Two mammalian DNA glycosylases, methyl-CpG binding domain protein 4 (MBD4) and thymine DNA glycosylase (TDG), are involved in active DNA demethylation via the base excision repair pathway. Both MBD4 and TDG excise the mismatch base from G:X, where X is uracil, thymine, and 5-hydroxymethyluracil (5hmU). In addition, TDG excises 5mC oxidized bases i.e. when X is 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) not 5-hydroxymethylcytosine (5hmC). A MBD4 inactive mutant and substrate crystal structure clearly explains how MBD4 glycosylase discriminates substrates: 5mC are not able to be directly excised, but a deamination process from 5mC to thymine is required. On the other hand, TDG is much more complicated; in this instance, crystal structures show that TDG recognizes G:X mismatch DNA containing DNA and G:5caC containing DNA from the minor groove of DNA, which suggested that TDG might recognize 5mC oxidized product 5caC like mismatch DNA. In mutation studies, a N157D mutation results in a more 5caC specific glycosylase, and a N191A mutation inhibits 5caC activity while that when X=5fC or T remains. Here I revisit the recent MBD4 glycosylase domain co-crystal structures with DNA, as well as TDG glycosylase domain co-crystal structures with DNA in conjunction with its mutation studies.

**Key words:** DNA glycosylase, methyl-CpG binding domain 4, Thymine DNA Glycosylase, active demethylation

---

Cytosine modifications

In the mammalian genome, cytosine residues can be dynamically modified at the C5 position, and can exist as either unmethylated cytosine (C), methyl-cytosine (5mC), hydroxymethylcytosine (5hmC), formylcytosine (5fC), or carboxyl-cytosine (5caC) (Fig. 1). Three DNA methyltransferases (DNMT1, DNMT3a, and DNMT3b) methylate cytosine (C) to 5mC, and three ten-eleven translocation (TET1, TET2, and TET3) dioxygenases oxidize 5mC to 5hmC, 5fC and 5caC. DNA methylation/unmethylation plays an important role in gene silencing, and DNA methylation levels vary in different tissues, at different ages, and in different genomic locations. So far, there is no reported enzyme which directly removes a methyl-group from 5-methylated cytosine (5mC), but N-methylated nucleic acids (3-methylated cytosine, 3-methylated thymine and 1-methylated adenine) can be removed. Oxidization of N-methylated nucleic acids involves in a 2-step demethylation processes. N-methylated nucleic acids are oxidized by AlkB proteins, and their hydroxymethyl-groups are spontaneously released from base with the byproduct, formaldehyde. However, the oxidative products of C-methylated nucleic acids are stable bases. 5mC oxidized products, 5hmC, and 5fC, act as substrates of TET proteins, which can be converted to 5caC as the final product in vitro.

There are two DNA demethylation pathways, one is a replication-dependent “passive demethylation” pathway, and the other is a replication-independent “active demethylation” pathway. In mammals, two “active” demethylation pathways via base excision repair pathway have been proposed. One is a deamination mediated demethylation pathway. DNA deaminases, like AID (activation-induced cytidine deaminase) and APOBEC (apolipoprotein B mRNA...
**MBD4 glycosylase structure and substrates specificity**

Mouse methyl-CpG binding domain 4 (MBD4) protein encodes 554 amino acids, and has two domains; the methyl-CpG DNA binding domain (MBD) domain at the amino-terminus and glycosylase domain at carboxy-terminus. The N-terminal MBD domain recognizes a G:T mismatch as well as fully methylated CpG dinucleotide. The C-terminal glycosylase domain excises uracil, thymine (also known as 5-methyluracil) and 5-hydroxymethyluracil (5hmU) paired with guanine, which forms deamination linked mismatches. MBD4 catalytic inactive mutant (Asp534Asn for mouse, Asp560Ala for human) and substrate DNA crystal structures clarifies how MBD4 recognizes substrates.

In the mouse MBD4 inactive protein (D534N) and G:T mismatch pair substrate complex structure (Fig. 2A; PDB: 4EVY), the G:T mismatch containing 11-bp DNA duplex was bent ~65°, between the G:T mismatch and a C:G base pair by the Leu482 side chain (Fig. 2C). Concurrently, the thymine nucleotide in the mismatch flips out, and an arginine finger (Arg442) penetrates into the space left by the flipped thymine (Fig. 2A). The intrahelical orphaned guanine is stabilized by hydrogen bonds with the main chain carbonyl oxygen atoms of Arg442 and Leu480 (Fig. 2D). Asn441 and Leu485 sit in the minor groove of DNA, and interact weakly with the neighboring G:C pair. No interaction to the neighboring G:C pair in the major groove of DNA is observed and neighboring cytosine modification does not affect glycosylase activities. The polar groups of the flipped thymine base along the Watson–Crick edge are all hydrogen bonded with Thr514 side chain, Gln423 side chain, and Val422 main chain (Fig. 2E). The O2 oxygen atom of thymine base accepts hydrogen bonds from the side chains of Tyr514 (hydroxyl group) and Gln423 (amino group), the N3 nitrogen atom of thymine donates a hydrogen bond to the carbonyl oxygen of Gln423 side chain, and Val422 main chain (Fig. 2E). The O2 oxygen atom of thymine base accepts hydrogen bonds from the main-chain amide nitrogen of Val422 (Fig. 2E). Because thymine and cytosine bases have different polar edges where MBD4 recognizes a flipped thymine base (Fig. 2E, 2F), MBD4 can discriminate against all cytosine derivatives, including 5mC, 5hmC, and 5caC. For this reason, the MBD4-mediated demethylation pathway requires a deamination process. C5 modifications of uracil are not recognized, while uracil and derivatives including 5-methyl uracil (thymine), and 5-hydroxymethyl uracil (5hmU) paired with guanine are substrates for the MBD4 glycosylase domain in vitro.

**TDG glycosylase**

Human thymine DNA glycosylase (TDG) encodes 410 amino acids, and belongs to the uracil DNA glycosylase (UDG) superfamily. TDG and *E. coli* MUG glycosylases are highly similar (31% identity, and 43% similarity) and both can excise 5caC paired with guanine. TDG excises uracil, thymine, 5-hydroxymethyl uracil, in addition to 5-formyl cytosine, 5-carboxyl cytosine and etheno-cytosine paired with guanine. An interesting substrate discrimination of TDG is that it excises 5fC, and 5caC but not 5hmC; yet it is still unknown how TDG discriminates 5-substituents (Fig. 1). Here we revisit the two “active” 5mC demethylation related glycosylases, MBD4 and TDG, their glycosylase domain DNA co-crystal structural studies and their mutation studies.

---

**Figure 1** Known and putative DNA demethylation pathways. DNMT3/DNMT3L complex produces new DNA methylation patterns by methylating cytosine (C) to 5-methylated cytosine (5mC), while the DNMT1/UHRF1 complex maintains previous cytosine methylation after the DNA replication. Ten-eleven translocation (TET) proteins consecutively oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC and 5caC can be removed by thymine DNA glycosylase (TDG) and replaced by cytosine via base excision repair (BER). AID (activation-induced cytidine deaminase)/APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide) enzymes deaminate 5-methylated cytosine (5mC) to thymine (T), and forms a G:T mismatch. Thymine is excised by TDG and MBD4, and replaced by cytosine via base excision repair (BER).
helix from the minor groove, occupying the space left by the flipped-out modified nucleotide (Fig. 3A). The intrahelical orphaned guanine is stabilized by hydrogen bonds with the main chain carbonyl oxygen atoms of Ala274 and Pro280 (Fig. 3D). Therefore, these structural studies suggest TDG potentially recognizes 5caC, 5fC as a mismatch like pair from minor groove of DNA. It was hypothesized that G:5fC/G:5caC pairs form a “wobble base pair” by amino/imino tautomeric forms33,34 (Fig. 3E), however, in which the amino tautomers of 5fC and 5caC are much more stable, the crystal structure of a 5fC containing 12-bp DNA duplex (PDB: 1VE8) shows that 5fC:G forms normal base pairs35–37. The crystal structure of the non-substrate 5hmC:G base pair containing DNA complexed with TDG would clarify how 5fC:G/5caC:G bases are recognized by TDG.

The carboxyl group of the 5caC base interacts with the minor groove of DNA, and not from major groove of DNA (Fig. 3A)38–40. Reported mutation and crystal studies, however, have yet to explain why TDG lacks 5hmC activity, but only excises 5fC and 5caC among 5mC oxidized bases. TDG distorts various substrate/product DNA ~18° (Fig. 3B), and there is no side chain interaction in a major groove of DNA where the modifications at the C5 position are positioned (Fig. 3C). However, the Pro198-Gly199-Ser200 (P-G-S) containing loop is located in a major groove of DNA and potentially can form interactions with C5 modifications such as in the mechanism proposed for MutM and human UNG31,32. This TDG loop can superimpose well with the corresponding loop of human UNG, but Ala-mutants in the P-G-S loop still discriminate substrate specificities. These results suggest that P-G-S loop is unlikely to play a strong role in discriminating between different 5-substituents26.

So far, in TDG and DNA complex structures, substrate/product DNA are recognized in the minor groove of DNA by Arg275. The side chain of Arg275 penetrates into the DNA groove of DNA, and not from major groove of DNA (Fig. 3A)38–40. Reported mutation and crystal studies, however, have yet to explain why TDG lacks 5hmC activity, but only excises 5fC and 5caC among 5mC oxidized bases. TDG distorts various substrate/product DNA ~18° (Fig. 3B), and there is no side chain interaction in a major groove of DNA where the modifications at the C5 position are positioned (Fig. 3C). However, the Pro198-Gly199-Ser200 (P-G-S) containing loop is located in a major groove of DNA and potentially can form interactions with C5 modifications such as in the mechanism proposed for MutM and human UNG31,32. This TDG loop can superimpose well with the corresponding loop of human UNG, but Ala-mutants in the P-G-S loop still discriminate substrate specificities. These results suggest that P-G-S loop is unlikely to play a strong role in discriminating between different 5-substituents26.

So far, in TDG and DNA complex structures, substrate/product DNA are recognized in the minor groove of DNA by Arg275. The side chain of Arg275 penetrates into the DNA groove of DNA, and not from major groove of DNA (Fig. 3A)38–40. Reported mutation and crystal studies, however, have yet to explain why TDG lacks 5hmC activity, but only excises 5fC and 5caC among 5mC oxidized bases. TDG distorts various substrate/product DNA ~18° (Fig. 3B), and there is no side chain interaction in a major groove of DNA where the modifications at the C5 position are positioned (Fig. 3C). However, the Pro198-Gly199-Ser200 (P-G-S) containing loop is located in a major groove of DNA and potentially can form interactions with C5 modifications such as in the mechanism proposed for MutM and human UNG31,32. This TDG loop can superimpose well with the corresponding loop of human UNG, but Ala-mutants in the P-G-S loop still discriminate substrate specificities. These results suggest that P-G-S loop is unlikely to play a strong role in discriminating between different 5-substituents26.

So far, in TDG and DNA complex structures, substrate/product DNA are recognized in the minor groove of DNA by Arg275. The side chain of Arg275 penetrates into the DNA groove of DNA, and not from major groove of DNA (Fig. 3A)38–40. Reported mutation and crystal studies, however, have yet to explain why TDG lacks 5hmC activity, but only excises 5fC and 5caC among 5mC oxidized bases. TDG distorts various substrate/product DNA ~18° (Fig. 3B), and there is no side chain interaction in a major groove of DNA where the modifications at the C5 position are positioned (Fig. 3C). However, the Pro198-Gly199-Ser200 (P-G-S) containing loop is located in a major groove of DNA and potentially can form interactions with C5 modifications such as in the mechanism proposed for MutM and human UNG31,32. This TDG loop can superimpose well with the corresponding loop of human UNG, but Ala-mutants in the P-G-S loop still discriminate substrate specificities. These results suggest that P-G-S loop is unlikely to play a strong role in discriminating between different 5-substituents26.

So far, in TDG and DNA complex structures, substrate/product DNA are recognized in the minor groove of DNA by Arg275. The side chain of Arg275 penetrates into the DNA groove of DNA, and not from major groove of DNA (Fig. 3A)38–40. Reported mutation and crystal studies, however, have yet to explain why TDG lacks 5hmC activity, but only excises 5fC and 5caC among 5mC oxidized bases. TDG distorts various substrate/product DNA ~18° (Fig. 3B), and there is no side chain interaction in a major groove of DNA where the modifications at the C5 position are positioned (Fig. 3C). However, the Pro198-Gly199-Ser200 (P-G-S) containing loop is located in a major groove of DNA and potentially can form interactions with C5 modifications such as in the mechanism proposed for MutM and human UNG31,32. This TDG loop can superimpose well with the corresponding loop of human UNG, but Ala-mutants in the P-G-S loop still discriminate substrate specificities. These results suggest that P-G-S loop is unlikely to play a strong role in discriminating between different 5-substituents26.

So far, in TDG and DNA complex structures, substrate/product DNA are recognized in the minor groove of DNA by Arg275. The side chain of Arg275 penetrates into the DNA groove of DNA, and not from major groove of DNA (Fig. 3A)38–40. Reported mutation and crystal studies, however, have yet to explain why TDG lacks 5hmC activity, but only excises 5fC and 5caC among 5mC oxidized bases. TDG distorts various substrate/product DNA ~18° (Fig. 3B), and there is no side chain interaction in a major groove of DNA where the modifications at the C5 position are positioned (Fig. 3C). However, the Pro198-Gly199-Ser200 (P-G-S) containing loop is located in a major groove of DNA and potentially can form interactions with C5 modifications such as in the mechanism proposed for MutM and human UNG31,32. This TDG loop can superimpose well with the corresponding loop of human UNG, but Ala-mutants in the P-G-S loop still discriminate substrate specificities. These results suggest that P-G-S loop is unlikely to play a strong role in discriminating between different 5-substituents26.

So far, in TDG and DNA complex structures, substrate/product DNA are recognized in the minor groove of DNA by Arg275. The side chain of Arg275 penetrates into the DNA
specificity but a N230D/N157D double mutant completely inhibits the 5caC activity but not U activity. A N191A single mutation specifically inhibits the 5caC activity while 5fC activity remains. 5fC activity is pH independent, while 5caC activity is pH dependent. These observations indicate that the catalytic mechanism of 5fC and 5caC are different. There are four asparagines (Asn140, Asn157, Asn191, and Asn230) in an active site and each play an individual role in substrate dependent catalysis. Asn140, and Asn157 are conserved between human TDG and E. coli Mug (PDB: 1MWI), but the other two are not. Asn191 of TDG position is Lys68 of eMug, and Asn230 of TDG position is Leu107 of eMug. However, eMug does have another Asn140 in its active site, equivalent to Ser271 of TDG. The catalytic mechanisms for 5caC substrates are also different between human TDG and E. coli Mug. Substrate preferences between TDG and eMug are

Figure 3 Structure of TDG in complex with DNA. (A) Structure of TDG wild-type in product complex with G:5caC base pair (PDB:4JGC). The surface charge at neutral pH is displayed as blue for positive (50 k_BT), red for negative (–50 k_BT), and white for neutral, where k_B is the Boltzmann’s constant and T is the temperature. (B) Entire 28 bp DNA structure in product complex with TDG (C) The Arg275-containing intercalation loop (cyan) and the P-G-S loop (magenta) approach the modified DNA strand from opposite directions. (D) The three hydrogen bonds formed with the intrahelical orphaned guanine (E) Proposed for amino-imino tautomerization. 5fC and 5caC exhibit an intramolecular hydrogen bond that could shift the amino/imino equilibrium toward the imino tautomeric form which would then base pair with guanine in a mismatch-like wobble pattern. (F) The interactions between 5caC base and TDG in its active site (G) Superimposition of N140A in complex with a 5caC-containing 28-bp DNA crystallized at pH 4.6 (colored in green and gray) suggests a base rotation of approximately 120° around the glycoside bond. Four asparagines (Asn140Ala, Asn191, Asn 157, Asn230) are also shown.
different as well. For instance, the eMug activities on 5hmU and thymine are extremely poor, even compared to the 5caC activities of human TDG and E. coli Mug, and the human TDG activity on 5caC is pH dependent while the E. coli Mug activity on 5caC is pH independent\(^\text{16-20}\). Therefore, it is still not clear how TDG discriminates substrates and how catalysis proceeds with the various substrates of TDG.

**Perspective**

DNA methylation and demethylation are dynamic processes and required for embryonic development\(^2\). 5mC glycosylases which directly remove a 5mC base are found in plants\(^3\), but so far, no 5mC demethylase as well as no 5mC DNA glycosylase, has been identified in mammals. Oxidation of 5mC to 5fC/5caC by TET proteins or deamination of 5mC to T by the deaminases, AID or APOBEC, are the initial step for active DNA demethylation. Afterward, TDG and/or MBD4 glycosylases remove oxidative 5mC and/or deaminated 5mC. Structural and biochemical studies have showed that oxidation or deamination of 5mC is required for removal of the 5mC base, following by the base excision repair pathway which is required to replace the resultant abasic site with cytosine. Therefore, targeting demethylated 5mC locations seem to be controlled by TET proteins, and 5hmC to 5fC oxidation might be a key conversion for 5mC locations seem to be controlled by TET proteins, and 5hmC to 5fC oxidation might be a key conversion for 5mC to 5fC/5caC by TET proteins or deamination of 5mC to T by the deaminases, AID or APOBEC, are the initial step for active DNA demethylation. Afterward, TDG and/or MBD4 glycosylases remove oxidative 5mC and/or deaminated 5mC. Structural and biochemical studies have showed that oxidation or deamination of 5mC is required for removal of the 5mC base, following by the base excision repair pathway which is required to replace the resultant abasic site with cytosine. Therefore, targeting demethylated 5mC locations seem to be controlled by TET proteins, and 5hmC to 5fC oxidation might be a key conversion for DNA demethylation. 5fC is potentially a mutagenic in mammalian cells\(^3\), DNA polymerase, and RNA polymerase also discriminate between C/5mC/5hmC and 5fC/5caC. DNA polymerases misincorporate a small amount of adenine using a 5fC template in vitro\(^8\). RNA polymerase II polymerization rates for GTP incorporation using 5fC/5caC containing DNA are significantly slower, and the discrimination of GTP over ATP is reduced by a factor of 30 for a 5fC template in comparison with a C template in vitro\(^4\). 5fC/5caC may introduce transition mutations during DNA replication and mRNA transcription. Therefore, 5fC, and 5caC might be recognized as damaged bases, and be distinct from 5hmC among 5mC oxidized products for DNA polymerase, RNA polymerase II and TDG glycosylases. TDG could recognize G:5fC, and G:5caC as a damaged base pair from the minor groove of DNA as same manner as G:X mismatch DNA. It appears that 5hmC might be not a simple DNA demethylation intermediate product, but potentially has its own biological function(s) and is maintained by DNMT3A, DNMT3B and TET proteins after DNA replication.

**References**

1. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
2. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68 (2014).
3. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. & Rao, A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
4. Ito, S., D’Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C. & Zhang, Y. Role of Tet proteins in 5mC to 5fC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
5. Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X. & Liu, J. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
6. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegreni, M., Jacobsen, S. E. & Reik, W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
7. Fu, Y. & He, C. Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.* **16**, 516–524 (2012).
8. Yang, C. G., Yi, C., Duguid, E. M., Sullivan, C. T., Jian, X., Rice, P. A. & He, C. Crystal structures of DNA:RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961–965 (2008).
9. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C. G., Cui, Q. & He, C. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468**, 330–333 (2010).
10. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21**, 1670–1676 (2011).
11. Hashimoto, H., Pias, J. E., Zhang, X., Saleh, L., Fu, Z. Q., Dai, N., Correa, I. R., Jr., Zheng, Y. & Cheng, X. Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395 (2014).
12. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P. & Xu, Y. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545–1555 (2013).
13. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
14. Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., Zhang, Y. & Kohli, R. M. AID/APOBEC deaminases favor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
15. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279**, 52353–52360 (2004).
16. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. & Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxycytosine. *Science* **333**, 1300–1303 (2011).
17. He, Y. F., Li, B. Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C. & Xu, G. L. Tet-mediated formation of 5-carboxycytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
18. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxycytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
19. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. & Bird, A. The thymine glycosylase MBD4 can bind to the product of...
25. Hashimoto, H., Liu, Y., Upadhyay, A.  
23. Moréra, S., Grin, I., Vigouroux, A., Couvé, S., Henriot, V.,  
22. Hashimoto, H., Zhang, X. & Cheng, X. Excision of thymine  
30. Maiti, A., Noon, M.  
68 BIOPHYSICS Vol. 10  
24. Wu, P., Qiu, C., Sohail, A., Zhang, X., Bhagwat, A.  
31. Qi, Y., Spong, M.  
26. Hashimoto, H., Hong, S., Bhagwat, A.  
29. Maiti, A., Morgan, M.  
27. Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G. L., Luo,  
S., MacKerell, A.  
766 (2009).  
8  
B., Bianchet, M.  
18, 6538–6547 (1998).  
17, 762–804 (2012).  
37. Maiti, A., Michelson, A.  
35. Irrera, S. & Portalone, G. First X-ray diffraction and quantum  
chemical study of proton-acceptor and proton-donor forms of 5-carboxylcytosine, the last-discovered nucleobase. J. Mol. Struct. 1050, 140–150 (2013).  
43. Munzel, M., Lischke, U., Stathis, D., Pfaffeneder, T., Gnerlich, F. A., Deiml, C. A., Koch, S. C., Karaghiosoff, K. & Carell, T. Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. Chemistry 17, 13782–13788 (2011).  
44. Kellinger, M. W., Song, C. X., Chong, J., Lu, X. Y., He, C. & Wang, D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat. Struct. Mol. Biol. 19, 831–833 (2012).  
