Enzymatic bypass and the structural basis of miscoding opposite the DNA adduct 1,N^2-ethenodeoxyguanosine by human DNA translesion polymerase η

Pratibha P. Ghodke[1,‡], Jyotirling R. Mali[2,‡], Amritraj Patra[1,‡], Carmelo J. Rizzo[1,2,3], F. Peter Guengerich[1,3], and Martin Egli[1,3,*]

From the [1]Department of Biochemistry, School of Medicine, [2]Department of Chemistry, College of Arts and Science, Vanderbilt University, Nashville, Tennessee, USA; [3]Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, USA

Edited by Patrick Sung

Etheno (ε)-adducts, *e.g.*, 1,N^2-ε–guanine (1,N^2-ε-G) and 1,N^6-ε–adenine (1,N^6-ε-A), are formed through the reaction of DNA with metabolites of vinyl compounds or with lipid peroxidation products. These lesions are known to be mutagenic, but it is unknown how they lead to errors in DNA replication that are bypassed by DNA polymerases. Here we report the structural basis of misincorporation frequencies across from 1,N^2-ε-G by human DNA polymerase (hpol) η. In single-nucleotide insertions opposite the adduct 1,N^2-ε-G, hpol η preferentially inserted dGTP, followed by dATP, dTTP, and dCTP. This preference for purines was also seen in the first extension step. Analysis of full-length extension products by LC-MS/MS revealed that G accounted for 85% of nucleotides inserted opposite 1,N^2-ε-G in single base insertion, and 63% of bases inserted in the first extension step. Extension from the correct nucleotide pair (C) was not observed, but the primer with A paired opposite 1,N^2-ε-G was readily extended. Crystal structures of ternary hpol η insertion-stage complexes with nonhydrolyzable nucleotides dAMPnPP or dCMPnPP showed a *syn* orientation of the adduct, with the incoming A staggered between adducted base and the 5'-adjacent T, while the incoming C and adducted base were roughly coplanar. The formation of a bifurcated H-bond between incoming dAMPnPP and 1,N^2-ε-G and T, compared with the single H-bond formed between incoming dCMPnPP and 1,N^2-ε-G, may account for the observed facilitated insertion of dGTP and dATP. Thus, preferential insertion of purines by hpol η across from etheno adducts contributes to distinct outcomes in error-prone DNA replication.

Exocyclic etheno (ε) nucleobase adducts, *e.g.*, 1,N^6-ε-A, 3,N^4-ε-C, N^2,3-ε-G, and 1,N^2-ε-G (1, 2), arise from the action of various *bis*-electrophiles with cellular DNA (3) (Fig. 1). Reactive species that generate such lesions and originate from exogenous sources include epoxides derived from the oxidation of vinyl chloride and other vinyl monomers (4, 5) (Fig. 2A), urethane, certain nitrosamines (6), and mucochloric

acid (7). In addition, etheno lesions are formed endogenously through interaction of nucleobases with lipid peroxidation-derived aldehydes and hydroxyalkenals (8–10) (Fig. 2B). Thus, etheno-DNA lesions have been detected in the human livers with nonalcoholic fatty liver disease, a cause of end-stage liver disease and hepatocellular carcinoma (11). Reactive oxygen species constitute an important factor in hepatocellular carcinoma pathogenesis and can react with polyunsaturated fatty acids that are derived from membrane phospholipids, *e.g.*, aldehydes such as 2,3-epoxy-4-hydroxynoneal and 4-hydroperoxynonenal react with DNA to form exocyclic etheno adducts (8, 10, 11). Other potential sources of ROS are inflammatory processes and cytochrome P450 (P450) 2E1. However, it has been established that hepatic etheno DNA adducts correlated significantly with 4-hydroxynonenal but not with levels of P450 2E1 (12).

1,N^2-ε-G was first synthesized in 1977 (13), and the 1,N^2-ε-G base was later found to be excreted in human urine, indicating its presence in cellular DNA and removal by glycosylases (14). 1,N^2-ε-G caused misincorporation and mutations with model DNA polymerases (pols) *in vitro* (*e.g.*, G→T and G→C base changes) and in bacterial (G→T, G→C and G→A base changes in *Escherichia coli*) and mammalian cells (G→A and G→T base changes) (15–17). However, the mutation rates following introduction of the 1,N^2-ε-G lesion into bacterial or mammalian cell lines were relatively low (16, 17), a finding that was subsequently attributed to some degree to excision of the lesioned base by glycosylases such as *E. coli* uracil DNA glycosylase (UDG) and human alkyl-N-purine DNA glycosylase (18). These results leave open the question of the mechanism of miscoding and the potential involvement of a wider spectrum of pols beyond the replicative enzymes initially tested. The relevance of the etheno adducts in mutagenesis and carcinogenesis (1, 19–21) provided a motivation to more closely examine the consequences of these lesions for DNA damage-inducible pols. The 1,N^2-ε-G adduct in which the two exocyclic carbon atoms prevent the formation of normal Watson–Crick H-bonds served as a useful model system for other exocyclic G lesions in such pol structure/function investigations (3, 22, 23).
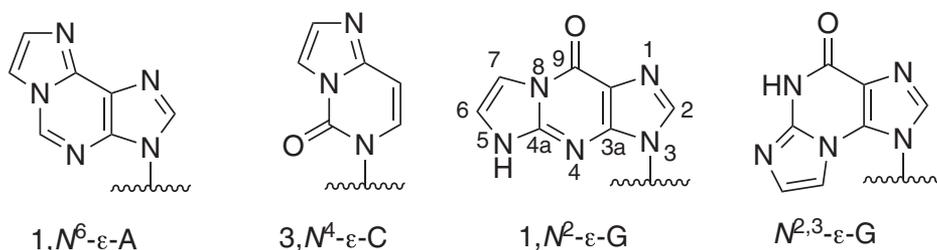
Figure 1. Major etheno (ε) adducts. The numbering pattern is shown for 1,N^2-ε-G.

In addition to the high-fidelity replicative pols, eubacteria, eukaryotes, and archaebacteria all have at their disposal the so-called translesion synthesis (TLS) or bypass polymerases (pols) (24–29). These Y-Family pols play major roles in replication past DNA damage, whereby TLS can proceed in both an error-prone and an error-free fashion. Y-family pols share the right-handed palm, finger, and thumb domains with replicative ones but feature an additional little finger domain and unique makeups of their active sites. *E. coli* possesses five pols, with pols IV and V being members of the Y-Family (30). Humans have at least 19 pols, of which pols η, κ, and ι and REV1 belong to the Y-Family. In the model archaeal organism *Sulfolobus solfataricus*, DNA pol IV (Dpo4) exhibits bypass behavior that is in some cases similar to that of human pol η (29). In initial work directed at the activity of a bypass pol past the 1,N^2-ε-G adduct, we
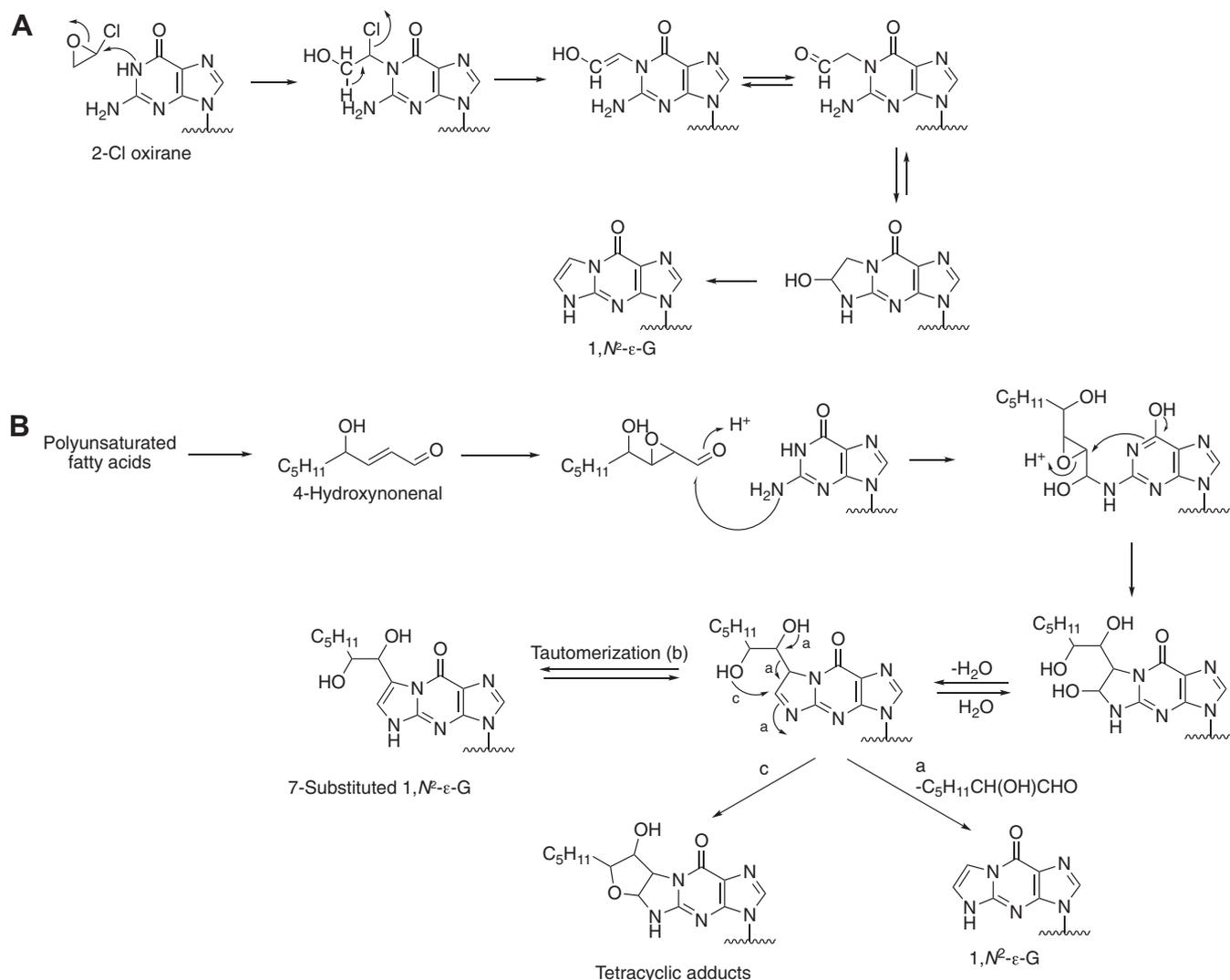
Figure 2. Mechanisms of formation of 1,N^2-ε-G in DNA and RNA. *A*, reaction with 2-chlorooxirane (vinyl chloride epoxide), based on ^{13}C labeling (5); *B*, reaction of the lipid peroxidation product 4-hydroxynonenal (11).

investigated the steady-state kinetics of Dpo4-catalyzed insertion and extension reactions (31). Experiments with individual dNTPs showed that A was the base preferentially inserted opposite $1,N^2$-ε-G. The products from full-length extension reactions were identified by an LC-MS/MS approach that revealed Dpo4 used multiple bypass mechanisms. Dpo4 extended primers containing $1,N^2$-ε-G to full length such that frameshifts predominated over A incorporation opposite the lesion. Thus, for a 3′-(1,N^2-ε-G)CACT-5′ template, 5′-GTGA-3′ (frameshift, no insertion opposite $1,N^2$-ε-G) was the major extension product and 5′-AGTGA-3′ was a minor product. Generation of the former entails a frameshift such that primer G base pairs with template C. Crystal structures of a binary Dpo4-DNA and three ternary Dpo4-DNA-ddNTP complexes revealed an *anti* conformation of $1,N^2$-ε-G that was stacked between adjacent base pairs but unopposed by a residue from the primer strand (31). In these so-called Type 2 structures, Dpo4 is able to accommodate two template bases in the active site, thereby trapping a frameshift intermediate and shedding light on the mechanism of incorporation into the growing primer of either A or G opposite T or C, respectively, located on the 5′-side of the lesion.

A subsequent study investigated the abilities of four human pols to synthesize past the $1,N^2$-ε-G adduct and their tendencies for error-prone or error-free bypass (32). One of them, the replicative human pol δ was completely blocked by the etheno lesion. The human Y-family pols κ and ι exhibited similar rates of either dTTP or dCTP incorporation. In comparison, hpol η was more efficient at bypassing the lesion than hpols κ and ι, but it also showed the highest error frequency, incorporating dGTP > dATP > dCTP. This order was maintained irrespective of whether the 5′-adjacent residue to the lesion was C or T. Therefore, Dpo4 and hpol η use different mechanisms of bypassing the $1,N^2$-ε-G lesion, whereby the latter pol appears to consistently insert a base opposite $1,N^2$-ε-G and then extend the primer in an error-free fashion.

In the present work, we conducted further insertion and extension experiments with the most active of the human bypass pols, pol η, opposite the $1,N^2$-ε-G lesion. The bypass behavior was assessed both with the 18mer 5′-d(CAT [1,N^2-ε-G]AT GAC GCT TCC CCC)-3′ (T(εG)A, **1**) and 5′-d(TCA T [1,N^2-ε-G]G AAT CCT TCC CCC)-3′ (T(εG)G, **2**) template strands, positioning the lesion between A and T and G and T, respectively. hpol η favored insertion of dGTP opposite the adduct in both sequence contexts, followed by dATP ≅ dTTP > dCTP for the 5′-T(εG)A-3′ (**1**) template and dATP > dTTP > dCTP for the 5′-T(εG)G-3′ (**2**) template. The ability of hpol η to extend from the lesion was tested after the correct (1,N^2-ε-G:C) base pair and a (1,N^2-ε-G:A) mispair. Interestingly, the polymerase was unable to extend from the former but capable of carrying out postlesion extension from the latter.

We determined crystal structures of two insertion-stage hpol η complexes with the template/primer duplex 5′-d(CAT [1,N^2-ε-G]AT GAC GCT)-3′/3′-d(TA CTG CGA)-5′ and either incoming dAMPnPP or dCMPnPP (nonhydrolyzable analogs of dATP and dCTP, respectively) in the presence of Mg^{2+}. The incoming dAMPnPP adopts a staggered orientation opposite $1,N^2$-ε-G in the *syn* orientation, with H-bond formation between the N6 atom of A and both the O6 atom of $1,N^2$-ε-G and the O4 atom of the 5′-T. In the complex with incoming dCMPnPP, the lesion also assumes the *syn* orientation but pairs opposite C with formation of an H-bond between the N4 atom of C and the O6 atom of $1,N^2$-ε-G.

The combined hpol η bypass and structural data with DNA template–primer duplexes containing the $1,N^2$-ε-G adduct contribute to a better understanding of the diverse mechanisms underlying the preferred insertion of dATP and dGTP by this pol opposite distinct lesions. In previous work directed at hpol η bypass of abasic sites and the $1,N^6$-ε-A adduct, we established the preferential insertion of dATP and dGTP opposite these lesions (33, 34), as seen here with the bypass of $1,N^2$-ε-G. However, structures of insertion stage complexes reveal that the "purine rule" with an abasic site involves interactions between the incoming nucleotide base and the template strand backbone that can also result in frameshifts (33). Both incoming dGTP and dATP stack onto the $1,N^6$-ε-A adduct, which adopts the *anti* conformation at the hpol η active site, whereas incoming dTTP is coplanar with the lesion in the *syn* conformation. Conversely, the preferred insertion of dGTP and dATP by hpol η occurs with the $1,N^2$-ε-G lesion in the *syn* conformation without the involvement of stacking between the lesioned and incoming nucleotide base moieties. $1,N^2$-ε-G maintains the *syn* orientation even opposite incoming dCTP, and unlike the case of bypass of abasic sites, the preferred insertion of purine nucleotides opposite $1,N^2$-ε-G is devoid of frameshifts.

Results

Translesion synthesis across from the 1,N²-ε-G adduct by hpol η

hpol η-mediated bypass of $1,N^2$-ε-G was performed using two different sequence contexts, 5′-T(εG)A-3′ (**1**) and 5′-T(εG)G-3′ (**2**) (see Table S1 for oligonucleotide sequences), employing full-length extension and single-nucleotide insertion assays. The oligonucleotide sequence context 5′-T(εG)A-3′ (**1**) was designed on the basis of the sequence used for crystallography. The 5′-T(εG)G-3′ (**2**) oligonucleotide was studied previously (32). Full-length primer extension reactions were performed in the presence of all four dNTPs using the respective primers (see Table S1 for oligonucleotide sequences). Under our experimental conditions, the bypass of $1,N^2$-ε-G was slower compared with unmodified templates (Fig. 3A and Fig. S4A) for both sequence contexts.

Single-nucleotide insertion assays were done utilizing individual dNTPs to determine the insertions across the adduct. The incorporation preference for the control templates was dCTP > dTTP > dGTP > dATP (Fig. 3B, Fig. S4B lanes 1–12). hpol η preferentially added dGTP across from $1,N^2$-ε-G, at almost similar rates for both sequence contexts. The incorporation preference for the 5′-T(εG)A-3′ (**1**) template

A

Primer_1 : 5′-FAM-GGGGGAAGCGTCAT-3′
Template : 3′-CCCCCTTCGCAGTAXTAC-5′

X = G   X = 1,N²-ε-G

B

X = G

C

X = 1,N²-ε-G
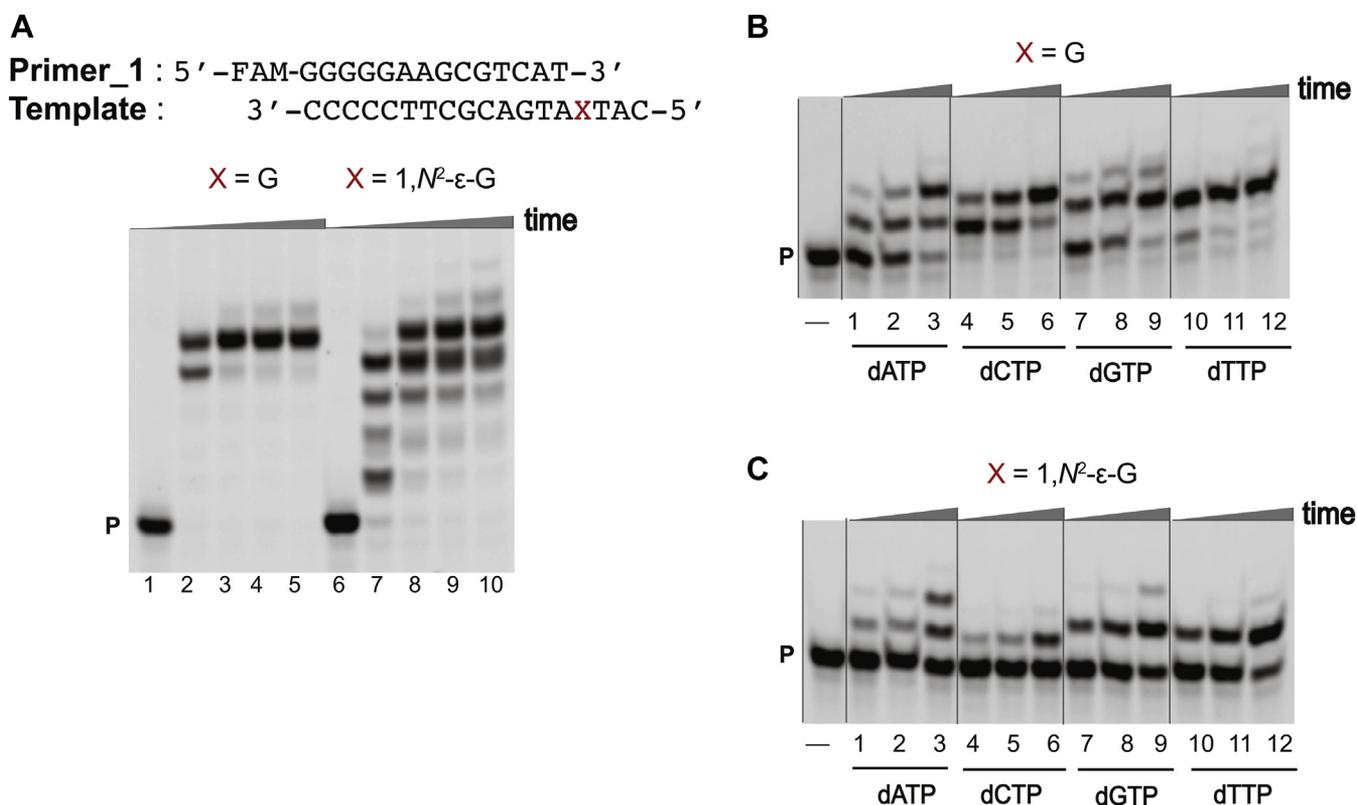
Figure 3. hpol η-mediated bypass across from 1,N²-ε-G in template 5′-T(εG)A-3′ (1). PAGE: 20%, 7 M urea. *A*, full-length extension assay: hpol η (120 nM) elongated Primer_1 opposite G, and 1,N²-ε-G-containing DNA templates in the presence of a mixture of dNTPs (500 μM). All reactions were done at 37 °C for 5-, 30-, 60-, and 120-min (time gradients indicated with wedges). Lanes: 1 to 5 for unmodified template; 6 to 10 for 1,N²-ε-G-modified template. Single-nucleotide insertion assays: hpol η (10 nM) was incubated with *B*, Primer_2/control template 1, and *C*, Primer_2/1,N²-ε-G modified template 1 (5′-T(εG)A-3′) as well as individual dNTPs (100 μM). Lanes: 1 to 3 for dATP, 4 to 6 for dCTP, 7 to 9 for dGTP, 10 to 12 for dTTP. All reactions were done at 37 °C for 5-, 10-, and 30-min. P indicates the FAM-labeled Primer_1. See Experimental procedures and Table S1 for the oligonucleotide sequences used.

was dGTP > dATP ≅ dTTP > dCTP (Fig. 3*C*, lanes 1–12) while for the 5′-T(εG)G-3′ (**2**) template, the preference was dGTP > dATP > dTTP > dCTP (Fig. S4*C*, lanes 1–12). These results show that hpol η tolerates 1,N²-ε-G in both sequence contexts.

Postlesion extension past the 1,N²-ε-G adduct by hpol η

hpol η-mediated postlesion extension for 1,N²-ε-G was performed for both the sequence contexts 5′-T(εG)A-3′ (**1**) and 5′-T(εG)G-3′ (**2**) (see Table S1 for oligonucleotide sequences) employing full-length extension and single-nucleotide insertion assays. We utilized two different primers for postlesion extension after the correct base pair (1,N²-ε-G:C) and a mispair (1,N²-ε-G:A). For a mis-incorporated base, A was used opposite the lesion because it showed the greatest misincorporation frequency, following that for dGTP.

Full-length primer extension reactions were performed in the presence of all four dNTPs using the respective primers (see Table S1 for oligonucleotide sequences). As a result, for both the sequence contexts, under our experimental conditions, extension from the 1,N²-ε-G:C pair was slower (Fig. 4*A* and Fig. S5*A*, lane 7) compared with the unmodified templates (Fig. 4*A* and Fig. S5*A*, lane 2). For the 1,N²-ε-G:A pair,

extension was faster (Fig. 5*A* and Fig. S6*A*, lane 7) compared with unmodified templates (Fig. 5*A* and Fig. S6*A*, lane 2).

Single-nucleotide insertion assays were done utilizing individual dNTPs to analyze postlesion insertions after the correct base pair (1,N²-ε-G:C) as well as mispairing (1,N²-ε-G:A).

For correct base pairing, the incorporation preference for control template 1 was dTTP > dGTP > dCTP > dATP (Fig. 4*B*, lanes 1–12), and the incorporation preference for the control template 2 was dATP > dTTP >dGTP > dCTP (Fig. S5*B* lanes 1–12). In the case of the 1,N²-ε-G:C template–primer pairs, postlesion insertion was drastically retarded, indicating resistance to extension past the correct base pair C. Overall, hpol η-mediated formation of very small amounts of inserted products for both sequence contexts was observed, at similar rates. The postlesion insertion preference for 5′-T(εG)A-3′ (**1**) template was dTTP > dGTP > dATP > dCTP (Fig. 4*C*, lanes 1–12) while for the 5′-T(εG)G-3′ (**2**) template, the preference was dGTP > dATP ≅ dTTP ≅ dCTP (Fig. S5*C*, lanes 1–12).

For a mispair (G:A), postlesion insertion was drastically affected for the control template 1 (Fig. 5*B*, lanes 1–12), while the incorporation preference for the control template 2 was dTTP > dATP ≅ dGTP > dCTP (Fig. S6*B*, lanes 1–12). In the case of the 1,N²-ε-G:A template–primer pairs, the postlesion insertions were unaffected (Fig. 5*C*, Fig. S6*C*). The order of
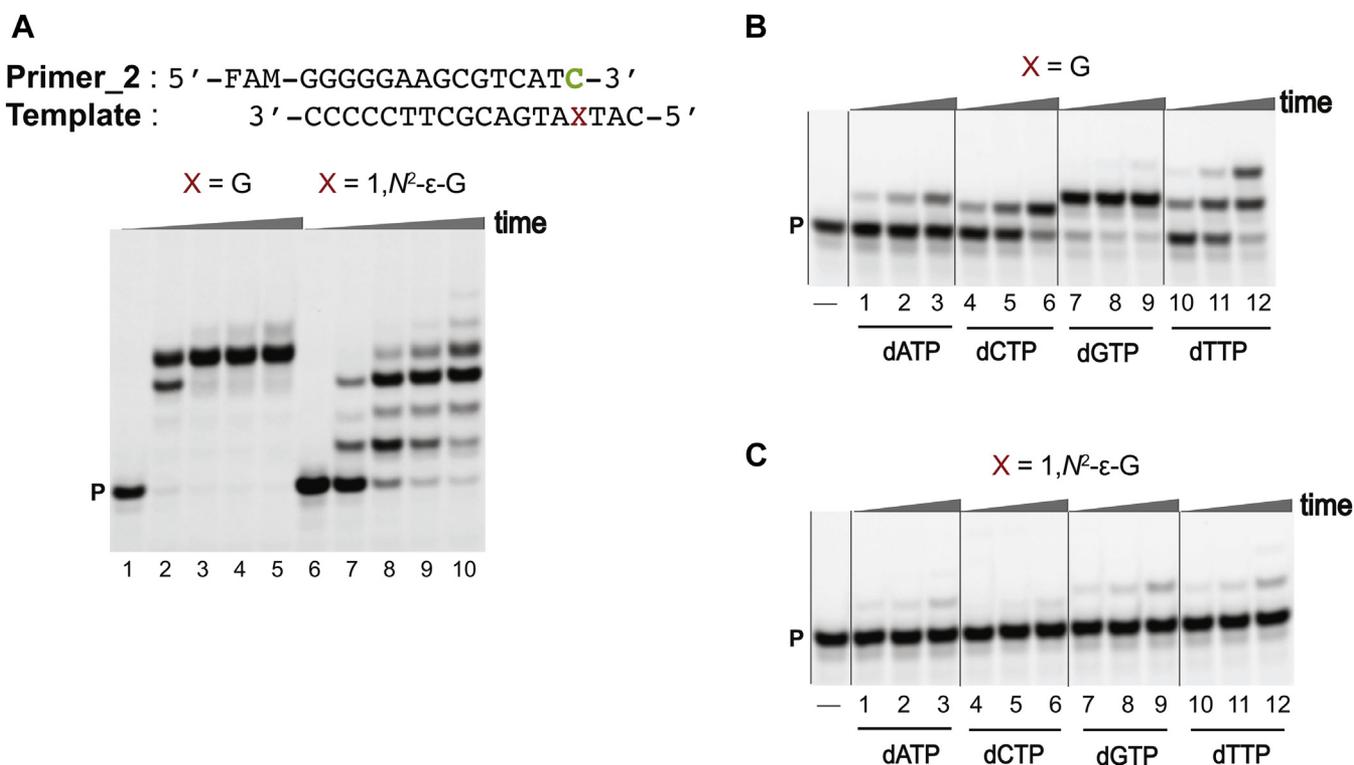
A

Primer_2 : 5′–FAM–GGGGGAAGCGTCATC–3′
Template :   3′–CCCCCTTCGCAGTAXTAC–5′

B

X = G

C

X = 1,N²-ε-G

Figure 4. hpol η-mediated postlesion full-length and single-nucleotide insertion assays using 1,N²-ε-G in template 5′-T(εG)A-3′ (1) and Primer_2. PAGE (20%, 7 M urea): *A*, full-length extension assay: hpol η (120 nM) elongated Primer_2 opposite G, and 1,N²-ε-G-containing DNA templates in the presence of a mixture of dNTPs (500 μM). All reactions were done at 37 °C for 5-, 30-, 60-, and 120-min (time gradients indicated with wedges). Lanes: 1 to 5 for unmodified template; 6 to 10 for 1,N²-ε-G-modified template. Single-nucleotide insertion assays: hpol η (10 nM) was incubated with *B*, Primer_2/control template 1, and *C*, Primer_2/1,N²-ε-G modified template 1 (5′-T(εG)A-3′), as well as individual dNTPs (100 μM). Lanes: 1 to 3 for dATP, 4 to 6 for dCTP, 7 to 9 for dGTP, 10 to 12 for dTTP. All reactions were done at 37 °C for 5-, 10-, and 30-min. P indicates the FAM-labeled Primer_2. See Experimental procedures and Table S1 for the oligonucleotide sequences used.

postlesion insertion for both the sequence contexts was dTTP > dGTP > dATP > dCTP (Fig. 5*C*, Fig. S6*C*, lanes 1–12). The preference for T insertion during hpol η-mediated postlesion bypass was also observed previously for unmodified (2′-F-G:C pair) as well as modified (N⁷-CH₃ 2′-F-G:C pair) templates (35).

Overall, hpol η was unable to perform postlesion extension for the correct pair (1,N²-ε-G:C). hpol η was able to perform postlesion extension for a mispair (1,N²-ε-G:A), indicating that the mispair was preferentially extended past the lesion.

LC-ESI-MS/MS analysis for hpol η-mediated primer extension across from 1,N²-ε-G

To elucidate details about other possible miscoding events, hpol η-mediated full-length extension reactions were performed for both the sequence contexts, 5′-T(εG)A-3′ (**1**) and 5′- T(εG)G-3′ (**2**) for LC-MS/MS analysis (see Table S1 for oligonucleotide sequences). The 2′-deoxyuridine (dU)-containing primers (see Table S1 for oligonucleotide sequences) were used for the extension reaction. Next, the fully extended reaction mixtures were treated with UDG and piperidine as described earlier (31, 36). Reaction mixtures were also analyzed by gel electrophoresis to confirm the elongations (Fig. S7). The relative yields of fully extended products were calculated by LC-MS/MS analysis and are summarized in

Table 1 and Table S6. CID spectra of the products are shown in Figs. S8–S19 and Tables S3–S10.

Replication of the control templates gave only error-free products, with blunt end addition of A and G (Table 1 and Table S6).

In case of the 5′-T(εG)A-3′ (**1**) template, hpol η replicated past the adduct in an error-prone manner only, resulting in three main products (Table 1). The first two products showed the same *m/z* (*i.e.*, 836.64, −3 ion) and coeluted together at same retention time (*t*R 4.34 min, Figs. S8 and S10). In these two products, the first product corresponded to misincorporation of A opposite the adduct, followed by misinsertion of G with blunt end addition of A (m/z 836.64, −3 ion: 5′-pCAT**AG**TGA-3′; 15%; Fig. 6 and Table S3), and the second product corresponded to misincorporation of G with blunt end addition of A (m/z 836.64, −3 ion: 5′-pCAT-**G**ATGA-3′; 38%; Fig. S9 and Table S4). The last product corresponded to misincorporation of G opposite the adduct, followed by misinsertion of G and with blunt end addition of A (m/z 842.00, −3 ion (*t*R 4.34 min): 5′-pCAT**GG**TGA-3′; 47%; Fig. S11 and Table S5). The insertion of G (opposite 1,N⁶-ε-A) followed by misinsertion of another G has been observed in previous studies (37).

For the 5′-T(εG)G-3′ (**2**) template, hpol η-mediated error-prone bypass resulted in four main products (Table S6). The first two products showed the same *m/z* (836.64, −3 ion) and

A

Primer_3 : 5'-FAM-GGGGGAAGCGTCAT**A**-3'
Template :   3'-CCCCCTTCGCAGTA**X**TAC-5'

B

X = G

C
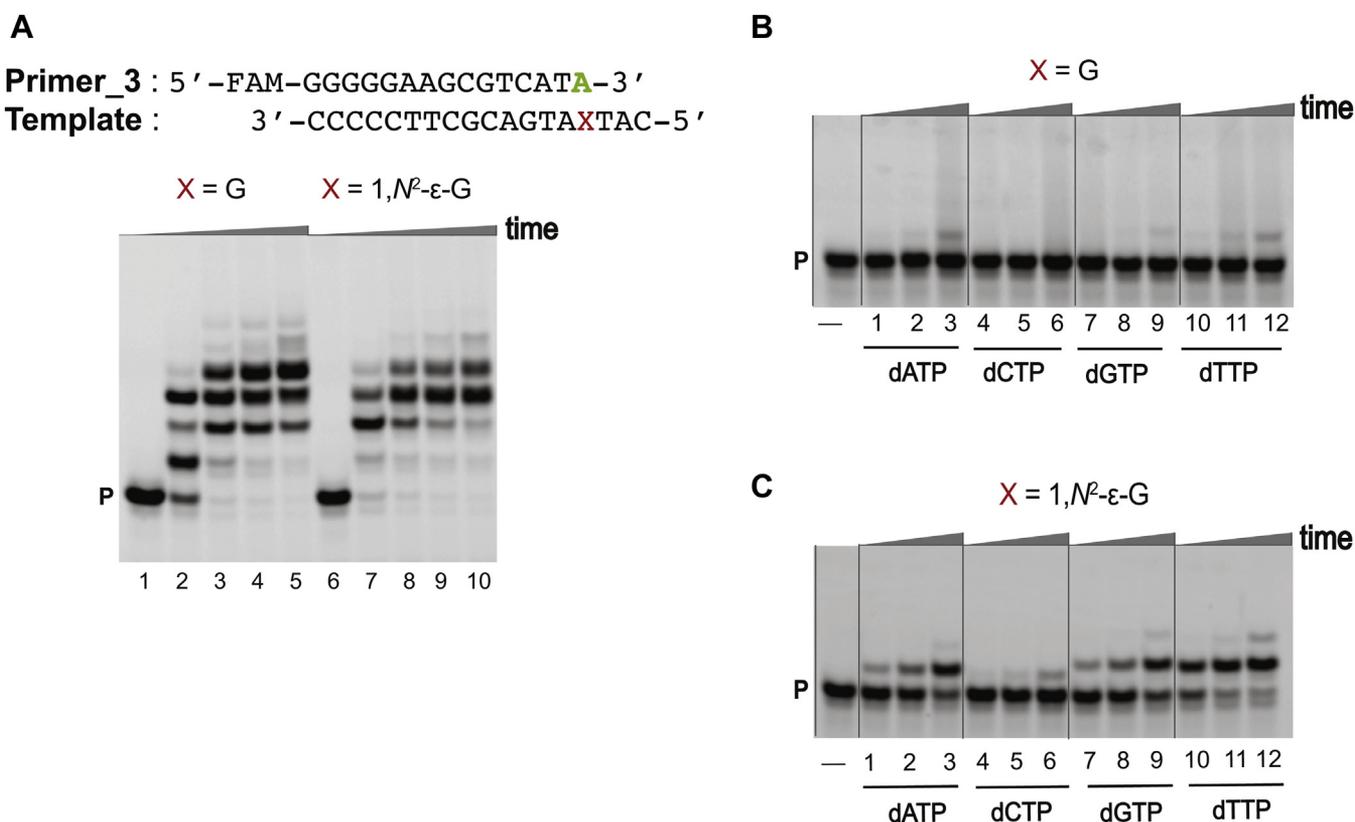
X = 1,N²-ε-G

Figure 5. hpol η-mediated postlesion full-length and single-nucleotide insertion assays using 1,*N²*-ε-G in template 5'-T(εG)A-3' (1) and Primer_3. PAGE (20%, 7 M urea). *A*, full-length extension assay: hpol η (120 nM) elongated Primer_3 opposite G, and 1,*N²*-ε-G-containing DNA templates in the presence of a mixture of dNTPs (500 μM). All reactions were done at 37 °C for 5-, 30-, 60-, and 120-min (time gradients indicated with wedges). Lanes: 1 to 5, unmodified template; 6 to 10, 1,*N²*-ε-G-modified template. Single-nucleotide insertion assays: hpol η (10 nM) was incubated with *B*, Primer_2/control template 1, and *C*, Primer_3/1,*N²*-ε-G modified template 1 (5'-T(εG)A-3'), as well as individual dNTPs (100 μM). Lanes: 1 to 3 for dATP, 4 to 6 for dCTP, 7 to 9 for dGTP, 10 to 12 for dTTP. All reactions were done at 37 °C for 5-, 10-, and 30-min. P indicates the FAM-labeled Primer_3. See Experimental procedures and Table S1 for the oligonucleotide sequences used.

coeluted at the same retention time (t_R 4.33 min, Figs. S12 and S15). Interestingly, the next two products also coeluted at same t_R and *m/z* (842.00, −3 ion, t_R 4.31 min, Figs. S16 and S19). Among all the products, the first corresponded to misincorporation of A opposite to the adduct with blunt end addition of G (m/z 836.64, −3 ion: 5'-pTC**A**ATGAG-3'; 20%; Fig. S13 and Table S7). The second product corresponded to misincorporation of G with blunt end addition of A (m/z 836.64, −3 ion: 5'-pTC**G**ATGAA-3'; 25%; Fig. S14 and Table S8). The third product corresponded to misincorporation of A followed by misinsertion of G, with misinsertion of G opposite the 5' T and blunt end addition of A (m/z 842.00, −3 ion: 5'-pTC**A**GTG**G**A-3'; 26%; Fig. S17 and Table S9). The last product corresponded to misincorporation of G opposite the

adduct, misinsertion of G opposite the 5' T, with blunt end addition of A (m/z 842.00, −3 ion: 5'-pTC**G**ATG**G**A-3'; 29%; Fig. S18 and Table S10).

No products were observed containing the incorporation of correct base C (or its stalled products), indicating that correct base C can be inserted opposite but that hpol η is not able to extend the primer to obtain full-length products, as reflected in postlesion extension assays.

Crystal structures of ternary hpol η insertion stage complexes with 1,N²-ε-G opposite dAMPnPP and dCMPnPP

We determined two crystal structures of hpol η·DNA (oligonucleotide)·dNTP ternary complexes trapped in the

Table 1
Summary of products of extension of 1,*N²*-ε-G-modified template 5'-T(εG)A-3' (1) and Primer_4 by hpol η analyzed by LC-ESI-MS/MS

	Primer_4: 5'-FAM-GGGGGAAGCG**U**CAT-3' Template: 3'-CCCCCTTCGCAGTA**X**TAC-5'			
X	**Sequence**	**Yield**	**Observed *m/z* (charge)**	**Base added**
G	5'-pCAT**C**ATGA-3'	70%	823.64 (−3)	C, plus, blunt end addition of A, and G
	5'-pCAT**C**ATGG-3'	30%	828.64 (−3)	
1,*N²*-ε-G	5'-pCAT**A**GTGA-3'	15%	836.64 (−3)	A, followed by misinsertion of G, plus, blunt end addition of A
	5'-pCAT**G**ATGA-3'	38%	836.64 (−3)	G, plus, blunt end addition of A
	5'-pCAT**G**GTGA-3'	47%	842.00 (−3)	G, followed by misinsertion of G plus, blunt end addition of A
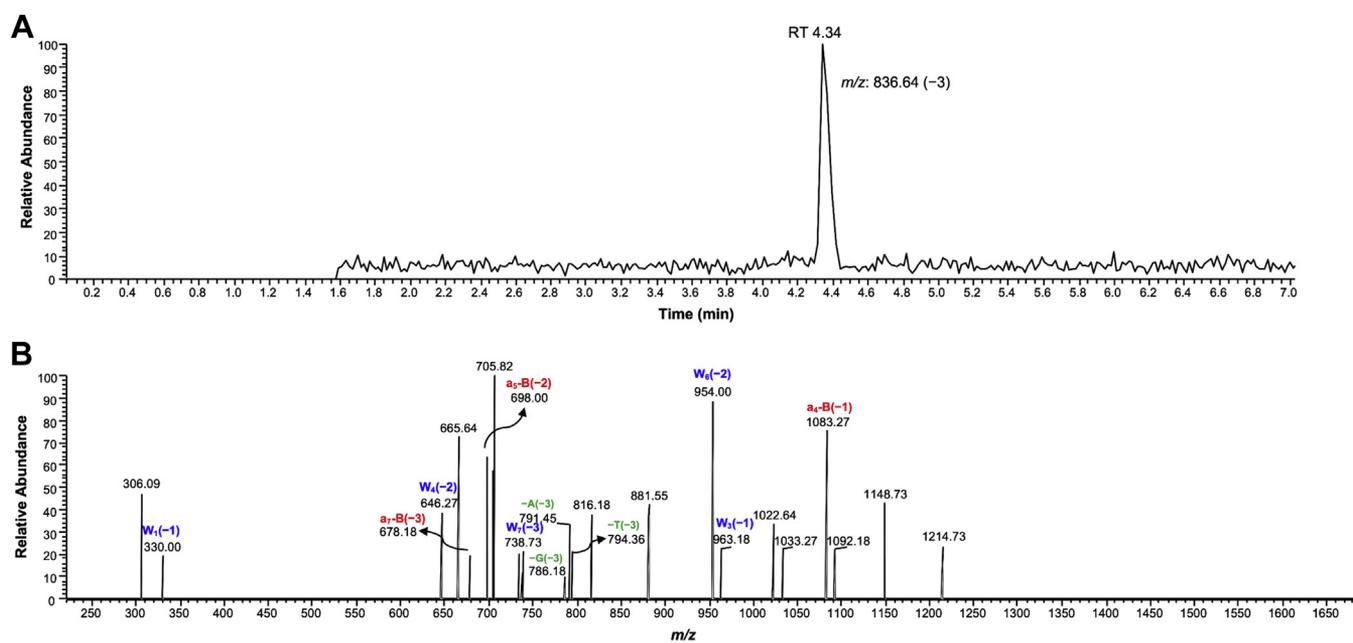
Figure 6. Extracted ion chromatogram and CID spectrum of *m/z* 836.64 ion. *A*, chromatogram; *B*, mass spectrum. The *m/z* 836.64 ion (−3, t_R 4.34) is associated with the extended product sequence 5′-pCAT**AG**TGA-3′ for 1,N²-ε-G template 5′-T(εG)A-3′ (**1**)–Primer_4 complex. a-B fragments represented in *red*, W fragments in *blue*, and base losses in *green*.

insertion state. Both complexes feature the same 8mer oligonucleotide primer and a 12mer oligonucleotide template with the 1,N²-ε-G adduct flanked by 5′-T and 3′-A (Table S2). The structure of the complex with incoming dAMPnPP was refined to 2.59 Å resolution, and the complex with incoming dCMPnPP was refined to 2.23 Å resolution. Selected crystal data, data collection, and refinement parameters are listed in Table 2. Examples of the quality of the electron density in the final models are depicted in Figs. S20 and S21. Along with the polymerase, oligonucleotide duplex, and nucleotide triphosphates, the dAMPnPP and dCMPnPP structures revealed 163 and 273 water molecules, respectively, and two active site Mg²⁺ ions. dAMPnPP and dCMPnPP are nonhydrolyzable nucleotide triphosphates that allow cocrystallization of pols with oligonucleotide template–primer constructs in the presence of Mg²⁺ without the use of a primer with a 3′-terminal 2′,3′-dideoxynucleoside and/or divalent metal ions that are not catalytically competent.

At the active sites of both complexes, 1,N²-ε-G adapts a *syn* orientation with the lesioned Watson–Crick edge of G jutting into the major groove (Figs. 7 and 8). The nucleobase of the incoming dAMPnPP is staggered relative to the 1,N²-ε-G base plane and sits halfway between the adduct and the 5′-adjacent thymine (Fig. 7A). This structure positions the N6 atom of adenine 2.9 Å from O6 of the adduct and 2.5 Å from O4 of thymine, consistent with the formation of two H-bonds. The projection, approximately perpendicular to the 1,N²-ε-G and adenine planes, shows that there is no stacking between the incoming nucleotide and the adduct (Fig. 7B). The 3′-hydroxyl group of the primer 3′-terminal dT is poised for attack at the α-phosphate of dAMPnPP with an O3′···Pα distance of 3.7 Å. Among hpol η side chains that project into the active site, Arg-61 plays a key role in that it is often engaged in multiple

interactions with nucleobase atoms and/or phosphate groups of the incoming nucleotide (38, 39). In the case of the ternary complex with dAMPnPP, the guanidino moiety of arginine does not stack on adenine or engage in contacts with the Hoogsteen edge of the base, as Arg-61 is shifted slightly into the major groove vis-à-vis the incoming nucleotide (Fig. 7B).

Table 2

Crystal data, data collection parameters, and structure refinement statistics[a]

Complex	1,N²-ε-G:dAMPnPP	1,N²-ε-G:dCMPnPP
Data Collection		
Wavelength [Å]	0.97856	0.97856
Space group	$P6_1$	$P6_1$
Resolution [Å]	50.0–2.59 (2.63–2.59)[a]	42.9–2.23 (2.29–2.23)[a]
Unit cell *a, b, c* [Å]	99.11, 99.11, 81.86	99.08, 99.08, 81.89
Unique reflections	14,272 (713)	22,384 (1659)
Completeness [%]	99.7 (100)	100 (100)
I/σ(I)	12.0 (1.8)	16.6 (2.9)
Wilson B-factor [Å²]	22.1	26.9
R-merge	0.152 (0.910)	0.093 (0.625)
Redundancy	5.7 (5.7)	5.7 (5.7)
Refinement		
R-work	0.191 (0.235)	0.158 (0.193)
R-free	0.248 (0.301)	0.210 (0.280)
Number of atoms		
Protein/DNA	3426/391	3394/391
dNTP/Water/Mg²⁺	30/163/2	28/273/2
Protein residues	431	430
B-factor [Å²]		
Average	38.6	33.5
Protein/DNA	37.8/45.9	32.6/37.5
dNTP/M²⁺/Water	41.3/33.4/38.7	34.4/26.5/38.5
R.m.s. deviations		
Bonds [Å]	0.003	0.008
Angles [deg.]	0.6	1.0
Ramachandran		
Favored (%)	96.8	97.0
Allowed (%)	3.0	2.8
Outliers (%)	0.2	0.2
PDB ID Code	5F9I	5F9N

[a] Statistics for the highest-resolution shell are shown in parentheses.

However, Arg-61 forms a somewhat long (3.4 Å) salt bridge with the α-phosphate and a water-mediated interaction with the γ-phosphate. Another amino acid side chain, Gln-38, that is positioned near the minor groove and occasionally engages in direct or water-mediated contacts with template base and/or incoming nucleotide (38, 39), is too far removed from both in the active site of this complex and does not form any with either.

Unlike dAMPnPP, the base moiety of the incoming dCMPnPP is coplanar with the base plane of $1,N^2$-ε-G (Fig. 8A). A single H-bond is formed between the N4 atom of cytosine and the O6 atom of the adduct. The O4 atom of the T 5′-adjacent to $1,N^2$-ε-G is too far removed from N4 of the incoming cytosine to engage in an H-bond interaction (3.5 Å). The 3′-hydroxyl group of the primer 3′-terminal T sits at a distance of 3.5 Å from the α-phosphate of dCMPNPP. As in the case of the complex with dAMPnPP, Arg-61 is shifted slightly into the major groove vis-à-vis the incoming cytosine (Fig. 8B), which precludes a stacking interaction. The guanidino moiety of Arg-61 is too far removed from the α-phosphate (4.1 Å) to engage in an effective interaction but forms a water-mediated contact with the γ-phosphate. In the minor groove, a water molecule bridges the Gln-38 side chain and O2 atom of the incoming dCMPnPP.

Discussion

The Y-Family hpol η plays a key role in translesion synthesis in human cells and features an active site composition that can accommodate bulky adducts (24, 40). Among the best-known base lesions that hpol η can cope with are those that arise as a consequence of UV damage, *e.g.*, cyclic pyrimidine dimers, as well as cisplatin intrastrand cross-links (41). However, hpol η can also synthesize past a variety of other adducts with various proficiencies in a more or less error-prone fashion. Thus, hpol η can bypass the major oxidative lesion 8-oxoG efficiently and virtually error-free (42). Additional lesions that this polymerase can overcome include abasic sites (33), exocyclic etheno A

and G adducts (32, 34), O^6-methyl G (39), and O^4-alkyl-T (43) as well as, in a remarkably efficient and mostly error-free mode, the methyl FAPY-G lesion (44), and a bulky aminobenzanthrone DNA adduct (45).

The insertion stage assays with both sequences, 5′-T(εG)A-3′ (**1**) and 5′-T(εG)G-3′ (**2**), demonstrate that hpol η preferentially inserts dGTP and dATP relative to the pyrimidine nucleoside triphosphates. In the case of sequence (**1**), dATP and dTTP were incorporated to a similar degree, but the correct dCTP nucleotide was the least preferred opposite $1,N^2$-ε-G in both sequence contexts (Fig. 3 and Fig. S4). The two crystal structures of hpol η ternary complexes reveal clear differences between the orientation of incoming dAMPnPP and dCMPnPP. The nucleobase of the former assumes a staggered orientation relative to the $1,N^2$-ε-G base plane and is thus positioned halfway between the adduct and template dT in the spacious active site of hpol η (Fig. 7). This allows formation of a bifurcated H-bond between incoming nucleotide (N6H₂) and template dT (O4) and $1,N^2$-ε-G (O6). The overall preferred dGTP is likely to adopt a similar orientation and may use its N1H atom to establish a bifurcated H-bond to the O4 and O6 keto oxygens of dT and the adduct, respectively. Unlike an incoming dATP, dGTP features the exocyclic amino function in the minor groove that may also allow a direct or water-bridged H-bond interaction with the side chain of Gln-38. In the structure of the complex with incoming dCMPnPP, cytosine is virtually coplanar with the $1,N^2$-ε-G base plane, resulting in a single H-bond between N4H₂ and the O6 keto oxygen of the adduct, but precluding an H-bond interaction with the 5′-adjacent template dT (Fig. 8). dTTP is preferentially inserted relative to dCTP in both sequence contexts and —assuming similar orientations of thymine and cytosine vis-à-vis $1,N^2$-ε-G—will likely establish an H-bond to the N7 atom of the adduct in the *syn* conformation *via* its N3 hydrogen donor function. As with an incoming dCMPnPP in the crystal structure, this pose would position the thymine O2 atom in the minor groove in the vicinity of the Gln-38 side chain so that
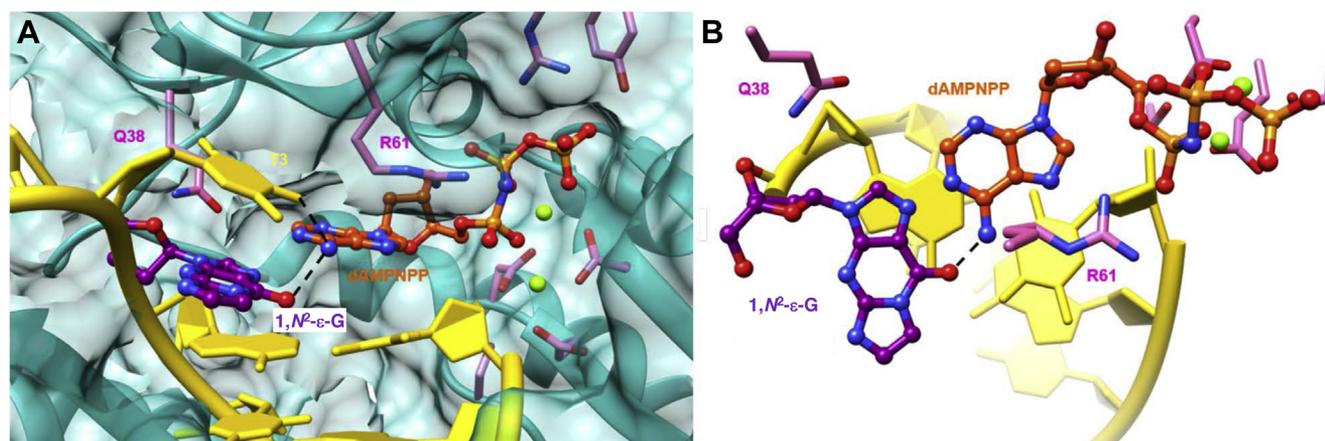
Figure 7. Active site conformation in the ternary hpol η insertion step complex with dAMPnPP opposite $1,N^2$-ε−G in the 5′-T(εG)A-3′ template sequence context (oligonucleotide 1). *A,* view into the DNA major groove. *B,* rotated by 90° and viewed perpendicular to the adenine and adduct base planes. Selected active site residues are colored by atom with carbon atoms shown in *purple* ($1,N^2$-ε−G), *orange* (incoming dAMPnPP), or *pink* (Arg-61 and Gln-38 from the finger domain and Asp/Glu coordinating Mg²⁺ ions that are shown as *light green spheres*). The remaining template and primer residues are colored in *yellow* and H-bonds involving the incoming nucleotide are drawn with *dashed lines*.
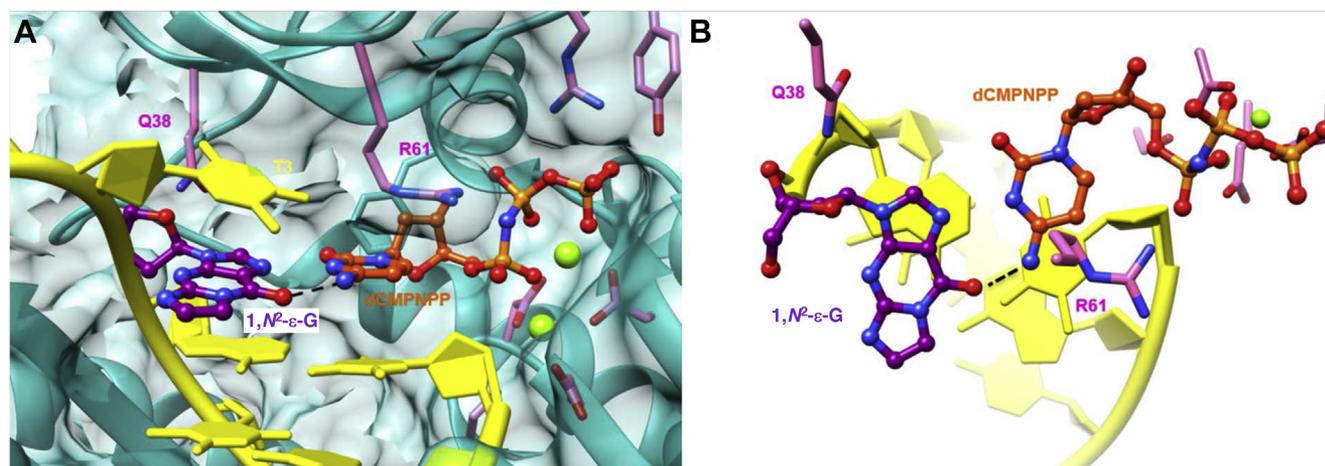
Figure 8. Active site conformation in the ternary hpol η insertion step complex with dCMPnPP opposite 1,N^2-ε–G in the 5′-T(εG)A-3′ template sequence context (oligonucleotide 1). *A*, view into the DNA major groove. *B*, rotated by 90° and viewed perpendicular to the cytosine and adduct base planes. Selected active site residues are colored by atom with carbon atoms shown in *purple* (1,N^2-ε–G), *orange* (incoming dCMPnPP), or *pink* (Arg-61 and Gln-38 from the finger domain and Asp/Glu coordinating Mg^{2+} ions that are shown as *light green spheres*). The remaining template and primer residues are colored in *yellow* and H-bonds involving the incoming nucleotide are drawn with *dashed lines*. It is unlikely that N1 of 1,N^2-ε–G and N3 of dCMPnPP are H-bonded because the pH of the crystallization solution is too high for cytosine to be protonated at N3 and a tautomeric form of the adduct with the hydrogen on N1 cannot be invoked.

the two could be linked by a bridging water molecule. Alternatively, it is possible that thymine shifts slightly toward the major groove, thus placing the N3 hydrogen donor roughly between the N7 and O6 acceptors of the adduct and potentially resulting in the formation of a bifurcated H-bond. The preferred insertion of dGTP and dATP by hpol η opposite 1,N^2-ε-G is most likely a consequence of their ability to reach two template nucleotides, as stacking differences between incoming purine and pyrimidine nucleoside triphosphates do not appear to play a significant role in the observed discrimination. However, the staggered orientation of dAMPnPP (and presumably dGTP) does not result in frameshifting, and once inserted, the purine may settle opposite the adduct in a more or less coplanar fashion, thereby allowing hpol η to bypass 1,N^2-ε-G without skipping a base.

Extension reactions were tested with primers that featured either A or C opposite the adduct. Not only was insertion of dCTP opposite the adduct disfavored relative to dGTP and dATP, but extension from the correct base opposite 1,N^2-ε-G was severely hampered as well (Fig. 4*A* and Fig. S5*A*, lane 7). hpol η was able to extend from a G:A pair with one of the control templates and then inserted dTTP > dATP ≅ dGTP > dCTP, *i.e.*, the extension proceeded in an error-prone fashion. Conversely, extensions from 1,N^2-ε-G:A proceeded more efficiently than the control reactions from G:A, and irrespective of the template sequence, hpol η inserted dTTP > dGTP > dATP > dCTP opposite T (Fig. 5 and Fig. S6). In a previous study from our laboratories (35), we observed that dTTP misincorporation was sixfold more efficient during hpol η-mediated postlesion bypass for the N^7-CH_3 2′-F-G lesion compared with 2′-F-G unmodified template, when paired with the correct base C. In a similar way, hpol η catalyzed TLS past the 1,N^2-ε-G adduct in an efficient but highly error-prone manner. The results of the full-length extension reactions analyzed by LC-MS/MS attest to the absence of frameshift

products but are supportive of untemplated addition of A and G at the end of the primer. The polymerase preferentially incorporated dGTP opposite the adduct and in the first extension step (85% and 63%, respectively, Table 1).

In a previous study from our laboratories (32), we also reported that hpol η preferred to insert G and A opposite 1,N^2-ε-G, with some C insertion. LC-MS extension analysis, with what is referred to here as oligonucleotide **2**, yielded only the product with G incorporated opposite 1,N^2-ε-G (and no frameshifts), in contrast to the results shown in Table S6. The only major differences in the earlier paper (32) were that the (longer) hpol η was expressed in a baculovirus-based system (instead of *E. coli*) and that a less sensitive mass spectrometer was used (Thermo DecaXP). In the earlier paper, a slippage mechanism was included that could have accounted for G insertion in the product (32). In reviewing the results of our present study, it is clear that both G and A were present in the extended primer product (Table 1 and Table S6, Fig. 6, Figs. S9, S11, S13, S14, S17 and S18), as evidenced by the presence of both in the products generated by hpol η from distinct oligonucleotides. A slippage mechanism is not very likely to explain the incorporation of A.

A result observed in the present work with oligonucleotide **2** is the insertion of A followed by misinsertion of G, apparently opposite the T immediately 5′ to the 1,N^2-ε-G adduct (Table S6). A slippage mechanism would require insertion of A followed by a 2-base loop-out to the 5′ penultimate C and then realignment to complete the polymerization. Another anomaly was the apparent misincorporation of G opposite the 5′ T in about one-half of the product (but not with the unmodified oligonucleotide) (Table S6).

Presently, there are no crystal structures of ternary complexes with hpol η trapped at the extension step after 1,N^2-ε-G bypass. The insertion stage complexes with either incoming dAMPnPP (Fig. 7) or dCMPnPP (Fig. 8) both reveal the adduct

in the *syn* orientation. The absence of frameshifting in the full-length extension reactions supports the notion that the adduct maintains the *syn* orientation following TLS by hpol η. A coplanar arrangement of either G or A paired to $1,N^2$-ε-G at the −1 position can be expected to result in a locally distorted duplex geometry in order to accommodate the adducted and inserted purine moieties. Although extensions can proceed, they do so in an error-prone way. What is clear, however, is that an arrangement with the adduct in the *anti* orientation and either incoming dGTP or dATP stacked on top of $1,N^2$-ε-G during bypass is neither supported by the structural data nor by the sequences identified from the extension assays.

Our observations of hpol η-catalyzed TLS past the $1,N^2$-ε-G lesion reveal similarities in terms of the preferential insertion of purine nucleoside triphosphates by this pol opposite an array of other lesions. However, they also demonstrate contrasting underlying bypass mechanisms as well as the distinct means of hpol η to accommodate similar adducts, *e.g.*, the exocyclic $1,N^2$-ε-G and $1,N^6$-ε-A, at its active site. For example, with abasic sites the purine rule applies—hpol η rarely inserts T and C (33)—and A and G are also preferred for insertion over T opposite $1,N^6$-ε-A by hpol η (34). However, crystal structures of ternary complexes showed that $1,N^6$-ε-A is in the *anti* conformation when the preferred A and G are inserted. Also, distinct from the bypass of $1,N^2$-ε-G, the bypass of $1,N^6$-ε-A by hpol η involves frameshifts. Another distinct feature of the bypass reactions opposite the two exocyclic etheno adducts is that structural data showed that $1,N^6$-ε-A is in the *syn* conformation opposite dTTP (34). Conversely, we demonstrated here that $1,N^2$-ε-G is in the *syn* orientation both with incoming dAMPnPP and dCMPnPP. The origins of the different behaviors of the two adducts in terms of adopting either the *syn* or *anti* orientation at the active site of hpol η are currently not clear.

Along with shared and distinct features of the bypass of similar adducts by hpol η, it is also worth noting in the context of $1,N^2$-ε-G that *S. solfataricus* Dpo4 and hpol η, two Y-family pols that have much in common in terms of both activity and structure, behave very differently vis-à-vis this adduct. Thus, both pols have active sites with high ceilings that allow accommodation of two template bases and efficient bypass of CPDs and cisplatin intrastrand cross-links. However, bypass reactions opposite $1,N^2$-ε-G by Dpo4 involve the adduct in the *anti* orientation with a stacked incoming nucleotide that result in −1 frameshifts (31). These differences contrast sharply with our previous finding that both Dpo4 and hpol η are capable of mostly correct bypass of 8-oxoG (42, 46). In the case of Dpo4, the preference of 8-oxoG in the *anti* conformation opposite incoming dCTP *versus* in the *syn* conformation opposite incoming dATP is about 19:1. For hpol η the discrimination is even larger, with the percentage of insertion of dATP (and therefore 8-oxoG in *syn*) only between 1 and 3%. It is therefore remarkable that the bypass of $1,N^2$-ε-G is so different for the two pols, *i.e.*, in Dpo4, the adduct is always in the *anti* conformation but in hpol η it seems to be always in the *syn* conformation as per the two new structures.

Experimental procedures

hPol η catalytic core protein expression and purification

The hpol η plasmid (pET28a) comprising residues 1 to 432 was a generous gift from Dr. Wei Yang, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health. The polymerase was expressed in *E. coli* and purified as described previously (47), and the protein solution was concentrated to ∼5 mg/ml.

Materials

The $1,N^2$-ε-G-modified phosphoramidite (Fig. S1) was synthesized using a previously reported procedure (48). Unlabeled dNTPs and uracil DNA glycosylase (UDG) were purchased from New England Biolabs. C_{18} Sep-Pak columns were purchased from Waters. Piperidine was from Sigma-Aldrich. Unmodified oligonucleotides and FAM-labeled oligonucleotide primers were purchased from Integrated DNA Technologies. dAMPnPP and dCMPnPP were obtained from Jena Bioscience.

1,N²-ε-G-oligonucleotide synthesis, purification, and characterization

Solid-phase synthesis of the $1,N^2$-ε-G-modified oligonucleotides was done on a PerSeptive Biosystems Model 8909 DNA synthesizer (see Table S1 for in-house synthesis of modified oligonucleotide sequences) (Figs. S2 and S3). The modified DNAs were synthesized on a 1-μmol scale using the appropriate controlled pore glass (CPG) as solid support. Manual coupling for the modified phosphoramidite was carried out for 30 min. The modified oligonucleotides were deprotected using a reported procedure (49). The modified oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis (20% PAGE, 7 M urea) at 55 W for 2.5 h using 1X TBE buffer (89 mM Tris-HCl buffer (pH 8.0) containing 89 mM boric acid and 2 mM EDTA). The gel was visualized under a UV lamp at 260 nm, and the desired oligonucleotide bands were isolated from the gel and extracted using TEN buffer (10 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA and 300 mM NaCl) overnight at room temperature. The modified oligonucleotides were desalted using C_{18} Sep-Pak columns, and their integrity was confirmed by MALDI-TOF mass spectrometry (positive linear mode, Figs. S2 and S3).

Full-length extension assays

A primer–template complex consisting of a FAM-labeled oligonucleotide primer and an unmodified or modified template was annealed (1:1 M ratio) at 95 °C for 5 min followed by slow cooling overnight (see Table S1 for oligonucleotide sequences) (Fig. 3A–5A and Fig. S4A–S6A). Full-length extension reactions were carried out using 50 mM Tris-HCl buffer (pH 7.5) containing 50 mM NaCl, 5 mM MgCl₂, 5% glycerol (v/v), 5 mM DTT, and 100 μg/ml BSA at 37 °C for 2 h. The final concentration of the primer–template complex was 150 nM, and 120 nM hpol η was used to obtain fully extended primers. Reactions were initiated by adding a 1.25 μl mixture of dNTPs to a total volume of 25 μl. Aliquots (3.5 μl) of

reaction mixtures were removed at times of 0, 5, 30, 60, and 120 min and quenched with 6.5 µl of 10 mM EDTA (pH 8.0) in 95% deionized formamide (v/v). Products were separated using 20% PAGE (7 M urea). Results were visualized on a Typhoon scanner (GE Healthcare) and analyzed with ImageJ software.

Single-nucleotide incorporation assays

Each primer–template oligonucleotide complex containing a FAM-labeled oligonucleotide primer and an unmodified or modified template was annealed (1:1 M ratio) at 95 °C for 5 min, followed by slow cooling overnight (see Table S1 for oligonucleotide sequences) (Figs. 3, *B* and *C*, 4, *B* and *C* and 5, *B* and *C*, Figs. S4, *B* and *C*, S5, *B* and *C*, and S6, *B* and *C*). All single-nucleotide insertion reactions were performed using 50 mM Tris-HCl buffer (pH 7.5) containing 50 mM NaCl, 5 mM $MgCl_2$, 5% glycerol (v/v), 5 mM DTT, and 100 µg/ml BSA, at 37 °C. The final concentration of primer–template complex was 150 nM and 10 nM hpol η was used, followed by the addition of 1.25 µl of an individual dNTP to a total volume of 25 µl. Aliquots (3.5 µl) of reaction mixtures were removed at 0, 5, 10, and 30 min and quenched with 6.5 µl of 10 mM EDTA (pH 8.0) in 95% deionized formamide (v/v). Products were separated using 20% PAGE (7 M urea). Results were visualized on a Typhoon scanner (GE Healthcare) and analyzed with ImageJ software.

Full-length extension reactions followed by UDG and piperidine treatment for LC-ESI-MS/MS analysis

A dU-containing FAM-labeled primer and an unmodified or modified template were annealed (1:1 M ratio) at 95 °C for 5 min followed by slow cooling overnight (see Table S1 for oligonucleotide sequences) (Fig. S7). The full-length extension reactions were carried out under similar reaction conditions as described in full-length extension assays, except that the final concentrations were as follows: the primer–template complex concentration was 2.5 µM and hpol η was 1.35 µM for the oligonucleotide 1,N^2-ε-G template–primer complex and 0.75 µM for the unmodified template–primer complex, in a total reaction volume of 85 µl. Reactions were started using a mixture of dNTPs (1 mM) at 37 °C for 4 h. The resulting products were desalted using Biospin columns. The fully extended products were treated with 25 U of UDG at 37 °C for 4 h, followed by 0.25 M piperidine at 95 °C for 1 h (31, 35), and the reaction mixture was evaporated to dryness by lyophilization. The dried pellet was resuspended in 500 µl of nuclease-free water and concentrated to dryness by lyophilization. The samples were dissolved in 30 µl of nuclease-free water for LC-ESI-MS/MS analysis.

LC-ESI-MS/MS analysis

LC-MS/MS analyses were performed on a Waters Acquity UPLC system connected directly to a Thermo-Finnigan LTQ mass spectrometer (electrospray ionization) (Fig. 6, Figs. S8–S19 and Tables S3–S10). The analysis was carried out in the negative ion mode using an octadecylsilane (C_{18}) column (Acquity UPLC BEH, 1.7 µm, 2.1 mm × 100 mm) as reported

previously (37). The extended product sequences were identified (Fig. 6, Figs. S8–S19, Table 1, Tables S3–S10) by comparing the theoretical *m/z* values using a Mongo Oligo Mass Calculator v2.06 with the observed CID fragments. For the coeluted products, relative peak areas were calculated on the basis of the intensity of representative fragments (as shown in Fig. S10 for a_5-B(−2) and Figs. S15 and S19 for a_4-B(−1) fragments). The relative yields of extended product sequences were calculated on the basis of relative peak areas of extracted ion chromatograms.

Crystallization of complexes

Primer and template sequences used in the crystallization experiments are shown in Table S2 (Figs. 7 and 8, Figs. S20 and S21). The DNA template–primer duplex was mixed with the protein in a 1.2:1 M ratio in the presence of excess 50 mM Tris-HCl buffer (pH 7.5) containing 450 mM KCl and 3 mM DTT. After adding 5 microl of 100 mM $MgCl_2$, the complex was concentrated to a final concentration of 2 to 3 mg/ml by ultrafiltration. Nonhydrolyzable nucleotide triphosphates were added last to form the ternary complexes. Crystallization experiments were performed by the hanging drop vapor diffusion technique at 18 °C using a sparse matrix screen (Hampton Research) (50). One microliter of the complex solution was mixed with 1 µl of reservoir solution and equilibrated against 500 µl reservoir wells. Crystals appeared in droplets containing 0.1 M MES buffer (pH 5.5) containing 5 mM $MgCl_2$ and 21 to 23% (w/v) PEG 2000 MME within 1 day and were harvested after 1 week. Crystals were mounted in nylon loops, cryoprotected in reservoir solution containing 22% glycerol (v/v), and frozen in liquid nitrogen.

X-ray diffraction data collection, structure determination, and refinement

Diffraction data were collected on the 21-ID-G beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source, Argonne National Laboratory. All data were integrated and scaled with the program HKL2000 (51). The structures were determined by the molecular replacement technique with the program MOLREP (52, 53), using the protein portion of structure of the hpol η complex with PDB ID 4O3N (42) as the search model. Structure refinement and model building were carried out with the programs Refmac (54) and COOT (55), respectively. Illustrations were prepared with the program UCSF Chimera (56). Final refinement parameters are summarized in Table 2. The quality of the model of the dAMPnPP complex structure is somewhat inferior compared with that of the dCMPnPP complex structure, in terms of both resolution and refinement parameters. However, we are confident that this structure too provides valid insights into the mechanism of hpol η-catalyzed bypass of the 1,N^2-ε-G lesion.

Data availability

With the exception of the Protein Data Bank coordinates (PDB 5F9I, 5F9N), all data is contained within the article and

Human pol η and 1,N²-ethenodeoxyguanosine

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: 1,N^2-ε-G, 1,N^2-ethenodeoxyguanosine or 1,N^2-ethenoguanine; CID, collision-induced dissociation; dAMPnPP, 2′-deoxyadenosine-5′-[(α,β)-imido]triphosphate; dCMPnPP, 2′-deoxycytosine-5′-[(α,β)-imido]triphosphate; dNMPnPP, 2′-deoxynucleoside-5′-[(α,β)-imido]triphosphate; EIC, extracted ion chromatogram; ESI, electrospray ionization; FAM, 6-carboxyfluorescein; h, human; pol, DNA polymerase; UDG, uracil DNA glycosylase; UPLC, ultraperformance liquid chromatography.

References

1. Barbin, A. (2000) Etheno-adduct-forming chemicals: From mutagenicity testing to tumor mutation spectra. *Mutat. Res.* **462**, 55–69
2. Jahnz-Wechmann, Z., Framski, G. R., Januszczyk, P. A., and Boryski, J. (2016) Base-modified nucleosides: Etheno derivatives. *Front. Chem.* **4**, 19
3. Singer, B., Bartsch, H., eds. (1999) *Exocyclic DNA Adducts in Mutagenesis and Carcinogenesis, No. 150.* International Agency for Research on Cancer Scientific Publications, Lyon, France
4. Guengerich, F. P., Kim, D.-H., and Iwasaki, M. (1991) Role of human cytochrome P-450 IIE1 in the oxidation of many low molecular weight cancer suspects. *Chem. Res. Toxicol.* **4**, 168–179
5. Guengerich, F. P., Persmark, M., and Humphreys, W. G. (1993) Formation of 1,N^2- and N^2,3-ethenoguanine derivatives from 2-halooxiranes: Isotopic labeling studies and formation of a hemiaminal derivative of N^2-(2-oxoethyl)guanine. *Chem. Res. Toxicol.* **6**, 635–648
6. Hecht, S. S., Young-Sciame, R., and Chung, F. L. (1992) Reaction of α-acetoxy-*N*-nitrosopiperidine with deoxyguanosine: Oxygen-dependent formation of 4-oxo-2-pentenal and a 1,N^2-ethenodeoxyguanosine adduct. *Chem. Res. Toxicol.* **5**, 706–712
7. Kronberg, L., Sjöholm, R., and Karlsson, S. (1992) Formation of 3,N^4-ethenocytidine, 1,N^6-ethenoadenosine, and 1,N^2-ethenoguanosine in reactions of mucochloric acid with nucleosides. *Chem. Res. Toxicol.* **5**, 852–855
8. Sodum, R. S., and Chung, F. L. (1988) 1,N^2-Ethenodeoxyguanosine as a potential marker for DNA adduct formation by *trans*-4-hydroxy-2-nonenal. *Cancer Res.* **48**, 320–323
9. Morinello, E. J., Ham, A. J. L., Ranasinghe, A., Sangaiah, R., and Swenberg, J. A. (2001) Simultaneous quantitation of N^2,3-ethenoguanine and 1,N^2-ethenoguanine with an immunoaffinity/gas chromatography/high resolution mass spectrometry assay. *Chem. Res. Toxicol.* **14**, 327–334
10. 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
11. Petrova, K. V., Jalluri, R. S., Kozekov, I. D., and Rizzo, C. J. (2007) Mechanism of 1,N^2-etheno-2′-deoxyguanosine formation from epoxyaldehydes. *Chem. Res. Toxicol.* **20**, 1685–1692
12. Linhart, K.-B., Glassen, K., Peccerella, T., Waldherr, R., Linhart, H., Bartsch, H., and Seitz, H. K. (2015) The generation of carcinogenic etheno-DNA adducts in the liver of patients with nonalcoholic fatty liver disease. *Hepatobiliary Surg. Nutr.* **4**, 117–123
13. Sattsangi, P. D., Leonard, N. J., and Frihart, C. R. (1977) 1,N^2-Ethenoguanine and N^2,3-ethenoguanine. Synthesis and comparison of the electronic spectral properties of these linear and angular triheterocycles related to the Y bases. *J. Org. Chem.* **42**, 3292–3296
14. Chen, H. J., and Chiu, W. L. (2005) Association between cigarette smoking and urinary excretion of 1,N^2-ethenoguanine measured by isotope dilution liquid chromatography-electrospray ionization/tandem mass spectrometry. *Chem. Res. Toxicol.* **18**, 1593–1599
15. Langouët, S., Müller, M., and Guengerich, F. P. (1997) Misincorporation of dNTPs opposite 1,N^2-ethenoguanine and 5,6,7,9-tetrahydro-7-hydroxy-9-oxoimidazo[1,2-*a*]purine in oligonucleotides by *Escherichia coli* polymerases I exo⁻ and II exo⁻, T7 polymerase exo⁻, human immunodeficiency virus-1 reverse transcriptase, and rat polymerase β. *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) Misincorporation of nucleotides opposite five-membered exocyclic ring guanine derivatives by *Escherichia coli* polymerases *in vitro* and *in vivo*: 1,N^2-ethenoguanine, 5,6,7,9-tetrahydro-9-oxoimidazo[1,2-a]purine, and 5,6,7,9-tetrahydro-7-hydroxy-9-oxoimidazo[1,2-a]purine. *Biochemistry* **37**, 5184–5193
17. 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
18. Saparbaev, M., Langouët, S., Privezentzev, C. V., Guengerich, F. P., Cai, H., Elder, R. H., and Laval, J. (2002) 1,N^2-Ethenoguanine, a mutagenic DNA adduct, is a primary substrate of *Escherichia coli* mismatch-specific uracil-DNA glycosylase and human alkyl-*N*-purine DNA glycosylase. *J. Biol. Chem.* **277**, 26987–26993
19. Bartsch, H., Nair, J., and Velic, I. (1997) Etheno-DNA base adducts as tools in human cancer aetiology and chemoprevention. *Eur. J. Cancer Prev.* **6**, 529–534
20. Nair, J., Barbin, A., Velic, I., and Bartsch, H. (1999) Etheno DNA-base adducts from endogenous reactive species. *Mutat. Res.* **424**, 59–69
21. Dechakhamphu, S., Yongvanit, P., Nair, J., Pinlaor, S., Sitthithaworn, S., and Bartsch, H. (2008) High excretion of etheno adducts in liver fluke-infected patients: Protection by praziquantel against DNA damage. *Cancer Epidemiol. Biomarkers Prev.* **17**, 1658–1664
22. VanderVeen, L. A., Druckova, A., Riggins, J. N., Sorrells, J. L., Guengerich, F. P., and Marnett, L. J. (2004) Differential DNA recognition and cleavage by EcoRI dependent on the dynamic equilibrium between two

ASBMB

forms of the malondialdehyde-deoxyguanosine adducts. *Biochemistry* **44**, 5024–5033

23. Guengerich, F. P. (2018) Exocyclic adducts. In: Wells, R. D., Bond, J. S., Klinman, J., Masters, B. S. S., Bell, E., Kaguni, L. S., eds. *Molecular Life Sciences*, Springer, New York, NY: 379–383

24. Vaisman, A., and Woodgate, R. (2017) Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* **52**, 274–303

25. Yang, W., and Wodgate, R. (2007) What a difference a decade makes: Insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15591–15598

26. Prakash, S., Johnson, R. E., and Prakash, L. (2005) Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem.* **74**, 317–353

27. Yang, I.-Y., Miller, H., Wang, Z., Frank, E. G., Ohmori, H., Hanaoka, F., and Moriya, M. (2003) Mammalian translesion DNA synthesis across an acrolein-derived deoxyguanosine adduct. Participation of DNA polymerase η in error-prone synthesis in human cells. *J. Biol. Chem.* **278**, 13989–13994

28. 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., *et al.* (2001) The Y-family of DNA polymerases. *Mol. Cell* **8**, 7–8

29. Boudsocq, F., Iwai, S., Hanaoka, F., and Woodgate, R. (2001) *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): An archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic pol eta. *Nucleic Acids Res.* **29**, 4607–4616

30. Eoff, R. L., Egli, M., and Guengerich, F. P. (2010) Impact of chemical adducts on translesion synthesis in replicative and bypass polymerases: From structure to kinetics. In: Geacintov, N. E., Broyde, S., eds. *The Chemical Biology of DNA Damage*, Wiley-VCH, Weinheim, Germany: 299–330

31. Zang, H., Goodenough, A. K., Choi, J.-Y., Irminia, 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 product with the adduct 1,*N*²-etheno-guanine. *J. Biol. Chem.* **280**, 29750–29764

32. 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

33. Patra, A., Zhang, Q., Lei, L., Su, Y., Egli, M., and Guengerich, F. P. (2015) Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem.* **290**, 8028–8038

34. Patra, A., Su, Y., Zhang, Q., Johnson, K. M., Guengerich, F. P., and Egli, M. (2016) Structural and kinetic analysis of miscoding opposite the DNA adduct 1,*N*⁶-ethenodeoxyadenosine by human translesion DNA polymerase η. *J. Biol. Chem.* **291**, 14134–14145

35. Njuma, O. J., Su, Y., and Guengerich, F. P. (2019) The abundant DNA adduct *N*⁷-methyl deoxyguanosine contributes to miscoding during replication by human DNA polymerase η. *J. Biol. Chem.* **294**, 10253–10265

36. Chowdhury, G., and Guengerich, F. P. (2011) Liquid chromatography-mass spectrometry analysis of DNA polymerase reaction products. *Curr. Protoc. Nucleic Acid Chem.* **47**, 7.16.1–7.16.11

37. Ghodke, P. P., and Guengerich, F. P. (2020) Impact of 1,*N*⁶-ethenoadenosine, a damaged ribonucleotide in DNA, on translesion synthesis and repair. *J. Biol. Chem.* **295**, 6092–6107

38. Su, Y., Patra, A., Harp, J., Egli, M., and Guengerich, F. P. (2015) Roles of residues Arg-61 and Gln-38 of human DNA polymerase η in bypass of deoxyguanosine and 7,8-dihydro-8-oxo-2'-deoxyguanosine. *J. Biol. Chem.* **290**, 15921–15933

39. Patra, A., Zhang, Q., Guengerich, F. P., and Egli, M. (2016) Mechanisms of insertion of dCTP and dTTP opposite the DNA lesion *O*⁶-methyl-2'-deoxyguanosine by human DNA polymerase η. *J. Biol. Chem.* **291**, 24304–24313

40. Yang, W. (2014) An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* **53**, 2793–2803

41. Cruet-Hennequart, S., Gallagher, K., Sokòl, A. M., Villalan, S., Prendergast, A. M., and Carty, M. P. (2009) DNA polymerase η, a key protein in translesion synthesis in human cells. *Subcell. Biochem.* **50**, 189–209

42. Patra, A., Nagy, L. D., Zhang, Q., Su, Y., Müller, L., Guengerich, F. P., and Egli, M. (2014) Kinetics, structure and mechanism of 8-oxo-7,8-dihydro-2'-deoxyguanosine bypass by human DNA polymerase η. *J. Biol. Chem.* **289**, 16867–16882

43. O'Flaherty, D. K., Patra, A., Su, Y., Guengerich, F. P., Egli, M., and Wilds, C. J. (2016) Lesion orientation of O4-alkylthymidine influences replication by human polymerase η. *Chem. Sci.* **7**, 4896–4904

44. Patra, A., Banerjee, S., Johnson Salyard, T., Malik, C., Christov, P., Rizzo, C. J., Stone, M. P., and Egli, M. (2015) Structural basis for error-free bypass of the 5-*N*-methylformamidopyrimidine-dG lesion by human DNA polymerase η and *Sulfolobus solfataricus* P2 polymerase IV (Dpo4). *J. Am. Chem. Soc.* **137**, 7011–7014

45. Patra, A., Politica, D. A., Chatterjee, A., Tokarsky, E. J., Suo, Z., Basu, A., Stone, M. P., and Egli, M. (2016) Mechanism of error-free bypass of the environmental carcinogen *N*-(2'-deoxyguanosin-8-yl)-3-aminobenzanthrone adduct by human DNA polymerase η. *ChemBioChem* **17**, 2033–2037

46. Zang, H., Irimia, A., Choi, J.-Y., Angel, K. C., Loukachevitch, L. V., Egli, M., and Guengerich, F. P. (2006) Efficient and high fidelity incorporation of dCTP opposite 7,8-dihydro-8-oxodeoxyguanosine by *Sulfolobus solfataricus* DNA polymerase Dpo4. *J. Biol. Chem.* **281**, 2358–2372

47. Biertümpfel, C., Zhao, Y., Kondo, Y., Ramon-Maiques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F., and Yang, W. (2010) Structure and mechanism of human DNA polymerase eta. *Nature* **465**, 1044–1048

48. Goodenough, A. K., Kozekov, I. D., Zang, H., Choi, J.-Y., Guengerich, F. P., Harris, T. M., and Rizzo, C. J. (2005) Site specific synthesis and polymerase bypass of oligonucleotides containing a 6-hydroxy-3,5,6,7-tetrahydro-9*H*-imidazo[1,2-*a*]purin-9-one base, an intermediate in the formation of 1,*N*²-etheno-2'-deoxyguanosine. *Chem. Res. Toxicol.* **18**, 1701–1714

49. Shanmugam, G., Kozekov, I. D., Guengerich, F. P., Rizzo, C. J., and Stone, M. P. (2007) Structure of the 1,*N*²-etheno-2'-deoxyguanosine adduct in duplex DNA at pH 8.6. *Chem. Res. Toxicol.* **20**, 1601–1611

50. Jancarik, J., and Kim, S.-H. (1991) Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409–411

51. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326

52. Vagin, A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025

53. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763

54. Vagin, A. A., Steiner, R. S., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2284–2295

55. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132

56. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612