HHS Public Access

Author manuscript

Bioessays. Author manuscript; available in PMC 2017 October 01.

Published in final edited form as:

Bioessays. 2016 October ; 38(10): 1038–1047. doi:10.1002/bies.201600126.

Are there specific readers of oxidized 5-methylcytosine bases?

Jikui Song[1] and **Gerd P. Pfeifer**[2]

[1]Department of Biochemistry, University of California Riverside, Riverside, CA, USA

[2]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI, USA

Abstract

5-methylcytosine (5mC) was long thought to be the only enzymatically created modified DNA base in mammalian cells. The discovery of 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine as reaction products of the TET family 5mC oxidases has prompted extensive searches for proteins that specifically bind to these oxidized bases. However, only a few of such 'reader' proteins have been identified and verified so far. In this review, we discuss potential biological functions of oxidized 5mC as well as the role the presumed reader proteins may play in interpreting the genomic signals of 5mC oxidation products.

Keywords

5-carboxylcytosine; 5-formylcytosine; 5-hydroxymethylcytosine; 5-methylcytosine; chromatin readers; TET proteins

Introduction

5-methylcytosine (5mC) is the first modified DNA base to be identified. Hotchkiss, who detected a small and unexpected fifth peak upon paper chromatography of the four standard DNA bases derived from calf thymus [1], referred to this base as "epi-cytosine." He had discovered the first epigenetic modification. In 1975, several potential functional roles of 5mC were proposed. The methylated cytosine, which pairs normally with guanine, was viewed as potentially important in regulating protein-DNA interactions, gene expression, X chromosome inactivation and cellular differentiation and development [2, 3]. Cytosine methylation in vertebrates occurs predominantly at CpG dinucleotide sequences, which due to their symmetry provide a system for mother to daughter strand heritability of methylation patterns through cell division cycles [3]. In mammals, DNA methylation is essential because genetic deletion of DNA methyltransferases leads to developmental failure [4, 5]. DNA methylation at CpG-dense promoters is generally incompatible with gene expression and is often used for perpetuating the silenced state long-term, for example by negatively controlling the expression of repetitive or parasitic DNA elements, genes on the inactive X chromosome of female mammals, one of the two alleles of imprinted genes, and germ line-specific genes in somatic tissues [6]. There has been a long-standing debate as to whether

Corresponding authors: Jikui Song, jikui.song@ucr.edu, Gerd P. Pfeifer, gerd.pfeifer@vai.org.

The authors have declared no conflict of interest.

DNA methylation is the cause or the consequence of gene silencing. Probably both statements are true, if one considers that active transcription complexes or other DNA-bound, sequence-specific proteins exclude DNA methylation [7] but also that once methylation has occurred, the DNA becomes inaccessible to the transcription machinery. The methylation state at most CpG sequences (methylated or unmethylated) is generally quite stable but can change globally and site-specifically during development and disease. A cell-specific DNA methylation pattern is the combined result of several processes: de novo methylation of previously unmethylated sequences; maintenance of the methylated state; and loss of methylation, also referred to as DNA demethylation (Fig. 1A).

The 5mC oxidation - demethylation cycle

For many decades, 5-methylcytosine (5mC) had been viewed as the only enzymatically produced modified cytosine base in mammalian genomes. The situation changed in 2009, when 5-hydroxymethylcytosine (5hmC) was discovered in embryonic stem cells and in Purkinje cell neurons of the brain [8, 9]. Tahliani et al reported that 5hmC is produced from 5mC by a 5mC dioxygenase enzyme (Ten-Eleven Translocation 1, TET1) that requires alpha-ketoglutarate, oxygen and Fe2+ as cofactors [9]. Two related proteins, TET2 and TET3, were also identified and characterized as having the same catalytic activity [9, 10]. Since these initial reports were published, numerous studies have further investigated the properties and genomic distribution of 5hmC, and the TET proteins were analyzed in greater detail. Soon it was found that 5hmC is not always the final reaction product of TET enzymatic activity. Rather, these enzymes carry out sequential oxidation steps that lead to the formation of 5hmC, 5-formylcytosine (5fC) and then 5-carboxylcytosine (5caC) starting from 5mC in DNA as the initial substrate (Fig. 1B) [11, 12]. From the beginning, enzymatic 5mC oxidation was viewed as a logical pathway leading towards DNA demethylation [9].

Demethylation of CpG sequences, the principal target sequence of mammalian DNA methyltransferases, can occur either in a passive or in an active manner (Fig. 1A). Passive demethylation will simply occur by DNA replication in the absence of the maintenance DNA methyltransferase DNMT1 or its cofactor UHRF1. The hemimethylated site created immediately after DNA replication remains as such, and after a second round of replication, one daughter cell will carry a completely unmethylated CpG site. Passive demethylation plays important roles in certain biological contexts such as epigenome reprogramming after fertilization [13–15], DNA demethylation in primordial germ cells [16, 17], and conversion of embryonic stem cells to a naïve pluripotent state [18].

DNA strands that contain a 5hmC base at a CpG site on one strand are very poor substrates for DNMT1-mediated maintenance methylation [19, 20], suggesting that oxidation of 5mC to 5hmC represents a signal that strongly promotes passive DNA demethylation. The alternative pathway is active DNA demethylation, which is defined as a loss of 5mC in the absence of DNA replication (Fig. 1A). Before 2009, active DNA demethylation had remained an obscure phenomenon with many of the published and proposed mechanisms lacking further validation [21]. In 2011, He et al. identified a complete TET-initiated, active DNA demethylation pathway that consists of TET-mediated oxidation of 5mC to 5fC and 5caC and enzymatic removal by base excision repair of the latter two modified DNA bases

(Fig. 1B) [11]. The DNA repair process for 5fC and 5caC excision is initiated by thymine DNA glycosylase (TDG), a protein that can also remove thymine from T/G mismatches [22–24]. Curiously, unlike 5hmC, 5fC and 5caC occur at much lower levels in mammalian tissues or cell types where they have been quantitated [25, 26] suggesting that these oxidized bases are produced less efficiently by TET activity in vivo, or that they are removed by TDG once they have been formed. Both 5fC and 5caC are detectable in non-negligible quantities in embryonic stem cells, in particular upon depletion of TDG [11, 27–29], and there are indications that 5fC and 5caC accumulate in developing mammalian embryos [30, 31].

The limited accumulation of 5fC and 5caC has made it difficult to study active DNA demethylation processes in vivo. Although the TET- and TDG-dependent DNA demethylation pathway has been biochemically reconstituted in vitro [32], it is still difficult to pinpoint the occurrence of this mechanism in living cells. Currently, perhaps the best available evidence suggests that active DNA demethylation occurs during the preferential oxidation and remodeling processes of the paternal genome shortly after fertilization [13–15,33–35], during development and reprogramming of primordial germ cells [36], and during somatic cell reprogramming to create induced pluripotent stem cells [37]. The situation may be more complex, however, than initially anticipated because there is evidence for TDG-independent DNA demethylation in zygotes [13, 15].

The relatively high abundance of 5hmC in many cell types, most notably embryonic stem cells and neuronal cells, clearly indicates that TET-initiated 5mC oxidation does not always lead to active DNA demethylation. For example, in neurons of the adult human brain, 5hmC replaces 20–30% of all 5mC [38, 39] and appears to be a stable DNA base as measured by isotope labeling [40]. This situation implies that there is a regulatory step that normally prevents conversion of 5hmC to 5fC and 5caC and that this blockage is only rarely overcome leading to active DNA demethylation by subsequent removal of 5fC and 5caC. In vitro, TET enzymes are easily capable of converting 5mC to 5caC. Nevertheless, a recent study provided structural and enzymatic evidence that TET2-mediated oxidation has an intrinsic substrate preference for 5mC over 5hmC and 5fC [41], which may in part explain the relative abundance of the 5hmC mark in cells. Other factors that specify 5hmC as the predominant reaction product could perhaps depend on association of TET proteins with other proteins, which act as regulatory factors perhaps akin to regulation of DNMT1 by UHRF1 (e.g. by inducing a change in the catalytic pocket to accommodate 5hmC instead of 5mC or a change in configuration of the exocyclic groups on the cytosine base [41]). Posttranslational modification of TET proteins such as the recently described phosphorylation and O-GlcNacetylation [42] might also play a role in regulating TET protein substrate preference.

The 5mC oxidation pathway has relevance for human diseases and has been best studied in cancer. *TET2* is one of the most frequently mutated genes in several hematological malignancies including myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) [43]. The metabolic pathway leading to the production of α-ketoglutarate, a cofactor of TET enzymes, is aberrant in tumors with mutations occurring in isocitrate dehydrogenase 1 (IDH1), for example in gliomas, in which the IDH1 mutation strongly correlates with a CpG island hypermethylation phenotype [44]. Furthermore, the levels of 5hmC are

drastically lower in tumor tissue than in normal tissue, a phenomenon observed for many types of human solid tumors [38]. The data clearly suggest that 5mC oxidation pathways are perturbed in tumors although it remains to be determined in more detail if these aberrations have a direct causal relationship to tumorigenesis.

Several proteins function as readers of 5mC

Although the focus of this review is on potential reader proteins of oxidized 5mC bases, we provide a brief overview of proteins that specifically recognize 5mC. The first such protein complex was identified in 1989 by Adrian Bird's group and was named MECP1 [45]. The protein composition of the MECP1 complex was clarified later and this research showed that the protein MBD2 was the component of the complex that conferred binding specificity for 5mC [46]. In 1992, a single protein, MECP2, was found to bind to methylated DNA [47]. MECP2 and MBD2 are members of a small protein family that contain a common methyl-CpG binding domain (MBD). This family includes the confirmed 5mC binding proteins MBD1, MBD2, MBD4, and MECP2 [48]. A similar MBD domain is also found in MBD3, MBD5 and MBD6. However, these three proteins have not been described as selective 5mC binding proteins, either because their MBD domains are structurally sufficiently different or because these proteins may have a distinct DNA sequence context specificity that has not been elucidated yet.

MBD1, MBD2, MBD3 and MECP2 are components of transcriptional repressor complexes that also can include histone deacetylases and chromatin remodelers. Recruitment of these repressor complexes to methylated DNA will impose an inactive chromatin configuration onto the methylated genomic regions with the MBD proteins acting as readers of 5mC-enriched target sequences. *MBD4* is an outlier within this gene family. The MBD4 protein has been characterized as a DNA glycosylase that is specific for G/T mispairs resulting from deamination of 5mC at methylated CpG sequences [22]. Of note, mouse knockout phenotypes of 5mC readers are relatively mild [49–51] although removal of MECP2 in mice produces neurological symptoms similar to Rett syndrome [52].

Importantly, the MBD domains of MBD1, MBD2, MBD4 are incapable of binding to DNA sequences containing 5hmC [53] (please see below for data on MECP2). This finding suggests that the reader function of these domains is blocked by oxidation of 5mC to 5hmC and that transcriptional repression imposed by 5mC may be relieved by the oxidation step.

One other protein domain capable of binding to 5mC is the SRA domain. Plant genomes encode several proteins with SRA domains, but mammals have only two, UHRF1 and UHRF2 [54]. UHRF1, a ubiquitin ligase, appears to act as an integrator of epigenetic information encoded by DNA and histone modifications and is critical for maintenance of DNA methylation patterns through DNA replication [55]. The related protein UHRF2 has been characterized as a protein capable of binding to 5mC and 5hmC (see below).

Other categories of proteins with 5mC-binding activity include specific subsets of zinc finger proteins and a few other transcription factors [56–58]. These proteins are generally sequence-specific and could activate transcription from methylated promoters, in particular

within CpG-poor sequence contexts. The recently described uncoupling between methylated DNA and repression of transcription seems particularly relevant for certain enhancer sequences and might operate at specific stages of development or differentiation [57, 59].

Further discussion of the 5mC reader properties and biological functions of the 5mC-binding proteins is beyond the scope of this article but may be found in relevant reviews [48, 57, 60, 61].

Is 5hmC a negative mark?

The high levels of 5hmC in certain tissues of mammals, most notably the brain, suggest that this base plays a specific functional role. One such role may simply be a negative one where the formation of 5hmC from 5mC abolishes recognition of methylated CpG sequences by MBD proteins [53] or other readers of 5mC. In order to appreciate such a potential function of 5hmC, one needs to consider at what locations in the genome 5hmC is formed. There are three types of genomic regions at which 5hmC accumulation has been most often observed and has been well characterized: gene bodies (intragenic regions), enhancers, and sequences flanking promoters or CpG islands.

Mapping of 5hmC in the brain identified intragenic regions (gene bodies) of active genes as preferentially marked with this modified base [62–64]. Generally, there is a positive correlation between gene body 5hmC levels and gene expression levels. The reasons for this sequence preference of 5mC oxidation and its functional meaning is currently not understood. The earlier observed association of gene expression with 5mC levels in intragenic regions [65], has been interpreted as a mechanism to prevent low level antisense or noise transcription initiation within gene bodies. In this setting, formation of 5hmC may be even more effective in opposing such aberrant, spurious transcription that could interfere with the normal sense transcript. In such a scenario, 5hmC would likely represent a repelling mark that interferes with transcription initiation complexes.

Presence of 5hmC at enhancer regions marks poised or active enhancers [66–69]. Enhancer sequences are usually less CpG-rich than gene promoters, which often are embedded within CpG islands. However, methylation of CpGs within enhancers may also result in enhancer inactivity. Therefore, there is a need to counteract developmentally imposed or inappropriately occurring DNA methylation at enhancers by 5mC oxidation. At enhancers, 5hmC may be a negative mark that opposes the binding of repressors to 5mC sequences. Alternatively, although evidence is currently lacking, enhancer 5hmC may play a unique role in attracting activating factors that switch the enhancer into an either active or poised configuration. Finally, it is also possible that 5hmC at enhancers is an intermediate that occurs transiently (but continuously) in a pathway leading towards active demethylation, i.e. 5hmC is oxidized further to 5fC or 5caC. This scenario is supported by mapping studies in which 5fC was found to accumulate at poised enhancer sequences lending support to a model of ongoing TET-dependent DNA demethylation at enhancers [28, 29].

At sequences flanking promoters, 5hmC may mark regions of enhanced TET activity that functions to keep CpG island promoters free of DNA methylation. Over time, DNA

methylation errors in the form of DNMT-induced de novo methylation events will occur inevitably and could eventually lead to inappropriate gene silencing if there is no mechanism to oppose this process [70]. It is likely that such methylation errors will initially take place at the border regions of CpG islands (also referred to as island shores) due to enhanced accessibility of the shores relative to the centers of CpG islands. TET-mediated enzymatic oxidation of inappropriately placed 5mC at regions flanking CpG islands may lead to emergence of 5hmC at these shores and would therefore also represent mostly a negative signal that might interfere with binding of repressor protein complexes that interact with methylated DNA.

There may be a few readers of 5hmC

The alternative view is that there are specific reader proteins that recognize 5hmC and translate this base modification into a functional output. In fact, earlier studies have suggested that MBD3 [71], MECP2 [72] and the UHRF1 SRA domain [73] may be specific "readers" of 5hmC. However, more detailed follow-up work and also earlier studies have failed to support these findings [19, 74–79]. Mass spectrometry analysis has provided a relatively small list of proteins that seem to interact preferentially with 5hmC when comparing to 5mC or C-containing oligonucleotides [78, 79]. However, subsequent work to characterize and verify these interactions has been very limited. In some cases, binding of a putative 5hmC reader may not only be modification-specific but also DNA sequence-specific, which makes an overall assessment of the in vivo biological significance of the recognition specificity quite challenging.

One mass spectrometry study revealed that the SRA domain of UHRF2 specially interacts with 5hmC [78]. This discovery was subsequently supported by a structural and biochemical study [74], which demonstrated that the UHRF2-SRA domain, unlike the equivalent domain in UHRF1, does not preferentially bind to hemi-methylated DNA. Instead, it binds to fully-hydroxymethylated and hemi-hydroxymethylated DNA 3.2- and 1.5-fold more tightly than it does to hemi-methylated DNA, respectively. The structure of the UHRF2 SRA domain in complex with a 12-bp DNA with a central hemi-hydroxymethylated CpG site at 2.2 Å resolution further reveals that the 5hmC base is flipped out of the DNA helix and inserts into a UHRF2-SRA pocket, in which a hydrogen bond is formed between the hydroxyl group of 5hmC and the backbone carbonyl group of UHRF2 T508 or G509 (Fig. 2). Meanwhile, the aromatic side chains of UHRF2 F495 and Y507 in the pocket permit the accommodation of the bulky side chain of 5hmC (Fig. 2). In addition, the NHR finger loop, which plays an important role in the UHRF1-DNA contact, appears to be disordered in UHRF2, which therefore allows for its strong binding to fully-hydroxymethylated DNA [74].

In summary, except for the UHRF2-5hmC interaction, which is supported by a structural study, there is currently only very limited evidence for the existence of specific 5hmC readers encoded by mammalian genomes. This surprising situation contrasts with the abundance of 5hmC in certain cell types such as neurons in which its biological function remains enigmatic. A so far neglected approach would be the rigorous biochemical purification of 5hmC binding proteins from relevant tissues including the brain.

Are there readers of 5fC?

Previous mass spectrometry analysis based on an oligonucleotide pulldown approach has identified a number of candidate proteins that may recognize the 5fC and 5caC modifications [78,79]. Interestingly, the number of identified readers for 5fC and 5caC is much greater than the number of readers for 5hmC. However, none of these interactions have yet been confirmed by detailed biochemical evidence. Therefore, it still remains to be determined whether a true reader of 5fC exists. Nevertheless, genome-wide profiling of 5fC in mouse embryonic stem cells and in mouse embryos has previously revealed that its genomic location pattern is distinct from those of 5hmC and 5mC [28–30, 80]; 5fC preferentially appears at enhancer regions, where its production might coordinate with p300 in remodeling epigenetic states [28, 30]. Consistent with the idea of a specific function of 5fC rather than just being an intermediate in active DNA demethylation, recent studies have indicated that 5fC is a stable DNA modification in cells [26], and more importantly, it can facilitate the conformational transition of DNA from B form to a so-called F form leading to helical unwinding (Fig. 3A,B) [81]. This finding suggests that the 5fC modification may influence protein-DNA interactions by changing the shape of DNA, therefore implying a role of the 5fC modification in chromatin remodeling.

Readers of 5caC have recently been discovered and characterized

Despite the low abundance of 5caC in cells, emerging evidence has suggested that this modification may have significant impact on genome stability and gene regulation [29, 82–84]. Accordingly, a number of studies have demonstrated that 5caC can be specifically recognized by nuclear proteins that are involved in transcription or genomic maintenance [78,82–86]. Among these, human thymine DNA glycosylase (hTDG) was found to bind and remove the 5caC base through its DNA glycosylase activity [11, 86, 87]. Structural analysis of the hTDG catalytic mutant (N140A) with a 23-mer (3.0 Å resolution) or a 28-mer DNA duplex (2.6 Å resolution), containing an A·caC or G·caC site, respectively, revealed that the 5caC base is flipped out of the DNA helix and inserts into a pocket formed by A145, H150, H151, Y152 and N157 of hTDG (Fig. 4A) [85, 86]. Subsequently, its 5-carboxyl group forms hydrogen bonds with the backbone amide of H150 and the side chain of N157 under pH 7 buffer conditions [86] or N230 under pH 4.5 buffer conditions [85], and initiates a van der Waals contact with the side chain of A145 [86]. In addition, the hydrogen bond interactions of the O2 and N4 atoms of 5caC with the main chain amides of I139, A140 and N191, respectively, as well as the hydrophobic interaction between 5caC and Y152, further stabilize the configuration of the 5caC nucleotide within the hTDG pocket (Fig. 4A). In accordance with these structural observations, a binding affinity assay showed that hTDG has a 2-fold binding preference for 5caC-containing DNA over 5fC-containing DNA, and even stronger preference over the 5mC or 5hmC modifications [86].

Our recent studies further identified the CXXC domain of TET3 as a specific reader for 5caC in the context of CcaCG [83]. Using electrophoretic mobility shift assay (EMSA), we first confirmed a previous observation [88] that the CXXC domain of mouse TET3 (mTET3-CXXC) binds to unmodified cytosine at CpG sequences. Next, we demonstrated that mTET3-CXXC has an even stronger binding affinity for 5caC-containing DNA, but a much

lower binding affinity for other cytosine modifications, including 5mC, 5hmC and 5fC. Furthermore, isothermal titration calorimetry (ITC) assays indicated that mTET3-CXXC binds to the CcaCG DNA 3-fold more strongly than it does to the CCG DNA. This 3-fold binding preference for 5caC over unmethylated DNA is similar to the two-fold binding preferences of the UHRF1 SRA domain for hemi- over unmethylated DNA, for example [89] and of TDG for 5caC over 5fC [86], thereby establishing this domain as a specific reader for 5caC. On this basis, we have solved the crystal structure of mTET3-CXXC with a palindromic 12-mer DNA containing a central CcaCG site at 1.3 Å resolution, which provides atomic details on the interaction between mTET3-CXXC and CcaCG DNA. In particular, the 5-carboxyl group of 5caC from one strand forms direct and water-mediated hydrogen bonds with the side chain amino group of mTET3 K88 and the backbone amide of mTET3 I83, respectively (Fig. 4B). Meanwhile, the 5-carboxyl group of the 5caC from the complementary strand forms direct and water-mediated hydrogen bonds with the side chain hydroxyl groups of mTet3 T80 and S74, respectively, while its N4 atom forms a hydrogen bond with the side chain carbonyl oxygen of mTET3 Q82 (Fig. 4B). In addition, the C·G pair preceding the CpG dinucleotide is recognized by mTET3 H81 through hydrogen bond interactions. Note that these CcaCG-interacting residues are highly conserved within the TET protein-associated CXXC domains, but not in the subfamily of DNMT1 CXXC domains [83, 88], which were shown to exclusively recognize unmodified CpG dinucleotides [88, 90–93].

The low abundance of 5caC in genomes has limited a detailed investigation of the functional readout of 5caC. Nevertheless, genomic mapping of full-length TET3 in neuronal cells and embryonic mouse brain showed that TET3 is recruited to the transcription start sites of a limited set of genes [83]. These genes are mainly involved in lysosomal activities, RNA splicing and DNA base excision repair. These observations therefore prompted us to propose that the specific recognition of 5caC by TET3 CXXC plays an important role in maintenance of the methylation-free state of CpG islands of the TET3 target promoters. At a global genome scale, the 3-fold preference of TET3 for 5caC would probably be less relevant. However, in the context of localized 5mC oxidation by TET3, the creation of 5caC at the sites of TET3 activity would allow immediate anchoring of TET3 to 5caC, its reaction product [83]. The simultaneous presence of a writer domain (the C-terminal catalytic domain) and reader domain (the N-terminal CXXC domain) in TET3 is not unique among chromatin modifier proteins. This dual writer-reader domain architecture has also been observed in histone methyltransferases and histone demethylases [94–102]. Such a peculiar domain arrangement has been proposed as a mechanism to aid localized, processive action and to enforce genomic specificity of the modifiers in chromatin. Furthermore, it is of interest that CXXC-type zinc finger domains are found in only twelve mammalian proteins all of them being proven or presumed epigenetic regulators [103]. It will be interesting to see if a subset of these CXXC domains also has the capacity to bind to 5caC, or perhaps to other oxidized 5mC derivatives. Another unexplored possibility is that the binding of TET3 (or related CXXC domains) may shield 5caC from excision by TDG and may therefore aid in maintenance of this mark in certain biological contexts.

A small number of studies have now reported that 5caC may regulate gene transcription through modulating the recognition of transcriptional proteins with target DNA [82,84]. For

instance, Wilms tumor protein 1 (WT1), a transcription factor with a recognition sequence 5'-GCG(T/G)GGGCG-3', binds to the target sequence when C, 5mC or 5caC, but not 5hmC or 5fC, is present [82]. The structure of the WT1 in complex with a 5caC-containing DNA at 2.1 Å resolution reveals that the binding preference of WT1 for 5caC over 5hmC and 5fC is attributed to a hydrogen bond formed between WT1 Q369 and the 5-carboxyl group of 5caC (Fig. 4C). Such differential recognition of DNA modifications may provide a mechanism of fine-tuning gene expression at different genomic loci [82]. In addition, another study has recently shown that the presence of 5caC modifications in gene bodies leads to retarded RNA polymerase II-mediated transcription [84]. The crystal structure of the PolII-5caC DNA complex at 3.5 Å resolution revealed that the 5-carboxyl group of the 5caC base, with 50% of population, forms a hydrogen bond with residue Q531 of PolII Rbp2, which is located in a loop region termed "epi-DNA recognition loop" (Fig. 4D). Such recognition in turn leads to a positional shift for the incoming nucleotide and compromised RNA synthesis. This observation suggests that the presence of 5caC in gene bodies may have significant impact on gene transcription, different from its biological effect when present in promoter regions.

Conclusions

A few years after their initial discovery, oxidized 5-methylcytosine bases are still not understood in terms of their biological function. Considerable evidence suggests that these bases are intermediate products of an active, replication-independent DNA demethylation pathway. However, in many settings these oxidized 5mC bases are produced to directly promote replication-dependent loss of DNA methylation by inhibiting DNMT1. On the other hand, it has been difficult to understand why 5hmC (and 5fC) for the most part appear to be stable DNA modifications and not just intermediates that promote demethylation. One view has been that oxidized 5mC is chiefly a negative or repellent epigenetic mark that prevents the binding of a set of proteins that would otherwise interact with 5mC. Among the proteins identified as 'readers' of 5hmC, only few have been verified in follow-up studies, e.g. UHRF2, although a downstream biological effect of this binding is not apparent at this time. Even though 5fC and 5caC generally occur at low levels, they tend to accumulate in certain genomic locations, e.g. enhancers, where specific reader proteins may recognize these modifications. Proteomics approaches as well as targeted studies have identified a set of proteins that can bind to 5fC and 5caC. A few protein domains capable of binding to oxidized 5mC bases at CpG sequences have recently emerged (e.g. the SRA domain for 5hmC and CXXC domains for 5caC) and it will be interesting to see if such domains are unique or if additional heretofore unknown binding modules will be identified that could help solve the enigma of oxidized 5mC modifications in the genome.

Acknowledgments

Research of the authors is supported by NIH grant CA160965 (to G.P.P.), a Kimmel Scholar Award from the Sidney Kimmel Foundation for Cancer Research and by the March of Dimes Foundation (1-FY15-345) (to J.S.).

Abbreviations

5caC 5-carboxylcytosine

5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
MBD	methyl-binding domain
(h)TDG	(human) thymine DNA glycosylase

References

1. Hotchkiss RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem. 1948; 175:315–332. [PubMed: 18873306]
2. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975; 187:226–232. [PubMed: 1111098]
3. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975; 14:9–25. [PubMed: 1093816]
4. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992; 69:915–926. [PubMed: 1606615]
5. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999; 99:247–257. [PubMed: 10555141]
6. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13:484–492. [PubMed: 22641018]
7. Stadler MB, Murr R, Burger L, Ivanek R, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]
8. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009; 324:929–930. [PubMed: 19372393]
9. Tahiliani M, Koh KP, Shen Y, Pastor WA, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324:930–935. [PubMed: 19372391]
10. Ito S, D'Alessio AC, Taranova OV, Hong K, et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466:1129–1133. [PubMed: 20639862]
11. He YF, Li BZ, Li Z, Liu P, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011; 333:1303–1307. [PubMed: 21817016]
12. Ito S, Shen L, Dai Q, Wu SC, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011; 333:1300–1303. [PubMed: 21778364]
13. Guo F, Li X, Liang D, Li T, et al. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell stem cell. 2014; 15:447–458. [PubMed: 25220291]
14. Peat JR, Dean W, Clark SJ, Krueger F, et al. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep. 2014; 9:1990–2000. [PubMed: 25497087]
15. Shen L, Inoue A, He J, Liu Y, et al. Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. Cell Stem Cell. 2014; 15:459–470. [PubMed: 25280220]
16. Kagiwada S, Kurimoto K, Hirota T, Yamaji M, et al. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J. 2013; 32:340–353. [PubMed: 23241950]
17. Seisenberger S, Andrews S, Krueger F, Arand J, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012; 48:849–862. [PubMed: 23219530]

18. von Meyenn F, Iurlaro M, Habibi E, Liu NQ, et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. Mol Cell. 2016; 62:848–861. [PubMed: 27237052]
19. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012; 40:4841–4849. [PubMed: 22362737]
20. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 2007; 67:946–950. [PubMed: 17283125]
21. Ooi SK, Bestor TH. The colorful history of active DNA demethylation. Cell. 2008; 133:1145–1148. [PubMed: 18585349]
22. Bellacosa A, Drohat AC. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair. 2015; 32:33–42. [PubMed: 26021671]
23. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013; 502:472–479. [PubMed: 24153300]
24. Schuermann D, Weber AR, Schar P. Active DNA demethylation by DNA repair: Facts and uncertainties. DNA repair. 2016 in press.
25. Globisch D, Munzel M, Muller M, Michalakis S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PloS One. 2010; 5:e15367. [PubMed: 21203455]
26. Bachman M, Uribe-Lewis S, Yang X, Burgess HE, et al. 5-Formylcytosine can be a stable DNA modification in mammals. Nat Chem Biol. 2015; 11:555–557. [PubMed: 26098680]
27. Raiber EA, Beraldi D, Ficz G, Burgess HE, et al. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. Genome Biol. 2012; 13:R69. [PubMed: 22902005]
28. Song CX, Szulwach KE, Dai Q, Fu Y, et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell. 2013; 153:678–691. [PubMed: 23602153]
29. Shen L, Wu H, Diep D, Yamaguchi S, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell. 2013; 153:692–706. [PubMed: 23602152]
30. Iurlaro M, McInroy GR, Burgess HE, Dean W, et al. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. Genome Biol. 2016; 17:141. [PubMed: 27356509]
31. Wheldon LM, Abakir A, Ferjentsik Z, Dudnakova T, et al. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. Cell Rep. 2014; 7:1353–1361. [PubMed: 24882006]
32. Weber AR, Krawczyk C, Robertson AB, Kusnierczyk A, et al. Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nat Commun. 2016; 7:10806. [PubMed: 26932196]
33. Iqbal K, Jin SG, Pfeifer GP, Szabo PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci USA. 2011; 108:3642–3647. [PubMed: 21321204]
34. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011; 2:241. [PubMed: 21407207]
35. Gu TP, Guo F, Yang H, Wu HP, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011; 477:606–610. [PubMed: 21892189]
36. Kawasaki Y, Lee J, Matsuzawa A, Kohda T, et al. Active DNA demethylation is required for complete imprint erasure in primordial germ cells. Sci Rep. 2014; 4:3658. [PubMed: 24413819]
37. Hu X, Zhang L, Mao SQ, Li Z, et al. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell. 2014; 14:512–522. [PubMed: 24529596]
38. Jin SG, Jiang Y, Qiu R, Rauch TA, et al. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res. 2011; 71:7360–7365. [PubMed: 22052461]

39. Münzel M, Globisch D, Bruckl T, Wagner M, et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl. 2010; 49:5375–5377. [PubMed: 20583021]
40. Bachman M, Uribe-Lewis S, Yang X, Williams M, et al. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem. 2014; 6:1049–1055. [PubMed: 25411882]
41. Hu L, Lu J, Cheng J, Rao Q, et al. Structural insight into substrate preference for TET-mediated oxidation. Nature. 2015; 527:118–122. [PubMed: 26524525]
42. Bauer C, Gobel K, Nagaraj N, Colantuoni C, et al. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). J Biol Chem. 2015; 290:4801–4812. [PubMed: 25568311]
43. Ko M, An J, Rao A. DNA methylation and hydroxymethylation in hematologic differentiation and transformation. Curr Opin Cell Biol. 2015; 37:91–101. [PubMed: 26595486]
44. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010; 17:510–522. [PubMed: 20399149]
45. Meehan RR, Lewis JD, McKay S, Kleiner EL, et al. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 1989; 58:499–507. [PubMed: 2758464]
46. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, et al. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999; 13:1924–1935. [PubMed: 10444591]
47. Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. 1992; 20:5085–5092. [PubMed: 1408825]
48. Du Q, Luu PL, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics. 2015; 7:1051–1073. [PubMed: 25927341]
49. Hendrich B, Guy J, Ramsahoye B, Wilson VA, et al. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001; 15:710–723. [PubMed: 11274056]
50. Wong E, Yang K, Kuraguchi M, Werling U, et al. Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. Proc Natl Acad Sci USA. 2002; 99:14937–14942. [PubMed: 12417741]
51. Zhao X, Ueba T, Christie BR, Barkho B, et al. Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc Natl Acad Sci USA. 2003; 100:6777–6782. [PubMed: 12748381]
52. Guy J, Hendrich B, Holmes M, Martin JE, et al. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet. 2001; 27:322–326. [PubMed: 11242117]
53. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010; 38:e125. [PubMed: 20371518]
54. Johnson LM, Bostick M, Zhang X, Kraft E, et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007; 17:379–384. [PubMed: 17239600]
55. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015; 16:519–532. [PubMed: 26296162]
56. Buck-Koehntop BA, Defossez PA. On how mammalian transcription factors recognize methylated DNA. Epigenetics. 2013; 8:131–137. [PubMed: 23324617]
57. Spruijt CG, Vermeulen M. DNA methylation: old dog, new tricks? Nat Struct Mol Biol. 2014; 21:949–954. [PubMed: 25372310]
58. Hu S, Wan J, Su Y, Song Q, et al. DNA methylation presents distinct binding sites for human transcription factors. eLife. 2013; 2:e00726. [PubMed: 24015356]
59. Charlet J, Duymich CE, Lay FD, Mundbjerg K, et al. Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. Mol Cell. 2016; 62:422–431. [PubMed: 27153539]
60. Dhasarathy A, Wade PA. The MBD protein family-reading an epigenetic mark? Mutation Res. 2008; 647:39–43. [PubMed: 18692077]

61. Wood KH, Zhou Z. Emerging molecular and biological functions of MBD2, a reader of DNA methylation. Front Genet. 2016; 7:93. [PubMed: 27303433]
62. Song CX, Szulwach KE, Fu Y, Dai Q, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011; 29:68–72. [PubMed: 21151123]
63. Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. 2011; 39:5015–5024. [PubMed: 21378125]
64. Hahn MA, Qiu R, Wu X, Li AX, et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep. 2013; 3:291–300. [PubMed: 23403289]
65. Rauch TA, Wu X, Zhong X, Riggs AD, et al. A human B cell methylome at 100-base pair resolution. Proc Natl Acad Sci USA. 2009; 106:671–678. [PubMed: 19139413]
66. Stroud H, Feng S, Morey Kinney S, Pradhan S, et al. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 2011; 12:R54. [PubMed: 21689397]
67. Hon GC, Song CX, Du T, Jin F, et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell. 2014; 56:286–297. [PubMed: 25263596]
68. Lu F, Liu Y, Jiang L, Yamaguchi S, et al. Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. 2014; 28:2103–2119. [PubMed: 25223896]
69. Serandour AA, Avner S, Oger F, Bizot M, et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic Acids Res. 2012; 40:8255–8265. [PubMed: 22730288]
70. Williams K, Christensen J, Helin K. DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep. 2012; 13:28–35.
71. Yildirim O, Li R, Hung JH, Chen PB, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011; 147:1498–1510. [PubMed: 22196727]
72. Mellen M, Ayata P, Dewell S, Kriaucionis S, et al. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell. 2012; 151:1417–1430. [PubMed: 23260135]
73. Frauer C, Hoffmann T, Bultmann S, Casa V, et al. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS One. 2011; 6:e21306. [PubMed: 21731699]
74. Zhou T, Xiong J, Wang M, Yang N, et al. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell. 2014; 54:879–886. [PubMed: 24813944]
75. Otani J, Kimura H, Sharif J, Endo TA, et al. Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. PLoS One. 2013; 8:e82961. [PubMed: 24340069]
76. Khrapunov S, Warren C, Cheng H, Berko ER, et al. Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. Biochemistry. 2014; 53:3379–3391. [PubMed: 24828757]
77. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, et al. 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. [PubMed: 15302911]
78. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013; 152:1146–1159. [PubMed: 23434322]
79. Iurlaro M, Ficz G, Oxley D, Raiber EA, et al. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol. 2013; 14:R119. [PubMed: 24156278]
80. Booth MJ, Marsico G, Bachman M, Beraldi D, et al. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nat Chem. 2014; 6:435–440. [PubMed: 24755596]
81. Raiber EA, Murat P, Chirgadze DY, Beraldi D, et al. 5-Formylcytosine alters the structure of the DNA double helix. Nat Struct Mol Biol. 2015; 22:44–49. [PubMed: 25504322]
82. Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, et al. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes Dev. 2014; 28:2304–2313. [PubMed: 25258363]

83. Jin SG, Zhang ZM, Dunwell TL, Harter MR, et al. Tet3 reads 5-carboxylcytosine through its CXXC domain and is a potential guardian against neurodegeneration. Cell Rep. 2016; 14:493–505. [PubMed: 26774490]
84. Wang L, Zhou Y, Xu L, Xiao R, et al. Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature. 2015; 523:621–625. [PubMed: 26123024]
85. Hashimoto H, Zhang X, Cheng X. Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. DNA Repair. 2013; 12:535–540. [PubMed: 23680598]
86. Zhang L, Lu X, Lu J, Liang H, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol. 2012; 8:328–330. [PubMed: 22327402]
87. Maiti A, Drohat AC. 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. [PubMed: 21862836]
88. Xu Y, Xu C, Kato A, Tempel W, et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell. 2012; 151:1200–1213. [PubMed: 23217707]
89. Rottach A, Frauer C, Pichler G, Bonapace IM, et al. The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res. 2010; 38:1796–1804. [PubMed: 20026581]
90. Allen MD, Grummitt CG, Hilcenko C, Min SY, et al. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. EMBO J. 2006; 25:4503–4512. [PubMed: 16990798]
91. Cierpicki T, Risner LE, Grembecka J, Lukasik SM, et al. Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat Struct Mol Biol. 2010; 17:62–68. [PubMed: 20010842]
92. Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011; 331:1036–1040. [PubMed: 21163962]
93. Xu C, Bian C, Lam R, Dong A, et al. The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. Nat Commun. 2011; 2:227. [PubMed: 21407193]
94. Dumesic PA, Homer CM, Moresco JJ, Pack LR, et al. Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell. 2015; 160:204–218. [PubMed: 25533783]
95. Hansen KH, Bracken AP, Pasini D, Dietrich N, et al. A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol. 2008; 10:1291–1300. [PubMed: 18931660]
96. Hassan AH, Prochasson P, Neely KE, Galasinski SC, et al. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell. 2002; 111:369–379. [PubMed: 12419247]
97. Klein BJ, Piao L, Xi Y, Rincon-Arano H, et al. The histone-H3K4-specific demethylase KDM5B binds to its substrate and product through distinct PHD fingers. Cell Rep. 2014; 6:325–335. [PubMed: 24412361]
98. Lan F, Collins RE, De Cegli R, Alpatov R, et al. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature. 2007; 448:718–722. [PubMed: 17687328]
99. Liu N, Zhang Z, Wu H, Jiang Y, et al. Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. Genes Dev. 2015; 29:379–393. [PubMed: 25637356]
100. Margueron R, Justin N, Ohno K, Sharpe ML, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 2009; 461:762–767. [PubMed: 19767730]
101. Torres IO, Kuchenbecker KM, Nnadi CI, Fletterick RJ, et al. Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism. Nat Commun. 2015; 6:6204. [PubMed: 25686748]

102. Zhang K, Mosch K, Fischle W, Grewal SI. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat Struct Mol Biol. 2008; 15:381–388. [PubMed: 18345014]

103. Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans. 2013; 41:727–740. [PubMed: 23697932]

104. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]

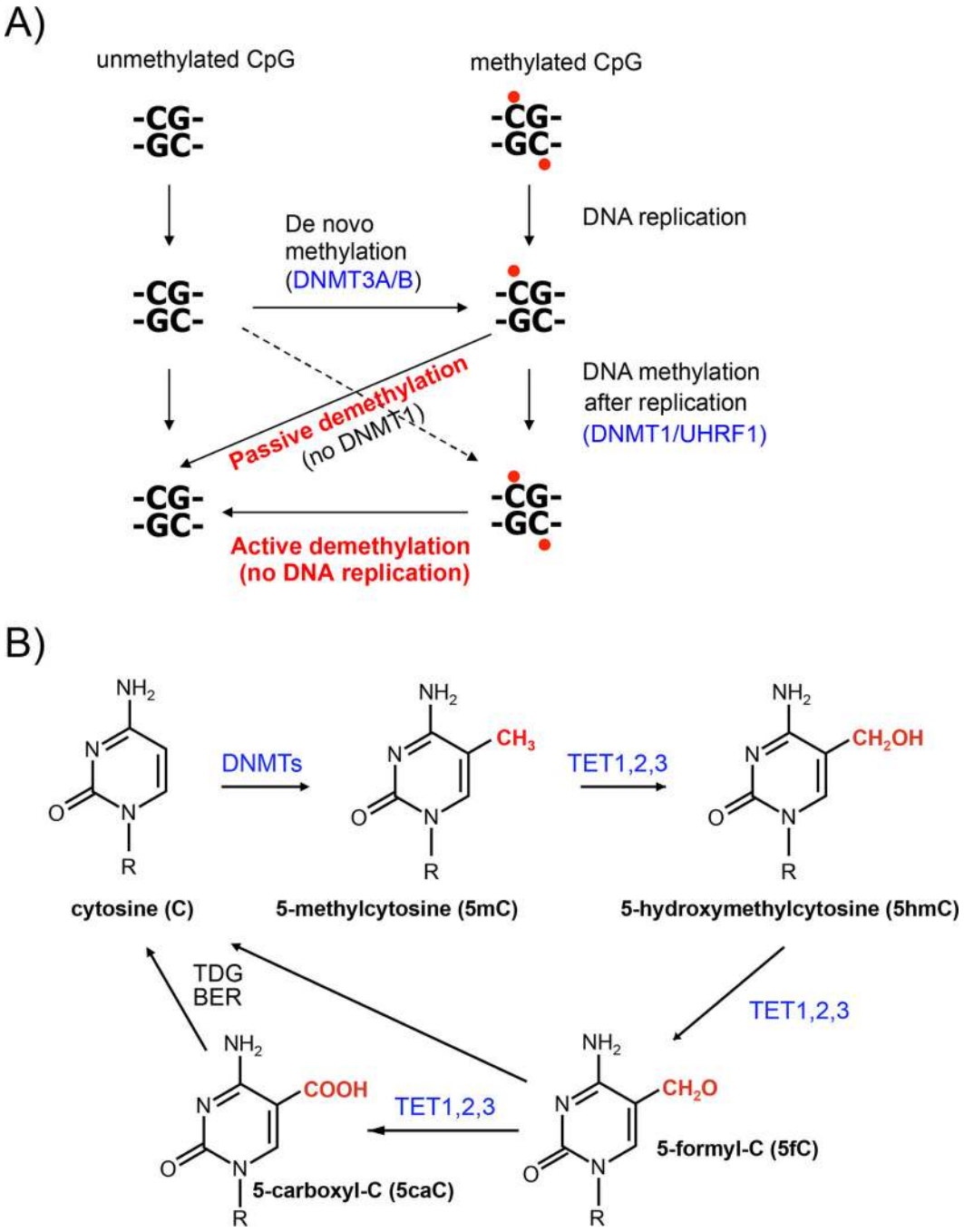

Figure 1.

Outline of DNA demethylation pathways and 5mC oxidation. **A:** Models for passive, DNA replication-dependent DNA demethylation and active, DNA replication-independent DNA demethylation. **B:** The model shows details of the TET-initiated, active DNA demethylation pathway. TET proteins oxidize 5mC leading to the formation of 5hmC, 5fC and 5caC. The oxidized bases 5fC and 5caC are recognized and excised from DNA by thymine DNA glycosylase (TDG) as part of a base excision repair (BER) process.

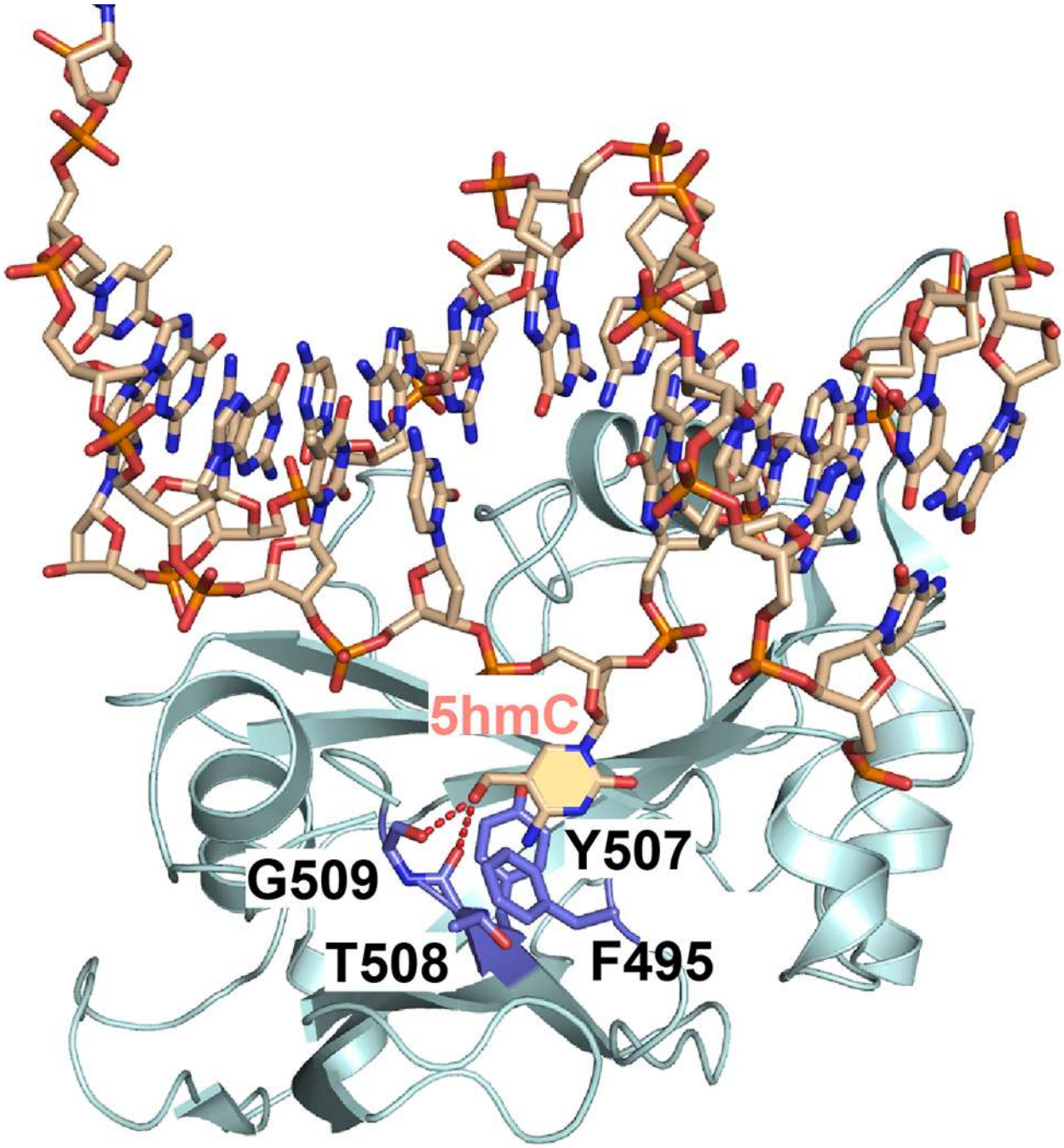

Figure 2.
Crystal structure of the UHRF2 SRA domain in complex with 5hmC-containing DNA (PDB 4PW5). The 5hmC base ring is shaded in beige.

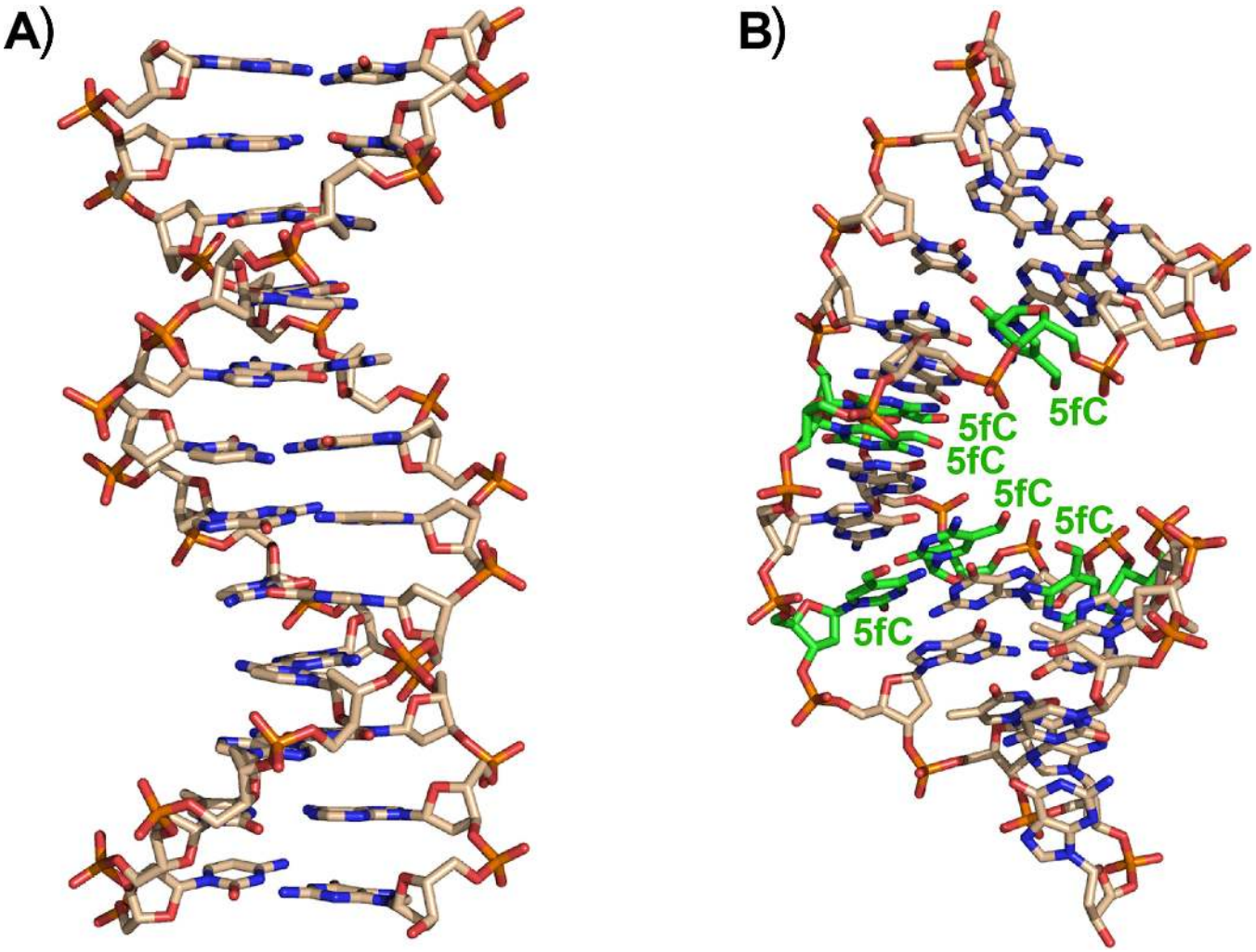

Figure 3.

Structural comparison of 5fC-containing DNA with B-form DNA. **A:** Model for standard B-form DNA generated using the Coot software [104]. **B:** Crystal structure of the 5fC-containing DNA (PDB 4QKK). Note that 5fC does not affect the overall B-DNA structure of a hemi-formylated (two hemi-5fC sites) Dickerson-Drew DNA (PDB 1VE8). However, the 5fC sites in this hemi-5fC structure show unusual local rotational and translocational parameters similar to the six 5fC-containing DNA [81].

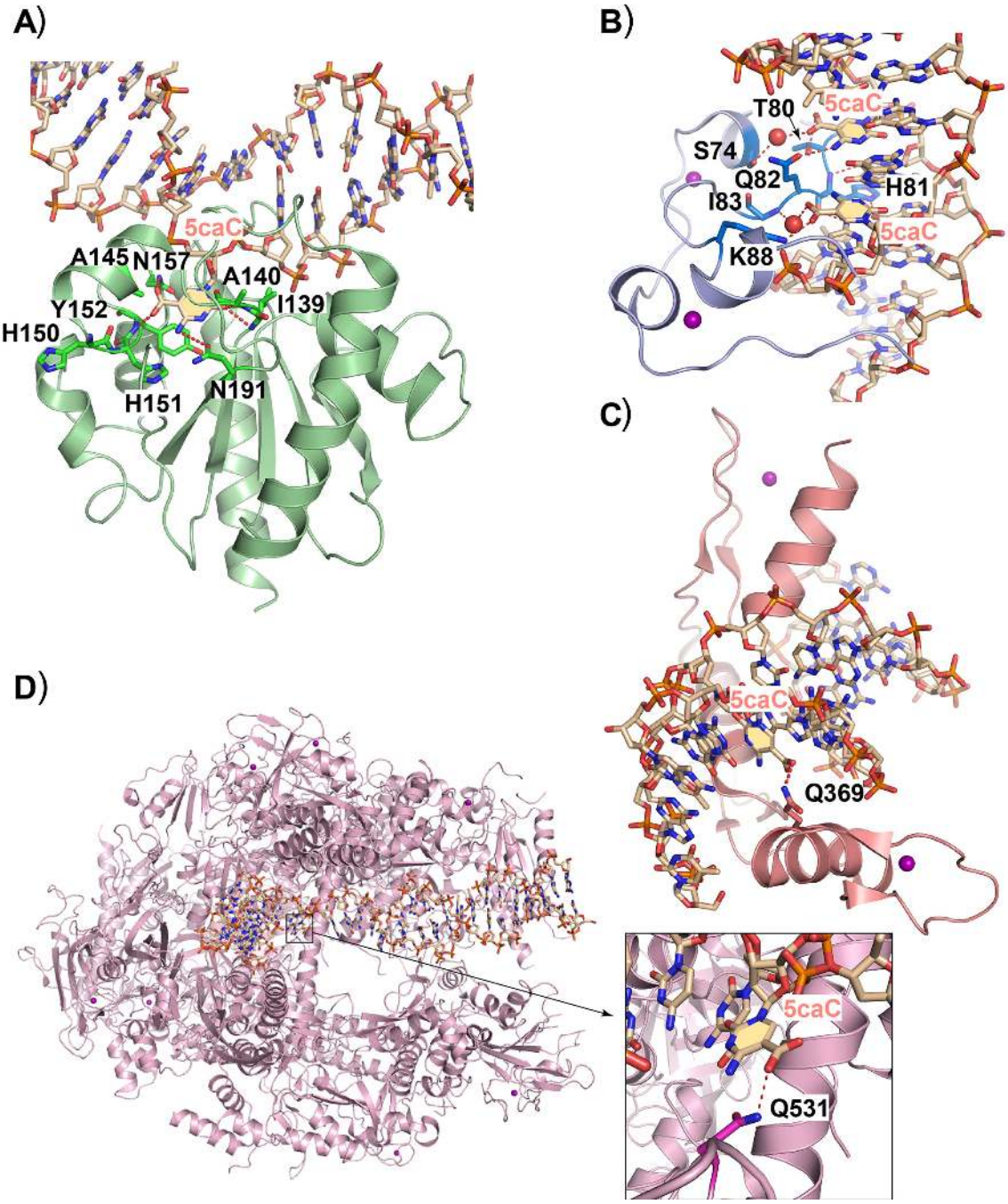

Figure 4.

Crystal structures of proteins in complex with 5caC-containing DNA. **A:** hTDG-5caC DNA (PDB 3UO7). **B:** Tet3 CXXC-5caC DNA (PDB 5EXH). **C:** WT1-5caC DNA (PDB 4R2R). **D:** RNA PolII-5caC DNA (PDB 4Y52). The 5caC base ring is shaded in beige. The water molecules and zinc ions are shown in sphere representation.