www.sciencemag.org/cgi/content/full/315/5813/859/DC1

Supporting Online Material for

CPD Damage Recognition by Transcribing RNA Polymerase II

Florian Brueckner, Ulrich Hennecke, Thomas Carell,* Patrick Cramer*

*To whom correspondence should be addressed. E-mail: thomas.carell@cup.uni-muenchen.de (T.C.); cramer@lmb.uni-muenchen.de (P.C.)

Published 9 February 2007, *Science* **315**, 859 (2007)
DOI: 10.1126/science.1135400

This PDF file includes:

Supporting Online Material

"CPD damage recognition by transcribing RNA polymerase II"
by F. Brueckner, U. Hennecke, T. Carell, and P. Cramer.

Materials and methods

Preparation of Pol II-nucleic acid complexes. Endogenous *S. cerevisiae* ten-subunit Pol II core enzyme and recombinant Rpb4/7 heterodimer were purified as described (*S1*). The thymine-thymine CPD formacetal analogue-containing DNA oligonucleotides were synthetized as described (*S2*). 5'-Bromo uracil phosphoamidite was purchased from Glen Research and incorporated as described by the manufacturer. Nucleic acid scaffolds (Figs. 1A, S3) were annealed by mixing equimolar amounts of synthetic template DNA, nontemplate DNA, and RNA in TE buffer pH 7.4 at a final concentration of 100 μM, heating the mixture to 95°C for 2 min, and slow-cooling to room temperature. Stoichiometric transcribing Pol II-nucleic acid complexes were assembled by incubating core Pol II for 10 min with 1.5 molar equivalents of nucleic acid scaffold, followed by 20 min incubation with 5 molar equivalents of recombinant Rpb4/7 in assembly buffer (100 mM Hepes pH 7.5, 40 mM ammonium sulphate, 1 mM $MgCl_2$, 10 μM $ZnCl_2$, 5 % glycerol, 10 mM DTT) at 20°C. The complexes were purified by size exclusion chromatography (Superose 6 10/300 GL) in Pol II buffer (5 mM Hepes pH 7.25, 40 mM ammonium sulphate, 10 μM $ZnCl_2$, 10 mM DTT).

Crystallization and crystal treatment. Purified Pol II-nucleic acid complexes were concentrated to 3.5-4.5 mg/mL and an additional amount of the nucleic acid scaffold was added prior to crystallization to a final concentration of 2 μM. Crystals were grown at 22°C with the hanging drop vapor diffusion method by mixing 2 μL of sample solution with 1 μL of reservoir solution (200 mM ammonium acetate, 150 mM magnesium acetate, 50 mM Hepes pH 7.0, 5% PEG 6000, 5 mM TCEP). Crystals were harvested in mother solution after 10-20 days of growth, when they had reached their maximum size (0.4 x 0.3 x 0.2 mm). Subsequently the crystals were transferred stepwise to mother solution containing additionally 0-20 % glycerol over 5 hr, slowly cooled to 8°C, and flash-cooled by plunging into liquid nitrogen.

X-ray structural analysis. Diffraction data were collected in 0.25° increments at the protein crystallography beamline X06SA of the Swiss Light Source (Table S1). Raw data were processed with Denzo and Scalepack (S*3*). Structures were solved by molecular replacement with the program Phaser (S*4*), using the structure of the complete 12-subunit Pol II elongation complex without nucleic acids as a search model (PDB 1Y1W (S*5*)). The molecular replacement solution was subjected to rigid body refinement with CNS (S*6*).

Model building was done using the program O (S*7*). The nucleic acids were built into the initial F_o-F_c electron density maps (Fig. S1A). The register of the nucleic acids was unambiguously defined by bromine labeling. A thymine residue in the template strand was replaced for 5-Bromo uracil, diffraction data of the resulting complex were recorded at the wavelength of the bromine K absorption edge, and the resulting anomalous difference Fourier maps revealed single peaks demarking the positions of the bromine atom (Fig. S1C). After minor adjustments to the protein model, atomic positions and B-factors were refined with CNS. Refinement was monitored carefully with the free R factor, calculated from the same set of excluded reflections as in refinement of complete Pol II (S*8*) and the complete Pol II elongation complex (S*5*).

Transcript extension assay. Stoichiometric transcribing complexes of complete Pol II, containing the ten-subunit core and Rpb4/7, were assembled and purified as described above, but in transcription buffer (20 mM Hepes pH 7.6, 60 mM ammonium sulphate, 8 mM magnesium sulphate, 10 μM $ZnCl_2$, 10 % glycerol, and 10 mM DTT). Transcribing complexes of ten-subunit core Pol II were assembled by incubating core Pol II for 20 min with 1.5-2 molar equivalents of the respective nucleic acid scaffold in transcription buffer at 20°C. Core Pol II elongation complexes were used for measurements presented in Figs. 2E and S2C-D. In these assays, core Pol II is generally indistinguishable from the complete Pol II. For all other measurements, stoichiometric elongation complexes of complete Pol II were used. The RNAs used for extension assays were identical to the ones used for structural studies, except five additional nucleotides (5'-UGCAU-3') and a fluoresceine label at their 5'-end. For the transcript extension and cleavage reactions, 0.2-0.4 μM of Pol II-nucleic acid complex were incubated with NTPs or TFIIS, respectivelyl, at 28°C in transcription buffer. The reaction was stopped by adding 50 mM EDTA pH 7.4. Complexes were disassembled and protein was denatured by gradually heating to 95°C in 12 min and then incubating at 95°C for 4 min. After removal of protein by centrifugation, the nucleic acid samples were desalted using C18 ZipTips (Millipore). Fluoresceine-labeled RNA extension products were analyzed using a Beckmann Coulter PACE-MDQ DNA system with laser-induced fluorescence detection (excitation and detection wavelengths were 488 nm and 520 nm, respectively). Separations were carried out using fused silica capillaries filled with polyacrylamide gel (Beckmann Coulter) and Tris borate buffer (0.1 M Tris borate, 1 mM EDTA, 7 M urea, pH 8.4). Samples were injected by a voltage pulse (10 kV for 10 s) and separated at 8 kV for 50 min. The nucleic acids of the transcript extension reactions were additionally analysed by mass spectrometry using a Bruker Daltonics autoflex 2 MALDI-TOF spectrometer. This confirmed the identity of the RNA extension products.

Supporting text

Use of a CPD analogue. A thymine-thymine CPD formacetal analogue was used that carries a methylene group instead of a phosphate group between the two thymines (*S10*). As a consequence there is a missing negative charge in the DNA template strand backbone between positions +2 and +3, which may be partially responsible for downstream DNA mobility in complex A. The analogue was used because it can be synthesized in large quantities, as required for crystallographic studies. All current evidence argues that the CPD analogue is structurally and functionally equivalent to the natural CPD. First, high-resolution small molecule crystallographic data of the CPD formacetal analogue show that it adopts the same structure and conformation than the normal CPD with a phosphate backbone (*S10-S11*). Second, the base pairing properties of the lesion and the lesion analogue, as judged from melting point studies, are the same (*S12*). Third, the formacetal linkage is a nearly exact bioisostere of the natural phosphate (*S13*). Forth, the small differences, which are in the range of only around 0.1 Å are without significance for our results and conclusions. The change to the formacetal does neither affect the nucleobases nor the sugar units, the only difference is the missing charge of the backbone. Finally, the CPD analogue is well-accepted as an analogue for recognition by a photolyase (*S14*).

Structural similarity of lesion-containing and lesion-free elongation complexes. The RMS deviation of Cα atoms between the lesion-free complete Pol II elongation complex (PDB code 1Y1W) and the CPD-containing complexes is 0.4 Å in all four cases. The RMS deviation of Cα atoms between a lesion-free core Pol II elongation complex (PDB code 1R9S) and the CPD-containing complexes is 0.9 Å in all four cases.

Minimal nucleic acid scaffolds and hybrid length in elongation complexes. The structure of the complete Pol II elongation complex showed seven DNA-RNA base pairs at positions –1 to –7, with Watson-Crick edges of the bases at position –8 at the upstream end of the hybrid about 4 Å apart (*S5*). This slightly increased base distance may have resulted from a bulky A-A DNA-RNA mismatch at position –9. The current structures A and C contain a smaller A-C mismatch at position -9, and show a hybrid length of eight base pairs (positions –1 to –8). To test if the scaffold design influences the outcome of the biochemical results, we tested complete scaffolds that included a nontemplate strand in biochemical assays. The results were essentially not influenced by the nontemplate strand, showing that the use of minimal scaffolds in the functional studies is justified. The use of the minimal scaffolds in structural studies is justified by the highly similar nucleic acid structure observed in the complete elongation complex structure (*S5*).

Rate constants and kinetic measurements. The rate constant for normal elongation *in vitro* is greater than 20 sec^{-1} (S*15*), too fast to be monitored by the kinetic methods used here. This is why the first incorporation into RNA of complex A appears to be instantaneous.

Detection of apparently "back-stepped" complexes. The nucleic acid scaffolds are designed to give post-translocated elongation complexes. The observed back-stepping in complexes B and D leads to the pre-translocated state, and occurs although it results in an A-C mismatch at position –8 at the upstream end of the hybrid. When the anomalous difference Fourier maps in complexes B and D (Fig. S1C) were contoured at the low level of 2.5 σ, there were no peaks at potential bromine positions in neighboring nucleotides. Thus there were no significantly populated alternative states with different nucleic acid registers, such as non-back-stepped states. In all pre-translocated complexes, electron density for the 3'-terminal RNA nucleotide in position +1 was fragmented or absent, presumably since the phosphodiester bond between positions -1 and +1 was cleaved by the intrinsic 3'-exonuclease activity of Pol II under crystallization conditions (initially 50 mM magnesium ions, 10-20 days incubation time at 22ºC).

Further support of specific misincorporation opposite the CPD 5'-thymine. Consistent with misincorporation opposite the CPD 5'-thymine, RNA in complex A was extended by only two residues, instead of three, when only ATP was present (fig. S2A, fig 2A).

Attempts to visualize the CPD 5'-thymine-uridine mismatch crystallographically. We included a uridine at the RNA 3'-end of complex C (Scaffold C^{+U}, fig. S3), and re-solved the structure, but the resulting electron density for the uridine was fragmented (not shown). Soaking UTP into complex C crystals also resulted in poor density for the uridine, and did not result in translocation, as seen from an unaltered location of the bromine label in the DNA template (not shown). The modeling of the CPD thymine-uridine wobble base pair must be taken with care because of the limited resolution of the new data and of previous data with bound nucleotide substrates. In structure D the uracil is present at position +1 but could not be modeled due to fragmented electron density (see above).

A DNA-RNA hybrid is destabilized by a mismatch at the CPD. Consistent with the model that a CPD 5'-thymine-uridine mismatch destabilizes the DNA-RNA hybrid, the melting point of a 16 base-pair hybrid duplex with a CPD 5'-thymine-uridine mismatch is substantially lower than that of the same duplex with a 5'-thymine-adenine base pair (50.1°C versus 60.0°C). The complete melting point data are as follows. Undamaged DNA-RNA heteroduplex (5'-TT-3'): 3'-AA-5', 63.8°C; 3'-AU-5', 53.5°C; 3'-UA-5', 54.0°C. Damaged DNA-RNA heteroduplex (5'-CPD-3'): 3'-AA-5', 60.0°C, 3'-AU-5', 47.5°C; 3'-UA-5', 50.1°C. Thus a CPD alone only leads to a minor destablization of the DNA-RNA duplex, consistent with earlier data (*S16*). However, a T-U mismatch leads to substantial destabilization.

Pol II binds the hybrid in a defined register. In structures B and D, the bromine peak in the template strand is observed precisely at register -3, instead of -4 (Figs. 1A-B, S1C). This indicates that Pol II binds the hybrid in a defined, one base pair register, consistent with movement of Pol II in discrete one base pair steps (*S17*).

Additional support of lesion-bypass transcription. When an additional adenosine is included at the RNA 3'-end of scaffold C (scaffold C^{+A}, Fig. S3), the RNA was readily extended to the run-off transcript (Fig. S2C). As a control, scaffold C with an additional uridine at the RNA 3'-end was used (Scaffold C^{+U}, Fig. S3) and no extension of the RNA could be observed (fig. S2C).

TFIIH location, accessibility of excision sites, and DNA repositioning. TFIIH contains two subunits with helicase activity, XPB and XPD. XPB is required for transcription and repair and is located at the downstream part of the cleft during transcription initiation (S*18-20*). The observed repositioning of downstream DNA in the CPD-containing structures B and D shows that downstream DNA in the elongation complex can occupy different positions, and a change in the position of downstream DNA may be required for TFIIH action during TCR. There is evidence that the upstream incision site is rendered accessible by TFIIH action (S*21*) and this may be due to the second helicase of TFIIH, XPD.

TFIIS induces backtracking of the CPD-stalled complex and RNA cleavage in discrete dinucleotide steps. When complex C with an additional uridine at the RNA 3'-end (scaffold C^{+U}, fig. S3) was incubated with TFIIS, repeated dinucleotide cleavage from the RNA 3'-end was observed (fig. S2E). This indicates Pol II backtracking and movement of the CPD lesion out of the active site into the downstream direction. Full-length yeast TFIIS was purified as described (*S22*), except for an additional size exclusion chromatography step (Superose 6 10/300 GL) in transcription buffer (Materials and methods).

Comparison of CPD recognition by different polymerases. The mechanism of CPD recognition by Pol II is likely the same in Pol I and Pol III, since their active centers are conserved (S*5, S23*). However, DNA polymerases recognize and process a CPD differently. The DNA polymerase of bacteriophage T7 and the lesion-bypass polymerase Dpo4 incorporate adenosine opposite the CPD 5'-thymine (S24*, S25*), in contrast to Pol II, which incorporates uridine. In addition, dATP binds differently opposite the CPD 5'-thymine in these two DNA polymerases (*S24-S25*). Therefore, detailed structure-function studies of each type of DNA lesion in the context of each polymerase family are required to unravel the mechanisms of lesion recognition and processing.

Supporting online figures

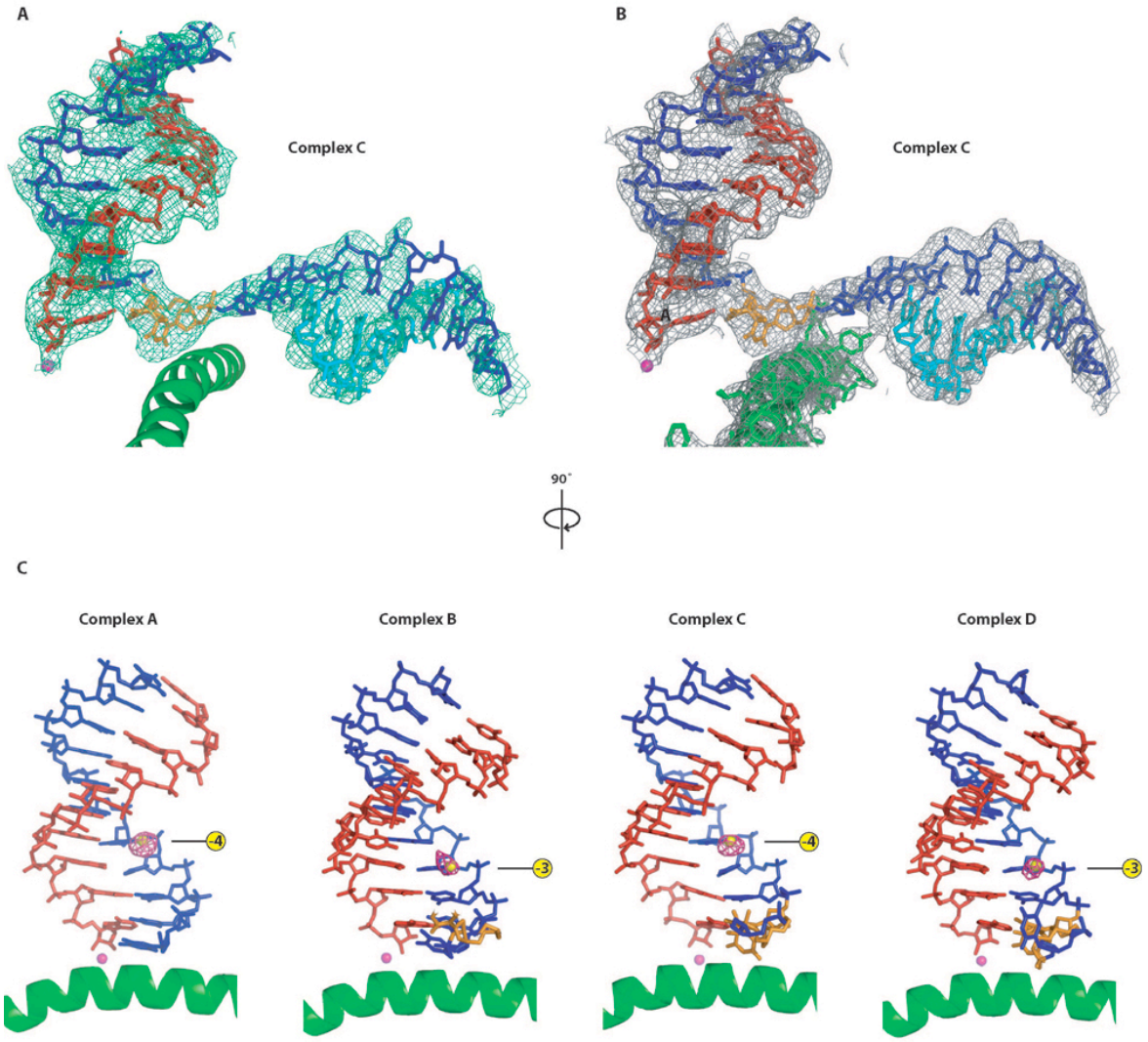

Fig. S1 Electron density maps.
Electron density maps were calculated from the datasets shown in Table S1. The final model of the nucleic acids, metal A and the bridge helix are superimposed (color code as in Fig. 1A).
(**A**) Unbiased F_o-F_c difference Fourier map (green) of complex C calculated after rigid body refinement with phases from the complete Pol II elongation complex without nucleic acids, contoured at 2.5 σ. The view is as in Fig. 1B (side view).
(**B**) $2F_o$-F_c map (grey) of complex C calculated with phases from the refined structure C and contoured around the nucleic acids and the bridge helix (side chains shown as sticks) at 1.0 σ.

(**C**) Bromine anomalous difference Fourier maps (pink) calculated with phases from the refined structures A, B, C, and D, contoured at 6.0 σ, 3.8 σ, 6.0 σ, and 3.8 σ, respectively. The observed peak heights are between 5.3 σ and 11.4 σ (Table S1), and coincide with the location of the bromine atom in a 5-bromo uracil residue incorporated in the DNA template strand, thereby defining the register of nucleic acids. When the anomalous difference Fourier maps in complexes B and D were contoured at the low level of 2.5 σ, there were no peaks at potential bromine positions in neighboring nucleotides. Thus there were no significantly populated alternative states with different nucleic acid registers, such as non-back-stepped states in complexes B and D. However we cannot exclude that a very minor portion of the complexes adopt a non-backstepped state. We can, however, exclude that the lower bromine peak heights in complexes B and D result from diffraction data collection at a wavelength different from the bromine anomalous peak since crystals were subjected to fluorescence scans and measured at the highly wavelength-stable SLS beamline PX1. With respect to (B), the models in (C) are rotated by 90° around a vertical axis.

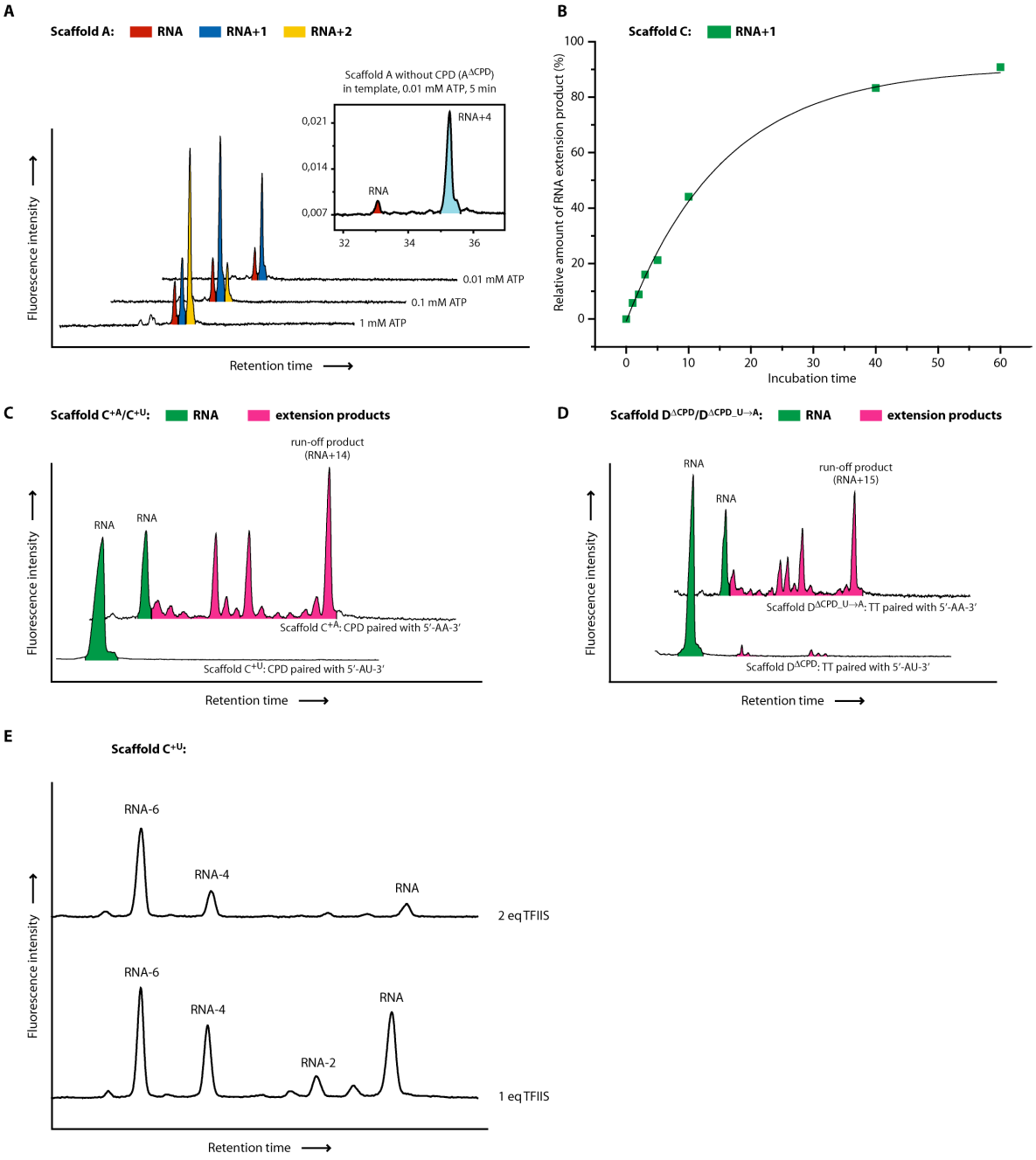

Fig. S2 Additional selected RNA extension assays.
(**A**) RNA extension at varying NTP concentration. Stoichiometric complexes of complete Pol II and scaffold A were incubated for 5 min with different concentrations of ATP. At 1 mM ATP, the translocation barrier is overcome, but only two extension steps are possible since adenine is not incorporated opposite the CPD 5'-thymine (Fig. 2A, 2C). At 0.01 mM ATP, the translocation barrier is not overcome and the RNA is extended by only one

nucleotide, whereas an undamaged template with the same sequence (scaffold $A^{\Delta CPD}$, Fig. S3) is readily extended at this ATP concentration and the same incubation time, providing a positive control (box).
(**B**) Kinetics of the final incorporation step. A stoichiometric complex of complete Pol II and scaffold C was incubated with 1mM of ATP, CTP, GTP and UTP. The relative amount of RNA extension product is plotted against the incubation time (compare Fig. 2B). This reaction corresponds to the third incorporation event in complex A (compare Fig. 2A, B, C).
(**C**) Lesion bypass transcription using an alternative nucleic acid scaffold. To exclude construct bias (mismatch at position -8 of scaffold D), lesion bypass was also examined using scaffold C with an additional adenine at the RNA 3'-end (5'-AA-3' opposite the CPD, scaffold C^{+A}, Fig. S3). Pol II could efficiently bypass the lesion in this scaffold after 20 min incubation with 1 mM of ATP, CTP, GTP and UTP. When scaffold C with an additional uridine was used (5'-AU-3' opposite the CPD, scaffold C^{+U}, Fig. S3), no bypass was observed after identical conditions.
(**D**) Highly inefficient T-U mismatch extension. To examine bypass of a T-U mismatch, the CPD in the template DNA of scaffold D was replaced by two thymines (Scaffold $D^{\Delta CPD}$). After 20 min incubation with 1 mM of ATP, CTP, GTP and UTP, a small fraction of elongation complexes had bypassed the mismatch (approx. 5 %). Without the mismatch (Scaffold $D^{\Delta CPD\_U\rightarrow A}$) efficient RNA extension to the run-off product was observed under identical conditions, providing a positive control.
(**E**) TFIIS-induced transcript cleavage in dinucleotide increments. Backtracking and transcript cleavage by TFIIS was analyzed using scaffold C with an additional uridine at the RNA 3'-end (scaffold C^{+U}, Fig. S3). The resulting elongation complex corresponds to a complex naturally stalled at a CPD lesion. Incubation time was 10 min, and 1 or 2 molar equivalents of TFIIS were used. Shortened RNA products are indicated (RNA-2, RNA-4, RNA-6, initial RNA shortened by 1, 2, and 3 dinucleotides, respectively).

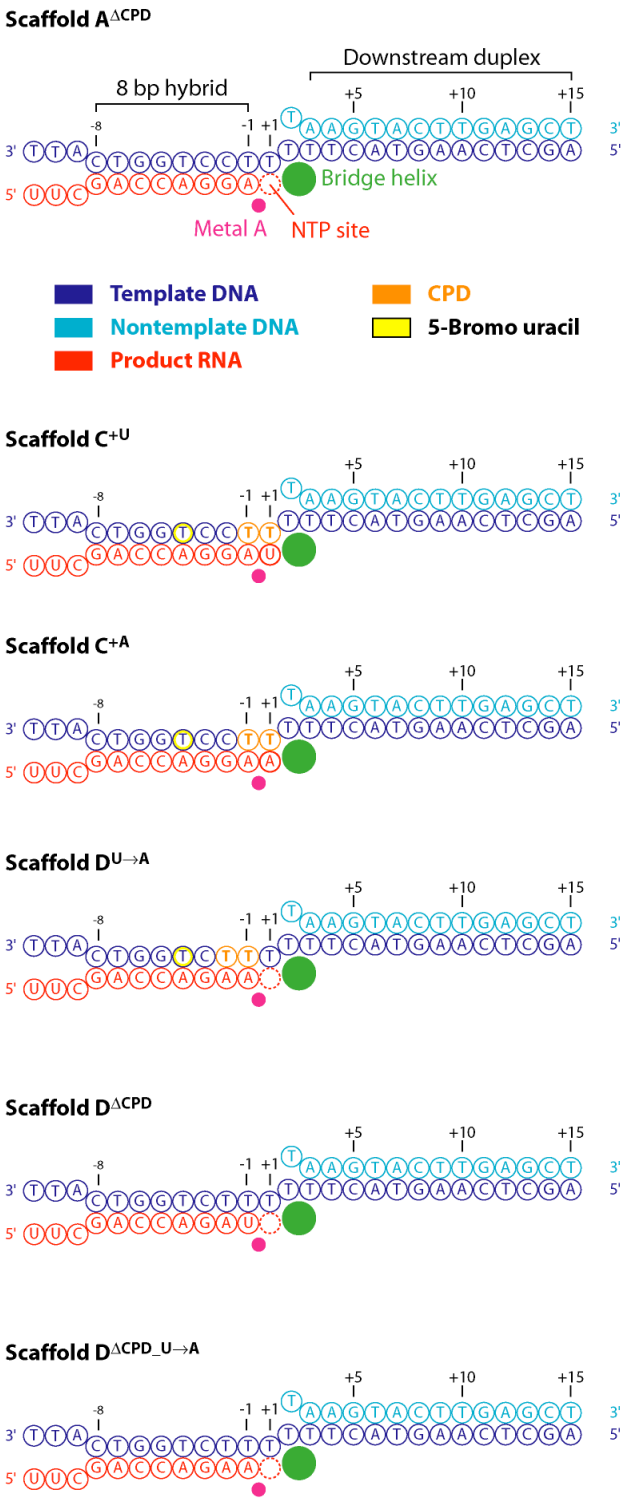

Fig. S3 Additional scaffolds used in RNA extension assays.

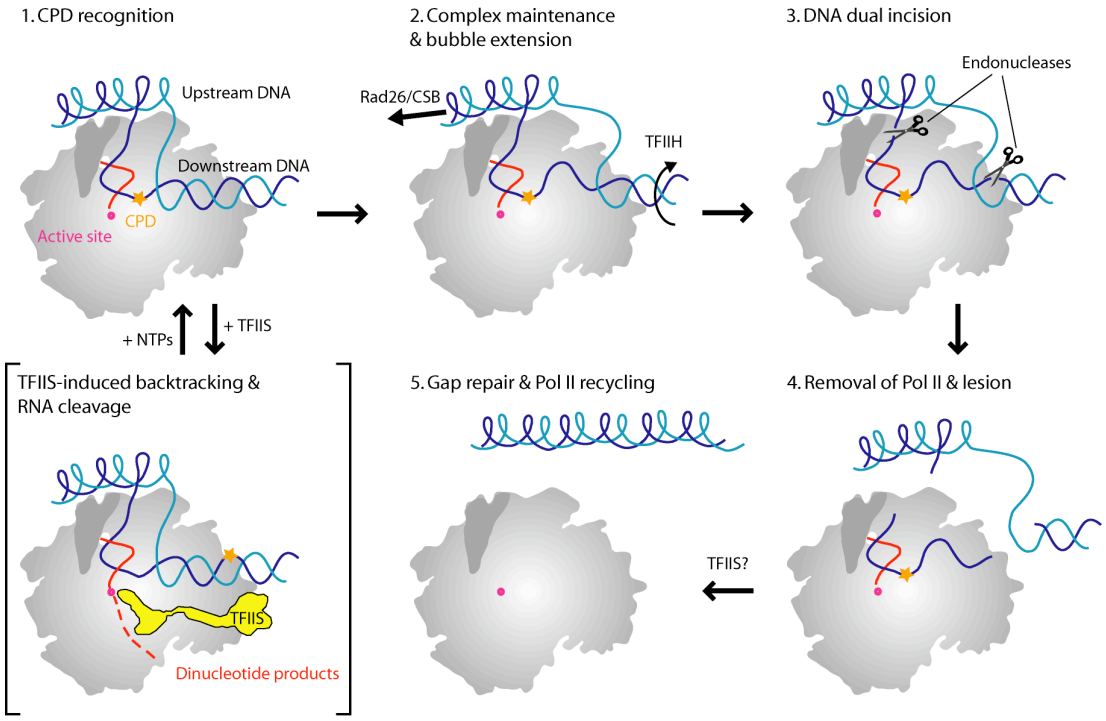

Fig. S4 Topological model for transcription-coupled DNA repair.
The lesion-containing Pol II elongation complex is shown schematically. The view and color code are as in Fig. 1C. An orange star and a magenta sphere indicate the CPD lesion and the active site, respectively. The upstream DNA was placed on Pol II on the basis of its location in the bacterial RNA polymerase-promoter complex (S9). The nontemplate strand in the bubble region was modeled. TFIIH helicase subunits XPD and XPB may be involved in extending the transcription bubble in the upstream and downstream direction, respectively, to enable dual incision. Modeling shows that incisions could occur at the two edges of an extended bubble around 10-15 nucleotides upstream (3') and around 15-20 nucleotides downstream (5') of the lesion.

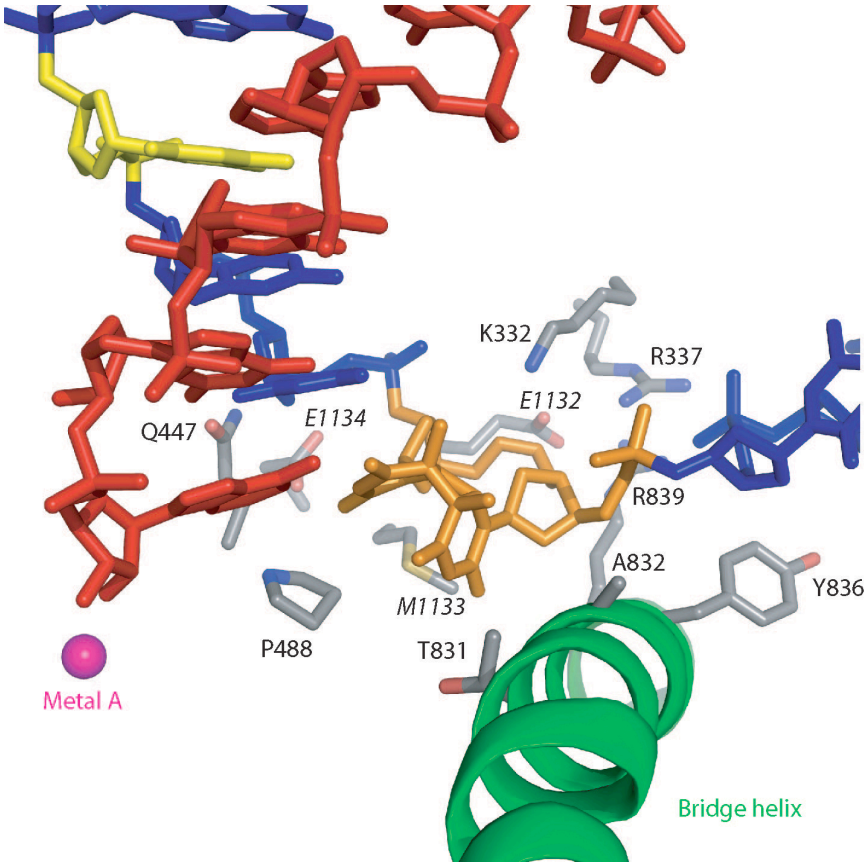

Fig. S5 Active site of Pol II in complex C. View as in fig. 1B. Color code as in fig 1A. Protein side chains within 5 Å of the CPD are shown as grey sticks. Italics are used to distinguish Rpb2 from Rpb1 residues.

Table S1 X-ray diffraction data and refinement statistics.

CPD-containing elongation complex	A	B	C	D
Data collection[1]				
Space group	$C222_1$	$C222_1$	C2	$C222_1$
Unit cell axes (Å)	222.2, 392.2, 284.0	223.6, 393.5, 283.5	394.4, 221.9, 283.1	222.6, 392.8, 282.9
Unit cell β angle (°)	90	90	90.6	90
Wavelength (Å)	0.9191	0.9190	0.9192	0.9189
Resolution range (Å)	50-3.8 (3.94-3.80)[2]	50-4.0 (4.14-4.00)	50-3.8 (3.94-3.80)	50-3.80 (3.94-3.80)
Unique reflections	120,477 (11,664)	104,987 (10,232)	238,284 (23,709)	121,341 (11,999)
Completeness (%)	99.0 (96.8)	99.8 (98.2)	99.6 (99.6)	99.9 (99.7)
Redundancy	4.6 (3.4)	5.8 (4.5)	3.3 (3.2)	5.1 (4.0)
Mosaicity (°)	0.50	0.50	0.47	0.34
R_{sym} (%)	12.0 (45.4)	8.7 (41.1)	9.8 (43.1)	8.3 (37.3)
I/σ(I)	10.3 (2.1)	15.8 (3.1)	10.2 (2.3)	17.1 (3.2)
Refinement				
Nonhydrogen atoms	31,665	32,016	63,934	32,005
RMSD bonds	0.010	0.009	0.010	0.009
RMSD angles	1.55	1.52	1.60	1.55
R_{cryst} (%)	27.0	29.2	25.7	26.9
R_{free} (%)	28.0	30.1	27.5	28.8
Br peak in anomalous Fourier (σ)	11.4	5.3	10.7, 10.6	6.4
PDB accession code	2ja5	2ja6	2ja7	2ja8

[1]All diffraction data were collected at the Swiss Light Source, Paul-Scherrer-Institut, Villigen, Switzerland.
[2]Numbers in parenthesis correspond to the highest resolution shell.

References

S1. K.-J. Armache, H. Kettenberger, P. Cramer, *Proc. Natl. Ac. Sc. USA* **100**, 6964 (2003).
S2. J. Butenandt, A. P. M. Eker, T. Carell, *Chemistry-a European Journal*, **642** (1998).
S3. Z. Otwinowski, W. Minor, *Meth. Enzym.* **276**, 307 (1996).
S4. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr D Biol Crystallogr* **61**, 458 (2005).
S5. H. Kettenberger, K.-J. Armache, P. Cramer, *Mol. Cell* **16**, 955 (2004).
S6. A. T. Brunger *et al.*, *Acta Crystallogr D Biol Crystallogr* **54**, 905 (1998).
S7. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Cryst.* **A47**, 110 (1991).
S8. K.-J. Armache, S. Mitterweger, A. Meinhart, P. Cramer, *J. Biol. Chem.* **280**, 7131 (2005).
S9. K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, *Science* **296**, 1285 (2002).
S10. J. Butenandt *et al.*, *Chem. Eur. J.* **6**, 62 (2000).
S11. J. Cadet, L. Voituriez, F. E. Hruska, A. Grand. *Biopolymers* **24**, 897-903 (1985).

S12. J. Butenandt, L.T. Burgdorf, T. Carell. *Angew. Chem. Int. Ed.* **38**, 708-711 (1999), J.S. Taylor, D.S. Garrett, I.R. Brockie, D.L. Svoboda, J. Telser. *Biochemistry* **29**, 8858-8866 (1990).
S13. X. Gao, *et al. Biochemistry* **31**, 6228-6236 (1992), J.M. Veal *et al.*, *J. Am. Chem. Soc.* **115**, 7139-7145 (1993).
S14. A. Mees *et al.*, *Science* **306**, 1789 (2004).
S15. A. M. Edwards, C. M. Kane, R. A. Young, R. D. Kornberg, *J. Biol. Chem.* **266**, 71 (1991).
S16. J. Butenandt, L.T. Burgdorf, T. Carell. *Angew. Chem. Int. Ed.* **38**, 708-711 (1999).
S17. E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, S. M. Block, *Nature* **438**, 460 (2005).
S18. D. Tantin, *J. Biol. Chem.* **273**, 27794 (1998).
S19. T. K. Kim, R. H. Ebright, D. Reinberg, *Science* **288**, 1418 (May 26, 2000).
S20. G. Miller, S. Hahn, *Nat. Struct. Mol. Biol.* **13**, 603 (2006).
S21. A. H. Sarker *et al.*, *Mol. Cell* **20**, 187 (2005).
S22. H. Kettenberger, K.-J. Armache, P. Cramer, *Cell* **114**, 347 (2003).
S23. A. J. Jasiak, K. J. Armache, B. Martens, R. P. Jansen, P. Cramer, *Mol. Cell* **23**, 71 (2006).
S24. H. Ling, F. Boudsocq, B. S. Plosky, R. Woodgate, W. Yang, *Nature* **424**, 1083 (2003).
S25. Y. Li *et al.*, *Nat. Struct. Mol. Biol.* **11**, 784 (2004).