

Basis of Miscoding of the DNA Adduct $N^2,3$ -Ethenoguanine by Human Y-family DNA Polymerases*

Received for publication, July 20, 2012, and in revised form, August 17, 2012. Published, JBC Papers in Press, August 21, 2012, DOI 10.1074/jbc.M112.403253

Linlin Zhao^{‡§}, Matthew G. Pence^{‡§}, Plamen P. Christov^{§¶}, Zdzislaw Wawrzak^{||}, Jeong-Yun Choi^{**}, Carmelo J. Rizzo^{§¶}, Martin Egli^{‡§}, and F. Peter Guengerich^{‡§1}

From the Departments of [‡]Biochemistry and [¶]Chemistry and [§]Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146, ^{||}Northwestern University Synchrotron Research Center, Life Sciences Collaborative Access Team, Argonne, Illinois 60439, and ^{**}Division of Pharmacology, Department of Molecular Cell Biology, Samsung Biomedical Research Institute, Sungkyunkwan University School of Medicine, Gyeonggi-do 440-746, Republic of Korea

Background: The miscoding of $N^2,3$ -etheno(ϵ)guanine(G) is of interest regarding cancer.

Results: $N^2,3$ - ϵ G:T mispairing was found with Y-family human DNA polymerases, and crystal structures of polymerase ι revealed Hoogsteen base pairing.

Conclusion: Structural similarity for $N^2,3$ - ϵ G:C and $N^2,3$ - ϵ G:T underlies similar catalytic efficiencies for polymerase ι .

Significance: The structural basis of $N^2,3$ - ϵ G miscoding is revealed.

$N^2,3$ -Ethenoguanine ($N^2,3$ - ϵ G) is one of the exocyclic DNA adducts produced by endogenous processes (e.g. lipid peroxidation) and exposure to bioactivated vinyl monomers such as vinyl chloride, which is a known human carcinogen. Existing studies exploring the miscoding potential of this lesion are quite indirect because of the lability of the glycosidic bond. We utilized a 2'-fluoro isostere approach to stabilize this lesion and synthesized oligonucleotides containing 2'-fluoro- $N^2,3$ - ϵ -2'-deoxyarabinoguanosine to investigate the miscoding potential of $N^2,3$ - ϵ G by Y-family human DNA polymerases (pols). In primer extension assays, pol η and pol κ replicated through $N^2,3$ - ϵ G, whereas pol ι and REV1 yielded only 1-base incorporation. Steady-state kinetics revealed that dCTP incorporation is preferred opposite $N^2,3$ - ϵ G with relative efficiencies in the order of pol κ > REV1 > pol η \approx pol ι , and dTTP misincorporation is the major miscoding event by all four Y-family human DNA pols. Pol ι had the highest dTTP misincorporation frequency (0.71) followed by pol η (0.63). REV1 misincorporated dTTP and dGTP with much lower frequencies. Crystal structures of pol ι with $N^2,3$ - ϵ G paired to dCTP and dTTP revealed Hoogsteen-like base pairing mechanisms. Two hydrogen bonds were observed in the $N^2,3$ - ϵ G:dCTP base pair, whereas only one appears to be present in the case of the $N^2,3$ - ϵ G:dTTP pair. Base pairing mechanisms derived from the crystal structures explain the slightly favored dCTP insertion for pol ι in steady-state

kinetic analysis. Taken together, these results provide a basis for the mutagenic potential of $N^2,3$ - ϵ G.

The integrity of DNA is continually challenged by environmental factors (e.g. UV irradiation and radiation), exogenous and endogenous chemicals, and suboptimal repair processes (1). DNA damage produces modified DNA bases (i.e. DNA lesions or DNA adducts), abasic sites, DNA inter- and intra-strand cross-links, and DNA-protein cross-links that, if not properly repaired, can lead to genomic instability and ultimately disease (e.g. cancer).

DNA polymerases (pols)² are crucial in maintaining genome integrity. Fifteen human DNA pols, varying in their functions in replication, repair, and tolerance of DNA damage, are known (2). The Y-family DNA polymerases (pol η , pol ι , pol κ , and REV1) are specialized in translesion synthesis (3, 4). For example, pol η is known for its unique role in correctly bypassing UV irradiation-induced cyclobutane pyrimidine dimer (5, 6). Pol ι , on the other hand, is unable to copy past cyclobutane pyrimidine dimer but can proficiently insert T or C opposite adducted purines that are impaired in their capability of forming Watson-Crick base pairs (7–9). Pol κ has a specialized role in bypassing bulky N^2 -G adducts (10) and interstrand cross-links (11) and is distinct in its moderate processivity, extending beyond the lesion, possibly due to the use of its N-clasp domain. REV1 is highly selective for inserting C opposite normal (12) and adducted template G (10, 13). Crystal structures of Y-family pols provide insight into their diverse functions in bypassing normal and adducted templates (14). Pol ι adopts an induced fit mechanism by flipping template purines into the *syn* conforma-

* This work was supported, in whole or in part, by National Institutes of Health Grants R01 ES010546 (to F. P. G.), R01 ES010375 (to F. P. G. and M. E.), P01 ES05355 (to M. E. and C. J. R.), T32 ES007028 (to F. P. G. and M. G. P.), and P30 ES000267 (to M. E., F. P. G., and C. J. R.) from the United States Public Health Service. This work was also supported by National Research Foundation Grant 2010-0006538 from the Ministry of Education, Science and Technology Korea (to J.-Y. C.).

The atomic coordinates and structure factors (codes 4FS2 and 4FS1) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence should be addressed: Dept. of Biochemistry, Vanderbilt University School of Medicine, 638B Robinson Research Bldg., 2200 Pierce Ave., Nashville, TN 37232-0146. Tel.: 615-322-2261; Fax: 615-322-4349; E-mail: f.guengerich@vanderbilt.edu.

² The abbreviations used are: pol, DNA polymerase; 1, N^6 - ϵ A, 1, N^6 -ethenoaddene; 1, N^2 - ϵ G, 1, N^2 -ethenoguanine; 3, N^4 - ϵ C, 3, N^4 -ethenocytidine; 2'-F-dG, 2'-fluoro-2'-deoxyarabinoguanosine; 2'-F- $N^2,3$ - ϵ dG, 2'-F- $N^2,3$ - ϵ -2'-deoxyarabinoguanosine; ddC, dideoxy-CMP; dNTP, deoxyribonucleotide triphosphate; Dpo4, *S. solfataricus* P2 DNA polymerase IV; ϵ , etheno; $N^2,3$ - ϵ G, $N^2,3$ -ethenoguanine; UPLC, ultraperformance liquid chromatography; hREV1, human REV1; dG, 2'-deoxyguanosine.

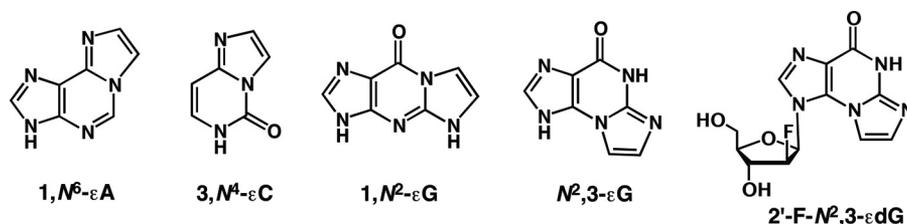


FIGURE 1. Etheno DNA adduct structures.

tion, forming Hoogsteen base pairs (7, 8, 15, 16). REV1 features pairing between dCTP and template G but uses its G-loop to hydrogen bond with the template G and an Arg in another segment (N-digit) to ensure the incorporation of dCTP (12). A high degree of functional and structural differences underlies the diverse but specialized roles in lesion bypass by Y-family human DNA polymerases (17).

Etheno (ϵ) DNA adducts comprise a series of exocyclic adducts, including $1,N^6$ -ethenoadenine ($1,N^6$ - ϵ A), $3,N^4$ -ethenocytidine ($3,N^4$ - ϵ C), $N^2,3$ -ethenoguanine ($N^2,3$ - ϵ G), and $1,N^2$ -ethenoguanine ($1,N^2$ - ϵ G) (Fig. 1). These were first identified as reaction products of nucleobases with chloroacetaldehyde (18) and were used as fluorescent analogs in biochemical studies and as probes for nucleic acid structures (19–21). The ϵ DNA adducts were subsequently recognized as reaction products of DNA with reactive metabolites of several genotoxic chemicals, including the carcinogens vinyl chloride and vinyl carbamate (an oxidation product of urethane). The detection of etheno DNA adducts in various tissues of unexposed rodents (22) and humans (23) led to the discovery of the endogenous pathways of formation (e.g. via reaction with *trans*-4-hydroperoxy-2-nonenal, a lipid peroxidation product (24)). In livers of unexposed rats or humans, the steady-state amounts of $N^2,3$ - ϵ G, $1,N^2$ - ϵ G, and $1,N^6$ - ϵ A have been estimated to be ~ 36 , 30, and 12 lesions/cell, respectively (25).

The mutagenic potentials of etheno adducts have been established in *in vitro* bypass assays (26–30) and site-specific mutagenesis in bacteria (31, 32), Chinese hamster ovary cells (33), and simian kidney cells (34). $N^2,3$ - ϵ G has been less well studied in terms of its replication and repair mechanisms because of the lability of its glycosidic bond. In a polyribo(G/ $N^2,3$ - ϵ G) template, both C and T were incorporated opposite $N^2,3$ - ϵ G by avian myeloblastosis virus reverse transcriptase (35). $N^2,3$ - ϵ -Deoxyguanosine triphosphate was reported to be inserted opposite T by *Escherichia coli* DNA polymerase I (Klenow fragment), *Drosophila melanogaster* polymerase α -primase complex, and human immunodeficiency virus-I reverse transcriptase (30). An indirect assay in *E. coli* showed an estimated mutation frequency of 13% for $N^2,3$ - ϵ G, resulting in G to A transitions (32). The long half-life of $N^2,3$ - ϵ G in rat liver and lung (150 days) and in rat kidney (75 days) in vinyl chloride-exposed rats suggests inefficient repair of this lesion. In human glycosylase assays (*in vitro*), $N^2,3$ - ϵ G was released at a much slower rate compared with $1,N^6$ - ϵ A and $3,N^4$ - ϵ C (36). The mutagenicity and persistence of $N^2,3$ - ϵ G suggest a high miscoding potential *in vivo*. $N^2,3$ - ϵ G is generally considered to contribute to the carcinogenesis of vinyl chloride and inflammation-driven malignancies (37). The dominance of GC to AT transitions in five of six *K-ras* (oncogene) tumors found in vinyl

chloride workers (25) suggests the importance of G adducts, and the miscoding pattern of $1,N^2$ - ϵ G is not consistent with this transition (26–35).

We recently investigated the miscoding of $N^2,3$ - ϵ G using a stabilized 2'-fluoro-substituted analog, 2'-fluoro- $N^2,3$ - ϵ -2'-deoxyarabinoguanosine (2'-F- $N^2,3$ - ϵ dG) (38). The presence of the electronegative fluorine atom destabilizes the transition state leading to an oxocarbenium-like intermediate and hydrolysis of the glycosidic bond. This analog was site-specifically incorporated into oligonucleotides; the stability of 2'-F- $N^2,3$ - ϵ dG permitted steady-state kinetics, primer extension assays, and crystallographic studies. Catalytic insertions opposite 2'-F- $N^2,3$ - ϵ dG were examined using five DNA polymerases, including *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4), the replicative bacteriophage pol T7 DNA exonuclease⁻, (*E. coli* pol I) Klenow fragment exonuclease⁻, yeast pol η , and human DNA pol κ , where a consistent miscoding pattern (2'-F- $N^2,3$ - ϵ dG:T) was found. Crystal structures of Dpo4 with 2'-F- $N^2,3$ - ϵ dG paired with dCTP showed a Watson-Crick-like structure, whereas the complex with 2'-F- $N^2,3$ - ϵ dG:T revealed a "sheared" base pair (38).

To further understand the miscoding potential of $N^2,3$ - ϵ G by Y-family human DNA polymerases, which are highly relevant to translesion synthesis, we carried out a series of primer extension and steady state-kinetic analyses using human pol ι , human pol η , and human REV1 with a template containing 2'-F- $N^2,3$ - ϵ dG. The extension products formed by pol ι were identified using LC-MS/MS. A consistent mispairing pattern was observed (2'-F- $N^2,3$ - ϵ dG:T), and base pairing mechanisms were revealed in two pol ι crystal structures with either dCTP or dTTP paired with 2'-F- $N^2,3$ - ϵ dG but with individual differences.

EXPERIMENTAL PROCEDURES

Materials—All commercial chemicals were of the highest quality available and were used without further purification. Unlabeled dNTPs, T4 polynucleotide kinase, uracil-DNA glycosylase, and restriction endonucleases were from New England Biolabs (Ipswich, MA). [γ - 32 P]ATP (specific activity, 3×10^3 Ci mmol⁻¹) was purchased from PerkinElmer Life Sciences. Biospin columns were from Bio-Rad. Unmodified oligonucleotides were from Midland Certified Reagents (Midland, TX). 2-Amino-9-(2-deoxy-2-fluoro- β -D-arabinofuranosyl)-guanine was from Metkinen (Kuopio, Finland). Modified oligonucleotides containing 2'-F- $N^2,3$ - ϵ dG were synthesized as described earlier (38) followed by HPLC purification and desalting with Sephadex G-25 (Sigma-Aldrich). The modified 23-mer template used for extension and kinetic assays was 5'-TCATXGAATCCTTACGAGCCCC-3' where X = 2'-F-

Miscoding of $N^2,3$ -Ethenoguanine

$N^2,3$ - ϵ dG (MALDI-TOF MS (3-hydroxypicolinic acid) m/z calculated for $[M - H]^-$, 6986.5; found, 6985.6) or 2'-fluoro-2'-deoxyarabinoguanosine (2'-F-dG) (MALDI-TOF MS (3-hydroxypicolinic acid) m/z calculated for $[M - H]^-$, 6962.5; found, 6963.5). The 18-mer oligomer used for crystallographic studies was 5'-TCT(2'-F- $N^2,3$ - ϵ dG)GGGTCCTAGGACC-(ddC)-3' (ddC, dideoxy-CMP) (MALDI-TOF MS (3-hydroxypicolinic acid) m/z calculated for $[M - H]^-$, 5514.9; found, 5515.2). Human DNA pol ι catalytic fragments (residues 1–420) (16), pol η (39), and pol κ (38) were purified following protocols described previously.

Preparation of Recombinant Catalytic Core of Human REV1—The gene fragment covering the catalytic core (residues 330–833) (12) of wild-type human REV1 was obtained by PCR amplification from the vector pET-22b(+)/hREV1 (13) as template using *Pfu* DNA polymerase (Stratagene, La Jolla, CA) with a pair of primers (5'-GGATCCATGTCTACGTTTAGCAAGGCAG-3' and 5'-GCGGCCGCTTATGTGGAAGGGTTCA-GATTAG-3'). The resulting PCR product of the 1.5-kb *hREV1* fragment was cloned into the vector pSC-B-Amp/Kan (Stratagene). Following sequence confirmation, the *hREV1* gene fragment was cloned into the BamHI and NotI sites of the vector pBG101 (obtained from the Center for Structural Biology, Vanderbilt University) to generate the cleavable glutathione *S*-transferase (GST)-tagged protein. The GST-tagged hREV1(330–833) was expressed in *E. coli* BL21 (DE3) cells, which were grown at 37 °C and 220 rpm to an OD_{600} of 0.6 and then induced with isopropyl β -D-1-thiogalactopyranoside (0.2 mM) for 12 h at 16 °C. The harvested pellets were resuspended in lysis buffer containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl, 10% (v/v) glycerol, 5 mM β -mercaptoethanol, 1 mg ml⁻¹ lysozyme, and protease inhibitor mixture (Roche Applied Science). Suspensions were sonicated, and the cell lysate was clarified by centrifugation at $4 \times 10^4 \times g$ for 60 min at 4 °C. The resulting supernatants were loaded onto a 1-ml GSTrap 4B column (GE Healthcare), and the column was washed with 20 ml of Buffer A (50 mM Tris-HCl (pH 7.4) containing 150 mM NaCl, 10% (v/v) glycerol, and 5 mM β -mercaptoethanol). The GST-tagged REV1(330–833) bound on the column was cleaved with PreScission protease (GE Healthcare) for 14 h at 4 °C. Cleaved REV1(330–833) was eluted with Buffer A, and the purity was analyzed by SDS-polyacrylamide gel electrophoresis with Coomassie Brilliant Blue R-250 staining. A typical yield was ~760 μ g from 1 liter of culture.

Primer Extension and Steady-state Kinetic Assays—An 18-mer oligomer (5'-GGGGGCTCGTAAGGATTC-3') was 5' [γ -³²P]ATP end-labeled and annealed to a 23-mer template (5'-TCATXGAATCCTTACGAGCCCC-3' where $X = 2'$ -F- $N^2,3$ - ϵ dG, 2'-F-dG, or 2'-deoxyguanosine (dG)). Primer extension experiments were performed in 50 mM Tris-HCl buffer (pH 7.5) containing 60 nM primer-template complex, 250 μ M dNTPs, 20 nM polymerase, 5% (v/v) glycerol, 5 mM DTT, 50 mM NaCl, 5 mM MgCl₂, and 50 μ g ml⁻¹ bovine serum albumin (BSA) at 37 °C. Steady-state kinetic experiments were carried out under the same conditions except using 5–20 nM pol, varying dNTP concentrations, and 2–10-min incubation times. Reactions were quenched with 9 μ l of 20 mM EDTA (pH 9.0) in 95% (v/v) formamide. Products were separated using 20% (w/v)

acrylamide electrophoresis gels, and results were visualized using a phosphorimaging system (Bio-Rad, Molecular Imager® FX) and analyzed by Quantity One software as described (38).

LC-MS Analysis of Full-length Extension Products—An 18-mer primer (5'-GGGGGCTCGTAAGGAT(dU)C-3') was annealed to the same 23-mer oligomer as described above at a 1:1 molar ratio. Reaction conditions were similar to those used in steady-state kinetic assays except that the final concentrations were as follows: 10 μ M pol ι , 12.5 μ M primer-template complex, and 2% (v/v) glycerol in a total volume of 80 μ l. Reactions were carried out in the presence of four dNTPs (10 mM each) for 3.5 h at 37 °C. The reactions were terminated by spin column separations to extract dNTPs and Mg²⁺, and the resulting product was treated with 50 units of uracil-DNA glycosylase and then 0.25 M hot piperidine (40). LC-MS/MS analyses were performed using an ACQUITY ultraperformance liquid chromatography (UPLC) system (Waters Corp.) connected to a Finnigan LTQ mass spectrometer (Thermo Scientific Corp., San Jose, CA) operating in the electrospray ionization negative ion mode and using an ACQUITY UPLC system BEH octadecylsilane (C₁₈) column (1.7 μ m; 1.0 \times 100 mm). UPLC conditions were as described (38).

Crystallization of Pol ι 2'-F- $N^2,3$ - ϵ dG-DNA Ternary Complexes—The sequence of the 18-mer oligomer used for co-crystallization with pol ι , 5'-TCT(2'-F- $N^2,3$ - ϵ dG)GGGTCCTAGGACC(ddC)-3', was designed based on previous studies (15, 16). Crystals were obtained under conditions similar to those described previously (15, 16). Specifically, 210 μ M pol ι , 253 μ M annealed DNA, 10 mM MgCl₂, and 20 mM dCTP (or dTTP) were mixed and preincubated on ice. Droplets (a 1:1 (v/v) mixture of pol-DNA complex mixture and the reservoir solution) were equilibrated against a reservoir solution containing 0.10 M 2-(*N*-morpholino)ethanesulfonic acid (MES) (sodium salt, pH 6.5), 0.3 M (NH₄)₂SO₄, and 15% (w/v) polyethylene glycol 5000 for pol ι -1 (pol ι 2'-F- $N^2,3$ - ϵ dG:dCTP) or 17% (w/v) polyethylene glycol 5000 for pol ι -2 (pol ι 2'-F- $N^2,3$ - ϵ dG:dTTP). Crystals were grown using the hanging drop vapor diffusion method at 4 °C and mounted at 4 °C with step soaking in mother liquor solutions containing 0–25% (w/v) glycerol prior to flash freezing in liquid nitrogen.

Structure Determination and Refinement—X-ray diffraction data were collected at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) on the 21-ID-F and 21-ID-G (Life Sciences Collaborative Access Team) beam lines. All data sets were recorded from cryoprotected crystals using a single wavelength at 100 K. Data were indexed and scaled with the program HKL2000 (41). Both crystal types belonged to space group *P6₅22*. X-ray diffraction data collection and processing statistics are listed in Table 3 (see below). Phases were calculated using MOLREP as a part of the CCP4 program suite (42, 43) based on a previously refined model (Protein Data Bank code 3OSN) (16). Refinements were performed using Refmac 6.0 with restrained and rigid body refinement (44, 45). Repeated cycles of manual rebuilding were performed in Coot (46). Structural images were generated in PyMOL (47).

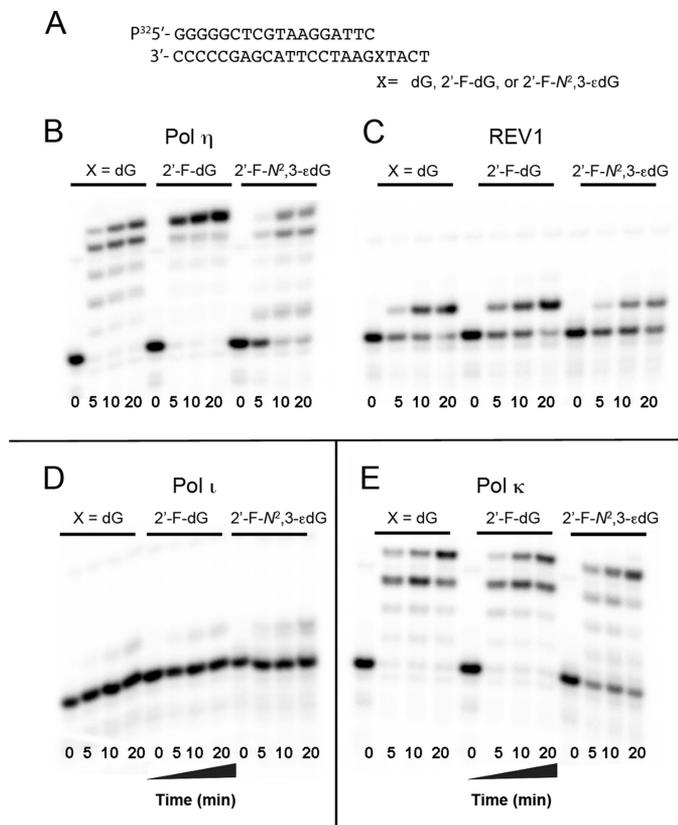


FIGURE 2. ^{32}P -labeled primer extension in the presence of four dNTPs by different Y-family human DNA pols. A, primer and template sequences where X is dG, 2'-F-dG, or 2'-F- $N^2,3$ -edG. B, pol η . C, REV1. D, pol ι . E, pol κ . Reactions were carried out for increasing times (as indicated) in 50 mM Tris-HCl buffer (pH 7.5) containing 60 nM primer-template complex, 250 μM dNTPs, 20 nM pol, 5% (v/v) glycerol, 5 mM DTT, 50 mM NaCl, 5 mM MgCl_2 , and 50 μg ml^{-1} BSA at 37 $^\circ\text{C}$.

RESULTS

Primer Extensions by Human Y-family Pols—Primer extension experiments were performed using a template containing stabilized 2'-F- $N^2,3$ -edG and compared with those using templates containing 2'-F-dG and unmodified dG (Fig. 2). Among the four pols, pol η appeared to be most efficient in terms of producing full-length products (5 bases extended). Although similar to pol η in efficiency, pol κ only produced products with 4 or fewer bases extended in a 20-min reaction. REV1 and pol ι extended the primer by 1 base within the same time frame. Quantitation of the extension products showed that the pol η -extended products constituted 93, 96, and 83% of the substrate for dG, 2'-F-dG, and 2'-F- $N^2,3$ -edG templates, respectively, upon 20-min incubation. The corresponding respective values were 73, 76, and 50% for REV1; 19, 10, and 10% for pol ι ; and 94, 94, and 70% for pol κ . The resistance to extension observed here is consistent with the known low processivities of pol ι and REV1 (9, 13).

Steady-state Kinetics of Nucleotide Incorporation—To determine the catalytic efficiency of incorporation of each dNTP catalyzed by different pols, steady-state kinetic analysis was performed. The catalytic efficiency (*i.e.* k_{cat}/K_m) and misinsertion frequency (*i.e.* $f = (k_{\text{cat}}/K_m(\text{dNTP})_{\text{incorrect}})/(k_{\text{cat}}/K_m(\text{dCTP}))$) are two useful parameters for comparing different polymerase reactions (Table 1). For the template containing 2'-F- $N^2,3$ -edG,

all three pols preferred to insert dCTP. Pol ι and pol η had relatively high misincorporation frequencies for T because of their marginally lower catalytic efficiencies compared with those for C insertion. For REV1, the order of preference for dNTP misinsertion opposite the lesion was $G > T > A$ (based on k_{cat}/K_m) but with much lower misincorporation frequencies compared with the three other pols (Table 1). When comparing the k_{cat}/K_m values for dCTP insertion opposite the lesion with those obtained for the dG template, pol η and pol ι showed ~ 40 -fold attenuation followed by about 10-fold attenuation for REV1, suggesting that the presence of the lesion did not significantly affect the deoxycytidyltransferase activity of REV1. The catalytic efficiencies for different dNTPs were similar in magnitude for the 2'-F-dG and dG templates, indicating that 2'-F modification did not dramatically perturb pol recognition. Among all four pols examined (including pol κ (38)), pol κ exhibited the highest relative efficiency (0.24) of nucleotide (dCTP) incorporation opposite 2'-F- $N^2,3$ -edG (compared with dCTP incorporation opposite unmodified dG) followed by REV1 (0.11), pol η (0.027), and pol ι (0.026) (Table 1). Pol η showed the highest absolute value of catalytic efficiencies for all three types of template, consistent with results seen in product extension experiments (see above).

Analysis of Pol ι Primer Extension Products by LC-MS/MS—Because the highest misinsertion frequency was observed for pol ι , we examined the extension products of pol ι reactions using LC-MS/MS. Previous procedures were followed using a uracil-containing primer (Fig. 3A), and the product was cleaved using uracil-DNA glycosylase to simplify the sequencing results obtained with collision-induced dissociation fragmentation. The most abundant species (-2 or -3 charge) were chosen for collision-induced dissociation analysis, and the identity of the product was established by matching the fragmentation pattern to the theoretical pattern obtained from a program linked to the mass spectrometry group at the University of Utah (48). By using a longer incubation time (3.5 h) and higher enzyme and substrate concentrations compared with those in the primer extension gel analyses, a greater portion of primer was extended by pol ι , and the products could be identified. As shown in Fig. 3, B–D, three products were identified with C, T, and A incorporated opposite 2'-F- $N^2,3$ -edG. Two additional products containing C and T opposite the lesion (Fig. 3, E and F, respectively) were identified as having an extra A at the end. The confirmed fragment ions are illustrated in the spectra with fragmentation patterns in the insets.

The relative yields of various products were calculated based on the peak areas of extracted LC-MS chromatograms (data not shown). The sum of the peak areas was used for the product, which existed in more than one charge state. Consistent with the low efficiency of pol ι seen in the primer extension gel analysis, the amount total of extended products (based on the total peak areas; data not shown) only accounted for 6% of the total products by Dpo4 formed under the same conditions (38). The major extension products were those containing T and C opposite 2'-F- $N^2,3$ -edG, and the other three are minor products (Table 2). Pol ι produced similar yields of extension products with C (41%) and T (52%) incorporated opposite the lesion. In addition, pol ι readily extended the 2'-F-dG and dG templates

Miscoding of *N*²,3-Ethenoguanine

TABLE 1
Steady-state kinetic analysis of polymerase-catalyzed single base insertion

The insertion was opposite *X* in the template sequence of 3'-CCCCGAGCATTCCTAAGXTACT-5' where *X* is 2'-F-*N*²,3- ϵ dG, 2'-F-dG, or dG.

Polymerase/template	dNTP	k_{cat} <i>min</i> ⁻¹	$K_{m(dNTP)}$ μ M	$k_{cat}/K_{m(dNTP)}$ <i>min</i> ⁻¹ μ M ⁻¹	f^a	Relative efficiency ^b	
pol ι	2'-F- <i>N</i> ² ,3- ϵ dG	C	0.26 ± 0.03	153 ± 58	0.0017		0.026
		T	0.33 ± 0.02	280 ± 36	0.0012	0.71	
		A	0.016 ± 0.001	956 ± 83	0.000017	0.010	
	2'-F-dG	G	0.007 ± 0.001	223 ± 71	0.000031	0.018	0.085
		C	0.079 ± 0.1	14 ± 4	0.0056		
		T	0.049 ± 0.005	156 ± 46	0.00031	0.055	
	dG	A	0.0067 ± 0.0011	88 ± 47	0.000076	0.014	1
		G	0.0040 ± 0.0006	44 ± 25	0.000091	0.016	
		C	0.093 ± 0.001	1.4 ± 0.5	0.066		
		T	0.63 ± 0.18	1090 ± 510	0.00058	0.0088	
		A	0.0052 ± 0.0007	111 ± 53	0.000047	0.00071	
		G	0.0098 ± 0.0009	365 ± 69	0.000027	0.00041	
Pol η	2'-F- <i>N</i> ² ,3- ϵ dG	C	0.88 ± 0.04	11 ± 2	0.08		0.027
		T	1.3 ± 0.06	26 ± 6	0.05	0.63	
		A	0.40 ± 0.03	63 ± 21	0.0063	0.079	
		G	0.30 ± 0.02	45 ± 9	0.0067	0.084	
	2'-F-dG	C	1.7 ± 0.06	0.51 ± 0.12	3.3		1.1
		T	0.70 ± 0.05	49 ± 12	0.014	0.0042	
		A	0.96 ± 0.07	15 ± 6	0.064	0.019	
	dG	G	0.63 ± 0.05	18 ± 7	0.035	0.011	1
		C	2.3 ± 0.14	0.76 ± 0.23	3.0		
		T	0.51 ± 0.03	57 ± 11	0.0089	0.0030	
		A	0.88 ± 0.03	74 ± 9	0.012	0.0040	
		G	0.51 ± 0.03	50 ± 9	0.010	0.0033	
REV1	2'-F- <i>N</i> ² ,3- ϵ dG	C	0.083 ± 0.005	4.8 ± 1.4	0.017		0.11
		T	0.024 ± 0.003	24 ± 2	0.0010	0.059	
		A	0.012 ± 0.001	32 ± 8	0.00038	0.022	
		G	0.060 ± 0.018	44 ± 13	0.0013	0.080	
	2'-F-dG	C	0.17 ± 0.01	1.6 ± 0.5	0.11		0.73
		T	0.37 ± 0.02	90 ± 13	0.0041	0.037	
		A	0.043 ± 0.004	27 ± 10	0.0016	0.015	
	dG	G	0.89 ± 0.03	10 ± 2	0.0089	0.081	1
		C	0.12 ± 0.01	0.81 ± 0.37	0.15		
		T	0.024 ± 0.001	7.8 ± 4.2	0.0031	0.021	
		A	0.043 ± 0.004	27 ± 11	0.0016	0.011	
		G	0.49 ± 0.03	130 ± 14	0.0037	0.025	
Pol κ^c	2'-F- <i>N</i> ² ,3- ϵ dG	C	1.6 ± 0.1	73 ± 13	0.022		0.24
		T	0.90 ± 0.04	111 ± 14	0.0082	0.37	0.091
		A	0.063 ± 0.002	55 ± 7	0.0011	0.05	
		G	0.22 ± 0.01	210 ± 22	0.0010	0.045	
	2'-F-dG	C	1.9 ± 0.1	2.8 ± 0.3	0.68		1
		dG	1.8 ± 0.1	20 ± 1	0.090		

^a Misinsertion frequency: $f = (k_{cat}/K_{m(dNTP)})_{incorrect} / (k_{cat}/K_{m(dCTP)})$.

^b Relative efficiency calculated as the ratio of the k_{cat}/K_{m} for dNTP insertion opposite the adduct to the k_{cat}/K_{m} for dCTP insertion opposite dG.

^c From Ref. 38.

in an error-free manner. The base insertion pattern obtained from LC-MS/MS analysis agrees with the steady-state kinetic analysis (Table 1) with T insertions being the major misincorporation events for pol ι -catalyzed bypass.

Crystal Structures of Pol ι with Oligonucleotides Containing 2'-F-*N*²,3- ϵ dG and dCTP or dTTP—To understand the base pairing mechanisms for dCTP and dTTP observed above, we conducted co-crystallization experiments with pol ι , a template containing 2'-F-*N*²,3- ϵ dG, and dCTP or dTTP. Two types of crystals were obtained, *i.e.* pol ι -1 (pol ι 2'-F-*N*²,3- ϵ dG:dCTP; Protein Data Bank code 4FS2) and pol ι -2 (pol ι 2'-F-*N*²,3- ϵ dG:dTTP; Protein Data Bank code 4FS1). Structures of these two ternary complexes were determined by molecular replacement using a previously refined model (Protein Data Bank code 3OSN) (16) without the lesion and the incoming nucleotide (Table 3). Clear electron densities around the 2'-F-*N*²,3- ϵ dG and incoming nucleotide facilitated the unbiased determina-

tion of the base pairing conformations at the active site. Similar to several structures seen previously for template native purines (7, 15) and adducted purines (8, 9, 16), the electron density around 2'-F-*N*²,3- ϵ dG indicated that the lesion was rotated from the *anti* to the *syn* conformation (Fig. 4, A and C). This substrate-induced conformational change of template purine is thought to be dictated by the rigid active site of pol ι (7).

In both the pol ι -1 (Fig. 4A) and pol ι -2 (Fig. 4C) structures, the incoming nucleotide served as a donor in hydrogen bonds with the Hoogsteen edge of 2'-F-*N*²,3- ϵ dG (*i.e.* the O6 and N7 atoms). When 2'-F-*N*²,3- ϵ dG was paired with dCTP (Fig. 4B), one hydrogen bond was observed between the N4 atom of dCTP and O6 atom of 2'-F-*N*²,3- ϵ dG as indicated by a 2.5-Å distance. The possibility of a second hydrogen bond cannot be ruled out based on the distance (3.0 Å) between the N3 atom of dCTP and the N7 atom of 2'-F-*N*²,3- ϵ dG provided that the N3 atom of dCTP is protonated. This mechanism was proposed in

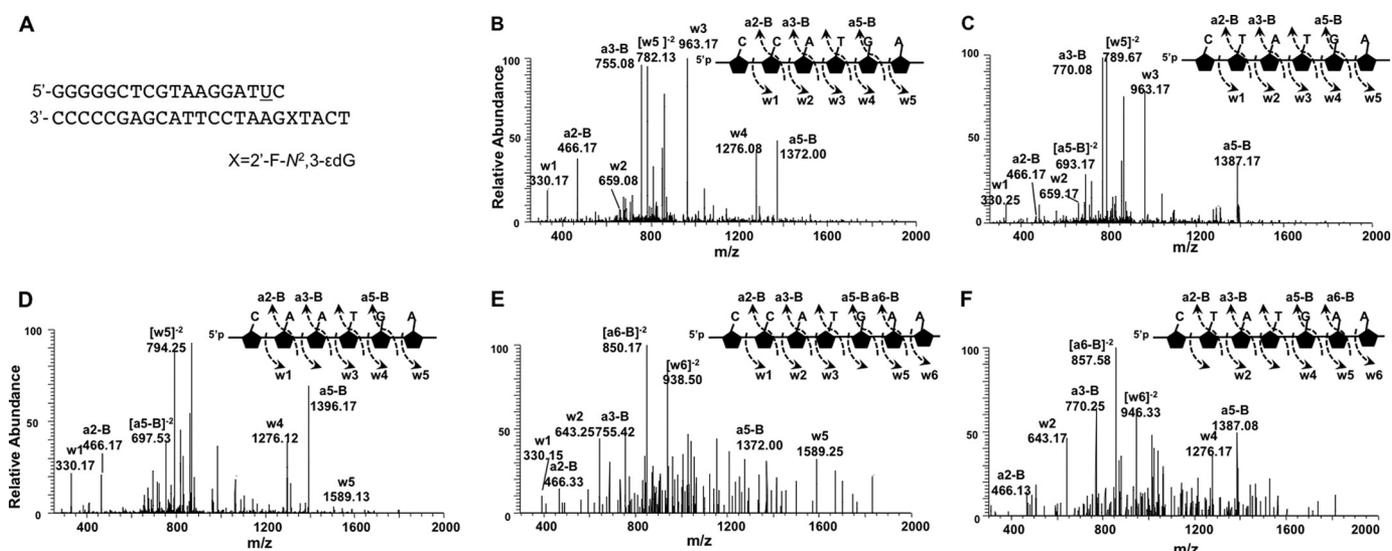


FIGURE 3. Collision-induced dissociation spectra from LC-MS/MS analysis of the full-length extension assays using pol ι and a 23-mer oligomer template containing 2'-F- $N^2,3$ - ϵ dG in the presence of all four dNTPs. A, template and primer sequences. The confirmed product sequences with the fragmentation patterns shown are 5'-CCATGA-3' (B), 5'-CTATGA-3' (C), 5'-CAATGA-3' (D), 5'-CCATGAA-3' (E), and 5'-CTATGAA-3' (F). The reaction contained 12.5 μ M DNA complex, 10 mM dNTPs, 10 μ M pol ι , 5 mM DTT, 50 mM NaCl, 5 mM MgCl₂, and 50 μ g ml⁻¹ BSA and were incubated at 37 °C for 3.5 h. Underlined U indicates the cutting site by uracil-DNA glycosylase after DNA polymerase reactions.

TABLE 2

Summary of pol ι extension products from LC-MS/MS analysis

The results were obtained using template · primer complexes containing 2'-F- $N^2,3$ - ϵ dG, 2'-F-dG, or dG. Underlined nucleotides indicate the base incorporated opposite the lesion.

3'-CCCCCGAGCATTCCTAAGXTACT	% of product	Base inserted
5'-GGGGGCTCGTAAGGATUC		
X= 2'-F- $N^2,3$ - ϵ -dG		
CCATGA	36	
CCATGAA	5	C
CTATGA	43	
CTATGAA	9	T
CAATGA	7	A
2'-F-dG		
CCATGA	100	C
dG		
CCATGA	100	C

previous studies for dCTP paired with G (15) or adducted G (16).

In the case of 2'-F- $N^2,3$ - ϵ dG-paired dTTP (Fig. 4D), it is likely that only one hydrogen bond exists between the O6 atom of 2'-F- $N^2,3$ - ϵ dG and the N3 atom of dTTP with a distance of 2.8 Å. Although a distance of 3.2 Å between N7(2'-F- $N^2,3$ - ϵ dG) and N3(dTTP) can also be interpreted as a potential hydrogen bonding distance, the asymmetry between the two distances (2.8 *versus* 3.2 Å) makes it unlikely that a bifurcated hydrogen bonding structure is present. The slightly longer distance (2.8 Å) in the case of 2'-F- $N^2,3$ - ϵ dG:T hydrogen bond may be an indication of its weaker strength compared with the 2.5-Å distance seen in the 2'-F- $N^2,3$ - ϵ dG:C pair, although the resolution limit of 2.5 Å does not permit a firm conclusion in this respect. Together with the possibility of two hydrogen bonds in the 2'-F- $N^2,3$ - ϵ dG:C pair, the base pair modes observed here are consistent with the slightly favorable insertion of C observed in the steady-state kinetic analysis.

Irrespective of the incoming nucleotide, the pol ι -1 and pol ι -2 structures are quite similar with a root mean square deviation

value of 0.27 Å for all atom pairs upon superimposition. The superimposition of the incoming nucleotides suggested a movement of 2'-F- $N^2,3$ - ϵ dG toward the minor groove for the 2'-F- $N^2,3$ - ϵ dG:T base pair (Fig. 5A). The pol ι -1 structure superimposes with the native G:C complex (Protein Data Bank code 2ALZ; Ref. 15) with a root mean square deviation value of 0.27 Å (Fig. 5B), indicating that the presence of the lesion (2'-F- $N^2,3$ - ϵ dG) did not significantly affect the conformations of the protein and the nucleic acid. The conformation of the 2'-F- $N^2,3$ - ϵ dG:C pair also resembles that of an N^2 -ethylguanine:C pair crystallized with pol ι (9) (Fig. 5C). When the conformation of the 2'-F- $N^2,3$ - ϵ dG:T mispair is compared with an O^6 -methylguanine:T base pair (16) (because of the lack of a structure with the pol ι G:T pair in the Protein Data Bank), the conformations of both base pairs are similar except that the lesion is slightly moved toward the minor groove in the case of the 2'-F- $N^2,3$ - ϵ dG:T pair (Fig. 5D, green). This shift could be due to the bulkier size of 2'-F- $N^2,3$ - ϵ dG compared with O^6 -methylguanine. Overall, pol ι appears to be able to accommodate the 2'-F- $N^2,3$ - ϵ dG pair rather well at the active site without significant protein and nucleic acid conformational changes. The structures showed that both C and T pair with the lesion in a similar fashion in line with the observation that pol ι promoted both error-free and error-prone bypass in steady-state kinetic and LC-MS/MS analyses.

DISCUSSION

The DNA adduct $N^2,3$ - ϵ dG is a ubiquitous modification produced from endogenous processes (*e.g.* lipid peroxidation) or exposure to environmental pollutants (*e.g.* vinyl chloride or urethane). We recently developed an isostere approach to incorporate the stabilized analog (2'-F- $N^2,3$ - ϵ dG) into oligonucleotides and investigated the miscoding potential of $N^2,3$ - ϵ dG using several prokaryotic and eukaryotic DNA pols (38). In the present work, we extended our previous investigation into the

TABLE 3

Crystal data collection and refinement statistics for the ternary complexes pol ι -1 (pol ι -2'-F- $N^2,3$ - ϵ dG:dCTP, Protein Data Bank code 4FS2) and pol ι -2 (pol ι -2'-F- $N^2,3$ - ϵ dG:dTTP, Protein Data Bank code 4F51)

	Pol ι -1 (2'-F- $N^2,3$ - ϵ dG:dCTP)	Pol ι -2 (2'-F- $N^2,3$ - ϵ dG:dTTP)
Data collection		
Beamline	21-ID-F	21-ID-G
Space group	$P6_322$	$P6_322$
Unit cell (a, b, c) (Å)	97.30, 97.30, 202.91	97.47, 97.47, 203.54
Unit cell (α, β, γ) (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å) ^a	2.05 (2.05–2.09)	2.49 (2.49–2.53)
No. of measured reflections	36,517	23,120
No. of unique reflections	36,444	23,074
Percent possible (%)	99.8 (98.4)	99.8 (100)
Redundancy	7.8 (6.3)	6.9 (6.8)
R_{linear}^b	0.051 (0.581)	0.067 (0.505)
Signal to noise ($I/\sigma I$)	32.5 (2.5)	23.0 (3.2)
Coordinate composition (asymmetric unit)		
No. of protein molecules	1	1
No. of amino acid residues	383	383
No. of water molecules	179	65
No. of Mg ²⁺ ions	3	2
No. of template nucleotides	8	9
No. of primer nucleotides	7	7
No. of dCTP	1	0
No. of dTTP	0	1
Refinement		
Resolution range (Å)	30.00–2.05	30.00–2.50
Reflections	34,557	19,448
R_{work} (%) ^c	21.4	21.5
R_{free} (%) ^d	26.0	27.0
Root mean square deviation bond length (Å)	0.018	0.014
Root mean square deviation bond angle (°)	2.05	1.99
Mean B-factor	51.3	56.7
Wilson B-factor	42.8	49.4
Ramachandran summary		
In preferred regions	354 (93.40%)	350 (92.35%)
In allowed regions	16 (4.22%)	23 (6.07%)
Outliers	9 (2.37%)	5 (1.58%)

^a Values for highest resolution bin are given in parentheses.^b $R_{\text{linear}} = \sum |I - \langle I \rangle| / \sum I$ where I is the integrated intensity of a given reflection.^c $R_{\text{work}} = \sum |F_{\text{observed}} - F_{\text{calculated}}| / \sum |F_{\text{observed}}|$.^d R_{free} was calculated using 5% test size with random selection.

other three human Y-family DNA pols and provided the structural basis of the most error-prone bypass enzyme, pol ι .

Primer extension gel analysis generated a qualitative comparison of the capability of bypassing 2'-F- $N^2,3$ - ϵ dG by Y-family pols (Fig. 1). The order of bypassing efficiency (from the percentage of total product extended) is pol η > pol κ > REV1 \approx pol ι . Compared with pol κ and pol ι , the higher activity of pol η copying past $N^2,3$ - ϵ G observed here is similar to that seen previously for other etheno adducts, *i.e.* 1, N^2 - ϵ G (29), 1, N^6 - ϵ A (26, 49), and 3, N^4 - ϵ C (50). With regard to DNA polymerases, the extension pattern is particularly similar to that of bypass of 1, N^2 - ϵ G; *i.e.* pol η readily extended the primer into full-length products, whereas pol ι and pol κ showed some single base incorporation (29).

Steady-state kinetic analysis established the preferred base incorporated opposite the lesion and provided a kinetic rationale for primer extension experiments (Table 1). For all four human Y-family DNA pols, the correct base C is marginally preferred opposite 2'-F- $N^2,3$ - ϵ dG with similar relative efficiencies in comparison with the insertion of C opposite a regular G (Table 1). The misinsertion of T is consistent for all four human Y-family DNA pols as well as for several other prokaryotic and eukaryotic DNA polymerases (38). The highest absolute value of catalytic efficiency (k_{cat}/K_m) seen (for pol η) is in line with primer extension results, which may be partly explained by the more open active site of pol η compared with other poly-

merases (51). The pattern of fidelity for pol η bypassing different etheno lesions is similar: both error-free and error-prone syntheses have been observed. Pol η inserted a C opposite $N^2,3$ - ϵ G in a marginally error-free manner with a misinsertion frequency of 0.63 for T (Table 1). Similarly, pol η copied past 1, N^6 - ϵ A in the order of preference T > C > A > G (49). The order was G > A > C for 1, N^2 - ϵ G (29) and A \approx G > C \approx T for 3, N^4 - ϵ C (50). Pol ι has the highest misincorporation frequency (although C is preferred 1.5-fold compared with T), which is consistent with the view that pol ι generally catalyzes error-prone bypass (3). The incorporation patterns seen for pol ι bypassing other etheno DNA adducts are as follows: pol ι somewhat prefers to incorporate C opposite 1, N^6 - ϵ A (8) and inserts both C and T opposite 1, N^2 - ϵ G with almost the same catalytic efficiencies (29). The fact that REV1 prefers to catalyze dCTP insertion is not surprising in that REV1 utilizes its G-loop to hydrogen bond with template G and an Arg in another segment (N-digit) to ensure the incorporation of dCTP (12). When comparisons are made with the catalytic efficiency of dCTP insertions opposite native G in the template, the order of relative efficiency is pol κ (0.24) > REV1 (0.11) > pol η (0.027) \approx pol ι (0.026) (Table 1), suggesting that 2'-F- $N^2,3$ - ϵ dG affects the DNA syntheses of the four Y-family pols to a similar extent.

LC-MS/MS analysis of the primer extension products by pol ι provided further insight into the nature of the bases inserted beyond the lesion in these error-prone reactions. With pol ι ,

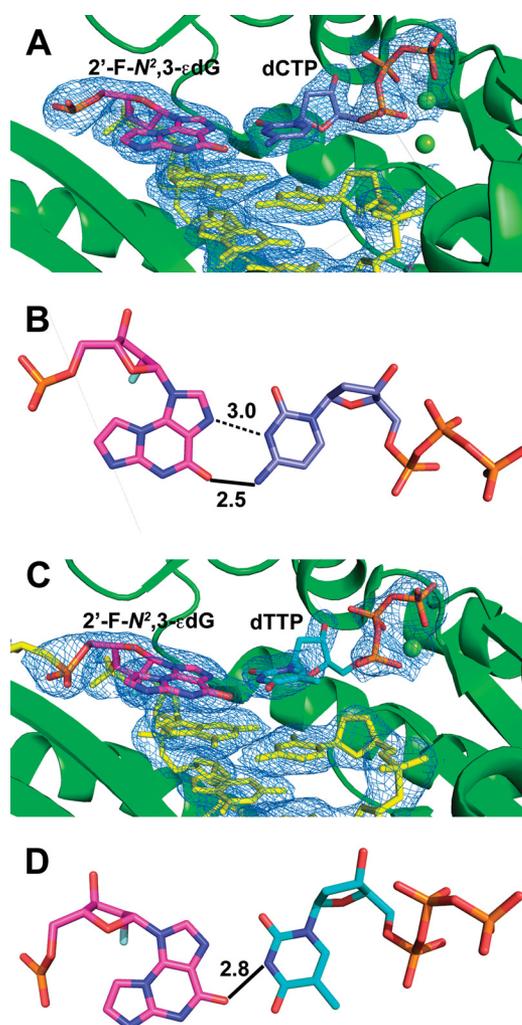


FIGURE 4. Crystal structures of ternary complexes pol ι -1 (pol ι -2'-F- $N^2,3$ - ϵ dG:dCTP, Protein Data Bank code 4FS2) and pol ι -2 (pol ι -2'-F- $N^2,3$ - ϵ dG:dTTP, Protein Data Bank code 4FS1). A, pol ι -1 active site with template 2'-F- $N^2,3$ - ϵ dG pairing with the incoming dCTP. Green spheres show the two observed Mg^{2+} ions. B, 2'-F- $N^2,3$ - ϵ dG in *syn* conformation in pol ι -1 with one hydrogen bond formed between the O6 atom of 2'-F- $N^2,3$ - ϵ dG and the N4 atom of dCTP (2.5 Å); the other potential hydrogen bond formed is indicated with a dashed line between the N7 atom of 2'-F- $N^2,3$ - ϵ dG and the N3 atom of dCTP (3.0 Å). C, pol ι -2 active site with template 2'-F- $N^2,3$ - ϵ dG pairing with incoming dTTP. The green sphere shows the observed single Mg^{2+} ion. D, 2'-F- $N^2,3$ - ϵ dG in the *syn* conformation in pol ι -2 with one hydrogen bond formed between the O6 atom of 2'-F- $N^2,3$ - ϵ dG and the N3 atom of dTTP (2.8 Å). The quality of the data is demonstrated using the Fourier $2F_o - F_c$ sum electron density map displayed (blue mesh) at 1.0 σ in A and C.

approximately half of the products contained T with a high fidelity extension beyond the lesion (Fig. 3 and Table 2). The observation of almost equal amounts of products containing C and T opposite 2'-F- $N^2,3$ - ϵ dG (with LC-MS/MS analysis) is in line with results from kinetic analysis (Table 1). The much lower amount of total extended products (6%) compared with Dpo4 (38) agrees with the low bypass efficiency of pol ι seen in the primer extension gel analysis (Fig. 2). These extension products are similar to the products generated by Dpo4 (38); however, the pattern of miscoding is considerably different from that generated by 1, N^2 - ϵ G, which yields mainly products with G inserted by human pol η (29) and -1 deletion products by Dpo4 (28).

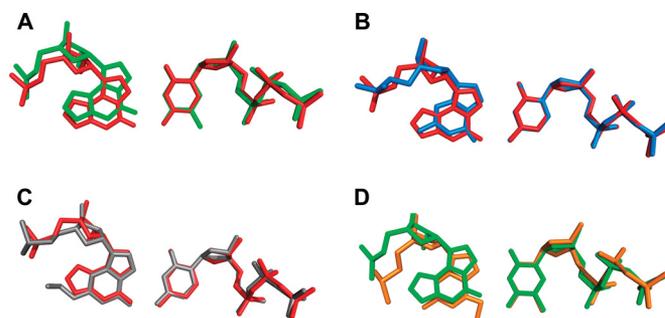


FIGURE 5. Comparisons of base pair positions at the active site of pol ι complexes based on the superimposition of incoming nucleotides. A, shearing of 2'-F- $N^2,3$ - ϵ dG toward the minor groove for the 2'-F- $N^2,3$ - ϵ dG:dTTP pair (green) compared with the 2'-F- $N^2,3$ - ϵ dG:dCTP pair (red). B, similarity between the 2'-F- $N^2,3$ - ϵ dG:dCTP pair (red) and the native G:dCTP pair (blue; Protein Data Bank code 2ALZ). C, structural similarity between the 2'-F- $N^2,3$ - ϵ dG:dCTP pair (red) and the N^2 -ethylguanine:dCTP pair (gray; Protein Data Bank code 3EPG) crystallized with pol ι ; D, shearing of 2'-F- $N^2,3$ - ϵ dG towards the minor groove for the 2'-F- $N^2,3$ - ϵ dG:dTTP pair (green) compared to O^6 -methylguanine:dTTP pair (16) (orange, Protein Data Bank code 3OSN).

The hydrogen bonding patterns of 2'-F- $N^2,3$ - ϵ dG:C and 2'-F- $N^2,3$ - ϵ dG:T base pairs seen in the crystal structures provided molecular explanations for the error-free and error-prone bypass of pol ι . The distance of 2.5 Å is a clear indication that a hydrogen bond is established between the O6 atom of 2'-F- $N^2,3$ - ϵ dG and the N4 atom of dCTP. The possibility of a second hydrogen bond also exists, *i.e.* between the N7 atom of 2'-F- $N^2,3$ - ϵ dG and the N3 atom of dCTP, provided that the N3 atom of dCTP is protonated. The tendency for protonation of the N3 atom of dCTP has been discussed in several other pol ι -DNA structures with both native and adducted purines in the templates (8, 15, 16). Although the N3 atom of free cytosine has a $pK_a \sim 4.5$, the local molecular environment could elevate the pK_a to 6.2–7.2 at a terminal position or >8.5 at an internal position in DNA triple helices (52, 53). Nair *et al.* (8) suggested that an elevation of the pK_a of dCTP could be due to the base-stacking and long range electrostatic interactions of the active site residues Asp-126 and Glu-127. The one hydrogen bond observed in the 2'-F- $N^2,3$ - ϵ dG:T pair is an indication that the 2'-F- $N^2,3$ - ϵ dG:T pair might be less stable compared with the 2'-F- $N^2,3$ - ϵ dG:C pair. Our crystallization attempts are consistent with this view in that pol ι -1 type crystals (with dCTP) grew more easily and diffracted to higher resolution than the pol ι -2 crystal (with dTTP). Collectively, the difference in hydrogen bonding may explain a slightly higher catalytic efficiency for dCTP by pol ι .

The typical strategy that pol ι uses a Hoogsteen base pairing mechanism to accommodate native and adducted purines was once again demonstrated in both the pol ι -1 and pol ι -2 structures albeit with different hydrogen bonding schemes. The similarity of the two structures consists of their use of the Hoogsteen edge of 2'-F- $N^2,3$ - ϵ dG to hydrogen bond with the incoming nucleotide. The conformation of 2'-F- $N^2,3$ - ϵ dG:C also resembles G:dCTP (15) and N^2 -ethylG:dCTP (9) at the pol ι active site.

However, the base pair conformations seen here are quite different from what has been observed at the active site of Dpo4 (38). Specifically, the 2'-F- $N^2,3$ - ϵ dG:C pair adopts a Watson-

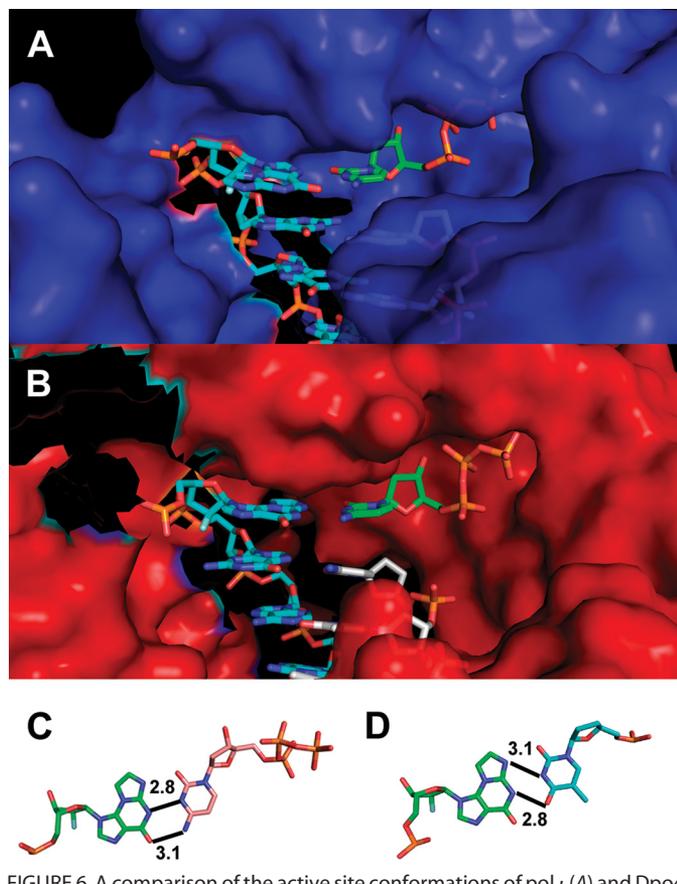


FIGURE 6. A comparison of the active site conformations of pol ι (A) and Dpo4 (B) (protein shown in surface view) is shown. The conformations of 2'-F- $N^2,3$ - ϵ dG paired with incoming nucleotides at the Dpo4 active site are shown in C (dCTP) and D (dTTP) (from Ref. 38).

Crick-like conformation, and the 2'-F- $N^2,3$ - ϵ dG:T structure contains a sheared base pair at the Dpo4 active site (Fig. 6, C and D, and Ref. 38). That 2'-F- $N^2,3$ - ϵ dG was observed to be positioned in the *anti* conformation by Dpo4 is likely due to the relatively open active site compared with pol ι (17) (surface view shown in Fig. 6, A and B). Particularly, the residues adjacent to the template base are bulkier (Leu-62, Val-64, and Gln-59) in pol ι compared with Dpo4 (Ala-42, Ala-44, and Val-32) (54). In the Dpo4 structure, hydrophobic interactions are likely to exist between Val-32 and the imidazole ring of 2'-F- $N^2,3$ - ϵ dG (*anti*). Conversely, residues (Leu-62, Val-64, and Gln-59) may force 2'-F- $N^2,3$ - ϵ dG to rotate into the *syn* conformation, which would otherwise clash with these residues if the lesion were positioned in the *anti* conformation. Despite these conformational differences, similar extents of T misinsertion are observed in both cases.

As mentioned in the Introduction, Singer and co-workers (30, 32, 35) reported three studies on the miscoding of $N^2,3$ - ϵ G more than 20 years ago. These studies were limited by the general methods available for studying miscoding at the time as well as the inherent lability of the glycosidic bond of $N^2,3$ - ϵ dG. The uncorrected mutation frequency for $N^2,3$ - ϵ dG inserted into an M13 phage system was only 0.5%, but in that study (32), an *in vitro* study with a polyribo(G/ $N^2,3$ - ϵ G) template and a reverse transcriptase (35), and a study on "reverse" incorporation of $N^2,3$ - ϵ dG triphosphate (30), the general pattern was

pairing of $N^2,3$ - ϵ G with T and C. This pattern, despite any deficiencies in the earlier work, is similar to those seen in our own studies (Ref. 38 and the present work). The $N^2,3$ - ϵ G:T wobble pairing proposed in that early work (35) had no experimental basis and has not been observed in our crystal structures with Dpo4 (38) or human pol ι (Fig. 4).

More recently, theoretical studies (55) have predicted that G should be the base most likely to pair with $N^2,3$ - ϵ G followed by T > A > C, a prediction that is clearly inconsistent with the results obtained with all DNA polymerases thus far (Tables 1 and 2) (38). The pairing patterns predicted in the theoretical study (55) are also inconsistent with our $N^2,3$ - ϵ G:C and $N^2,3$ - ϵ G:T structures observed in the Dpo4 (38) and human pol ι (Fig. 4) crystals.

As mentioned in the Introduction, the goal of the 2'-fluoro substitution was to stabilize the glycosidic bond by destabilizing the transition state leading to an oxocarbenium-like intermediate in hydrolysis. The substitution was clearly successful in stabilizing the residue in oligonucleotides (38). Although miscoding by $N^2,3$ - ϵ G (specifically, 2'-F- $N^2,3$ - ϵ dG) was clearly demonstrated relative to both dG and 2'-F-dG (Table 1), it should be noted that the substitution of fluorine for hydrogen at the C2' sugar position is not without effect; *i.e.* the substitution caused up to a 12-fold change in k_{cat}/K_m (primarily in the K_m parameter) among four Y-family DNA polymerases: an ~8-fold decrease of k_{cat}/K_m with pol κ and a ~12-fold increase of k_{cat}/K_m with pol ι but no changes with pol η or REV1. Therefore, 2'-fluoro substitution seems to slightly interfere with pol ι activity but to facilitate pol κ activity, which might be related to a possible stabilizing effect of 2'-fluorine to exert a (intra- and/or inter-residual) pseudo-hydrogen bonding interaction with purine H8 as shown previously with 2'-fluoroarabinonucleic acid (56–58). Such a conformational effect (preferentially to an *anti* conformation) by 2'-F at dG might affect catalysis differently with the various polymerases by interfering with the (*syn-anti*) Hoogsteen base pairing adopted by pol ι but facilitating the (*anti-anti*) Watson-Crick base pairing utilized by pol κ (albeit not with pol η). Nevertheless, these points regarding the influence of the fluorine do not affect our conclusions about the miscoding properties of $N^2,3$ - ϵ G reported here.

In conclusion, we have utilized a recently developed stabilized analog, 2'-F- $N^2,3$ - ϵ dG, to discern the mutation potential of a ubiquitous but unstable DNA lesion, $N^2,3$ - ϵ dG. Kinetic and extension analyses allow qualitative and quantitative assessments of the miscoding pattern of this lesion for Y-family DNA polymerases, which are particularly relevant to translesion synthesis. Structural insights provided the molecular bases of error-free and error-prone synthesis by pol ι . The consistency of T misinsertion with all polymerases studied thus far underscores the miscoding potential of $N^2,3$ - ϵ G. The miscoding for T suggests the relevance of $N^2,3$ - ϵ G to vinyl chloride-induced angiosarcomas in which prevailing GC to AT transition mutations were found in the second base of codon 13 of the *K-ras* gene (59). Our study supports the hypothesis that $N^2,3$ - ϵ G may contribute to the carcinogenesis of vinyl chloride and inflammation-driven malignancies (25, 37).

Acknowledgment—The use of the Advanced Photon Source and Life Sciences Collaborative Access Team Sector 21 was supported by United States Department of Energy Grant DE-AC02-06CH11357 and Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor Grant 085P1000817.

REFERENCES

- Lord, C. J., and Ashworth, A. (2012) The DNA damage response and cancer therapy. *Nature* **481**, 287–294
- Bebenek, K., and Kunkel, T. A. (2004) Functions of DNA polymerases. *Adv. Protein Chem.* **69**, 137–165
- Hubscher, U., Maga, G., and Spadari, S. (2002) Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **71**, 133–163
- Lange, S. S., Takata, K., and Wood, R. D. (2011) DNA polymerases and cancer. *Nat. Rev. Cancer* **11**, 96–110
- Johnson, R. E., Washington, M. T., Prakash, S., and Prakash, L. (2000) Fidelity of human DNA polymerase η . *J. Biol. Chem.* **275**, 7447–7450
- Yang, W. (2011) Surviving the sun: Repair and bypass of DNA UV lesions. *Protein Sci.* **20**, 1781–1789
- Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2006) An incoming nucleotide imposes an *anti* to *syn* conformational change on the templating purine in the human DNA polymerase- ι active site. *Structure* **14**, 749–755
- Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2006) Hoogsteen base pair formation promotes synthesis opposite the $1,N^6$ -ethenodeoxyadenosine lesion by human DNA polymerase ι . *Nat. Struct. Mol. Biol.* **13**, 619–625
- Pence, M. G., Blans, P., Zink, C. N., Hollis, T., Fishbein, J. C., and Perrino, F. W. (2009) Lesion bypass of N^2 -ethylguanine by human DNA polymerase ι . *J. Biol. Chem.* **284**, 1732–1740
- Fukuda, H., Takamura-Enya, T., Masuda, Y., Nohmi, T., Seki, C., Kamiya, K., Sugimura, T., Masutani, C., Hanaoka, F., and Nakagama, H. (2009) Translesional DNA synthesis through a C8-guanyl adduct of 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) *in vitro*: REV1 inserts dC opposite the lesion, and DNA polymerase κ potentially catalyzes extension reaction from the 3'-dC terminus. *J. Biol. Chem.* **284**, 25585–25592
- Minko, I. G., Harbut, M. B., Kozekov, I. D., Kozekova, A., Jakobs, P. M., Olson, S. B., Moses, R. E., Harris, T. M., Rizzo, C. J., and Lloyd, R. S. (2008) Role for DNA polymerase κ in the processing of N^2 - N^2 -guanine inter-strand cross-links. *J. Biol. Chem.* **283**, 17075–17082
- Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009) Structure of the human Rev1-DNA-dNTP ternary complex. *J. Mol. Biol.* **390**, 699–709
- Choi, J. Y., and Guengerich, F. P. (2008) Kinetic analysis of translesion synthesis opposite bulky N^2 - and O^6 -alkyl guanine DNA adducts by human DNA polymerase REV1. *J. Biol. Chem.* **283**, 23645–23655
- Yang, W., and Woodgate, R. (2007) What a difference a decade makes: insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15591–15598
- Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2005) Human DNA polymerase ι incorporates dCTP opposite template G via a G-C+ Hoogsteen base pair. *Structure* **13**, 1569–1577
- Pence, M. G., Choi, J. Y., Egli, M., and Guengerich, F. P. (2010) Structural basis for proficient incorporation of dTTP opposite O^6 -methylguanine by human DNA polymerase ι . *J. Biol. Chem.* **285**, 40666–40672
- Sale, J. E., Lehmann, A. R., and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **13**, 141–152
- Kochetko, N. K., Shibaev, V. N., and Kost, A. A. (1971) New reaction of adenine and cytosine derivatives, potentially useful for nucleic acids modification. *Tetrahedron Lett.* **12**, 1993–1996
- Secrist, J. A., 3rd, Barrio, J. R., Leonard, N. J., and Weber, G. (1972) Fluorescent modification of adenosine-containing coenzymes: biological activities and spectroscopic properties. *Biochemistry* **11**, 3499–3506
- Sattsangi, P. D., Leonard, N. J., and Frihart, C. R. (1977) $1,N^2$ -Ethenoguanine and $N^2,3$ -ethenoguanine synthesis and comparison of electronic spectral properties of these linear and angular triheterocycles related to Y bases. *J. Org. Chem.* **42**, 3292–3296
- Leonard, N. J. (1993) Etheno-bridged nucleotides in enzyme reactions and protein binding. *Chemtracts Biochem. Mol. Biol.* **4**, 251–284
- Fedtke, N., Boucheron, J. A., Walker, V. E., and Swenberg, J. A. (1990) Vinyl chloride-induced DNA adducts. 2. Formation and persistence of 7-(2'-oxoethyl)guanine and $N^2,3$ -ethenoguanine in rat-tissue DNA. *Carcinogenesis* **11**, 1287–1292
- Gonzalez-Reche, L. M., Koch, H. M., Weiss, T., Müller, J., Drexler, H., and Angerer, J. (2002) Analysis of ethenoguanine adducts in human urine using high performance liquid chromatography-tandem mass spectrometry. *Toxicol. Lett.* **134**, 71–77
- Lee, S. H., Arora, J. A., Oe, T., and Blair, I. A. (2005) 4-Hydroperoxy-2-nonenal-induced formation of $1,N^2$ -etheno-2'-deoxyguanosine adducts. *Chem. Res. Toxicol.* **18**, 780–786
- Swenberg, J. A., Lu, K., Moeller, B. C., Gao, L., Upton, P. B., Nakamura, J., and Starr, T. B. (2011) Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol. Sci.* **120**, S130–S145
- Hang, B., Chenna, A., Guliaev, A. B., and Singer, B. (2003) Miscoding properties of $1,N^6$ -ethanoadenine, a DNA adduct derived from reaction with the antitumor agent 1,3-bis(2-chloroethyl)-1-nitrosourea. *Mutat. Res.* **531**, 191–203
- Shibutani, S., Suzuki, N., Matsumoto, Y., and Grollman, A. P. (1996) Miscoding properties of $3,N^4$ -etheno-2'-deoxycytidine in reactions catalyzed by mammalian DNA polymerases. *Biochemistry* **35**, 14992–14998
- Zang, H., Goodenough, A. K., Choi, J. Y., Irimia, A., Loukachevitch, L. V., Kozekov, I. D., Angel, K. C., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2005) 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. *J. Biol. Chem.* **280**, 29750–29764
- Choi, J. Y., Zang, H., Angel, K. C., Kozekov, I. D., Goodenough, A. K., Rizzo, C. J., and Guengerich, F. P. (2006) Translesion synthesis across $1,N^2$ -ethenoguanine by human DNA polymerases. *Chem. Res. Toxicol.* **19**, 879–886
- Singer, B., Kuśmierk, J. T., Folkman, W., Chavez, F., and Dosanjh, M. K. (1991) Evidence for the mutagenic potential of the vinyl-chloride induced adduct, $N^2,3$ -etheno-deoxyguanosine, using a site-directed kinetic assay. *Carcinogenesis* **12**, 745–747
- Basu, A. K., Wood, M. L., Niedernhofer, L. J., Ramos, L. A., and Essigmann, J. M. (1993) Mutagenic and genotoxic effects of three vinyl chloride-induced DNA lesions: $1,N^6$ -ethanoadenine, $3,N^4$ -ethenocytosine, and 4-amino-5-(imidazol-2-yl)imidazole. *Biochemistry* **32**, 12793–12801
- Cheng, K. C., Preston, B. D., Cahill, D. S., Dosanjh, M. K., Singer, B., and Loeb, L. A. (1991) The vinyl-chloride DNA derivative $N^2,3$ -ethenoguanine produces G→A transitions in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9974–9978
- Akasaka, S., and Guengerich, F. P. (1999) Mutagenicity of site-specifically located $1,N^2$ -ethenoguanine in Chinese hamster ovary cell chromosomal DNA. *Chem. Res. Toxicol.* **12**, 501–507
- Moriya, M., Zhang, W., Johnson, F., and Grollman, A. P. (1994) Mutagenic potency of exocyclic DNA adducts: marked differences between *Escherichia coli* and simian kidney cells. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11899–11903
- Singer, B., Spengler, S. J., Chavez, F., and Kuśmierk, J. T. (1987) The vinyl chloride-derived nucleoside, $N^2,3$ -ethenoguanosine, is a highly efficient mutagen in transcription. *Carcinogenesis* **8**, 745–747
- Dosanjh, M. K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., and Singer, B. (1994) All four known cyclic adducts formed in DNA by the vinyl chloride metabolite chloroacetaldehyde are released by a human DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1024–1028
- Nair, U., Bartsch, H., and Nair, J. (2007) Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans. *Free Radic. Biol. Med.* **43**, 1109–1120
- Zhao, L., Christov, P. P., Kozekov, I. D., Pence, M. G., Pallan, P. S., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2012) Replication of $N^2,3$ -ethenoguanine by DNA polymerases. *Angew. Chem. Int. Ed. Engl.* **51**, 5466–5469

Miscoding of $N^2,3$ -Ethenoguanine

39. Choi, J. Y., and Guengerich, F. P. (2005) Adduct size limits efficient and error-free bypass across bulky N^2 -guanine DNA lesions by human DNA polymerase η . *J. Mol. Biol.* **352**, 72–90
40. Sagripanti, J. L., and Kraemer, K. H. (1989) Site-specific oxidative DNA damage at polyguanosines produced by copper plus hydrogen peroxide. *J. Biol. Chem.* **264**, 1729–1734
41. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
42. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
43. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760–763
44. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321
45. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D. Biol. Crystallogr.* **57**, 122–133
46. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132
47. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York
48. Rozenski, J., and McCloskey, J. A. (2002) SOS: a simple interactive program for *ab initio* oligonucleotide sequencing by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **13**, 200–203
49. Levine, R. L., Miller, H., Grollman, A., Ohashi, E., Ohmori, H., Masutani, C., Hanaoka, F., and Moriya, M. (2001) Translesion DNA synthesis catalyzed by human pol η and pol κ across $1,N^6$ -ethenodeoxyadenosine. *J. Biol. Chem.* **276**, 18717–18721
50. Singer, B., Medina, M., Zhang, Y., Wang, Z., Guliaev, A. B., and Hang, B. (2002) 8-(Hydroxymethyl)- $3,N^4$ -etheno-C, a potential carcinogenic glycidaldehyde product, miscodes in vitro using mammalian DNA polymerases. *Biochemistry* **41**, 1778–1785
51. Silverstein, T. D., Johnson, R. E., Jain, R., Prakash, L., Prakash, S., and Aggarwal, A. K. (2010) Structural basis for the suppression of skin cancers by DNA polymerase η . *Nature* **465**, 1039–1043
52. Asensio, J. L., Lane, A. N., Dhese, J., Bergqvist, S., and Brown, T. (1998) The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J. Mol. Biol.* **275**, 811–822
53. Plum, G. E., and Breslauer, K. J. (1995) Thermodynamics of an intramolecular DNA triple-helix: a calorimetric and spectroscopic study of the pH and salt dependence of thermally-induced structural transitions. *J. Mol. Biol.* **248**, 679–695
54. Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) Replication by human DNA polymerase- ι occurs by Hoogsteen base-pairing. *Nature* **430**, 377–380
55. Srinivasadesikan, V., Sahu, P. K., and Lee, S. L. (2011) Model calculations for the misincorporation of nucleotides opposite five-membered exocyclic DNA adduct: $N^2,3$ -ethenoguanine. *J. Phys. Chem. B* **115**, 10537–10546
56. Berger, I., Tereshko, V., Ikeda, H., Marquez, V. E., and Egli, M. (1998) Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymines: implications for conformational preorganization for duplex stability. *Nucleic Acids Res.* **26**, 2473–2480
57. Li, F., Sarkhel, S., Wilds, C. J., Wawrzak, Z., Prakash, T. P., Manoharan, M., and Egli, M. (2006) 2'-Fluoroarabino- and arabinonucleic acid show different conformations, resulting in deviating RNA affinities and processing of the heteroduplexes with RNA by RNase H. *Biochemistry* **45**, 4141–4152
58. Anzahae, M. Y., Watts, J. K., Alla, N. R., Nicholson, A. W., and Damha, M. J. (2011) Energetically important C–H \cdots F–C pseudohydrogen bonding in water: evidence and application to rational design of oligonucleotides with high binding affinity. *J. Am. Chem. Soc.* **133**, 728–731
59. Marion, M. J., De Vivo, I., Smith, S., Luo, J. C., and Brandt-Rauf, P. W. (1996) The molecular epidemiology of occupational carcinogenesis in vinyl chloride exposed workers. *Int. Arch. Occup. Environ. Health* **68**, 394–398