Supplementary Information for

RNA polymerase II stalls on oxidative DNA damage via a "torsion-latch" mechanism involving lone pair-pi and CH-pi interactions

Juntaek Oh[a], Aaron M. Fleming[b], Jun Xu[a], Jenny Chong [a], Cynthia J. Burrows[b], Dong Wang[a,*]

[a]Division of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093; Department of Cellular and Molecular Medicine, School of Medicine, University of California, San Diego, La Jolla, CA 92093

[b]Department of Chemistry, University of Utah, 315 S 1400 East, Salt Lake City, UT 84112

[*] Corresponding author: dongwang@ucsd.edu

This PDF file includes:

Materials and Methods

Synthesis, purification, and characterization of damaged oligonucleotides

All OG-containing oligodeoxynucleotides (ODN) were synthesized and deprotected by the DNA-peptide core facility at the University of Utah following manufacturer's protocols (Glen Research, Sterling, Virginia). Crude samples were purified by semi-preparative anion-exchange HPLC running the following mobile phases: A = 1.5 M NaOAc (pH 7) in 10% MeCN and 90% ddH_2O, and B = 10% MeCN and 90% ddH_2O while running a flow rate of 3 mL/min and monitoring the elution via the absorbance at 260 nm. The purification salt was removed by dialysis against ddH_2O for 36 hr using a 3500 molecular weight cutoff (MWCO) membrane cassette. The purified samples were used in the synthesis of Sp and Gh via the following methods. Synthesis of Gh was achieved by mixing a 20-μM solution of OG-containing ODN in ddH_2O at 20 ºC with 120 μM Na_2IrCl_6, while letting the reaction proceed for 30 min. Synthesis of Sp was achieved by mixing a 20-μM solution of OG-containing ODN in 20 mM NaP_i (pH 7.5) buffer preincubated at 45 ºC, after which 120 μM Na_2IrCl_6 was added, while letting the reaction proceed for 30 min. After completion of the reaction to synthesize either Sp or Gh, the reactions were quenched by adding 2 mM EDTA (pH 8). Next, the reacted samples were purified using an analytical anion-exchange HPLC setup running the following mobile phases: A = 1.5 M NaOAc (pH 7) in 10% MeCN and 90% ddH_2O, and B = 10% MeCN and 90% ddH_2O while running a flow rate of 1 mL/min and monitoring the elution via the absorbance at 260 nm. The purification salt was removed by dialysis against ddH_2O for 36 hr using 3500 MWCO membrane cassettes. Product purity was determined by analytical anion-exchange HPLC and product identity was determined by ESI-MS. The OG-containing ODN calcd mass = 8736.4, expt mass = 8736.0; Gh-containing ODN calcd mass = 8726.4, expt mass = 8725.6; Sp-containing ODN calcd mass = 8752.4 expt mass = 8752.0. Long

template strand DNA for the full-bubble transcription assay was prepared by ligation and PAGE purification. Other undamaged DNA and RNA oligomers were purchased from IDT.

Protein purification

Ten-subunit Pol II from *Saccharomyces cerevisiae* was purified as previously described (1-3). Briefly, 10 subunit Pol II containing recombinant protein A tagged Rpb3 was purified by using IgG resin (GE Healthcare). After cleavage of the protein A tag by TEV protease, 10-subunit complex was further purified by Hitrap Heparin and anion exchange chromatography (GE Healthcare). Twelve-subunit Pol II was prepared by adding four molar ratios of Rpb4 and Rpb7 to 10 subunit Pol II and further purified by size exclusion chromatography. Final buffer conditions were 20 mM Tris (pH 7.5), 40 mM KCl, 5 mM $MgCl_2$, and 5 mM DTT (elongation buffer). Recombinant Rad26 protein was expressed and purified as described (3).

In vitro transcription assay

All transcription assays were performed by using 12-subunit Pol II as described (3-5). Briefly, the mini-scaffold was prepared by annealing 200 nM of 5' ^{32}P-labeled RNA, 600 nM of template strand DNA, and 800 nM of non-template strand DNA in elongation buffer. The prepared mini-scaffold was preincubated with RNAP II in elongation buffer at room temperature for 10 min and added to an equal volume of rNTP or TFIIS to start the reaction. The final concentration of each component in the reaction mixture was 20 nM of mini-scaffold, 120 nM of RNAP II, 0-1 mM of rNTP, 0-1 μM of TFIIS. At each timepoint, the reaction mixture was added to the quench-loading buffer (90% formamide, 50 mM EDTA, 0.05% xylene cyanol and 0.05% bromophenol blue) at the volume ratio of 1: 4. All samples were denatured at 95 °C for 10 min and analyzed by

12% of denaturing urea/TBE PAGE gel. For the full-bubble scaffold assay, tsDNA and RNA were annealed and allowed to form a complex with Pol II. EC was formed by adding biotin-labeled ntsDNA. The assembled elongation complex was incubated with 20 μL of Streptavidin magnetic beads (NEB) for 30 min at room temperature (23 °C) and subsequently washed with elongation buffer (EB), followed by EB with 0.3 M NaCl, EB with 1 M NaCl, EB with 0.3 M NaCl and finally EB. Transcription was initiated by adding 1 mM of rNTPs with additional 3 mM dATP to facilitate ATPase activity of Rad26. The final concentration for Rad26 was 100 nM and TFIIS was 300 nM. Sequences for full-bubble scaffold were: tsDNA, 5'-CCTTCTCTCTCTCGCTGAXCCTCTCGATGCGGTCACGCTCCATCACATAAGGGATAAC ACC-3' (X = G, Gh, *R*-Sp or *S*-Sp); ntsDNA 5' Biotin-TTTGGTGTTATCCCTTATGTGATGGAGCGTGACCGCATCGAGAGGCTCAGCGAGAGA GAGAAGG-3'; RNA, 5'- GAGCGUGACC-3'. All transcription assays were performed three times.

Crystallization and structure determination

Crystallization of 10-subunit Pol II was performed as previously described (4, 5). Briefly, a mini-scaffold was prepared by annealing 1:2:2 molar ratio of tsDNA, RNA, and ntsDNA in elongation buffer. The EC was formed by incubating 2.5 μM of Pol II and 12.5 μM of the mini-scaffold (tsDNA concentration) at room temperature for 30 min and 4 °C for 30 min. To remove excess scaffold and change the buffer to 25 mM Tris pH 7.5, 20 mM NaCl, 5 mM DTT, 1 μM $Zn(OAc)_2$, 100 μM EDTA, ultrafiltration was performed. Finally, 6-8 mg/mL of protein was subjected to crystallization by the hanging drop vapor diffusion method with 390 mM ammonium phosphate pH 6.0, 5 mM DTT, 5 mM dioxane, and 9-13 % (w/v) PEG 6,000. After incubation at

22 °C for up to 14 days, crystals were subsequently moved to a cryo solution (100 mM MES pH 6.0, 350 mM NaCl, 5mM DTT, 5mM Dioxane, 16 % PEG 6,000, and 17 % PEG400), incubated at 4 °C overnight, and flash-frozen in liquid nitrogen. For soaking, we added 5 mM of AMPCPP, 10 mM of ATP, 2 mM of CMPCPP or 10 mM of UTP, respectively. The same concentration of $MgCl_2$ was included together for ligand soaking.

X-ray diffraction datasets were collected at beamlines 5.0.3, 8.2.1 and 8.2.2, Advanced Light Source, Lawrence Berkeley National Laboratory. Collected images were processed by XDS, followed by pointless and aimless (6, 7). CC1/2, half correlation coefficient around 0.3 was considered as a criterion for high resolution determination (8). All crystals had a space group C2, with one elongation complex in the asymmetric unit. Molecular replacement was performed by using the EC structure without damage (PDB code: 2E2J) (1) as a search model, using Phenix (9). Several rounds of manual model building and refinement was performed by using COOT and Phenix (10). As part of quality check, all models were evaluated using composite simulated annealing omit map to eliminate potential model bias.

For the model building and refinement of state 1 and 1*, we initially tried to build a single conformer to fit the electron density corresponding to the Gh lesion. We first built a model for a single conformer state 1* based on the observed electron density over the bridge helix. However, we observed residual electron density in the +1-loading site after we built the state 1*. Similarly, when we built in a single conformer as state 1, we observed residual electron density above the bridge helix. These results suggest that a single conformer in state 1* or state 1 cannot fully explain the electron density. We then decided to model and refine the two conformers state 1* and state 1

with partial occupancy. The occupancy of state 1 and 1* was 0.45 and 0.55, respectively, after the final refinement. To remove model bias, a polder omit map was prepared by Phenix, confirming the presence of both state 1 and state 1* (Fig. S5A) (11). Data collection and refinement statistics are summarized at Table S2. Figures with structural information are prepared by using Pymol (12).

Structure modeling

For modeling of various hydantoin lesions, ligand definition of *S*-Sp (PDB: SDH) was downloaded. Definitions for *R*-Sp, Hyd-T, Hyd-C and Hyd-5mC were prepared by using ChemDraw software, followed by eLBOW in Phenix for optimization (9). Each model was prepared by replacing the Gh lesion in state 1, 1*, 2E, 3 or 5 with various hydantoin lesion, superposing and replacing the hydantoin moiety of each lesion in COOT (10). Bypass models in Figure S9B was prepared by superimposing and replacing the N3 and C4 amine groups of the -1 cytosine in dG structure with N3 and C4 carboxyl group of each hydantoin lesion, using same plane of cytosine in COOT. All the coordinates of the structure models are deposited in the supplementary information. Data are deposited as Datasets S1-S15.

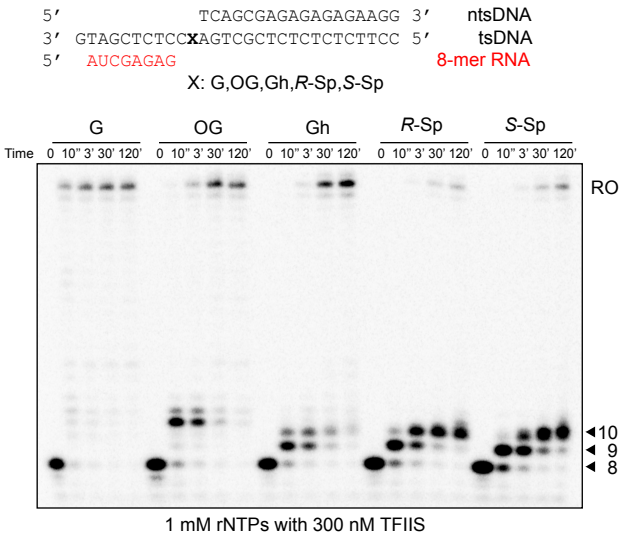

Figure S1. Transcription assay in the presence of 300 nM of TFIIS.

Experimental conditions are the same as in Figure 1C.

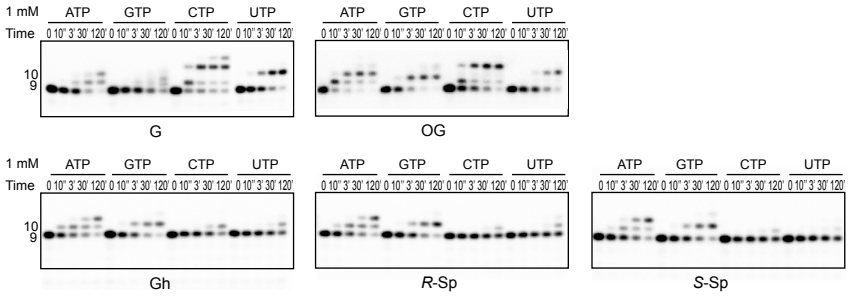

Figure S2. Single nucleotide incorporation assay opposite oxidative DNA lesions.

Experimental conditions are the same as in Figure 3B, except 1 mM of each rNTP was used as final concentration.

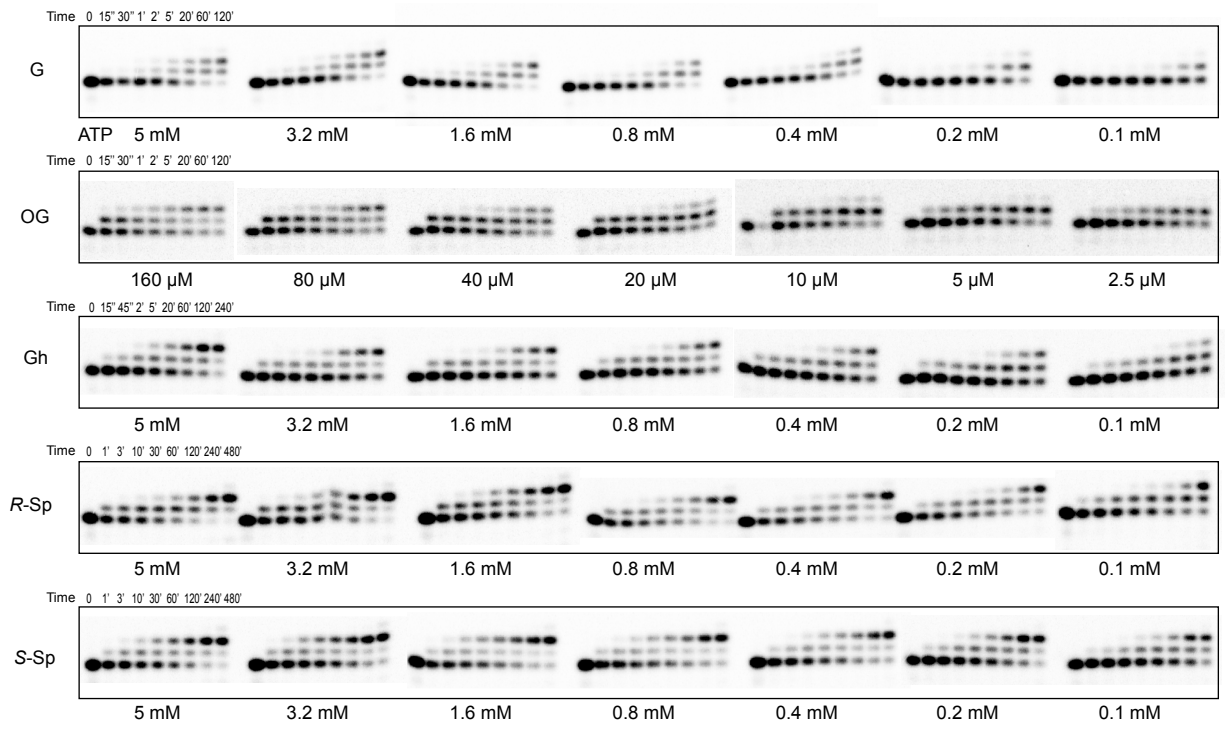

Figure S3. Transcription assay gels for kinetic analysis.

Raw gel data of kinetic analysis. We used different time points and concentrations of ATP, due to the difference of activity between lesions. Time points for G and 8OG: 0 (control), 15s, 30s, 1m, 2m, 5m, 20m, 1hr, and 2hrs. Time points for Gh: 0, 15s, 45s, 2m, 5m, 20m, 1hr, 2hrs, and 4hrs. Time points for *R*-Sp and *S*-Sp: 0, 1m, 3m, 10m, 30m, 1hr, 2hrs, 4hrs, 8hrs. Kinetic parameters are shown in Table S1.

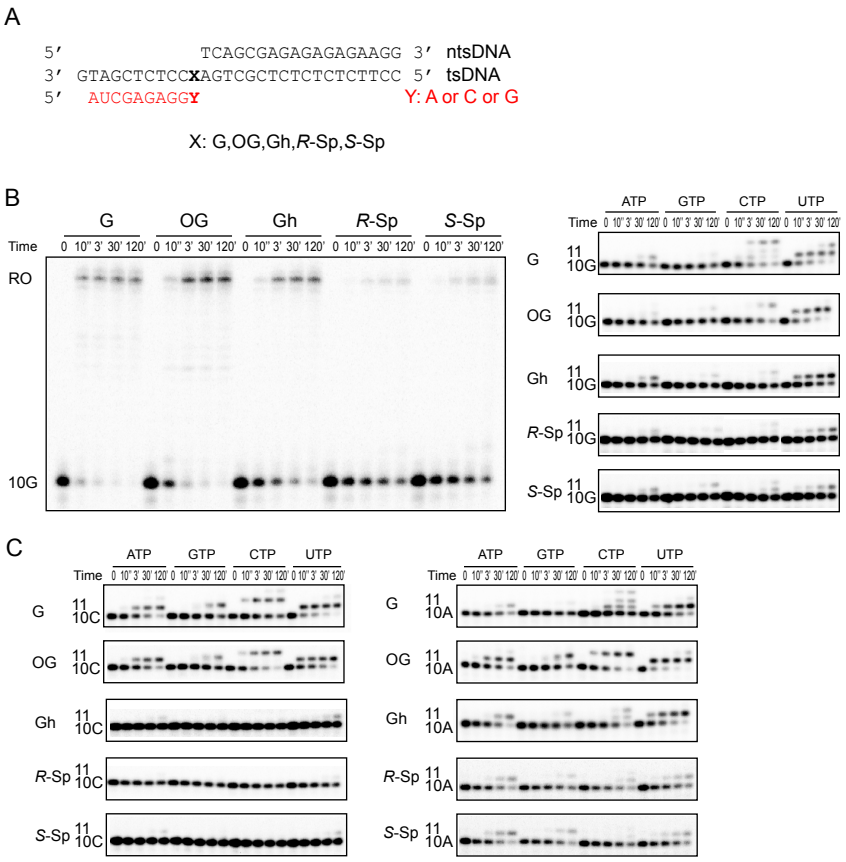

Figure S4. Impact of base pairing combination on extension step.

(A) Scaffold used in the transcription extension assay with 10-mer RNA. (B) Extension (left panel) and single nucleotide incorporation assay (right panel) of 10G scaffold, as described in Figure 4. (C) Original gels of Figure 4 C and E, which show single nucleotide incorporation using different rNTPs. Each reaction was incubated with either a mixture of four 1 mM rNTPs or individual rNTPs as labeled.

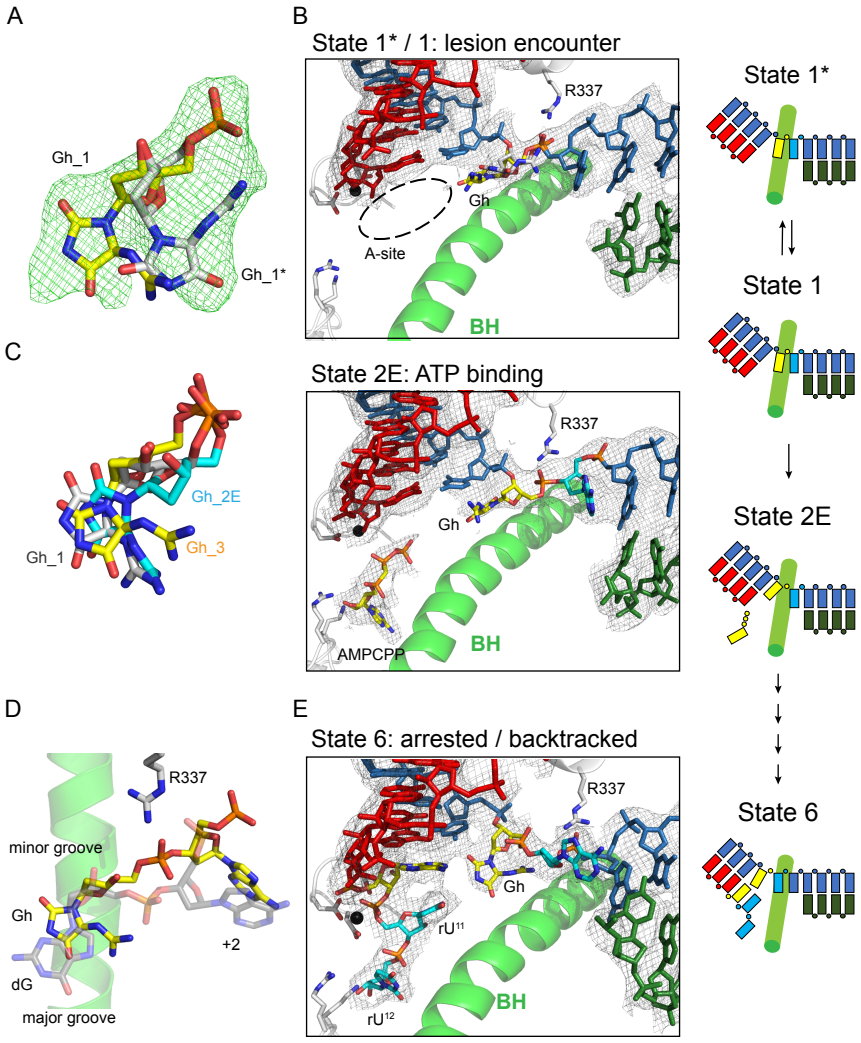

Figure S5. Crystal structure of Pol II with Gh-containing scaffold.

(A) Two alternative conformations of Gh base in apo structure. Fo-Fc polder omit map (omit selection: Gh lesion) is contoured at 7 σ (11). (B) Active site of EC in State 1 (1*) and State 2E. Color codes and orientation are the same as in Figure 8. 2Fo-Fc electron density map is contoured at 1.1 σ. (C) Comparison of Gh base in three different structures, apo (State 1, white), AMPCPP soaked (State 2E, cyan) and ATP soaked (State 3, yellow). (D) Comparison of Gh base (yellow) and dG (gray). +2 tsDNA bases are shown to highlight interaction between Gh base and R337 in Rpb1. (E) Prolonged incubation of UTP results in two successive incorporation of UTP (state 6). Note that second UTP is located in a backtracked state.

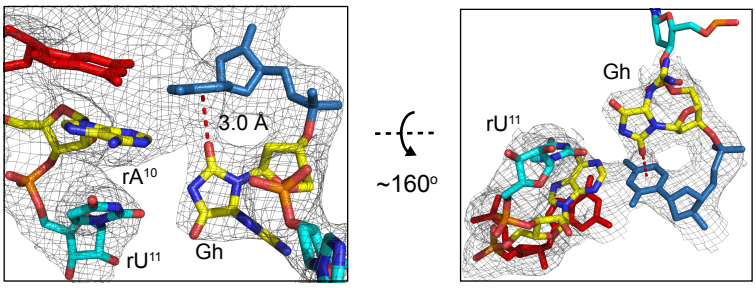

Figure S6. Hydantoin ring of Gh lesion forms lone pair-π interaction with upstream base.

Enlarged view of active site of State 5 to highlight lone pair-π interaction with the similar orientation from Fig. 8B. Distance between carbonyl group of Gh base and center of cytosine ring was 3.0 Å. 2Fo-Fc electron density map is contoured at 1.3 σ.

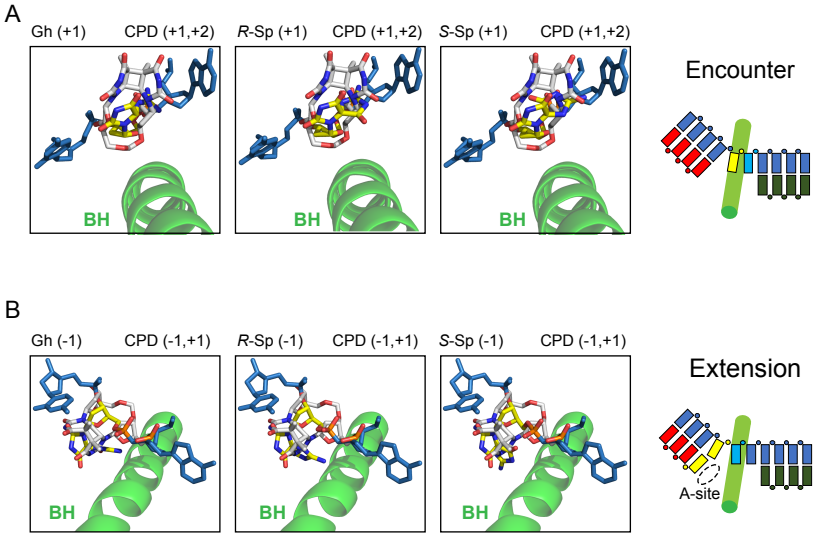

Figure S7. Structural superposition of Pol II EC with Gh or Sp to Pol II EC with CPD lesion.

Pol II ECs containing Gh or Sp lesions are superposed with Pol II EC CPD lesion (using Pol II residues for alignment). The DNA lesions have similar spatial arrangement. PDB code for CPD (+1,+2) and CPD (-1,+1) is 4A93 and 2JA7, respectively (13, 14).

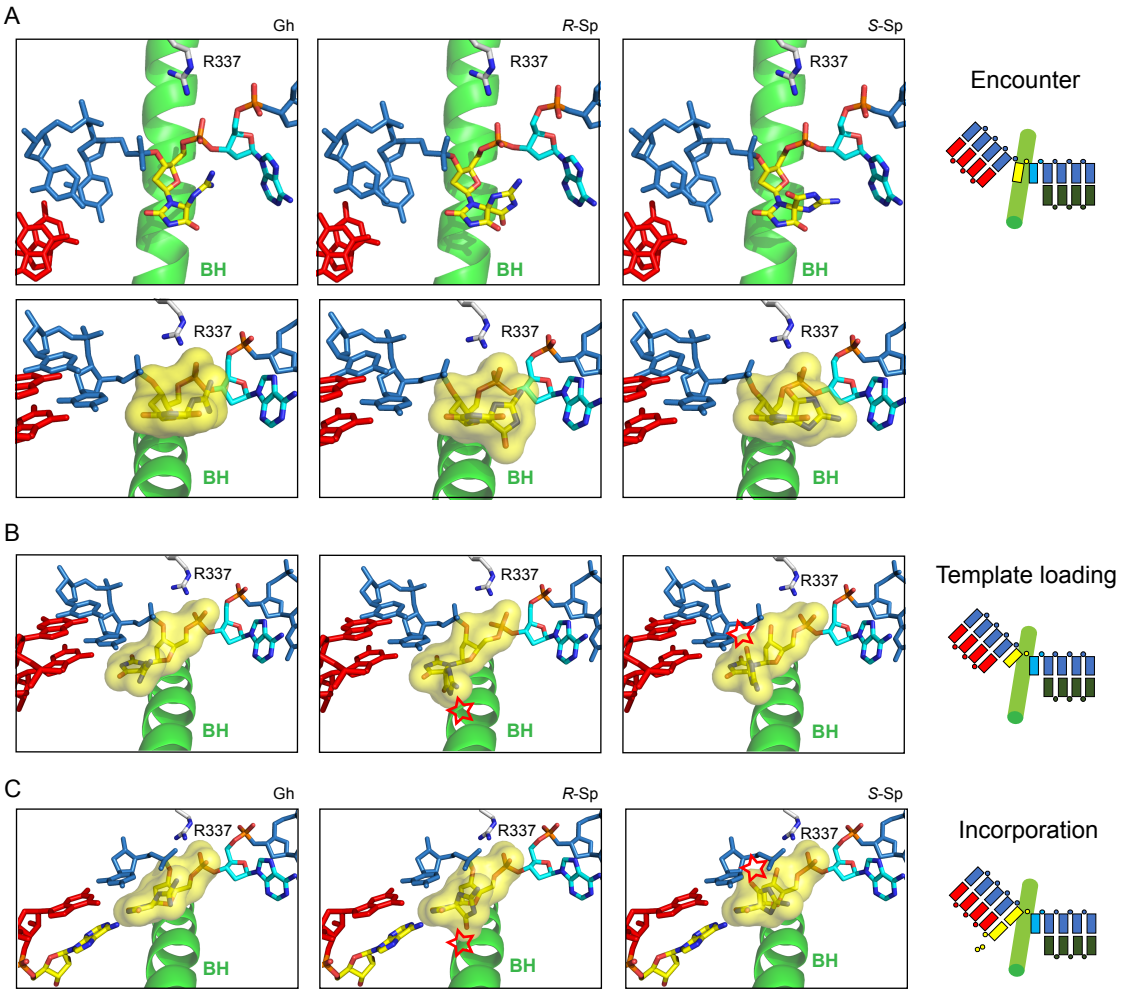

Figure S8. Structural model of *R*-Sp and *S*-Sp lesion during encounter, loading and insertion step. Sp lesion models are generated by superposing hydantoin group of Gh in state 1* (A) or state 1 (B) or state 3 (C). Surface of each lesion is shown in yellow. Note that *R*-Sp overlaps (clash) with bridge helix during loading and insertion step, while *S*-Sp overlaps with base of 1bp upstream template strand. Coordinates of each model are uploaded as; Dataset S1: S8A_RSp, Dataset S2: S8A_SSp, Dataset S3: S8B_RSp, Dataset S4: S8B_SSp, Dataset S5: S8C_RSp and Dataset S6: S8C_SSp, respectively.

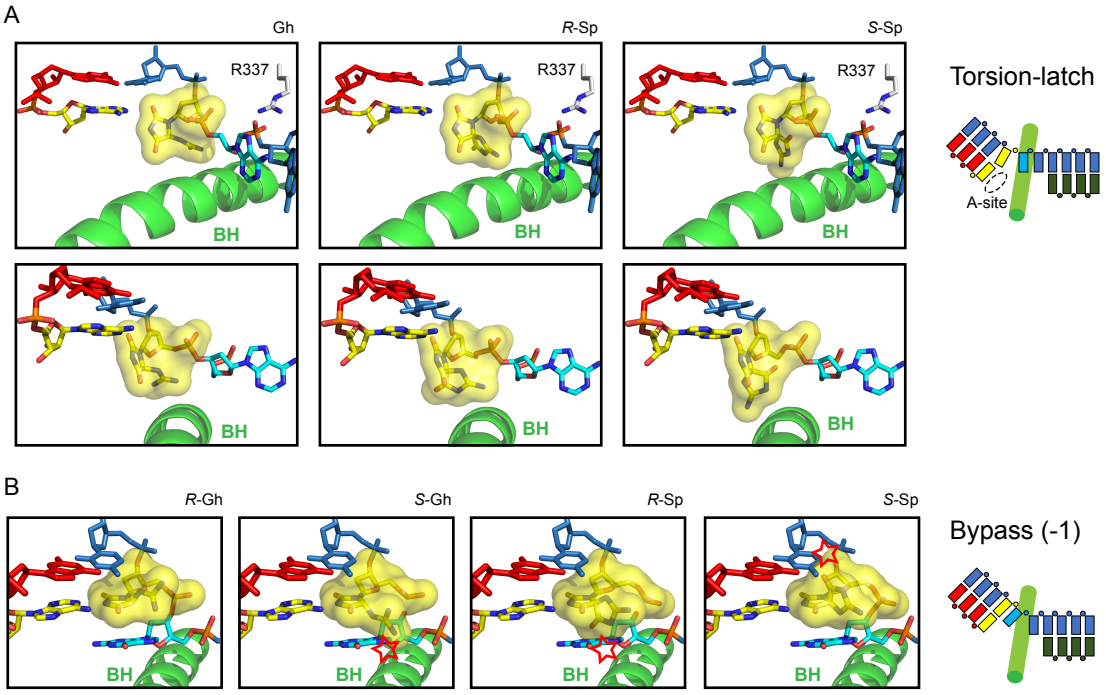

Figure S9. Structural model of *R*-Sp and *S*-Sp lesion during extension and bypass step.

(A) Sp lesion models are generated by superposing hydantoin ring of Gh in state 5. (B) Hydantoin ring of four Gh and Sp lesions are superposed to -1 Cytosine base in non-damaged dG structure (Fig. 6B). Carboxy group and N4 of cytosine base was used for superposition. Note that except *R*-Gh, all other lesions collide with base of template strand. Coordinates of each model are uploaded as; Dataset S7: S9A_RSp, Dataset S8: S9A_SSp, Dataset S9: S9B_RGh, Dataset S10: S9B_SGh, Dataset S11: S9B_RSp and Dataset S12: S9B_SSp, respectively.

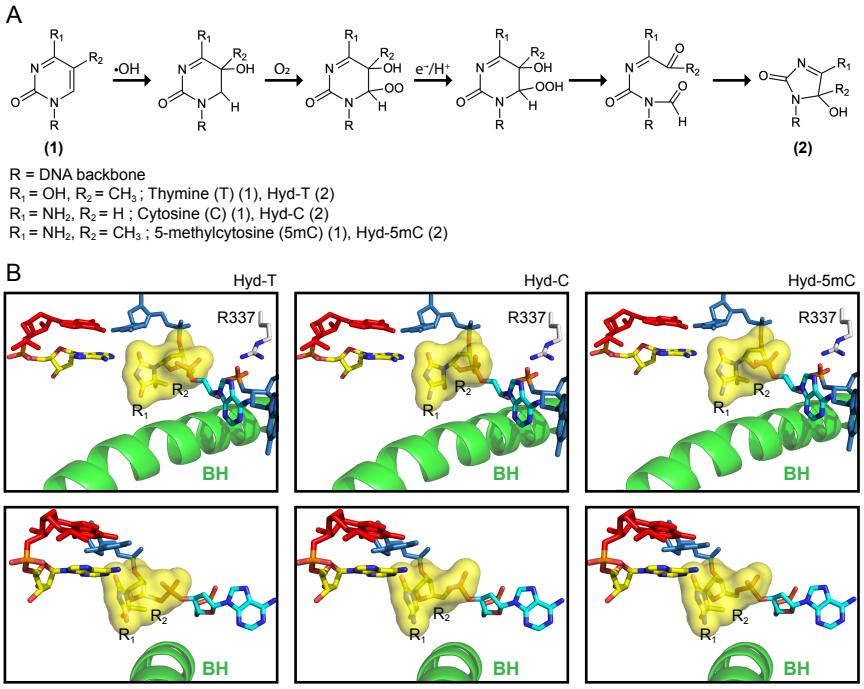

Figure S10. Structural model of pyrimidine-derived hydantoin lesions.

(A) Schematic representation of formation of three hydantoin lesions, Hyd-T, Hyd-C and Hyd-5mC by hydroxyl radical induced oxidation. (B) Structural model of hydantoin lesions in (A), where hydantoin ring was superposed to that of Gh in state 5. Note that there is no obvious steric clash in all kinds of lesions derived from T, C, and 5mC. Coordinates of each model are uploaded as; Dataset S13: S10_Hyd_T, Dataset S14: S10_Hyd_C and Dataset S15:S10_Hyd_5mC, respectively.

Table S1. Kinetic analysis of ATP incorporation efficiency.

	k_{pol} (min^{-1})	K_d (μM)	k_{pol}/K_d ($min^{-1}μM^{-1}$)	Relative activity
G	0.12 ± 0.01	1100 ± 310	(1.09 ± 0.32) × 10^{-4}	1
OG	6.3 ± 0.2	27.5 ± 2.4	0.229 ± 0.021	2100
Gh	3.7 ± 0.2	240 ± 51	(1.54 ± 0.34) × 10^{-2}	141
R-Sp	1.8 ± 0.1	547 ± 50	(3.29 ± 0.35) × 10^{-3}	30
S-Sp	0.60 ± 0.05	320 ± 87	(1.88 ± 0.53) × 10^{-3}	17
AP	0.43 ± 0.01	814 ± 35	(5.28 ± 0.26) × 10^{-4}	5
T	642 ± 6	15 ± 2	42.8 ± 0.5	390000

To obtain rate constants (k_{pol}) and apparent substrate binding affinity ($K_{d.,app}$), intensities of each band in the raw data (Fig. S3) were quantified using Image Lab and converted to relative ratio of extended RNA transcript (%). Initial velocity at each ATP concentration was calculated using Prism6. Initial velocity vs ATP concentration was further analyzed by fitting Michaelis-Menten equation to calculate k_{pol} and $K_{d,app}$. We compared k_{pol}, $K_{d.,app}$, $k_{pol}/K_{d.,app}$ and relative activity where ATP misincorporation to template G is set to 1. Kinetic parameters of dT and abasic site are referred from our previous paper (4). Summarizing graph is shown in Figure 3C.

Table S2. Data collection and refinement statistics.

	Gh_state 1	Gh_state 2E	Gh_state 3	Gh_state 4	Gh_state 5	Gh_state 6	dG_state 1
PDB ID	**6UPX**	**6UPY**	**6UPZ**	**6UQ0**	**6UQ3**	**6UQ1**	**6UQ2**
Data collection[a]							
Resolution, Å	48.2 – 3.4 (3.52 – 3.4)[b]	49.2 - 3.4 (3.52 - 3.4)	49.0 - 3.1 (3.21 -3.1)	49.1 - 3.56 (3.69 - 3.56)	48.3 - 3.47 (3.59 - 3.47)	48.5 - 3.6 (3.73 - 3.6)	49.8 - 3.2 (3.31 - 3.2)
Space group				C 1 2 1			
Unit cell (a b c)	169.0 223.3	168.5 223.1	167.7 222.9	167.9 223.3	169.2 222.6	170.5 223.5	165.7 223.3
β (α, γ = 90°)	193.1	193.1	193.5	193.1	194.6	194.9	193.5
	101.0	100.9	100.9	100.9	101.3	101.6	99.4
Unique reflections	96348 (9614)	95681 (9488)	126050 (12517)	83439 (8323)	91066 (9096)	82622 (8208)	114068 (11384)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.9 (99.7)	99.5 (98.7)	99.9 (99.9)	99.8 (99.6)	99.9 (99.7)	99.8 (99.8)	99.8 (99.5)
Mean I/sigma(I)	7.3 (0.9)	6.7 (1.0)	8.3 (1.2)	3.7 (1.0)	5.0 (0.9)	5.4 (1.1)	7.8 (0.7)
Wilson B-factor	86.13	89.9	64.3	75.6	75.4	93.3	80.5
R-merge	0.14 (1.00)	0.15 (1.07)	0.12 (0.84)	0.27 (0.90)	0.23 (0.99)	0.18 (0.83)	0.11 (1.02)
CC1/2[c]	0,99 (0.37)	0.99 (0.41)	0.99 (0.46)	0.93 (0.36)	0.95 (0.36)	0.98 (0.34)	0.99 (0.46)
Refinement							
R-work / R-free	0.220 / 0.266	0.227 / 0.282	0.220 / 0.264	0.236 / 0.279	0.234 / 0.285	0.242 / 0.288	0.229 / 0.280
RMS (bonds)	0.006	0.004	0.006	0.006	0.005	0.005	0.006
RMS (angles)	1.09	0.96	1.16	1.12	1.08	1.04	1.20
Ramachandran outliers (%)	0.07	0.03	0.09	0.63	0.00	0.03	0.46
Average	112.5	110.1	89.6	91.6	98.3	111.0	97.5
B-factor							
protein	109.8	107.8	82.7	89.4	95.5	108.0	94.8
RNA	125.7	118.4	125.4	101.5	119.0	148.4	121.8
DNA	199.4	184.4	183.0	161.4	161.7	196.3	182.7

[a]This table was prepared by PHENIX and CCP4i2, with minor modifications (9, 15).
[b]Numbers in parentheses indicate the highest resolution shell.
[c]CC1/2, half correlation coefficient suggested by Karplus and Diederichs, is used to define high resolution cutoff (8).

Movie S1. Movie showing overall structure and Gh lesion encounter complex. Two conformers of Gh lesions, state 1* and 1, are shown.

Movie S2. Direct comparison between damaged Gh (state 1 and 1*) and undamaged dG at state 1.

Movie S3. Movie showing serial transition from state 1* -> 2E -> 3 -> 4 and finally to 5, which contains states of lesion encounter, ATP binding, incorporation and half-translocation (torsion-latch).

References

1. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941-954 (2006).
2. D. Wang *et al.*, Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203-1206 (2009).
3. J. Xu *et al.*, Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **551**, 653-657 (2017).
4. W. Wang, C. Walmacq, J. Chong, M. Kashlev, D. Wang, Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc Natl Acad Sci U S A* **115**, E2538-E2545 (2018).
5. J. Oh, J. Xu, J. Chong, D. Wang, Structural and biochemical analysis of DNA lesion-induced RNA polymerase II arrest. *Methods* 10.1016/j.ymeth.2019.02.019 (2019).
6. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
7. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).
8. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).
9. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
10. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
11. D. Liebschner *et al.*, Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol* **73**, 148-157 (2017).
12. Schrodinger, LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.
13. C. Walmacq *et al.*, Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. *Mol Cell* **46**, 18-29 (2012).
14. F. Brueckner, U. Hennecke, T. Carell, P. Cramer, CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859-862 (2007).
15. L. Potterton *et al.*, CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* **74**, 68-84 (2018).