Lila, Y. Liu, J. Mason, J. Mellad, J. M. Monahan, O. Nentwich, A. Palmer, M. Steward, M. Taipale, A. Vandomme, R. S. San-Bento, A. Singhal, J. Vivian, N. Wojtowicz, N. Williams, N. J. Walker, N. C. H. Wong, G. N. Yalloway, J. D. Holbrook and S. Balasubramanian, Simultaneous sequencing of genetic and epigenetic bases in DNA, *Nat. Biotechnol.*, 2023, **41**, 1457–1464.
45 D. Bai, X. Zhang, H. Xiang, Z. Guo, C. Zhu and C. Yi, Simultaneous single-cell analysis of 5mC and 5hmC with SIMPLE-seq, *Nat. Biotechnol.*, 2025, **43**, 85–96.
46 B. Searle, M. Muller, T. Carell and A. Kellett, Third-Generation Sequencing of Epigenetic DNA, *Angew Chem. Int. Ed. Engl.*, 2023, **62**, e202215704.
47 E. Parasyraki, M. Mallick, V. Hatch, V. Vastolo, M. U. Musheev, E. Karaulanov, A. Gopanenko, S. Moxon, M. Mendez-Lago, D. Han, L. Schomacher, D. Mukherjee and C. Niehrs, 5-Formylcytosine is an activating epigenetic mark for RNA Pol III during zygotic reprogramming, *Cell*, 2024, **187**, 6088–6103.
48 E. A. Raiber, G. Portella, S. Martinez Cuesta, R. Hardisty, P. Murat, Z. Li, M. Iurlaro, W. Dean, J. Spindel, D. Beraldi, Z. Liu, M. A. Dawson, W. Reik and S. Balasubramanian, 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells, *Nat. Chem.*, 2018, **10**, 1258–1266.
49 F. Li, Y. Zhang, J. Bai, M. M. Greenberg, Z. Xi and C. Zhou, 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles, *J. Am. Chem. Soc.*, 2017, 10617–10620.
50 H. Hashimoto, Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang and X. Cheng, Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence, *Genes Dev.*, 2014, **28**, 2304–2313.
51 S. G. Jin, Z. M. Zhang, T. L. Dunwell, M. R. Harter, X. Wu, J. Johnson, Z. Li, J. Liu, P. E. Szabo, Q. Lu, G. L. Xu, J. Song and G. P. Pfeifer, Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration, *Cell Rep.*, 2016, **14**, 493–505.
52 D. X. Wang, H. Hashimoto, X. Zhang, B. G. Barwick, S. Lonial, L. H. Boise, P. M. Vertino and X. D. Cheng, MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma, *Nucleic Acids Res.*, 2017, **45**, 2396–2407.
53 M. W. Kellinger, C. X. Song, J. Chong, X. Y. Lu, C. He and D. Wang, 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription, *Nat. Struct. Mol. Biol.*, 2012, **19**, 831–833.
54 L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X. D. Fu and D. Wang, Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex, *Nature*, 2015, **523**, 621–625.
55 W. S. Li, X. Zhang, X. Y. Lu, L. You, Y. Q. Song, Z. G. Luo, J. Zhang, J. Nie, W. W. Zheng, D. N. Xu, Y. P. Wang, Y. Q. Dong, S. L. Yu, J. Hong, J. P. Shi, H. K. Hao, F. Luo, L. C. Hua, P. Wang, X. P. Qian, F. Yuan, L. H. Wei, M. Cui, T. P. Zhang, Q. Liao, M. H. Dai, Z. W. Liu, G. Chen, K. Meckel, S. Adhikari, G. F. Jia, M. B. Bissonnette, X. X. Zhang, Y. P. Zhao, W. Zhang, C. He and J. Liu, 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers, *Cell Res.*, 2017, **27**, 1243–1257.
56 C. X. Song, S. Yin, L. Ma, A. Wheeler, Y. Chen, Y. Zhang, B. Liu, J. Xiong, W. Zhang, J. Hu, Z. Zhou, B. Dong, Z. Tian, S. S. Jeffrey, M. S. Chua, S. So, W. Li, Y. Wei, J. Diao, D. Xie and S. R. Quake, 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages, *Cell Res.*, 2017, **27**, 1231–1242.
57 K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber and K. Helin, TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity, *Nature*, 2011, **473**, 343–348.

© 2026 The Author(s). Published by the Royal Society of Chemistry

58 G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews and W. Reik, Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation, *Nature*, 2011, **473**, 398–402.

59 H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun and Y. Zhang, Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells, *Genes Dev.*, 2011, **25**, 679–684.

60 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 and A. Rao, Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells, *Nature*, 2011, **473**, 394–397.

61 A. B. Robertson, J. A. Dahl, R. Ougland and A. Klungland, Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads, *Nat. Protoc.*, 2012, **7**, 340–350.

62 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 and C. He, Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine, *Nat. Biotechnol.*, 2011, **29**, 68–72.

63 B. C. Buchmuller, J. Droden, H. Singh, S. Palei, M. Drescher, R. Linser and D. Summerer, Evolved DNA Duplex Readers for Strand-Asymmetrically Modified 5-Hydroxymethylcytosine/5-Methylcytosine CpG Dyads, *J. Am. Chem. Soc.*, 2022, **144**, 2987–2993.

64 L. Engelhard, D. Schiller, M. S. Zambrano-Mila, K. Keliuotyte, B. Buchmuller, S. Tiwari, J. Imig, A. Simeone, C. Schroeter, S. Becker and D. Summerer, HM-DyadCap – Capture and Mapping of 5-Hydroxymethylcytosine/5-Methylcytosine CpG Dyads in Mammalian DNA, *bioRxiv*, 2025, preprint, DOI: 10.1101/2025.10.29.685270.

65 L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown and C. J. Schofield, Structural insights into how 5-hydroxymethylation influences transcription factor binding, *Chem. Commun.*, 2014, **50**, 1794–1796.

66 D. Renciuk, O. Blacque, M. Vorlickova and B. Spingler, Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine, *Nucleic Acids Res.*, 2013, **41**, 9891–9900.

67 M. W. Szulik, P. S. Pallan, B. Nocek, M. Voehler, S. Banerjee, S. Brooks, A. Joachimiak, M. Egli, B. F. Eichman and M. P. Stone, Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson-Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine, *Biochemistry*, 2015, **54**, 1294–1305.

68 S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schroder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld and T. Carell, Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine, *J. Am. Chem. Soc.*, 2013, **135**, 14593–14599.

69 V. Tereshko, G. Minasov and M. Egli, The Dickerson-Drew B-DNA dodecamer revisited at atomic resolution, *J. Am. Chem. Soc.*, 1999, **121**, 470–471.

70 T. Fu, L. Liu, Q. L. Yang, Y. Wang, P. Xu, L. Zhang, S. Liu, Q. Dai, Q. Ji, G. L. Xu, C. He, C. Luo and L. Zhang, Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine, *Chem. Sci.*, 2019, **10**, 7407–7417.

71 F. Battistini, P. D. Dans, M. Terrazas, C. L. Castellazzi, G. Portella, M. Labrador, N. Villegas, I. Brun-Heath, C. Gonzalez and M. Orozco, The Impact of the HydroxyMethylCytosine epigenetic signature on DNA structure and function, *PLoS Comput. Biol.*, 2021, **17**, e1009547.

72 R. C. A. Dubini, A. Schon, M. Muller, T. Carell and P. Rovo, Impact of 5-formylcytosine on the melting kinetics of DNA by 1H NMR chemical exchange, *Nucleic Acids Res.*, 2020, **48**, 8796–8807.

73 R. C. A. Dubini, E. Korytiakova, T. Schinkel, P. Heinrichs, T. Carell and P. Rovo, (1)H NMR Chemical Exchange Techniques Reveal Local and Global Effects of Oxidized Cytosine Derivatives, *ACS Phys. Chem. Au*, 2022, **2**, 237–246.

74 A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown and C. J. Schofield, Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation, *Chem. Commun.*, 2011, **47**, 5325–5327.

75 C. M. Lopez, A. J. Lloyd, K. Leonard and M. J. Wilkinson, Differential effect of three base modifications on DNA thermostability revealed by high resolution melting, *Anal. Chem.*, 2012, **84**, 7336–7342.

76 M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein and M. Drndic, Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules, *J. Am. Chem. Soc.*, 2011, **133**, 486–492.

77 E. A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi and S. Balasubramanian, 5-Formylcytosine alters the structure of the DNA double helix, *Nat. Struct. Mol. Biol.*, 2015, **22**, 44–49.

78 L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown and C. J. Schofield, Structural insights into how 5-hydroxymethylation influences transcription factor binding, *Chem. Commun.*, 2014, **50**, 1794–1796.

79 P. M. Severin, X. Zou, H. E. Gaub and K. Schulten, Cytosine methylation alters DNA mechanical properties, *Nucleic Acids Res.*, 2011, **39**, 8740–8751.

80 Q. X. Song, Z. M. Qiu, H. J. Wang, Y. M. Xia, J. Shen and Y. Zhang, The effect of methylation on the hydrogen-

© 2026 The Author(s). Published by the Royal Society of Chemistry

bonding and stacking interaction of nucleic acid bases, *Struct. Chem.*, 2013, **24**, 55–65.
81 L. C. Sowers, B. R. Shaw and W. D. Sedwick, Base Stacking and Molecular Polarizability - Effect of a Methyl-Group in the 5-Position of Pyrimidines, *Biochem. Biophys. Res. Commun.*, 1987, **148**, 790–794.
82 T. T. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev and T. Ha, Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability, *Nat. Commun.*, 2016, **7**, 10813.
83 D. Ji, K. Lin, J. Song and Y. Wang, Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation, *Mol. BioSyst.*, 2014, **10**, 1749–1752.
84 M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X. J. Cao, B. A. Garcia, G. A. Cisneros and R. M. Kohli, Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine, *Nat. Chem. Biol.*, 2017, **13**, 181–187.
85 A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kusnierczyk, C. B. Vagbo, D. Schuermann, A. Klungland and P. Schar, Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism, *Nat. Commun.*, 2016, **7**, 10806.
86 T. Zhou, J. Xiong, M. Wang, N. Yang, J. Wong, B. Zhu and R. M. Xu, Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2, *Mol. Cell*, 2014, **54**, 879–886.
87 S. Ji, H. Shao, Q. Han, C. L. Seiler and N. Tretyakova, Reversible DNA-protein cross-linking at epigenetic DNA marks, *Angew Chem. Int. Ed. Engl.*, 2017, **56**, 14130–14134.
88 G. P. Pfeifer, P. E. Szabo and J. K. Song, Protein Interactions at Oxidized 5-Methylcytosine Bases, *J. Mol. Biol.*, 2020, **432**, 1718–1730.
89 C. Rausch, F. D. Hastert and M. C. Cardoso, DNA Modification Readers and Writers and Their Interplay, *J. Mol. Biol.*, 2020, **432**, 1731–1746.
90 R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner and A. P. Bird, Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs, *Cell*, 1989, **58**, 499–507.
91 Q. Du, P. L. Luu, C. Stirzaker and S. J. Clark, Methyl-CpG-binding domain proteins: readers of the epigenome, *Epigenomics*, 2015, **7**, 1051–1073.
92 K. L. Ho, L. W. Mcnae, L. Schmiedeberg, R. J. Klose, A. P. Bird and M. D. Walkinshaw, MeCP2 binding to DNA depends upon hydration at methyl-CpG, *Mol. Cell*, 2008, **29**, 525–531.
93 B. C. Buchmuller, B. Kosel and D. Summerer, Complete Profiling of Methyl-CpG-Binding Domains for Combinations of Cytosine Modifications at CpG Dinucleotides Reveals Differential Read-out in Normal and Rett-Associated States, *Sci. Rep.*, 2020, **10**, 4053.
94 I. Ohki, N. Shimotake, N. Fujita, J. Jee, T. Ikegami, M. Nakao and M. Shirakawa, Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA, *Cell*, 2001, **105**, 487–497.
95 M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis and N. Heintz, MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system, *Cell*, 2012, **151**, 1417–1430.
96 O. Yildirim, R. W. Li, J. H. Hung, P. B. Chen, X. J. Dong, L. S. Ee, Z. P. Weng, O. J. Rando and T. G. Fazzio, Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells, *Cell*, 2011, **147**, 1498–1510.
97 C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Muller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Muller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell and M. Vermeulen, Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, *Cell*, 2013, **152**, 1146–1159.
98 M. Iurlaro, G. Ficz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian and W. Reik, A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation, *Genome Biol.*, 2013, **14**, R119.
99 V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird and L. C. Sowers, Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2), *Nucleic Acids Res.*, 2004, **32**, 4100–4108.
100 S. G. Jin, S. Kadam and G. P. Pfeifer, Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine, *Nucleic Acids Res.*, 2010, **38**, e125.
101 H. Zhu, G. H. Wang and J. Qian, Transcription factors as readers and effectors of DNA methylation, *Nat. Rev. Genet.*, 2016, **17**, 551–565.
102 J. Xiong, Z. Q. Zhang, J. Y. Chen, H. Huang, Y. L. Xu, X. J. Ding, Y. Zheng, R. Nishinakamura, G. L. Xu, H. L. Wang, S. Chen, S. R. Gao and B. Zhu, Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine, *Mol. Cell*, 2016, **64**, 913–925.
103 H. Hashimoto, Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang and X. D. Cheng, Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence, *Genes Dev.*, 2014, **28**, 2304–2313.
104 S. Khund-Sayeed, X. He, T. Holzberg, J. Wang, D. Rajagopal, S. Upadhyay, S. R. Durell, S. Mukherjee, M. T. Weirauch, R. Rose and C. Vinson, 5-Hydroxymethylcytosine in E-box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4, *Integr. Biol.*, 2016, **8**, 936–945.
105 M. Iurlaro, G. Ficz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian and W. Reik, A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation, *Genome Biol.*, 2013, **14**, R119.
106 L. Bai, G. Yang, Z. Qin, J. Lyu, Y. Wang, J. Feng, M. Liu, T. Gong, X. Li, Z. Li, J. Li, J. Qin, W. Yang and C. Ding,

© 2026 The Author(s). Published by the Royal Society of Chemistry

Proteome-Wide Profiling of Readers for DNA Modification, *Adv. Sci.*, 2021, **8**, e2101426.

107 L. Engelhard, Z. V. Cakil, S. Eppmann, T. Gonzalez, R. Linser, P. Janning and D. Summerer, Mammalian Proteome Profiling Reveals Readers and Antireaders of Strand-Symmetric and -Asymmetric 5-Hydroxymethylcytosine-Modifications in DNA, *bioRxiv*, 2025, preprint, DOI: 10.1101/2025.06.27.661915.

108 G. Song, G. Wang, X. Luo, Y. Cheng, Q. Song, J. Wan, C. Moore, H. Song, P. Jin, J. Qian and H. Zhu, An all-to-all approach to the identification of sequence-specific readers for epigenetic DNA modifications on cytosine, *Nat. Commun.*, 2021, **12**, 795.

© 2026 The Author(s). Published by the Royal Society of Chemistry