Supporting Information

Schorr et al. 10.1073/pnas.1008894107

SI Text

Cloning, Protein Expression, and Purification. The DNA sequence coding for residues 1–513 of the DNA Polymerase η from *Saccharomyces cerevisiae* was PCR amplified (forward primer 5' GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GTC AAA ATT TAC TTG GAA GGA G 3' reverse primer 5'GGG GAC GAC TTT GTA CAA GAA AGC TGG GTC TCA TTT TTG TAA ATC TAT AAT ATC GAA ATT AG 3') and cloned in a two-step Gateway® reaction via the vector pDONR201 (Invitrogen) into the expression vector pDEST007, in frame with an N-terminal *Strep-Tag II* (IBA). The enzyme was expressed in *Escherichia coli* (Rosetta, Novagen) cells grown in Teriffic Broth (TB)-medium at 37 °C supplemented with carbenicillin (100 μg/mL) and chloramphenicol (5 μg/mL) until an $OD_{600} = 1.0$ was reached. Expression was induced by adding anhydrotetracycline to a final concentration of 200 ng/mL and the incubation was continued at 16 °C for 4–5 h. Cells were harvested by centrifugation, resuspended in lysis buffer [100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM ethylenediaminetetraacetic (EDTA)] with a cocktail of protease inhibitors (Roche) added, and lysed using a high pressure homogenizer (Avestin Europe GmbH). The cell lysate was cleared by centrifugation and applied onto a Strep-Tactin column (IBA). The protein was eluted from the column with the same buffer containing 2.5 mM desthiobiotin. The fractions were pooled and concentrated in 50 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM EDTA, 5 mM DTT, 5% glycerol and loaded onto a HiTrap Heparin column (GE Healthcare). The protein was eluted with a linear gradient of the same buffer containing 800 mM NaCl. The buffer was exchanged to 20 mM Tris-HCl pH 7.6, 50 mM KCl, 10 mM β-mercaptoethanol, 10% glycerol using centrifugal filter devices (Millipore). The enzyme was stored at a concentration of 10 mg/mL at −80 °C.

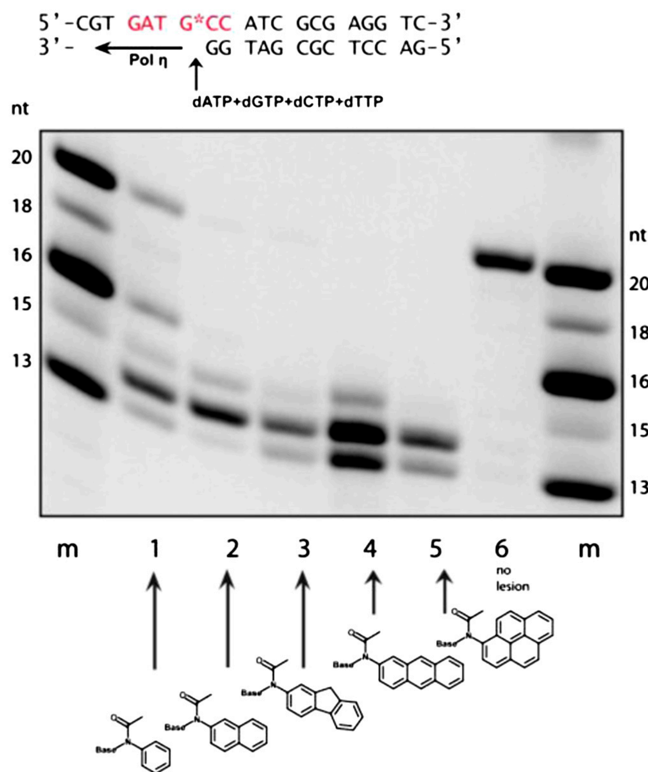

Fig. S1. Primer extension reaction with various bulky adducts in the same sequence context at lower enzyme concentrations than shown in Fig. 2*A* in the main text. Reaction conditions are 150 nM Pol η, 1 μM template DNA, 200 μM deoxynucleotides, and 5 min incubation, 30 °C. 1 = AAB-dG, 2 = AAN-dG, 3 = AAF-dG, 4 = AAA-dG, 5 = AAP-dG, and 6 = undamaged template strand. Here all bulky adducts except the small AAB-dG adduct (lane 1), are a strong block for Pol η. The polymerase inserts the next base opposite the lesions (dCTP) but extension from the lesions is problematic. Only for the benzene adduct, the 20mer full extension product is observed.

5'-CGT GAT G*CC ATC GCG AGG TC-3'
3'- ← Pol η GG TAG CGC TCC AG-5'
↑
ddATP or ddGTP or ddCTP or ddTTP

nt 20 18 16 15 13 — A T G C all — m 1 2 3 4 5

Fig. S2. Incorporation of individual nucleotides with AAA-dG containing oligonucleotides and lesion-free oligonucleotides (200 nM Pol η, 1 μM template DNA, 200 μM of each dideoxynucleotide of all four dNTPs; 1 = ddATP, 2 = ddTTP, 3 = ddGTP, 4 = ddCTP, 5 = all four dNTPs).

A B C D M74 R73 Q55 I60 L128 S58

Fig. S3. Exemplary electron density of Pol η in complex with DNA containing AAF-dG and AAA-dG. (*A*) Stereo figure of the $2F_o$-DF_c composite-omit electron density map of the region surrounding the acetylaminofluorene lesion contoured at 1σ level. F_o-DF_c omit electron density map contoured at 3σ of the AAF-dG complex B (*B*), AAA-dG in complex B (*C*) and AAA-dG in complex A (*D*). The active site residues are shown in gray and the DNA in gold, with the lesion highlighted in blue (AAF-dG) and cyan (AAA-dG). In order to highlight the relative location of the AAF/AAA adduct in respect to the DNA Figs. *B–D* were arranged along the base pair 3′ to the lesion.

A B

Fig. S4. Schematic representation of Pol η binding AAF-dG (*A*) and AAA-dG (*B*) containing DNA. In the complexes B, the enzyme rotates the DNA in respect to the fluorene moiety in order to free its active site for an incoming dCTP. The protein is shown as semitransparent surface representation (gray) with the DNA as stick model (gold) and the lesion highlighted in orange. Active site residues are also depicted as stick model (blue). The catalytic calcium ions in the active site are displayed as red spheres.

A

B

C

Fig. S5. Structural superposition of known yeast Pol η structures. The proteins were least-square fitted using the main-chain atoms of residues number 1–230 [LSQKAB (1)], which were chosen as reference frame according to a distance matrix plot [DDMP (2)] (*A*) Closing of the PAD domain and slight rotation of the DNA helix in the Pol η structure in complex with the CPD lesion (PDB code 3MFI, red) compared to the AAF (this work, golden). In the 90° rotated view the finger domain (resi 1–122) was removed for clarity. The polypeptide chain is shown as ribbon and the DNA as cartoon representation. (*B*) Superposition of all known yeast Pol η—DNA complex structures: the cisplatin GG adduct (PDB codes 2R8J and 2R8K, green and blue); the cisplatin-GTG adduct (PDB code 2WTF, gray); and CPD lesion (PDB code 3MHI, red), AAF (PDB code 2XGP, gold). (*C*) Close-up in the active site with the same color scheme as in *B*. The respect triphosphates drawn as stick model and the ions as spheres. For clarification the DNA is not shown. The superposed structures are Pol η bound to lesion containing DNA: the cisplatin GG adduct (PDB codes 2R8J and 2R8K, green and blue); the cisplatin-GTG adduct (PDB code 2WTF, gray); and CPD lesion (PDB code 3MHI, red). Note in all structures with a bound triphosphate, the ions in the active site are in structural similar positions, regardless if there are the catalytic competent magnesium (Pol η—CPD) or calcium ions (all other Pol η—DNA complex stuctures). Despite closing in of the PAD domain, no structural changes occur in the active site.

1 CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
2 Richards FM, Kundrot CE (1998) Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins* 3:71–84.

Table S1. Data collection, processing, and structure refinement statistics

	Pol η—AAA-dG	Pol–AAF-dG
Data collection		
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a, *b*, *c* (Å)	103.5, 103.5, 292.7	103.44, 103.44, 292.35
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.919	0.9129
Resolution (Å)	48.0 - 2.7 (2.85-2.7)	49-2.7 (2.85-2.7)
R_{merge}	0.067 (0.377)	0.107 (0.523)
Mean $I/\sigma I$	23.0 (6.0)	26.84 (5.99)
No. of observations	215,900 (30,490)	584,408 (92,986)
No. of unique reflections	44,732 (6,421)	44,372 (6,887)
Completeness (%)	99.9 (99.8)	98.2 (96.5)
Redundancy	4.8 (4.7)	13.1 (13.6)
Refinement		
Resolution (Å)	46.6 - 2.7	48.76 – 2.7
No. reflections	42,326	44,599
R_{work}/R_{free}	21.5/25.6	22.0/27.2
No. atoms	9,076	9,070
Protein	8,147	8,152
DNA	826	824
Waters	103	89
Ions	6 calcium	6 calcium
B-factors		
Protein	45.0	29.6
DNA	65.3	45.1
Waters	36.0	20.5
rmsd		
Bond lengths (Å)	0.0126	0.0120
Bond angles (°)	1.439	1.442

Numbers in parentheses correspond to the high resolution shell.