Structure and Activity of Y-class DNA Polymerase DPO4 from *Sulfolobus solfataricus* with Templates Containing the Hydrophobic Thymine Analog 2,4-Difluorotoluene*[S]

Adriana Irimia[1]**, Robert L. Eoff**[1]**, Pradeep S. Pallan, F. Peter Guengerich, and Martin Egli**[2]

From the Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146

The 2,4-difluorotoluene (DFT) analog of thymine has been used extensively to probe the relative importance of shape and hydrogen bonding for correct nucleotide insertion by DNA polymerases. As far as high fidelity (A-class) polymerases are concerned, shape is considered by some as key to incorporation of A(T) opposite T(A) and G(C) opposite C(G). We have carried out a detailed kinetic analysis of *in vitro* primer extension opposite DFT-containing templates by the trans-lesion (Y-class) DNA polymerase Dpo4 from *Sulfolobus solfataricus*. Although full-length product formation was observed, steady-state kinetic data show that dATP insertion opposite DFT is greatly inhibited relative to insertion opposite T (~5,000-fold). No products were observed in the pre-steady-state. Furthermore, it is noteworthy that Dpo4 strongly prefers dATP opposite DFT over dGTP (~200-fold) and that the polymerase is able to extend an A:DFT but not a G:DFT pair. We present crystal structures of Dpo4 in complex with DNA duplexes containing the DFT analog, the first for any DNA polymerase. In the structures, template-DFT is either positioned opposite primer-A or -G at the −1 site or is unopposed by a primer base and followed by a dGTP:A mismatch pair at the active site, representative of a −1 frameshift. The three structures provide insight into the discrimination by Dpo4 between dATP and dGTP opposite DFT and its inability to extend beyond a G:DFT pair. Although hydrogen bonding is clearly important for error-free replication by this Y-class DNA polymerase, our work demonstrates that Dpo4 also relies on shape and electrostatics to distinguish between correct and incorrect incoming nucleotide.

Recent research on *in vitro* primer extension reactions catalyzed by a range of DNA polymerases and using the hydrophobic T isostere 2,4-difluorotoluene (DFT)[3] (Fig. 1) and other analogs with substituents of increasing size at the 2- and 4-positions of the aromatic moiety appear to support different mechanisms of nucleotide insertion by high fidelity (A-class) and trans-lesion (Y-class) DNA polymerases (reviewed in Refs. 1 and 2). Thus, accurate replication by A-class polymerases may be more dependent on a close steric match between the active site and the shape of a Watson-Crick base pair ("active site tightness") than on the formation of complementary hydrogen bonds between incoming nucleotide base and template residue (3–6). Conversely, their more open active sites may render Y-class polymerases (6–10) less sensitive to changes in base pair dimensions but more dependent on formation of hydrogen bonds between nucleotide pairs at the replicative position (6–9). A-class DNA polymerases whose activities were assessed using apolar T analogs to date include *Escherichia coli* pol I (3–5) and the polymerase from phage T7 (6), and the tested Y-class DNA polymerases consisting of yeast pol η (7), human pol κ (8), and the Dbh (DinB homolog (5)) and Dpo4 (9) polymerases from *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, respectively.

Careful review of the accumulated data regarding A-class polymerases (*i.e. E. coli* DNA pol I Klenow fragment) also reveals that in contrast to the relatively modest reduction in efficiency for inserting dATP opposite DFT, the efficiencies of extension after pairs comprising the hydrophobic analog are drastically inhibited (5). Moreover, there is a significant asymmetry in the efficiency of the extension reaction, an ~30-fold reduction for insertion of dATP opposite template DFT and ~900-fold reduction for insertion of d(DFT)TP opposite template A (relative to the corresponding processes with T and dTTP, respectively (5)). The latter loss is comparable with the effects on insertion reactions involving the DFT analog seen with Y-class polymerases (5, 7–9). Together, these observations appear inconsistent with the conclusion that A- and Y-class polymerases use different mechanisms of replication and that the former may rely chiefly on shape for efficient and correct nucleotide insertion (1, 2). An important limitation of the large body of work involving the use of the hydrophobic T isostere DFT for probing enzyme

---

[3] The abbreviations used are: DFT, 2,4-difluorotoluene; Dpo4, DNA polymerase IV; CID, collision-induced dissociation; DTT, dithiothreitol; ESI, electrospray ionization; LC, liquid chromatography; MS, mass spectrometry; MS/MS, tandem mass spectrometry; O^6-MeG, O^6-methyl-G; WC, Watson-Crick; BSA, bovine serum albumin; PDB, Protein Data Bank; pol, polymerase; NSLS, National Synchrotron Light Source; nt, nucleotide.

FIGURE 1. **Nucleotide structures.** Structures of thymidine (*left*) and the 2′-deoxyribo-2,4-difluorotoluyl nucleoside analog (DFT, *right*).

mechanism and the role of hydrogen bonding in DNA replication is constituted by the fact that the analog has never been visualized at the active site of any DNA polymerase.

The Dpo4 trans-lesion DNA polymerase from *S. solfataricus* (10) has been investigated in more detail both in terms of its function (10–14) and structure (15–17) than any other representatives of the Y-class family of polymerases. Numerous crystal structures of Dpo4 in complex with DNA template-primer constructs containing adducted nucleotides have been determined during the last 4 years (18–24). The structural data reveal a versatile enzyme that in some cases defies the view of Y-class polymerases as low processivity and/or fidelity catalysts of DNA replication. For example, Dpo4 was found to synthesize past the 8-oxoG adduct efficiently and with relatively good fidelity, incorporating ≥95% dCTP instead of dATP and exhibiting faster rates for pre-steady-state kinetics of dCTP incorporation opposite 8-oxoG than G (21). As far as the different roles of sterics *versus* hydrogen bonding in the replications catalyzed by A-class and Y-class DNA polymerases are concerned, recent crystal structures of a the high fidelity *Bacillus stearothermophilus* DNA polymerase I large fragment (BF) and Dpo4 in complex with DNAs containing the O^6-methyl-G (O^6-MeG) adduct opposite C and T may be instructive (see Refs. 25 and 22, respectively). At the constrained active site of BF, both the O^6-MeG:C and O^6-MeG:T pairs adopt a Watson-Crick conformation, whereby the cytosine in the former is protonated. This finding is consistent with the preferential incorporation of T over C opposite O^6-MeG by high fidelity polymerases. The Watson-Crick mode for correctly and mispaired combinations was also observed at the −1- and −2-bp positions. On the other hand, the O^6-MeG:C pair at the more spacious active site of Dpo4 adopts a wobble geometry, and analysis of Dpo4-catalyzed extension products reveals that the enzyme accurately bypasses O^6-MeG with C being the major product (>70%) and T and A being the minor species (22). T exhibited multiple conformations opposite O^6-MeG at the Dpo4 active site that most likely include both the Watson-Crick and wobble geometries.

Here we report the results of detailed pre-steady-state and steady-state kinetic analyses as well as LC-MS/MS characterizations (20, 21) of full-length products of Dpo4-catalyzed primer extension reactions opposite template strands containing DFT either at the replicative or −1-bp positions. Our

steady-state kinetic data confirm that the catalytic efficiency of Dpo4-catalyzed insertion opposite DFT was strongly inhibited, consistent with previously published data by others (9). However, it is noteworthy that Dpo4 is ~200-fold more efficient (k_{cat}/K_m) at inserting dATP opposite DFT than dGTP. The pre-steady-state kinetic analysis revealed no product formation for insertion of dATP opposite DFT, and next-base extension reactions were slower following an A:DFT pair and prevented by a G:DFT pair. To interpret the activity data, we determined crystal structures of Dpo4 in complex with primer-template DNA duplexes featuring dGTP opposite DFT at the active site as well as G or A paired with DFT at the −1 position. The structures provide insight into the different efficiencies of Dpo4 for inserting A and G opposite template-DFT and allow a rationalization of the inability of Dpo4 to further extend the primer strand after a G:DFT pair. The structure featuring an A:DFT pair also challenges the usual assumption that this pair virtually matches A:T in shape. Thus, the absence of hydrogen bonds in the former actually leads to significant changes in the local geometry of the template-primer duplex. Although Dpo4 may rely more heavily on hydrogen bonds between incoming nucleotide and template base for error-free replication compared with high fidelity polymerases, sterics clearly play a role in discriminating between A and G as the pairing partner of DFT.

EXPERIMENTAL PROCEDURES

Materials—Dpo4 was expressed in *E. coli* and purified to electrophoretic homogeneity as described previously (20). All unlabeled dNTPs were obtained from Amersham Biosciences, and [γ-^{32}P]ATP was purchased from PerkinElmer Life Sciences. All unmodified oligonucleotides used in this work were synthesized by Integrated DNA Technologies (Coralville, IA). The DFT phosphoramidite was purchased from Glen Research (Sterling, VA), and DFT-modified template strands were synthesized using an ABI 381A oligonucleotide synthesizer on a 1 μM scale following standard solid phase synthesis and purification protocols.

Full-length Extension Assay—A ^{32}P-labeled primer was annealed to template oligonucleotide by heating a 1:1 solution of oligonucleotide to 95 °C for 5 min and then slow cooling to room temperature. The primer was then incubated with Dpo4 and extended in the presence of a mixture of all four dNTPs. Each reaction was initiated by adding dNTP·Mg^{2+} (1 mM of each dNTP and 5 mM MgCl$_2$) solution to a preincubated Dpo4·DNA complex (100 nM Dpo4 and 200 nM DNA). The reaction was carried out at 37 °C in 50 mM Tris-HCl (pH 7.4), 50 mM NaCl, 5 mM DTT, 100 μg ml^{-1} BSA, and 5% (v/v) glycerol. At the indicated time, 5-μl aliquots were quenched with 50 μl of 500 mM EDTA (pH 9.0). The samples were then mixed with 100 μl of a 95% formamide, 20 mM EDTA solution and were separated on a 20% polyacrylamide (w/v), 7 M urea gel.

Steady-state Kinetics—Dpo4-catalyzed single nucleotide incorporation was measured over a range of dNTP concentrations. All reactions were carried out at 37 °C in 50 mM Tris-HCl (pH 7.4) buffer containing 50 mM NaCl, 5.0 mM DTT, 50 μg ml^{-1} BSA, and 5% glycerol (v/v). Dpo4 (10 nM) was preincubated with radiolabeled DNA (100 nM), and the reaction was initiated by adding dNTP·Mg^{2+}. Aliquots were quenched with

500 mM EDTA (pH 9.0) after varying incubation times. Substrate and product DNA were separated by electrophoresis on a 20% polyacrylamide (w/v), 7 M urea gel. The products were then visualized using a PhosphorImager and quantitated using Quantity One™ software (Bio-Rad). The initial portion of the velocity curve was fit to a linear equation in the program GraphPad Prism (GraphPad, San Diego). The resulting velocity was plotted as a function of dNTP concentration and then fit to a hyperbola, correcting for enzyme concentration, to obtain estimates of k_{cat} and $K_{m,dNTP}$ (Table 1).

Transient-state Kinetics—All pre-steady-state experiments were performed using a KinTek RQF-3 model chemical quench-flow apparatus (KinTek Corp., Austin, TX) with 50 mM Tris-HCl buffer (pH 7.4) in the drive syringes. All RQF experiments were carried out at 37 °C in a buffer containing 50 mM Tris-HCl buffer (pH 7.4) containing 50 mM NaCl, 5 mM DTT, 100 μg ml^{-1} BSA, and 5% (v/v) glycerol. Polymerase catalysis was stopped by the addition of 500 mM EDTA (pH 9.0). Substrate and product DNA was separated by electrophoresis on a 20% polyacrylamide (w/v), 7 M urea gel. The products were then visualized using a PhosphorImager and quantitated using Quantity One™ software. Results obtained under single-turnover conditions were fit to Equation 1,

$$y = A(1 - e^{-k_{obs}t})$$ (Eq. 1)

where A is product formed in first binding event; k_{obs} is rate constant defining polymerization under the conditions used for the experiment being analyzed, and t indicates time.

LC-MS/MS Analysis of Oligonucleotide Products from Dpo4 Reactions—Dpo4 (5 μM) was preincubated with primer-template DNA (10 μM), and the reaction was initiated by addition of dNTP (1 mM each) and MgCl$_2$ (5 mM) in a final volume of 100 μl. Dpo4 catalysis was allowed to proceed at 37 °C for 4 h in 50 mM Tris-HCl buffer (pH 7.8 at 25 °C) containing 50 mM NaCl, 1 mM DTT, 50 μg ml^{-1} BSA, and 5% glycerol (v/v). The reaction was terminated by extraction of the remaining dNTPs using a size-exclusion chromatography column (Bio-Spin 6 chromatography column, Bio-Rad). Concentrated stocks of Tris-HCl, DTT, and EDTA were added to restore the concentrations to 50, 5, and 1 mM, respectively. Next, *E. coli* uracil DNA glycosylase (20 units, Sigma) was added, and the solution was incubated at 37 °C for 6 h to hydrolyze the uracil residue on the extended primer. The reaction mixture was then heated at 95 °C for 1 h in the presence of 0.25 M piperidine, followed by removal of the solvent by centrifugation under vacuum. The dried sample was resuspended in 100 μl of H$_2$O for MS analysis.

LC-MS/MS analysis (20, 21) was performed on a Waters Acquity UPLC system (Waters) connected to a Finnigan LTQ mass spectrometer (Fisher), operating in the ESI negative ion mode (Table 2). An Acquity UPLC BEH octadecylsilane (C$_{18}$) column (1.7 μm, 1.0 × 100 mm) was used with the following LC conditions: Buffer A contained 10 mM NH$_4$CH$_3$CO$_2$ plus 2% CH$_3$CN (v/v), and Buffer B contained 10 mM NH$_4$CH$_3$CO$_2$ plus 95% CH$_3$CN (v/v). The following gradient program was used with a flow rate of 150 μl min^{-1}: 0–2.5 min, linear gradient from 100% A to 95% A, 5% B (v/v); 2.5–6.0 min, linear gradient to 75% A, 25% B (v/v); 6–6.5 min, linear gradient to 100% B;

6.5–8.0 min, hold at 100% B; 8.0–9.0 min, linear gradient to 100% A; 9.0–12.0 min, hold at 100% A. The temperature of the column was maintained at 50 °C. Samples were injected with an autosampler system. ESI conditions were as follows: source voltage 4 kV; source current 100 μA; auxiliary gas flow rate setting 20; sweep gas flow rate setting 5; sheath gas flow setting 34; capillary voltage −49 V; capillary temperature 350 °C; tube lens voltage −90 V. MS/MS conditions were as follows: normalized collision energy 35%; activation Q 0.250; activation time 30 ms. The doubly (negatively) charged species were generally used for CID analysis. The calculations of the CID fragmentations of oligonucleotide sequences were done using a program linked to the Mass Spectrometry Group of Medicinal Chemistry at the University of Utah. The nomenclature used in supplemental Tables S1–S6 has been described previously (26).

Dpo4 Pyrophosphorolysis Activity—The effect of DFT upon the pyrophosphorolysis activity of Dpo4 was tested by incubating Dpo4 (100 nM) with a radiolabeled DNA substrate (100 nM) in 50 mM Tris-HCl buffer (pH 7.4) containing 50 mM NaCl, 10.0 mM DTT, 200 μg ml^{-1} BSA, and 2.5% glycerol (v/v) at 60 °C. Two DNA templates were used, a control 18-mer template containing T and the 18-mer containing DFT. A 13-mer primer containing A at the 3′ terminus was annealed to the template DNA so that it paired with T or DFT. Pyrophosphorolysis was initiated by addition of sodium pyrophosphate (pH 7.4; 500 μM final concentration). Aliquots were quenched with 20 mM EDTA (pH 9.0) in 95% formamide (v/v) after varying incubation times. Substrate and product DNA were separated and imaged as described for the steady-state assays. The initial portion of the velocity curve was fit to a linear equation in the program GraphPad Prism.

Crystallization and X-ray Diffraction Data Collection—The 18-nt template 5′-TTCAG(DFT)AGTCCTTCCCCC-3′ was annealed with the 13-nt primer 5′-GGGGGAAGGACT*X*-3′, with *X* either G in the Dpo4(DFT:13G) complex or A in the Dpo4(DFT:13A) complex. For the insertion complex Dpo4(dGTP), the 18-nt template 5′-TTCA(DFT)TAGTCCT-TCCCCC-3′ was annealed with the 13-nt primer 5′-GGGG-GAAGGACTA-3′. The Dpo4 protein was mixed with the DNA duplex in a buffer containing 60 mM NaCl, 4% glycerol, 20 mM Tris (pH 7.4), 5 mM CaCl$_2$, and 1 mM d(d)NTP (either dGTP or ddCTP, see Table 3). Droplets consisted of a 1:1 mixture of the protein·DNA complex and reservoir solutions. In the case of the Dpo4(DFT:13A) complex crystals were obtained by equilibrating droplets against reservoir solution containing 12% polyethylene glycol 3350, 0.2 M ammonium acetate, 0.1 M calcium acetate, and 20 mM Tris (pH 7.5). Crystals of the Dpo4(DFT:13G) and Dpo4(dGTP) complexes were grown from droplets equilibrated against reservoir solutions containing 16–24% polyethylene glycol 3350, 0.1 M Ca(OAc)$_2$, and 20 mM Tris (pH 7.5).

X-ray diffraction data for the Dpo4(DFT:13A) and Dpo4(DFT:13G) complex crystals were collected at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) on the 5-ID (DND-CAT) and 22-ID (SER-CAT), beam lines, respectively. Diffraction data for the Dpo4(dGTP) complex crystal were acquired on the X25 beam line at the National Synchrotron Light Source (NSLS, Brookhaven National Laboratory, Upton, NY). All data were recorded using

a wavelength of 1.0 Å at 110 K from cryo-protected crystals. The crystals diffracted to between 2.8 and 3.0 Å resolution (Table 3). Individual frames were indexed and scaled with the programs XDS (27) (Dpo4(DFT:13A) or HKL2000 (28) ((Dpo4(DFT:13G) and Dpo4(dGTP)). The Dpo4(DFT:13A) and Dpo4(DFT:13G) complex crystals belong to space group $P2_1$, and the Dpo4(dGTP) crystal belongs to space group $P2_12_12$. Various CCP4 package programs (29), including TRUNCATE (30), were used for further processing of the data. X-ray diffraction data collection and processing statistics are listed in Table 3.

Crystal Structure Determination and Refinement—The three structures were solved by molecular replacement. The solutions for the Dpo4(DFT:13A) and Dpo4(DFT:13G) structures were obtained with Phaser (31) using the Dpo4 complex with the Protein Data Bank (PDB) accession code 2bq3 (20) minus the solvent molecules as a search model. The structure with the PDB accession code 2uvv (24) minus solvent molecules and dGTP was used as the starting model for Dpo4(dGTP). The initial positions of the models were optimized by several rounds of rigid body refinement while gradually increasing the resolution of the diffraction data. Rebuilding of the models was performed with the program TURBO-FRODO (42), and refinements were carried out with the program CNS (32) by performing simulated annealing, gradient minimization, and refinement of individual isotropic temperature factors. Selected refinement statistics, including values for R-work and R-free (33) for the current models, are listed in Table 3. The crystallographic figures were prepared with the program PyMOL (43).
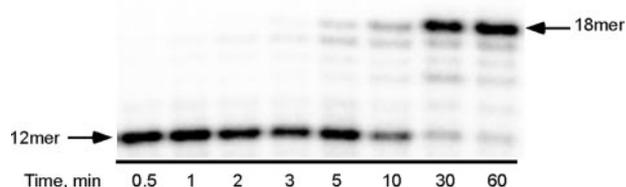
FIGURE 2. **Dpo4-catalyzed polymerization opposite DFT-containing template DNA.** Dpo4-catalyzed (100 nM) extension of 12/18-mer DNA (200 nM) containing DFT was allowed to proceed in the presence of a 4 mM dNTP mix. For comparisons regarding the time frame of full-length extensions opposite native DNA, please see Fig. 2 of Ref. 22 and/or Fig. 1 of Ref. 24.

RESULTS

Extension of Oligonucleotide Primers by Dpo4 in the Presence of All Four dNTPs—Dpo4 catalysis opposite DFT and extension past the modification was allowed to proceed in the presence of all four dNTPs (Fig. 2). Fully extended products begin to appear after ~5 min. These results are in contrast to similar experiments with unmodified DNA in which full-length extension was completed in ~3 min (22, 24), indicating that Dpo4 must undergo multiple rounds of binding and dissociation for product to form when DFT is in the template. Use of a primer containing a uracil at the 3′ terminus does not alter the extension profile. Based on these results, the presence of DFT in the template DNA inhibits catalysis by Dpo4, but full-length extension does occur.

Steady-state Kinetic Analysis of Dpo4 Catalysis Opposite DFT—The relative catalytic efficiency of insertion opposite DFT by Dpo4 was measured by varying the concentration of dNTP in the reaction solution (Table 1). The Dpo4 catalytic efficiency (k_{cat}/K_m) of insertion for dATP was strongly inhibited (~5,300-fold) by the presence of DFT in the template DNA. However, Dpo4 remains 6-, 15-, and 189-fold more efficient at incorporation of A opposite DFT compared with incorporation of C, T, or G opposite the modified base. Steady-state experiments were performed with a primer containing uracil at the 3′ terminus to determine whether the results obtained during LC-MS/MS analysis of the extension products were altered by the presence of uracil in the primer. The overall trend in the steady-state parameters is unchanged by replacing thymine with uracil (Table 1).

Transient-state Kinetic Analysis of Dpo4 Catalysis Opposite DFT—Pre-steady-state analysis of Dpo4-catalyzed insertion of dATP opposite DFT was performed. No product was observed in the pre-steady-state (Fig. 3A), even at high (2.5 mM) concentrations of dATP, consistent with the perturbations observed in the full-length extension time course and in the steady-state kinetic parameters.

Next-base extension of both A:DFT and G:DFT pairings was tested. Incorporation of the next correct nucleotide proceeded at a moderately slower rate for the A:DFT pair (Fig. 3B, *circles*)

TABLE 1
Steady-state kinetic parameters for 1-base incorporation by Dpo4

Oligomer pair	Primer-template pair	dNTP	k_{cat}	$K_{m,dNTP}$	ΔEfficiency relative to dCTP:G
			min^{-1}	μM	
13-mer		dCTP	34.8 ± 0.6^a	3.0 ± 0.2^a	
18-mer	-G				
12-mer		dATP	1.8 ± 0.10	816 ± 78	5,300-Fold less
18-mer	-DFT				
12-mer		dTTP	0.19 ± 0.04	1320 ± 560	80,000-Fold less
18-mer	-DFT				
12-mer		dCTP	0.15 ± 0.01	440 ± 96	34,000-Fold less
18-mer	-DFT				
12-mer		dGTP	0.010 ± 0.002	870 ± 430	10^6-Fold less
18-mer	-DFT				
12-U		dATP	1.8 ± 0.1	745 ± 31	4,800-Fold less
18-mer	-DFT				
12-U		dTTP	0.10 ± 0.01	890 ± 200	103,000-Fold less
18-mer	-DFT				
12-U		dCTP	0.27 ± 0.02	550 ± 95	24,000-Fold less
18-mer	-DFT				
12-U		dGTP	0.015 ± 0.003	780 ± 340	604,000-Fold less
18-mer	-DFT				

^{a} Data are from Ref. 22.

FIGURE 3. **Pre-steady-state analysis of Dpo4-catalyzed nucleotide incorporation and next-base extension of DFT-modified DNA.** *A*, Dpo4-catalyzed (200 nM) extension of 12/18-mer template DNA (100 nM) containing DFT modification was allowed to proceed in the presence of 2.5 mM dATP. *B*, Dpo4-catalyzed (200 nM) extension of 13-A/18-mer (●) and 13-G/18-mer (■) template DNA (100 nM) containing DFT modifications was allowed to proceed in the presence of 2.5 mM dCTP. The results for extension of the 13-A/18-mer were fit to a single exponential equation to yield the following kinetic parameters: $A = 50 \pm 3$ nM, $k_{obs} = 0.33 \pm 0.06$ s^{-1}. The results for 13-G/18-mer (■) were fit to a linear equation to give a steady-state velocity of 0.15 ± 0.01 nM s^{-1}.

relative to what has been observed previously with Dpo4 catalysis opposite unmodified DNA (21). The slower rate indicates that residual inhibition occurs after Dpo4 has inserted dATP opposite DFT. No product was observed for extension of the G:DFT pair (Fig. 3*B*, *squares*).

LC-MS/MS Analysis of Full-length Extension Products—Unambiguous identification of full-length extension products resulting from Dpo4 catalysis was carried out using an LC-MS/MS approach described previously (20). Six ions were observed in the MS spectrum that corresponded to extension products (Fig. 4, Table 2, and supplemental Figs. S2–S6). Selected ion traces are shown in Fig. 4*C*, and CID analysis of the *m*/*z* 939 ion resulted in the fragmentation pattern shown in Fig. 4*D*. The major ions in the fragmentation pattern are consistent with the sequence 5′-pACTGAA-3′, which corresponds to the insertion of A opposite DFT and accurate full-length extension of the primer (supplemental Table S1).

Dpo4-catalyzed incorporation of all four dNTPs was detected during LC-MS/MS analysis of the extension products with incorporation of A comprising ~74% of the total. Each of these products was followed by error-free extension. Additionally, a −1 deletion product was observed in which Dpo4 skips DFT and then extends the primer in an error-free manner. The

−1 deletion product represents ~10% of the total products observed in the reaction, which indicates that Dpo4 is just as likely to skip DFT completely as it is to "misincorporate" C, T, or G opposite the modified base. The trend observed in the LC-MS/MS analysis (Table 2) is consistent with what is observed in the steady-state parameters (Table 1), but it is worth noting that the steady-state analysis could not analyze the proclivity of Dpo4 to skip the modified base and generate the −1 deletion products.

Based on the presence of −1 frameshift products in the LC-MS results, it is possible that the steady-state assays could be measuring a dCTP incorporation event in which Dpo4 skips over DFT and pairs dCTP with the G to the 5′-side of the lesion. However, it is noteworthy that the k_{cat} values for dTTP and dCTP incorporation are quite similar (Table 1). Previous work with the etheno-G modification, where skipping the lesion is a predominant product, showed a much reduced k_{cat} value for Dpo4-catalyzed incorporation of dATP opposite etheno-G even when the next template base was T (20). Based on this comparison, it seems likely that the steady-state values obtained with dCTP largely represent incorporation opposite DFT, but "base skipping" in the steady-state measurements cannot be completely ruled out.

DFT Inhibits the Pyrophosphorolysis Activity of Dpo4—The ability of Dpo4 to reverse the polymerization reaction through pyrophosphorolysis is well documented and apparently unique among the Y-family polymerases studied thus far (12, 13). We tested the effect of DFT upon the pyrophosphorolysis activity of Dpo4 (supplemental Fig. S7). Interestingly, DFT strongly inhibits the pyrophosphorolysis activity of Dpo4. The rate of pyrophosphorolysis for A:DFT is decreased 11-fold relative to the A:T control reaction. The strong inhibition of both the forward and reverse reactions indicates an important role for hydrogen bonds in the stabilization of nascent base pairs in the Dpo4 active site.

Crystal Structures of Dpo4-DNA Complexes with DFT-modified Template Strands—To gain a better understanding of the kinetic data obtained from Dpo4-catalyzed nucleotide insertion and extension reactions opposite and following DFT, respectively, we determined three x-ray structures of the enzyme in complex with DNA duplexes of which the template strands contained a single DFT residue. The primers and templates were designed so as to place a 13th primer-G or -A nucleotide opposite template-DFT at the −1 position (post-insertion complexes Dpo4(DFT:13A) and Dpo4(DFT:13G)) or dGTP opposite DFT at the active site (insertion complex Dpo4(dGTP)).

Crystals of the Dpo4(DFT:13A) and Dpo4(DFT:13G) complexes belong to space group $P2_1$. Although they feature different cell constants, the crystallographic asymmetric unit of both encompasses two complex molecules, whereby one of them displays clear electron density for all DNA nucleotides (Fig. 5, *A* and *C*, and Table 3). For the second complex, all 13 primer nucleotides could be located in the electron density maps but only 16 of the 18 template nucleotides (Fig. 5*D* and Table 3). In the Dpo4(DFT:13G) complexes the electron density around the sugar and base portions of the 12th primer nucleotides is relatively poorly defined, and the DNA duplex exhibits local distor-
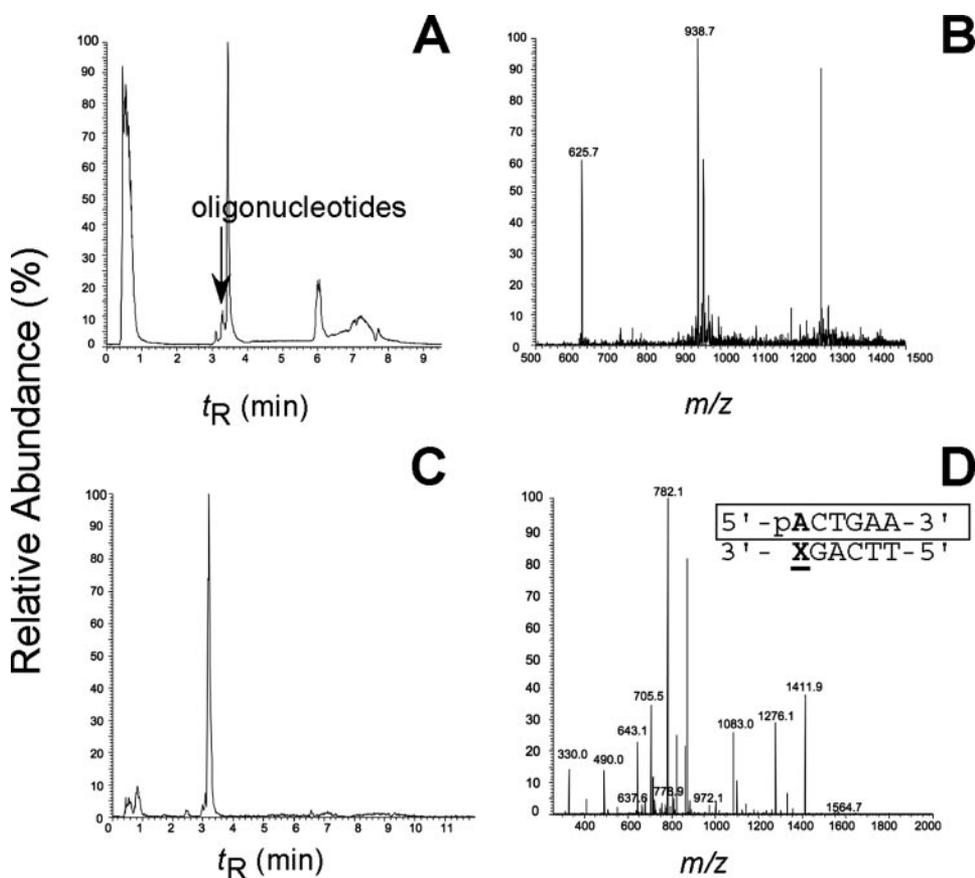
FIGURE 4. **LC-MS analysis of Dpo4-catalyzed full-length extension products.** *A,* total ion current trace of products derived from Dpo4-catalyzed extension of 12/18-mer DNA containing DFT. *B,* ESI mass spectrum of the oligonucleotide peaks that elute at 3.2 min. *C,* total ion current trace of ion *m/z* 939. *D,* CID mass spectrum of ion *m/z* 939. *X,* denotes DFT. This product contained A inserted opposite DFT and extended in an error-free manner.

TABLE 2

Results of LC-MS analysis of Dpo4-catalyzed full-length extension products

Product	% of total
5'-p**A**CTGAA-3'	64
5'-p**A**CTGAAA-3'	10
5'-p**C**CTGAA-3'	8
5'-p**T**CTGAA-3'	5
5'-p**G**CTGAA-3'	3
5'-pCTGAA-3'	10

tions (Fig. 5, *C* and *D*). However, the electron density maps were of sufficient quality to locate the ddCTPs in both Dpo4(DFT: 13A) and Dpo4(DFT:13G) complexes. Peaks corresponding to three Ca^{2+} ions were observed for Dpo4(DFT:13A) (Fig. 5*A*), but density for only two Ca^{2+} ions was present in the Dpo4(DFT:13G) structures (Fig. 5, *C* and *D*).

Crystals of the Dpo4(dGTP) complex belong to space group $P2_12_12$ with a single protein·DNA complex per asymmetric unit (Table 3). The $3F_o - 2F_c$ electron density map shows 16 template nucleotides, 13 primer nucleotides, 1 dGTP, and 3 Ca^{2+} ions (Fig. 5*B*). Two template nucleotides at the 5'-end were disordered and missing from the structure.

Pairing Modes of the DFT Analog and Effects of DFT on the Dpo4 Active Site Configuration—Dpo4(DFT:13A) is a so-called type I complex (using the convention of Ling *et al.* (15)). Only one template nucleotide (tG5) is accommodated in the active

site cleft and is paired with ddCTP (Fig. 5*A*). The tDFT6 residue faces the 3'-terminal primer nucleotide pA13 in a Watson-Crick like fashion, but the fluorine and nitrogen atoms are too far away from each other (>5.0 Å; Fig. 6, *A* and *B*) to interact in a stabilizing fashion. Thus, this arrangement differs markedly from the one in the complex between Dpo4 with the native DNA duplex of identical sequence (Fig. 7*A*) (PDB accession code 1s9f (16)). The root mean square deviation between the two independent complexes in the Dpo4(DFT:13A) structure amounts to less than 0.5 Å and the conformation of the template-primer duplexes, and the geometry of the tDFT:pA pairs (Fig. 6*B*) differs only minimally. Therefore, we will limit the discussion of the Dpo4(DFT:13A) structure to only one of the two complex molecules. In Dpo4(DFT:13A), the ddCTP phosphates adopt a chair-like configuration (12). The γ-phosphate is anchored by hydrogen bonds to Asp-7 (via Ca^{2+}), Tyr-48, Arg-51, and Lys-159. The β-phosphate is stabilized by interactions with Arg-51, Asp-7, and Asp-105 (via Ca^{2+}). The ddCTP α-phos-

phate is not interacting with any of the Dpo4 residues, and its position is moved by ~1.6 Å away from the position it occupies in the Dpo4(DFT:13G) complexes (Fig. 7, *D* and *E*) and in the Dpo4(dGTP) complex. It is also noteworthy that the A site Ca^{2+} ion is displaced by ~4.0 Å from the positions it occupies in the Dpo4 complex with native DNA (PDB accession code 2agq (12)) (Fig. 7*C*) or in the Dpo4(DFT:13G) (Fig. 7, *D* and *E*) and Dpo4(dGTP) complexes. The sugar 3'-OH of the pA13 nucleotide interacts with Ser-103 and Glu-106, and this last residue also stabilizes the phosphate group of the pA13 nucleotide (Fig. 6*A*).

Dpo4(dGTP) has a type II complex configuration of the active site (15) with two template nucleotides (tA4 and tDFT5) simultaneously accommodated inside the catalytic pocket. The tDFT5 residue is left unpaired, and the incoming dGTP makes a distorted wobble base pair with tA4 (Fig. 6, *C* and *D*). The dGTP phosphates adopt a chair-like configuration with the γ-phosphate interacting with Asp-7 (via Ca^{2+}), Tyr-48, Arg-51, and Lys-159. The β-phosphate forms hydrogen bonds to Arg-51 and Asp-7 and Asp-105 (via Ca^{2+}), and the α-phosphate interacts with Asp-7 and Asp-105 (via Ca^{2+}). The Dpo4(dGTP) complex is representative of a −1 frameshift; and compared with the Dpo4(DFT:13A) and Dpo4(DFT:13G) complexes, the primer is translocated toward the −1 position, and the space left between dGTP and the 3'-terminal nucleotide of
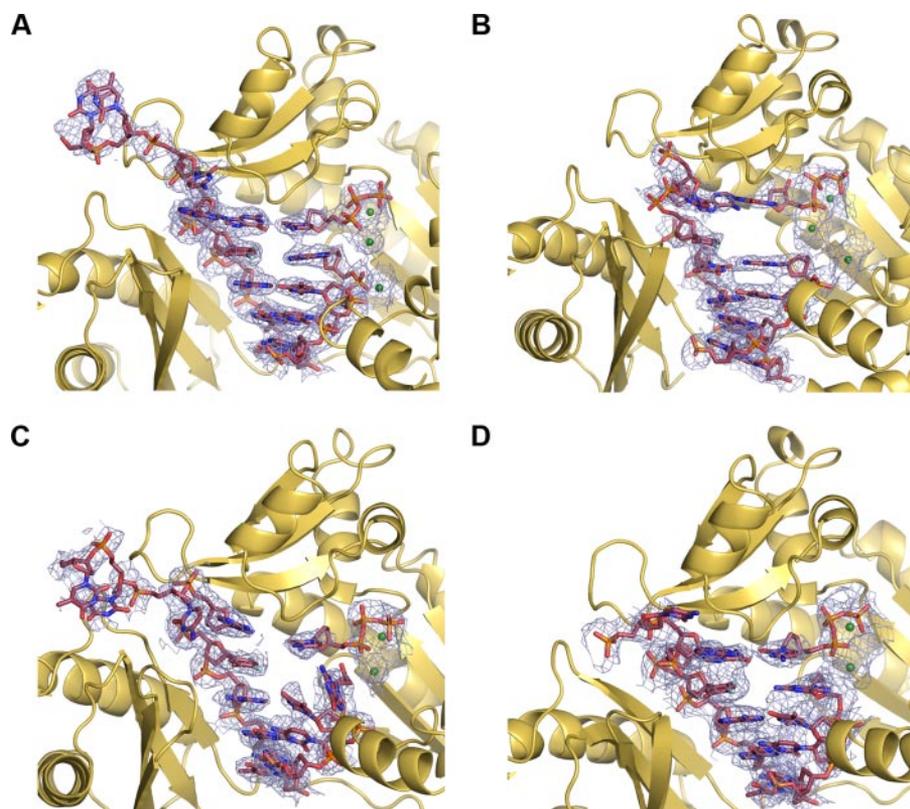
FIGURE 5. **Examples of the quality of the final electron density.** Electron density around the DNA duplex at the active site is shown. *A,* Dpo4(DFT:13A); template 5′-TTCAG(**DFT**)AGTCCTTCCCCC-3′ and primer 5′-GGGG-GAAGGACT**A**-3′, ddCTP. *B,* Dpo4(dGTP); template 5′-TTCA(**DFT**)TAGTCCTTCCCCC-3′ and primer 5′-GGGG-GAAGGACT**A**-3′, dGTP. *C,* one of the Dpo4(DFT:13G) complexes per asymmetric unit; *D,* second Dpo4(DFT:13G) complex; template 5′-TTCAG(**DFT**)AGTCCTTCCCCC-3′ and primer 5′-GGGGGAAGGACT**G**-3′, ddCTP. The $3F_o − 2F_c$ electron density maps (*blue chicken wire*) are contoured at the 1σ level. The Dpo4 protein is shown as *ribbons* and the DNA duplex as *sticks.* The Ca^{2+} ions are depicted as *green spheres.*

the primer is wider. Therefore, Ser-103, Asp-105, and Glu-106 are too far removed to interact with the pA13 nucleotide.

The Dpo4(DFT:13G) complex has type I active site configuration with the tG5 nucleotide pairing with ddCTP (Fig. 6, *E* and *G*). Base pairs at −1, −2, and −3 positions relative to the active site (tDFT5:pG13, tA6:pT12, and tG7:pC11) exhibit a distorted geometry in both complexes per asymmetric unit, and the DNAs in the two complexes also adopt different conformations (Fig. 5, *C* and *D*, respectively). The 13th primer nucleotide pG13 is positioned far from tDFT5 (Fig. 6, *F* and *H*). One of the complexes shows an opening of the tDFT5:pG13 pair with guanine oriented in a different plane than the difluorotoluyl moiety (Fig. 6*F*). Here the nitrogen (G) and fluorine (DFT) atoms are located at >7 Å from each other. In addition to the aforementioned opening of the aromatic portions, there are shifts and altered torsion angles in the sugar-phosphate backbone of the pG13 nucleotide compared with pA13 in the Dpo4(DFT:13A) complex (Fig. 7*D*).

In the second Dpo4(DFT:13G) complex per asymmetric unit, the aromatic moieties of tDFT5 and pG13 are virtually coplanar, but their separation (>4.5 Å) still exceeds the regular hydrogen bonding distance (~2.8 Å) by far (Fig. 6*H*). The position of the sugar of pG13 is much closer to that observed for pA13 in the Dpo4(DFT:13A) complex, although the position of the pG13 phosphate group is still shifted by ~2.5 Å relative to

that of pA13 (Fig. 7*E*). In the Dpo4(DFT:13G) complexes, ddCTP phosphates assume a chair-like conformation similar to the situation in the other structures present here. The γ-phosphate is kept in place by hydrogen bonding interactions with Asp-7 (via Ca^{2+}), Tyr-48, Arg-51, and Lys-159. Arg-51 and Asp-105 (via Ca^{2+}) interact with the β-phosphate. Asp-7 and Asp-105 (via Ca^{2+}) stabilize the α-phosphate in both complexes per asymmetric unit, whereas Glu-106 shows two configurations (Fig. 6, *E* and *G*) and, via Ca^{2+}, interacts with the α-phosphate of ddCTP only in one complex (Fig. 6*G*). None of the Dpo4 residues interact with the 3′-OH group of pG13 in the complex where this 3′-terminal primer nucleotide is flipped (Fig. 6*E*). On the other hand, Ser-103 contacts the pG13 3′-OH group in the second complex (Fig. 6*G*), similar to what was seen with Dpo4(DFT:13A) (Fig. 6*A*).

DISCUSSION

Catalytic differences have been observed between so-called "replicative" and trans-lesion polymerases, but a mechanistic understanding of these differences is only now becoming apparent. A hydrophobic T isostere, DFT, which lacks significant hydrogen bonding capability, was used to probe the relative importance of hydrogen bonds to Dpo4 catalytic efficiency and fidelity. Both the pre-steady-state and steady-state kinetic data reported here show that replacement of T by DFT in the template strand seriously impairs single-nucleotide insertion by Dpo4. In the steady state the efficiency (k_{cat}/K_m) of inserting A opposite DFT is 5,300-fold lower than that for insertion opposite T. A previous study with Dpo4 found that insertion of A opposite DFT is inhibited ~200-fold (9). The difference in the observed level of inhibition is largely because of differences in the measured K_m values and does not alter the general conclusions derived from the steady-state data. No product from single-A insertion experiments opposite DFT was seen in the pre-steady-state (Fig. 3*A*), in line with the slow down in full-length product formation (Fig. 2; compare with Fig. 6 in Ref. 10 or Fig. 4 in Ref. 13), despite the use of a high concentration of dNTPs (4 mM). Fidelity in the steady state is also affected by the change from T to DFT; the relative efficiencies or nucleotide misincorporation ratios ((k_{cat}/K_m)$_{incorrect}$/(k_{cat}/K_m)$_{correct}$) with DFT are significantly diminished at least in the cases of dTTP and dCTP (6.6×10^{-2}, T; 1.6×10^{-1}, C; 5.3×10^{-3}, G; see Table 1) compared with T (5.3×10^{-4}, T; 7.2×10^{-4}, C; 1.3×10^{-3}, G; see Ref. 9). Interestingly, the relative efficien-

TABLE 3

Selected crystal data, data collection, and refinement parameters

Parameter	Dpo4(DFT:13A)	Dpo4(dGTP)	Dpo4(DFT:13G)
Type of complex	Postinsertion	Insertion	Postinsertion
Crystal data, data collection			
X-ray source	APS(DND-CAT)	NSLS	APS(SER-CAT)
Beamline	ID-5	X29	ID-22
Detector	MARCCD	Quantum CCD	MARCCD
Wavelength (Å)	1.00	1.00	1.00
Temperature (K)	110	110	110
No. of crystals	1	1	1
Space group	$P2_1$	$P2_12_12$	$P2_1$
Unit cell (a, b, and c; Å)	52.14,101.87,111.06	94.62,103.56,52.64	52.7,186.74,52.73
(α, β, and γ; °)	90,94.9,90	90,90,90	90,110.1,90
Resolution range (Å)	50.0–2.8	50.0–2.98	50.0–3.0
Highest resolution shell[a]	(2.9-2.8)	(3.09-2.98)	(3.19-3.0)
No. of measurements	113,315 (10,980)	31,795 (1,726)	61,192 (3,171)
No. of unique reflections	27,671 (2,611)	9,628 (863)	16,865 (1,510)
Redundancy	4.1 (4.2)	3.3 (2.0)	3.6 (2.1)
Completeness (%)	96.6 (91.2)	84.5 (81.0)	87.5 (81.0)
R_{merge}[b]	7.0 (40.1)	9.1 (30.0)	14.2 (52.5)
Signal to noise ($\langle I/\sigma I\rangle$)	16.3 (3.42)	14.3 (4.05)	9.2 (2.0)
Solvent content (%)	58.9	54.2	52.05
Refinement			
Model composition (asymmetric unit)			
No. of amino acid residues	342/343	342	343/342
No. of water molecules	216	66	159
No. of Ca^{2+} ions	3/3	3	2/2
No. of template nucleotides	16/18	16	18/16
No. of primer nucleotides	13/13	13	13/13
No. of dGTP		1	
No. of ddCTP	1/1		1/1
R_f[c] (%)	21.1	23.0	21.8
R_{free}[d] (%)	24.9	28.2	29.0
Estimated coordinate error (Å)			
Luzatti plot	0.36	0.38	0.38
Luzatti plot (c-v)	0.44	0.53	0.55
sA plot	0.5	0.53	0.57
σA plot (c-v)	0.6	0.72	0.84
Temperature factors			
Wilson plot (Å^2)	69.8	77.5	75.33
Mean isotropic (Å^2)	51.5	58.9	58.2
Root mean square deviation in temperature factors			
Bonded main chain atoms (Å^2)	1.3	1.3	1.3
Bonded side chain atoms (Å^2)	1.7	1.8	1.6
Root mean square deviation from ideal values			
Bond lengths (Å)	0.007	0.01	0.009
Bond angles (°)	1.1	1.8	1.2
Dihedral angles (°)	22.4	23.0	23.0
Improper angles (°)	1.0	1.3	1.0

[a] Values in parentheses correspond to the highest resolution shells.
[b] $R_{merge} = \Sigma_{hkl} \Sigma_{j=1,N} |\langle I_{hkl}\rangle - I_{hklj}|/\Sigma_{hkl} \Sigma_{j=1,N} |I_{hklj}|$, where the outer sum (*hkl*) is taken over the unique reflections.
[c] $R_f = \Sigma_{hkl} ||F_{o,hkl}| - k|F_{c,hkl}||/\Sigma_{hkl} |F_{o,hkl}|$, where $|F_{o,hkl}|$ and $|F_{c,hkl}|$ are the observed and calculated structure factor amplitudes, respectively.
[d] R_{free} idem, for the set of reflections (5% of the total) omitted from the refinement process.
[e] c-v is cross-validation.

cies of dGTP misincorporation opposite DFT and T differ only by a factor of 4.

Steady-state kinetic data for the high fidelity *E. coli* DNA polymerase I Klenow fragment showed that the efficiency of A insertion opposite DFT was reduced just 50-fold relative to T (4). In the pre-steady state, a similar difference was observed, but d(DFT)TP incorporation opposite template A was 900-fold decreased relative to dTTP (5) (for d(DFT)TP *versus* dTTP insertion steady-state data see Ref. 34). The relatively small reductions in efficiency for incorporation of dATP opposite DFT compared with T seen with both *E. coli* DNA polymerase I (5) and T7 DNA polymerase (6) are remarkable and are clearly different from the effects observed with *S. solfataricus* Dpo4 (>5000-fold, steady state; this work and see Ref. 9) and *S. acidocaldarius* Dbh (1,700-fold, pre-steady state (5)). Two items need to be distinguished at this point to try and infer mechanis-

tic differences between A-family and Y-family polymerases based on the kinetic results. The first item concerns the relationship between hydrogen bonding and catalytic efficiency, and the second item relates to the ability of a given DNA polymerase to selectively incorporate the "correct" dNTP opposite a thymidine isostere with no hydrogen bonding capability. Regarding the concept of efficiency, all Y-family polymerases tested to date (Dpo4, Dbh, pol κ, and pol η) show much more pronounced decreases in efficiency (220- to ~5,000-fold relative to unmodified DNA) when incorporating opposite DFT compared with the A-family polymerases (*i.e.* DNA pol I Klenow fragment ~30-fold). Such a trend may indicate a more important role for hydrogen bonds between DNA nucleobases during Y-family polymerase catalytic progression, which, it should be noted, includes multiple steps and is not easily dissected.
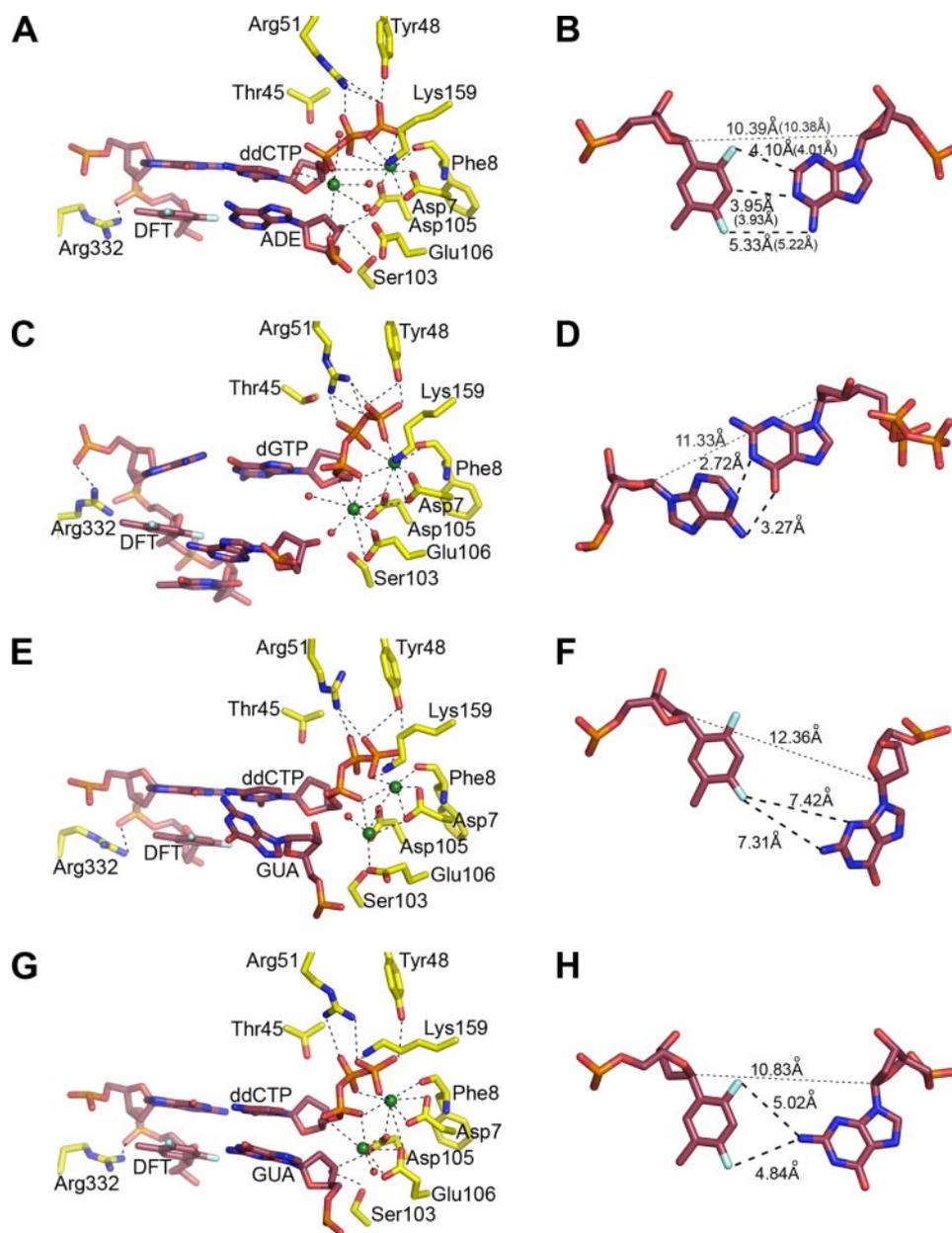
FIGURE 6. **Configuration of the Dpo4 active site and base pairing modes.** Amino acids (carbon atoms colored in *yellow*) interacting with the DNA duplex and d(d)NTP (carbon atoms colored in *brown*) at the active site of Dpo4 are shown in *stick* representation. The Ca^{2+} atoms are depicted as *green spheres*. *A* and *B* correspond to Dpo4(DFT:13A). *C* and *D* correspond to Dpo4(dGTP). *E* and *F* correspond to one of the Dpo4(DFT:13G) complexes per asymmetric unit. *G* and *H* correspond to the second Dpo4(DFT:13G) complex.

tion (in this case, dGTP). Upon inclusion of DFT in the template DNA, Klenow fragment is only ~50-fold more efficient at insertion of dATP opposite DFT than it is at insertion of dTTP opposite DFT. The overall selectivity for Klenow fragment is diminished roughly 36-fold. It is also interesting to note a change in the misincorporation spectrum for Dpo4 and Klenow fragment from A > G > C or T during catalysis opposite T to A > C or T > G during catalysis opposite DFT, indicative of an interaction between DFT and dGTP that is not conducive to polymerization in both systems. The trend concerning nucleoside selectivity is clear in that Dpo4 fidelity is more strongly affected by DFT. However, Dpo4 still retains preference for correct insertion of dATP opposite DFT, which indicates a mechanism that remains similar in some respects to that employed by A-family polymerases.

We used the LC-MS/MS approach previously applied to a dissection of products of Dpo4-catalyzed *in vitro* primer extension reactions opposite adducted templates (20–24) to analyze the composition of products obtained with a DFT-modified 18-mer template (Table 2 and Fig. 4). Accordingly, about 75% of the primers extended to full-length are the result of incorporation of dATP opposite DFT. Furthermore, the data show that a −1 frameshift is about as common as misincorporation of dTTP, dCTP, or dGTP opposite DFT combined. The crystal structure of the Dpo4(dGTP) complex has trapped a frameshifting event, whereby dGTP avoids DFT and instead pairs with the 5′-proximal template A under formation of two hydrogen bonds (Fig. 6, *C* and *D*). Accommodation of the wider purine-purine pair with pseudo-Watson-Crick (WC) geometry is accompanied by severe buckling (particularly of the adenine; Fig. 6*C*) and shearing of bases (Fig. 6*D*). These deformations also affect the positions of the DFT and T residues in the template strand 3′-adjacent to the paired A (see the superimposition of the Dpo4(dGTP) active site with that in the type II complex between Dpo4 and native DNA in supplemental Fig. S8). So-called type II complexes are commonly observed in crystal structures of Dpo4 with native (15, 17) and adducted DNA

Regarding the second item of nucleoside triphosphate selectivity, inserting DFT into the template DNA alters Dpo4 selectivity by making the enzyme only ~6–10-fold better at inserting dATP opposite DFT than it is at inserting dCTP opposite DFT. This is in contrast to the ~750-fold more efficient insertion of dATP opposite T relative to the next most efficient reaction (dGTP) (9). The overall selectivity is therefore decreased ~75–125-fold for Dpo4 in the steady state. It should be stated that pre-steady-state analysis of Dpo4 insertion of dATP opposite unmodified T was ~930-fold more efficient than the next best misincorporation event (dCTP) (11). Klenow fragment, on the other hand, is ~1800-fold more efficient at insertion of dATP opposite T than insertion of the next most efficient reac-

FIGURE 7. **Structural comparison between Dpo4 complexes containing a DFT-modified DNA template strand and Dpo4 in complex with native DNA.** The DNA duplex, d(d)NT(D)P and Dpo4 active site residues are shown as *sticks*. Metal ions are depicted as spheres labeled *A* and *B* for A site and B site, respectively. The Dpo4(DFT:13A) and Dpo4(DFT:13G) complexes are colored by atom with carbon atoms in *green* and *brown*, respectively. Structures of Dpo4 in complex with native DNA are depicted in *orange* (PDB accession code 1s9f (16)) and *pink* (PDB accession code 2agq (12)). The *dashes* show the distance between the metal ions. *A,* superimposition of Dpo4(DFT:13A) (*green*) with the Dpo4·native DNA complex (PDB code 1s9f, *orange*) containing exactly the same DNA primer and (unmodified) template strands and ddCTP (hydrolyzed to ddCDT during crystallization (16)). *B,* superimposition of the Dpo4·native DNA complex (PDB code 1s9f, *orange*) with the other Dpo4·native DNA complex (PDB code 2agq, *pink*), the latter containing G template and C primer base at the −1 position and template T and dATP at the active site. *C,* superimposition of Dpo4(DFT:13A) (*green*) with the Dpo4·native DNA complex (PDB code 2agq, *pink*). In the *right panel*, DNA primer, metal ions, and d(d)NTP are shown after a 90° rotation relative to the *left panel*. *D,* superimposition of Dpo4(DFT:13A) (*green*) with the first Dpo4(DFT:13G) complex (*brown*) per asymmetric unit. *E,* superimposition of Dpo4(DFT:13A) (*green*) with the second Dpo4(DFT:13G) complex (*brown*) per asymmetric unit.

(18–23). We did not pursue crystallizations of ternary Dpo4·DNA·dNTP complexes with either dCTP or dTTP opposite DFT (constituting 8 and 5%, respectively, of the full-length products; see Table 2). However, because of the propensity of type II complex formation in crystals of Dpo4 with DNA and

the observed preference for A over DFT by the incoming dGTP in the Dpo4(dGTP) complex structure, we are assuming that such crystals would likely also have trapped frameshifted states under formation of either A:dCTP or A:dTTP pairs.

In the structure of the Dpo4(DFT:13A) complex with template-DFT opposite primer-A at the post-replicative (−1) site, the two bases are only loosely associated, with distances between atom pairs engaged in hydrogen bonds in a regular T:A pair extended to >5 Å (N-6(A) ⋯ F-4(DFT)) and ∼4 Å (N-1(A) ⋯ C-3(DFT)) (Fig. 6*B*). Interestingly, these increased distances do not translate into a larger spacing between C-1′ atoms across strands (which would amount to a larger diameter of the primer-template duplex). The C-1′ ⋯ C-1′ distance seen in the case of the DFT:A pair is similar to those observed in ternary Dpo4 complexes with native DNA (tG5:ddCTP = 10.79 Å and tT6: pA13 = 10.25 Å, PDB 1sf9 (16), and tT4:dATP = 10.56 Å and tG5: pC13 = 10.62 Å, PDB 2agq (12)). Rather than in a simple stretching of the base pair, the geometric changes triggering a looser association of DFT and A compared with a canonical T:A pair are the result of an opening (toward the major groove; Fig. 6*B*) and an out-of-plane shift by the adenine base (Fig. 6*A*). The geometry of the DFT:A pair seen here is relatively similar to that of ribo-2,4-difluorotoluyl:A (rDFT:A) pairs in a recent crystal structure of an RNA duplex (35), although the distances between N-6(A) and F-4(DFT) do not exceed 4 Å there. Gradually increasing the size of hydrophobic analogs of T (by replacing the O-2 and O-4 atoms with fluorine, chlorine, bromine, or iodine (9)) may not simply lead to an expansion of the corresponding pairs with A along the direction of WC hydrogen bonds (which only exist in the native T:A pair). Instead, rotations (*i.e.* opening, as observed in the Dpo4(DFT:13A) complexes) or translations (along the stacking direction, staggering (Fig. 6*A*), or perpendicular to the direction of WC hydrogen bonds, shearing) of the two base moieties relative to one another are likely sufficient to

relieve potential steric repulsions because of geometric constraints of the polymerase active site. Comparison of the active site geometry in the Dpo4(DFT:13A) complex with the geometries in complexes between Dpo4 and native DNA reveals surprisingly small changes (Fig. 7, *A* and *C*). In particular, there is very little movement of polymerase active site residues as a result of the presence of DFT in the template strand.

The only obvious change apart from slight movements of bases in the DNA template and primer strands and the position of the ddCTP relative to dCDP (Fig. 7*A*) or dATP (Fig. 7*C*) is the shift of the A site Ca^{2+} ion in the Dpo4(DFT:13A) complex (Fig. 7*C*, *right panel*). In the structure of the complex with native DNA used for the superimposition shown in Fig. 7*A*, the dCTP had hydrolyzed to dCDP during crystallization (16), and metal ion A is absent as a result of the missing phosphate group. Except for this dCDP, the structure with PDB accession code 1s9f (16) features template and primer strands with sequences that are identical with those in the Dpo4(DFT:13A) complex. We therefore superimposed the 1s9f complex on the structure of another complex with native DNA (PDB 2agq (12); Fig. 7*B*) to demonstrate that except for the conformations of the d(d)NT(D)P phosphate groups the positions of primer and template nucleotides all match very well. The superimposition of the Dpo4(DFT:13A) and 2agq complexes depicted in Fig. 7*C* then allows for a comparison between the relative positions of the A site Ca^{2+} ion and the α-phosphate group of ddCTP (Dpo4(DFT:13A)) and dATP (PDB 2agq). Thus, the shift of the Ca^{2+} ion is most likely because of the different orientations of ddNTPs and dNTPs at the active site of Dpo4 that go along with deviating positions of phosphate groups. Therefore, it is unlikely that the shift of the metal ion is a consequence of the presence of DFT alone, although slight adaptations in the orientation of the primer strand opposite DFT may contribute to the different position of the metal ion. Interestingly, in the Dpo4(DFT:13G) complexes with ddCTP present at the active sites, the positions of the A site Ca^{2+} ion deviate from the position in the Dpo4(DFT:13A) structure (Fig. 7, *D* and *E*) but are close to that seen in the 2agq structure with dATP.

Compared with the geometry of the DFT:A pair in the Dpo4(DFT:13A) structure, DFT:G pairs in the Dpo4(DFT:13G) complexes show more significant deviations relative to a canonical WC base pair (Fig. 6, *E–H*, and Fig. 7, *D* and *E*). The C-1′···C-1′ distances are once again similar to the distance in native DNA, but the pG13 residues are rotated into the major groove to various degrees (Fig. 6*H*), and in one complex guanine is strongly tilted (Fig. 6*E*). These observations provide further support that lack of hydrogen bonding and mispairing does not result in a stretched geometry of the base pair. Instead the geometric constraints of the polymerase active site push one of the bases into the major groove. In principle it should also be possible to accommodate the base in the minor groove that is relatively open in Dpo4 because of the absence of side chains from the palm and finger domains protruding into that space compared with high fidelity polymerases. However, the crystal structures reveal that G opposite DFT is shifted into the major groove. This adjustment is accompanied by a rotation of the

2′-deoxyribose sugar of the 3′-terminal pG13 nucleotide (Fig. 6, *F* and *H*), which alters the relative orientations of the 3′-hydroxyl group and both the α-phosphate of (d)dNTP and particular Dpo4 residues (Fig. 6, *E* and *G*, and Fig. 7, *D* and *E*). For example, as shown in Fig. 6*G*, the position of the 3′-OH of pG13 is similar to that of the 3′-OH of pA13 in Dpo4(DFT:13A) (Fig. 7*E*) and in the complex of Dpo4 with native DNA (Fig. 7*C*), allowing it to interact with both Ser-103 and Glu-106. On the other hand, these interactions are missing in the first Dpo4(DFT:13G) complex (Fig. 6*E*) (and also in the Dpo4(dGTP) complex (Fig. 6*C*)). Such changes are consistent with the LC-MS/MS data that show only 3% of the full-length product as resulting from dGTP incorporation opposite DFT (Table 2) and the pre-steady-state kinetic data that revealed no extension after a tDFT:pG pair (Fig. 3*B*). Can the structures shed light on the ability of Dpo4 to discriminate between dATP and dGTP opposite DFT (∼200-fold, Table 1)? Addition of an amino group at the 2-position of pA13 in the Dpo4(DFT:13A) complex (Fig. 6*B*) leads to a clash with F-2 of DFT. In fact the position of G opposite DFT in the second Dpo4(DFT:13G) complex (Fig. 6*H*) can be viewed as a direct result of avoiding this clash. This provides support for the notion that Dpo4 relies fully on sterics to discriminate between dATP and dGTP opposite DFT and that the change in shape allows the polymerase to distinguish between nucleotides to a similar degree as that afforded by different hydrogen bonding patterns (in the case of the two purines).

The geometries of DFT:G pairs at the −1 position in the structures of the two Dpo4(DFT:13G) complexes also differ from the *reverse*-wobble mode of a T:G mismatch at the equivalent position in a previous structure of a Dpo4 ternary complex (PDB 1s97 (16)). Superimposition of the Dpo4(DFT:13G) complexes with the complex containing the T:G mismatch (supplemental Fig. S8, *B* and *C*) also reveals different orientations of phosphate groups from ddCTPs, boat-like (12, 16) and chair-like (our structures). Looking at the Dpo4(DFT:13A) and Dpo4(DFT:13G) structures, it appears that in DNA, DFT is a better mimic of T pairing with A than T pairing with G. Surprisingly, we recently found the opposite for rDFT and U in the case of RNA. Whereas the geometries of rDFT:A (35) and DFT:A pairs (this work) are relatively similar and consistent with the absence of hydrogen bonds, rDFT:G pairs virtually match U:G wobble pairs geometrically, and the stability and hydration of the latter indicate that fluorine atoms are acting as weak hydrogen bond acceptors (36). The structural data are also supported by stability data that demonstrate that rDFT:G pairs are less punishing thermodynamically than rDFT:A pairs in RNA duplexes ((this is different for duplex DNA where the loss of stability for DFT:A is smaller than for DFT:G (37)). Moreover, the structural data are in line with studies of enzyme activity that show that an RNA endonuclease tolerates an rDFT:A pair adjacent to the cleavage site but that cleavage is prevented by the presence of a structurally more distorted and rigid rDFT:G wobble pair (36). The finding that the effects of replacement of T (or U) by the 2,4-difluorotolyl nucleoside analog should be so different for DNA and RNA is

surprising, and a rationalization remains elusive at the moment.

An important conclusion from our crystal structures of Dpo4 complexes with DFT incorporated in the template strands is that the enzyme conformational changes in the active site, both among the Dpo4(DFT:13A), Dpo4(DFT:13G), and Dpo4(dGTP) structures, as well as in comparison to earlier structures of complexes with native DNA, are actually very small. Although one needs to keep in mind that the resolutions of the crystallographic data for complexes of Dpo4 are typically in the 2.4–2.8 Å range, and therefore not atomic, it is safe to say that there are no drastic (*i.e.* >1 Å) deviations in the positions of side chains (Fig. 7 and supplemental Fig. S8). Therefore, the structural data do not support the notion of a loose Dpo4 active site (9). Rather, our observations are more indicative of a spacious active site that does, however, provide limits as far as the width of a duplex is concerned. The similar C-1′···C-1′ distances attest to this constraint and, instead of the protein adapting to the increased dimensions of DFT:A or DFT:G pairs by expanding the size of its active site, the DNA itself is forced to change its conformation at the replication fork as a result of steric constraints (different shapes) arising from the lack of hydrogen bonds. Evidently, a position of G in the major groove as seen in one of the Dpo4(DFT:13G) complexes (Fig. 6, *E* and *F*) leads to a loss of base stacking and argues against a flexible active site. The structures also demonstrate that DFT itself may be an isostere of T, but paired with either A or G in the active site of Dpo4, the similarity to T:A or T:G pairs, respectively, is obviously lost. The lack of hydrogen bonding interactions goes along with a significant change in shape and therefore hydrogen bonding and geometry may not be readily separable entities as far as probing the activities of DNA polymerases is concerned.

There are currently no crystal structures available for DFT:A or DFT:G pairs at the active site of a high fidelity (A-class) polymerase or, for that matter, in an isolated DNA duplex. Such structures could shed light on potential differences in the ways that A- and Y-class polymerases deal with the DFT analog and whether the tighter active sites of the former will result in a geometry of the DFT:A pair that resembles the native T:A more closely than does the DFT:A pair in our Dpo4(DFT:13A) structure. However, even in the absence of such structural data, there are indications that the differences between the two classes of polymerases may be less significant than has been assumed based on the work with DFT and other hydrophobic nucleoside analogs. The difference in the magnitude of the changes in catalytic efficiency (when template-T is replaced by DFT) observed between A- and Y-family polymerases could be interpreted as a more significant role played by hydrogen bonds between template and incoming nucleotide at the active sites of the Y-class polymerases. On the other hand, the similar trends observed by varying base pair size on the efficiencies of a high (DNA polymerase I Klenow) and a low fidelity enzyme (Dpo4) (9) argue against fundamentally different mechanisms employed by A- and Y-class polymerases.

What is different between the two classes of polymerases, however, is the hydrogen bonding interaction between amino acid side chains and both bases of DNA base pairs at the repli-

cative and −1 sites from the minor groove side (present in A-class (25, 38, 39) and absent in Y-class enzymes, *i.e.* Dpo4 (15–24)). In fact, the asymmetric loss of efficiency observed with Klenow fragment (large for d(DFT)TP:A and small for dATP:DFT) could be due to different influences of hydrogen bonds from polymerase residues to template and dNTP bases. Thus, loss of the hydrogen bond to DFT in the template strand (O-2(T) → F-2 in the minor groove) would have a relatively small effect. Conversely, the inability of an A-class polymerase to contact the minor groove base edge of the incoming nucleotide could affect dNTP orientation and the rate of the chemical step more severely. Clearly, the asymmetry in the efficiency of the enzyme cannot be explained by the absence of hydrogen bonds between A and DFT. Template base and nucleobase of dNTP at the replicative site of A-class polymerases also exhibit different stacking interactions in that the latter is sandwiched between the base of the 3′-terminal primer nucleotide and a tyrosine (typically the template base lacks a stacking interaction on the 5′-side (25, 38, 39)). This binding mode combined with the fact that DFT is a more efficient stacking system (40) than T might result in a bigger adjustment of the d(DFT)TP position compared with that of DFT in the template strand as a consequence of the loss of the respective interactions with the protein in the minor groove. It is noteworthy in this context that the asymmetric efficiency and fidelity by Y-class polymerases with regard to dCTP incorporation opposite template 8-oxo-7,8-dihydroguanosine (8-oxoG; dCTP ≫ dATP) relative to d(8-oxoG)TP incorporation opposite template A or C (little discrimination between the two) (21, 24, 41) can be rationalized on the basis of different interactions between protein and template 8-oxoG or d(8-oxoG)TP. Thus, crystal structures of Dpo4 in complex with DNA duplexes containing 8-oxoG in the template revealed formation of a hydrogen bond from a residue in the little finger domain (Arg-332) to O-8 of 8-oxoG (21, 24). This hydrogen bond cannot be established to d(8-oxoG)TP and locks 8-oxoG in the *anti* conformation, consistent with the preferred incorporation of dCTP.

In summary, there are several strong arguments based on activity and structural data (crystal structures presented here and of complexes between high fidelity enzymes and native DNA) for Y- and A-class polymerases that argue against a uniqueness of the latter in terms of relying more on (base pair) shape than hydrogen bonds for efficiently and accurately replicating DNA. The most basic among them is that shape and hydrogen bonding are intimately related and that the lack of hydrogen bonds in DFT:A and DFT:G pairs renders their shapes significantly different from those of T:A and T:G. In addition, to explain actual differences between the two classes of polymerases as far as tolerance toward the hydrophobic DFT analog is concerned, undue emphasis may have been placed on hydrogen bonds between template base and incoming nucleotide. It would appear that differential interactions between A-class polymerases and DFT either in the template or the incoming nucleotide triphosphate are more important in this respect. This is because such hydrogen bonds that allow a scanning of the template-primer duplex minor groove are absent at

the more open active sites of Y-class polymerases. As a result, the differences in the geometries of DFT:A pairs in the Dpo4(DFT:13A) complexes and T:A could exceed those between the corresponding pairs at the tighter active sites of A-class polymerases.

REFERENCES

1. Kool, E. T. (2002) *Annu. Rev. Biochem.* **71,** 191–219
2. Kool, E. T., and Sintim, H. O. (2006) *Chem. Comm.* 3665–3675
3. Morales, J. C., and Kool, E. T. (1998) *Nat. Struct. Biol.* **5,** 950–954
4. Kim, T. W., Delaney, J. C., Essigmann, J. M., and Kool, E. T. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15803–15808
5. Potapova, O., Chan, C., DeLucia, A. M., Helquist, S. A., Kool, E. T., Grindley, N. D. F., and Joyce, C. M. (2006) *Biochemistry* **45,** 890–898
6. Kim, T. W., Brieba, L. G., Ellenberger, T., and Kool, E. T. (2006) *J. Biol. Chem.* **281,** 2289–2295
7. Washington, M. T., Helquist, S. A., Kool, E. T., Prakash, L., and Prakash, S. (2003) *Mol. Cell. Biol.* **23,** 5107–5112
8. Wolfle, W. T., Washington, M. T., Kool, E. T., Spratt, T. E., Helquist, S. A., Prakash, L., and Prakash, S. (2005) *Mol. Cell. Biol.* **25,** 7137–7143
9. Mizukami, S., Kim, T. W., Helquist, S. A., and Kool, E. T. (2006) *Biochemistry* **45,** 2772–2778
10. Boudsocq, F., Iwai, S., Hanaoka, F., and Woodgate, R. (2001) *Nucleic Acids Res.* **29,** 4607–4616
11. Fiala, K. A., and Suo, Z. (2004) *Biochemistry* **43,** 2116–2125
12. Vaisman, A., Ling, H., Woodgate, R., and Yang, W. (2005) *EMBO J.* **24,** 2957–2967
13. Irimia, A., Zang, H., Loukachevitch, L. V., Guengerich, F. P., and Egli, M. (2006) *Biochemistry* **45,** 5949–5956
14. Guengerich, F. P. (2006) *Chem. Rev.* **106,** 420–452
15. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) *Cell* **107,** 91–102
16. Trincao, J., Johnson, R. E., Wolfle, W. T., Escalante, C. R., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) *Nat. Struct. Mol. Biol.* **11,** 457–462
17. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2004) *Mol. Cell* **13,** 751–762
18. Ling, H., Boudsocq, F., Plosky, B. S., Woodgate, R., and Yang, W. (2003) *Nature* **424,** 1083–1087
19. Ling, H., Sayer, J. M., Plosky, B. S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D. M., and Yang, W. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2265–2269
20. Zang, H., Goodenough, A. K., Choi, J.-Y., Irimia, A., Loukachevitch, L. V., Kozekov, I. D., Angel, K. C., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2005) *J. Biol. Chem.* **280,** 29750–29764
21. Zang, H., Irimia, A., Choi, J.-Y., Angel, K. C., Loukachevitch, L. V., Egli, M., and Guengerich, F. P. (2006) *J. Biol. Chem.* **281,** 2358–2372
22. Eoff, R. L., Irimia, A., Egli, M., and Guengerich, F. P. (2007) *J. Biol. Chem.* **282,** 1456–1467
23. Eoff, R. L., Angel, K. C., Egli, M., and Guengerich, F. P. (2007) *J. Biol. Chem.* **282,** 13573–13584
24. Eoff, R. L., Irimia, A., Angel, K. C., Egli, M., and Guengerich, F. P. (2007) *J. Biol. Chem.* **282,** 19831–19843
25. Warren, J. J., Forsberg, L. J., and Beese, L. S. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 19701–19706
26. Christian, N. P., Reilly, J. P., Mokler, V. R., Wincott, F. E., and Ellington, A. D. (2001) *J. Am. Soc. Mass. Spec.* **12,** 744–753
27. Kabsch, W. (1993) *J. Appl. Crystallogr.* **26,** 795–800
28. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
29. Collaborative Computing Project Number 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
30. French, S., and Wilson, K. (1978) *Acta Crystallogr. Sect. A* **34,** 517–525
31. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40,** 658–674
32. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
33. Brünger, A. T. (1992) *Nature* **355,** 472–475
34. Moran, S., Ren, R. X. F., and Kool, E. T. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 10506–10511
35. Xia, J., Noronha, A., Toudjarska, I., Li, F., Akinc, A., Braich, R., Rajeev, K. G., Egli, M., and Manoharan, M. (2006) *ACS Chem. Biol.* **1,** 176–183
36. Li, F., Pallan, P. S., Maier, M. A., Rajeev, K. G., Mathieu, S. L., Kreutz, C., Fan, Y., Sanghvi, J., Micura, R., Rozners, E., Manoharan, M., and Egli, M. (2007) *Nucleic Acids Res.* **35,** in press
37. Guckian, K. M., Krugh, T. R., and Kool, E. T. (1998) *Nat. Struct. Biol.* **5,** 954–959
38. Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) *Science* **260,** 352–355
39. Brieba, L. G., Eichman, B. F., Kokoska, R. J., Doublié, S., Kunkel, T. A., and Ellenberger, T. (2004) *EMBO J.* **23,** 3452–3461
40. Guckian, K. M., Schweitzer, B. A., Ren, R. X., Sheils, C., Paris, P. L., Tahmassebi, D. C., and Kool, E. T. (1996) *J. Am. Chem. Soc.* **118,** 8182–8183
41. Einolf, H. J., Schnetz-Boutaud, N., and Guengerich, F. P. (1998) *Biochemistry* **37,** 13300–13312
42. Cambillau, C., and Roussel, A. (1997) *Turbo Frodo*, Version OpenGL.1, Université Aix-Marseille II, Marseille, France
43. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, CA