Mechanism of error-free and semi-targeted mutagenic bypass of an aromatic amine lesion by Y-family polymerase Dpo4

Olga Rechkoblit[1], Alexander Kolbanovskiy[2], Lucy Malinina[1,3], Nicholas E. Geacintov[2], Suse Broyde[4], and Dinshaw J. Patel[1]

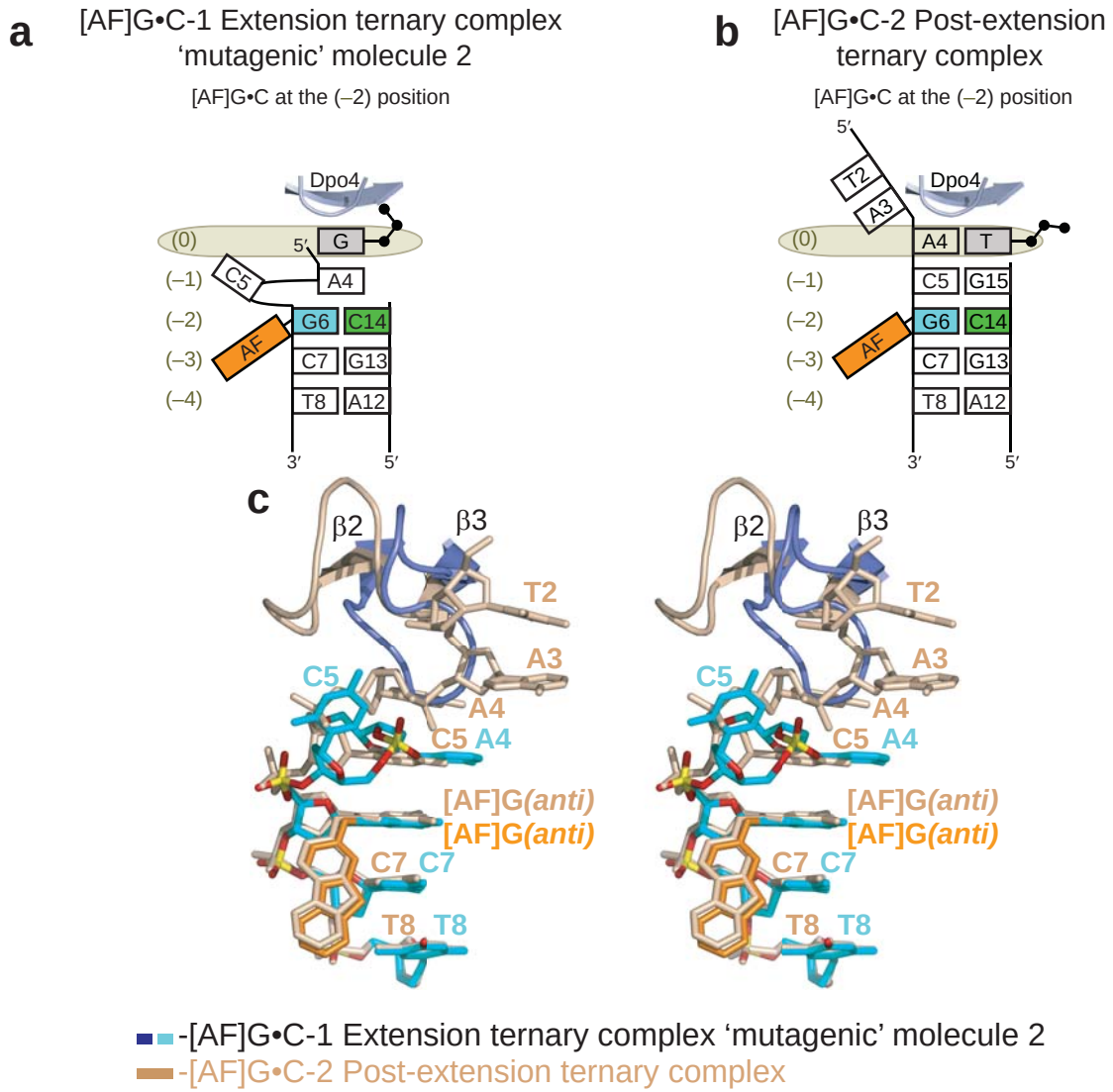
[1]Structural Biology Program
Memorial Sloan-Kettering Cancer Center
New York, NY, 10021

[2]Chemistry Department
New York University
New York, NY, 10003

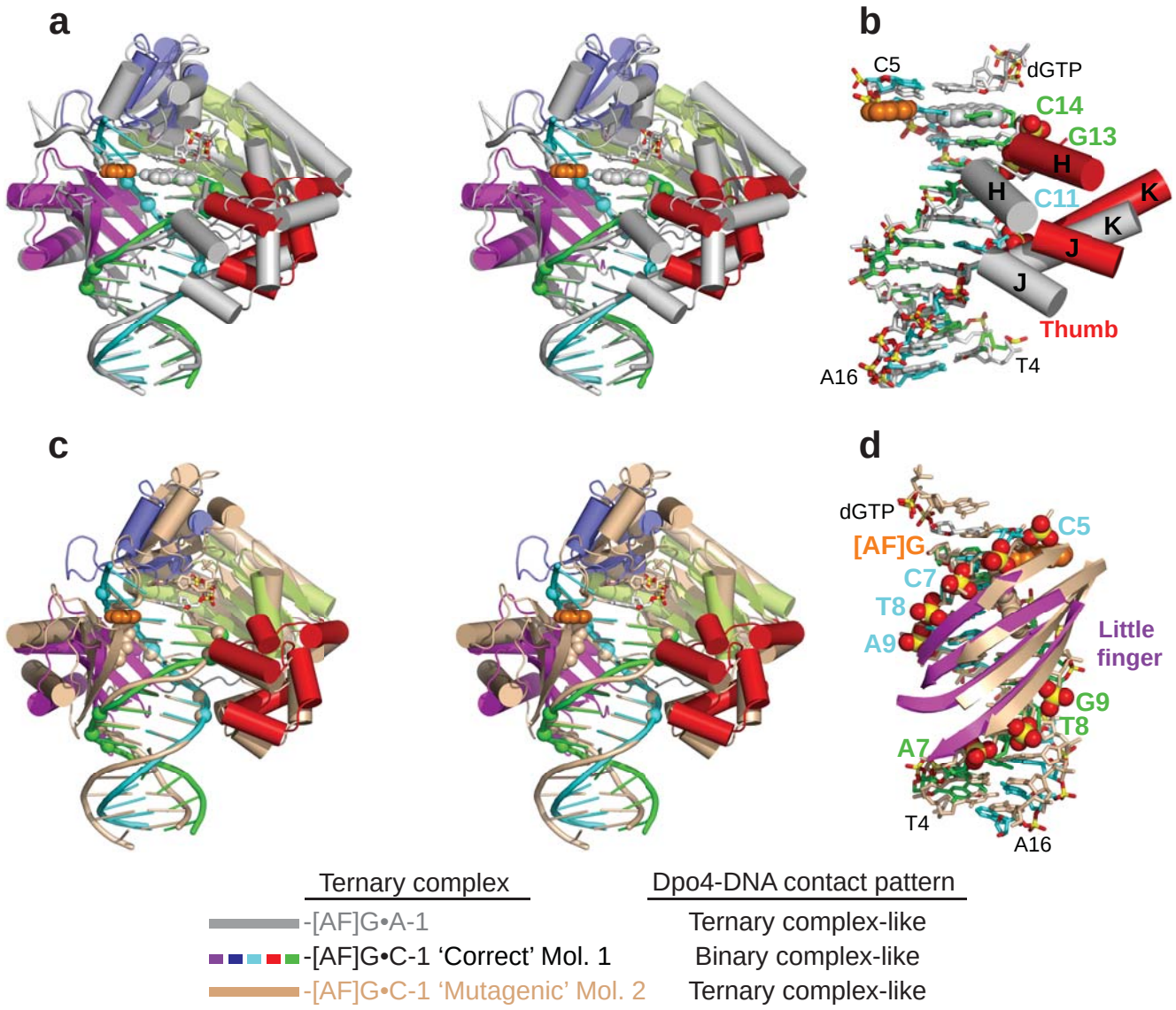
[3]Structural Biology Unit
CIC bioGUNE
Bizkaia Technology Park
48160 Derio, Spain

[4]Biology Department
New York University
New York, NY, 10003

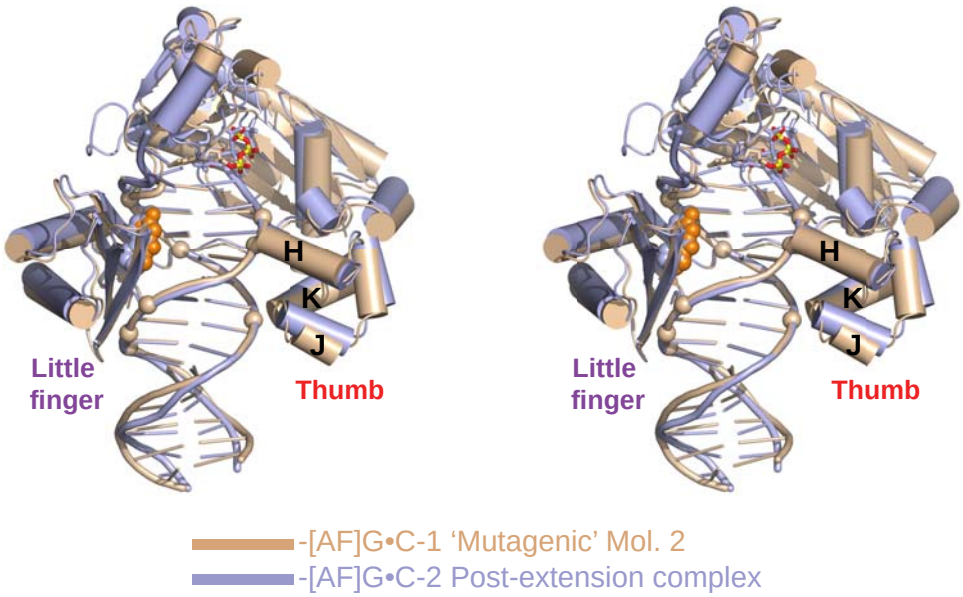
Supplementary Information



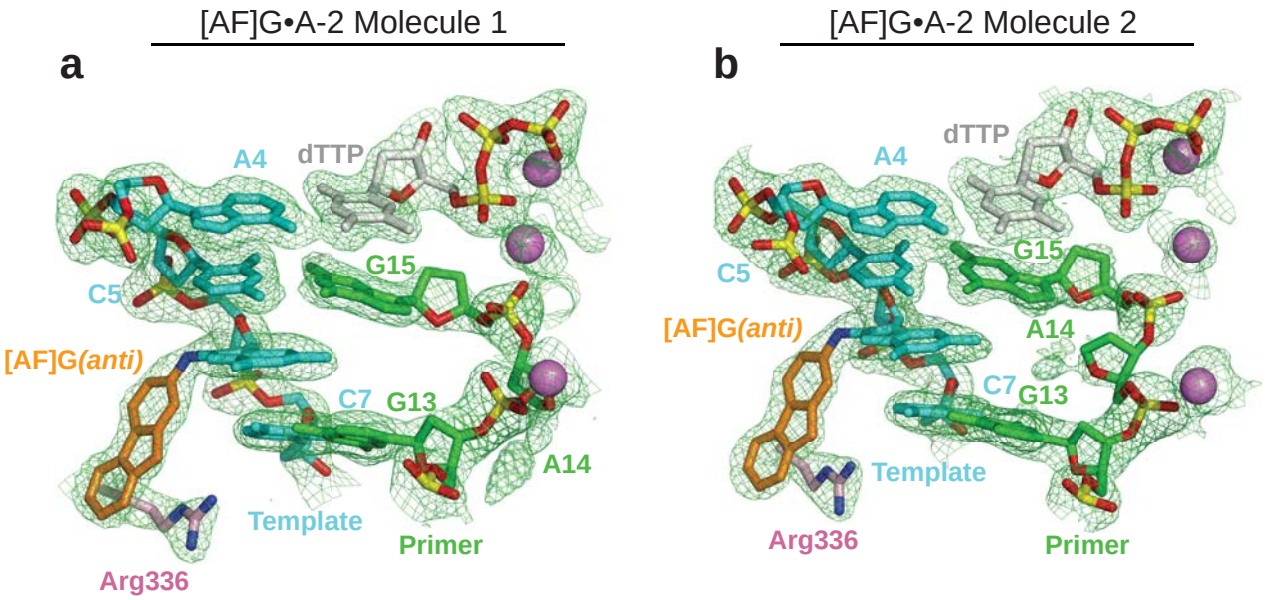
Supplementary Figure 1. Relocated and normal positions of the loop connecting β2 and β3 of the Dpo4 finger domain in molecule 2 of the [AF]G•C-1 extension ternary complex and the [AF]G•C-2 post-extension ternary complex, respectively. (**a**) Schematics of the base pairing arrangement within the Dpo4 active site of molecule 2 of the [AF]G•C-1 extension ternary complex. (**b**) Schematics of the base pairing arrangement within the Dpo4 active site of the [AF]G•C-2 post-extension ternary complex. (**c**) Stereo view comparing different conformations of the loop connecting β2 and β3 of the Dpo4 finger domain in molecule 2 of [AF]G•C-1 (multi-color representation) and [AF]G•C-2 (beige color). The structures are superimposed using the Dpo4-bound portions of the DNA duplexes. In molecule 2 of [AF]G•C-1 the loop connecting β2 and β3 fills the space normally taken by the sugar phosphate backbone of a template base at the (0) position (A4 in [AF]G•C-2) and the next 5′ base of the single stranded template overhang exiting the active site (A3 in [AF]G•C-2). Note similar arrangement of the [AF]G6-T8 segments of the template strand and distinct arrangement of the A4-C5 bases in these complexes. The C1-A3 bases are disordered in molecule 2 of the [AF]G•C-1 complex.



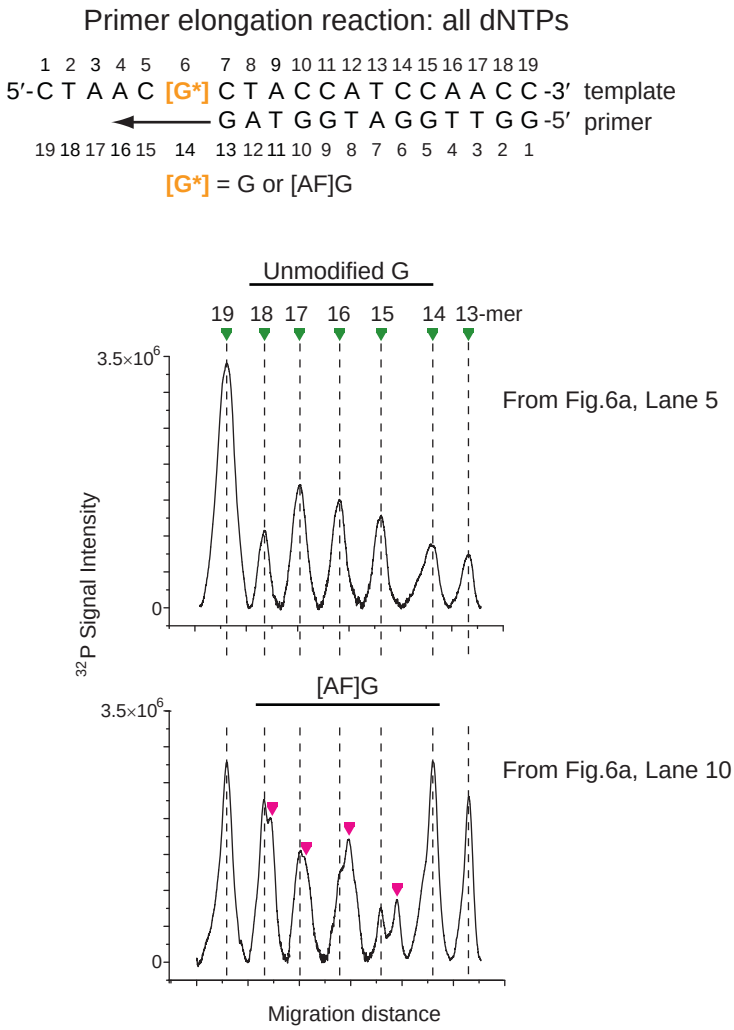
Supplementary Figure 2. Stepwise translocations of the Dpo4 thumb and little finger domains produce 'correct' molecule 1 and 'mutagenic' molecule 2 alignments during extension from the [AF]G*(anti)*•C base pair. (**a**) Stereo views of the translocated position of the thumb domain following superposition [AF]G•A-1 and [AF]G•C-1 molecule 1. The [AF]G•A-1 complex, with ternary complex-like Dpo4/DNA contact pattern, is shown in gray. Molecule 1 of the [AF]G•C-1 complex is shown in color, and has a binary complex-like Dpo4/DNA contact pattern. The complexes are superimposed based on the Dpo4-bound portions of the DNA duplexes. The DNA phosphate groups in contact with Dpo4 are indicated by spheres. (**b**) In molecule 1 of the [AF]G•C-1 complex, the thumb domain interacts with the phosphate groups of the template C11 and primer G13-C14, while in the [AF]G•A-1 complex, the thumb domain interacts with the phosphate groups of template A12 and primer A12-G13. CPK spheres are shown for molecule 1 of the [AF]G•C-1 complex. (**c**) Stereo views of the translocation of the little finger domain following superposition of [AF]G•C-1 molecule 1 and molecule 2 complexes. Molecule 2 of [AF]G•C-1 has ternary complex-like Dpo4/DNA contact patterns and is shown in beige. (**d)** Molecule 1 of the [AF]G•C-1 complex showing the little finger domain contacting the phosphates of template C5-A9 and primer A7-G9, and molecule 2 showing the little finger domain contacting the phosphates of template [AF]G6-T8 (C5 is looped out and A4 phosphate disordered) and primer T8-G10. CPK spheres are shown for molecule 1 of the [AF]G•C-1 complex.



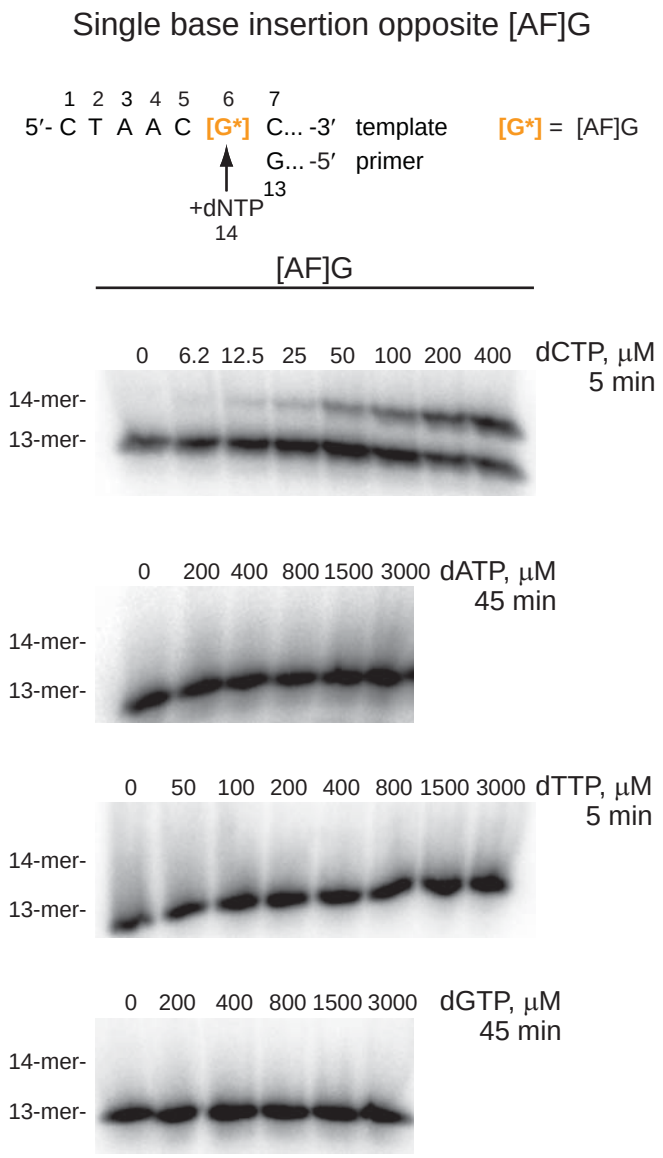
Supplementary Figure 3. Superposition of the 'mutagenic' molecule 2 of the [AF]G•C-1 extension ternary complex with the post-extension [AF]G•C-2 complex. The two complexes were aligned in stereo following superpositioning of their DNA components. The DNA phosphate groups in contact with Dpo4 are indicated by spheres.



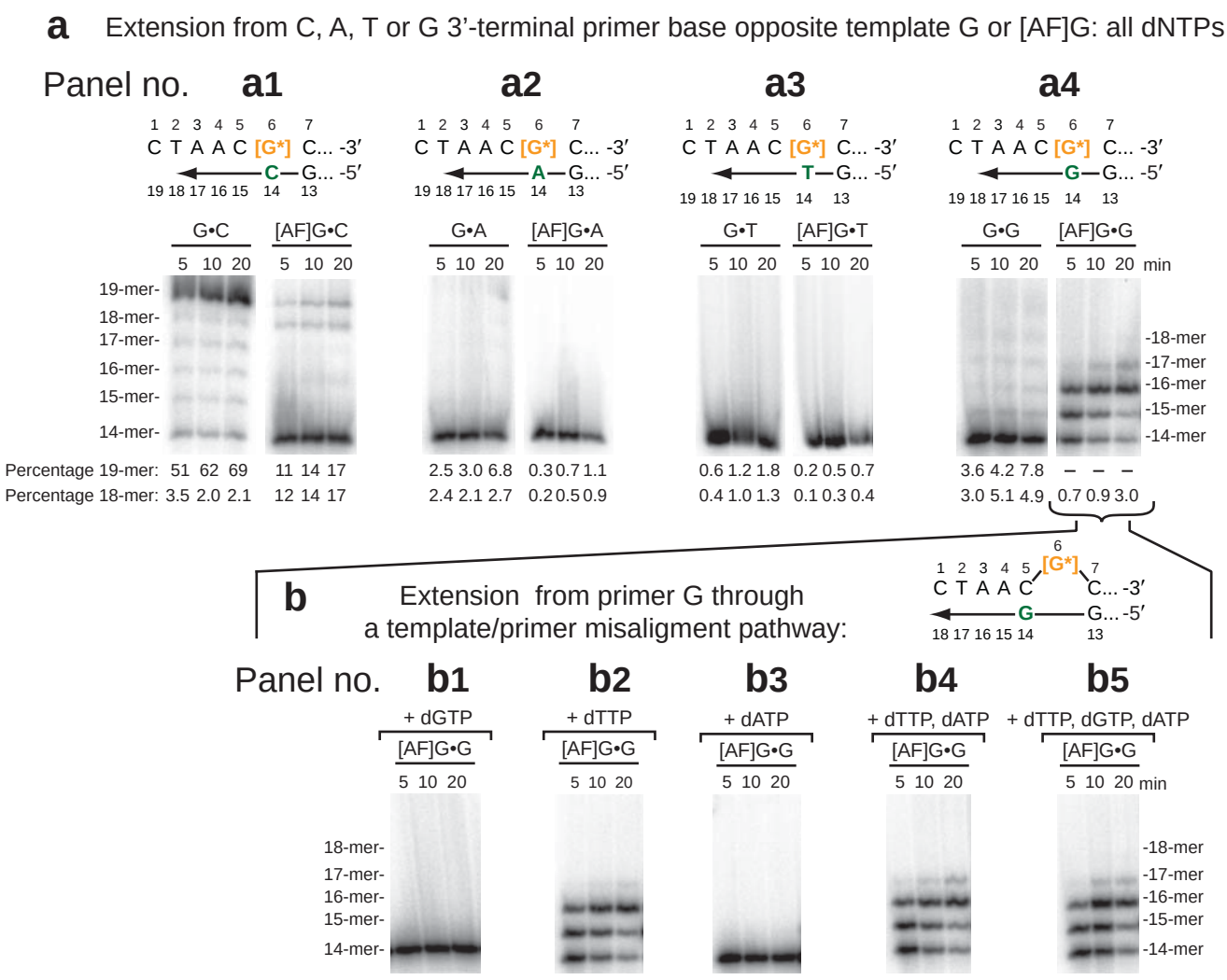
Supplementary Figure 4. 2Fo-Fc electron density map for the template/primer DNA and dTTP at the active site of the [AF]G•A-2 Dpo4 post-extension ternary complex. The [AF]G•A-2 complex has two distinct molecules per asymmetric unit (AU) with different positions of the partner A14 base. 2Fo-Fc electron density map contoured at 1σ level is colored in green (2.10 Å resolution). (**a**) Molecule 1. The [AF]G*(anti)* partner A14 is disordered but appears to be positioned outside the helix on the minor groove side. (**b**) Molecule 2. The [AF]G*(anti)* partner A14 is disordered but appears to be positioned inside the helix.



Supplementary Figure 5. Profile analysis of ^{32}P-signal intensity of time course of 13-mer primer extension on the unmodified G (lane 5 in **Fig. 6a**) and on the [AF]G-modified (lane 10 in **Fig. 6a**) 19-mer templates in the presence of all four dNTPs. The 13-mer primer has the 3′-end positioned one base before the unmodified-G, or [AF]G; the 14-mer extends to the position opposite the G or [AF]G. The green triangles represent the correct products; the magenta triangles represent mutagenic extension.

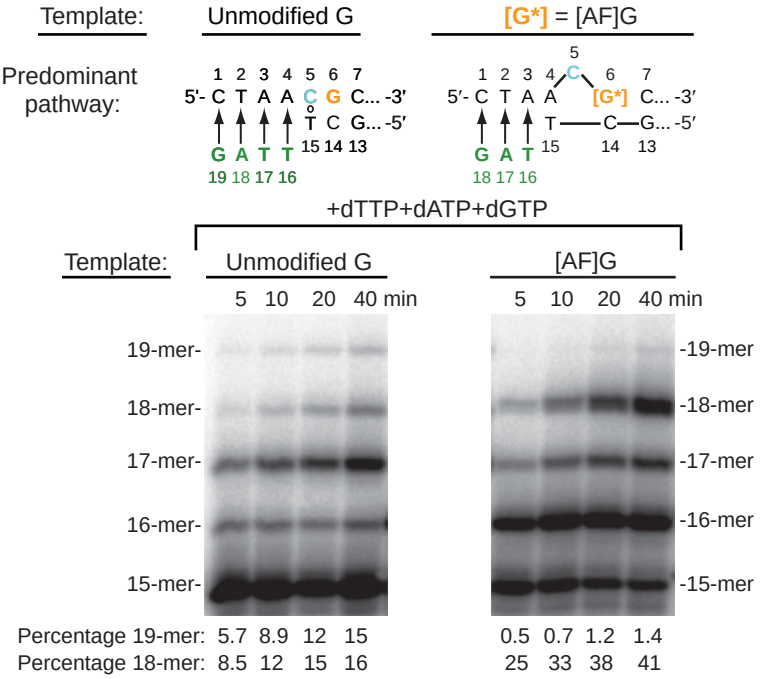


Supplementary Figure 6. Kinetics of single nucleotide insertion on the [AF]G-modified template. The examples of the gels show data as a function of dNTP concentration; the reaction times are indicated on the top right corner of each panel. The primer extension data recorded under single-hit polymerization conditions (less than 20% of primer extended) were used for estimation of Michaelis-Menten parameters listed in **Supplementary Table 1**.

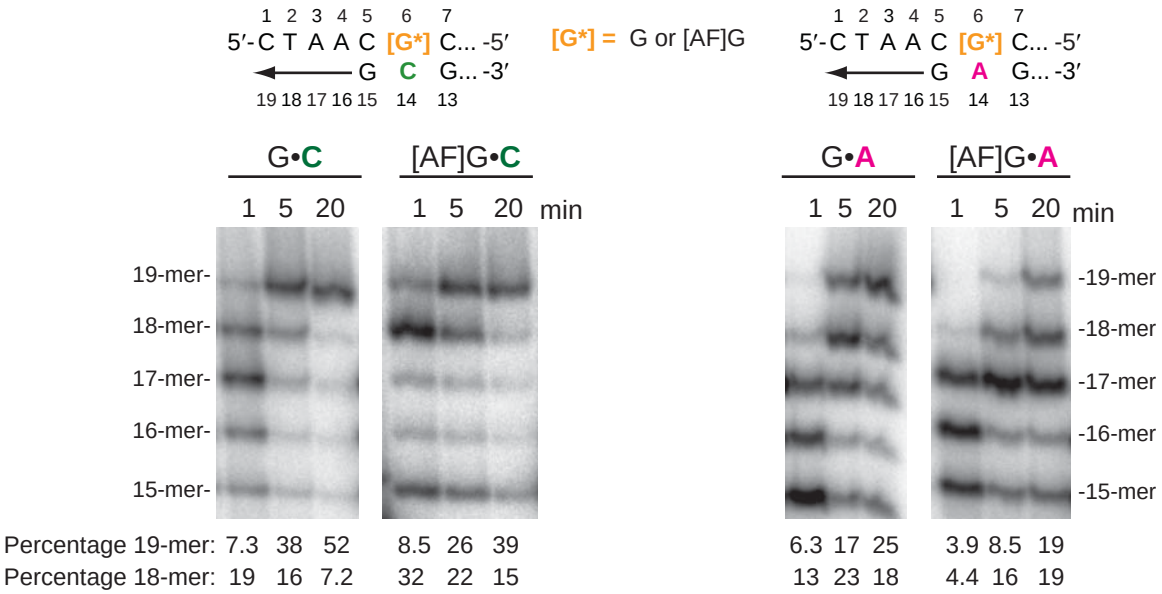


Supplementary Figure 7. Efficiency and fidelity of extension from the primer C, A, T or G base opposite the template G or [AF]G. **(a)** Time course of ^{32}P 5′-end-labeled 14-mer primer

extension with terminal C, A, T or G base on 19-mer G- and [AF]G-templates in the presence of all dNTPs. The fractions of fully extended 19-mer product and shorter 18-mer products are indicated at the bottom of the panels. Note the approximately equal fractions of 19-mer and 18-mer products accumulating over the reaction time during extension from a C base opposite [AF]G (the left set of panels). In contrast, during extension from a G•C pair almost all 18-mers are elongated to 19-mers. Extension from a G base opposite a modification site on the [AF]G template is faster than on the unmodified template (the right set of panels). Furthermore, the 18-mer is the longest product observed during extension from G base opposite the [AF]G-adduct, while small amounts of the 19-mer product are observed with the unmodified template due to a weak extension from the C5•T mismatch. (**b**) Time course of ^{32}P 5′-end-labeled 14-mer primer extension with terminal G base on [AF]G-templates in the presence of either dGTP, dTTP, or dATP, a mixture of dTTP with dATP or a mixture of dTTP, dATP and dGTP. Insertion of dGTP is inefficient, while the insertion of dTTP occurs readily. This insertion preference is consistent with dTTP insertion opposite template A4 base in an intermediate containing the C5•G14 terminal template/primer pair and looped out [AF]G. dNTP-misinsertion misalignment mechanism has been demonstrated previously for aromatic amine adducts, other lesions and unmodified sequences[21,37-42]. However, this misalignment pathway is expected to produce a very small fraction of extended products, due to an inefficient insertion of a G base opposite the [AF]G adduct that occurs with a ~0.05 % probability relative to the insertion of the other dNTPs (**Supplementary Table 1**). (**c**) Kinetics of extension from G on unmodified-G and [AF]G-templates by dTTP under steady-state conditions. The gels show data as a function of dTTP concentration. dTTP insertion on unmodified-G template apparently represents extension from a G6•G14 template/primer mismatch via an 'incorrect' nucleotide (dTTP) and occurs with V_{max}=0.38 ± 0.04 nM/min, and K_m=660 ± 90 μM. On the [AF]G modified template, extension occurs through a misaligned intermediate, where primer G14 forms a base pair with template C5 and the modified base is looped out, and insertion of dTTP occurs opposite the template A4 base. Michaelis-Menten parameters on the [AF]G-template are: V_{max}=1.9 ± 0.24 nM/min and K_m=56 ± 8 μM. Thus, dTTP incorporation to G14-terminated primer strand occurs 62-fold faster on the [AF]G-modified templates, than on the unmodified template. The efficiency of extension beyond an [AF]G•G terminus via dTTP incorporation is ~5-fold higher than extension beyond the matched [AF]G•C pair via correct dGTP (**Supplementary Table 1**).

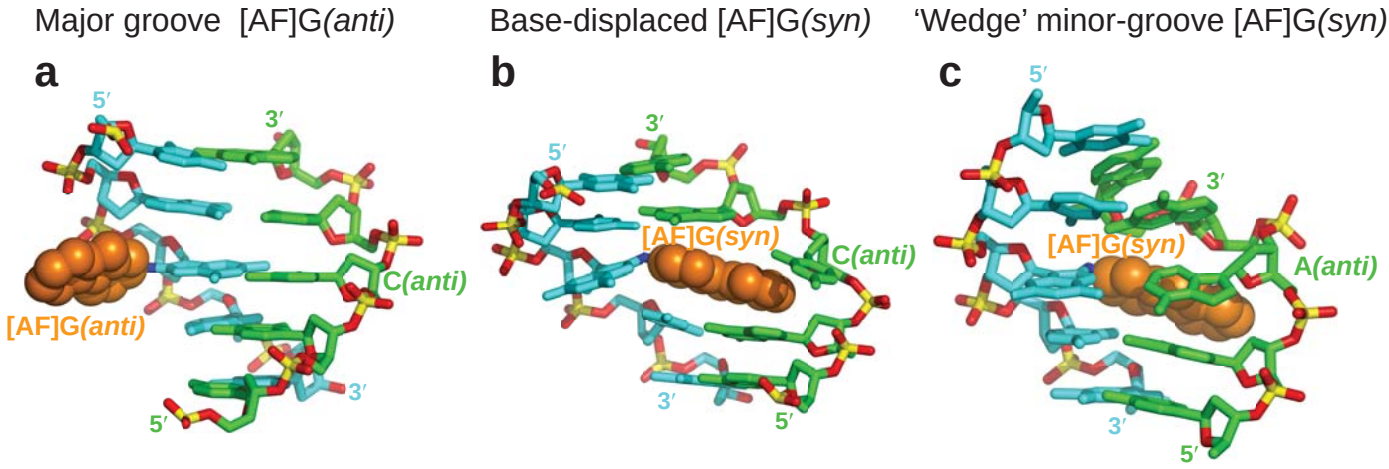


Supplementary Figure 8. Efficiency of extension from a 15-mer primer that either forms a mismatched C5•T15 template/primer terminus or creates a misaligned structure with correct A4•T15 template/primer terminus and a looped out C5 base 5′ to the modification site. Time course of ^{32}P 5′-end-labeled 15-mer primer extension on 19-mer G- and [AF]G-templates in the presence of all dNTPs necessary for correct extension (dTTP, dATP and dGTP). The fractions of fully extended 19-mer product and shorter 18-mer products are indicated at the bottom of the panels. In the case of the [AF]G-template, the 18-mer is the predominantly observed product that is consistent with the looped out template C5 and extension from the A4•T pair. With the unmodified template, the 19-mer product is also observed consistent with some extension from the C5•T mismatch.



Supplementary Figure 9. Extension efficiency from the cognate 15-mer primer with either C or A base opposite the modification site. Time course of ^{32}P 5′-end-labeled 15-mer primer extension on 19-mer G- and [AF]G- templates in the presence of all dNTPs necessary for correct extension (dTTP, dATP and dGTP). The fractions of the fully extended 19-mer products and shorter 18-mer products are indicated at the bottom of the panels. The data demonstrate that further extension from the C5•G15 base pair with properly aligned template/primer strands using a 15-mer primer with a C or A opposite the [AF]G adduct proceeds with similar rates as in the case of unmodified DNA; this is evident from the similar amounts of fully extended 19-mer primer bands that are observed in all cases. Interestingly, extension from the G15 next to the [AF]G6*(anti)*•A14 pair in the [AF]G•A-2 complex is reduced by ~2 fold compared to the [AF]G•C-2 complex. We hypothesize that the conformational heterogeneity of A14 opposite [AF]G (**Fig. 5**) might induce the mobility of the 3′-terminal base G15 in the primer strand, and, thus, decrease reaction efficiency.

The alignments for [AF]G adduct in free DNA duplexes in solution



Supplementary Figure 10. The alignments for the [AF]G adduct in free DNA duplexes in solution. (**a**) A major groove structure with the [AF]G*(anti)* opposite a partner C base[64]. The AF-moiety of [AF]G*(anti)* is positioned externally in the major groove with the modified G in the normal *anti* conformation forming a Watson-Crick base pair with the partner C. (**b**) A base-displaced structure with the [AF]G*(syn)* opposite the partner C base[62]. The AF-moiety is intercalated between intact neighboring base pairs, the modified G adopts the *syn* conformation and is displaced into the major groove, and Watson-Crick pairing is disrupted. (**c**) A 'wedge' minor groove conformation with the [AF]G*(syn)* opposite a partner A base[35]. The AF ring system is wedged into the minor groove and the modified-G*(syn)* is inserted inside the helix. The coordinates for these structures, provided by Dr. Lihua Wang (Biology Department, New York University), were obtained from molecular dynamic simulations based on the NMR solution structures[33].

Supplementary Table 1. Kinetic parameters of insertion and extension catalyzed by Dpo4.

Substrate: 5′-CTAAC-**X**-CTACCATCCAACC-3′
(**Y**)GATGGTAGGTTGG-5′

dNTP Insertion opposite:	dNTP	V_{max}, nM/min	K_M, μM	V_{max}/K_M	f_{ins}*#
X = G, no Y	dCTP‡	2.8 ± 0.3	1.96 ± 0.4	1.5	1
	dATP‡	0.29 ± 0.02	2860 ± 300	10^{-4}	6.7×10^{-5}
	dTTP‡	0.48 ± 0.03	2550 ± 300	2.0×10^{-4}	1.3×10^{-4}
	dGTP‡	0.04 ± 0.008	1960 ± 600	2.0×10^{-5}	1.3×10^{-5}
X = [AF]G, no Y	dCTP	2.8 ± 0.3	15.8 ± 4	0.18	0.12
	dATP	0.12 ± 0.02	1040 ± 120	1.1×10^{-4}	7.3×10^{-5}
	dTTP	0.67 ± 0.1	1370 ± 290	4.9×10^{-4}	3.3×10^{-4}
	dGTP	0.049 ± 0.005	600 ± 90	8.2×10^{-5}	5.5×10^{-5}
Extension from terminal base pair with 'correct' dNTP:					f_{ext}*†
Y = C opposite X = G	dGTP‡	0.32 ± 0.02	1.2 ± 0.2	0.28	1
Y = A opposite X = G	dGTP‡	0.12 ± 0.006	280 ± 50	4.0×10^{-4}	1.4×10^{-3}
Y = T opposite X = G	dGTP	0.26 ± 0.05	660 ± 160	3.9×10^{-4}	1.4×10^{-3}
Y = G opposite X = G	dGTP	0.11 ± 0.05	720 ± 190	1.5×10^{-4}	5.5×10^{-4}
Y = C opposite X = [AF]G	dGTP	0.11 ± 0.03	580 ± 60	1.9×10^{-4}	6.8×10^{-3}
Y = A opposite X = [AF]G	dGTP	0.017 ± 0.03	450 ± 70	3.8×10^{-5}	1.4×10^{-4}
Y = T opposite X = [AF]G	dGTP	0.009 ± 0.001	620 ± 120	1.5×10^{-5}	5.4×10^{-4}
Y = G opposite X = [AF]G	dGTP	0.003 ± 0.0003	450 ± 130	6.6×10^{-6}	2.4×10^{-5}

*$f = V_{max}/K_m$

All f_{ins} values normalized to the X = G, dCTP insertion value.

† All f_{ext} values normalized to the Y = C opposite X = G value.

‡ These data were obtained by us previously[49].

SUPPLEMENTARY DATA

Translocations of the thumb and little finger domains

The contacts of the little finger and thumb domains of Dpo4 with the template/primer DNA duplex[48] undergo stepwise changes that produce the alignments observed in molecule 1 and molecule 2 of the [AF]G•C-1 complex. These changes are similar to the ones observed previously during normal Dpo4 transitions through the catalytic cycle[49].

The [AF]G•A-1 complex has a pattern of Dpo4 interactions with DNA similar to the one in an unmodified ternary complex[48-49], and thus serves as a starting point for comparison. In the [AF]G•A-1 complex, the little finger domain contacts the phosphates spanning the C5-A9 segment of the template and A7-G9 of the primer (**Fig. 3a, top**), while the thumb domain interacts with the phosphate groups of A12 of the template and A12-G13 of the primer (**Fig. 3a, bottom**). However, in molecule 1 of the [AF]G•C-1 complex, the little finger domain contacts the same residues as in the [AF]G•A-1 complex (**Fig. 3b, top**), while the thumb domain is shifted towards the active site and interacts with the phosphate groups of C11 of the template and G13-C14 of the primer (see arrow, **Fig. 3b, bottom**). The thumb, the finger and palm domains are rotated 18-19° counterclockwise around the DNA, and the little finger domain is rotated by 4° on proceeding from [AF]G•A-1 to [AF]G•C-1 complexes (stereo view of translocation at the full structure level in **Supplementary Fig. 2a**, as well as thumb domain level in **Supplementary Fig. 2b**, following superposition of DNAs). Similar changes were observed after covalent nucleotide incorporation in the unmodified complex, and the pattern of Dpo4 interactions with DNA in molecule 1 of the [AF]G•C-1 extension ternary complex resembles that of an unmodified binary complex (without dNTP)[49].

The Dpo4-DNA interaction pattern in the 'mutagenic' molecule 2 of [AF]G•C-1 corresponds to that of a ternary complex. The little finger domain has translocated from the position observed in molecule 1 to contact the phosphates of the [AF]G6-T8 segment of the template (C5 is looped out and the A4 phosphate is disordered) and of the T8-G10 segment of the primer (see arrow, **Fig. 3c, top**), while the thumb domain maintains its contacts as in the molecule 1 complex (**Figs. 3c, bottom**). The Dpo4 undergoes the continued counterclockwise rotation around the DNA with the thumb, palm and finger domains rotated by 11°, and the little finger domain rotated by 26° on

proceeding from [AF]G•C-1 molecule 1 to molecule 2 complexes (stereo view of translocation at the full structure level in **Supplementary Fig. 2c** and little finger level in **Supplementary Fig. 2d**, following superposition of DNAs). This conformational change normally occurs during the dNTP binding step, when the little finger, palm and finger domains translocate relative to the template/primer duplex, in order to create space for the next template base and the dNTP to enter the active site[49].

The palm, finger and thumb domains are rotated an additional 4-5° on proceeding from the [AF]G•C-1 'mutagenic' molecule 2 extension complex to the [AF]G•C-2 post-extension complex (**Supplementary Fig. 3**), to allow the correct template/primer-dTTP alignment in the latter complex.

Examples of semi-targeted mutations

It has been reported that 10% of all mutations observed with a γ-radiation induced guanine-thymine intrastrand crosslink in human embryonic kidney cells were semi-targeted base substitutions, predominantly immediately 5′ to the lesion site[52]. Up to 23% of the observed mutations were semi-targeted to the nearest nucleotides flanking the TT cyclobutane dimer (CPD) in Pol η-deficient human cells[53]. In the case of the bulky aflatoxin B1-N7-guanine adduct, 13% of the mutations in *E.coli* were found at the base 5′ to the lesion[54]. Ethyl- and benzyl-O6-G adducts, but not the smaller methyl-O6-G lesions, caused semi-targeted mutations in rodent cells[55]. Moreover, in the repetitive DNA sequence 5′-CCC$G_1G_2G_3$ with an [AAF]G adduct positioned at G_1 or G_2 (but not G_3), up to 10% of one-base deletions were induced in a run of C's 5′ to the adduct site in SOS-induced *E.coli*[40]. About 25% of all mutations in *E.coli* caused by a BPDE adduct at N2-G_1 in the 5′-CG_1G_2C-3′ sequence were G_2 to A substitutions[56].

In a related study of Dpo4 replication past the 2-amino-3-methylimidazo[4,5-f]quinoline (IQ)-C8-guanine adduct, a semi-targeted mutation with several adjacent mismatches/deletions 5′ to the [IQ]G adduct was characterized by tandem mass-spectroscopy methods[42]. Recent biochemical and molecular modeling studies showed that during replication past a bulky BPDE-N^2-dG adduct by Dpo4, the unmodified G 3′ to

the lesion can be skipped in 5′-…T[BPDE-N^2-G] G-…[61] and in 5′-…C[BPDE-N^2-G]G…[57] sequence contexts.

[AF]G conformations in Dpo4-free and Dpo4-bound states

In previous NMR solution structural studies it was shown that the [AF]G lesion paired with C in double-stranded DNA, within the same 5′-…C-[AF]G-C… sequence context as employed in this study, exists in two conformations: a major groove structure in the case of the adduct in the [AF]G*(anti)* conformation (**Supplementary Fig. 10a)** and a base-displaced [AF]G*(syn)* structure (**Supplementary Fig. 10b**) with a 30 : 70 ratio, respectively[62]. In other sequence contexts, the fraction of the [AF]G*(anti)* conformer varies from 10 to 90%[33-34,59,63]. Interestingly, within the Dpo4 active site we observed the [AF]G adduct only in the *anti* conformation. In solution, the AF-moiety of [AF]G*(anti)* is solvent exposed and the base-base stacking of the modified-G*(anti)*•C pair with adjacent base pairs is unperturbed[64]. However, within Dpo4, the AF-moiety is shielded from the solvent either by stacking with the A4 and C5 bases 5′ to the [AF]G*(anti)* at the (–1) position of the active site, or by insertion into the pocket on the surface of the little finger domain at the (–2) position of the active site. In the former case, base stacking is disrupted, but in the latter case base stacking is preserved and the structure of the [AF]G*(anti)*-modified DNA segment is remarkably similar to that observed in solution[64]. Thus, the interactions with Dpo4 polymerase stabilize the [AF]G*(anti)* conformer in a Watson-Crick arrangement with a partner C base.

The [AF]G adduct in the *syn* glycosidic conformation opposite A in a free DNA duplex in solution, adopts a ‘wedge’ conformation with the AF-moiety in the minor groove[35] (**Supplementary Fig. 10c**); a flanking C base 3′ to the lesion site promotes this alignment[36]. However, despite the open minor groove of the Dpo4 active site, the [AF]G adduct is observed either in the *syn* conformation with the AF-moiety intercalated inside the template/primer helix at the (–1) position of the active site, or in the *anti* conformation with the AF-moiety in the pocket on a surface of the little finger domain at the (–2) position. Thus, Dpo4 does not favor the formation of an incorrect modified-G*(syn)*•A*(anti)* base pair that can occur in the case of the ‘wedge’ [AF]G*(syn)* alignment.

Interestingly, the structure of the DNA segment with [AF]G*(syn)* opposite A within Dpo4 is similar to the one observed with the [AF]G*(syn)* opposite C in solution[62].

SUPPLEMENTARY METHODS

Preparation and Purification of Template and Primer DNA Strands

The DNA 19-mer template 5′-CTAACGCTACCATCCAACC-3′, 13-mer primer 5′-GGTTGGATGGTAG-3′, 14-mer primer 5′-GGTTGGATGGTAGX-3′ (X = C, A, T, or G), 15-mer primer 5′-GGTTGGATGGTAGXG-3′ (X = C or A) with 3′-OH, were synthesized using an automated Applied Biosystem 392 DNA synthesizer with phosphoramidite chemistry. The oligomers were cleaved from the support, deprotected with ammonium hydroxide for 5 hours at 55°C, purified on 20% polyacrylamide gel electrophoresis in the presence of 8 M urea, electro-eluted, and desalted. 2′,3′-dideoxy-A and 2′,3′-dideoxy-G at the 3′ end were introduced into the 13-mer primers 5′-GTTGGATGGTAGA-3′ and 5′-TTGGATGGTAGXG-3′ (X = C or A), respectively, by reverse 5′ to 3′ synthesis using 5′-CE phosphoramidites. The 13-mer primer with 3′-terminal 2′,3′-dideoxy-C 5′-GTTGGATGGTAGC-3′ was synthesized using 2′,3′-ddC columns by regular 3′ to 5′ synthesis, and purified as described.

The purified unmodified 19-mer template was converted into the [AF]G -modified sequence 5′-CTAAC-[AF]G-CTACCATCCAACC-3′ as described previously[35]. Briefly, the oligonucleotide was first converted into its acetylaminofluorene [AAF]G adduct by treating the 19-mer oligomer, dissolved in 0.002 M sodium citrate buffer, pH 7.1, with an excess of N-acetoxy-2-acetylaminofluorene, dissolved in absolute ethanol (10 mg/mL), in a molar ratio of 1:8. The mixture was shaken at 37°C in the dark for 3 hr and then extracted with water-saturated diethyl ether. The crude [AAF]G-modified oligomer contained in the remaining aqueous solution was purified by C3 reverse-phase HPLC using a linear gradient of acetonitrile against 0.01 M triethylammonium acetate buffer, pH 7.1 (0-296 in 30 min). The pure [AAF]G-modified oligomer was converted into the [AF]G-modified oligomer by dissolution in 1 M NaOH containing 0.3% (v/v) 2-mercaptoethanol at a concentration of 2 mg/mL. The reaction was allowed to proceed for 45 min at room temperature and then neutralized with dilute hydrochloric acid. After

extensive purging of the solution with nitrogen and desalting on a Sephadex G-25 column, the [AF]G-modified oligomer was purified by HPLC, as in the case of the [AAF]G-modified oligomer. The pure [AF]G-modified oligomer was desalted on Sephadex G-25 and converted to sodium form on a Dowex 50x8 cation exchange resin. All phosphophoramidites were purchased from Glen Research. The reagents for crystallization were obtained from Hampton Research.

Preparation and Purification of Dpo4

The DNA fragments encoding the 353 amino acid full-length Dpo4 (plus 5 glycines at the N-terminus) were obtained by PCR and inserted into pET-28a between *Nde*I and *EcoR*I cleavage sites. The His-tag version of Dpo4 was expressed in *E.coli* BL21-CodonPlus (DE3)-RIL strain (Stratagene), and then subjected to a Ni-chelating column. The 6-His tag was next cleaved with thrombin, and the Dpo4 protein was then further purified by heparin column chromatography, followed by passage through a Superdex-200 column. Dpo4 was concentrated to 26 mg/ml in 5 mM DTT, 25 mM HEPES, pH 7.5 and 300 mM NaCl.

Crystallization

The crystals of the Dpo4 extension ternary complexes containing [AF]G-modified 19-mer template 5′-CTAAC-[AF]G-CTACCATCCAACC-3′ and the 13-mer primers terminated at the 3′-end with 2′,3′-dideoxy-C or 2′,3′-dideoxy-A 5′-GTTGGATGGTAGX-3′ (X = C or A) were grown in the presence of dGTP under conditions described previously[49]. The crystals of the post-extension complexes were grown with 2′,3′-dideoxy-G 13-mer 5′-TTGGATGGTAGXG-3′ (X = C or A) in the presence of dTTP. Multiple attempts to crystallize the [AF]G-modified insertion ternary complexes with the 13-mer 2′,3′-dideoxy terminated primer 5′-GGTTGGATGGTAG-3′ a dCTP, dATP or dTTP opposite the lesion site were unsuccessful. Briefly, template/primer DNA's were annealed and mixed with Dpo4 in 1.2:1 molar ratio to a final concentration 0.15 mM in 20 mM HEPES pH 7.0, 60 mM NaCl, 5 mM $MgCl_2$, 1mM DTT, and dNTP (1 mM) was added to produce the ternary complexes. The protein-DNA complexes were than incubated at 37°C for 5 minutes and centrifuged at 11,000 rpm for 5 minutes at 4°C.

The crystals were grown by the hanging drop method against a reservoir solution containing 100 mM HEPES, pH 7.0, 100 mM calcium acetate and 10% PEG 4000 at 20°C. Several rounds of micro seeding (Hampton Research kit) were employed to produce the diffraction quality crystals of the [AF]G•C-1 and [AF]G•A-1 extension ternary complexes. The crystals were transferred to a cryosolution containing the mother liquor with 15% PEG 4000 and 15% ethylene glycol and flash-frozen in liquid nitrogen for X-ray data collection.

Structure determination and refinement

X-ray diffraction data for [AF]G•A-1, [AF]G•A-2 and [AF]G•C-2 were collected at the NE-CAT 24-ID-C beam line at the Advanced Photon Source (Argonne National Laboratory, Chicago). The data from the [AF]G•C-1 crystal were collected at the NE-CAT 24-ID-E micro-focus beam line. The data were processed and scaled using the HKL2000 suite. The structures were solved by molecular replacement (AMoRe[65]) using our published oxoG-modified insertion ternary Dpo4-DNA-dCTP structure (PDB ID: 2ASD[49], as a search model. The thumb (residues 167-233) and the little finger (residues 244-341) domains were split from the complex and its position refitted by AMoRe. The model building, including substitution of the DNA sequence, was manually finished in TURBO-FRODO (http://www.afmb.univ-mrs.fr/-TURBO-) based on the electron density maps calculated in REFMAC, and the resulting models were refined in REFMAC. Necessary replacements of DNA bases and some conformational changes were then introduced, and models were refined. The crystal data, together with the data collection and refinement statistics for all structures are summarized in **Supplementary Table 1**. The simulated annealing omit maps were calculated in CNS with the [AF]G, partner base and Arg336 omitted from the models after they were heated to 2,000 K and then slowly cooled to remove model bias. Figures were prepared with PyMol (http://www.pymol.org).

Primer elongation standing start assay

The template/primer DNA complexes formed by annealing a 19-mer template containing unmodified-G or [AF]G at the same site with a ^{32}P 5′-end labeled 3′-OH terminated

primer (13-mer ending one base before the modification site, 14-mers with C, A, T or G opposite G, [AF]G or [AAF]G and 15-mers ending 5′ to the modification site), were incubated with Dpo4 polymerase in the presence of all four dNTPs. Aliquots were withdrawn from the reaction mixture after incubation times indicated in the figures and quenched by a gel-loading buffer (95% formamide with 20 mM EDTA, 45 mM Tris-borate, 0.1% bromophenol blue, 0.1% xylene cyanol). For a typical experiment 25 µl of the solution of all four dNTPs was added to the 25 µl of polymerase plus template/primer DNA mixture, both solutions were in 100 mM HEPES, pH 8.0, 5 mM $MgCl_2$, 60 mM NaCl, 5 mM dithiothreitol. The polymerization reactions were conducted at 30°C, final template/primer DNA concentration was 10 nM, Dpo4 was 10 nM, and 100 µM of each dNTP. The reaction products were resolved on 20% polyacrylamide gels in the presence of 8 M urea on Bio-Rad Sequi-Gen GT System (38 x 50 cm gels), the gels were dried before radiography with Fuji image plate. The images were scanned on Fuji PhosphoImager and the bands were quantified using profile analysis mode in ImageGauge software.

Steady state kinetic analysis of one base insertion and extension

Steady state kinetics parameters were analyzed for incorporation of each deoxynucleotide opposite the [AF]G and [AAF]G with the 13-mer primer and for extension of the 14-mer primers with C, A, T or G 3′-terminal opposite G and [AF]G in the presence of the next correct nucleotide (dGTP)[66]. The extension from the 14-mer primer with 3′-terminal G was also studied in a presence of dTTP to evaluate efficiency of reaction pathway with looped out [AF]G (see text). In a typical experiment, 6 µL solutions with increasing concentrations of single dATP, dTTP, dGTP or dCTP were added to 6 µL of polymerase/template-primer DNA mixture, and the reactions were stopped after 5 minutes by the addition of 12 µL of denaturing loading buffer. To insure single hit polymerization conditions (less than 20% of primer extended), the nucleotide concentration interval and, in some cases incubation time, was adjusted for every experiment. The template/primer DNA concentrations were 50 nM, and Dpo4 was 2 nM. The gel band intensities of extended and unreacted primer strands were quantified as described. The reaction rates (ν, nM/min) were plotted as a function of the dNTP

concentration, and the data were fit by nonlinear regression of the Michaelis-Menten equation,

$$\nu = V_{max} \times [dNTP]/(K_M + [dNTP])$$

to calculate apparent K_M and V_{max} steady state parameters[49,66-67]. The frequency of nucleotide incorporation (f_{inc}) and extension (f_{ext}) were defined as followed:

$$f_{inc} \text{ or } f_{ext} = (V_{max}/K_M)_f.$$

References:

61. Xu, P., Oum, L., Geacintov, N.E. & Broyde, S. Nucleotide selectivity opposite a benzo[a]pyrene-derived N2-dG adduct in a Y-family DNA polymerase: a 5'-slippage mechanism. *Biochemistry* **47**, 2701-9 (2008).

62. Mao, B., Hingerty, B.E., Broyde, S. & Patel, D.J. Solution structure of the aminofluorene [AF]-intercalated conformer of the syn-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry* **37**, 81-94 (1998).

63. Meneni, S.R. et al. Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. *Nucleic Acids Res* **34**, 755-63 (2006).

64. Mao, B., Hingerty, B.E., Broyde, S. & Patel, D.J. Solution structure of the aminofluorene [AF]-external conformer of the anti-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry* **37**, 95-106 (1998).

65. Navaza, J. AMoRe: an Automated Package for Molecular Replacement. *Acta Crystallographica* **A50**, 157-163 (1994).

66. Creighton, S., Bloom, L.B. & Goodman, M.F. Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol* **262**, 232-56 (1995).

67. Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L. & Prakash, S. Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. *Nat Genet* **25**, 458-61 (2000).