Molecular basis for the faithful replication of 5-methylcytosine and its oxidized forms by DNA Polymerase β

Michael J. Howard¹, K. Grace Foley¹, David D. Shock, Vinod K. Batra, Samuel H. Wilson²

Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709

1. *Running Title: DNA replication of epigenetically modified cytosine

²To whom correspondence should be addressed: Samuel H. Wilson, Genome Integrity and Structural Biology Laboratory, National Institutes of Health, National Institute of Environmental Health Sciences, 111 T. W. Alexander Dr., P.O. Box 12233 MD F1-12, Research Triangle Park, NC 27709. Tel.: 984-287-3451; E-mail: wilson5@niehs.nih.gov

Keywords: DNA Polymerase, Base excision repair, Epigenetics, 5-methylcytosine, DNA repair

ABSTRACT

DNA methylation is an epigenetic mark that regulates gene expression in mammals. One method of methylation removal is through TET-catalyzed oxidation and the base excision repair pathway. The iterative oxidation of 5-methylcytosine (5mC) catalyzed by TET enzymes produces three oxidized forms of cytosine: 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC). The effect these modifications have on the efficiency and fidelity of the base excision repair pathway during the repair of opposing base damage, and in particular DNA polymerization, remains to be elucidated. Using kinetic assays, we show that the catalytic efficiency for the incorporation of dGTP catalyzed by human DNA Polymerase β is not affected when 5mC, 5hmC, 5fC are in the DNA template. In contrast, the catalytic efficiency of dGTP insertion decreases ~20-fold when 5caC is in the templating position, as compared to unmodified cytosine. However, DNA polymerase fidelity is unaltered when these modifications are in the templating position. Structural analysis reveals that the methyl, hydroxymethyl, and formyl modifications are easily accommodated within the polymerase active site. However, to accommodate the carboxy modification, the phosphate backbone on the templating nucleotide shifts ~2.5 Å to avoid a potential steric/repulsive clash. This altered conformation is stabilized by lysine 280 which makes a direct interaction with the carboxy modification and the phosphate backbone of the templating strand. This work provides the molecular basis for the accommodation of epigenetic base modifications in a polymerase active site and suggests that these modifications are not mutagenically copied during BER.

Epigenetic modifications have a significant role in maintaining cellular identity by regulating transcript levels of the human genome (1). Methylated cytosine (5mC) is one such epigenetic modification, which consists of a methyl group addition to the fifth carbon of the cytosine base. DNA methyltransferases (DNMT1-3) catalyze this methylation of cytosine, usually within the context of CpG dinucleotide clusters or islands (2). In general, 5mC in CpG islands down-regulates gene expression by weakening transcription factor binding or by promoting repressor binding (3,4).

Cytosine methylation can be erased to allow for differential gene expression. This necessitates demethylation pathways that can either remove the methyl group from the cytosine base or remove the base and replace it with normal cytosine (active demethylation pathway). At this point, an enzyme capable of catalyzing the direct removal of a C5 methyl group from cytosine has not been
identified. However, in vivo isotopic labeling studies have suggested the existence of such an enzyme (5). Nonetheless, the most well described demethylation pathway involves the removal and replacement of 5mC by action of Ten-eleven translocation (TET) enzymes and base excision repair (BER) (6). Humans encode three TET enzymes which are non-heme Fe(II) alpha-ketoglutarate dependent dioxygenases that have multiple domains (7,8). These enzymes oxidize the methyl group on 5C iteratively, forming three intermediates: 5-hydroxymethyl cytosine (5hmC); 5-formyl cytosine (5fC); and 5-carboxy cytosine (5caC) (9). The base excision repair enzyme Thymine DNA glycosylase (TDG) catalyzes removal of 5fC and 5caC from DNA generating an abasic site product (10,11). Apurinic/apyrimidinic endonuclease1 (APE1) hydrolyzes the phosphodiester backbone of the abasic site, generating a 5’ deoxyribose phosphate (dRP) flap in a 1-nucleotide gap (12). DNA polymerase (pol) β removes the dRP group using its lyase activity and catalyzes gap-filing DNA synthesis (13,14). Finally, DNA ligase III/XRCC1 complex or DNA ligase I catalyzes ligation of the DNA backbone (15).

Although a role of 5mC and 5hmC in epigenetic gene regulation has been established, such a role for 5fC and 5caC has yet to be uncovered (16). However, 5fC and 5caC was detected in genomes from neuronal tissue and stem cells, and proteins specific for binding to 5fC and 5caC in DNA have been reported (17-20). Information on how the cytosine modifications might influence DNA repair pathways, such as BER, is relatively limited (21,22). Furthermore, the potential for these modified cytosine bases to alter the activity and fidelity of DNA polymerases during repair and replication has not been quantitatively assessed and conflicting reports have been published.

In general, these modifications in the template base position can have two effects on DNA polymerization. They may reduce the catalytic efficiency of dGTP insertion; and, these modifications may alter dNTP selectivity, leading to a decrease in fidelity. It has been hypothesized that the formyl and carboxy modifications promote the imino tautomeric form of cytosine (23). The imino form of cytosine is potentially mutagenic as it can form a Watson-crick base pair with adenine (24). However, analytical experiments have revealed that these modifications do not promote the tautomeric form of cytosine and call into question the potential of these bases to encode for adenine insertion into DNA (25,26).

Previous replication studies in Escherichia coli using a shuttle vector assay suggested that 5mC and its oxidized forms could induce C to T transition mutations at a low frequency (~0.2-1.1%) (27). Another in vivo study, using cos-7 cells, found 5fC to induce C to T mutations at a frequency of 0.03-0.28 % (28). Qualitative in vitro kinetic studies using DNA polymerases η, κ, and Klenow exo- and the DNA template base 5fC revealed a frequency of 1% for the insertion of dATP vs. dGTP (24). Another qualitative kinetic study revealed that 5mC and its oxidized forms had no effect on insertion of dGTP or dATP with Klenow exo- and human pol η, as compared to unmodified templating cytosine (29).

To provide clarity to the disparate results observed with DNA polymerases regarding the templating properties of oxidized cytosine, we measured single-turnover and steady-state kinetics of nucleotide insertion by human DNA pol β. The results indicate that the oxidized forms of 5mC do not significantly alter the fidelity of pol β. A series of crystal structures reveal the molecular basis for the accommodation of these epigenetic modifications in the active site of pol β.

Results

We first measured the single-turnover rate constants of pol β-catalyzed nucleotide insertion with 1-nt gap DNA substrates containing C, 5mC, 5hmC, 5fC, or 5caC in the templating base position (Figure 1). In these reactions, pol β and dGTP are at saturating concentrations allowing for the determination of the observed rate constant, $k_{pol}$. The $k_{pol}$ values are comparable among these varying templating base DNA substrates (Figure 2 and Table 1), suggesting that the modifications do not alter the rate-limiting step of insertion. Although the $k_{pol}$ values are similar, the reaction amplitudes vary between the substrates (Figure 2). Longer reaction timepoints revealed that a fraction of the 5fC (~25%) and 5caC (~40%) substrates were not extendable, suggesting a population of these templates were altered.

To determine if the cytosine modifications alter the fidelity of pol β, catalytic efficiencies were
DNA replication of epigenetically modified cytosine

measured for correct (dGTP) and incorrect (dATP, dTTP, and dCTP) nucleotide insertions. The catalytic efficiency ($k_{cat}/K_m$) for nucleotide insertion is a measure of the productive associations between the binary complex (Polβ-DNA) and dNTP. Templating bases 5mC, 5hmC, or 5fC did not significantly alter the catalytic efficiencies for insertion of dGTP (Figure 3 and Table 2). In contrast, templating 5caC decreased the catalytic efficiency of dGTP insertion by ~20-fold as compared to a 1-nt gap templating C substrate (Figure 3 and Table 2). DNA polymerase fidelity can be determined by comparing the catalytic efficiencies of the correct insertion to those of incorrect insertion (correct/incorrect). The catalytic efficiencies for misinsertion were measured for each template (Table 2), and the results are shown on a discrimination plot (Figure 3). The fidelity among the various templates can be visually appreciated by comparing the distance between the correct and incorrect nucleotide insertion on the plot (30). Figure 3 shows that the fidelity of pol β is not significantly altered when replicating these templating cytosine modifications.

To provide structural insight into the positioning of 5-modified cytosine within the active site of pol β, we solved the structure of pol β in complex with 5hmC and 5caC in the templating position and compared the results with cytosine. The crystallographic and refinement data statistics are shown in Table S1.

First, a binary complex structure of pol β with 5caC in the template position was solved. The structure revealed clear density for the carboxy group and positioning of the 5caC base in the canonical anti-conformation (Figure S1). To form a ternary complex, the binary complex crystals were soaked with a non-hydrolysable dGTP analog, dGP(CH$_2$)PP. Binding of dGP(CH$_2$)PP resulted in Pol β undergoing the well-recognized open to close conformational change; Watson-Crick base pairing between the incoming dGP(CH$_2$)PP base and templating 5caC base is observed (Figure 4). Overlay of ternary complex structures of templating C and 5caC reveals that the 5'-phosphate backbone of the 5caC templating base is repositioned (Figure 4). This repositioning of the phosphate backbone is a consequence of the repulsion between the 5'-carboxylic group and non-bridging oxygen of the templating phosphate. Lysine 280 is positioned to make electrostatic and hydrogen bonding interactions with the 5-carboxy group and the non-bridging oxygen of the templating phosphate backbone (Figure 4 inset).

Based on these results, we suspected the formyl and hydroxymethyl modifications containing only one oxygen could be positioned in either two conformations, or somewhere in between. To provide insight into the positioning of the oxygen atom on 5hmC and 5fC in the active site of pol β, we solved a ternary complex structure of 5hmC with the non-hydrolysable dGTP. Like the structure of 5caC, Watson-Crick base pairing was observed between the base of the incoming dGTP analog and 5hmC, and clear density for the 5hmC modification was present. The 5hmC modification is positioned to form a stabilizing hydrogen bond with N4 of cytosine and avoiding a steric clash with the templating phosphate backbone (Figure 5).

Discussion

The epigenetic modification of cytosine is found throughout vertebrates and is used to regulate transcription. An estimated 70%-80% of CpG sites are methylated in human somatic cells, making it likely that DNA repair pathways would encounter such modifications. Here we show that 5mC and its three oxidized forms in the templating base position do not alter the single-turnover rate constant, $k_{pol}$, or the fidelity of DNA synthesis. However, the efficiency of insertion was decreased ~20-fold with templating 5caC compared with the other base modifications. Structural analysis revealed backbone repositioning of the DNA template due to a steric clash between the carboxy group and the non-bridging phosphate oxygen of the template, and this correlates with the reduced catalytic efficiency observed with the 5caC template base. Lysine 280 stabilizes the altered backbone conformation by directly interacting with the carboxy modification and the backbone phosphate non-bridging oxygens.

Contrary to previous reports of 5fC and 5caC as being pro-mutagenic, we observe no evidence for their mutagenicity. Using catalytic efficiencies, we can measure mutation frequencies of ~0.0001% (for every million insertions one mistake). A 1% mutation frequency would correspond to a catalytic efficiency of ~0.01 $\mu$M$^{-1}$ s$^{-1}$. Interestingly, we do observe a ~1% mutation frequency ($k_{cat}/K_m$ ~0.01 $\mu$M$^{-1}$s$^{-1}$) for the insertion of dATP, but this is observed for all C templates,
indicating that the modifications do not enhance dATP insertion. This is consistent with previous studies that revealed the formyl and carboxy modifications do not promote the tautomeric form of cytosine (25,26). It is tempting to extrapolate these findings to other eukaryotic polymerases, however, since the fidelity of dATP insertion opposite templating cytosine is low with pol β it may not be the best exemplar. Polymerases that have high fidelity for the insertion of dATP opposite templating C may show differential effects with these modified templating cytosines. Accordingly, further studies with other eukaryotic polymerases would need to be performed before such broad conclusions are made.

The methyl, hydroxymethyl, and formyl modifications of cytosine are accommodated within the active site of Pol β without the need for side chain or DNA repositioning, consistent with the results of single-turnover kinetics displaying no effect with these modifications. However, the accommodation of the carboxyl modification in the active site requires template repositioning to avoid a steric/repulsive clash. This required repositioning likely leads to a reduction of productive ternary complexes resulting in a decrease in catalytic efficiency for dGTP insertion. Because the methyl, hydroxymethyl, and formyl are positioned toward the N4 exocyclic amine of cytosine, a steric/repulsive clash with the backbone is avoided. However, the carboxy modification has two oxygens, so if an intramolecular hydrogen bond is maintained with N4 and one oxygen, a steric/repulsive clash will occur with the other carbonyl oxygen and the phosphate backbone. This is clearly illustrated in Figure 4.

It is worthwhile to consider the alternative conformations that the carboxy modification or cytosine base may occupy. For instance, the carboxy modification could rotate relative to the cytosine base (not planar with the cytosine base) breaking the intramolecular H-bond with N4 but avoid a steric clash with the phosphate backbone. However, it must be energetically more favorable to maintain the intramolecular hydrogen bound and shift the template backbone. Another potential position that could occur to avoid the steric clash is rotation of the N-glycosidic bond, positioning the cytosine base from the canonical anti-conformation and into the syn-conformation. Presumably, this would happen in the binary state before binding nucleotide, as this is what is observed in the templating 8-oxoguanine binary structure (31). However, in the 5caC binary structure, the carboxy modification is positioned far enough away from the backbone phosphate to avoid the clash. Only upon base pairing with the incoming dGTP does the templating 5caC shift into the steric clash position.

In previous structures, K280 can be observed in different positions depending on the templating base. However, the methylene portion of the sidechain is generally positioned to make Van der Waals interactions with the templating base and the epsilon amino group typically interacts with a non-bridging oxygen of the templating phosphate backbone (31,32). The 5caC structure reported here is the first observation of K280 making a specific interaction with a templating base. This highlights the versatility of the polymerase active site.

The repositioned conformation observed in the ternary 5caC structure reveals the dynamic nature of the steric block that is created by a kink in the templating DNA strand when bound to the polymerase. This steric block lies on the major groove side of the templating base and discriminates against modifications by shaping the allowable space (Figure 6). Previous studies revealed that this steric block can be repositioned (“opened”) when 8-oxoguanine is in the templating position, allowing the base to sample both anti- and syn-conformations in the binary complex. As observed here in the 5caC ternary complex structure, opening of the steric-block mitigates a potential steric clash with the phosphate backbone of the templating strand. However, in contrast to the binary complex structure of pol β with templating 8-oxoguanine, the carboxy group of 5caC is accommodated within the binary complex structure because it can occupy available downstream space, highlighting a difference between major groove facing modifications of templating purines and pyrimidines.

DNA polymerases will encounter 5C modified cytosines in the templating strand during repair and replication. The consequences these modifications have on the efficiency and the fidelity of different polymerases are important for understanding overall genome integrity. Here we demonstrate that the fidelity of DNA pol β is maintained when encountering these modified cytosines in the template strand. This suggests that
the fidelity of BER remains unaltered in the presence of epigenetically modified cytosine.

Methods
The oligonucleotides used are from Integrated DNA Technologies (IDT; Coralville, IA), and the modified cytosine oligos are from the Midland Reagent Company Inc (Midland, Texas). The FAM-labeled oligos were purified using high performance liquid chromatography (HPLC), and the unlabeled oligos were purified using polyacrylamide gel electrophoresis (PAGE). Concentrations of the oligos were determined by UV absorbance at 260 nm using the extinction coefficients given by the manufacturer. Double stranded DNA substrates with a 1-nucleotide (nt) gap were prepared by annealing the primer, downstream, and template oligos at a 1:1:1:1.1 molar ratio respectively. The sequences of the various oligonucleotides can be found in the supplementary data (Table S1). Substrate annealing was performed using a thermocycler, with heating at 95°C for 5 min, followed by a temperature drop of 1°C per min until 10°C was reached.

Pol β was expressed and purified as described for DNA polymerase μ, with some alterations to the purification strategy (33). Full-length GST-TEV-pol β was expressed from a pGEX4T3 TEV-modified vector in Rosetta2 (DE3) cells (33). After sonication and centrifugation, the lysate was incubated with glutathione 4B resin, washed, and then treated with TEV protease overnight at 4°C. Eluted pol β was further purified by cation exchange (SP sepharose) and size exclusion chromatography. TEV cleaved pol β contains seven non-cognate amino acids on the N-terminus (GSNSRVD). The concentration was determined by absorbance at 280 nm using an extinction coefficient of 23,380 M$^{-1}$ cm$^{-1}$. The pol β concentrations reported throughout reflect the value determined by this method. Pol β was dialyzed into storage buffer (25 mM Tris, pH 7.4, 100 mM KCl, 1 mM TCEP, and 10% glycerol), concentrated to 185 µM, frozen in liquid nitrogen, and stored at -80°C.

Kinetic parameters for gap-filling catalyzed by pol β were determined by two different methods depending on whether a correct insertion or misinsertion of a nucleotide was being measured. The correct insertion reactions consisted of a single time-point measurement that was collected to determine the initial velocity of Pol β, as previously described (34). The single time-point measurement contained 1 nM pol β, 200 nM DNA substrate, and varying concentrations of dGTP (0.2 - 100 µM). Reactions were performed at 37°C and contained 50 mM Tris-HCl, pH 7.4, 100 mM KCl, 1 mM DTT, 0.1 mg/mL BSA, 10% glycerol, and 5 mM MgCl₂. Reactions (45 µl) were initiated by adding 5 µl of pol β and quenched at the appropriate time with 50 µl of 200 mM EDTA, 80% formamide, ~0.1% bromophenol blue, and ~0.1% xylene cyanol. Reactions were then heated to 95°C for 5 minutes, placed on ice, and loaded into a 22 % denaturing polyacrylamide gel. The resulting gels were scanned using a Typhoon scanner and quantified using ImageQuant TL using the rolling ball method. The initial velocity normalized by the enzyme concentration ($V_0/|E|$) was calculated by multiplying the fraction product by the concentration of substrate (200 nM) and dividing by the time-point and concentration of Pol β (1 nM). The resulting $V_0/|E|$ value was plotted against the respective dGTP concentration and the data were fit to the Michaelis-Menten equation (Figure S2 and S3). Because the templates containing 5fC and 5caC were not fully extendable (~75 and ~60 %) the substrate concentration was corrected (i.e. 150 nM and 120 nM for 5fC and 5caC, respectively).

The misinsertion kinetic assays were performed under the same standard conditions described above. Reactions were performed with sub-saturating concentrations of dNTP, 2 or 20 nM pol β, and 200 nM DNA substrate. An aliquot was removed at the appropriate time to be quenched in an equal volume of 300 mM EDTA. Reactions were then heated to 95°C for 5 minutes and then loaded into a 18 % or 22% denaturing polyacrylamide gel. Gels were visualized and analyzed as described above. Initial velocities were determined by fitting each time course to a straight line (Figure S2). The resulting slopes were then divided by the pol β concentration and plotted vs their respective dNTP concentration and fit to a straight line. The resulting slope is the catalytic efficiency ($k_{cat}/K_M$). In some cases, the catalytic efficiency was measured using a single sub-saturating concentration of nucleotide (Figure S4). Under these circumstances the catalytic efficiencies were determined by dividing the slope by the enzyme concentration and multiplying by the nucleotide concentration.
Single-turnover reactions were performed with a KinTek rapid quench-flow apparatus, as previously described (35,36). The single turnover assays were performed under the following conditions: 50 mM Tris, pH 7.4, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 10% glycerol, and 0.1 mg/ml BSA at 37 °C. Reactions were initiated by mixing a solution containing 100 nM DNA and 100 µM dGTP, with a 2 µM concentration of pol β equilibrated in reaction buffer. Reactions were quenched by addition of 100 mM EDTA. Samples were resolved by 22% denaturing PAGE and quantified using ImageQuant TL with the rolling ball method. Plots of fraction product versus time were fit to a single -exponential equation. Data points from two experiments are shown for the 5mC substrate and the data are fitted together. These experiments are independent with regards to time (performed on separate days), and the enzyme and substrate mixtures were prepared separately, but from the same stocks. The $k_{obs}$ values reported here are also comparable to our lab’s previously published work: 2.8 s⁻¹ (37), 2.9 s⁻¹ (38), 3.3 s⁻¹ (39). The $k_{obs}$ values for templating C, 5mC, 5hmC, 5fC, and 5caC are considered to be within experimental error of each other. Kinetic data were analyzed using Graphpad (prism) using a non-linear regression fitting procedure.

Binary complex crystals of Human DNA polymerase β with the modified dC as the templating base in a 1-nucleotide gapped DNA were grown as described. The sequence of the template strand (16-mer) was 5′-CCGACC*GCGCATCAG C-3′ (C* = 5caC or 5hmC). The primer strand (10-mer) sequence was 5′-GCTGATGCGC-3′. The downstream oligonucleotide (5-mer) was phosphorylated, and the sequence was 5′-GTCCC-3′. These oligonucleotides were annealed in a ratio of 1:1:1 by heating at 90°C for 10 minutes and cooling to 4 °C (1 °C/min) using a PCR thermocycler resulting in a 1 mM mixture of 1-nucleotide gapped duplex DNA with the modified dC base in the templating position. This annealed mixture was further incubated with an equal volume of pol β (14 mg/ml). The pol β-DNA complex was crystallized by sitting-drop vapor diffusion at 18°C by mixing 2 µl of the complex with 2 µl of the crystallization buffer. The crystallization buffer consisted of 14-16 % PEG2000 monomethyl ether, 350 mM sodium acetate, and 50 mM imidazole, pH 7.5. Crystals grew in approximately 2-4 days after seeding. The ternary complexes were obtained by soaking crystals of binary complex in artificial mother liquor with 50 mM MgCl₂, 20% PEG2000 monomethyl ether, 15% ethylene glycol, and 2 mM (α,β)-CH₂-dGTP for 1-2 hours. The crystals were then flash-frozen to 100 K in a nitrogen stream. Diffraction quality data were then collected for the binary and ternary complex crystals as described below.

Data were collected at 100 K on a CCD detector system mounted on a MiraMax®-007HF (Rigaku Corporation) rotating anode generator. Data were integrated and reduced with HKL2000 software (40). All crystals belong to the space group $P2_1$. Binary complex and ternary complex structures were solved by molecular replacement using 3ISB, and 2FMS as a reference models respectively. The structure was refined using PHENIX (41) and manual model building using Coot (42). The crystallographic statistics are reported in supplementary Table 2.

Structure factors and the coordinates for the human DNA polymerase β complexes with 5caC (6N2R Binary and 6N2S ternary) and 5hmC (6N2T ternary) have been deposited in the Protein Data Bank.
Acknowledgements: This research was supported by research project numbers Z01ES050158 and Z01ES050159 (S.H.W.) in the intramural research program of the National Institutes of Health, NIEHS. We thank Lars Pedersen and Andrea Moon of the x-ray crystallography core laboratory at NIEHS for discussion.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: MJH, KGF, and DDS designed, performed, and analyzed the enzyme kinetics experiments. VKB preformed the x-ray crystallography experiments and analyses. SHW provided the research direction, commented on the results and the manuscript at all stages, and revised the manuscript. MJH and KGF wrote the paper.
DNA replication of epigenetically modified cytosine

References
1. Bernstein, B. E., Meissner, A., and Lander, E. S. (2007) The mammalian epigenome. Cell 128, 669-681
2. Hermann, A., Gowher, H., and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. Cell Mol Life Sci 61, 2571-2587
3. Jabbari, K., and Bernardi, G. (2004) Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene 333, 143-149
4. Holliday, R., and Pugh, J. E. (1975) DNA modification mechanisms and gene activity during development. Science 187, 226-232
5. Iwan, K., Rahimoff, R., Kirchner, A., Spada, F., Schroder, A. S., Kosmatchev, O., Ferizaj, S., Steinbacher, J., Parsa, E., Muller, M., and Carell, T. (2018) 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. Nat Chem Biol 14, 72-78
6. Zhu, J. K. (2009) Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet 43, 143-166
7. Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., and Elf, J. (2012) The lac repressor displays facilitated diffusion in living cells. Science 336, 1595-1598
8. Rasmussen, K. D., and Helin, K. (2016) Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 30, 733-750
9. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333, 1300-1303
10. He, Y. F., Li, B. Z., Li, 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., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333, 1303-1307
11. Maiti, A., and Drohat, A. C. (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
12. Doetsch, P. W., Helland, D. E., and Haseltine, W. A. (1986) Mechanism of action of a mammalian DNA repair endonuclease. Biochemistry 25, 2212-2220
13. Matsumoto, Y., and Kim, K. (1995) Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. Science 269, 699-702
14. Singhal, R. K., and Wilson, S. H. (1993) Short gap-filling synthesis by DNA polymerase beta is processive. J Biol Chem 268, 15906-15911
15. Ellenberger, T., and Tomkinson, A. E. (2008) Eukaryotic DNA ligases: structural and functional insights. Annu Rev Biochem 77, 313-338
16. Branco, M. R., Ficz, G., and Reik, W. (2011) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* **13**, 7-13
17. Kriaucionis, S., and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930
18. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935
19. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935
20. Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., and Yi, C. (2017) Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* **20**, 720-731 e725
21. Kasymov, R. D., Grin, I. R., Endutkin, A. V., Smirnov, S. L., Ishchenko, A. A., Saparbaev, M. K., and Zharkov, D. O. (2013) Excision of 8-oxoguanine from methylated CpG dinucleotides by human 8-oxoguanine DNA glycosylase. *FEBS Lett* **587**, 3129-3134
22. Sassa, A., Caglayan, M., Dyrkkeeva, N. S., Beard, W. A., and Wilson, S. H. (2014) Base excision repair of tandem modifications in a methylated CpG dinucleotide. *J Biol Chem* **289**, 13996-14008
23. Karino, N., Ueno, Y., and Matsuda, A. (2001) Synthesis and properties of oligonucleotides containing 5-formyl-2'-deoxycytidine: in vitro DNA polymerase reactions on DNA templates containing 5-formyl-2'-deoxycytidine. *Nucleic Acids Res* **29**, 2456-2463
24. Munzel, M., Lischke, U., Stathis, D., Paffeneder, T., Gnerlich, F. A., Deiml, C. A., Koch, S. C., Karaghiosoff, K., and Carell, T. (2011) Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry* **17**, 13782-13788
25. Szulik, M. W., Pallan, P. S., Nocek, B., Voehler, M., Banerjee, S., Brooks, S., Joachimiak, A., Egli, M., Eichman, B. F., and Stone, M. P. (2015) Differential stabilities and sequence-dependent base pair opening dynamics of Watson-Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochemistry* **54**, 1294-1305
26. Dai, Q., Sanstead, P. J., Peng, C. S., Han, D., He, C., and Tokmakoff, A. (2016) Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chem Biol* **11**, 470-477
27. Xing, X. W., Liu, Y. L., Vargas, M., Wang, Y., Feng, Y. Q., Zhou, X., and Yuan, B. F. (2013) Mutagenic and cytotoxic properties of oxidation products of 5-methylcytosine revealed by next-generation sequencing. *PLoS One* **8**, e72993
28. Kamiya, H., Tsuchiya, H., Karino, N., Ueno, Y., Matsuda, A., and Harashima, H. (2002) Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells. *J Biochem* **132**, 551-555
29. Shibutani, T., Ito, S., Toda, M., Kanoa, R., Collins, L. B., Shibata, M., Urabe, M., Koseki, H., Masuda, Y., Swenberg, J. A., Masutani, C., Hanaoka, F., Iwai, S., and Kuraoka, I. (2014) Guanine-5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci Rep* **4**, 5220
30. Cavanaugh, N. A., Beard, W. A., Batra, V. K., Perera, L., Pedersen, L. G., and Wilson, S. H. (2011) Molecular insights into DNA polymerase deterrents for ribonucleotide insertion. *J Biol Chem* **286**, 31650-31660
31. Batra, V. K., Shock, D. D., Beard, W. A., McKenna, C. E., and Wilson, S. H. (2012) Binary complex crystal structure of DNA polymerase beta reveals multiple conformations of the templating 8-oxoguanine lesion. *Proc Natl Acad Sci U S A* **109**, 113-118
32. Beard, W. A., Shock, D. D., Yang, X. P., DeLauder, S. F., and Wilson, S. H. (2002) Loss of DNA polymerase beta stacking interactions with templating purines, but not pyrimidines, alters catalytic efficiency and fidelity. *J Biol Chem* **277**, 8235-8242

33. Moon, A. F., Pryor, J. M., Ramsden, D. A., Kunkel, T. A., Bebenek, K., and Pedersen, L. C. (2014) Sustained active site rigidity during synthesis by human DNA polymerase mu. *Nat Struct Mol Biol* **21**, 253-260

34. Beard, W. A., Shock, D. D., and Wilson, S. H. (2004) Influence of DNA structure on DNA polymerase beta active site function: extension of mutagenic DNA intermediates. *J Biol Chem* **279**, 31921-31929

35. Werneburg, B. G., Ahn, J., Zhong, X., Hondal, R. J., Kraynov, V. S., and Tsai, M. D. (1996) DNA polymerase beta: pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity. *Biochemistry* **35**, 7041-7050

36. Johnson, K. A. (1995) Rapid quench kinetic analysis of polymerases, adenosinetriphosphatases, and enzyme intermediates. *Methods Enzymol* **249**, 38-61

37. Howard, M. J., Rodriguez, Y., and Wilson, S. H. (2017) DNA polymerase beta uses its lyase domain in a processive search for DNA damage. *Nucleic Acids Res* **45**, 3822-3832

38. Howard, M. J., and Wilson, S. H. (2017) Processive searching ability varies among members of the gap-filling DNA polymerase X family. *J Biol Chem*

39. Beard, W. A., Shock, D. D., Batra, V. K., Prasad, R., and Wilson, S. H. (2014) Substrate-induced DNA polymerase beta activation. *J Biol Chem* **289**, 31411-31422

40. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307-326

41. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221

42. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612

**FOOTNOTES**

1 Co-first authors.

*This work was supported by NIH grants Z01-ES50158 and Z01-ES050161 (Intramural Research Program, NIEHS).

2 The abbreviations used are: pol, DNA polymerase; 5mC, 5-methylcytosine; 5hmC, 5-hydromethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxycytosine; TET, ten-eleven transfer; BER, base excision repair

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Table 1. Single-turnover kinetic summary for the insertion of dGTP opposite indicated templating base. Data were collected as described in experimental methods and Figure legend 2. The reported error is from the fit.

| Template | $k_{pol}$ (s$^{-1}$) |
|----------|----------------------|
| C        | 2.8 ± 0.2            |
| 5mC      | 2.9 ± 0.1            |
| 5hmC     | 2.7 ± 0.1            |
| 5fC      | 2.6 ± 0.1            |
| 5caC     | 2.2 ± 0.3            |

Table 2. Steady-state kinetic parameters for pol β catalyzed nucleotide insertion with modified templating cytosine. Data were collected as described in experimental methods and Figure S2. The reported error for dGTP insertions is the standard deviation of the indicated number of replicates. The standard error from fitting is reported for the misinsertion data. “nd” is not determined.

| Template | dNTP  | $k_{cat}$ (s$^{-1}$) | $K_M$ (µM) | $k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$) | n |
|----------|-------|----------------------|------------|-------------------------------------|---|
| C        | dGTP  | 0.65 ± 0.15          | 0.52 ± 0.14| 1.25 ± 0.42                        | 3 |
| C        | dATP  | nd                   | nd         | 0.0039 ± 0.0002                     | 1 |
| C        | dTTP  | nd                   | nd         | (9.9 ± 0.5) x 10$^{-6}$             | 1 |
| C        | dCTP  | nd                   | nd         | (3.75 ± 0.2) x 10$^{-6}$            | 1 |
| 5mC      | dGTP  | 1.05 ± 0.53          | 1.27 ± 0.28| 0.83 ± 0.43                        | 3 |
| 5mC      | dATP  | nd                   | nd         | 0.0059 ± 0.0002                     | 1 |
| 5mC      | dTTP  | nd                   | nd         | (2.54 ± 0.2) x 10$^{-5}$            | 1 |
| 5mC      | dCTP  | nd                   | nd         | (7.9 ± 0.2) x 10$^{-6}$             | 1 |
| 5hmC     | dGTP  | 0.56 ± 0.28          | 1.54 ± 0.38| 0.36 ± 0.2                         | 3 |
| 5hmC     | dATP  | nd                   | nd         | 0.0024 ± 0.0001                     | 1 |
| 5hmC     | dTTP  | nd                   | nd         | (2.39 ± 0.5) x 10$^{-5}$            | 1 |
| 5hmC     | dCTP  | nd                   | nd         | (1.32 ± 0.1) x 10$^{-5}$            | 1 |
| 5fC      | dGTP  | 0.87 ± 0.15          | 0.58 ± 0.26| 1.50 ± 0.71                        | 3 |
| 5fC      | dATP  | nd                   | nd         | 0.0029 ± 0.0001                     | 1 |
| 5fC      | dTTP  | nd                   | nd         | (4.82 ± 0.6) x 10$^{-5}$            | 1 |
| 5fC      | dCTP  | nd                   | nd         | (1.19 ± 0.1) x 10$^{-5}$            | 1 |
| 5caC     | dGTP  | 0.18 ± 0.01          | 2.75 ± 0.15| 0.065 ± 0.01                       | 2 |
| 5caC     | dATP  | nd                   | nd         | 0.00028 ± 0.00003                   | 1 |
| 5caC     | dTTP  | nd                   | nd         | (2.6 ± 0.6) x 10$^{-5}$             | 1 |
**DNA replication of epigenetically modified cytosine**

| 5caC | dCTP | nd | nd | (5.0 ± 0.5) x 10^{-5} | 1 |

**Figures**

**Figure 1. DNA substrates used in this study.** A 1-nt gap DNA substrate containing templating C or position 5-modified C (x) as the templating base. The primer oligo contains fluorescein (FAM) on the 5’ end and the downstream oligo contains a 5’ phosphate.

**Figure 2. Single-turnover analysis of Pol β catalyzed nucleotide insertion with modified cytosine templates.** Reactions were performed with saturating concentrations of Pol β (1 µM), indicated DNA substrate (50 nM), and dGTP (100 µM). Data are fit to a single-exponential equation yielding observed rate constants of 2.8 ± 0.2, 2.9 ± 0.1, 2.7 ± 0.1, 2.6 ± 0.1, and 2.2 ± 0.3 for templating C, 5mC, 5hmC, 5fC, and 5caC, respectively. The reported error is from the fit.
Figure 3. Discrimination plot representing the catalytic efficiencies for nucleotide insertion catalyzed by Pol β with modified cytosine templates. The identity of the nucleotide is indicated by color and the template identity is represented on the x-axis. Error bars are only represented for dGTP insertions and represent the standard deviation of three experiments, except 5caC which was performed twice.
Figure 4. Structural comparison of nascent base pairs in ternary complexes containing templating cytosine and 5-caC. Nascent base pair as for templating cytosine and incoming dGTP (PDB 4UB4) shown in cyan color. Nascent base pair between templating 5caC and incoming nonhydrolyzable dGTP is shown in gold color. The Helix N of the binary complex structure (open conformation) and the ternary complex structure (closed conformation) are shown in pink and gold color respectively. The repositioning of the 5’ phosphate backbone relative to templating cytosine ternary structure is depicted with a solid black line. The simulated-annealing fo-fc omit map for the 5caC moiety contoured at 3σ is shown in red.
Figure 5. Structural comparison of nascent base pairs in ternary complexes containing templating cytosine and 5-hmC. Nascent base pair as for templating cytosine and incoming dGTP (PDB 4UB4) shown as in pink color. Nascent base pair between templating 5-hmC and incoming nonhydrolyzable dGTP is shown in gold color. The repositioning of the 5’ phosphate backbone relative to templating cytosine ternary structure is not observed. The simulated-annealing fo-fc omit map for the 5-hmC moiety contoured at 3σ is shown in red.
Figure 6. Major groove steric block. Pol β bends the templating DNA strand, shaping the allowable space for major groove facing modifications of the templating base.
