Recent Advances on DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications

12

Ren Ren, John R. Horton, Samuel Hong, and Xiaodong Cheng

Abstract

The modification of DNA bases is a classic hallmark of epigenetics. Four forms of modified cytosine—5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine—have been discovered in eukaryotic DNA. In addition to cytosine carbon-5 modifications, cytosine and adenine methylated in the exocyclic amine—N4-methylcytosine and N6-methyladenine—are other modified DNA bases discovered even earlier. Each modified base can be considered a distinct epigenetic signal with broader biological implications beyond simple chemical changes. Since 1994, several crystal structures of proteins and enzymes involved in writing, reading, and erasing modified bases have become available. Here, we present a structural synopsis of writers, readers, and erasers of the modified bases from prokaryotes and eukaryotes. Despite significant differences in structures and functions, they are remarkably similar regarding their engagement in flipping a target base/nucleotide within DNA for specific recognitions and/or reactions. We thus highlight base flipping as a common structural framework broadly applied by distinct classes of proteins and enzymes across phyla for epigenetic regulations of DNA.

Keywords

Epigenetic methylation · DNA base flipping · Reader, writer and eraser of DNA methyl marks · Methyltransferases · Demethylases

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5ghmC	Glucosylated 5-hydroxymethylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
αKG	α-ketoglutarate
AdoHcy	*S*-adenosyl-L-homocysteine (SAH)
AdoMet	*S*-adenosyl-L-methionine (SAM)
AlkB	*E. coli* Alkylated DNA repair protein AlkB
ALKBH5	Alkylated DNA repair protein AlkB homolog 5 in human
CMT2	Chromomethylase 2 (plant specific)

Ren Ren and John R. Horton contributed equally with all other contributors.

R. Ren · J. R. Horton · X. Cheng (✉)
Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX, USA
e-mail: RRen1@mdanderson.org; JRHorton@mdanderson.org; xcheng5@mdanderson.org

S. Hong
McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA
e-mail: samuel.hong@uth.tmc.edu

A. Jeltsch, R. Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 1389, https://doi.org/10.1007/978-3-031-11454-0_12

CMT3	Chromomethylase 3 (plant specific)
DME	Demeter (plant)
DML3	Demeter-like protein 3 (plant)
DNMT1	Mammalian DNA methyltransferase 1
DNMT3A	Mammalian DNA methyltransferase 3A
DNMT3L	Mammalian DNA methyltransferase 3-like
DRM2	Domain rearranged methyltransferase 2 (plant)
FTO	Fat mass and obesity-associated protein
HhH	Helix-hairpin-helix
JBP	J-binding protein
MBD	Methyl-CpG binding domain
McrB	Modified cytosine restriction B
Met1	DNA methyltransferase 1 (plant)
MTase	Methyltransferase
N4mC	N4-methylcytosine
N6mA	N6-methyladenine
NOG	*N*-oxalylglycine
ROS1	Repressor of silencing 1 (plant specific)
SRA	SET- and RING-associated
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TRD	Target recognition domain
Uhrf1	Ubiquitin-like-containing PHD and RING finger domains protein 1
WH	Winged helix

12.1 Introduction

Chemical modifications of DNA bases have fundamental biological roles in virtually every living organism. In both prokaryotes and many eukaryotes, cytosine can be methylated at carbon-5 (C5) position by cytosine C5 methyltransferases (MTases) to generate 5-methylcytosine (5mC) (Kumar et al. 1994; Bestor 2000) (Fig. 12.1a). In higher eukaryotes, ten-eleven translocation (TET) 5mC dioxygenase enzymes utilize α-ketoglutarate (αKG) and Fe (II) to oxidize the methyl group of 5mC to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) via discrete reactions (Kriaucionis and Heintz 2009; Tahiliani et al. 2009; Ito et al. 2011; He et al. 2011). In prokaryotes, 5mC and 5hmC can be introduced de novo into the genome during phage invasions, as both modified bases can be synthesized prior to incorporation into the phage genome during DNA synthesis (Warren 1980). After DNA synthesis, phage glucosyltransferases can modify 5hmC within the genome to generate glucosylated 5hmC (5ghmC) (Lehman and Pratt 1960; Kornberg et al. 1961; Lunt et al. 1964). Beyond cytosine C5 modifications, exocyclic amine groups of adenine and cytosine can be methylated in prokaryotes to generate N6-methyladenine (N6mA) and N4-methylcytosine (N4mC), respectively (Malone et al. 1995) (Fig. 12.1b, c). Crystal structures of DNA modification enzymes to date have consistently shown that the target nucleotide is flipped out of the double helix for reactions in a process called base flipping.

In addition to the modification writers, modified base readers have also been shown to flip the target base for recognition. Mammalian SET- and RING-associated (SRA) domains recognize 5mC within genome by base flipping (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008), and have been characterized as non-enzymatic base flippers. Since the first discovery in eukaryotes, SRA has been rediscovered in prokaryotes, recognizing 5mC, 5hmC, and/or 5ghmC to coordinate restriction activity in a modification-dependent manner (Horton et al. 2014a–c). In addition to SRA, the bacterial modified cytosine restriction B enzyme (McrB) also flips 5mC for recognition but is structurally distinct from other known base flippers (Sukackaite et al. 2012). Structural homologs of McrB across different phyla may recognize modified bases in a similar way.

A brief survey of DNA base modifications in both prokaryotes and eukaryotes reveals that two major families of enzymes, methyltransferases, and dioxygenases are involved in writing DNA modifications in the four forms of modified cytosine: 5mC, 5hmC, 5fC, and 5caC. In plants, 5mC DNA glycosylase repressor of silencing 1 (ROS1) can excise 5mC and 5hmC (in vitro) (Gong et al.

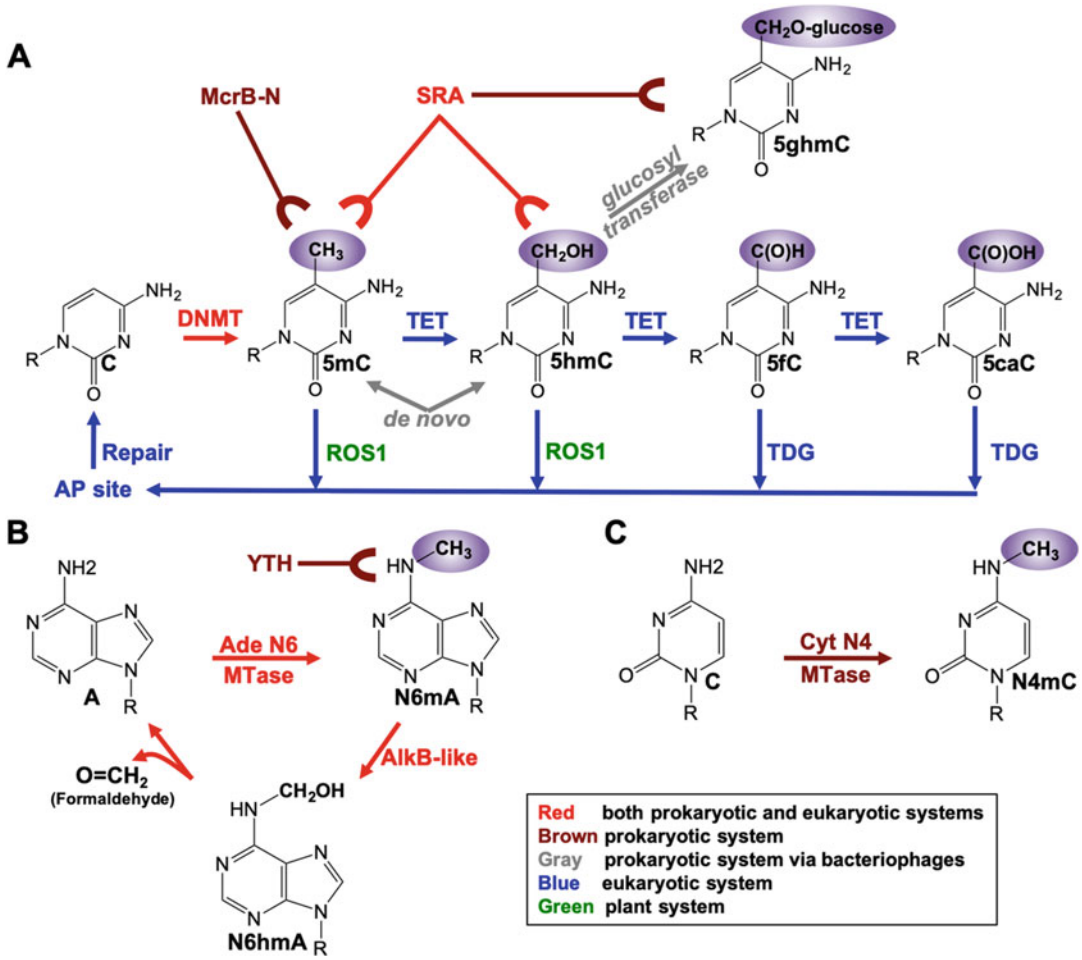

Fig. 12.1 Chemical modifications of nucleic acids. (**a**) DNA cytosine C5 modifications: enzymes and proteins involved in writing, reading, and erasing the modifications via base-flipping mechanisms. (**b**) Adenine N6 methylation in DNA or RNA: enzymes and proteins involved in writing, reading, and erasing adenine N6 methylation. (**c**) Cytosine N4 methylation

2002; Jang et al. 2014; Hong et al. 2014), and mammalian thymine DNA glycosylase (TDG) can excise 5fC and 5caC (He et al. 2011; Maiti and Drohat 2011; Zhang et al. 2012; Hashimoto et al. 2012a). These discoveries effectively link the base excision repair pathway, including AlkB homologs (see below), to DNA demethylation/demodification, by which epigenetic signals encoded in the modified cytosines can be reversed. DNA glycosylases represent the most structurally diverse family of enzymes that are involved in base flipping (also known as nucleotide flipping) (Brooks et al. 2013). Thus, base flipping is not restricted to writers and readers, but has been adopted by DNA glycosylases for erasing DNA modifications as well. Together, structural characterizations of writers, readers, and erasers of DNA base modifications in prokaryotes and eukaryotes effectively showcase base flipping as a general mechanism for regulating and translating fundamental epigenetic signals as well as for maintaining genome integrity (i.e., DNA damage repair).

12.2 Base Flipping for Methylation of DNA Bases

12.2.1 Bacterial DNMTs (HhaI, TaqI, Dam, CcrM, and CamA)

Biological methylation is widely engaged in various regulations, and it uses *S*-adenosyl-L-methionine (AdoMet or SAM) as a primary methyl donor. The methyl group of AdoMet is bound to a positively charged sulfur atom predisposed to a nucleophilic attack. During the methylation reaction, AdoMet loses the methyl group and becomes *S*-adenosyl-L-homocysteine (AdoHcy or SAH). A number of different families of MTases use AdoMet as cofactor, targeting diverse substrates ranging from small molecules to large macromolecules such as DNA, RNA, proteins, lipid, and polysaccharides. The atoms subjected to methylation also vary, including carbon (C), nitrogen (N), oxygen (O), sulfur (S), and several metals. AdoMet-dependent DNA MTases were first discovered in bacterial restriction-modification systems (Roberts et al. 2015). The known structures of AdoMet-dependent DNA MTases share a common "MTase fold" characterized by mixed seven-stranded β sheets (6↓ 7↑ 5↓ 4↓ 1↓ 2↓ 3↓) in which strand 7 is inserted between strands 5 and 6 antiparallel to the others (Cheng 1995; Schubert et al. 2003).

M.HhaI was the first DNA MTase to be structurally characterized (Cheng et al. 1993) (Fig. 12.2a). It contains an N-terminal MTase domain and a C-terminal target recognition domain (TRD). M.HhaI is a cytosine C5 MTase that methylates the first cytosine within 5′-GCGC-3′ recognition sequences and prevents R.HhaI restriction activity at this site (Roberts et al. 1976; Horton et al. 2020). Before the structure was available, the proposed mechanism predicted that catalytic Cys81 would make a nucleophilic attack on C6 of cytosine to form a covalent complex, followed by transferring the methyl group from AdoMet to cytosine C5 and releasing the covalent intermediate (Wu and Santi 1985, 1987). In 1994, the crystal structure of M.HhaI-DNA with AdoMet was solved as a trapped covalent enzyme-DNA intermediate using 5-fluorocytosine and directly supported the proposed mechanism, presenting the catalytic cysteine covalently linked to C6 and showing methylated C5 adjacent to AdoHcy (Klimasauskas et al. 1994). Yet, the most striking aspect of the structure was that both the MTase and the TRD of the enzyme work simultaneously to bind DNA and flip the target base into the active-site pocket. The mechanism of DNA base access by base flipping has since been described as the framework for other DNA MTases (Cheng and Roberts 2001).

After the first structure of M.HhaI-DNA was solved, many crystal structures of DNA MTase-DNA complexes have been solved. Besides cytosine C5 methylation, adenine exocyclic N6 methylation is also a critical modification in prokaryotic DNA (Fig. 12.2b–d) and in eukaryotic RNA (Anton and Roberts 2021; Wei and He 2021). The structure of the adenine N6 MTase M. TaqI in complex with DNA and a non-reactive AdoMet analog was solved in 2001 (Goedecke et al. 2001) (Fig. 12.2b). The enzyme methylates adenine within 5′-TCGA-3′ sequence and harbors a similar two-domain organization as M.HhaI, with the conserved N-terminal MTase domain, but a quite distinct C-terminal TRD. The ternary structure is remarkably reminiscent of M.HhaI, involving a flipped adenine in the active site, where the methyl group from the AdoMet analog is positioned near N6 of the flipped adenine. Instead of the catalytic cysteine residue as in M. HhaI, the asparagine 105 side chain and the following proline backbone oxygen make hydrogen bonds with the adenine N6 amine group, potentially modulating the direct transfer of the methyl group from AdoMet. A similar mode of interaction is also seen in the active site of the T4 phage DNA adenine MTase (T4 Dam) that flips adenine in 5′-GATC-3′ sequence, and an aspartate residue (Asp171) contacts adenine N6 (Horton et al. 2005).

Dam is an orphan MTase (Fig. 12.2c), a type of MTase that acts alone without associated cognate restriction endonuclease as part of R-M system (Roberts et al. 2015). Besides Dam, cell

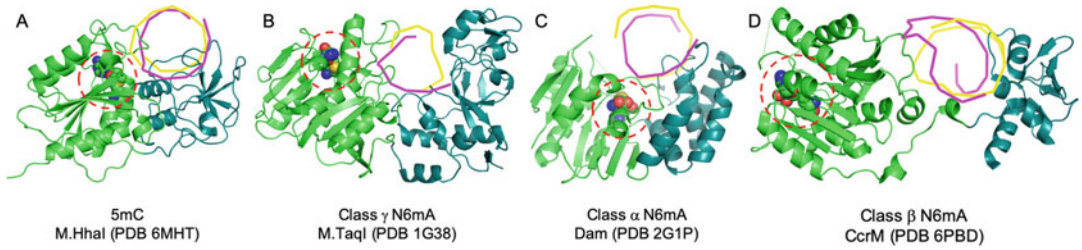

Fig. 12.2 Examples of bacterial DNA MTases. (**a**) M. HhaI, a 5mC MTase, (**b**) M.TaqI, a class γ MTase, (**c**) EcoDam, a class α MTase, and (**d**) CcrM, a class β MTase. The MTase domain (green) binds cofactor (in ball model), and the target recognition domain (TRD) is colored in dark blue. The DNA recognition strand containing the flipped target base is in magenta, and the complementary strand in yellow

cycle-regulated DNA MTase (CcrM) in *Caulobacter crescentus* (Zweiger et al. 1994) (Fig. 12.2d) and newly discovered *Clostridioides difficile* adenine MTase A (CamA) (Oliveira et al. 2020) are also orphan MTases, although they belong to different subgroups depending on the sequential order of conserved motifs (Malone et al. 1995; Woodcock et al. 2020a). The kinetics and structural studies revealed these orphan MTases shared similarities while functioning quite differently. Both Dam (α-group) and CcrM (β-group) are responsible for the post-replication maintenance of daughter strand adenine methylation of symmetrical GATC sequence or near-symmetric GAnTC (n=any base) (Messer and Noyer-Weidner 1988; Stephens et al. 1996). CamA (γ-group), predominantly presented in all *C. difficile* genomes, methylates an asymmetric 6-bp sequence: CAAAAA (underlining indicates the target A) (Oliveira et al. 2020; Zhou et al. 2021). Although the target adenine residue is flipped out during the catalysis for all three enzymes, each individual orphan MTase has its own distinct mechanism. Dam-bound DNA conformation has intact intrahelical paired bases (Horton et al. 2005), CcrM pulls the two DNA strands apart and creates a bubble comprising four unpaid bases for enzyme recognition (Horton et al. 2019), and CamA squeezes out the target adenine by base pair rearrangement (Zhou et al. 2021). In addition, Dam and CamA make base-specific contacts to both DNA strands, whereas CcrM only contacts the bases in the target strand. These unique features allow CcrM to methylate not only double-stranded but also single-stranded DNA (Reich et al. 2018; Konttinen et al. 2020), while both Dam and CamA are inactive on single-stranded DNA.

12.2.2 Mammalian DNMTs (DNMT1, DNMT3A/3L)

Structural features of classic prokaryotic cytosine C5 MTases are extensively shared by mammalian DNA MTases: DNMT1, DNMT3A, and DNMT3B. They are all cytosine C5 MTases containing the MTase domain with the catalytic cysteine and the TRD. DNMT1 is primarily implicated in methylation of the daughter strand during DNA replication to maintain the methylation pattern encoded in the mother strand by preferentially recognizing hemi-methylated DNA in CpG dinucleotide context (Li et al. 1992). On the other hand, DNMT3A and DNMT3B are considered de novo MTases that can methylate CpG sites as well as non-CpG sites (Okano et al. 1999; Ramsahoye et al. 2000; Gowher and Jeltsch 2001). Such differences in substrate specificities are partly due to the involvement of other domains outside the catalytic fragment. For example, the CXXC and BAH1 domains within DNMT1 hinder methylation of unmethylated CpG sites (Song et al. 2011), whereas DNMT3A and DNMT3B do not contain such domains and can readily methylate them.

Moreover, it appears there exist strong influences of flanking sequences on the CpG and non-CpG methylation activity of mammalian

DNA MTases and resulting patterns of methylation. Deep enzymology experiments, which utilize substrates with partly randomized sequences and then analyze the methylation levels by bisulfite conversion coupled with deep next-generation sequencing (NGS) readout, along with structural information, delineate an intricate interplay between flanking sequence, the enzyme-mediated base flipping, and the dynamic landscape of DNA methylation (Jeltsch et al. 2021). Flanking sequences of CpG directly influence catalysis by affecting base-flipping mechanisms which are accompanied by conformational changes in the MTase.

For DNMT1 (Fig. 12.3a), there are strong effects of the ± 2 flanking sequences of a CpG on its activity, with an about 100-fold difference in methylation rates of $N_{-2}N_{-1}CpGN_{+1}N_{+2}$ sites with the best and worst flanking sequences (Adam et al. 2020). Three crystal structures augment this study as a DNMT1 complex with DNA with a highly favorable flanking C_{-1} only showed flipping of the target base, while another complex with a less favorable A_{-1} showed the target base flipping together with rotation of the orphaned G. Finally, a structure with the least favorable G_{-1} displayed a rotation of the target base and invasion of the orphaned G into the −1 flank base pair followed by the formation of a GG non-canonical base pair and flipping of the C normally in the base pair with the G_{-1} residue (Song et al. 2011; Adam et al. 2020). Differences in positioning of the helix that follows the catalytic loop of DNMT1 could be observed with the different nucleotides at the N_{-1} position as well: with A and C at that position, the helix of DNMT1 predominantly adopted a kinked conformation, and with G, the helix mainly adopted a straight conformation moving the active-site loop away from the DNA. Differences of residues at the plus-side flank of the CpG appear to affect the minor groove width which may also influence the equilibrium between the two alternative conformations. Structural transition between these two states of active-site loop-helix is required for DNMT1 activity, suggesting that targeting this transition with small compounds could reduce DNA methylation in cancer cells (Ye et al. 2018). Excitingly, this has been shown to be the case with a new class of reversible DNMT1-selective inhibitors containing a dicyanopyridine moiety (Pappalardi et al. 2021) (Fig. 12.3c).

The flanking sequence preferences of DNMT3A and DNMT3B have also been investigated using the deep enzymology approach (Gao et al. 2020). Methylation levels were averaged for 4096 CpG sites with randomized ± 3 flanks, i.e., $N_{-3}N_{-2}N_{-1}CpGN_{+1}N_{+2}N_{+3}$ sites. The highly methylated sites of DNMT3A showed a preference for C at the +1 site, while DNMT3B preferred G/A at the +1 site and a G at the +2 site DNMT3B while both enzymes prefer T at the −2 site. In addition, the ratio of flanking sequence preferences of DNMT3A and DNMT3B for sites extended a more than 100-fold range, illustrating the noticeable divergence in flanking sequence preferences between the two enzymes.

12.2.3 Implications of DNA Methyltransferase Oligomers (DNMT3A/3L, DNMT3A/3B3, EcoP15I, CcrM, and MettL3-14)

The genomic targeting of DNMT3A and DNMT3B is further regulated by additional factors, including their N-terminal domains and DNMT3L. Besides being a catalytic domain, the MTase domain can participate in protein–protein interactions as exemplified by the DNMT3A MTase domain interacting with a naturally inactive MTase-like domain of DNMT3L, a scaffold protein that binds histone tail H3 to guide DNMT3A activities by forming a tetramer of 3L-3A-3A-3L (Jia et al. 2007; Ooi et al. 2007) (Fig. 12.4a). Moreover, DNMT3L can also form a similar linear assembly with DNMT3B (Lin et al. 2020) (Fig. 12.4b), and 3B-3B and 3B-3L interfaces share some conserved residues with those of the DNMT3A-3L complex.

As a key accessory protein of de novo DNA methylation, Dnmt3L predominantly exists in early embryos and embryonic stem cells and is silenced upon differentiation. Dnmt3b3, a catalytically inactive Dnmt3b isoform, which is

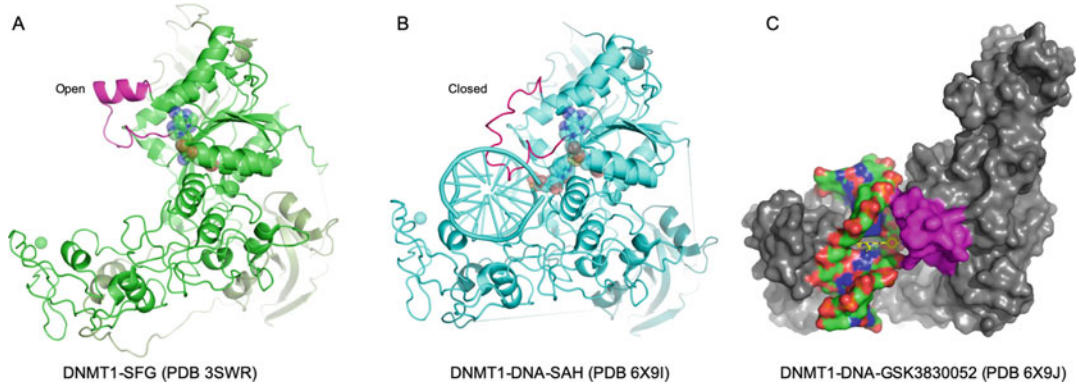

Fig. 12.3 Human DNMT1. (**a**) In the absence of DNA. (**b**) In the presence of DNA. (**c**) In the presence of DNA- and DNMT1-specicifc inhibitor

ubiquitously expressed in differentiated cells and contains the exact 63-residue deletion corresponding to that of Dnmt3L, carries out the regulatory role during late embryonic development and in somatic cells. Dnmt3b3 can positively regulate the catalytic activities of both Dnmt3a2 and Dnmt3b2 either in vitro or in vivo (Duymich et al. 2016; Zeng et al. 2020). The stimulatory effect of Dnmt3b3 is highly dependent on the direct interaction with active form of Dnmt3a or Dnmt3b proteins, and the optimal stimulation is reached at equal molar stoichiometry. The newly reported cryo-electron microscopy structure defined the architecture of a ternary complex of Dnmt3a2-Dnmt3b3 heterotetramer with a nucleosome core particle flanked by linker DNA (Xu et al. 2020) (Fig. 12.4c). This complex contains two monomers each of Dnmt3a2 and Dnmt3b3, forming a tetramer with 3b3-3a2-3a2-3b3, similar to the arrangement of 3L-3a-3a-3L complex. The 3b3-3a2-3a2-3b3 tetramer interacts asymmetrically with the nucleosome with one of the Dnmt3b3 molecules bound to the acidic patch of the nucleosome. The contact point orients both Dnmt3a2 catalytic domains and the second Dnmt3b3 to follow the path of the linker DNA and constrain the arrangement of Dnmt3a target recognition domain with linker DNA. This model suggested a crucial role of Dnmt3b3 in targeting the nucleosome core and driving the de novo methylation on a genome-wide scale.

Interestingly, a multi-subunit prokaryotic DNA N6mA methyltransferase, EcoP15I, contains a DNA MTase dimer in which one monomer is involved in target base flipping and the other in the recognition of DNA base context

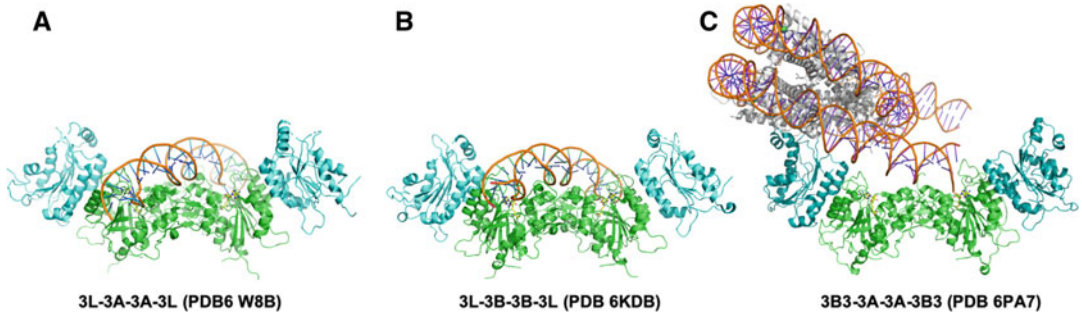

Fig. 12.4 Human DNMT3 family. (**a**) DNMT3A-3L tetramer in complex with DNA. (**b**) DNMT3B-3L tetramer in complex with DNA. (**c**) DNMT3A-3B3 tetramer in complex with nucleosome

(Gupta et al. 2015) (Fig. 12.5a). The "division of labor" might be a conserved feature among class-beta MTases, including M.EcoGII (Murray et al. 2018) and M.EcoP15I (Gupta et al. 2015) from *Escherichia coli*, *Caulobacter crescentus* cell cycle-regulated DNA methyltransferase (CcrM) (Horton et al. 2019) (Fig. 12.5b), the MTA1-MTA9 complex from the ciliate *Oxytricha* (Beh et al. 2019) (Fig. 12.5c), and the mammalian MettL3-MettL14 complex (Fig. 12.5d) (reviewed in Woodcock et al. 2020a). These MTases all generate N6-methyladenine in DNA, with some members having activity on single-stranded DNA as well as RNA. The beta class of MTases has a unique multimeric feature, forming either homo- or hetero-dimers, allowing the enzyme to use division of labor between two subunits in terms of substrate recognition and methylation. Thus, dimerization of two structurally comparable proteins for divergent functionalities may be a mechanism for fine-controlling DNA/RNA modifications.

12.2.4 Plant DNMTs

Plant DNA MTases show similar functionalities to their mammalian counterparts. Met1 is homologous to mammalian DNMT1 and is responsible for the maintenance CpG methylation, whereas domains rearranged methyltransferase 2 (DRM2) is involved in de novo DNA methylation (Law and Jacobsen 2010). DRM2 contains a rearranged MTase domain, such that its N-terminal half is equivalent to the C-terminal half of the conventional MTase fold and vice versa. A structural study of DRM2 family MTase domain has revealed that the rearranged domain still forms a classic MTase structure and functions as a homodimer (Zhong et al. 2014) analogous to the DNMT3A-3L heterodimer. In addition to Met1 and DRM2, plants also have plant-specific DNA MTases, such as CMT2 and CMT3 that are specifically involved in CNG methylation (Stroud et al. 2014; Lindroth et al. 2001; Zemach et al. 2013). The higher diversity of the MTase family within plants compared to the mammalian family suggests that DNA methylation may be more dynamically regulated in plants than in mammals.

12.3 Base Flipping in Oxidative Modifications of Methylated Bases

12.3.1 Eukaryotic TET Enzymes

The 5mC is by far the most widely studied modified base. Yet, if 5mC has been considered "the fifth" base of the genetic alphabet, 5hmC is increasingly being labeled as "the sixth" base and has garnered much attention. The existence of 5hmC in bacteriophage, modified from 2′-deoxycytidine before integration into the viral genome (Warren 1980), was first reported in the early 1950s (Wyatt and Cohen 1952, 1953). In 1993, a novel J base (β-D-glucosyl hydroxymethyluracil) was discovered in trypanosome, in which J-binding proteins (JBP1 and JBP2) are involved in oxidizing the C5 methyl group of thymine during J base synthesis by using αKG and Fe(II) as cofactors to generate 5-hydroxymethyluracil (Gommers-Ampt et al. 1993; Borst and Sabatini 2008). In 2009, mammalian JBP homolog TET enzymes were discovered to oxidize the methyl group of 5mC to generate 5hmC (Tahiliani et al. 2009). Further analysis revealed that TET enzymes could further oxidize 5hmC to 5fC and then to 5caC (Ito et al. 2011; He et al. 2011). Also, TET enzymes have been shown to convert thymine (5-methyluracil) to 5-hydroxymethyluracil by oxidizing the C5 methyl group of thymine (Pfaffeneder et al. 2011; Pais et al. 2015).

Eukaryotic JBP/TET homologs are present across many eukaryotic organisms including amoeboflagellate *Naegleria gruberi (*Iyer et al. 2013; Hashimoto et al. 2014a, 2015a*)*. Crystal structures of *Naegleria gruberi* TET-like (NgTET) (Fig. 12.6a) and human TET2 (hTET2) (Fig. 12.6b) in complex with 5mC-, 5hmC, and 5fC-containing DNA have been characterized (Hashimoto et al. 2014a, 2015a; Hu et al. 2013, 2015). All TET structures show a flipped base positioned in the active-site pocket close to *N*-oxalylglycine (NOG)—an inactive αKG analog—and a divalent metal such as Fe(II) or Mn(II) for stalling the enzyme in the pre-reaction state. Some of the features of the flipped base recognition observed in

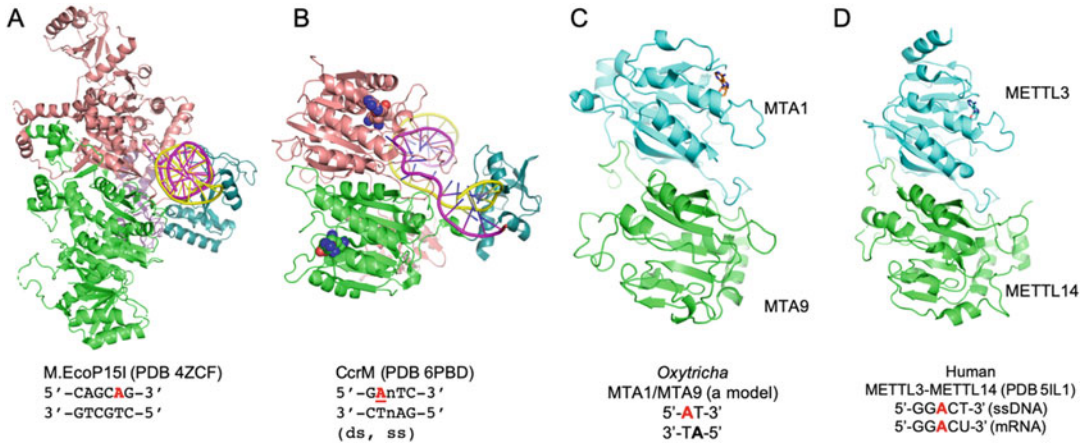

Fig. 12.5 Examples of dimeric class β MTases. (**a**) M.EcoP15I-DNA complex. (**b**) CcrM-DNA complex. (**c**) A model of Oxytricha MTA1-MTA9. (**d**) Human MettL3-MettL14 in the absence of substrate RNA or DNA

DNMT-DNA complex structures (Cheng and Roberts 2001) can also be seen in the structures of TET-DNA complexes. The flipped base in the active site of a TET enzyme in complex with DNA is stabilized by π stacking interactions involving an aromatic residue such as Phe295 in NgTET (Hashimoto et al. 2014a) and Tyr1902 in hTET2 (Hu et al. 2013). Also, polar residues such as Asn147, His297, and Asp234 in NgTET contact O2, N3, and N4, respectively, to guide substrate specificities (Hashimoto et al. 2014a), and the methyl or the hydroxymethyl group is oriented toward NOG and Fe(II)/Mn(II) (Hashimoto et al. 2015a; Hu et al. 2015). Often, active-site pockets for flipped bases not only contain residues for base recognition, but also specifically orient the base for distinct reactions depending on the type of enzymes. Base flipping is therefore a common mechanism applied by different classes of enzymes, such as AdoMet-dependent methyltransferases and αKG- and Fe(II)-dependent dioxygenases to recognize and stabilize the target base for specific reactions.

Like the mammalian DNMTs, flanking sequence preferences of TET1 and TET2 have been analyzed by a deep enzymology approach

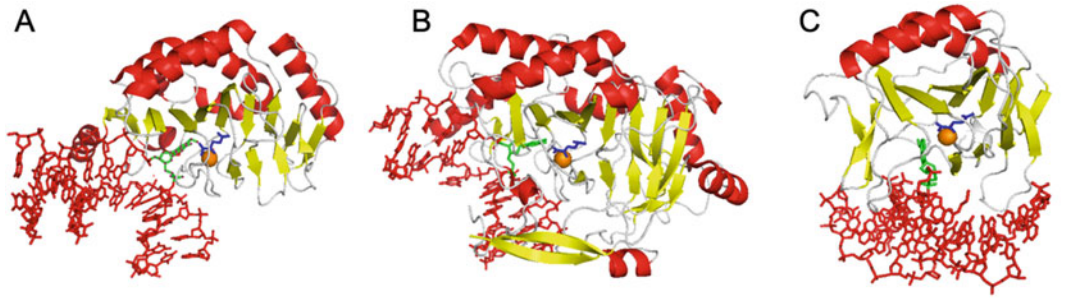

Fig. 12.6 Examples of Fe(II)- and αKG-dependent dioxygenases. (**a**) *Naegleria gruberi* Tet in complex with DNA. (**b**) Human Tet2 in complex with DNA. (**c**) *E. coli* AlkB in complex with DNA. The metal ion is shown as an orange ball in the active site, αKG is colored in blue, and the flipped nucleotide is in green

(Adam et al. 2022) and revealed that TET enzymes show up to 70-fold differences in oxidation rates of either 5mC or 5hmC in CpG target sites embedded in different $N_{-2}N_{-1}CpGN_{+1}N_{+2}$ flanking contexts. For TET1 and TET2 and both substrates, an A is strongly preferred at the −1 site and G is strongly disfavored. At the +1 site, C is generally disfavored, and at the +2 site, TET2 prefers T. TET1 prefers a TA dinucleotide at the −2/−1 site, particularly with 5mC substrates. Yet, a TG at the same place is disfavored especially by TET2. Moreover, a TT dinucleotide is preferred at the +1/+2 site, more so by TET2 and with 5hmC substrates. It was also found that sites with a high genomic 5hmC/5mC ratio are preferentially observed within an A_{-1} and T_{+1} context, while sites with a low genomic 5hmC/5mC ratio are associated with G_{-1} and C_{+1} flanks.

Inspection of structures of TET can help justify these findings. For instance, while only a water-mediated contact is formed with a TA base pair at the +1 site in a TET2 structure (Hu et al. 2015), in another complex with a CG base pair at the +1 site (Hu et al. 2013), a conserved arginine residue makes a direct hydrogen bond to G(+1′) in the minor groove. In turn, the CG(+1) base pair is shifted by 1.5 Å in the direction of the helix axis which may increase the stacking of the target C base in the helix, making base flipping of the target residue more difficult, and explaining the disfavor for C at the +1 flank position. Interestingly, CACGTG appears to be the best sequence which also is the canonical E-box motif, a well-known recognition site for many helix-loop-helix (bHLH) and basic zipper leucine domain (bZIP) transcription factors (Ravichandran et al. 2021). In the case of MAX, a binding partner of the oncogenic transcription factor MYC, MAX exhibits the greatest affinity for a 5caCpG containing E-box, and much reduced affinities for the corresponding 5mC, 5hmC, or 5fC forms (Wang et al. 2017). In the case of TCF4, which binds the E-box element in the context of CG_0-CA_1-CG_2-TG_3 (where the numbers indicate successive dinucleotides), modification of the central CG_2 has very little effect on TCF4 binding, the CA_1 modification has a negative influence on binding, while modification of the flanking CG_0, particularly carboxylation, has a strong positive impact on TCF4 binding to DNA (Yang et al. 2019).

12.3.2 AlkB and Homologs

Similar to TET enzymes, eukaryotic homologs of *E. coli* AlkB (Fig. 12.6c) such as FTO and ALKBH5 are also αKG- and Fe(II)-dependent dioxygenases that can oxidize the methyl group of N6mA within mRNA to yield demethylated adenine (Jia et al. 2011; Zheng et al. 2013; Zhu and Yi 2014). Another ALKBH family member, ALKBH1, appears to be a DNA N6mA demethylase in mammals with preference of N6mA nucleotide within a bubbled or bulged DNA with flanking duplex stems (Zhang et al. 2020). Interestingly, the human METTL3–METTL14 heterocomplex N6mA MTase (Liu et al. 2014a) has been shown by in vitro methylation assays to methylate ssDNA and unpaired DNA regions with flanking duplex DNA, as well as having activity on dsDNA containing cyclopyrimidine dimers, which are the major UV radiation-induced photoproducts (Woodcock et al. 2019; Yu et al. 2021; Qi et al. 2022).

TET-DNA complex structures are remarkably comparable to that of the AlkB-DNA complex, and both TET and AlkB enzymes are Fe(II)- and aKG-dependent dioxygenases using base flipping for reactions (Hashimoto et al. 2014a; Hu et al. 2013; McDonough et al. 2010). Common structural folds include two twisted β-sheets in the core where the active site is formed (Fig. 12.6). However, the two enzymes differ in an important mechanistic aspect. TET enzymes oxidize CH_3 attached to an inert carbon atom (cytosine or thymine C5). The resulting product (5hmC or 5hmU) is very stable and can undergo further oxidations in subsequent rounds of reactions to generate further oxidized products. On the other hand, FTO and ALKBH1/5 likely generate N6-hydroxymethyladenine intermediate in which the oxidized carbon is attached to a reactive nitrogen atom (adenine N6). This intermediate spontaneously releases the hydroxymethyl group as formaldehyde and decomposes to

adenine—the final "demethylated" product (Hashimoto et al. 2015b) (Fig. 12.1b). Therefore, AlkB and its homologs are demethylases, while TET enzymes should not be characterized as demethylases, but would rather be appropriately understood as "writers" that generate additional modifications on 5mC within genomes to alter epigenetic signals.

Several biochemical observations suggest that modified cytosines beyond 5mC may form distinct epigenetic signals. Many 5mCpG readers such as methyl-CpG binding domain (MBD) proteins have shown significantly reduced binding affinity toward 5hmC when compared to 5mC within CpG context (Hashimoto et al. 2012b, 2015b), whereas some proteins may preferentially bind 5hmC (Zhou et al. 2014). DNMT1 has a significantly reduced activity toward hemihydroxymethylated DNA substrate compared to hemi-methylated DNA (Hashimoto et al. 2012b), suggesting that methylation marks altered by TET enzymes can be lost in subsequent DNA replications. In addition, the RNA polymerase II transcription rate can be specifically reduced by 5fC and 5caC (Kellinger et al. 2012; Wang et al. 2015). These findings strongly point to the possibility that modifications beyond 5mC are distinct signals, and future work is needed to elucidate how the modified bases are differently implicated in larger biological contexts. Several 5caC reader proteins have been structurally characterized in the context of specific sequences (Wang et al. 2017; Yang et al. 2019; Hashimoto et al. 2014b).

12.4 Base Flipping in the Recognition of Modified Bases

12.4.1 Eukaryotic SRA Domains

The function of 5mC and N6mA in prokaryotes was classically understood in the context of restriction-modification systems, in which methylated bacterial DNA is protected from restriction digestion (Wilson and Murray 1991). Effects of DNA methylation are fundamentally determined by the way the methyl groups alter various protein–DNA interactions. In eukaryotes, genomic 5mC bases are considered widely involved in various regulatory processes to control gene expression, chromatin states, and genomic stability that are highly relevant in the human disease context (Robertson 2005). Such penetrating biological implications can be partly attributed to the large number of protein–DNA interactions that are potentially affected by DNA methylation in a direct manner. Evidence shows that several transcription factors are prevented from DNA binding when their binding site is methylated (Tate and Bird 1993), whereas several MBD family proteins are specific 5mCpG readers, as previously mentioned (Klose and Bird 2006). Furthermore, a few 5caC readers have been recently characterized (Yang et al. 2020). The interface between modified DNA and its biological effects can be further complicated by the involvement of histone modifications which DNA methylation is profoundly associated with (Cedar and Bergman 2009; Hashimoto et al. 2010).

The initial discovery of 5mC-binding proteins has raised the possibility of other readers involved in modified base recognitions. In 2007, another family of 5mC readers was discovered in plants and was termed SET- and RING-associated (SRA) domain as a part of VIM1 (Woo et al. 2007). A mammalian homolog to VIM1 is UHRF1, which can associate with DNMT1 during post-replicative maintenance of DNA methylation (Bostick et al. 2007; Sharif et al. 2007). In the following year, three crystal structures of the mammalian UHRF1 SRA domain in complex with 5mC-containing DNA were reported (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008). The structures have revealed that SRA recognizes 5mC by base flipping, although it is not a DNA-modifying enzyme such as MTases or dioxygenases. SRA is also structurally distinct from other base flippers and is characterized by a twisted β-sheet fold resembling a half-moon shape (Fig. 12.7a). Remarkably, the 5mC-binding pocket of SRA features familiar modes of base recognitions exemplified by π stacking interactions, recognitions of the N3 and N4 by Asp474 side chain, and a van der Waals'

contact of the C5-methyl group of flipped 5mC by Ser486 Cβ.

Interestingly, the SRA of UHRF2 binds 5hmC with a slightly higher preference compared to 5mC, and the crystal structure of UHRF2 SRA in complex with 5hmC-containing DNA is available (Zhou et al. 2014). In the structure, 5hmC is flipped and stabilized, and the OH moiety of the hydroxymethyl group is contacted by the backbone carbonyl groups of Thr508 and Gly509 in the active-site pocket which is slightly larger in size compared to that of UHRF1 SRA. Therefore, eukaryotic SRA has been characterized as a base-flipping domain that recognizes both 5mC and 5hmC.

In addition, SRA domains have been rediscovered in prokaryotes in families of modification-dependent restriction enzymes that recognize modified bases and introduce a double-stranded break in some distances away. MspJI was among the first such enzymes to be reported, which recognizes hemi-modified 5mC or 5hmC by the N-terminal SRA-like domain and restricts the DNA by the C-terminal endonuclease domain (Cohen-Karni et al. 2011). The crystal structure of MspJI has been solved with substrate DNA, revealing an SRA-like structure in the N-terminal modification recognition domain that flips the target 5mC (Fig. 12.7b) (Horton et al. 2014b). Despite the lack of amino acid sequence conservation between eukaryotic UHRF1/2 SRA and MspJI SRA, all SRA domains feature a twisted β-sheet fold with a half-moon shape.

As more modification-dependent restriction enzymes have been identified, some of them are found with different specificities toward 5mC, 5hmC, and 5ghmC. AbaSI, unlike MspJI, has an N-terminal Vsr-like endonuclease domain and a C-terminal SRA-like domain (Horton et al. 2014a; Borgaro and Zhu 2013). Its SRA domain seems to preferentially recognize 5ghmC and 5hmC compared to 5mC, as the relative rate of cleavage of DNA containing the corresponding modification is 5ghmC:5hmC:5mC=8000:500:1 (Wang et al. 2011). Structural features within SRA domains that fine-tune such specificities await future characterizations.

12.4.2 EcMcrB-N Homologs as 5mC and N6mA Readers

Modification-dependent restriction enzymes also utilize yet another 5mC recognition domain. The N-terminus of *Escherichia coli* McrB (EcMcrB-N) recognizes 5mC next to adenine within 5′-ACCGGT-3′ sequences, and McrC associates with McrB to provide endonuclease activity (Stewart et al. 2000). The crystal structure of EcMcrB-N in complex with 5mC-containing DNA shows a flipped 5mC in the active site (Fig. 12.7c), revealing a novel fold distinct from all other known base flippers (Sukackaite et al. 2012). The active site displays familiar π stacking of the flipped 5mC via aromatic residues and van der Waals' contact of the C5-methyl group via the side chain of Leu68.

The history of the discovery of base flippers suggests a strong possibility of its structural homologs present in a wide spectrum of phyla. EcMcrB-N is poorly conserved among the wide array of McrBC homologs and other domains exist in other homologs for binding modified bases. Indeed, *T. gammatolerans* N-McrB is structurally distinct from the EcMcrB DNA-binding domain, adopting a YTH domain fold commonly found in eukaryotic proteins (Wu et al. 2017). It binds and flips methylated base out of the DNA (Hosford et al. 2020) (Fig. 12.7d). Similarly, it has been observed that, in addition to N6mA in RNA, the N6mA reader domain of human YTHDC1 binds N6mA in either a single-stranded DNA or a single-base gap between two canonical DNA helices (Yu et al. 2021; Woodcock et al. 2020b) (Fig. 12.7e).

12.4.3 5mC and N6mA Readers Use Non-Base-Flipping Recognition

While base flipping seems to be a major mechanism by which a modified DNA base can be recognized, it should be noted that many DNA-binding proteins recognize modified bases in a sequence-dependent manner without

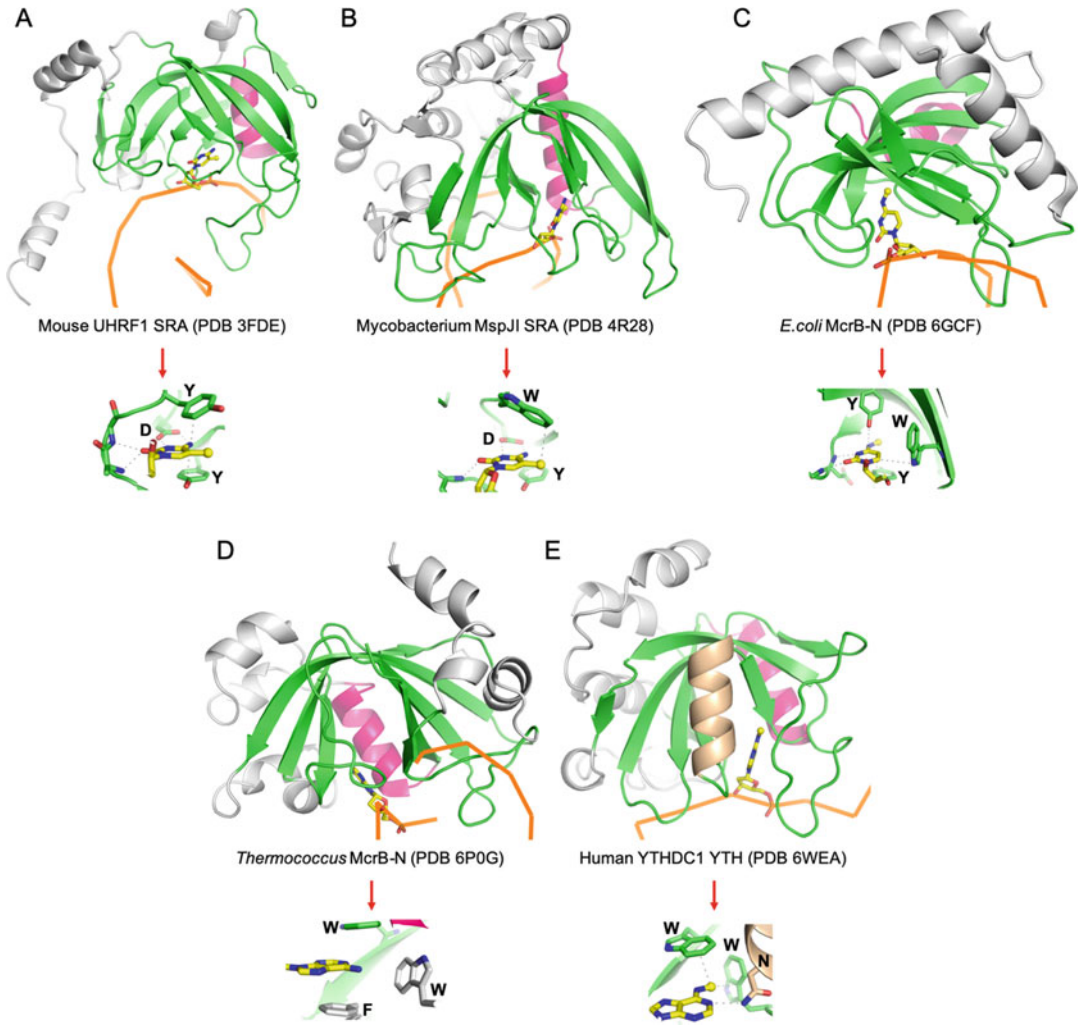

Fig. 12.7 Examples of reader domain proteins using base flipping. (**a**) Mouse UHRF1 SRA with 5mC, (**b**) MspJI SRA with 5mC, (**c**) *E. coli* McrB-N with N4mC, (**d**) *Thermococcus* McrB-N, and (**e**) human YTHDC1 YTH with N6mA. The conserved strands are in green and one conserved helix (colored in red) is behind the arch. The other helices are in gray. DNA strands are in ribbon with the flipped base in stick presentation. The modified bases are bound in a cage formed by 2–3 aromatic residues

involving base flipping. Along with previously mentioned MBD family proteins that recognize 5mC within the simple dinucleotide CpG sequence, mammalian DNA-binding proteins such as Kaiso (Buck-Koehntop et al. 2012), Zfp57 (Liu et al. 2012), Klf4 (Liu et al. 2014b), Egr1 (Hashimoto et al. 2014b; Zandarashvili et al. 2015), and AP-1 (Hong et al. 2017) bind 5mC within specific sequences via a common structural motif (Liu et al. 2013). In addition, transcription factors WT1 (Hashimoto et al. 2014b), MAX (Wang et al. 2017), Tcf3-Ascl1 heterodimer (Golla et al. 2014), and TCF4 (Yang et al. 2019) can specifically bind 5caC within their consensus sequences. In prokaryotes, DpnI harbors a C-terminal WH domain that recognizes the methyl group of N6mA within 5′-GATC-3′ sequence via Trp138 involving van der Waals' interactions (Mierzejewska et al. 2014). Therefore, DNA modifications may

regulate transcription-binding sites in much more dynamic and selective manners than previously understood.

12.5 Base Flipping in Removing Modified and Unmodified Bases

12.5.1 Mammalian Thymine DNA Glycosylase (TDG)

The discovery of TET-mediated modified cytosine bases has provided a fresh insight into a long-sought-after pathway of 5mC demethylation/demodification within mammalian genomes (Zhu 2009). In the base excision repair pathway, DNA glycosylases cleave the glycosidic bond between the ribose and the target base and represent the most structurally diverse family of base-flipping enzymes (Brooks et al. 2013). Initially, it was hypothesized that 5mC (and 5hmC) is removed by mammalian 5mC/5hmC DNA glycosylase activities (Vairapandi and Duker 1993; Cannon et al. 1988; Vairapandi et al. 2000; Vairapandi 2004). However, the glycosylase(s) involved was never identified. After the discovery of TET enzymes, mammalian TDG that generally removes uracil or thymine mismatched to guanine was surprisingly revealed to excise 5fC and 5caC to establish genome-wide DNA demethylation (He et al. 2011; Maiti and Drohat 2011; Hashimoto et al. 2012a). The crystal structure of the human TDG catalytic domain in complex with 5caC-containing DNA was solved (Fig. 12.8a), presenting the flipped base in the active site with an Asn140Ala mutation where the C5-carboxyl moiety of 5caC is specifically recognized by the side chain of Asn157 and the Tyr152 amide backbone (Zhang et al. 2012). In another crystal structure of TDG bound to DNA with a non-cleavable (2′-fluoroarabino) analog of 5-formyldeoxycytidine flipped into its active site (Pidugu et al. 2019), TDG provides a hydrogen bond (2.8 Å) from the Tyr152 backbone N−H to the 5fC formyl oxygen. A nucleophilic water molecule is bound by Asn140 and the backbone oxygen of Thr197, supporting an essential catalytic role for Asn140. The 1.6 Å high-resolution structures of TDG and its N140A mutant bound to DNA with 5caC flipped into the active site suggest that acid-catalyzed 5caC excision is facilitated by two water molecules and contact with Asn191, resulting in a protonated form of 5caC that would be ineffective for C, 5mC, or 5hmC (Pidugu et al. 2019). The discovery of TDG excising 5fC and 5caC has effectively linked the base excision repair pathway to DNA demethylation in mammalian systems.

12.5.2 Plant ROS1

In plants, paradoxically, bone fide 5mC DNA glycosylases were clearly demonstrated and identified in 2002 (Gong et al. 2002), approximately a decade before TET and TDG were implicated in DNA demethylation. In Arabidopsis, four closely related 5mC DNA glycosylases exist: ROS1, DME, DML2, and DML3 (Gehring et al. 2006; Morales-Ruiz et al. 2006; Ortega-Galisteo et al. 2008). They have a catalytic glycosylase domain homologous to *E. coli* endonuclease III (Fig. 12.8b), a helix-hairpin-helix (HhH) fold DNA glycosylase that harbors an iron-sulfur cluster-binding site (Ponferrada-Marin et al. 2009, 2010, 2011, 2012). Thus, plant ROS1 and mammalian TDG have mutually exclusive substrate specificities for 5mC, 5hmC, 5fC, and 5caC: the first two specific for ROS1 and the latter two for TDG (Hong et al. 2014; Hashimoto et al. 2012a) (as shown in Fig. 12.1a). One of the most surprising aspects of plant 5mC DNA glycosylases is that they excise the target base only when both the catalytic glycosylase domain and the C-terminal domain are present (Hong et al. 2014; Mok et al. 2010). The C-terminal domain of ROS1 is conserved only among plant 5mC DNA glycosylases and has been shown to strongly associate with the catalytic domain, suggesting that domain–domain interactions are important for target base recognition and excision (Hong et al. 2014).

While TDG and ROS1 have been clearly implicated in DNA demethylation pathways,

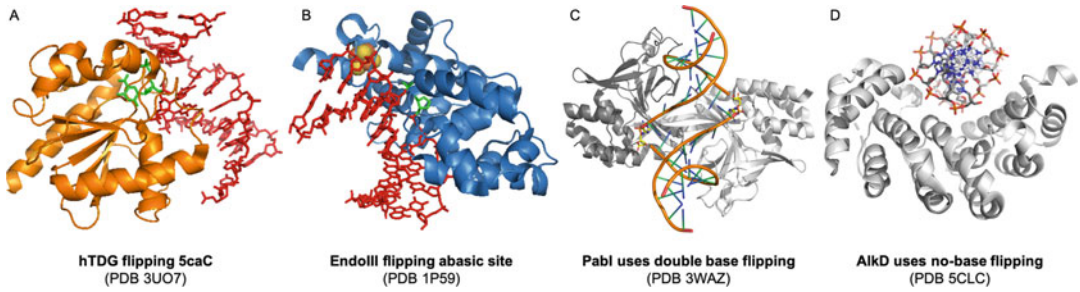

Fig. 12.8 Examples of DNA glycosylases. (**a**) Human TDG flipping 5caC opposite guanine. (**b**) *Geobacillus stearothermophilus* Endonuclease III in complex with DNA. Iron-sulfur cluster is colored in orange and yellow. (**c**) *Pyrococcus abyssi* PabI uses double base flipping. (**d**) *Bacillus cereus* AlkD uses no-base flipping

jury is still out on the possibility of other pathways to DNA demethylation. In addition to the previously mentioned mammalian 5mC DNA glycosylase activities, 5hmC DNA glycosylase activity was observed in a calf thymus extract (Cannon et al. 1988). A proteomic study has revealed that several mammalian DNA glycosylases such as NTH1, OGG1, NEIL1, and NEIL2 bind 5mC- and 5hmC-containing DNA in a modification-specific manner (Spruijt et al. 2013), though they by themselves do not have the glycosylase activity against 5mC or 5hmC (Hong et al. 2014).

The 5mC DNA glycosylase activity by ROS1 is interesting from a standpoint of historical characterization of DNA glycosylases as DNA damage repair enzymes. In a given genome, there can be many types of damaged bases, and their diversity is on par with many classes of DNA glycosylases that are structurally distinct (Brooks et al. 2013). On the other hand, 5mC in plants is not considered a damaged base and exists in substantial amounts in the *Arabidopsis* genome (Zhang et al. 2006). Thus, ROS1 must be regulated and specifically targeted to a certain genomic location to initiate DNA demethylation (Zheng et al. 2008; Qian et al. 2012). In addition to 5mC, ROS1 is comparably active on thymine mismatched to guanine and on some damaged pyrimidines, suggesting that ROS1 can be involved in both DNA demethylation and DNA damage repair (Ponferrada-Marin et al. 2009, 2010). Such dual functionality can be applied to TDG, which not only excises thymine or uracil mismatched to G during the process of DNA mismatch repair, but also excises 5fC and 5caC base-paired with guanine for DNA demodification in mammals.

12.5.3 Archaeon PabI Activity as Adenine DNA Glycosylase

Interestingly, the archaeal *Pyrococcus abyssi* PabI enzyme was initially thought to be a restriction endonuclease but has been re-characterized as a sequence-specific adenine DNA glycosylase (Miyazono et al. 2014, 2020). PabI is comparable to MutY family mismatch repair DNA glycosylases that excise target adenine mismatched to 8-oxoguanine (Fromme et al. 2004). However, PabI is remarkably distinct from MutY, because PabI excises adenine correctly base-paired to thymine in a targeted manner (Fig. 12.8c). It is therefore possible that DNA glycosylases have adapted to function in more processes than DNA damage repair by removing benign bases for various biological regulations.

12.6 Conclusions

First observed in 1994 in the crystal structure of M.HhaI with DNA, base flipping is now understood as a common mode of protein–DNA/RNA interactions adopted by structurally and functionally distinct classes of proteins across various

phyla. Base flipping is the only known mechanism for establishing DNA modifications in a targeted manner via DNA MTases and TET dioxygenases. What used to be considered a eukaryote-specific base-flipping SRA 5mC reader has later been shown to be a widely prevalent domain in prokaryotic systems for recognizing several modified bases including 5mC, 5hmC, and 5ghmC. In addition to SRA, more structurally diverse classes of modified base readers have been discovered in prokaryotes, such as the base-flipping McrB-N 5mC reader and the N6mA-recognizing WH domain of DpnI (using non-base-flipping mechanism). Also, DNA glycosylases are base flippers primarily characterized as DNA repair enzymes, though not all DNA glycosylases flip a base/nucleotide for base excision, as presented in the example of bacterial AlkD (Mullins et al. 2015) (Fig. 12.8d). Today, DNA demodification is considered a bone fide output of the base excision repair pathway through DNA glycosylases, such as mammalian TDG and plant ROS1 whose mechanism of action again involves base flipping. In an era in which DNA modifications are considered critical and increasingly complex epigenetic signals, this simple but elegant structural mechanism for protein–DNA interaction is preserved as a truly ubiquitous framework.

Acknowledgments We thank members of the Cheng laboratory during the last three decades of research on DNA base flipping. The work in the authors' laboratory is currently supported by grant from the National Institutes of Health (R35GM134744) and the Cancer Prevention and Research Institute of Texas (CPRIT RR160029). X.C. is a CPRIT Scholar in Cancer Research.

References

Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE, Bashtrykov P, Song J, Jeltsch A (2020) DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. Nat Commun 11:3723

Adam S, Bracker J, Klingel V, Osteresch B, Radde NE, Brockmeyer J, Bashtrykov P, Jeltsch A (2022) Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. Commun Biol 5:92

Anton BP, Roberts RJ (2021) Beyond restriction modification: epigenomic roles of DNA methylation in prokaryotes. Annu Rev Microbiol 75:129–149

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455:818–821

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455:822–825

Beh LY, Debelouchina GT, Clay DM, Thompson RE, Lindblad KA, Hutton ER, Bracht JR, Sebra RP, Muir TW, Landweber LF (2019) Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. Cell 177:1781–1796

Bestor TH (2000) The DNA methyltransferases of mammals. Hum Mol Genet 9:2395–2402

Borgaro JG, Zhu Z (2013) Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res 41:4198–4206

Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62:235–251

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760–1764

Brooks SC, Adhikary S, Rubinson EH, Eichman BF (2013) Recent advances in the structural mechanisms of DNA glycosylases. Biochim Biophys Acta 1834:247–271

Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, Wright PE (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci U S A 109:15229–15234

Cannon SV, Cummings A, Teebor GW (1988) 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. Biochem Biophys Res Commun 151:1173–1179

Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10:295–304

Cheng X (1995) DNA modification by methyltransferases. Curr Opin Struct Biol 5:4–10

Cheng X, Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 29:3784–3795

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74:299–307

Cohen-Karni D, Xu D, Apone L, Fomenkov A, Sun Z, Davis PJ, Kinney SR, Yamada-Mabuchi M, Xu SY, Davis T et al (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc Natl Acad Sci U S A 108:11040–11045

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7:11453

Fromme JC, Banerjee A, Huang SJ, Verdine GL (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature 427:652–656

Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S, Uryu H, Zhang ZM, Chen D, Yin J et al (2020) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11:3355

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124:495–506

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol 8:121–125

Golla JP, Zhao J, Mann IK, Sayeed SK, Mandal A, Rose RB, Vinson C (2014) Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. Biochem Biophys Res Commun 449:248–255

Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, Kowalak JA, Crain PF, Borst P (1993) beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell 75:1129–1136

Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111:803–814

Gowher H, Jeltsch A (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. J Mol Biol 309:1201–1208

Gupta YK, Chan SH, Xu SY, Aggarwal AK (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. Nat Commun 6:7363

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature 455:826–829

Hashimoto H, Vertino PM, Cheng X (2010) Molecular coupling of DNA methylation and histone methylation. Epigenomics 2:657–669

Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X (2012a) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res 40:10203–10214

Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, Zhang X, Cheng X (2012b) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res 40:4841–4849

Hashimoto H, Pais JE, Zhang X, Saleh L, Fu ZQ, Dai N, Correa IR Jr, Zheng Y, Cheng X (2014a) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature 506:391–395

Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X (2014b) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes Dev 28:2304–2313

Hashimoto H, Pais JE, Dai N, Correa IR Jr, Zhang X, Zheng Y, Cheng X (2015a) Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. Nucleic Acids Res 43:10713–10721

Hashimoto H, Zhang X, Vertino PM, Cheng X (2015b) The mechanisms of generation, recognition, and erasure of DNA 5-methylcytosine and thymine oxidations. J Biol Chem 290:20723–20733

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333:1303–1307

Hong S, Hashimoto H, Kow YW, Zhang X, Cheng X (2014) The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity. J Mol Biol 426:3703–3712

Hong S, Wang D, Horton JR, Zhang X, Speck SH, Blumenthal RM, Cheng X (2017) Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein-Barr virus Zta. Nucleic Acids Res 45:2503–2515

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. Cell 121:349–361

Horton JR, Borgaro JG, Griggs RM, Quimby A, Guan S, Zhang X, Wilson GG, Zheng Y, Zhu Z, Cheng X (2014a) Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res 42:7947–7959

Horton JR, Wang H, Mabuchi MY, Zhang X, Roberts RJ, Zheng Y, Wilson GG, Cheng X (2014b) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res 42:12092–12101

Horton JR, Nugent RL, Li A, Mabuchi MY, Fomenkov A, Cohen-Karni D, Griggs RM, Zhang X, Wilson GG, Zheng Y et al (2014c) Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. Sci Rep 4:4246

Horton JR, Woodcock CB, Opot SB, Reich NO, Zhang X, Cheng X (2019) The cell cycle-regulated DNA adenine

methyltransferase CcrM opens a bubble at its DNA recognition site. Nat Commun 10:4600
Horton JR, Yang J, Zhang X, Petronzio T, Fomenkov A, Wilson GG, Roberts RJ, Cheng X (2020) Structure of HhaI endonuclease with cognate DNA at an atomic resolution of 1.0 A. Nucleic Acids Res 48:1466–1478
Hosford CJ, Bui AQ, Chappie JS (2020) The structure of the Thermococcus gammatolerans McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs. J Biol Chem 295:743–756
Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, Shi YG, Zhu J, Wang P, Xu Y (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell 155:1545–1555
Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, Lou Z, Zhang L, Li W, Gong W et al (2015) Structural insight into substrate preference for TET-mediated oxidation. Nature 527:118–122
Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333:1300–1303
Iyer LM, Zhang D, Maxwell Burroughs A, Aravind L (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res 41:7635–7655
Jang H, Shin H, Eichman BF, Huh JH (2014) Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. Biochem Biophys Res Commun 446: 1067–1072
Jeltsch A, Adam S, Dukatz M, Emperle M, Bashtrykov P (2021) Deep enzymology studies on DNA methyltransferases reveal novel connections between flanking sequences and enzyme activity. J Mol Biol 433:167186
Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449: 248–251
Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG et al (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 7: 885–887
Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol 19:831–833
Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. Cell 76:357–369
Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31:89–97
Konttinen O, Carmody J, Pathuri S, Anderson K, Zhou X, Reich N (2020) Cell cycle regulated DNA methyltransferase: fluorescent tracking of a DNA strand-separation mechanism and identification of the responsible protein motif. Nucleic Acids Res 48: 11589–11601
Kornberg SR, Zimmerman SB, Kornberg A (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J Biol Chem 236:1487–1493
Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324:929–930
Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG (1994) The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 22:1–10
Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11:204–220
Lehman IR, Pratt EA (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J Biol Chem 235:3254–3259
Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926
Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS (2020) Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic Acids Res 48:3949–3961
Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science 292:2077–2080
Liu Y, Toh H, Sasaki H, Zhang X, Cheng X (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. Genes Dev 26: 2374–2379
Liu Y, Zhang X, Blumenthal RM, Cheng X (2013) A common mode of recognition for methylated CpG. Trends Biochem Sci 38:177–183
Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X et al (2014a) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol 10: 93–95
Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, Cheng X (2014b) Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res 42:4859–4867
Lunt MR, Siebke JC, Burton K (1964) Glucosylated nucleotide sequences from T2-bacteriophage deoxyribonucleic acid. Biochem J 92:27–36
Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem 286:35334–35338
Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases,

and suggests a catalytic mechanism for these enzymes. J Mol Biol 253:618–632
McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol 20:659–672
Messer W, Noyer-Weidner M (1988) Timing and targeting: the biological functions of Dam methylation in E. coli. Cell 54:735–737
Mierzejewska K, Siwek W, Czapinska H, Kaus-Drobek M, Radlinska M, Skowronek K, Bujnicki JM, Dadlez M, Bochtler M (2014) Structural basis of the methylation specificity of R.DpnI. Nucleic Acids Res 42:8745–8754
Miyazono K, Furuta Y, Watanabe-Matsui M, Miyakawa T, Ito T, Kobayashi I, Tanokura M (2014) A sequence-specific DNA glycosylase mediates restriction-modification in Pyrococcus abyssi. Nat Commun 5:3178
Miyazono KI, Wang D, Ito T, Tanokura M (2020) Distortion of double-stranded DNA structure by the binding of the restriction DNA glycosylase R.PabI. Nucleic Acids Res 48:5106–5118
Mok YG, Uzawa R, Lee J, Weiner GM, Eichman BF, Fischer RL, Huh JH (2010) Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. Proc Natl Acad Sci U S A 107:19225–19230
Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T (2006) DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci U S A 103:6853–6858
Mullins EA, Shi R, Parsons ZD, Yuen PK, David SS, Igarashi Y, Eichman BF (2015) The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. Nature 527:254–258
Murray IA, Morgan RD, Luyten Y, Fomenkov A, Correa IR Jr, Dai N, Allaw MB, Zhang X, Cheng X, Roberts RJ (2018) The non-specific adenine DNA methyltransferase M.EcoGII. Nucleic Acids Res 46:840–848
Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–257
Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, Sekulovic O, Mead EA, Pak T, Zhu S, Deikus G et al (2020) Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. Nat Microbiol 5:166–180
Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717
Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldan-Arjona T (2008) Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant molecular biology 67:671–681
Pais JE, Dai N, Tamanaha E, Vaisvila R, Fomenkov AI, Bitinaite J, Sun Z, Guan S, Correa IR Jr, Noren CJ et al (2015) Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. Proc Natl Acad Sci U S A 112:4316–4321
Pappalardi MB, Keenan K, Cockerill M, Kellner WA, Stowell A, Sherk C, Wong K, Pathuri S, Briand J, Steidel M et al (2021) Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. Nat Cancer 2:1002–1017
Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeier C, Carell T (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. Angewandte Chemie 50:7008–7012
Pidugu LS, Dai Q, Malik SS, Pozharski E, Drohat AC (2019) Excision of 5-carboxylcytosine by thymine DNA glycosylase. J Am Chem Soc 141:18851–18861
Ponferrada-Marin MI, Roldan-Arjona T, Ariza RR (2009) ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. Nucleic Acids Res 37:4264–4274
Ponferrada-Marin MI, Martinez-Macias MI, Morales-Ruiz T, Roldan-Arjona T, Ariza RR (2010) Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates demethylation in long substrates. J Biol Chem 285:23032–23039
Ponferrada-Marin MI, Parrilla-Doblas JT, Roldan-Arjona-T, Ariza RR (2011) A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. Nucleic Acids Res 39:1473–1484
Ponferrada-Marin MI, Roldan-Arjona T, Ariza RR (2012) Demethylation initiated by ROS1 glycosylase involves random sliding along DNA. Nucleic Acids Res 40:11554–11562
Qi S, Mota J, Chan SH, Villarreal J, Dai N, Arya S, Hromas RA, Rao MK, Correa IR Jr, Gupta YK (2022) RNA binding to human METTL3-METTL14 restricts N(6)-deoxyadenosine methylation of DNA in vitro. eLife:11
Qian W, Miki D, Zhang H, Liu Y, Zhang X, Tang K, Kan Y, La H, Li X, Li S et al (2012) A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. Science 336:1445–1448
Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97:5237–5242
Ravichandran M, Rafalski D, Ortega-Recalde O, Davies CI, Glanfield CR, Kotter A, Misztal K, Wang AH, Wojciechowski M, Rażew M et al (2021) Pronounced sequence specificity of the TET enzyme catalytic domain guides its cellular function. bioRxiv. https://doi.org/10.1101/2021.12.29.474486

Reich NO, Dang E, Kurnik M, Pathuri S, Woodcock CB (2018) The highly specific, cell cycle-regulated methyltransferase from Caulobacter crescentus relies on a novel DNA recognition mechanism. J Biol Chem 293:19038–19046

Roberts RJ, Myers PA, Morrison A, Murray K (1976) A specific endonuclease from Haemophilus haemolyticus. J Mol Biol 103:199–208

Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299

Robertson KD (2005) DNA methylation and human disease. Nat Rev Genet 6:597–610

Schubert HL, Blumenthal RM, Cheng X (2003) Many paths to methyltransfer: a chronicle of convergence. Trends Biochem Sci 28:329–335

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K et al (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450:908–912

Song J, Rechkoblit O, Bestor TH, Patel DJ (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331:1036–1040

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F et al (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152:1146–1159

Stephens C, Reisenauer A, Wright R, Shapiro L (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Natl Acad Sci U S A 93: 1210–1214

Stewart FJ, Panne D, Bickle TA, Raleigh EA (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol 298:611–622

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21:64–72

Sukackaite R, Grazulis S, Tamulaitis G, Siksnys V (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res 40:7552–7562

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935

Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 3:226–231

Vairapandi M (2004) Characterization of DNA demethylation in normal and cancerous cell lines and the regulatory role of cell cycle proteins in human DNA demethylase activity. J Cell Biochem 91:572–583

Vairapandi M, Duker NJ (1993) Enzymic removal of 5-methylcytosine from DNA by a human DNA-glycosylase. Nucleic Acids Res 21:5323–5327

Vairapandi M, Liebermann DA, Hoffman B, Duker NJ (2000) Human DNA-demethylating activity: a glycosylase associated with RNA and PCNA. J Cell Biochem 79:249–260

Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, Roberts RJ, Zhu Z, Zheng Y (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res 39:9294–9305

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, Chong J, Li H, He C, Fu XD et al (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature 523:621–625

Wang D, Hashimoto H, Zhang X, Barwick BG, Lonial S, Boise LH, Vertino PM, Cheng X (2017) MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. Nucleic Acids Res 45:2396–2407

Warren RA (1980) Modified bases in bacteriophage DNAs. Annu Rev Microbiol 34:137–158

Wei J, He C (2021) Chromatin and transcriptional regulation by reversible RNA methylation. Curr Opin Cell Biol 70:109–115

Wilson GG, Murray NE (1991) Restriction and modification systems. Annu Rev Genet 25:585–627

Woo HR, Pontes O, Pikaard CS, Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev 21: 267–277

Woodcock CB, Yu D, Hajian T, Li J, Huang Y, Dai N, Correa IR Jr, Wu T, Vedadi M, Zhang X et al (2019) Human MettL3-MettL14 complex is a sequence-specific DNA adenine methyltransferase active on single-strand and unpaired DNA in vitro. Cell Discov 5:63

Woodcock CB, Horton JR, Zhang X, Blumenthal RM, Cheng X (2020a) Beta class amino methyltransferases from bacteria to humans: evolution and structural consequences. Nucleic Acids Res 48(18): 10034–10044

Woodcock CB, Horton JR, Zhou J, Bedford MT, Blumenthal RM, Zhang X, Cheng X (2020b) Biochemical and structural basis for YTH domain of human YTHDC1 binding to methylated adenine in DNA. Nucleic Acids Res 48:10329–10341

Wu JC, Santi DV (1985) On the mechanism and inhibition of DNA cytosine methyltransferases. Prog Clin Biol Res 198:119–129

Wu JC, Santi DV (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262:4778–4786

Wu B, Li L, Huang Y, Ma J, Min J (2017) Readers, writers and erasers of N(6)-methylated adenosine modification. Curr Opin Struct Biol 47:67–76

Wyatt GR, Cohen SS (1952) A new pyrimidine base from bacteriophage nucleic acids. Nature 170:1072–1073

Wyatt GR, Cohen SS (1953) The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J 55:774–782

Xu TH, Liu M, Zhou XE, Liang G, Zhao G, Xu HE, Melcher K, Jones PA (2020) Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. Nature 586:151–155

Yang J, Horton JR, Li J, Huang Y, Zhang X, Blumenthal RM, Cheng X (2019) Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. Nucleic Acids Res 47:8375–8387

Yang J, Zhang X, Blumenthal RM, Cheng X (2020) Detection of DNA modifications by sequence-specific transcription factors. J Mol Biol 432:1661–1673

Ye F, Kong X, Zhang H, Liu Y, Shao Z, Jin J, Cai Y, Zhang R, Li L, Zhang YW et al (2018) Biochemical studies and molecular dynamic simulations reveal the molecular basis of conformational changes in DNA methyltransferase-1. ACS Chem Biol 13:772–781

Yu D, Horton JR, Yang J, Hajian T, Vedadi M, Sagum CA, Bedford MT, Blumenthal RM, Zhang X, Cheng X (2021) Human MettL3-MettL14 RNA adenine methyltransferase complex is active on double-stranded DNA containing lesions. Nucleic Acids Res 49:11629–11642

Zandarashvili L, White MA, Esadze A, Iwahara J (2015) Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. FEBS Lett 589: 1748–1753

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153:193–205

Zeng Y, Ren R, Kaur G, Hardikar S, Ying Z, Babcock L, Gupta E, Zhang X, Chen T, Cheng X (2020) The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev 34: 1546–1558

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell 126:1189–1201

Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol 8:328–330

Zhang M, Yang S, Nelakanti R, Zhao W, Liu G, Li Z, Liu X, Wu T, Xiao A, Li H (2020) Mammalian ALKBH1 serves as an N(6)-mA demethylase of unpairing DNA. Cell Res 30:197–210

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK (2008) ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature 455:1259–1262

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH et al (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 49:18–29

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell 157:1050–1060

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, Xu RM (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54:879–886

Zhou J, Horton JR, Blumenthal RM, Zhang X, Cheng X (2021) Clostridioides difficile specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. Nat Commun 12:3436

Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet 43:143–166

Zhu C, Yi C (2014) Switching demethylation activities between AlkB family RNA/DNA demethylases through exchange of active-site residues. Angewandte Chemie 53:3659–3662

Zweiger G, Marczynski G, Shapiro L (1994) A Caulobacter DNA methyltransferase that functions only in the predivisional cell. J Mol Biol 235:472–485