

## Structural Basis for Error-Free Bypass of the 5-N-Methylformamidopyrimidine-dG Lesion by Human DNA Polymerase $\eta$ and *Sulfolobus solfataricus* P2 Polymerase IV

Amritraj Patra,<sup>†</sup> Surajit Banerjee,<sup>‡,§</sup> Tracy L. Johnson Salyard,<sup>‡</sup> Chanchal K. Malik,<sup>‡</sup> Plamen P. Christov,<sup>‡</sup> Carmelo J. Rizzo,<sup>‡</sup> Michael P. Stone,<sup>\*,‡</sup> and Martin Egli<sup>\*,†</sup>

<sup>†</sup>Department of Biochemistry, Center in Molecular Toxicology, Vanderbilt-Ingram Cancer Center, Vanderbilt Institute of Chemical Biology, Center for Structural Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, United States

<sup>‡</sup>Department of Chemistry, Center in Molecular Toxicology, Vanderbilt-Ingram Cancer Center, Vanderbilt Institute of Chemical Biology, Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States

<sup>§</sup>Northeastern Collaborative Access Team and Department of Chemistry and Chemical Biology, Cornell University, Argonne National Laboratory, Building 436E, Argonne, Illinois 60439, United States

### Supporting Information

**ABSTRACT:** *N*<sup>6</sup>-(2-Deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) arises from N7-methylation of deoxyguanosine followed by imidazole ring opening. The lesion has been reported to persist in animal tissues. Previous in vitro replication bypass investigations of the MeFapy-dG adduct revealed predominant insertion of C opposite the lesion, dependent on the identity of the DNA polymerase (Pol) and the local sequence context. Here we report crystal structures of ternary Pol·DNA·dNTP complexes between MeFapy-dG-adducted DNA template:primer duplexes and the Y-family polymerases human Pol  $\eta$  and P2 Pol IV (Dpo4) from *Sulfolobus solfataricus*. The structures of the hPol  $\eta$  and Dpo4 complexes at the insertion and extension stages, respectively, are representative of error-free replication, with MeFapy-dG in the anti conformation and forming Watson–Crick pairs with dCTP or dC.

Alkylating agents are the earliest class of chemotherapy drugs and are still commonly used to treat different types of cancers. These include monofunctional methylating agents such as temozolomide and bifunctional alkylating agents such as nitrogen mustards or chloroethylating agents.<sup>1</sup>

The guanine N7 position constitutes the most nucleophilic site in DNA.<sup>2</sup> Thus, DNA methylation occurs predominantly at that site, resulting in a cationic N7-methyldeoxyguanosine adduct.<sup>3</sup> This product can undergo further hydrolysis, yielding an abasic (AP) site<sup>4,5</sup> or the imidazole-ring-fragmented lesion *N*<sup>6</sup>-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) (Figure 1).<sup>3,6</sup> Opening of the purine imidazole ring depends on the N7 substituent and the pH of the medium. Whereas the ring opening under physiological conditions is slow, it becomes accelerated at higher temperature and in alkaline solution.

MeFapy-dG has been characterized *in vivo* in the liver of rats<sup>7,8</sup> and has also been observed in the urine of healthy humans.<sup>9</sup> In general, the MeFapy-dG adduct is considered non-

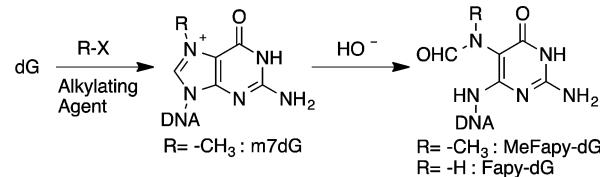


Figure 1. Formation of formamidopyrimidine lesions.

miscoding because the Watson–Crick face remains unaltered. However, various in vitro experiments using *Escherichia coli* DNA polymerase I Klenow fragment and T4 polymerase showed that MeFapy-dG blocks DNA chain elongation.<sup>10,11</sup> Replication of a site-specific MeFapy-dG lesion in primate cells gave complex mutational spectra with frequencies of 7–21% depending on the local sequence. Common mutations were G  $\rightarrow$  T transversions and deletions.<sup>12</sup> An in vitro replication study using the MeFapy-dG lesion with *Sulfolobus solfataricus* Dpo4 found miscoding, with the incorporation of all four nucleotides with various efficiencies depending on the DNA template sequence around the lesion.<sup>13</sup> Oligonucleotides with 5'-T-(MeFapy-dG)-G-3' resulted in error-free bypass, with insertion of dC opposite the adduct and full-length extension of the primer strand. By comparison, a 5'-T-(MeFapy-dG)-T-3' template triggered a one-base deletion, or misincorporation of dA opposite the MeFapy-dG lesion. The error-free bypass and extension efficiency by Dpo4 was estimated to be 74% for 5'-T-(MeFapy-dG)-G-3' and 51% for 5'-T-(MeFapy-dG)-T-3', along with 11% one-base deletion product for the latter template. Recent in vitro replication bypass experiments using human Y-family polymerases (hPols)  $\eta$ ,  $\kappa$ ,  $\iota$ , and Rev1 showed efficient translesion synthesis (TLS) by hPols  $\eta$  and  $\kappa$  with error-free insertion of dCTP opposite MeFapy-dG and extension in the above sequence contexts.<sup>14</sup> Among these TLS Pols, hPol  $\eta$  is the most efficient in the error-free bypass of MeFapy-dG (>70%).

Received: March 14, 2015

Published: May 19, 2015

To analyze the structural basis for the mostly error-free bypass of the MeFapy-dG adduct by hPol  $\eta$ , we determined the crystal structure of an hPol  $\eta$  complex trapped at the insertion stage, with MeFapy-dG opposite the nonhydrolyzable dCMPNPP analogue (in which an N atom bridges the  $\alpha$  and  $\beta$  P atoms), in the presence of Mg<sup>2+</sup> (Table 1). Further crystal

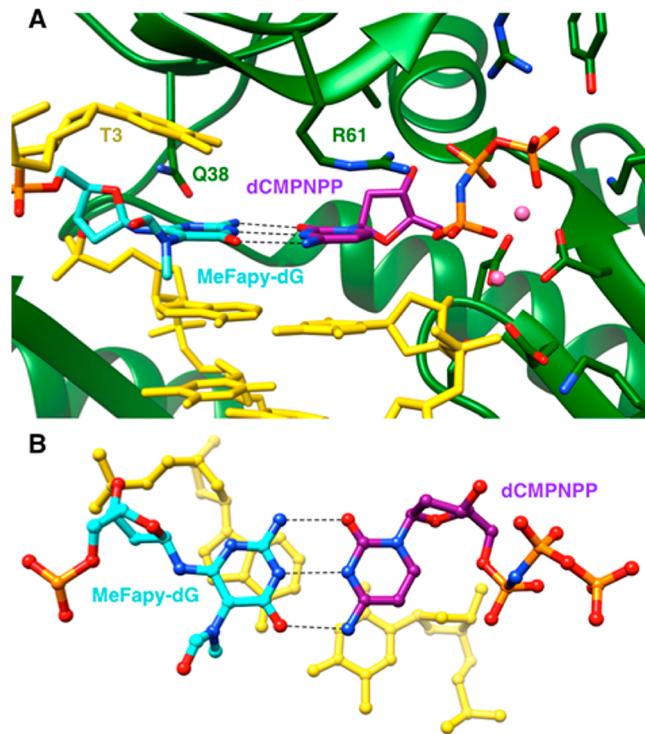
**Table 1.** DNA Sequences Used in the Crystallizations

complex	DNA sequence (X = MeFapy-dG)	incoming nucleotide
hPol $\eta$ (insertion)	3'-TCG CAG TAX TAC-5' 5'-AGC GTC AT-3'	dCMPNPP
ssDpo4 (insertion)	3'-CCC CCT TCC TAA GXT ACT-5' 5'-GGG GGA AGG ATT C-3'	dATP
ssDpo4 (extension)	3'-CCC CCT TCC TAA TXT ACT-5' 5'-GGG GGA AGG ATT AC-3'	dATP

structures concern the hPol  $\kappa$  homologue Dpo4 in complex with an MeFapy-dG-containing template–primer DNA duplex in the presence of Ca<sup>2+</sup> and trapped in two different phases of bypass. In the first complex, representative of the insertion stage and a  $-1$  frameshift, MeFapy-dG is unopposed by a residue from the primer, and instead the incoming dATP pairs with the T that is 5'-adjacent to the adduct on the template. In the second complex, representative of the extension stage and error-free bypass, MeFapy-dG pairs with dC at the  $-1$  position, and the incoming dATP pairs with the downstream T of the template.

The structure of the hPol  $\eta$  ternary complex with MeFapy-dG paired opposite incoming dCMPNPP was determined at a resolution of 2.48 Å (Figure 2 and Table 2). The DNA duplex consists of a 12-mer template containing MeFapy-dG and an 8-mer primer (Table 1). In the structure, all of the primer nucleotides were visible in the electron density maps along with 11 of the 12 template nucleotides (Table 2). An example of the quality of the final electron density is shown in Figure S1 in the Supporting Information. At the active site, the MeFapy-dG:dCMPNPP pair displays the expected Watson–Crick geometry, with the formamide moiety adopting an orientation that is more or less perpendicular to the plane of the six-membered ring (Figure 2A). Thus, the active-site configuration in the structure of the hPol  $\eta$ -MeFapy-dG complex is similar to that in the crystal structure of the complex between hPol  $\eta$  and native DNA with a G:dCMPNPP pair lodged at the active site.<sup>15</sup> Superimposition of the active sites of these two structures indicates that amino acids from the hPol  $\eta$  finger domain (i.e., Gln-38 and Arg-61) adopt similar orientations relative to the nascent base pair (Figure S2).

Two structures were determined for Dpo4 in complex with DNA duplexes containing MeFapy-dG-modified template strands (Table 1). Both crystals diffracted to ca. 3 Å and belong to space group P2<sub>1</sub>2<sub>1</sub>2 with a single complex per asymmetric unit (Table 2). The first structure features the template sequence 5'-T(MeFapy-dG)G-3' and represents a so-called type-II complex, in which the adduct is unopposed by a primer base, resulting in a  $-1$  frameshift (Figures 3 and S3). As is characteristic for a type-II complex,<sup>16</sup> Dpo4 simultaneously accommodates two template nucleotides inside the catalytic pocket, namely, MeFapy-dG and the 5'-adjacent T. Pairing between this T and the incoming dATP leaves a 6 Å gap between the  $\alpha$ -phosphate of the nucleotide triphosphate and the 3'-hydroxyl group at the primer terminus (Figure 3). The ability of Dpo4 to accommodate two template bases in its active



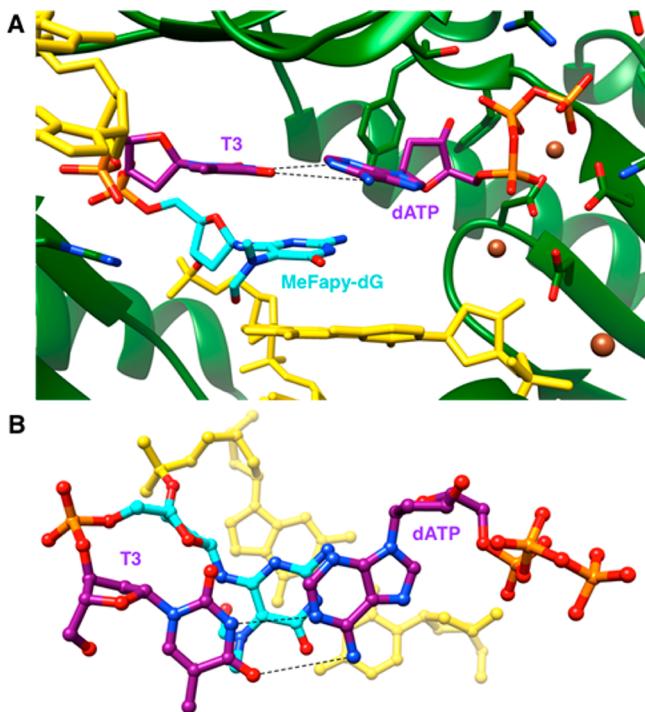
**Figure 2.** Active-site configuration in the ternary hPol  $\eta$  insertion-step complex with dCMPNPP opposite MeFapy-dG. (A) View into the DNA major groove. (B) View rotated by  $\sim 90^\circ$  around the horizontal axis and looking perpendicularly onto the nucleobase plane of the incoming dCMPNPP. Carbon atoms of MeFapy-dG and dCMPNPP are colored in cyan and purple, respectively, and Mg<sup>2+</sup> ions are pink spheres. Gln-38 forms a hydrogen bond to O4' of MeFapy-dG, but N3 in the minor groove is too far removed from the Gln-38 amide oxygen (3.9 Å) for hydrogen-bond formation.

**Table 2.** Selected Crystal Data, Diffraction Data Collection, and Refinement Parameter Statistics

	hPol $\eta$ (insertion)	ssDpo4 (insertion)	ssDpo4 (extension)
Data Collection			
space group	P6 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2
resolution [Å] <sup>a</sup>	50.0–2.65 (2.70–2.65)	50.0–3.10 (3.15–3.10)	30.0–2.90 (2.95–2.90)
<i>a</i> [Å]	98.97	95.45	94.38
<i>b</i> [Å]	98.97	102.72	103.97
<i>c</i> [Å]	81.62	53.43	52.56
completeness [%]	98.9 (100)	98.4 (84.7)	99.8 (100)
I/σ(I)	14.5 (1.9)	21.2 (2.9)	18.7 (1.8)
R <sub>merge</sub> [%]	14.7 (98.5)	11.2 (68.0)	14.3 (61.7)
redundancy	5.7 (5.6)	6.2 (3.1)	7.1 (7.1)
Refinement			
R <sub>work</sub> [%]	15.9 (22.1)	16.4 (25.9)	17.1 (26.9)
R <sub>free</sub> [%] <sup>b</sup>	22.9 (34.1)	25.6 (38.6)	23.5 (35.3)
avg <i>B</i> [Å <sup>2</sup> ]	47.7	88.2	69.8
RMSD bonds [Å]	0.011	0.010	0.010
RMSD angles [deg]	1.5	1.5	1.4
PDB ID <sup>c</sup>	4RU9	4RUA	4RUC

<sup>a</sup>Statistics for the highest-resolution shell are shown in parentheses.

<sup>b</sup>Based on 5% of the reflections. <sup>c</sup>Atomic coordinates and structure factors have been deposited in the Protein Data Bank (<http://wwpdb.org/>).

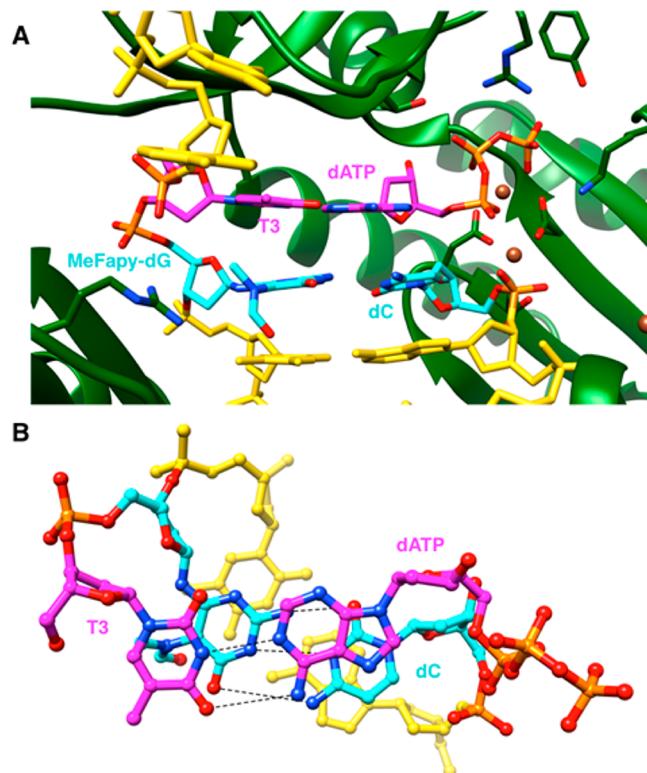


**Figure 3.** Active-site configuration in the ternary Dpo4 insertion-step complex with unpaired MeFapy-dG and incoming dATP opposite template T. (A) View into the DNA major groove. (B) View rotated by  $\sim 90^\circ$  around the horizontal axis and looking perpendicularly onto the nucleobase plane of the incoming dATP. Color codes match those in Figure 2 except that carbon atoms of the template T 5'-adjacent to the adduct are purple.  $\text{Ca}^{2+}$  ions are brown spheres.

site forms the basis for correct bypass of cyclic pyrimidine dimers (CPDs) by this Pol, with the type-II configuration and a misaligned template strand being consequences of the spacious active site.<sup>16</sup>

The second structure is of a postinsertion ternary complex involving the template sequence 5'-T(MeFapy-dG)T-3' and a 14-mer primer of which the 3'-terminal cytosine is intended to pair with the template MeFapy-dG (Figures 4 and S4). The incoming dATP pairs with the thymine base to the 5'-side of MeFapy-dG. Both the MeFapy-G:dC and dT:dATP base pairs are found in the standard Watson–Crick hydrogen-bonding configuration. As in the case of the hPol  $\eta$  complex (Figure 2), the formamide is rotated out of the plane of the six-membered ring in the two Dpo4 complexes (Figures 3 and 4). However, unlike in the hPol  $\eta$  complex, the formamide C=O is directed toward the residue 3'-adjacent to the adduct.

In vitro or in vivo formations of methylated or non-methylated formamidopyrimidines (MeFapy-dG or Fapy-dG, respectively) in DNA have been studied extensively.<sup>17</sup> Although the origins of these lesions are different (Fapy-dG is the result of oxidative damage), they are structurally very similar and only weakly mutagenic. Earlier studies established that Pols exhibit a strong preference for dCTP incorporation opposite the Fapy-dG lesion.<sup>18–21</sup> Recent work by Gehrke et al.<sup>22</sup> showed that the replication of the carbocyclic sugar analogue Fapy lesion (cFaPydG) by high-fidelity polymerase I from *Geobacillus stearothermophilus* (*Bst* Pol I) results in error-free bypass alongside a smaller amount of error-prone bypass. A previous NMR structure of a cFaPydG:dC-containing DNA duplex was



**Figure 4.** Active-site configuration in the ternary Dpo4 extension-step complex with primer dC opposite MeFapy-dG and dATP opposite template dT. (A) View into the DNA major groove. (B) View rotated by  $\sim 90^\circ$  around the horizontal axis and looking perpendicularly onto the nucleobase plane of the incoming dATP. Carbon atoms of the MeFapy-G:dC and dT:dATP pairs are colored in cyan and magenta, respectively.

also consistent with cFaPydG pairing with dC in a standard Watson–Crick fashion.<sup>23</sup>

N7-methyldeoxyguanosine (m7dG), the initial lesion formed by methylating agents (Figure 1), has also been the subject of several studies and classified as non-promutagenic.<sup>3,24</sup> Previous crystallographic studies using the chemically stable 2'-fluorom7dG analogue in a DNA sequence, in complex with either *E. coli* DNA glycosylase AlkA<sup>25</sup> or human DNA polymerase  $\beta$ ,<sup>26</sup> revealed that this polymerase bypasses m7dG accurately and that the lesion forms a canonical Watson–Crick base pair with incoming dCTP.

In summary, the three X-ray crystal structures of TLS Pols trapped either at the insertion or extension stages of MeFapy-dG bypass synthesis provide detailed insight into the basis of the mostly error-free replication of the adduct by hPol  $\eta$  and Dpo4, an hPol  $\kappa$  homologue. The structure of the hPol  $\eta$ -MeFapy-dG-dCMPNPP ternary complex is consistent with kinetic data of error-free bypass and extension by hPol  $\eta$ . The major groove is wide open at the active site of hPol  $\eta$ , allowing this Pol to potentially also bypass more bulky lesions such as BenzylFapy-dG (Figure S5).

In the hPol  $\eta$  complex, the formamide C=O points toward the 5'-adjacent T, whereas it is directed toward the 3'-adjacent G in the Dpo4 structure(s), perhaps as a result of the proximity of the formamide and 5'-template T in the latter case (Figures 2, 4, and S6). Thus, partial unstacking of the template T 5'-adjacent to the lesion at the active site of hPol  $\eta$  creates somewhat more room, allowing a virtually upright orientation

of the formamide. By comparison, the equivalent T and the six-membered ring of MeFapy-dG are tightly stacked at the Dpo4 active site, and a formamide orientation similar to that in the hPol  $\eta$  complex would create a tight spacing between C=O and the T C5-methyl substituent (Figure S6).

## ■ ASSOCIATED CONTENT

### S Supporting Information



## ■ AUTHOR INFORMATION

### Corresponding Authors

\*martin.egli@vanderbilt.edu

\*michael.p.stone@vanderbilt.edu

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by NIH Grants P01 CA160032, P30 ES00267, and P30 CA068485.

## ■ REFERENCES

- (1) Beranek, D. T. *Mutat. Res.* **1990**, *231*, 11–30.
- (2) Reiner, B.; Zamenhof, S. *J. Biol. Chem.* **1957**, *228*, 475–486.
- (3) Gates, K. S.; Noonan, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839–856.
- (4) Boiteux, S.; Guillet, M. *DNA Repair* **2004**, *3*, 1–12.
- (5) Loeb, L. A.; Preston, D. B. *Annu. Rev. Genet.* **1986**, *20*, 201–230.
- (6) Tudek, B. *J. Biochem. Mol. Biol.* **2003**, *36*, 12–19.
- (7) Beranek, D. T.; Weis, C. C.; Evans, F. E.; Chetsanga, C. J.; Kadlubar, F. F. *Biochem. Biophys. Res. Commun.* **1983**, *110*, 625–631.
- (8) Kadlubar, F. F.; Beranek, D. T.; Weis, C. C.; Evans, F. E.; Cox, R.; Irving, C. C. *Carcinogenesis* **1984**, *5*, 587–592.
- (9) Barak, R.; Vincze, A.; Bel, P.; Dutta, S. P.; Chedda, G. B. *Chem.-Biol. Interact.* **1993**, *86*, 29–40.
- (10) Boiteux, S.; Laval, J. *Biochem. Biophys. Res. Commun.* **1983**, *110*, 552–558.
- (11) O'Connor, T. R.; Boiteux, S.; Laval, J. *Nucleic Acids Res.* **1988**, *16*, 5879–5894.
- (12) Earley, L. F.; Minko, I. G.; Christov, P. P.; Rizzo, C. J.; Lloyd, R. S. *Chem. Res. Toxicol.* **2013**, *26*, 1108–1114.
- (13) Christov, P. P.; Angel, K. C.; Guengerich, F. P.; Rizzo, C. J. *Chem. Res. Toxicol.* **2009**, *22*, 1086–1095.
- (14) Christov, P. P.; Yamanaka, K.; Choi, J. Y.; Takata, K.; Wood, R. D.; Guengerich, F. P.; Lloyd, R. S.; Rizzo, C. J. *Chem. Res. Toxicol.* **2012**, *25*, 1652–1661.
- (15) Patra, A.; Nagy, L. D.; Zhang, Q.; Su, Y.; Muller, L.; Guengerich, F. P.; Egli, M. *J. Biol. Chem.* **2014**, *289*, 16867–16882.
- (16) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. *Mol. Cell* **2004**, *13*, 751–762.
- (17) Dizdaroglu, M.; Kirkal, G.; Jaruga, P. *Free Radical Biol. Med.* **2008**, *45*, 1610–1621.
- (18) Wiederholt, C. J.; Greenberg, M. M. *J. Am. Chem. Soc.* **2002**, *124*, 7278–7279.
- (19) Patro, J. N.; Wiederholt, C. J.; Jiang, Y. L.; Delaney, J. C.; Essigmann, J. M.; Greenberg, M. M. *Biochemistry* **2007**, *46*, 10202–10212.
- (20) Kalam, M. A.; Haraguchi, K.; Chandani, S.; Loechler, E. L.; Moriya, M.; Greenberg, M. M.; Basu, A. K. *Nucleic Acids Res.* **2006**, *34*, 2305–2315.
- (21) Ober, M.; Müller, H.; Pieck, C.; Gierlich, J.; Carell, T. *J. Am. Chem. Soc.* **2005**, *127*, 18143–18149.
- (22) Gehrke, T. H.; Lischke, U.; Gasteiger, K. L.; Schneider, S.; Arnold, S.; Müller, H. C.; Stephenson, D. S.; Zipse, H.; Carell, T. *Nat. Chem. Biol.* **2013**, *9*, 455–461.
- (23) Lukin, M.; Zaliznyak, T.; Attaluri, S.; Johnson, F.; de los Santos, C. *Chem. Res. Toxicol.* **2012**, *25*, 2423–2431.
- (24) Boysen, G.; Pachkowski, B. F.; Nakamura, J.; Swenberg, J. A. *Mutat. Res.* **2009**, *678*, 76–94.
- (25) Lee, S.; Bowman, B. R.; Ueno, Y.; Wang, S.; Verdine, G. L. *J. Am. Chem. Soc.* **2008**, *130*, 11570–11571.
- (26) Koag, M. C.; Kou, Y.; Ouzon-Shubeita, H.; Lee, S. *Nucleic Acids Res.* **2014**, *42*, 8755–8766.

## **Supporting Information**

# **Structural Basis for Error-Free Bypass of the 5-N-Methyl-formamidopyrimidine-dG Lesion by Human DNA Polymerase $\eta$ and *Sulfolobus solfataricus* P2 Polymerase IV**

Amritraj Patra, Surajit Baneerjee, Tracy L. Johnson Salyard, Chanchal K. Malik, Plamen P. Christov, Carmelo J Rizzo, Michael P. Stone\* and Martin Egli\*

## **Contents**

Page S2: Experimental Procedures

Page S4: Table S1. Crystal data, data collection parameters, and structure refinement statistics

Page S6. Figure S1. Quality of the final electron density (hPol  $\eta$  ternary complex)

Page S7. Figure S2: Superimposition of hPol  $\eta$ •MeFapy-dG•dCMPNPP and hPol  $\eta$ •G•dCMPNPP complexes

Page S8. Figure S3. Quality of the final electron density (Dpo4 insertion complex)

Page S9. Figure S4. Quality of the final electron density (Dpo4 extension complex)

Page S10. Figure S5. Active site configuration of hPol  $\eta$  bypassing BzFapy-dG (model)

Page S11. Figure S6. Alternative formamide conformations in the hPol  $\eta$  and Dpo4 complexes

Page S12. Acknowledgements

Page S13. References

## EXPERIMENTAL PROCEDURES

**hPol  $\eta$  Catalytic Core Protein Expression and Purification.** The hPol  $\eta$  plasmid (pET28a) comprising residues 1-432 was a generous gift from Dr. Wei Wang, NIDDK, NIH. The polymerase was expressed in *E. coli* and purified as described previously.<sup>1</sup> The protein solution was concentrated to 5 mg/mL.

**Dpo4 Protein Expression and Purification.** Dpo4 was expressed in *E. coli* and purified using heat denaturation, Ni-nitrilotriacetate affinity chromatography and ion-exchange chromatography as described previously.<sup>2</sup>

**Oligonucleotide Synthesis and Annealing.** DNA primer and template sequences used in the crystallization experiments are listed in Table 1 of the main paper. The synthesis of site-specifically modified oligonucleotides containing the MeFapy-dG lesion was previously reported.<sup>3</sup> All modified oligodeoxynucleotides were characterized by MALDI-TOF mass spectrometry. Capillary gel electrophoresis and C-18 HPLC confirmed their purities. Unmodified DNA primers were purchased from Integrated DNA Technologies (Coralville, IA). Template and primer strands were mixed in a 1:1 molar ratio and were annealed in the presence of 10 mM sodium HEPES buffer (pH 8.0), 0.1 mM EDTA, and 50 mM NaCl by heating for 10 min. at 85°C, followed by slow cooling to room temperature.

**Crystallization of the hPol  $\eta$ •MeFapy-dG DNA•dCMPNPP Insertion Complex.** The DNA template-primer duplex was mixed with the protein in a 1.2:1 molar ratio in the presence of excess 50 mM Tris-HCl, pH 7.5, 450 mM KCl, and 3 mM DTT. After adding 5  $\mu$ L of 100 mM MgCl<sub>2</sub> the complex was concentrated to a final concentration of ~2-3 mg/mL by ultrafiltration. Non-hydrolyzable nucleotide triphosphates were added last to form the ternary complexes. Crystallization experiments were performed by the hanging drop vapor diffusion technique at 18°C using a sparse matrix screen (Hampton Research, Aliso Viejo, CA).<sup>4</sup> One  $\mu$ L of the complex solution was mixed with 1  $\mu$ L of reservoir solution and equilibrated against 500  $\mu$ L reservoir wells. Crystals appeared in droplets containing 0.1 M MES (pH 5.5), 5 mM MgCl<sub>2</sub>, and 25% (w/v) PEG 2000 MME within one day and were harvested after a week. Crystals were

mounted in nylon loops, cryo-protected in reservoir solution containing 25% glycerol (v/v), and frozen in liquid nitrogen.

**Crystallization of Ternary Dpo4 Insertion and Extension Complexes.** The Dpo4 polymerase was concentrated to 50–70 mg/mL using a spin concentrator in 50 mM Tris-HCl (pH 7.4 at 25 °C) buffer containing 100 mM NaCl, 5 mM β-mercaptoethanol, and 50% glycerol (v/v). The polymerase was then combined with template:primer DNA duplex (1:1.2 molar ratio) and placed on ice for 1 h prior to incubation with 1 mM d(N)TP and 5 mM CaCl<sub>2</sub>. Crystals were grown using the sitting drop vapor diffusion method by mixing 1 μL of complex with 1 μL of a solution containing 50 mM Tris-HCl (pH 7.4 at 25 °C) buffer, 12–20% polyethylene glycol 3350 (w/v), 100 mM Ca(OAc)<sub>2</sub>, and 2.5% glycerol (v/v). Crystals were soaked in mother liquor containing an additional 25% polyethylene glycol 3350 (w/v) and 15% ethylene glycol (v/v) and frozen in liquid nitrogen.

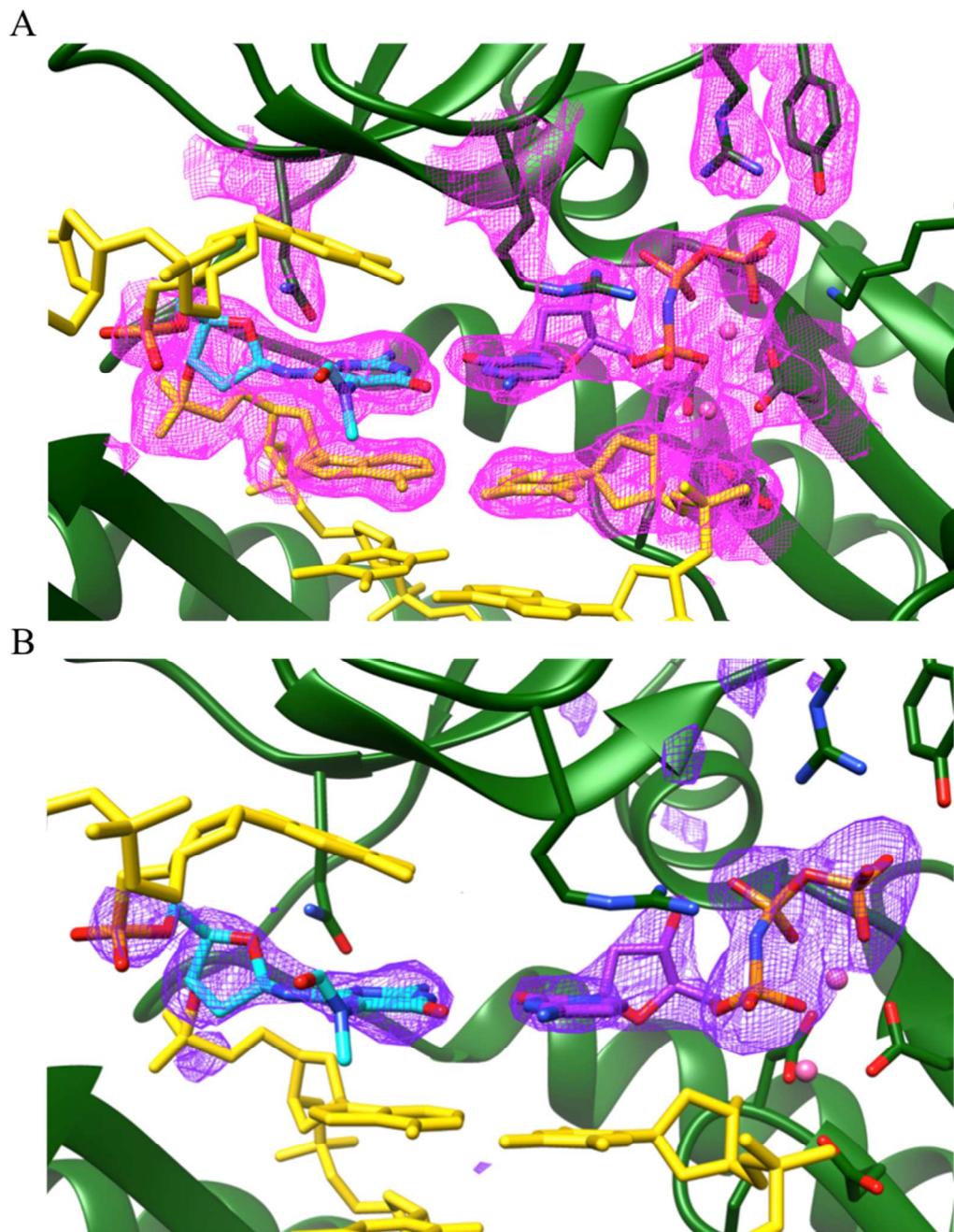
**X-ray Diffraction Data Collection, Structure Determination and Refinement.** Diffraction data were collected on the 21-ID-D beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL). All data were integrated and scaled with the program HKL2000.<sup>5</sup> The structures were determined by the Molecular Replacement technique with the program MOLREP.<sup>6,7</sup> For the complex of hPol  $\eta$  the structure with PDB entry 4O3N (protein only)<sup>8</sup> and for the complexes of the Dpo4 polymerase the structure with PDB ID code 2BQU (protein only)<sup>2</sup> were used as the search models. Structure refinement and model building were carried out with Refmac<sup>9</sup> and COOT,<sup>10</sup> respectively. The program Refmac is used to generate the omit maps. Illustrations were prepared with the program UCSF Chimera.<sup>11</sup>

**Table S1.** (Expanded version of Table 2 in the main paper) Crystal data, data collection parameters, and structure refinement statistics<sup>a</sup>

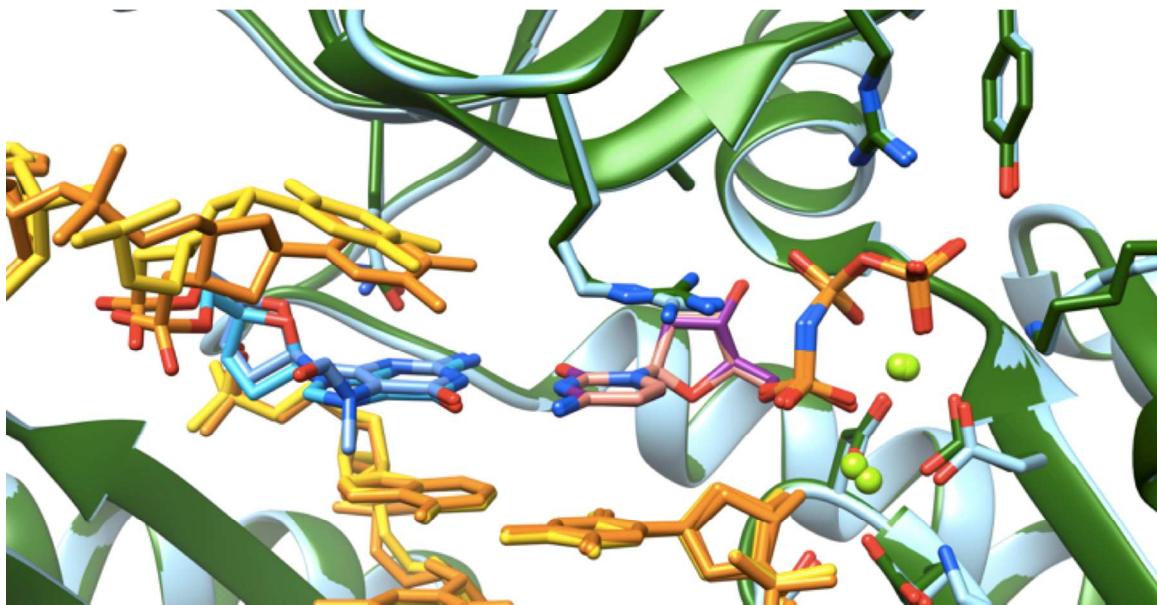
Complex	hPol $\eta$ : MeFapy-dG	Dpo4 : MeFapy-dG (Insertion)	Dpo4 : MeFapy-dG (Extension)
<i>Data Collection</i>			
Wavelength [Å]	1.07810	1.07810	1.07810
Space group	<i>P</i> 6 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2
Resolution [Å]	50.0 - 2.65 (2.70 - 2.65) <sup>a</sup>	50.0 - 3.10 (3.15 - 3.10)	30.0 - 2.90 (2.95 - 2.90)
Unit cell <i>a</i> , <i>b</i> , <i>c</i> [Å]	98.97, 98.97, 81.62	95.45, 102.72, 53.43	94.38, 103.97, 52.56
Unique reflections	13,176 (659)	10,163 (438)	12,098 (586)
Completeness [%]	98.9 (100)	98.4 (84.7)	99.8 (100)
I/σ(I)	14.5 (1.9)	21.2 (2.9)	18.7 (1.8)
Wilson B-factor [Å <sup>2</sup> ]	36.2	76.8	55.1
R-merge	0.147 (0.985)	0.112 (0.680)	0.143 (0.617)
Redundancy	5.7 (5.6)	6.2 (3.1)	7.1 (7.1)
<i>Refinement</i>			
R-work	0.159 (0.221)	0.164 (0.259)	0.171 (0.269)
R-free	0.229 (0.341)	0.256 (0.386)	0.235 (0.353)
Number of atoms			
Protein/DNA	3,358/391	2,744/611	2,744/630

dNTP/Water/ M <sup>2+</sup>	28/138/2(Mg <sup>2+</sup> )	30/36/3(Ca <sup>2+</sup> )	30/58/3(Ca <sup>2+</sup> )
Protein residues	427	341	341
B-factor [Å <sup>2</sup> ]			
Average	47.7	88.2	69.8
Protein/DNA	47.5/57.9	87.7/91.1	68.9/72.1
dNTP/M <sup>2+</sup> /Water	33.8/32.0/42.9	93.0/98.5/67.6	64.5/66.6/60.7
R.m.s. deviations			
bonds [Å]	0.011	0.010	0.010
angles [deg.]	1.5	1.5	1.4
Ramachandran plot			
Favored (%)	97	91	96
Allowed (%)	2.3	7.2	3.4
Outliers (%)	0.7	1.8	0.6
PDB ID Code	4RU9	4RUA	4RUC

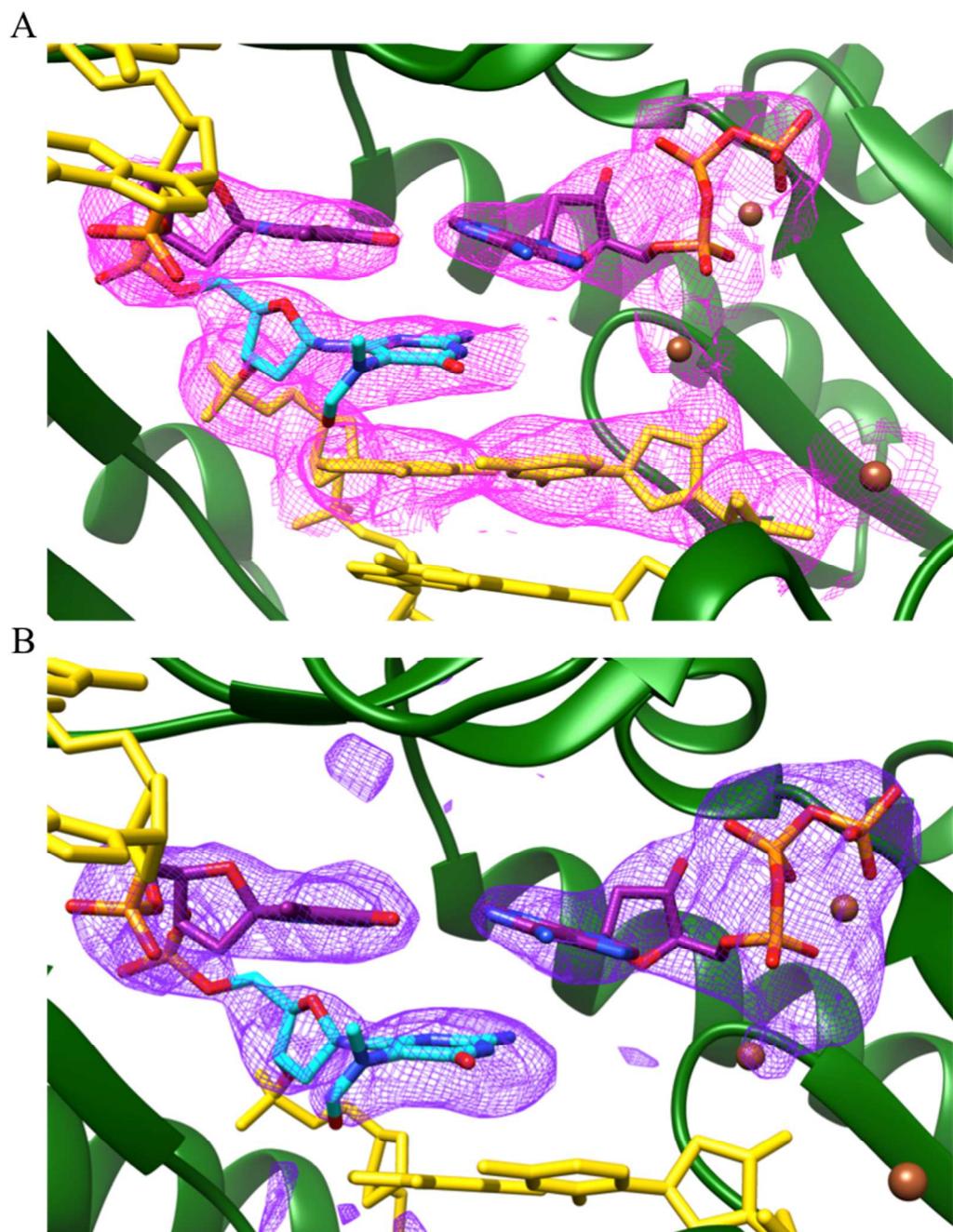
<sup>a</sup> Statistics for the highest-resolution shell are shown in parentheses.



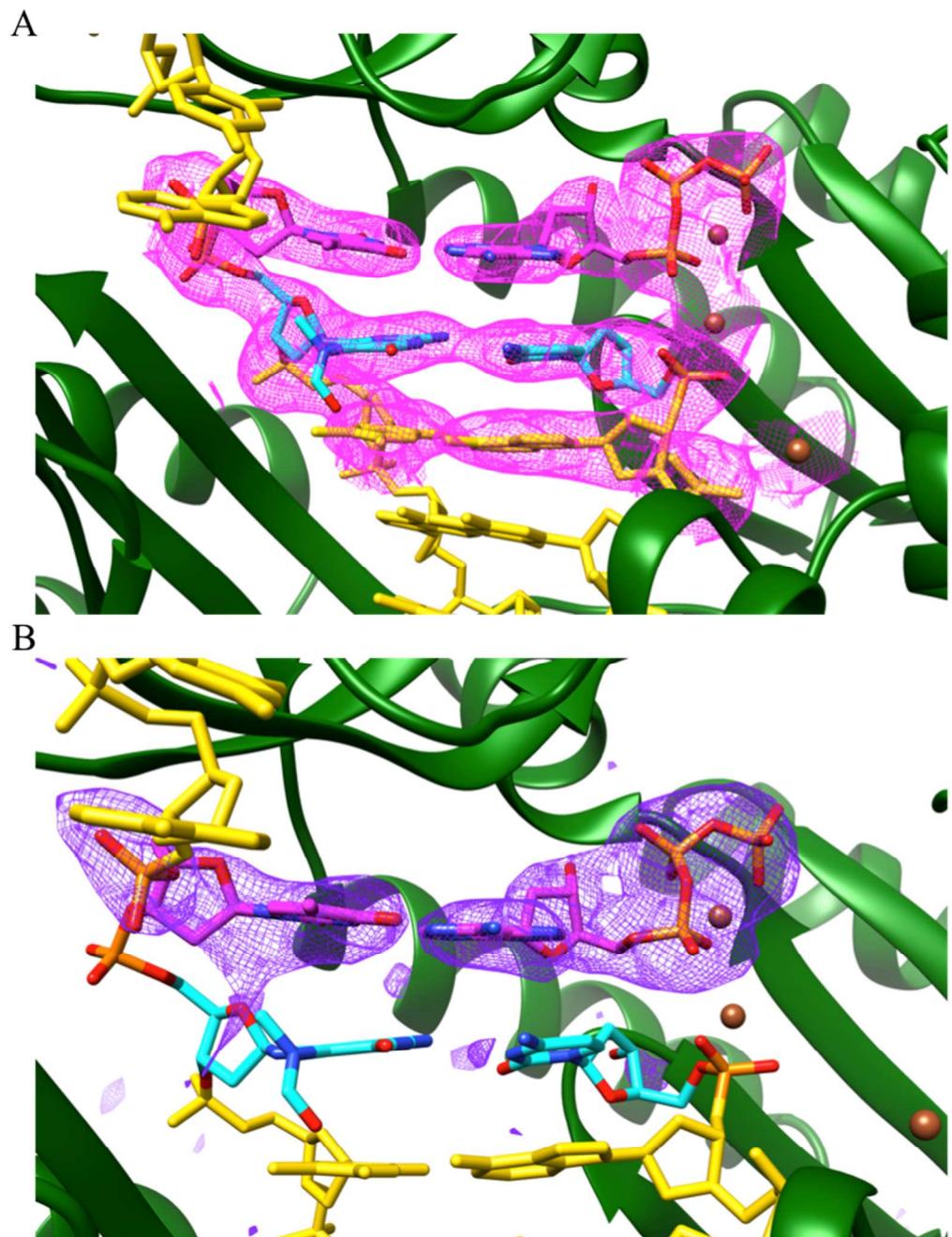
**Figure S1.** Electron density at the active site. (A) Quality of the final Fourier 2Fo-Fc sum electron density (1 $\sigma$  threshold) in the active site region of the ternary complex between hPol  $\eta$ , MeFapy-dG adducted template-primer duplex and dCMPNPP. (B) Omit ( $F_0 - F_c$ ) electron density map drawn at 3 $\sigma$  threshold around the template MeFapy-dG and the incoming dCMPNPP.



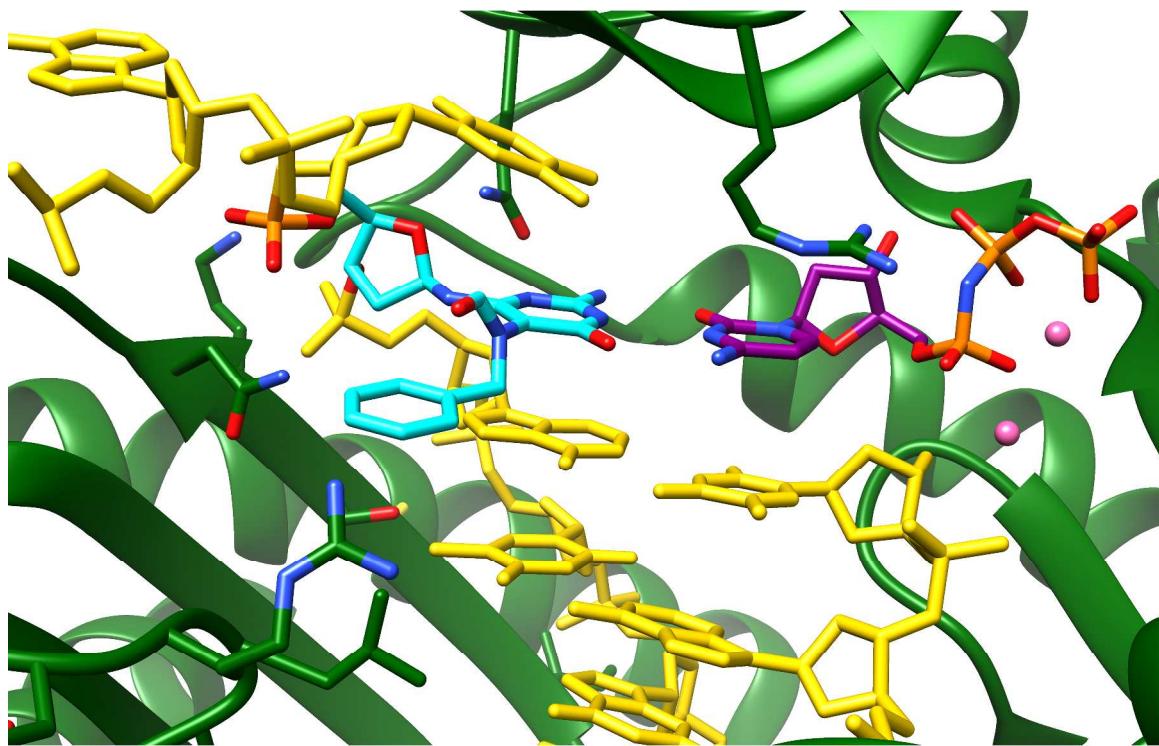
**Figure S2.** Overlay of the active sites in the hPol  $\eta$  ternary complexes with either MeFapy-dG modified DNA template-primer duplex or unmodified (dG) DNA and incoming dCMPNPP in the presence of Mg<sup>2+</sup>. The view is into the major groove, and the color codes for the MeFapy-dG containing complex match those used in **Figure 2**. The crystal structure of the hPol  $\eta$  ternary complex with native DNA is from reference 8 (PDB ID code 4O3N), with the protein colored in cyan, the incoming dCMPNPP in salmon, and the DNA duplex in orange, except for G which is highlighted in sky blue. The superimposition was performed by aligning the entire protein chains from both complexes, resulting in a root mean square deviation of 0.304 Å between 425 atom pairs.



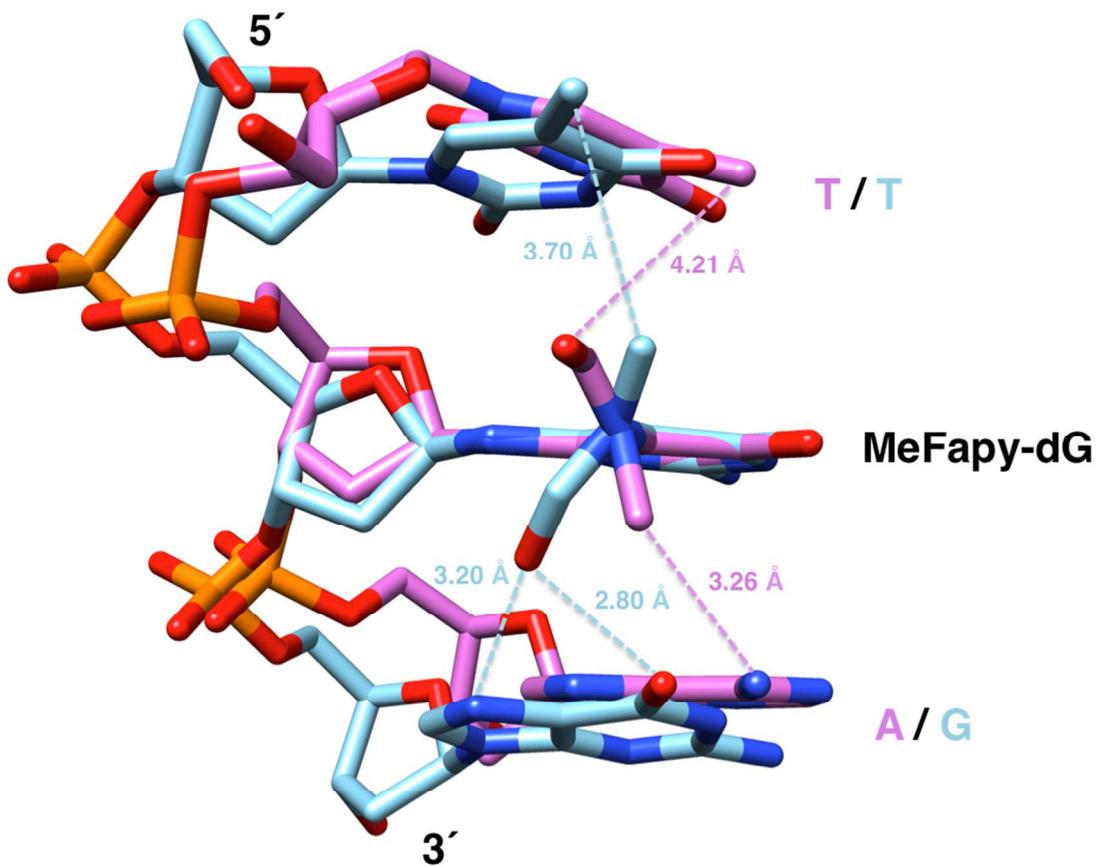
**Figure S3.** Electron density at the active site. (A) Quality of the final Fourier  $2F_o - F_c$  sum electron density ( $1\sigma$  threshold) in the active site region of the insertion complex between Dpo4, MeFapy-dG adducted template-primer duplex and dATP. (B) Omit ( $F_o - F_c$ ) electron density map drawn at the  $3\sigma$  threshold around the template residues MeFapy-dG and dT, and the incoming dATP.



**Figure S4.** Electron density at the active site. (A) Quality of the final Fourier  $2F_o - F_c$  sum electron density ( $1\sigma$  threshold) in the active site region of the extension complex between Dpo4, MeFapy-dG adducted template-primer duplex and dATP. (B) Omit ( $F_o - F_c$ ) electron density map drawn at the  $3\sigma$  threshold around the template dT and the incoming dATP.



**Figure S5.** Model of the active site configuration in the ternary hPol  $\eta$  insertion-step complex with dCMPNPP opposite BzFapy-dG (cyan carbon atoms). The benzyl moiety was modeled in place of the methyl group seen in the crystal structure of the complex with MeFapy-dG, protrudes into the major groove of the template-primer duplex, and can be accommodated without clashing with amino acid side chains in its vicinity.



**Figure S6.** Overlay of the DNA template nucleotides 5'-T[MeFapy-dG]A-3' from the hPol  $\eta$  (pink carbon atoms) and 5'-T[MeFapy-dG]G-3' from the Dpo4 (light blue carbon atoms) insertion stage complexes. Selected interatomic distances involving the formamide moiety are indicated with dashed lines.

The formamide moiety in an isolated MeFapy-dG nucleoside or nucleotide is expected to be conformationally quite flexible and able to adopt a range of orientations relative to the pyrimidine ring. In both insertion stage complexes, formamide is rotated out of the ring plane. In the hPol  $\eta$  complex it adopts an orientation that is more or less perpendicular to the ring and the C=O group is directed toward the 5'-adjacent T. Tight stacking between the MeFapy-dG six-membered ring and the 5'-T in the Dpo4 complex precludes an orientation of the formamide group similar to that in the hPol  $\eta$  complex and the C=O group points toward the 3'-adjacent G.

Alternatively, it is possible that the side chain of R332 from the Dpo4 little finger domain (visible in Figures 3A and 4A on the left-hand side) plays a role in influencing the

orientation of the formamide moiety and fixes it such that the C=O group preferentially points toward the 3'-adjacent template nucleotide. Previous crystal structures of Dpo4 in complex with templates that contained 8-oxoG revealed that the R332 guanidino moiety forms a H-bond with O8, thus stabilizing the *anti* conformation of 8-oxoG and enabling the Pol to efficiently and correctly bypass this major oxidative lesion.<sup>12,13</sup> In the structures of the Dpo4 complexes with MeFapy-dG described here, the distances between the guanidino moiety of R332 and the formamide oxygen exceed 5 Å. The electron density is also indicative of a considerable conformational flexibility of the arginine side chain. However, it is possible that a water molecule could mediate an interaction between R332 and formamide and that the limited resolution of the structures prevented visualization of many solvent molecules. Such an interaction involving water might be at the origin of the particular orientation of the formamide group relative to the pyrimidine ring observed in the structures of the Dpo4 complexes, i.e. C=O pointing in the direction of the 3'-adjacent nucleotide and toward R332.

## ACKNOWLEDGMENTS

Vanderbilt University is a member institution of the Life Sciences Collaborative Access Team at sector 21 of the Advanced Photon Source, Argonne, IL. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357.

## REFERENCES

1. Bierbaum, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W. *Nature* **2010**, *465*, 1044-1048.
2. Zang, H.; Goodenough, A. K.; Choi, J. Y.; Irimia, A.; Loukachevitch, L. V.; Kozekov, I. D.; Angel, K. C.; Rizzo, C. J.; Egli, M.; Guengerich, F. P. *J. Biol. Chem.* **2005**, *289*, 29750-29764.
3. Christov, P. P.; Brown, K. L.; Kozekov, I. D.; Stone, M. P.; Harris, T. M.; Rizzo, C. J. *Chem. Res. Toxicol.* **2008**, *21*, 2324-2333.
4. Jancarik, J.; Kim, S.-H. *J. Appl. Cryst.* **1991**, *24*, 409-411.
5. Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326.
6. Vagin, A.; Teplyakov, A. *J. Appl. Cryst.* **1997**, *30*, 1022-1025.
7. The CCP4 suite: programs for protein crystallography. *Acta Cryst. Section D, Biol. Cryst.* **1994**, *50*, 760-763.
8. Patra, A.; Nagy, L. D.; Zhang, Q.; Su, Y.; Muller, L.; Guengerich, F. P.; Egli, M. *J. Biol. Chem.* **2014**, *289*, 16867-16882.
9. Vagin, A. A.; Steiner, R. S.; Lebedev, A. A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N. *Acta Cryst. Section D, Biol. Cryst.* **2004**, *60*, 2284-2295.
10. Emsley, P.; Cowtan, K. *Acta Cryst. Section D, Biol. Cryst.* **2004**, *60*, 2126-2132.
11. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605-1612.
12. Zang, H.; Irimia, A.; Choi, J.-Y.; Angel, K. C.; Loukachevitch, L. V.; Egli, M.; Guengerich, F. P. *J. Biol. Chem.* **2006**, *281*, 2358-2372.
13. Eoff, R. L.; Irimia, A.; Angel, K. C.; Egli, M.; Guengerich, F. P. *J. Biol. Chem.* **2007**, *282*, 19831-19843.