DNA Adduct Bypass Polymerization by *Sulfolobus solfataricus* DNA Polymerase Dpo4

ANALYSIS AND CRYSTAL STRUCTURES OF MULTIPLE BASE PAIR SUBSTITUTION AND FRAMESHIFT PRODUCTS WITH THE ADDUCT 1,N^2-ETHENOGUANINE*⑤

Hong Zang‡§, Angela K. Goodenough¶‖, Jeong-Yun Choi‡§, Adriana Irimia‡§, Lioudmila V. Loukachevitch‡§, Ivan D. Kozekov§¶, Karen C. Angel‡§, Carmelo J. Rizzo§¶, Martin Egli‡§, and F. Peter Guengerich‡§**

From the Departments of ‡Biochemistry and ¶Chemistry and the §Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146

1,N^2-Etheno(ϵ)guanine is a mutagenic DNA lesion derived from lipid oxidation products and also from some chemical carcinogens. Gel electrophoretic analysis of the products of primer extension by *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) indicated preferential incorporation of A opposite 3′-(1,N^2-ϵ-G)TACT-5′, among the four dNTPs tested individually. With the template 3′-(1,N^2-ϵ-G)CACT-5′, both G and A were incorporated. When primer extension was done in the presence of a mixture of all four dNTPs, high pressure liquid chromatography-mass spectrometry analysis of the products indicated that (opposite 3′-(1,N^2-ϵ-G)CACT-5′) the major product was 5′-GTGA-3′ and the minor product was 5′-AGTGA-3′. With the template 3′-(1,N^2-ϵ-G)TACT-5′, the following four products were identified by high pressure liquid chromatography-mass spectrometry: 5′-AATGA-3′, 5′-ATTGA-3′, 5′-ATGA-3′, and 5′-TGA-3′. An x-ray crystal structure of Dpo4 was solved (2.1 Å) with a primer-template and A placed in the primer to be opposite the 1,N^2-ϵ-G in the template 3′-(1,N^2-ϵ-G)TACT 5′. The added A in the primer was paired across the template T with classic Watson-Crick geometry. Similar structures were observed in a ternary Dpo4-DNA-dATP complex and a ternary Dpo4-DNA-ddATP complex, with d(d)ATP opposite the template T. A similar structure was observed with a ddGTP adjacent to the primer and opposite the C next to 1,N^2-ϵ-G in 3′-(1,N^2-ϵ-G)CACT-5′. We concluded that Dpo4 uses several mechanisms, including A incorporation opposite 1,N^2-ϵ-G and also a variation of dNTP-stabilized misalignment, to generate both base pair and frameshift mutations.

The accurate replication of DNA is important in the maintenance of the genetic integrity of an organism, and DNA polymerases have biochemical mechanisms to copy with very high fidelity (1, 2). DNA is constantly subject to damage by physical and chemical agents, including compounds produced endogenously and those derived from the environment (3). More than 130 human proteins are involved in the repair of DNA damage (4). However, some damage escapes these surveillance systems, and DNA polymerases are forced to encounter multiple types of damaged DNA. The polymerases can be blocked, leading to cell cycle arrest or death, or can proceed to copy past the modified DNA. Copying past a DNA adduct raises the prospect of mutation. Mutation is a significant factor in the carcinogenicity of chemicals, and mutagens are generally regarded as cancer suspects (5).

1,N^2-ϵ-G,[1] originally reported by Leonard and co-workers (6), is a prototype for the so-called exocyclic DNA adducts, a series of lesions formed by various bis-electrophiles (7). Included in the list of compounds that generate etheno adducts are electrophilic products of industrial vinyl monomers (*e.g.* epoxides formed by oxidation of vinyl chloride, urethane, etc. (8, 9)), some nitrosamines (10), mucochloric acid (11), and also products of the peroxidation of endogenous lipids (12–14). 1,N^2-ϵ-G produces misincorporation and mutations in systems involving model DNA polymerases and bacterial and mammalian cells (15–17). 1,N^2-ϵ-G is an interesting lesion in that the two added exocyclic carbons block the normal coding "face" involved in Watson-Crick pairing and preclude the usual hydrogen bonding mode (Scheme 1). Most interestingly, the frequency of mutations is low in bacterial and mammalian cells in which the 1,N^2-ϵ-G lesion is placed (16, 17). This low frequency of errors may be attributed to the presence of bacterial and mammalian glycosylases that act on this lesion (18).

Until recently, biochemical studies on the mechanisms of misincorporation have been largely focused on replicative and other available DNA polymerases (2). However, many mutagens were known to be most effective in bacteria when the SOS response system was activated (19). The characterization of DNA damage-inducible polymerases in bacteria led to a better understanding of the mutagenic mechanism (20), and now

1,N^2-ε-G

A
```
5'GGGGGAAGGATTC        3' + dATP or dCTP
3'CCCCCTTCCTAAG XZACT 5'
```

B
```
5'GGGGGAAGGATTC A      3'  +  dATP
3'CCCCCTTCCTAAG XZACT 5'
```

C
```
5'GGGGGAAGGATTC AA     3'  + dATP or dTTP
3'CCCCCTTCCTAAG XZACT 5'
```
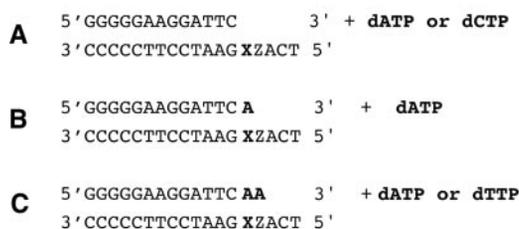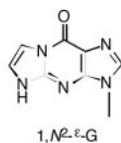
X = G or 1,N^2-ε-G
Z = T or C

SCHEME 1. **Oligonucleotides used for steady-state kinetic assays (Table II, Z = T; Table III, Z = C).** *A*, insertion experiments. *B*, next-base extension experiment. *C*, second-base extension experiment.

these so-called translesion DNA polymerases are considered to have major roles in the replication past DNA damage, both "error-free" and "error-prone," in eubacteria, eukaryotes, and archebacteria (21–25). Some of these translesion polymerases have been studied in considerable detail in recent years, in regard to function, structure, and mechanism (26–29).

One archebacter translesional DNA polymerase that has been used as a model of eukaryotic pol η and as a prototype is *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) (24, 30). Several crystal structures have been reported, including some with damaged DNA (31–34), and pre-steady-state kinetic analysis has helped define a general catalytic mechanism (for undamaged DNA) (35, 36). A number of questions still exist about the interactions of 1,N^2-ε-G with DNA polymerases, in that different misincorporations and frameshifts have been observed with individual polymerases, and in some cases undefined mixtures of products are obtained with a single polymerase (15, 16). In cells, complex biological effects are also observed (16, 17). The availability of Dpo4 provided opportunities to study the incorporation of dNTPs in a simple system and to obtain structures of crystals of ternary complexes (of the polymerase, primer-template complex, and dNTP).

We report a series of steady-state kinetic and other primer-template studies with Dpo4 and 1,N^2-ε-G in two template sequences. LC-MS/MS methods were used to define the products and provide insight into the multiple bypass mechanisms Dpo4 can use. Four crystal structures of Dpo4:1,N^2-ε-G oligonucleotide complexes provide direct visualization of some of the intermediates, particularly the −1 frameshift mechanism. This latter intermediate has normal hydrogen bonding to the base 5′ of the 1,N^2-ε-G but no distortion of the adducted oligonucleotide.

EXPERIMENTAL PROCEDURES

Materials—*Escherichia coli* UDG was purchased from Sigma. pol T7⁻ (exonuclease⁻) was expressed in *E. coli* and purified as described elsewhere (37–39).

Unmodified oligonucleotides were purchased from Midland Certified Reagents (Midland, TX). Oligonucleotides containing 1,N^2-ε-G were prepared in the Vanderbilt facility from a phosphoramidite derivative of 1,N^2-ε-G, using a modification of procedures described previously (15, 18). Purification was done by using HPLC (reversed-phase) (15).

Isolation of S. solfataricus P2 DNA Polymerase IV (Dpo4) Gene and Construction of Dpo4 Expression Plasmid—The *dpo4* gene was obtained by PCR amplification from the genomic DNA of *S. solfataricus* P2 (American Tissue Culture Collection, Manassas, VA; catalog number ATCC 35092) as template using *Pfu ultra* DNA polymerase (Stratagene, La Jolla, CA) with two primers, 5′-CATATGCATCACCATCAC-CATCACATGATTGTTCTTTTCGTTG-3′ and 5′-GGATCCTTAAGTA-

TCGAAGAACTTG-3′. The resulting 1.1-kb *Dpo4* gene PCR product was cloned into the vector pPCR-Script Amp (Stratagene, La Jolla, CA), and the nucleotide sequence was determined (Vanderbilt facility) to confirm the coding region. The 1.1-kb *Dpo4* gene fragment was then cloned into the NdeI and XhoI sites of the vector pET-22b(+) (Novagen, Madison, WI), generating the vector pET22b(+)/DPO4-NHis.

Expression and Purification of Dpo4—Dpo4 fused to an N-terminal His₆ tag was expressed in *E. coli* strain BL21(DE3). *E. coli* BL21(DE3) harboring the pET22b(+)/DPO4-NHis vector (6 liters) was grown in Luria-Bertani broth supplemented with ampicillin (100 μg ml⁻¹) at 37 °C, with aeration, to an A_{600} of 0.6. Isopropyl β-D-thiogalactopyranoside was added to 1 mM, and the incubation was continued for 3 h. The cells were harvested by centrifugation and resuspended in 60 ml of lysis buffer (50 mM Tris-HCl, pH 7.4, containing 300 mM NaCl, 10% glycerol (v/v), 5 mM β-mercaptoethanol, 1 mg of lysozyme ml⁻¹, and protease inhibitor mixture) (Roche Applied Science), cooled on ice for 30 min, and then lysed by sonication (12 × 10 s duration with a Branson digital sonifier (VWR, West Chester, PA), microtip, 45% amplitude, with intervening cooling time). The cell lysate was clarified by centrifugation at $4 \times 10^4 \times g$ for 45 min at 4 °C. The resulting supernatant was heated at 80 °C for 10 min and centrifuged to remove the denatured proteins; the soluble and heat-stable proteins were collected (including Dpo4).

The soluble supernatant was applied to a 5-ml HisTrap column (Amersham Biosciences) and washed sequentially with 50 ml of buffer A (50 mM Tris-HCl, pH 7.4, containing 300 mM NaCl, 10% glycerol (v/v), 5 mM β-mercaptoethanol) containing 20 mM imidazole, 50 ml of buffer A containing 40 mM imidazole, and 50 ml of buffer A containing 50 mM imidazole. Bound proteins were eluted with a linear gradient of 50–400 mM imidazole. Eluted proteins (detected at 280 nm) were dialyzed against buffer B (50 mM Tris-HCl, pH 7.5, containing 0.5 mM EDTA, 10% glycerol (v/v), and 5 mM β-mercaptoethanol) and loaded onto a 1-ml MonoS column (Amersham Biosciences) that had been equilibrated with buffer B. The column was washed with 20 ml of buffer B containing 100 mM NaCl to remove the contaminating proteins, and the protein was eluted with a 30-ml linear gradient of 100–500 mM NaCl in buffer C. The eluted fractions (280 nm absorbance) were analyzed by SDS-PAGE and silver staining (40). Dpo4 was eluted at 200 mM NaCl. By using this protocol, Dpo4 was purified to electrophoretic homogeneity, based on staining (see Supplemental Material). Protein concentration was determined using a calculated ϵ_{280} value of 22 mM⁻¹ cm⁻¹ for Dpo4 (41). The final yield of purified Dpo4 was ~10 mg.

Polymerization Assays and Gel Electrophoresis—A ³²P-labeled primer, annealed to either an unmodified or adducted template, was extended in the presence of single dNTPs (Scheme 1). Each reaction was initiated by adding 2 μl of dNTP-Mg²⁺ solution (final concentrations of 250 μM of each dNTP and 5 mM MgCl₂) to a preincubated *E*·DNA complex (final concentrations of 50 mM Tris-HCl, pH 7.8, 100 nM DNA duplex, 100 nM Dpo4 or pol T7⁻, 1 mM DTT, 50 μg of BSA ml⁻¹, 50 mM NaCl, and 10% glycerol (v/v))² at 37 °C, yielding a total reaction volume of 8 μl. When pol T7⁻ was used, *E. coli* thioredoxin was added in a 20-fold molar excess of pol T7⁻. After 30 min, reactions were quenched with 50 μl of 20 mM EDTA (pH 9.0) in 95% formamide (v/v). Aliquots (3 μl) were separated by electrophoresis on a denaturing gel containing 8.0 M urea and 16% acrylamide (w/v) (from a 19:1 acrylamide:bisacrylamide solution, AccuGel, National Diagnostics, Atlanta, GA) with 80 mM Tris borate buffer, pH 7.8, containing 1 mM EDTA. The gel was exposed to a PhosphorImager screen (Imaging Screen K, Bio-Rad) overnight. The bands (representing extension of the primer) were visualized with a PhosphorImaging system (Bio-Rad, Molecular Imager® FX) using the manufacturer's Quantity One software, version 4.3.0.

Steady-state Kinetics—Unless indicated otherwise, all Dpo4 reactions were performed at 37 °C in 50 mM Tris-HCl buffer, pH 7.8, containing 10% glycerol (v/v), 5 mM DTT, 50 mM NaCl, and 50 μg of BSA ml⁻¹.² For unmodified templates, the molar ratio of primer-template to enzyme was 40:1, and the reactions were done at eight dNTP concentrations (reaction time of 3 min). For modified templates, the molar ratio of primer-template to enzyme was usually 10:1, and reactions were done at eight dNTP concentrations (reaction time of 5 min).

LC-MS Analysis of Oligonucleotide Products from Dpo4 Reactions—Dpo4 reactions were performed at 37 °C for 6 h in 50 mM Tris-HCl buffer, pH 7.8, containing 5% glycerol (v/v), 5 mM DTT, 50 mM NaCl, 5 mM MgCl₂, and 100 μg of BSA ml⁻¹. The reactions were done with 10 μM oligonucleotide substrate, 5 μM Dpo4, and four dNTPs at 1 mM each,

---

² In some cases, alternative conditions were used, with 50 mM sodium MOPS (pH 7.5), 10 mM NaCl, 5 mM DTT, and 100 μM dNTP (no BSA), and similar results were obtained.

TABLE I

Crystal data and refinement parameters for the binary Dpo4-DNA and the three ternary Dpo4-DNA-d(d)NTP complexes

Parameter	Dpo4-1	Dpo4-2	Dpo4-3	Dpo4-4
X-ray source	ID-22 (SER-CAT/APS)	ID-22 (SER-CAT/APS)	ID-22 (SER-CAT/APS)	ID-5 (DND-CAT/APS)
Wavelength (Å)	0.999	0.991	0.999	0.920
Temperature (K)	110	110	110	110
No. crystals	1	1	1	1
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell (a, b, c; Å)	93.67, 102.85, 53.17	95.52, 103.02, 53.09	95.35, 103.02, 52.98	93.50,102.07, 52.53
Resolution range (Å)	46.84–2.0	42.32–2.37	42.25–2.5	29.81–2.17
Highest resolution shell	(2.13–2.0)	(2.52–2.37)	(2.66–2.5)	(2.31–2.17)
No. measurements	221,493	165,987	86,276	178,232
No. unique reflections	35,290	20,357	17,022	26,856
Redundancy	6.3	8.2	5.1	6.6
Completeness (%)	97.6 (87.3)[a]	92.6 (65.9)	90.7 (73.7)	98.2 (95.9)
R_{merge}[b] (%)	4.9	7.7	5.5	6.9
Signal to noise ($\langle I/\sigma^I \rangle$)	16.2	10.6	7.2	17.8
Solvent content (%)	54.0	52.26	52.63	54.49
Model composition (asymmetric unit)				
No. amino acid residues	341	341	341	341
No. protein atoms (non-H)	2816	2795	2808	2775
No. water molecules	166	109	116	118
No. Ca^{2+} ions	3	3	3	3
No. template nucleotides	17	17	17	17
No. primer nucleotides	14	13	13	13
No. ddATP			1	
No. dATP		1		
No. ddGTP				1
R_f[c] (%)	23.9	22.3	22.1	23.9
R_{free}[d] (%)	27.0	26.0	24.8	26.6
Estimated coordinate error (Å)				
From Luzatti plot	0.32	0.37	0.37	0.32
From Luzatti plot (cross-validation)	0.37	0.43	0.43	0.37
From σ_A plot	0.32	0.45	0.41	0.31
From σ_A plot (cross-validation)	0.34	0.52	0.35	0.30
Temperature factor from Wilson plot (Å^2)	46.9	65.3	60.8	50.0
Mean isotropic temperature factor (Å^2)	48.4	56.6	54.8	46.3
r.m.s. deviation in temperature factors				
Bonded main chain atoms (Å^2)	1.47	1.36	1.36	1.38
Bonded side chain atoms (Å^2)	2.13	2.10	2.03	2.11
r.m.s. deviation from ideal values				
Bond lengths (Å)	0.007	0.006	0.007	0.008
Bond angles (°)	1.33	1.16	1.17	1.42
Dihedral angles (°)	22.17	21.9	21.96	22.08
Improper angles (°)	1.07	0.97	0.99	1.05
Ramachandran plot (no. residues excluding Pro and Gly)				
Favored regions	281	273	272	283
Allowed regions	28	38	39	26
Unfavored regions	2			2

[a] Values in parentheses correspond to the highest resolution shells.

[b] $R_{merge} = \Sigma_{hkl} \Sigma_{j=1,N} |\langle I_{hkl} \rangle - I_{hkl,j}| / \Sigma_{hkl} \Sigma_{j=1,N} I_{hkl,j}$, 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.

in a final reaction volume of 100 μl. The reaction was terminated by extraction of excess dNTPs using a spin column (Bio-Spin 6 chromatography column, Bio-Rad). To the above filtrate (120 μl), concentrated Tris-HCl, DTT, and EDTA were added to restore the initial concentrations, and UDG solution was added (20 units). The reaction was incubated at 37 °C for 6 h to hydrolyze the uracil residues on the primer. The final reaction mixture was heated at 95 °C for 1 h in the presence of 0.25 M piperidine, followed by lyophilization. The dried residues were dissolved in 100 μl of H_2O for the following MS analysis.

MS was performed on a DecaXP ion trap instrument (ThermoFinnigan, San Jose, CA). Separation of oligonucleotide was carried out with a Jupiter microbore column (1.0 mm $\times$ 150 mm, 5 μm, Phenomenex, Torrance, CA). Buffer A contained 10 mM $NH_4CH_3CO_2$, pH 6.8, and 2% CH_3CN (v/v); buffer B contained 10 mM $NH_4CH_3CO_2$, pH 6.8, and 95% CH_3CN (v/v). The following gradient program was used with a flow rate of 1.0 ml min^{-1}: 0–2 min, hold at 100% A; 2–20 min, linear program to 100% B; 20–30 min, hold at 100% B; 30–31 min, linear program to 100% A; 31–40 min, hold at 100% A (for next injection). The desired oligonucleotide products were eluted at $\sim$ 7 min. A pre-column "Tee" set-up was applied, with only 10% of the total flow infused to the ion source. The fast flow rate helped reduce the equilibration time of the

column, and thus retention time of the oligonucleotides was consistent between runs. Samples were infused using an autosampler, with 8 μl withdrawn from a 50-μl reaction. ES conditions were as follows: source voltage 3.4 kV, source current 8.5 μA, sheath gas flow rate setting 28.2, auxiliary sweep gas flow rate setting 4.3, capillary voltage 49 V, capillary temperature 230 °C, tube lens voltage 67 V. MS/MS conditions were as follows: normalized collision energy 35%, activation Q 0.250, time 30 min, 1 scan. Product ion spectra were acquired over the range m/z 250–2000. The abundant ions from LC-MS spectra were selected for CID analysis, and the cut-off was set above $\sim$15% of the most abundant ion. When more than one ion came from a single species, the peak responding to the doubly charged parent ion was chosen for fragmentation. The calculations of the CID fragmentations of a certain oligonucleotide sequence were done using a program linked to the Mass Spectrometry Group of Medicinal Chemistry at the University of Utah (medlib.med.utah.edu/masspec/).

Crystallization of Dpo4-DNA Complexes—Primer and template sequences that were used in the crystallizations were similar to those utilized by Yang and co-workers (32, 34) as follows: 5'-GGG GGA AGG ATT CA-3' (14-mer primer; binary Dpo4-1 complex; see Table IV), 5'-GGG GGA AGG ATT C-3' (13-mer primer; ternary Dpo4-2, -3, and -4

```
5′ GGGGGCGATGCTCGTAAGGATTC        3′
3′ CCCCCGCTACGAGCATTCCTAAGXZACT 5′
```
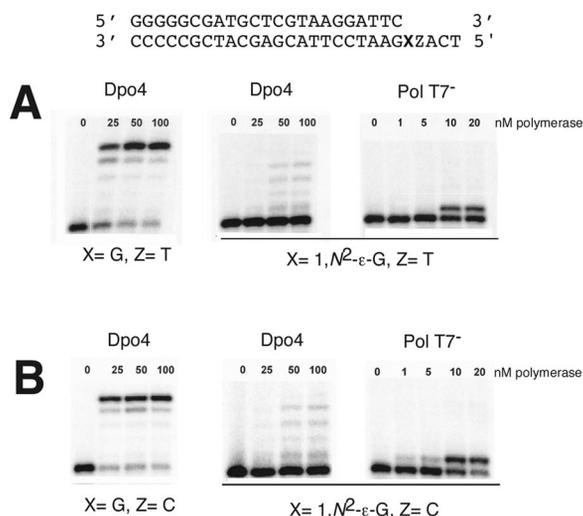

FIG. 1. Extension of ^{32}P-labeled primers by Dpo4 and pol T7$^-$ in the presence of all four dNTPs. Primers (13-mers, Scheme 1*A*) were annealed with a template containing either *Z* = G or 1,N^2-ε-G (Scheme 1) and either (*A*) *Z* = T or (*B*) *Z* = C at the position on the template 5′ to *X*. Reactions in each set were done for the same time period with the indicated concentration (in nM) of Dpo4 or pol T7$^-$. The concentration of primer-template complex was 100 nM, and the concentration of each dNTP was 250 µM. The reaction products were analyzed by denaturing gel electrophoresis with PhosphorImaging analysis. *A,* the percentage of total extended products was 76, 79, and 82% with *X* = G using 25, 50, and 100 nM Dpo4, respectively. The corresponding values with *X* = 1,N^2-ε-G and Dpo4 were 8, 16, and 17%, respectively. With *X* = 1,N^2-ε-G and 1, 5, 10, and 20 nM pol T7$^-$, the values were 2, 4, 32, and 33%, respectively. *B,* the percentage of total extended products was 65, 79, and 80% with *X* = G using 25, 50, and 100 nM Dpo4, respectively. The corresponding values with *X* = 1,N^2-ε-G and Dpo4 were 13, 20, and 23%, respectively. With *X* = 1,N^2-ε-G and 1, 5, 10, and 20 nM pol T7$^-$, the values were 13, 15, 53, and 62%, respectively.

```
5′ GGGGGCGATGCTCGTAAGGATTC3′
3′ CCCCCGCTACGAGCATTCCTAAG1,N²-ε-GCACT 5′
```
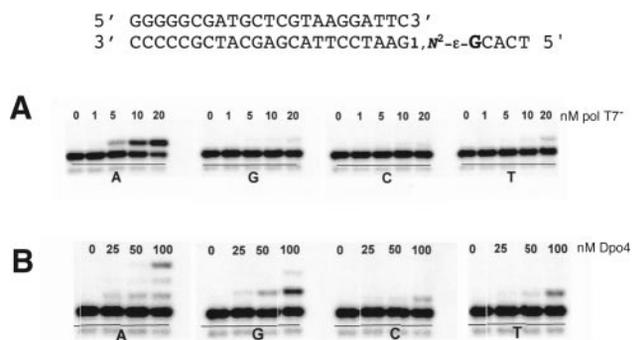

FIG. 2. Extension of a ^{32}P-labeled primer opposite template adduct by pol T7$^-$ and Dpo4 in the presence of single dNTPs. The concentrations of each polymerase used are indicated. *A,* the percentage of extension with 20 nM pol T7$^-$ was 59, 4, 5, and 9% for A, G, C, and T, respectively. *B,* the percentage of extension with 100 nM Dpo4 was 23, 26, 8, and 11% for A, G., C, and T, respectively.

complexes), 5′-TCA T(1,N^2-ε-G)G AAT CCT TCC CCC-3′ (18-mer template 1; Dpo4-1, -2, and -3 complexes), and 5′-TCA C(1,N^2-ε-G)G AAT CCT TCC CCC-3′ (18-mer template 2; Dpo4-4 complex). Annealed template 1 and 14-mer primer were used to make the binary complex with Dpo4 (Dpo4-1). Ternary complexes were prepared by annealing either template 1 (Dpo4-2 and Dpo4-3) or template 2 (Dpo4-4) and 13-mer primer, mixing the DNA with Dpo4, and adding either dATP and Ca^{2+} (Dpo4-2), ddATP and Mg^{2+} (Dpo4-3), or ddGTP and Ca^{2+} (Dpo4-4). Polymerase and the DNA duplex were mixed in a 1:1.2 molar ratio and incubated in 1 mM dATP, ddATP, or ddGTP in 20 mM Tris-HCl buffer, pH 7.5, containing 60 mM NaCl, 5 mM MgCl$_2$ (ddATP), or CaCl$_2$ (dATP and ddGTP), 1.0 mM tris(carboxyethyl)phosphine, and 2.5% glycerol (v/v) for 10 min at room temperature. The final protein concentration was kept at 5 mg ml^{-1}. Crystals were obtained by the hanging-drop vapor diffusion method at 20 °C. The well solution contained 10% polyethylene glycol 3350, 20 mM Tris-HCl, pH 7.5, and 100 mM

```
5′  GGGGGCGATGCTCGTAAGGATTC              3′
3′  CCCCCGCTACGAGCATTCCTAAG1,N²-ε-GTACT 5′
```
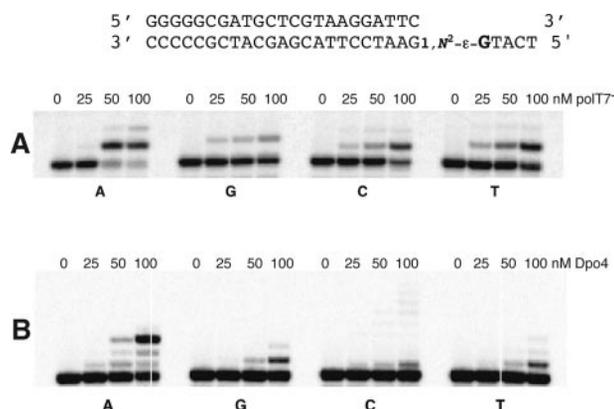

FIG. 3. Extension of a ^{32}P-labeled primer opposite the template adduct by pol T7$^-$ or Dpo4 in the presence of single dNTPs. The concentrations of each polymerase used are indicated. *A*, pol T7$^-$. *B*, Dpo4.

TABLE II
Steady-state kinetic parameters for 1-base incorporation by Dpo4 (Z = T, Scheme 1)

For G*, 1,N^2-ε-G.

dNTP inserted	Template	K_m	k_{cat}	k_{cat}/K_m	f^a
		µM	min^{-1}		
dCTP	5′- 3′-**GTA**-	4.7 ± 0.4	6.4 ± 0.13	1.4	
dCTP	5′- 3′-**G***TA-	96 ± 14	0.0057 ± 0.0002	0.6×10^{-4}	
dATP	5′- 3′-**GTA**-	33 ± 8	0.12 ± 0.01	0.0036	0.0026
dATP	5′- 3′-**G***TA-	33 ± 5	0.027 ±0.001	0.0008	14
dATP	5′-A 3′-G-**TA**-	46 ± 6	0.15 ± 0.01	0.0033	
dATP	5′-A 3′-G**G***TA-	200 ± 27	0.18 ± 0.01	0.0009	
dATP	5′-A A 3′-G**T***TA-	88 ± 11	0.29 ± 0.01	0.0033	
dTTP	5′-A A 3′-G**T***T**A**-	140 ± 22	2.0 ± 0.15	0.014	

^{a} Where f (misinsertion frequency) = $(k_{cat}/K_{mm})_{incorrect}/(k_{cat}/K_m)_{correct}$.

Ca(CH$_3$CO$_2$)$_2$. Crystals were transferred to mother liquor containing 25% polyethylene glycol 3350 (w/v) and 15% ethylene glycol (v/v) and were swiped through paratone-N before being flash-frozen in liquid N$_2$.

X-ray Diffraction Data Collection and Processing—Diffraction data sets for crystals of the Dpo4-1, -2, and -3 complexes were collected at a wavelength of 0.99 Å on the ID-22 SER-CAT beam line at the Advanced Photon Source (Argonne, IL). Data for the Dpo4-4 complex crystal were collected at a wavelength of 0.92 Å on the ID-5 DND-CAT beam line at the Advanced Photon Source. The resolutions for the data of the Dpo4-1, -2, -3, and -4 complexes were 2.0, 2.37, 2.5, and 2.17 Å, respectively (Table I). Diffraction data for the Dpo4-1, -2, and -3 complexes were indexed and scaled using the program HKL2000 (42), and those of the Dpo4-4 complex were processed with the program XDS (43). Autoindexing and systematic absences indicated P2$_1$2$_1$2 space group in all four cases. The data were further processed for each structure using CCP4 package programs, including the truncate procedure performed with TRUNCATE (44). The resulting data sets are of good quality with R_{merge} values ranging between 4.9 and 7.7%. Statistics of data processing and data quality are summarized in Table I.

Structure Determination and Refinement—The Dpo4-1 structure was solved by molecular replacement with the program AMoRe (45), using the 2.8 Å resolution structure of a Dpo4-DNA complex (Protein Data Bank accession code 1s0n (34)), devoid of all side chains and solvent molecules, as the search model. The positioned model gave a correlation coefficient between observed and calculated structure factor amplitudes of 0.469 at 3.5 Å resolution. The location of the Dpo4-1 model with side chains added was optimized by several rounds of rigid body refinement, including data of increasing resolution until the diffraction limit was reached. Due to the similar values of the unit cell parameters of the four crystal structures reported here (Table I), the refined structure of the

TABLE III

Steady-state kinetic parameters for 1-base incorporation by Dpo4
(Z = C, Scheme 1)

For G*, 1,N^2-ϵ-G.

dNTP inserted	Template	K_m	k_{cat}	k_{cat}/K_m	f^a
		μM	min^{-1}		
dCTP	5'- 3'-GCA-	0.028 ± 0.003	0.53 ± 0.02	19	
	5'- 3'-G*CA-	430 ± 94	0.028 ± 0.002	6.5×10^{-5}	
dATP	5'- 3'-GCA-	230 ± 42	0.090 ± 0.006	3.9×10^{-4}	2.0×10^{-5}
	5'- 3'-G*CA-	270 ± 46	0.056 ± 0.004	2.1×10^{-4}	3.2
dGTP	5'- 3'-GCA-	78 ± 6	0.053 ± 0.001	6.8×10^{-4}	3.4×10^{-5}
	5'- 3'-G*CA-	49 ± 11	0.072 ± 0.006	1.5×10^{-3}	22

^{a} f (misincorporation) = (k_{cat}/K_m)incorrect/(k_{cat}/K_m)correct.

Dpo4-1 protein-DNA complex without solvent molecules was used as a starting model for the Dpo4-2, -3, and -4 structures. The positions of the models were then optimized by several rounds of rigid body refinement by using data of increasing resolution. Manual model rebuilding was carried out with the TURBO-FRODO program (57) into σA maps computed using modified σA coefficients (46). The initial difference Fourier maps showed clear negative density (higher than 5.0 r.m.s. deviation) for the Ca^{2+} ions as well as for the nucleotides used for crystallization (*i.e.* dATP for Dpo4-2, ddATP for Dpo4-3, and ddGTP for Dpo4-4). Unambiguous density was also observed for the DNA duplex except for the first three template nucleotides (5'-TCA) that were disordered in all four structures. Therefore, in the absence of density, the 5'-terminal T of the template nucleotide was completely omitted from the models. The adjacent C2 has lower occupancy in all four structures, but A3 has lower occupancy only in Dpo4-3 and Dpo4-4.

Refinements were performed using the program CNS (47). To calculate the cross-validation residual R_{free}, 5% of amplitudes were excluded from refinement. During the model rebuilding steps, water oxygen atoms were added into positive regions of $(F_o - F_c)$ Fourier difference electron density (higher than 3.0 standard deviation). They were kept in the model after the final refinement round only if they explained the σA density maps and formed at least one hydrogen bond with any other atom in the model. Current statistics of the refined models for all four structures are presented in Table I.

The quality and stereochemistry of the four models were examined with standard procedures in CNS (47) and PROCHECK (48). All the crystallographic figures were prepared using CHIMERA (49).

Atomic coordinates and measured structure factor amplitudes for the Dpo4-1, -2, -3, and -4 complexes have been deposited in the Protein Data Bank with accession codes 2bq3, 2bqr, 2bqu, and 2br0, respectively.

RESULTS

Extension of Oligonucleotide Primers by Dpo4 in the Presence of All Four dNTPs—The template sequences selected for use here were based on crystallography work done by Ling *et al.* (31–34). Separate experiments were done with the pyrimidines T and C placed 5' to the 1,N^2-ϵ-G (Scheme 1A).

As expected from previous work with 1,N^2-ϵ-G (15), the model replicative DNA polymerase pol T7$^-$ was blocked considerably in its ability to extend past this DNA adduct (Fig. 1). The major product involved a 1-base extension, *i.e.* insertion of one dNTP opposite 1,N^2-ϵ-G. Dpo4 was not particularly efficient but did extend the primer to the apparent end of the template (Fig. 1). In the extension, several bands were seen, as a result of the distributive character of this translesion polymerase or due to frameshifts.

Insertion of Individual dNTPs by Dpo4—With the template containing a C residue 5' to 1,N^2-ϵ-G (Scheme 1A), pol T7$^-$ showed a tendency to insert a single A, presumably opposite the 1,N^2-ϵ-G adduct (Fig. 2A). Some similarity was observed in the pattern with Dpo4 (Fig. 2B), although (i) more incorporation of dGTP was detected (possibly opposite the C residue 5' of the
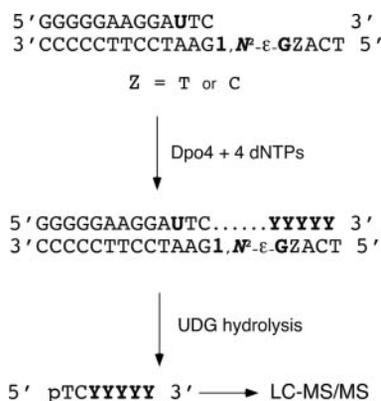
SCHEME 2. **LC-MS/MS method for sequencing primer extension products (Figs. 4–11).**

1,N^2-ϵ-G and suggesting a deletion in the product), and (ii) with both dATP and dGTP a tendency was seen for multiple incorporations.

By using the template with a T residue 5' of the 1,N^2-ϵ-G, pol T7$^-$ incorporated a single base in the apparent order of preference A $\gg$ T > C > G (Fig. 3A). With Dpo4 some incorporation of a single G, C, or T was observed, but the preferred reaction was the incorporation of three As (Fig. 3B). This latter result was unexpected and would appear to correspond to the insertion of an A opposite the 1,N^2-ϵ-G, the T, and the A (Scheme 1A). Some of this tendency was also seen with pol T7$^-$ under these conditions (Fig. 3A) but was less pronounced.

Steady-state Kinetic Analysis of dNTP Incorporation—The incorporation of individual dNTPs was analyzed in more detail using Dpo4 and the template with a T residue 5' of the 1,N^2-ϵ-G adduct (Scheme 1A). The individual experiments are outlined in Scheme 1 and Table II, with the results presented in Table II.

"Correct" incorporation of dCTP opposite 1,N^2-ϵ-G was very unfavorable, with an apparent catalytic efficiency 20,000-fold less than the incorporation of dCTP opposite G (Table II, 1st and 2nd rows). As indicated in Fig. 3B, A was the base preferentially incorporated into the primer. Three As were incorporated sequentially, even under limiting conditions, and we were unable to cleanly stop the reaction at the individual intermediates, even using rapid mixing approaches. The results for the incorporation of A are presented in Table II, 3rd and 4th rows, with the results being calculated for the 3-A addition product. The values shown for the insertion of dATP opposite G are based on incorporation of a single residue, and thus a direct comparison of the efficiency of incorporation of A opposite G and 1,N^2-ϵ-G is somewhat misleading. However, even with these caveats the apparent misinsertion frequency (*f*) is 14, indicating that a 1,N^2-ϵ-G adduct has considerable proclivity for directing the insertion of A > C.

The tendency for Dpo4 to add additional As to the primer beyond the 1,N^2-ϵ-G:A pair was also examined (Table II, 6th and 7th rows). The apparent catalytic efficiencies for these steps are low (0.0009 and 0.0033 min^{-1} μM^{-1}) and may limit the value for the reaction shown in row 4. The efficiency for adding the third A (row 7) is ~4-fold higher than for the second A (Table II, 6th row) and helps to explain the pattern of incorporation of three As in the gels shown in Fig. 3B.

When the addition of C or A to the primer was examined using rapid quench methods, no burst phase was observed (results not presented), indicating that the addition is an inherently slow process, and the k_{off} rate (for the oligonucleotide) is not rate-limiting, in contrast to results reported for normal base incorporation (35) and confirmed here.

The 1-base incorporation studies were also done with the template sequence containing C instead of T as the pyrimidine
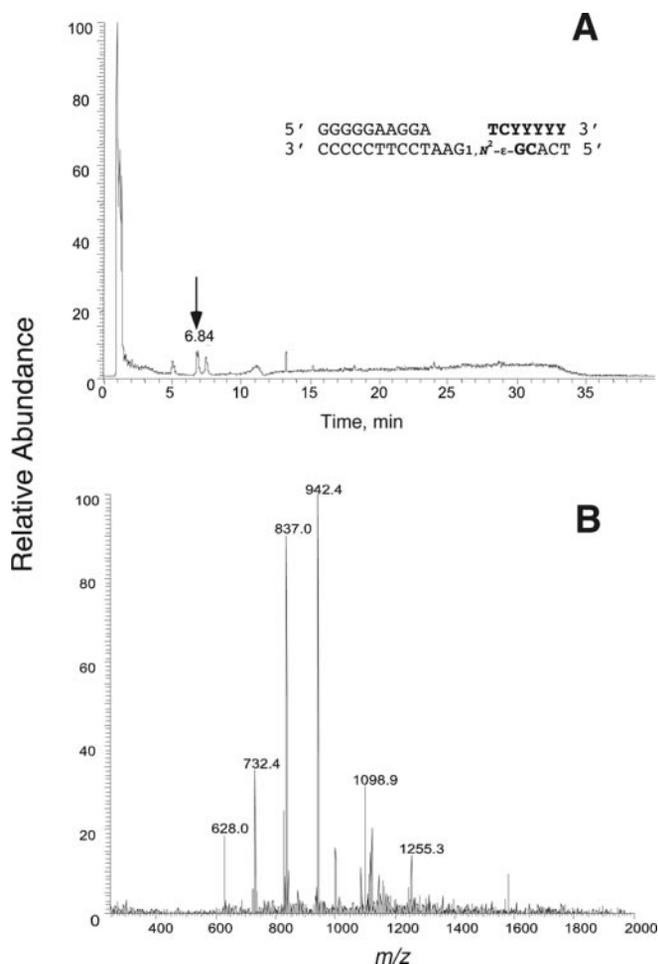
FIG. 4. **Analysis of the Dpo4 reaction product containing the oligonucleotide fragments obtained by UDG hydrolysis.** *A*, TIC spectrum. *B*, ES mass spectrum of the peaks eluted at 6.8 min in *A*.

FIG. 5. **Analysis of Dpo4 major reaction product resulting from extension using template 3′-(1,N^2-ε-G)C-′.** *A*, TIC spectrum of ion *m/z* 942.40. *B*, CID mass spectrum of ion *m/z* 942.40.

following 1,N^2-ε-G (Scheme 1). dCTP was also poorly incorporated opposite 1,N^2-ε-G (Table III). Both dATP and dGTP were incorporated more efficiently than dCTP (Table III, 4th and 6th rows). Dpo4 showed a preference for incorporation of a single G.

MS Analysis of Dpo4 Reaction Products—ES-MS/MS has been used in the past to sequence short oligonucleotides (50, 51). LC-MS/MS has advantages over traditional gel sequencing methods (15, 52), including the capability of handling an oligonucleotide mixture and better sensitivity, to facilitate the analysis of the oligonucleotide products of the Dpo4 reaction. A uracil-containing primer was designed (Scheme 2). After the polymerase extension was done (see Supplemental Material for an electrophoretogram of an incubation done in parallel), the extended primers were digested by UDG to generate short oligonucleotides, *i.e.* 6- and 7-mers in length. The short oligonucleotides can be separated easily from the complementary strand by HPLC, due to the reduced T_m. Furthermore, short oligonucleotides generate less complex CID spectra for sequence analysis.

After UDG digestion and piperidine treatment, a gap was generated in the DNA strand obtained from primer extension with the template containing a C residue 5′ of 1,N^2-ε-G (Scheme 2). This mixture was separated by HPLC, and the oligonucleotide-containing peak was eluted at 6.8 min. Preliminary analysis of the ES mass spectrum (Fig. 4*B*) indicated that if *m/z* 942.4, 1098.9, and 1255.3 were the M − 2H ions, then *m/z* 628.0, 732.4, and 837.0 should be the corresponding M − 3H ions. Thus, there were three major species present in the peak at 6.8 min, and the 942.4, 1098.9, and 1255.3 ions were selected for CID analysis

(Figs. 5 and 6). The TIC trace for the *m/z* 942.4 ion is shown in Fig. 5*A*. The apparent composition of the *m/z* 942.4 (M − 2H) oligonucleotide is 1 phosphate, 2 Ts, 1 C, 2 Gs, and 1 A (because the unknown oligonucleotide was a hydrolysis product obtained using UDG, it should contain a phosphate residue at the 5′-end) (see web page for the Oligo Composition Calculator at medlib.med.utah.edu/masspec/compo.htm). Given the sequence of the template, the unknown sequence was proposed to be 5′-pTCGTGA-3′. CID fragments calculated for this sequence are listed in Table IV, and all of the major a-B and w ions appeared in the CID spectra of the unknown oligonucleotide (Fig. 5*B*). Thus, 5′-pTCGTGA-3′ was identified as one of the major products from the Dpo4 reaction. The same analysis was carried out for the 1098.9 (M − 2H) ion, and this sequence was identified as 5′-pTCAGTGA-3′, with the calculated CID fragments shown in Table V. From comparisons of the TIC intensities, 5′-pTCGTGA-3′ corresponds to 84% of the total product, and 5′-pTCAGTGA-3′ accounts for 16%. Because the 1255.3 (M − 2H) ion could not be fragmented under the same CID conditions and was present throughout the HPLC gradient, this species was not considered for further analysis.

When the C on the 5′-side of 1,N^2-ε-G was replaced by a T (Z = T in Scheme 1), four major oligonucleotide products were generated and eluted at t_R 7.4 min as a mixture: *m/z* 1090.8 (M_1 − 2H), 1086.3 (M_2 − 2H), 934.4 (M_3 − 2H), and 777.8 (M_4 − 2H) (Fig. 7). The ions *m/z* 727.1 (M_1 − 3H) and 622.9 (M_2 − 3H) belong to the parent ions M_1 and M_2, and other peaks were regarded as minor products. The same sequence analysis procedure was used. For example, the *m/z* 1090.8 ion was assumed to be a M − 2H ion, and the M_r of the parent ion should be 2183.6. Thus, the unknown oligonucleotide contained
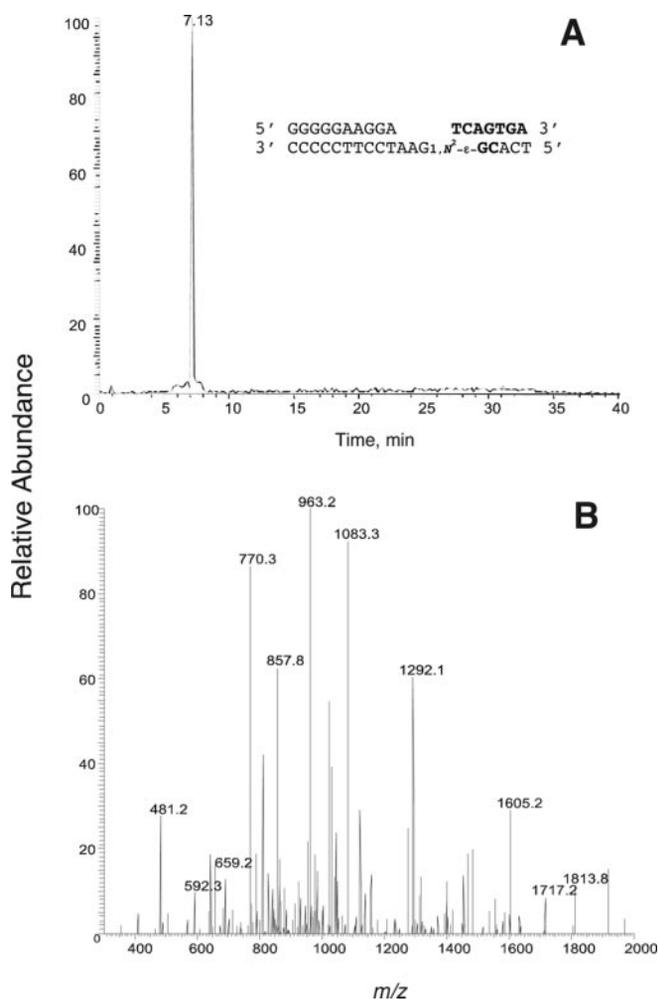
FIG. 6. **Analysis of Dpo4 minor reaction product resulting from extension using template 3′-(1,N^2-ε-G)C-5′.** *A*, TIC spectrum of ion *m/z* 1098.90. *B*, CID mass spectrum of ion *m/z* 1098.90.

TABLE IV
CID fragments of 5′-pTCGTGA-3′

n	Charge	a-B	w	y	d-H_2O
1	−1		**330.059**[a]	250.093	303.037
	−2				
2	−1	**481.040**	659.111	**579.145**	592.083
	−2	240.016	329.051	289.068	295.538
	−3				196.689
3	−1	**770.086**	**963.157**	883.191	921.136
	−2	384.539	481.074	441.091	460.064
	−3	256.023	320.380	293.725	306.373
	−4				229.528
4	−1	1099.138	**1292.209**	1212.243	1225.181
	−2	549.065	645.600	605.617	612.086
	−3	365.707	430.064	403.409	407.722
	−4	274.028	322.296	302.305	305.539
	−5				244.230
5	−1	**1403.183**	1581.255	**1501.289**	1554.233
	−2	**701.088**	790.123	750.140	776.613
	−3	467.056	526.413	499.757	517.406
	−4	350.040	394.558	374.566	387.802
	−5	279.830	315.444	299.451	310.040
	−6				258.199

[a] Boldface numbers refer to the peaks found in the CID spectrum in Fig. 5*B*.

1 phosphate, 2 Ts, 1 C, 1 G, and 3 As, based on the calculations. Given the sequence of the template, a candidate sequence of the unknown is 5′-pTCAATGA-3′. The CID spectra (Fig. 8*B*) matched well with that calculated based on this candidate

TABLE V
CID fragments of 5′-pTCAGTGA-3′

n	Charge	a-B	w	y	d-H_2O
1	−1		330.059	250.093	303.037
	−2				
2	−1	**481.040**[a]	**659.111**	579.145	**592.083**
	−2	240.016	329.051	289.068	295.538
	−3				196.689
3	−1	**770.086**	**963.157**	883.191	905.141
	−2	384.539	481.074	441.091	452.066
	−3	256.023	320.380	293.725	301.041
	−4				225.529
4	−1	**1083.143**	**1292.209**	1212.243	1234.193
	−2	541.067	645.600	605.617	616.592
	−3	360.375	430.064	403.409	410.725
	−4	270.029	322.296	302.305	307.792
	−5				246.032
5	−1	1412.195	**1605.266**	1525.300	1538.239
	−2	705.593	802.129	762.146	768.615
	−3	470.059	534.417	507.761	512.074
	−4	352.293	400.560	380.569	383.803
	−5	281.632	320.247	304.253	306.841
	−6				255.533
6	−1	**1716.241**	1894.312	**1814.346**	1867.291
	−2	**857.616**	946.652	906.669	933.141
	−3	571.408	630.765	604.110	621.758
	−4	428.304	472.822	452.830	466.066
	−5	342.442	378.056	362.063	372.652
	−6	285.200	314.878	301.551	310.375
	−7				265.892

[a] Boldface numbers refer to the peaks found in the CID spectrum in Fig. 6*B*.

sequence (see Supplemental Data for tables of calculated CID fragments). Thus, the four products were identified as 5′-pTCAATGA-3′ (M_1), 5′-pTCATTGA-3′ (M_2), 5′-pTCATGA-3′(M_3), and 5′-pTCTGA-3′ (M_4) (Figs. 7–11), and their yields were 31, 27, 24, and 18% respectively (Scheme 3).[3]

X-ray Crystallography—We determined the crystal structures of a binary Dpo4·DNA and three ternary Dpo4·DNA·d(d)NTP complexes at resolutions of between 2.0 and 2.5 Å (Table I). The lengths of the primer (14-mer for Dpo4-1 and 13-mers for the Dpo4-2, -3, and -4 complexes) and template strands (18-mers in all cases) are similar to those used previously for crystallizing complexes between DNA and Dpo4 (31–34). Alterations in the sequences of the primer and template strands used here compared with published work concern the 3′ terminus of the former and the 5′-half of the latter, including the adducted 1,N^2-ε-G residue (see Ref. 31 and "Experimental Procedures"). The structure of the complex between Dpo4 and a DNA primer-template duplex featuring an abasic site (at position 14 of the 18-mer template (34)) served as the model in the molecular replacement searches to phase the Dpo4-1 complex. Consistent with the excellent correlation factors and initial R factors resulting from the search, the two structures (and indeed all four structures described here) are very similar.

The active site of the polymerase in the Dpo4-1, -2, -3, and -4 complexes closely resembles that in the "type II" crystals of the complex between Dpo4 and native DNA (31). Thus, the 1,N^2-ε-G residue of the template as well as the 5′-adjacent T (Dpo4-1 to -3) and C (Dpo4-4) are both accommodated at the active site

[3] One interesting observation was that *two* major species resulted from the *m/z* 1086.3 ion (Fig. 9). Based on the CID analysis, the one eluted at 7.5 min was identified as 5′-pTCATTGA-3′ (M_2). Another species eluted at t_R 3 min and contained fragments from 5′-GGGG-GAAGGA-3′, which was the other half of the primer after the piperidine hydrolysis. In addition, 1086.3 matched the M − 3H ion of the species, 5′-GGGGGAAGGA-(2′,3′-dideoxy-2′,3′-dehydroribose)-H_2O. It can be assumed that this oligonucleotide may form a cyclic configuration by losing one H_2O molecule. Thus, this product can be eluted even faster than a 5-mer oligonucleotide such as 5′-pTCTGA-3′ (M_4) (Fig. 5).

FIG. 7. **Analysis of the Dpo4 reaction containing the oligonucleotide fragments obtained by UDG hydrolysis of the product of extension of the primer opposite the template 3′-(1,N^2-ε-G)T-5′.** *A*, TIC spectrum. *B*, electrospray ionization mass spectrum of the peaks eluted at 7.4 min in *A*.
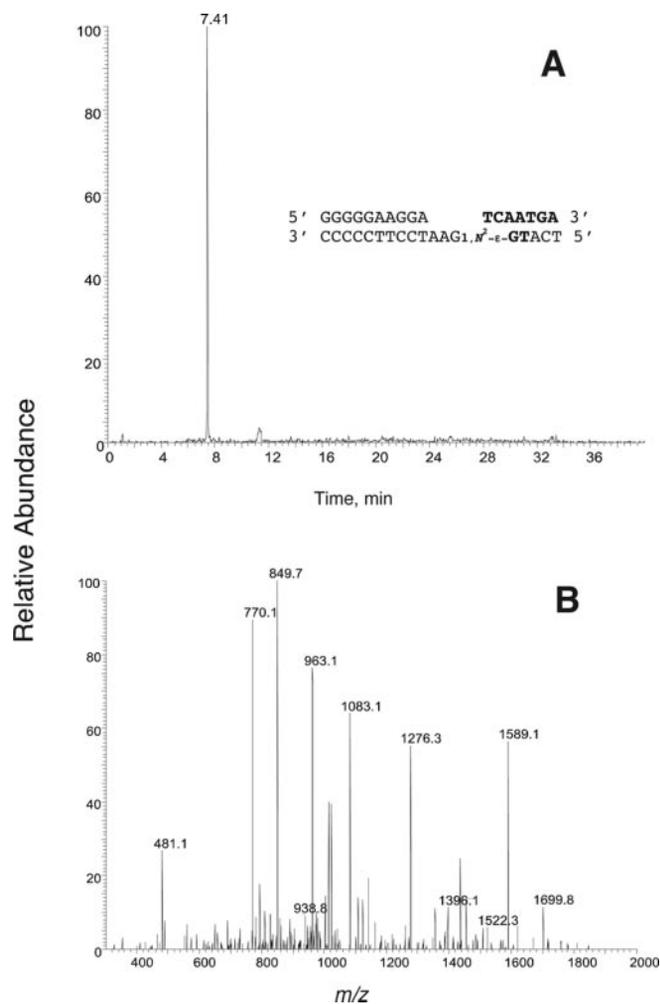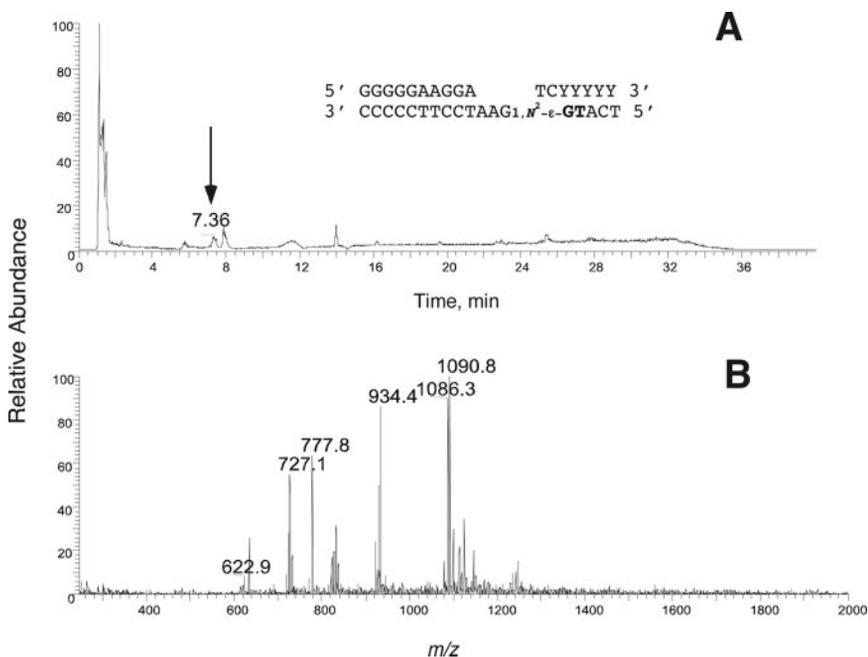
FIG. 8. **Analysis of the most abundant product (5′-AATGA-3′) of Dpo4 extension of primer with the template 5′-(1,N^2-ε-G)-T-3′.** *A*, TIC spectrum of ion *m/z* 1090.8. *B*, CID mass spectrum of ion *m/z* 1090.8.

FIG. 9. **Analysis of 5′-AATGA-3′ product derived from Dpo4 extension of primer with the template 3′-(1,N^2-ε-G)-T-5′.** *A*, TIC spectrum of ion *m/z* 1086.3. *B*, CID mass spectrum of ion *m/z* 1086.3.

(Fig. 12). In the complex with native DNA, the next residue of the template 5′-TT̲CG- (T, underlined) is also directed toward the active site (see Fig. 6*B* in Ref. 31), whereas in our com-

plexes with template sequences 5′-TCA̲T(C)-, the corresponding residue (A, underlined) is looped out and directed away from the active site (Figs. 12 and 13). This may indicate that a purine base is too bulky to be stacked above the second of the two bases lodging in the active site. However, the situation
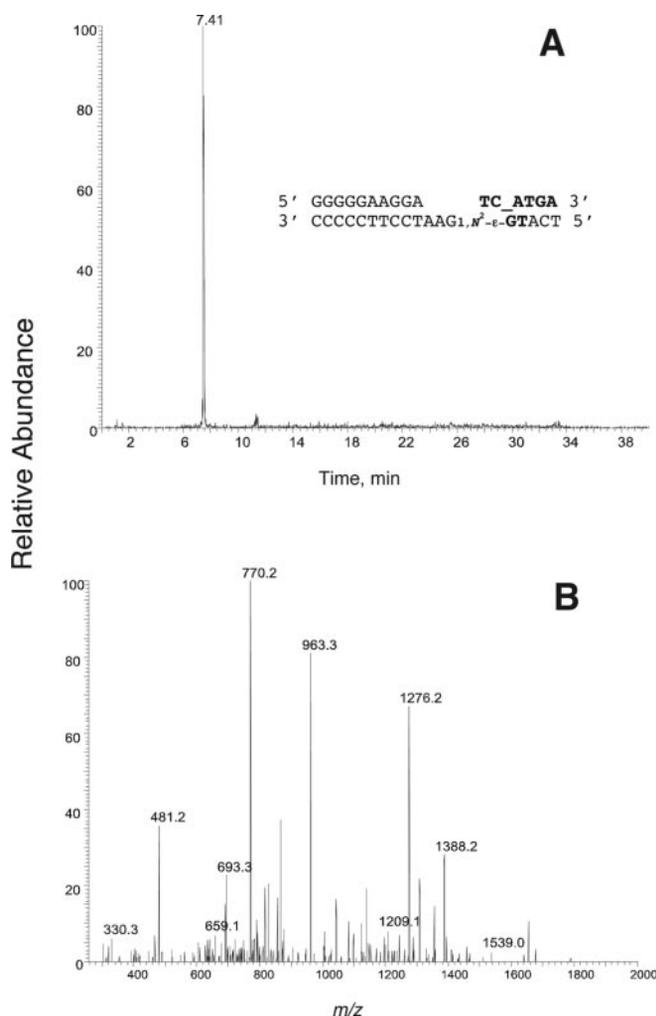
FIG. 10. **Analysis of 5′-ATGA-3′ product of Dpo4 extension of primer with the template 3′-(1,N^2-ϵ-G)-T-5′.** *A*, TIC spectrum of ion *m/z* 934.4. *B*, CID mass spectrum of ion *m/z* 934.4.

FIG. 11. **Analysis of the 5′-TGA-3′ product of Dpo4 extension of primer containing 3′-(1,N^2-ϵ-G)-T-5′.** *A*, TIC spectrum of ion *m/z* 777.8. *B*, CID mass spectrum of ion *m/z* 777.8.

encountered in the Dpo4-4 complex serves to demonstrate that the template strand is able to adopt a range of conformations at this site, as the adenine in that complex assumes a different orientation than in the other three complexes (Fig. 13*D*).

A shared feature of all four complexes is the orientation of the 1,N^2-ϵ-G adduct. The residue is stacked into the duplex but is not opposed by a base from the primer strand (Fig. 13). Instead, the cytosine and adenine bases of the extended primer (+1 and −1 positions relative to the position of the 1,N^2-ϵ-G in the template strand) in the Dpo4-1 complex are wrapped around the modified residue. To bridge the larger distance that results from omitting the template 1,N^2-ϵ-G, both primer bases are tilted in the product complex as well as the ternary Dpo4-2, -3, and -4 complexes. The positions of nucleobases at the active site vary only slightly as can be seen in the superpositions of the complexes with adducted DNA and the "type I" and type II complexes with native DNA (Fig. 14, *A* and *B*, respectively). The superposition of the DNAs in the Dpo4-1, -2, and -4 complexes with a portion of the native DNA primer-template duplex at and adjacent to the active site in the type I crystal shows that the 1,N^2-ϵ-G residue assumes a position that was previously occupied by a Watson-Crick T:pA pair (Fig. 14*B*). The superposition reveals practically no changes in the conformation of the sugar-phosphate backbone of the template. Deviations relative to the native DNA duplex are restricted to the primer residues that wrap around 1,N^2-ϵ-G. In addition to the aforementioned tilting of bases, there are shifts and altered

		TGA 3′	18%
		ATGA 3′	24%
		ATTGA 3′	27%
		AATGA 3′	31%
5′GGGGGAAGGAUTC			
3′CCCCCTTCCTAAG1,N-ϵ-GTACT 5′			
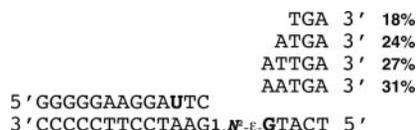
SCHEME 3. **Primer extension products derived from sequence of Scheme 1A (X = 1,N^2-ϵ-G, Z = T).**

torsion angles in the sugar-phosphate backbone of the primer to allow for the bridging of the gap resulting between T and A (G in the case of Dpo4-4; Fig. 14*A*).

The superposition of the binary Dpo4-1 complex, the ternary complexes Dpo4-2 and Dpo4-4, as well as the type II ternary complex containing native DNA also manifests only subtle variations between the complex with the product and those with dNTPs or ddNTPs occupying the active site (Fig. 14*B*). The 3′-OH groups of the last primer nucleotide are poised for attack at the α-phosphates of dATP (Dpo4-2; Fig. 14*B*) and ddATP (Dpo4-3; not shown). Moreover, the metal ions at the active site occupy virtually identical sites in all four complexes with 1,N^2-ϵ-G-modified templates. However, there is an interesting rearrangement of the triphosphate moiety of ddGTP in the Dpo4-4 complex (Fig. 13*D*) compared with the conformations of dATP and ddATP in the Dpo4-2 and Dpo4-3 complexes, respectively (Fig. 13, *B* and *C*). In the Dpo4-4 complex, the γ-phosphate is closest to the 3′-terminal nucleotide of the primer, whereas the α-phosphate now occupies a position that is close to those of the γ-phosphates in the Dpo4-2 and -3
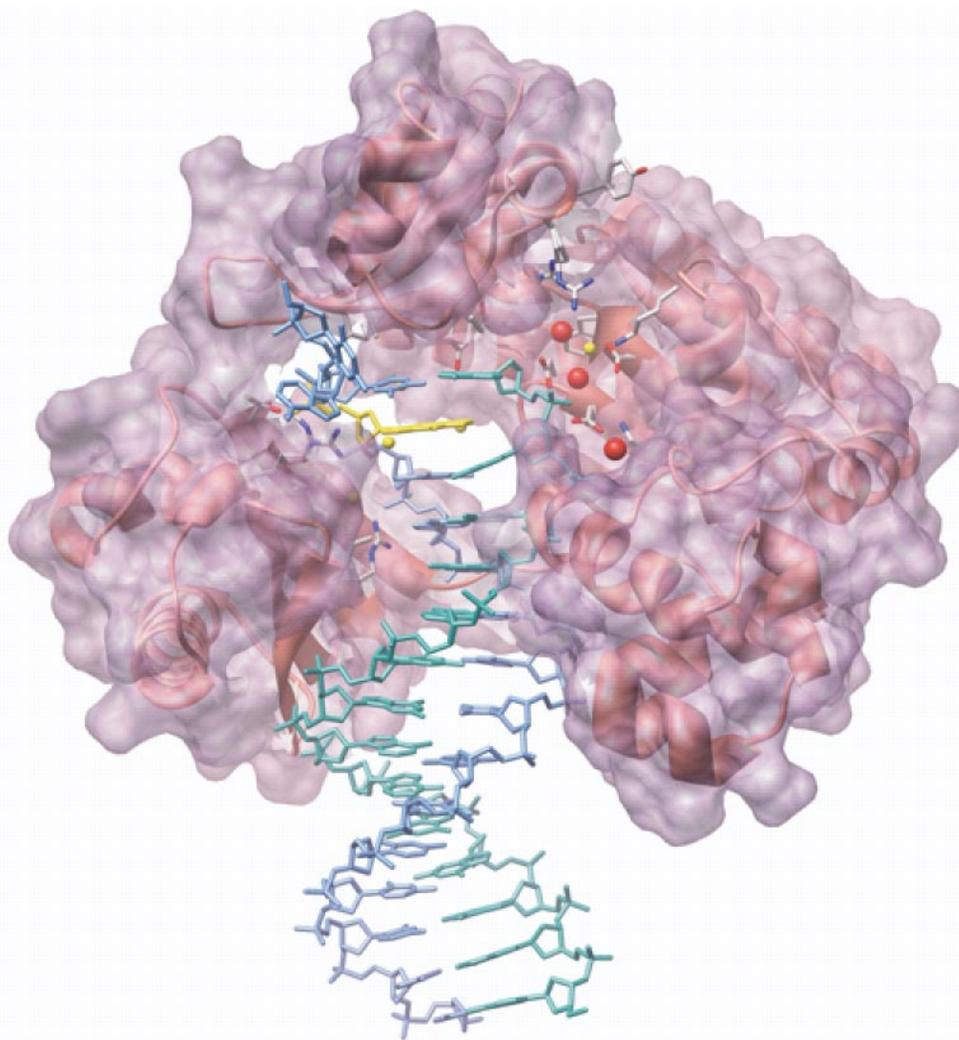
FIG. 12. **Overall structure of the Dpo4 binary complex with DNA.** The primer template complex is that shown in Scheme 1*A*, with $X =$ 1,N^2-ϵ-G and $Z =$ T. The Dpo4 backbone is depicted in a ribbon diagram with the molecular surface superimposed. Side chains of protein residues in the little finger (*left*), finger (*top*), and palm (*upper right*) domains surrounding the active site are shown. The DNA template (*blue*) and primer (*cyan*) strands are shown as wire models with the 1,N^2-ϵ-G nucleotide highlighted in *yellow* and Ca^{2+} ions are *red spheres*.

complexes. The resolutions of the structures allow unambiguous assignments of the conformations of nucleotides, and the altered orientation of the triphosphate in Dpo4-4 is not a consequence of poor electron density or local disorder. When the positions of the d(d)NTPs in the three structures are compared (results not shown), it is also evident that the dATP (Dpo4-2) occupies a position that is distinct from those of ddATP (Dpo4-3) and ddGTP (Dpo4-4). This is likely the result of a hydrogen bond formed between the 3'-OH of dATP and OD2 of Asp-105 that is absent in the two other complexes with either ddATP or ddGTP. Detailed analysis reveals that the slightly altered orientations of the nucleotide triphosphates go along with different sets of hydrogen bonding interactions with Dpo4 side chains in the respective complexes.

The observation that the presence of 1,N^2-ϵ-G in the template strand leads to virtually no structural changes relative to the type II Dpo4 complex with native DNA is remarkable. In particular, there is no evidence from the structural data for an extrahelical orientation of the 1,N^2-ϵ-G residue. Such a rearrangement into the major groove (toward the viewer in Figs. 12 and 13) would be necessary in order to allow a Hoogsteen-type interaction with adenine via the N-2 and N-3 positions of 1,N^2-ϵ-G from the minor groove side. However, the observed intrahelical orientation of 1,N^2-ϵ-G in all four complexes presumably offers energetic advantages in terms of stacking. As is evident

from Figs. 14 and 15, 1,N^2-ϵ-G provides an excellent platform for cross-strand stacking with an incoming dATP or dGTP. Independent of the sequence of the template 5' to 1,N^2-ϵ-G, it may therefore be easier to incorporate A or G, even if the template strand codes for a pyrimidine. This phenomenon would be consistent with the facile incorporation of multiple As catalyzed by Dpo4 with templates containing 1,N^2-ϵ-G (Figs. 2 and 3).

DISCUSSION

The interactions of the archebacter DNA polymerase Dpo4 and oligonucleotides containing 1,N^2-ϵ-G were studied as a model of a polymerase-DNA-carcinogen system, and 1,N^2-ϵ-G is an adduct of potential mutagenic concern itself. Typical gel electrophoresis assays and steady-state kinetic analysis indicated that a mixture of products was formed. The products formed by Dpo4 polymerization past the template 1,N^2-ϵ-G were analyzed by LC-ES-MS/MS. The six identified products, obtained with the two templates, correspond to substitution at the adduct site, -1 and -2 frameshifts, and a more complex mechanism involving a variation of a "dNTP-stabilized" slipped intermediate (53) (Schemes 4 and 5). The proposed mechanisms are supported by the results of analysis of x-ray crystal structures of three of the intermediates, including one binary complex and three different ternary complexes (Fig. 13).
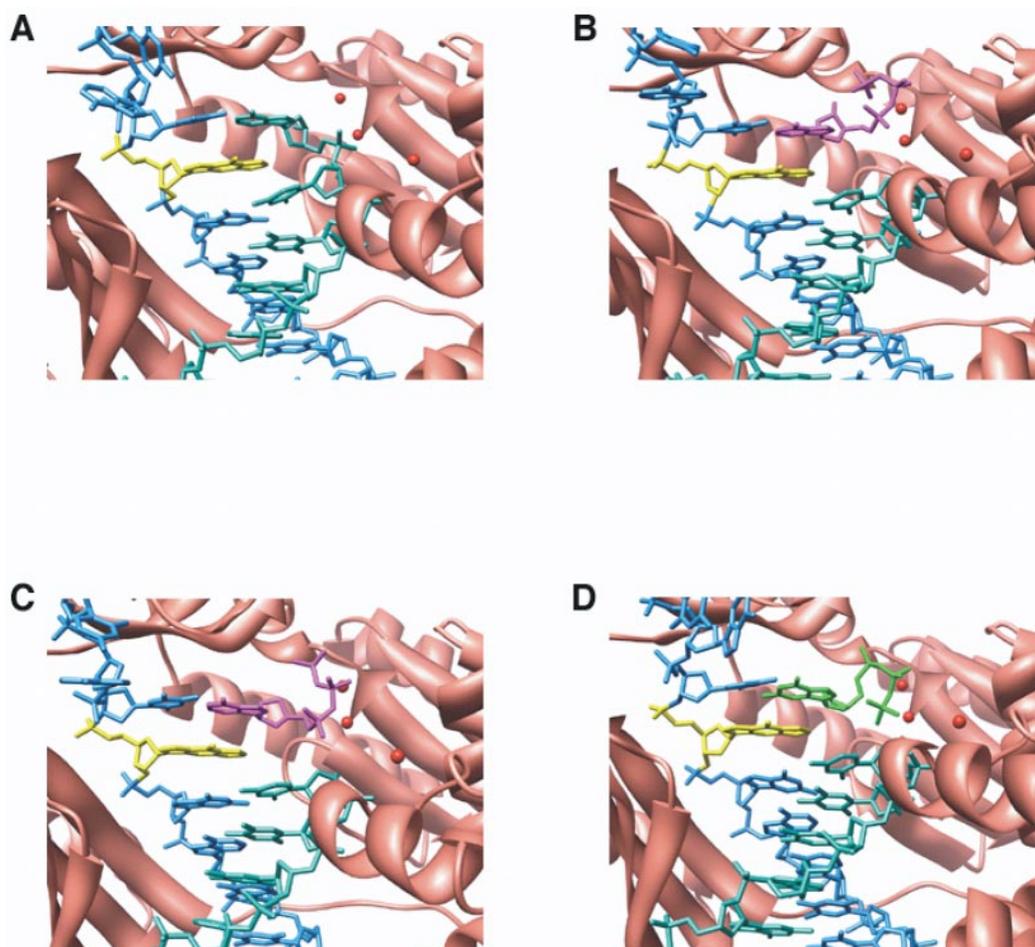
FIG. 13. **Close views of active site regions of Dpo4 crystal structures.** *A*, Dpo4-1. Dpo4·DNA·Ca^{2+} (primer-template complex of Scheme 1*B*, with $X = 1,N^2$-ϵ-G and Z = T). *B*, Dpo4-2. Dpo4·DNA·dATP·Ca^{2+} (primer-template complex of Scheme 1*A*, with $X = 1,N^2$-ϵ-G and Z = T). *C*, Dpo4-3. Dpo4·DNA·ddATP·low Mg^{2+} (primer-template complex of Scheme 1*A*, with $X = 1,N^2$-ϵ-G and Z = T). *D*, Dpo4-4. Dpo4·DNA·Ca^{2+} (primer-template complex of Scheme 1*B*, with Z = C). The color codes for protein and DNA are the same as in Fig. 12: dATP and ddATP are drawn in *pink*, ddGTP in *green*, and divalent metal ions are shown as *red spheres*.

A critical part of this work was the LC-ES-MS/MS analysis of the extended primer product (Figs. 4–11 and Scheme 3). The substitution of a uracil for T in the primer, 5′ of the start of polymerization, permitted better recovery, separation, and analysis of the products (Scheme 2). The MS methodology was optimized using the product obtained with the 3′-1,N^2-ϵ-G-5′ template. In ES-MS, the signal intensity is a reasonable measure of the amounts of each of similar materials, and we estimate that 84% of the product corresponded to an incorporation of G opposite the C residue 5′ of the 1,N^2-ϵ-G, thus skipping the adduct and resulting in a −1 deletion product (Scheme 4*A*). The minor product (16%) was due to the insertion of A opposite 1,N^2-ϵ-G, followed by accurate polymerization beyond this match and yielding a base pair substitution (Scheme 4*B*). ES-MS/MS was used instead of matrix-assisted laser desorption ionization methods, for which the signal response can be quite variable for similar compounds. The ES-MS/MS method was also applied to the analysis of products obtained with the 3′-1,N^2-ϵ-G-T-5′ template. The four products were determined (Scheme 3), and reasonable mechanisms are proposed in Scheme 5. The crystal structures of two of the intermediates were obtained (Fig. 13, *A–C*).

The MS methods for analysis of the products are far superior to Maxam-Gilbert analysis of oligonucleotide products, which was done previously in work with 1,N^2-ϵ-G (15, 16). The individual two purines and two pyrimidines are only distinguished by "subtractive analysis" in the reading of the gels (52). The electrophoretic assays are inappropriate for mixtures (*e.g.* Scheme 3) and particularly so for oligonucleotides of varying lengths with partial identical sequence. Furthermore, if longer products were formed and ligated into vectors for modern automated sequencing methods, the results obtained with a mixture (*e.g.* Scheme 3) would be uninterpretable.

The analysis of the products by MS/MS provides a more accurate analysis of the course of incorporation, in the presence of a mixture of all four dNTPs, than is obtained with the more traditional gel assays using a single dNTP in each setting (Figs. 2 and 3 and Tables II and III). With the 3′-1,N^2-ϵ-G-C-5′ template (Fig. 2), Dpo4 inserts multiple As and Gs (or 1 T). A similar pattern was seen with Dpo4 and the 3′-1,N^2-ϵ-G-T-5′ template, although apparently 3 As, 1 G, or 1 T was inserted (Fig. 3). The steady-state kinetic analyses are consonant with these results and provide some quantitative information, although the subsequent MS (and x-ray) analyses provide a superior view of the polymerase mechanism. In principle, the course of the nucleotide incorporation process opposite and beyond 1,N^2-ϵ-G might be discerned by analysis of sets of quantitative steady-state analyses (*e.g.* Tables II and III), although complex multiple experiments would be in order. The difficulty of doing this is seen in Scheme 5 and the myriad of possible events. Each 1-base incorporation event would have to be followed by analysis of the next base insertion, and identifying the formation of the product in Scheme 5*B* would be quite difficult. Thus the MS analysis has the advantage of providing a semi-
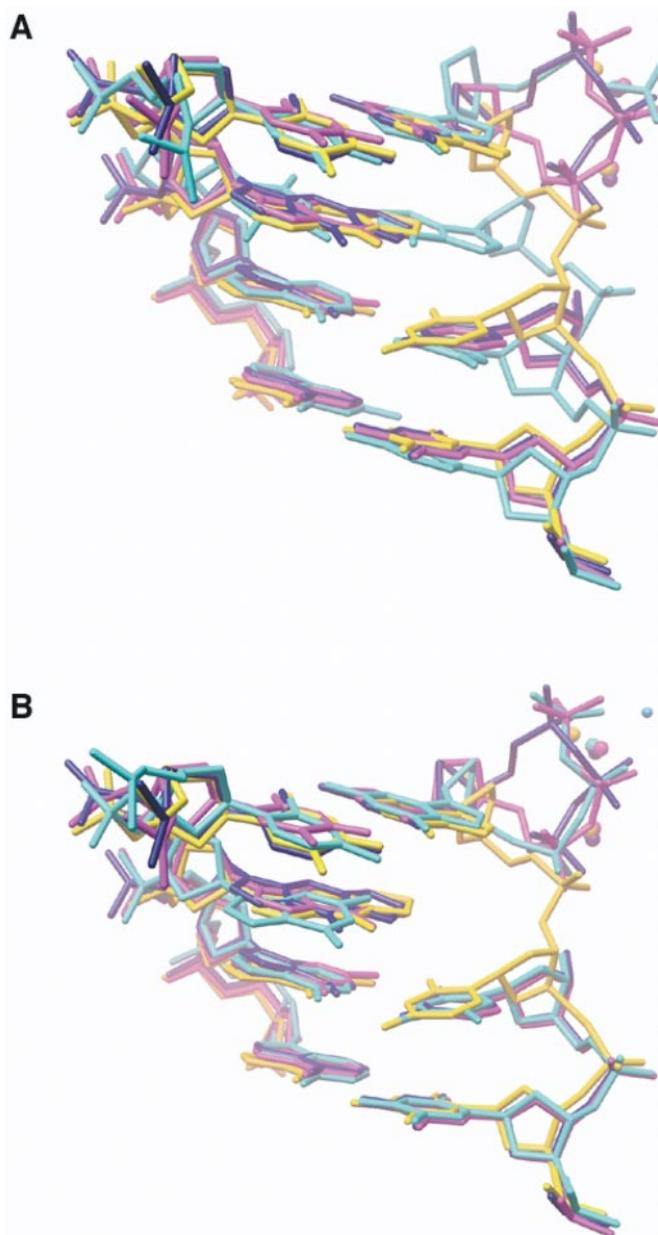
FIG. 14. **Refined structures of Dpo4 crystals.** *A*, superpositions of the primer-template DNA duplexes of Dpo4-1 (*yellow*), Dpo4-2 (*magenta*), Dpo4-4 (*purple*), and native DNA (*blue*, from Ref. 31). *A,* type I crystal. *B,* type II crystal.

quantitative description of the overall process in a single experiment.

Some comparisons of the Dpo4 results can be made with other DNA polymerases in regard to $1,N^2$-ϵ-G results, although (i) some of the earlier work with individual polymerases was done in different sequence contexts, and (ii) all of the earlier work with individual polymerases was done with simple gel analysis and has limitations. Previous work with *E. coli* pol I⁻ (Klenow fragment) and pol II⁻, pol T7⁻, human immunodeficiency virus, type 1, reverse transcriptase, and rat pol β have yielded incorporation of all four dNTPs opposite $1,N^2$-ϵ-G, as well as −1 and −2 deletions (15, 16). Both G to A transitions and G to T transversions were found in *E. coli* cells, corresponding to insertion of A and T opposite $1,N^2$-ϵ-G (16). In Chinese hamster ovary cells, the most common base pair substitution at the $1,N^2$-ϵ-G site corresponded to insertion of T (G to A transition) (17). One of the base pair mutants corresponded to insertion of dATP, and a number of complex mutants resulted from

rearrangements and deletions. A direct comparison of Dpo4 with another polymerase, pol T7⁻, is made here in Figs. 1–3, with the same sequences used. pol T7⁻ showed somewhat similar tendencies, as judged by these results, but there were some differences. For instance, Dpo4 was more likely to insert G than A (Fig. 2). With the other template sequence used (Fig. 3), both polymerases preferred to incorporate A, but Dpo4 catalyzed multiple incorporations. Thus, we conclude that the course of polymerization events is a function of the individual DNA adduct, the DNA sequence (Scheme 4), and the polymerase (Figs. 2 and 3).

Multiple mechanisms are required to explain the mixtures of products Dpo4 generated (Scheme 3). We first consider the mechanisms involved in extension of the primer opposite the template with a C residue 5′ of $1,N^2$-ϵ-G (Scheme 4). The major product, the −1 deletion, is explained by the mechanism shown in Scheme 4*A*, with a direct insertion of the dGTP opposite the C. This intermediate is confirmed by its crystal structure (Fig. 13*D*). This behavior of Dpo4 has been suggested previously in kinetic (35) and structural (31) studies with mispaired bases, without DNA adducts. The minor product has the A inserted opposite $1,N^2$-ϵ-G, followed by normal incorporation. The latter result is of significance in explaining the results obtained with the other template (Schemes 3 and 5) because it shows an inherent tendency to insert an A opposite $1,N^2$-ϵ-G.

The mechanisms for replication of the template with T in the position 5′ of $1,N^2$-ϵ-G are more complex, with four products recovered in similar yield (Scheme 5). The product shown in Scheme 5*A* can be derived via the intermediates for which crystal structures were obtained (in *boxes*) or from an initial binding of dATP opposite $1,N^2$-ϵ-G. The former path is supported by the crystal structures, but the result with the other template (depicted in Scheme 4*B*) is possible, with evidence from the Scheme 4*B* mechanism for the other sequence. The product produced in Scheme 5*B* is presumed to be bulged. The proposed mechanism begins with the intermediates identified in the crystal and proceeds with a slippage. This mechanism has some semblance to the "dNTP-stabilized misalignment" and "misinsertion misalignment" mechanisms discussed by Tippin *et al.* (53), although it is not exactly the same as either. The mechanism proposed in Scheme 5*C* is identical to that of Scheme 4*A* and is supported by the crystal structures. Finally, the product generated in Scheme 5*D* (−2 deletion) clearly must come from dTTP insertion opposite the A, although we have not searched for crystal structures of this type. Finally, it is of interest to note that the results of the gel assays with single dNTPs, by themselves (Figs. 2 and 3), are of limited usefulness in predicting the results obtained using MS and x-ray crystallography, as discussed above.

The first crystal structure of Dpo4, published by Ling *et al.* (31), showed a much more open cleft than known for the processive DNA polymerases, as visualized in Fig. 12 of the present work. Our crystal structures have the incoming d(d)ATP skipping the $1,N^2$-ϵ-G and pairing in a classic hydrogen-bonding mode with the T base 5′ to the $1,N^2$-ϵ-G (Fig. 13, *B* and *C*). Similarly, in the template sequence with the C, the incoming d(d)GTP skips the $1,N^2$-ϵ-G and pairs in a classic hydrogen-bonding mode with the C that is 5′ to the $1,N^2$-ϵ-G. In the binary complex with 1-base-extended primer plus template, the same Watson-Crick bonding of the A with the T base 5′ to the $1,N^2$-ϵ-G is observed (Figs. 13 and 14). A somewhat unexpected feature is that the $1,N^2$-ϵ-G base is not bulged or flipped, and the chains extend past the lesion for phosphodiester bond formation. This perturbed duplex structure might be expected to be associated with an altered Dpo4 structure, although we have not seen any major changes relative to the complex with un-
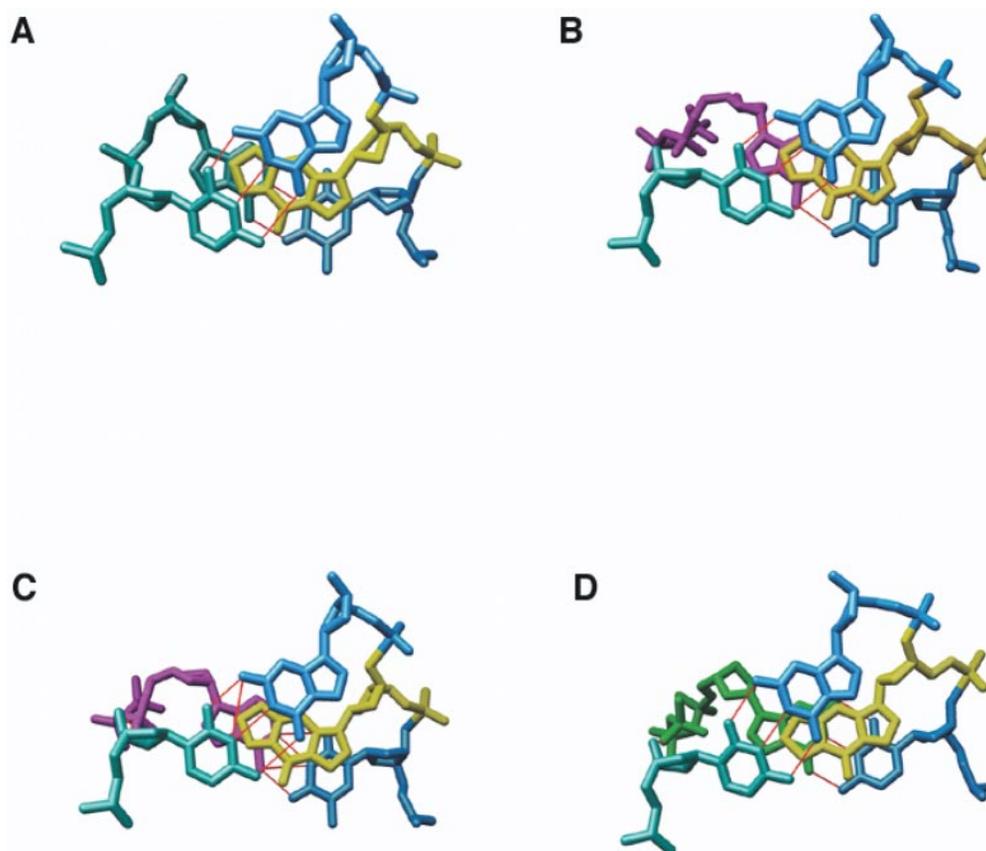
FIG. 15. **Stacking interactions between DNA residues at the active site of Dpo4 and the adjacent C:G base pair.** The C:G base pair is in the foreground, viewed approximately along the helical axis. *A*, Dpo4-1. *B*, Dpo4-2. *C*, Dpo4-3. *D*, Dpo4-4. The color codes for the template and primer strands including 1,N^2-ϵ-G are the same as in Fig. 12. Nucleotide triphosphates are colored *magenta* (dATP and ddATP in *B* and *C*, respectively) or *green* (ddGTP, *D*).
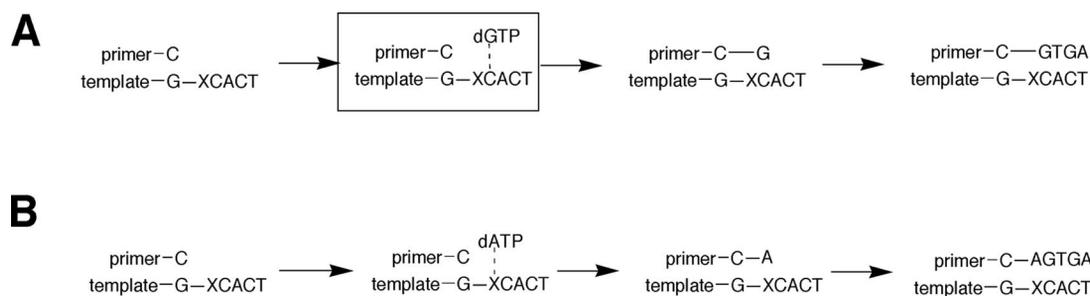
SCHEME 4. **Proposed mechanisms for primer extension with template 5′-1,N^2-ϵ-G·C-3′ (Scheme 1A).** A crystal structure was obtained for the intermediate shown with the *box* (Fig. 13D). *A*, major product. *B*, minor product.

modified DNA. The structures have some relationship to the type II Dpo4 complex of Ling *et al.* (31), with two adjacent template bases admitted into the active site simultaneously.

Several other Dpo4 structures are now available with adducted DNA or mismatched residues and may be compared. Two structures were observed with a *cis-syn*-cyclobutane (T-T) dimer (32). In this system, Dpo4 incorporates two As and thus preserves the coding information. An N^6-adenyl benzo-[*a*]pyrene diol epoxide complex yielded two crystal forms (BP-1 and BP-2), which correspond to stacked and extruded orientations (33). The considerable bulk of this lesion makes a comparison with our work difficult. Recently, Wang *et al.* (54) published a molecular dynamics simulation of the interaction of Dpo4 with an N^2-G adduct derived from benzo[*a*]pyrene diol epoxide and compared the predicted stabilities. A set of five structures related to the copying of Dpo4 past an abasic site, published by Ling *et al.* (34), provides very useful information on the course of the mechanism of several phenomena. Several

of these structures may have relevance for the structures we solved with the 1,N^2-ϵ-G adduct (Fig. 13) and the products obtained with the two different templates (Schemes 4 and 5). In the Ab-2B complex that displays the type II structure, the primer terminus is opposite the template abasic site, and the next template base 5′-adjacent to the lesion has no partner, and the incoming dTTP then base-pairs with the template A two bases from the lesion. The ensuing −1 frameshift is similar to those seen with 1,N^2-ϵ-G-modified templates (Schemes 4, *B* and *C*, and 5*A*; Fig. 13). If one were to consider the abasic site in that complex as a model for a sugar-phosphate backbone conformation that could potentially be adopted by 1,N^2-ϵ-G, one would arrive at a scenario that is consistent with the −2 frameshift product depicted in Scheme 5*D*. In the structures of the Ab-1 and Ab-2A complexes that are of type I, the abasic lesion in the template is looped out, indicating that a local distortion of the DNA backbone is acceptable at the active site of the Dpo4 polymerase. Incorporation of an adenine opposite
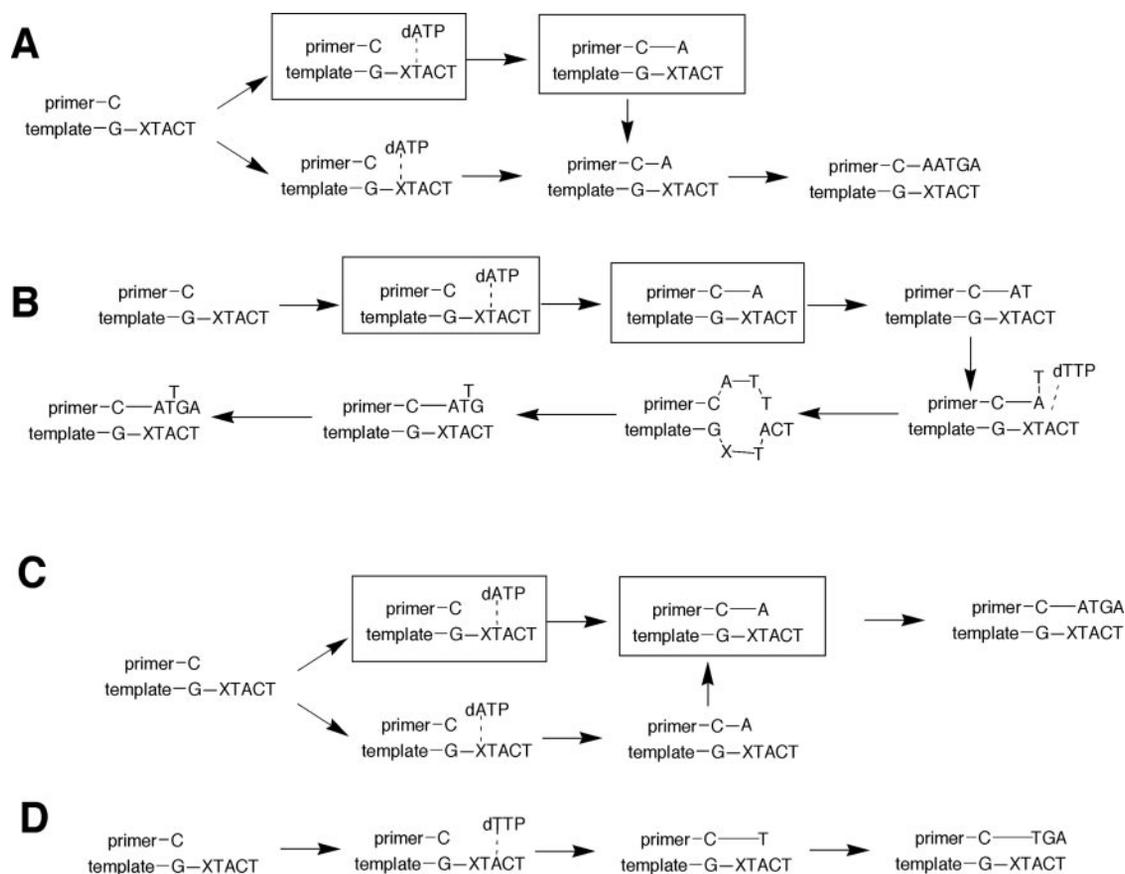
SCHEME 5. **Proposed mechanisms for primer extension with the template 5′-1, N^2-ϵ-G-G-T-3′ (Schemes 1A and 3).** Crystal structures were obtained for the intermediates shown in the *boxes* (Figs. 13, *A–C*). The products were characterized in the individual figures: *A*, Fig. 8. *B*, Fig. 9. *C*, Fig. 10. *D*, Fig. 11.

the 1,N^2-ϵ-G adduct (Scheme 4*B* and Scheme 5*A*) requires the latter to swing out of the helix, thus making way for the pairing interaction with A (presumably via the minor groove edge of 1,N^2-ϵ-G). Finally, the structure of the Ab-3 complex revealed the 3′-A of the primer opposite the intrahelical abasic site and the incoming dCTP opposite the template-G 5′ to the lesion. Therefore, both the template and the primer were in the correct base pair register. However, the terminal primer base has slid toward the template strand to increase stacking with the preceding bases from both strands and the 5′-adjacent guanine from the template. This latter interaction is reminiscent of the cross-strand stacking observed in the four structures of binary and ternary complexes of Dpo4 with 1,N^2-ϵ-G-modified DNAs. This observation serves to reiterate the importance of stacking for generating a stable intermediate of the polymerase-DNA complex and may provide a possible explanation for why we were unable thus far to isolate structures of complexes with Dpo4 that display partly extrahelical or fully extruded orientations of the 1,N^2-e-G residue.

Finally, Trincao *et al.* (55) analyzed the structure of a G:T mismatch. In this structure a wobble configuration (56) was observed for the G:T pair and is postulated to prevent extension of the primer. This latter structure does not seem relevant to those determined here with the 1,N^2-ϵ-G adduct.

In summary, analysis of the nucleotide sequences of the Dpo4 reaction products provides new insight into the complexity of the action of a polymerase on carcinogen-modified DNA. X-ray crystal structures of several of the proposed intermediates support their proposed roles in the process. Mechanisms include direct coding for A incorporation opposite the adduct (1,N^2-ϵ-G), what appears to be a dNTP-stabilized misalignment mechanism (35, 53), a variant of the latter that produces a

2-base deletion, and a more complex mode of the dNTP-stabilized misalignment that involves both the primer and template (Scheme 5). Two striking features of the −1 deletion crystal structures, both with the dNTP and the 1-base elongated product (Fig. 13), are (i) the lack of perturbation of the template 1,N^2-ϵ-G and (ii) the classic Watson-Crick hydrogen bonding of the "misaligned" bases. The application of these approaches to other types of DNA adducts, coupled with kinetic and other biochemical approaches, may provide more general insight into the nature of interactions of DNA-carcinogen adducts with DNA polymerases.

REFERENCES

1. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd Ed., W. H. Freeman & Co., New York
2. Friedberg, E., Walker, G. C., and Siede, W. (1995) *DNA Repair and Mutagenesis*, American Society for Microbiology, Washington, D. C.
3. Searle, C. E., ed (1984) *Chemical Carcinogens,* Vols. 1 and 2, American Chemical Society, Washington, D. C.
4. Christmann, M., Tomicic, M. T., Roos, W. P., and Kaina, B. (2003) *Toxicology* **193,** 3–34
5. Ames, B. N., Durston, W. E., Yamasaki, E., and Lee, F. D. (1973) *Proc. Natl. Acad. Sci. U. S. A.* **70,** 2281–2285

6. Sattsangi, P. D., Leonard, N. J., and Frihart, C. R. (1977) *J. Org. Chem.* **42,** 3292–3296
7. Singer, B., and Bartsch, H., (eds) (1999) *Exocyclic DNA Adducts in Mutagenesis and Carcinogenesis,* No. 150, International Agency for Research on Cancer Scientific Publications, Lyon, France
8. Guengerich, F. P., Kim, D.-H., and Iwasaki, M. (1991) *Chem. Res. Toxicol.* **4,** 168–179
9. Guengerich, F. P., Persmark, M., and Humphreys, W. G. (1993) *Chem. Res. Toxicol.* **6,** 635–648
10. Hecht, S. S., Young-Sciame, R., and Chung, F. L. (1992) *Chem. Res. Toxicol.* **5,** 706–712
11. Kronberg, L., Sjöholm, R., and Karlsson, S. (1992) *Chem. Res. Toxicol.* **5,** 852–855
12. Sodum, R. S., and Chung, F. L. (1988) *Cancer Res.* **48,** 320–323
13. Morinello, E. J., Ham, A. J. L., Ranasinghe, A., Sangaiah, R., and Swenberg, J. A. (2001) *Chem. Res. Toxicol.* **14,** 327–334
14. Lee, S. H., Arora, J. A., Oe, T., and Blair, I. A. (2005) *Chem. Res. Toxicol.* **18,** 780–786
15. Langouët, S., Müller, M., and Guengerich, F. P. (1997) *Biochemistry* **36,** 6069–6079
16. Langouët, S., Mican, A. N., Müller, M., Fink, S. P., Marnett, L. J., Muhle, S. A., and Guengerich, F. P. (1998) *Biochemistry* **37,** 5184–5193
17. Akasaka, S., and Guengerich, F. P. (1999) *Chem. Res. Toxicol.* **12,** 501–507
18. Saparbaev, M., Langouet, S., Privezentev, C. V., Guengerich, F. P., Cai, H., Elder, R. H., and Laval, J. (2002) *J. Biol. Chem.* **277,** 26987–26993
19. Walker, G. C. (1984) *Microbiol. Rev.* **48,** 60–93
20. Tang, M., Bruck, I., Eritja, R., Turner, J., Frank, E. G., Woodgate, R., O'Donnell, M., and Goodman, M. F. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9755–9760
21. Woodgate, R. (1999) *Genes Dev.* **13,** 2191–2195
22. Ding, J., Das, K., Baynton, K., and Fuchs, R. P. P. (2000) *Trends Biochem. Sci.* **25,** 74–79
23. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S., and Prakash, L. (2000) *Nature* **406,** 1015–1019
24. Ohmori, H., Friedberg, E. C., Fuchs, R. P. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z. G., and Woodgate, R. (2001) *Mol. Cell* **8,** 7–8
25. Friedberg, E. C., Wagner, R., and Radman, M. (2002) *Science* **296,** 1627–1630
26. Boudsocq, F., Kokoska, R. J., Plosky, B. S., Vaisman, A., Ling, H., Kunkel, T. A., Yang, W., and Woodgate, R. (2004) *J. Biol. Chem.* **279,** 32932–32940
27. Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) *Nature* **430,** 377–380
28. Uljon, S. N., Johnson, R. E., Edwards, T. A., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) *Structure (Camb.)* **12,** 1395–1404
29. Yang, W. (2005) *FEBS Lett.* **579,** 868–872
30. Boudsocq, F., Iwai, S., Hanaoka, F., and Woodgate, R. (2001) *Nucleic Acids Res.* **29,** 4607–4616
31. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) *Cell* **107,** 91–102
32. Ling, H., Boudsocq, F., Plosky, B. S., Woodgate, R., and Yang, W. (2003) *Nature* **424,** 1083–1087
33. 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
34. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2004) *Mol. Cell* **13,** 751–762
35. Fiala, K. A., and Suo, Z. (2004) *Biochemistry* **43,** 2106–2115
36. Fiala, K. A., and Suo, Z. (2004) *Biochemistry* **43,** 2116–2125
37. Patel, S. S., Wong, I., and Johnson, K. A. (1991) *Biochemistry* **30,** 511–525
38. Furge, L. L., and Guengerich, F. P. (1997) *Biochemistry* **36,** 6475–6487
39. Zang, H., Harris, T. M., and Guengerich, F. P. (2005) *J. Biol. Chem.* **280,** 1165–1178
40. Wray, W., Boulikas, T., Wray, V. P., and Hancock, R. (1981) *Anal. Biochem.* **118,** 197–203
41. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci.* **4,** 2411–2423
42. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
43. Kabsch, W. (1988) *J. Appl. Crystallogr.* **21,** 916–924
44. French, G. S., and Wilson, K. S. (1978) *Acta Crystallogr. Sect. A* **34,** 517–525
45. Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50,** 157–163
46. Vellieux, F. M. D., and Dijkstra, B. W. (1997) *J. Appl. Crystallogr.* **30,** 396–399
47. 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
48. Laskowski, L. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
49. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J. Comput. Chem.* **25,** 1605–1612
50. Ni, J., Pomerantz, S. C., Rozenski, J., Zhang, Y., and McCloskey, J. A. (1996) *Anal. Chem.* **68,** 1989–1999
51. Owens, D. R., Bothner, B., Phung, Q., Harris, K., and Siuzdak, G. (1998) *Bioorg. Med. Chem.* **6,** 1547–1554
52. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 560–564
53. Tippin, B., Kobayashi, S., Bertram, J. G., and Goodman, M. F. (2004) *J. Biol. Chem.* **279,** 45360–45368
54. Wang, L., Wu, M., Yan, S. F., Patel, D. J., Geacintov, N. E., and Broyde, S. (2005) *Chem. Res. Toxicol.* **18,** 441–456
55. 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
56. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, pp. 155–157, Springer-Verlag Inc., New York
57. Cambillau, C., and Roussel, A. (1997) *TURBO-FRODO,* Version OpenGL.1, University Aix-Marseille II, Marseille, France