A structural gap in Dpo4 supports mutagenic bypass of a major benzo[*a*]pyrene dG adduct in DNA through template misalignment

Jacob Bauer[†], Guangxin Xing[†], Haruhiko Yagi[‡], Jane M. Sayer[‡], Donald M. Jerina[‡], and Hong Ling[†§]

[†]Department of Biochemistry, University of Western Ontario, London, ON, Canada N6A 5C1; and [‡]Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892

Edited by Wei Yang, National Institutes of Health, Bethesda, MD, and accepted by the Editorial Board July 11, 2007 (received for review January 25, 2007)

Erroneous replication of lesions in DNA by DNA polymerases leads to elevated mutagenesis. To understand the molecular basis of DNA damage-induced mutagenesis, we have determined the x-ray structures of the Y-family polymerase, Dpo4, in complex with a DNA substrate containing a bulky DNA lesion and incoming nucleotides. The DNA lesion is derived from an environmentally widespread carcinogenic polycyclic aromatic hydrocarbon, benzo[*a*]pyrene (BP). The potent carcinogen BP is metabolized to diol epoxides that form covalent adducts with cellular DNA. In the present study, the major BP diol epoxide adduct in DNA, BP-N^2-deoxyguanosine (BP–dG), was placed at a template–primer junction. Three ternary complexes reveal replication blockage, extension past a mismatched lesion, and a −1 frameshift mutation. In the productive structures, the bulky adduct is flipped/looped out of the DNA helix into a structural gap between the little finger and core domains. Sequestering of the hydrophobic BP adduct in this new substrate-binding site permits the DNA to exhibit normal geometry for primer extension. Extrusion of the lesion by template misalignment allows the base 5′ to the adduct to serve as the template, resulting in a −1 frameshift. Subsequent strand realignment produces a mismatched base opposite the lesion. These structural observations, in combination with replication and mutagenesis data, suggest a model in which the additional substrate-binding site stabilizes the extrahelical nucleotide for lesion bypass and generation of base substitutions and −1 frameshift mutations.

cancer | DNA replication | mutagenesis | translesion synthesis | polycyclic aromatic hydrocarbon

Translesion DNA synthesis (TLS), the extension of primers opposite damaged DNA templates, is a fundamental process that enables a cell to survive DNA damage (1–3). Although high-fidelity replicative DNA polymerases are generally blocked by DNA lesions, specialized DNA polymerases belonging to the Y-family are capable of replicating past DNA damage. These translesion polymerases are much less accurate than the replicative polymerases and can cause mutations by introducing errors as they replicate DNA. Thus, error-prone TLS by the Y-family polymerases provides a molecular connection between DNA damage and mutagenesis. Many mutagenic lesions in DNA are caused by environmental agents such as UV light or exogenous chemicals or their metabolic products. Polycyclic aromatic hydrocarbons (PAHs) are widespread environmental contaminants derived from incomplete combustion. They are present in cigarette smoke, automobile exhaust, smoked foods, and industrial emissions, and they enter the body through the air or through the diet (4, 5). Benzo[*a*]pyrene (BP) is a prototypical PAH that is highly tumorigenic and implicated in specific mutations found in lung tumors of smokers (6). BP is metabolized to BP diol epoxides (BP DEs) in mammals (7). These highly reactive DEs form covalent adducts with the purine bases of DNA (8). The predominant BP DE isomer, (+)-(7*R*,8*S*,9*S*,10*R*)-BP DE, is the most tumorigenic in animal models (7). Its reaction at N^2 of guanine gives the major BP DE DNA adduct (trans 10*S* BP-N^2-dG, Fig. 1*A*) in mammalian cells (9, 10). This adduct induces mainly base substitutions as well as −1 frameshift mutations *in vivo* (11, 12). In this study, we describe the structural basis for incorrect bypass of the BP–dG adduct by a Y-family polymerase.

Y-family DNA polymerases are involved in the mutagenic bypass of BP adducts (1, 3, 13–15). There are multiple Y-family DNA polymerases in cells. The human enzymes polη and polκ can replicate past BP–dG (16). DNA polymerase IV (Dpo4) from *Sulfolobus solfataricus* is homologous to polκ, yet shows lesion bypass abilities akin to those of polη (17). It has become a model system for understanding of replication past DNA damage at an atomic level. Dpo4 consists of four domains, a three-domain classic polymerase core of palm, finger, and thumb domains and an additional "little finger" (LF) domain unique to the Y-family (18). The four domains are conserved in all of the Y-family DNA polymerases (18). The finger and thumb domains are unusually small and form an open active site with few DNA contacts. This spacious active site has been demonstrated to be responsible for replication slippage, mismatched primer extension, and lesion bypass (18–22). The LF domain of Y-family polymerases interacts with the major groove of the bound DNA. Attachment of this LF domain to the core domain creates a structural gap between the core and the LF domains. This structural gap harbors the extrahelical BP dG adduct at the template–primer junction in the protein–DNA–dNTP complexes reported here, solved at 2.5- and 2.25-Å resolution [supporting information (SI) Table 1]. The extrusion of the adducted base by template misalignment results in either base substitution or −1 frameshift. The present results reveal a previously unidentified structural basis for mutagenic replication by Dpo4.

Results and Discussion

Benzo[*a*]pyrene Intercalation Blocks G* Bypass. We have determined structures for the adducted DNA–Dpo4–dNTP complex in two crystal forms. In the first crystal (BPG-1), the DNA substrate has a mispaired G*:A at the template–primer junction (sequence in

Author contributions: J.M.S., D.M.J., and H.L. designed research; J.B., G.X., and H.L. performed research; H.Y. and J.M.S. contributed new reagents/analytic tools; J.B., J.M.S., D.M.J., and H.L. analyzed data; and J.B., J.M.S., D.M.J., and H.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. W.Y. is a guest editor invited by the Editorial Board.

Abbreviations: PAH, Polycyclic aromatic hydrocarbons; BP, benzo[*a*]pyrene; BP–dG, BP diol epoxide-N^2-deoxyguanosine, the major adduct of (+)-(7*R*,8*S*,9*S*,10*R*)-BP DE, formed by trans addition of N^2-dG to (+)-(7*R*,8*S*,9*S*,10*R*)-BP DE; Dpo4, DNA polymerase IV; TLS, translesion synthesis; LF domain, little finger domain.

Data deposition: The atomic coordinates and structural factors have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 2IA6 (BPG-1) and 2IBK (BPG-2)].

[§]To whom correspondence should be addressed. E-mail: hling4@uwo.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0700717104/DC1.

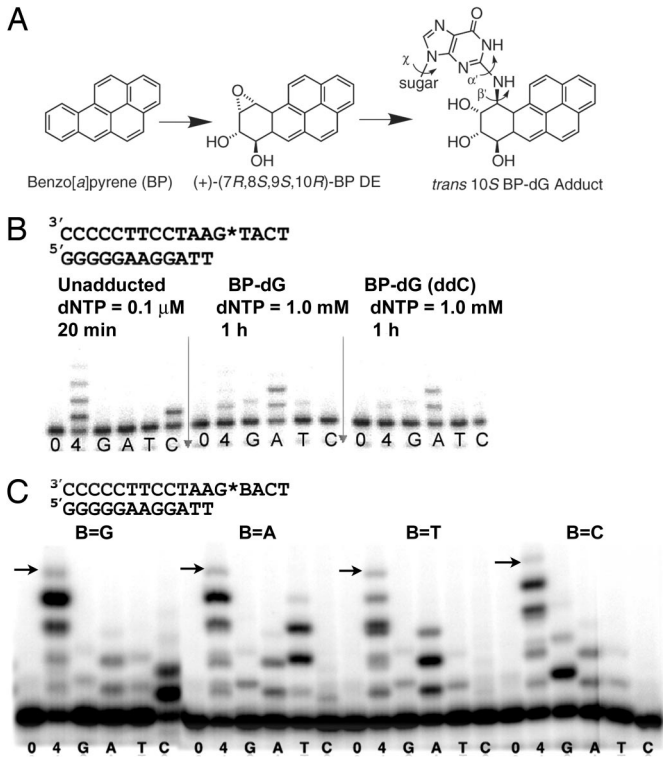

Fig. 1. Formation of a benzo[*a*]pyrene diol epoxide adduct (BP–dG) and its effect on DNA replication. (*A*) Structures of BP, its (+)-(7*R*,8*S*,9*S*,10*R*) diol epoxide (DE) metabolite and the trans 10*S* BP–dG adduct in DNA. The α', β', and χ torsion angles are labeled. (*B*) Incorporation of nucleotides opposite the BP–dG adduct in the template in comparison with incorporation opposite an unmodified dG residue. The BP–dG adduct is indicated as G*. Assays were performed with the dNTP concentrations and the times indicated. Lanes 0, 4, G, A, T, and C correspond to no dNTP, all four dNTPs, and each single dNTP, respectively. ddC (*Right*) indicates the 3′ end of the template strand is dideoxy-C, which is used in crystallization to prevent nucleotide addition at the blunt end (21). (*C*) Incorporation of dNTP opposite the BP–dG adduct in four sequence contexts by Dpo4. The arrows indicate the expected size of the full-length products. Reaction time was 1 hour with 100 μM dNTPs.

Fig. 1*B* plus an A base in the primer opposite the adducted base G*) and an incoming dATP matching the template dT 5′ to the lesion. In primer extension assays with the given template a dA residue was most efficiently incorporated opposite the dG* adduct (Fig. 1*B*), and this dA:dG* mispair was also most efficiently extended (Fig. 1*B* and SI Fig. 5*A*). With 1 mM dNTPs, bypass of the adduct was less efficient in the presence of all four dNTPs than with dATP alone (Fig. 1*B*). We suggest that the other dNTPs, which are poor substrates compared with dATP, may act as competitive inhibitors of dATP incorporation at high dNTP concentration. BPG-1 contains two ternary complexes, BPG-1A and BPG-1B, in the asymmetric unit. The structures represent intermediates for primer extension beyond the lesion site. The conformation of the Dpo4 is almost identical in the two complexes, with rmsd on Cα of 0.3 Å, but the conformation of the DNA differs in the structures.

BPG-1A has the pyrene ring system intercalated into the DNA helix. The hydrocarbon occupies the space corresponding to one base pair and inserts between the last two bases at the primer strand terminus (Fig. 2). Intercalation of the BP rings displaces the G* base into the minor groove in a syn conformation (SI Table 2). Base-pairing is disrupted both at the lesion (0) position (G*:A) and at the position 5′ to the lesion (+1 position, T:dATP). By pointing toward the 3′ side of the modified G base, the BP ring system stacks with the bases surrounding the lesion site and has its long axis at an angle of ≈40° with the DNA duplex axis. Accordingly the A base at the primer strand end tilts by 40° from its ideal position and overwinds the helix by ≈15° to maximize its stacking interactions with the pyrene rings (Fig. 2*D*). The intercalation also disturbs the −1 base pair (A:T) that is beneath the pyrene rings, introducing 32° of buckle and −5° of opening, lengthening the N^6(A)–O^4(T) and $N1$(A)–$N3$(T) hydrogen bonds to 3.1 and 3.5 Å, respectively. The single-stranded template 5′ to the lesion is disordered.

Disruption of the base-pairing at the template–primer junction and distortion of the template prevents the incoming dATP from binding in the correct conformation for primer extension. Without an ordered template base with which to pair, the incoming dATP is flipped over, with its Hoogsteen edge facing the template strand to maximize its stacking with the 3′-terminal base A of the primer (Fig. 3). In contrast to a normal cis base-paired nucleotide, the trans paired dATP has the sugar ring nearly upside down and its α phosphate shifting away from the 3′ end of the primer (Fig. 3). The distance between the 3′ OH of the primer strand and Pα of the dATP is 9.0 Å, too great for the phosphoryl transfer reaction to occur. In addition, only one Ca^{2+} was found in the active site, occupying the "B" site of the two catalytic metal ion-binding sites (18, 20). The complex BPG-1A thus represents a nonproductive conformation in which intercalation of the hydrocarbon prevents primer extension.

BP–dG Flipping-Out and Base Substitutions. In marked contrast, the BP adducted G* is flipped from the minor groove side into the gap between the LF and the core domains in the second complex (BPG-1B) (Fig. 2). In the enzyme-substrate complex, the conformation of the adduct with χ, α', and β' torsion angles (Fig. 1*A*) of −153.0, −160.0, and 161.7, respectively, does not lie in a low-energy region as predicted for the corresponding nucleoside adduct in the absence of an enzyme (23). In this conformation, the BP–dG adduct would be exposed completely to the solvent if there were no protein present. Full exposure of the hydrocarbon rings and aromatic base G is extremely unfavorable in an aqueous environment. None of the experimentally observed DNA structures containing BP–dG adducts have shown such a fully exposed conformation (24–27). The protein-free NMR structures have the BP ring system either intercalated or bound to DNA in the minor groove to avoid full exposure of the aromatic rings to water. However, such a BP ring-exposed conformation is stabilized in BPG-1B by burying the hydrophobic moieties in the gap between the core and the LF domains of Dpo4. The interface of the core and the LF domains is near the DNA template–primer junction and thus can accommodate this otherwise unfavorable conformation. When oriented in the gap, the pyrene ring packs against the LF domain, whereas the purine contacts the catalytic core domains (Fig. 4). The BP rings contact the hydrophobic parts of Lys 275 and Glu 271 (LF domain), with the outside edge being partially shielded by Tyr 274. The adducted base is stabilized by two H bonds between $N1$ of the base and the side chain of Glu 79 and between O^6 of G* and the main-chain N of Lys 78, as well as by hydrophobic interactions between the purine ring and Met 76 and the hydrophobic portion of Lys 78 (Fig. 4).

Flipping-out of the BP lesion from the intrahelical space reduces the structural distortion of the DNA duplex and allows base pairs around the lesion to stack with each other as in a normal B-form DNA (Fig. 2). The replicating base pair T:dATP and the two unpaired template nucleotides 5′ to the replicating base pair are structured, in contrast to a substantial disordering of this region in the intercalated structure BPG-1A. This allows the incoming dATP to bind normally in a standard Watson–Crick base pair (Fig. 3). The dATP sugar and phosphate moieties are now in their proper positions, and the distance between the primer 3′ OH and the Pα is 3.9 Å. This distance is small enough

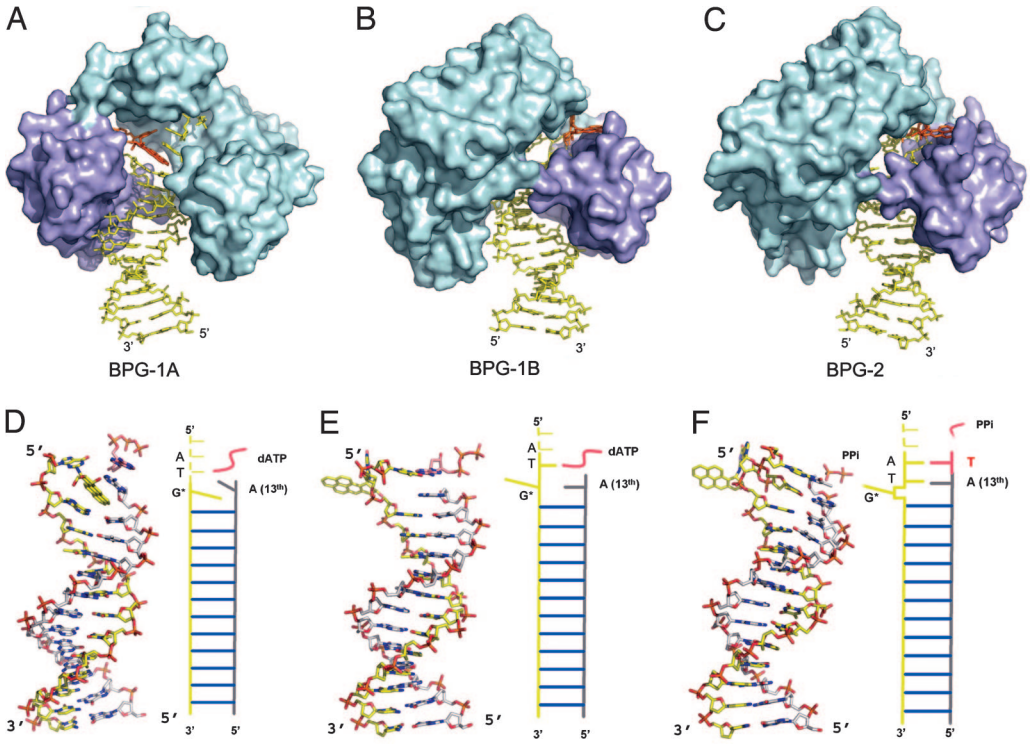

Fig. 2. Structures of BPG-1A, BPG-1B and BPG-2. (*A–C*) Dpo4 is represented as a molecular surface with the polymerase core in cyan and the LF domain in purple; DNA and nucleotide are shown as sticks, and BP–dG is highlighted in orange. BPG-1B and BPG-2 in *B* and *C* are rotated 180° relative to BPG-1A in *A* around the DNA helix axis, to show the extrahelical BP–dG in the gap between the core and LF domains. (*D–F*) The DNA conformations corresponding to (*A–C*) as stick models, all with the same orientations as in *A*. The primer strands are in gray, and incoming dATP is in pink. The single-stranded portion of the template DNA is not shown. Figs. 2, 3, and 4 were generated by using PYMOL (46).

to produce a catalytically productive conformation with only relatively minor conformational adjustments. Both metal ions were observed in the active site at their canonical binding sites (20, 28).

Extrusion of the BP from the helix leaves a noninstructional gap in the template strand. This gap is occupied by the terminal A of the primer strand, which would otherwise have been paired with the flipped-out G* (Fig. 2). A water molecule (red sphere, Fig. 3*B*) fills the remaining space between the A at the primer terminus and the opposite strand of the template. This noninstructional space in the template strand potentially allows the incorporation of any nucleotide opposite the lesion, which would

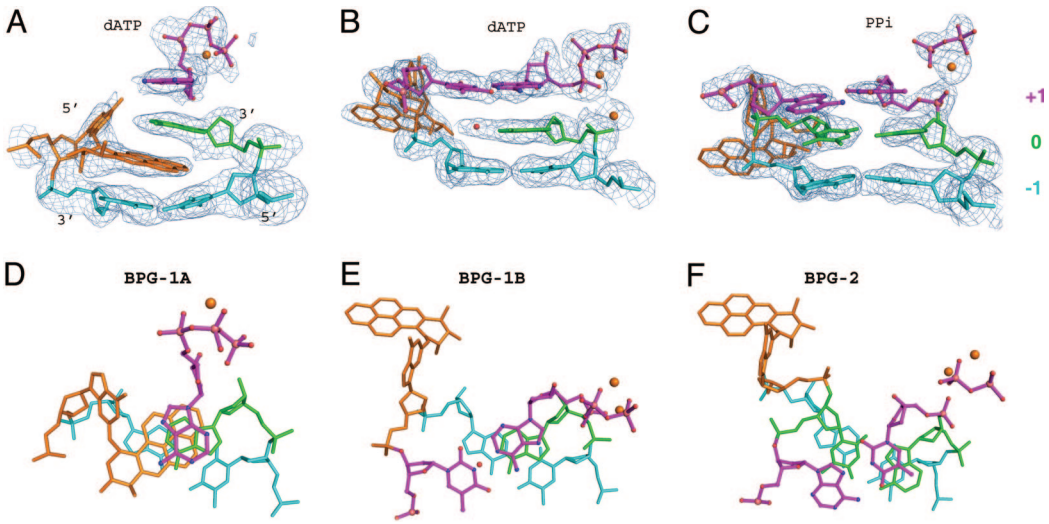

Fig. 3. The BP–dG adduct and its surrounding base pairs in the active site. The base pairs are colored according to their position relative to the lesion site: The lesion site (0) is green, +1 (above) is purple, and −1 is light cyan. The adducted nucleotide itself is orange. The DNA and dATP are shown as ball-and-stick models with Ca^{2+} ions and water molecules shown as orange and red spheres. (*A–C*) Side views of the base pairs with the simulated annealing omit map contoured at 2.5σ at 2.5 Å and 2.25 Å resolution, respectively. (*D–F*) Views rotated 90° relative to the corresponding image in *A–C*. The oxygen, nitrogen, and phosphorus atoms are colored red, blue, and pink, respectively, for the top layer.

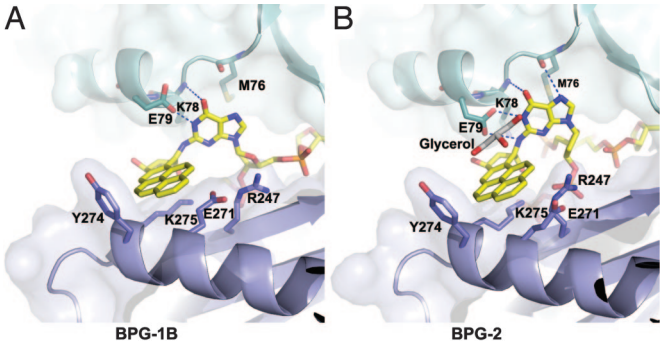

Fig. 4. Close-up views of BP–dG in the structure gap between the core and the LF domains. (*A*) BPG-1B. (*B*) BPG-2. The protein is in ribbon models covered by a transparent molecular surface. The key residues interacting with the adduct G* are shown as stick models. The BP ring system is in van der Waals contact with the LF domain (purple); the adducted G base interacts with the core domain (cyan). The glycerol molecule is in gray.

account for base substitutions induced by the BP-adduct. The conformation in BPG-1B could arise from untemplated incorporation of dA, or it could arise from a pairing of the dA with the dT 5′ to the lesion and a subsequent rearrangement. To examine this possibility with Dpo4, a second crystal was prepared (see next section).

BP–dG Looping-Out and −1 Frameshift Mutations. The second crystal (BPG-2) was cocrystallized Dpo4 with the same adducted DNA but with dTTP as the incoming nucleotide to match the A base that is the second template base 5′ to G*. In this structure, dA in the primer strand has skipped the dG* and paired with the dT 5′ to the lesion. This complex shows a successful extension of a misaligned adduct in the template–primer junction (Fig. 2). The dT has been incorporated into the primer strand, with PPi remaining in the active site. The remaining PPi was observed in the ternary complex of Dpo4 with a mismatched base pair (G:T) in the active site (22). The slow release of PPi has been suggested as facilitating pyrophosphorolysis, a reversal of DNA polymerization that might serve as a proofreading step for Y-family polymerases. Instead of flipping out of the adducted base as in BPG-1B, the whole BP-adducted nucleotide unit, including the sugar–phosphate backbone, is looped out of the DNA helix by template misalignment in BPG-2. The looping-out was secured by two base pairs 5′ to the lesion site. This conformation of BP–dG (with χ, α', and β' torsion angles of 87.0, −100.0, and 179.0, respectively) is energetically unfavorable based on molecular simulation (23). The BP–dG adduct again is extruded from the minor groove side and occupies the cleft between the core and LF domains. The BP ring system is stabilized by the same residues from the LF domain as in BPG-1B, with the solvent accessible face covered by a glycerol molecule (Fig. 4). The carboxyl group of Glu-79 and main chain *N* and *O* from Lys-78 and Met-76 in the palm domain form four H-bonds with $N1$, N^2, O^6 and $N7$ of the G* base. The G* stacks against Met 76 and Lys 78 as in BPG-1B. Therefore, the bulky adduct is sandwiched between the LF and the palm domains in both structures.

With the BP–dG completely looping out, the template dT 5′ to the lesion is brought down to fill the gap left by the BP–dG and to pair with the A at the 13th position of the primer strand, resulting in a −1 frameshift (Fig. 2*F*). The very weak top bands for the full-size product of primer extension in Fig. 1*C* indicate that the major product is one base shorter than the template strand as a result of template misalignment. When the 5′ dT is shifted down to fill the gap, it is pushed much farther into the center of the helix, closer to the primer strand (green base pair in Fig. 3 *C* and *F*). To retain its normal backbone position in the protein active site, the complementary dA at the 13th position in the primer strand rotates from an anti to a syn conformation and exposes its Hoogsteen edge to the shifted dT next to G*. Interestingly, the shifting of the dT 5′ to the lesion propagates to its neighboring dA at the replicating (+1) position, such that this dA also assumes a syn conformation to compensate for the shortened distances between the template and primer strands in BPG-2 (Fig. 3). As a result, the two base pairs after the looped-out adduct are Hoogsteen base pairs (Fig. 3*F*). The C1′–C1′ distances in the Hoogsteen base-pairs are 7.7 and 8.3 Å, respectively, shorter than the corresponding distance (≈10.5 Å) for a Watson–Crick base pair. The Hoogsteen base pairs accommodate the distortion caused by the template misalignment and permit the primer backbone to fit into the active site of the protein for extension. The incorporation of dT into the growing primer strand in BPG-2 indicates that the loop-out conformation is productive. In this state, the polymerase is in the process of moving along the template strand to the next base.

Template Misalignment and the 5′ Rule. The loop-out of BP–dG provides a structural basis for 5′ sequence-dependent base substitutions and −1 frameshift mutations induced by the major BP–dG adduct *in vivo* (11, 12, 29, 30). Once the lesion is looped out by template misalignment, primer extension proceeds by using the nucleotide 5′ to the lesion as template for incoming nucleotides. In consequence, the base 5′ determines the identity of the base incorporated opposite the lesion, which is referred as the 5′ rule (20). The mutations induced by BP–dG in *Escherichia coli* are >95% G → T in a 5′-T*G*C sequence and >95% G → A in a 5′-A*G*A sequence, respectively, which follow the 5′ rule and are Y-family polymerase-dependent (12, 29, 30). The 5′ rule is also clearly demonstrated by the present replication assays on templates that have four different bases 5′ to the BP–dG lesion (Fig. 1*C*). In the presence of a single dNTP, primer extension was most efficient when the incoming dNTP was complementary to the base 5′ to the adducted G* (Fig. 1*C*). For the 5′-T*G*A-3′ sequence in the crystal structures, the preference for A incorporation opposite the lesion is a result of looping-out of the modified base (BPG-2) and a subsequent strand realignment (BPG-1B). A −1 frameshift will result if the entire adducted residue loops out, and strand realignment does not occur (BPG-2). In addition, the pronounced 5′ dependency for base incorporation by Dpo4 opposite BP–dG has not been observed for Dpo4 with other types of lesions. Bypass of the minor cis-(*10R*)-BP-N^6-dA (BP–dA) adduct by Dpo4 strongly follows the A rule, which shows preferential insertion of dA opposite a lesion (21, 31). This is consistent with the Dpo4/BP–dA structure in which the dA* is directly paired with dA at the active site (21). With an abasic lesion, both dependences on the 5′-base and the A rule were observed, indicating possible direct insertion of dA opposite the lesion (20). In contrast, replication past the BP–dG lesion by Dpo4 strictly follows the 5′ rule, not the A rule (Fig. 1*C*). This mechanism clearly supports a link between base substitutions and frameshift mutations and suggests a straightforward explanation for 5′ sequence dependence in base substitutions induced by the bulky adducts in *E. coli* (2, 12, 30, 32).

Additional Substrate-Binding Site Provides a Previously Uncharacterized TLS Mechanism. Previous structural analyses of Dpo4 have attributed its lesion bypass ability to its open and solvent-accessible active site that can accommodate DNA lesions in productive conformations (19–21). No direct stabilization was provided by Dpo4 to the DNA lesions in the previous studies. For instance, Dpo4 bypasses a minor *cis*-(*10R*)-BP-N^6-dA adduct by placing the PAH moiety into the solvent-accessible major groove (N^6 is on the major-groove side). The BP–dA adduct is stabilized by packing one face of the PAH against base pairs in

the major groove (21). Loop-out bypass by Dpo4 was also observed for abasic lesions (20), but its structural basis differs from that for bypass of the present BP–dG adduct. In the structure of Dpo4 complexed with an abasic lesion (Ab-2A), the looped-out backbone is in the open active site and exposed to solvent (20). A small and hydrophilic abasic lesion (a sugar–phosphate unit without a base moiety) does not need to be shielded from water. In the absence of an enzyme, the hydrocarbon portion of the BP-N^2-dG adduct either intercalates with base pairs or lies in the minor groove (N^2 is on the minor groove side) in NMR structures (24–26). PAH intercalation blocks the primer extension past BP–dG in BPG-1A. The minor groove-bound bulky PAH in the NMR model would be in steric conflict with Dpo4 if it were put into the active site. A modeling study predicted either an altered Dpo4 structure with a wider gap between the palm and finger domains to accommodate the minor groove adduct or a DNA model with the PAH shifted into the major groove to fit BP–dG into the active site (33). In the present crystal structures, Dpo4 stabilizes this hydrophobic lesion in a pocket outside the active site. The stabilization of a productive conformation allows the adducted DNA and the incoming dNTP to fit into the catalytic site and permits bypass synthesis, which reveals a structural basis different from all previous observations for translesion synthesis. There is no significant difference among the Dpo4 conformations in the three structures here, which closely resemble previously determined structures of Dpo4. The rmsd on Cα superimposition is within 0.6 Å on overlapping the present three structures with the previous structures of Dpo4 in complex with undamaged or damaged DNA substrates (18–21). Dpo4 is in a ready-to-go state and does not need to undergo conformational changes to accept a variety of DNA substrates. The additional site is an intrinsic part of Dpo4 and is well positioned near the template–primer junction to harbor bulky lesions, potentially resulting in template misalignment, regardless of the specific DNA lesion. In contrast, a crystal structure of a DNA substrate containing the same major BP–dG adduct in complex with DNA polymerase I from *Bacillus stearothermophilus* shows that this high-fidelity polymerase has no room to accommodate a productive conformation of BP–dG-adducted DNA, so that replication was completely blocked by the lesion (34).

The LF domain is conserved in all Y-family polymerases, and the structure gap between the LF and the core domains has been observed in many Y-family polymerase structures (35). The human Y-family polκ bypasses the major BP adduct with correct insertion of C (15). Notably, efficiency, $(V_{max}/K_m)_{ins}$, for this correct insertion is increased by ≈4 to 11-fold when the base 5′ to the adduct is a G (36). To account for this observation, a misalignment mechanism whereby G* is looped out, and the 5′-neighboring G serves as a template for C insertion was suggested (36). Structural superposition of human Y-family polκ (37) and Dpo4 indicates that polκ may use the same additional binding site to harbor the looped-out adduct, which cannot fit into the active site without intercalation. Polκ could stabilize the adducted base better than Dpo4 by stacking the G base with Phe 171, which is equivalent to Lys 78 in Dpo4 (SI Fig. 5*C*). Polκ also has a unique "N-clasp" to encircle the DNA for tight DNA binding (37). These two structural features of polκ may contribute to its efficient bypass of the BP adduct. Furthermore, strand misalignment is a spontaneous process that induces frameshift mutations (38, 39). The frequency of frameshift mutations in replication depends on the stability of misaligned intermediates at a template–primer junction. The additional binding site that stabilizes extrahelical bases may also facilitate generation of frameshift mutations in DNA replication. The study reported here demonstrates a mechanism for translesion synthesis by a Y-family polymerase past bulky and hydrophobic DNA adducts and extends our understanding of carcinogen-induced mutagenesis.

Materials and Methods

Adducted Oligonucleotide. The template oligonucleotide (17-mer) used for the present crystal structures, 5′-(TC ATG* AAT CCT TCC CCC)-3′, with the trans (10*S*)-BP–dG adduct indicated by the asterisk and a dideoxycytidine residue at the 3′ terminus, was synthesized essentially as described (40). The adducted oligonucleotide was purified by reverse phase HPLC and annealed with a slight excess of its 13-mer primer strand, 5′-(GGG GGA AGG ATT A)-3′. The resulting duplex was separated from the excess primer by HPLC on a hydroxyapatite column (for further details see *SI Materials and Methods*).

Crystallization and Structure Determination. Dpo4 was expressed and purified as described (18). Crystals were produced by vapor diffusion. The details of crystallization are described in *SI Materials and Methods*. The diffraction data were collected at APS beamline 19-ID and CHESS beamline F-1. The data sets were processed with HKL (41). The structures were solved by molecular replacement with CNS (42) or PHASER (43) by using the type I structure as a search model (18). The structures were iteratively adjusted with O (44) and refined by using CNS and Refmac5 (45) with all residues in the allowed regions of the Ramachandran plot.

Primer Extension. Primer extension by Dpo4 was performed at 37°C with 10 nM DNA substrate with 5′ ^{32}P-labeled primer and 100 μM dNTPs (or as indicated in Fig. 1). The buffer used contained 40 mM Tris·HCl (pH 8.0), with 5 mM $MgCl_2$, 10 mM DTT, 0.35 mg/ml BSA, and 2.5% glycerol, along with 10 nM Dpo4. Oligonucleotides were prepared by preannealing of 150 nM template and 100 nM primer. Products were separated on 20% polyacrylamide 7 M urea gels and visualized by PhosphorImager analysis.

We thank Dr. R. Zhang (Structural Biology Center, Argonne National Laboratory, Argonne, IL) who collected BPG-1 data. The CHESS in which BPG-2 data were collected is supported by the National Science Foundation and the National Institutes of Health (NIH)/National Institute for General Medical Sciences under award DMR-0225180. We thank Dr. M. Cai (NIH) for some of the primer extension studies. This work was supported by the Terry Fox Foundation and the Intramural Research Program of the NIH (National Institute of Diabetes and Digestive and Kidney Diseases). H.L. acknowledges the New Investigators Award from the Canadian Institutes of Health Research.

1. Friedberg EC, Wagner R, Radman M (2002) *Science* 296:1627–1630.
2. Goodman MF (2002) *Annu Rev Biochem* 71:17–50.
3. Lehmann AR (2002) *Mutat Res* 509:23–34.
4. Committee on the Biological Effects of Atmospheric Pollutants (1972) (Natl Acad Sci, Washington DC).
5. Phillips DH (1999) *Mutat Res* 443:139–147.
6. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P (2002) *Oncogene* 21:7435–7451.
7. Thakker DR, Yagi H, Levin W, Wood AW, Conney AH, Jerina DM (1985) *Bioactivation of Foreign Compounds* (Academic, New York).
8. Jerina DM, Chadha A, Cheh AM, Schurdak ME, Wood AW, Sayer JM (1991) *Adv Exp Med Biol* 283:533–553.
9. Vock EH, Wolfe AR, Meehan T (2001) *Mutat Res* 478:199–206.
10. Koreeda M, Moore PD, Wislocki PG, Levin W, Conney AH, Yagi H, Jerina DM (1978) *Science* 199:778–781.
11. Lenne-Samuel N, Janel-Bintz R, Kolbanovskiy A, Geacintov NE, Fuchs RP (2000) *Mol Microbiol* 38:299–307.
12. Shukla R, Geacintov NE, Loechler EL (1999) *Carcinogenesis* 20:261–268.
13. Seo KY, Nagalingam A, Miri S, Yin J, Chandani S, Kolbanovskiy A, Shastry A, Loechler EL (2006) *DNA Repair (Amsterdam)* 5:515–522.
14. Chiapperino D, Kroth H, Kramarczuk IH, Sayer JM, Masutani C, Hanaoka F, Jerina DM, Cheh AM (2002) *J Biol Chem* 277:11765–11771.
15. Rechkoblit O, Zhang Y, Guo D, Wang Z, Amin S, Krzeminsky J, Louneva N, Geacintov NE (2002) *J Biol Chem* 277:30488–30494.
16. Zhang Y, Wu X, Guo D, Rechkoblit O, Geacintov NE, Wang Z (2002) *Mutat Res* 510:23–35.

17. Boudsocq F, Iwai S, Hanaoka F, Woodgate R (2001) *Nucleic Acids Res* 29:4607–4616.
18. Ling H, Boudsocq F, Woodgate R, Yang W (2001) *Cell* 107:91–102.
19. Ling H, Boudsocq F, Plosky BS, Woodgate R, Yang W (2003) *Nature* 424:1083–1087.
20. Ling H, Boudsocq F, Woodgate R, Yang W (2004) *Mol Cell* 13:751–762.
21. Ling H, Sayer JM, Plosky BS, Yagi H, Boudsocq F, Woodgate R, Jerina DM, Yang W (2004) *Proc Natl Acad Sci USA* 101:2265–2269.
22. Vaisman A, Ling H, Woodgate R, Yang W (2005) *EMBO J* 24:2957–2967.
23. Xie XM, Geacintov NE, Broyde S (1999) *Biochemistry* 38:2956–2968.
24. Cosman M, de los Santos C, Fiala R, Hingerty BE, Singh SB, Ibanez V, Margulis LA, Live D, Geacintov NE, Broyde S, Patel DJ (1992) *Proc Natl Acad Sci USA* 89:1914–1918.
25. Cosman M, Hingerty BE, Geacintov NE, Broyde S, Patel DJ (1995) *Biochemistry* 34:15334–15350.
26. Geacintov NE, Cosman M, Hingerty BEAS, Broyde S, Patel DJ (1997) *Chem Res Toxicol* 10:111–146.
27. Zhang N, Lin C, Huang X, Kolbanovskiy A, Hingerty BE, Amin S, Broyde S, Geacintov NE, Patel DJ (2005) *J Mol Biol* 346:951–965.
28. Steitz TA (1999) *J Biol Chem* 274:17395–17398.
29. Mackay W, Benasutti M, Drouin E, Loechler EL (1992) *Carcinogenesis* 13:1415–1425.
30. Yin J, Seo KY, Loechler EL (2004) *DNA Repair (Amsterdam)* 3:323–334.
31. Strauss BS (2002) *DNA Repair (Amsterdam)* 1:125–135.
32. Rodriguez H, Loechler EL (1995) *Mutat Res* 326:29–37.
33. Perlow-Poehnelt RA, Likhterov I, Scicchitano DA, Geacintov NE, Broyde S (2004) *J Biol Chem* 279:36951–36961.
34. Hsu GW, Huang X, Luneva NP, Geacintov NE, Beese LS (2005) *J Biol Chem* 280:3764–3770.
35. Yang W (2005) *FEBS Lett* 579:868–872.
36. Huang X, Kolbanovskiy A, Wu X, Zhang Y, Wang Z, Zhuang P, Amin S, Geacintov NE (2003) *Biochemistry* 42:2456–2466.
37. Lone S, Townson SA, Uljon SN, Johnson RE, Brahma A, Nair DT, Prakash S, Prakash L, Aggarwal AK (2007) *Mol Cell* 25:601–614.
38. Streisinger G, Owen J (1985) *Genetics* 109:633–659.
39. Garcia-Diaz M, Kunkel TA (2006) *Trends Biochem Sci* 31:206–214.
40. Kroth H, Yagi H, Sayer JM, Kumar S, Jerina DM (2001) *Chem Res Toxicol* 14:708–719.
41. Otwinowski Z, Minor W (1997) *Methods Enzymol* 276:307–326.
42. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, *et al.* (1998) *Acta Crystallogr D* 54:905–921.
43. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) *Acta Crystallogr D* 61:458–464.
44. Jones TA, Zou JY, Cowan SW (1991) *Acta Crystallogr A* 47:110–119.
45. Winn MD, Isupov MN, Murshudov GN (2001) *Acta Crystallogr D* 57:122–133.
46. DeLano WL (2002) *Pymol* (DeLano Scientific, San Carlos, CA).