nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-019-0255-5

Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link

Petria S. Thompson[1,3], Katherine M. Amidon[2,3], Kareem N. Mohni[1], David Cortez[1*] and Brandt F. Eichman[1,2*]

Abasic (AP) sites are one of the most common DNA lesions that block replicative polymerases. 5-hydroxymethylcytosine binding, embryonic stem cell-specific protein (HMCES) recognizes and processes these lesions in the context of single-stranded DNA (ssDNA). A HMCES DNA-protein cross-link (DPC) intermediate is thought to shield the AP site from endonucleases and error-prone polymerases. The highly evolutionarily conserved SOS-response associated peptidase (SRAP) domain of HMCES and its *Escherichia coli* ortholog YedK mediate lesion recognition. Here we uncover the basis of AP site protection by SRAP domains from a crystal structure of the YedK DPC. YedK forms a stable thiazolidine linkage between a ring-opened AP site and the α-amino and sulfhydryl substituents of its amino-terminal cysteine residue. The thiazolidine linkage explains the remarkable stability of the HMCES DPC, its resistance to strand cleavage and the proteolysis requirement for resolution. Furthermore, its structure reveals that HMCES has specificity for AP sites in ssDNA at junctions found when replicative polymerases encounter the AP lesion.

Apurinic and apyrimidinic (abasic or AP) site repair via base excision repair (BER) depends on an intact DNA duplex[1–3]. While most AP sites form in double-stranded DNA (dsDNA), base loss is actually more rapid in ssDNA[4]. Furthermore, the action of the DNA helicase in replicating cells will convert dsDNA AP sites that have not been repaired into ssDNA AP sites. In this case, the replicative polymerases will stall at the AP site leaving a 3′ dsDNA-ssDNA junction. Until recently, the major mechanism to overcome this replication challenge was thought to be translesion synthesis by error-prone polymerases, including DNA polymerase ζ[5]. However, we recently discovered an alternative pathway dependent on the SRAP domain protein HMCES that improves cell viability and reduces mutation frequency[6].

SRAP proteins are conserved in organisms from bacteria to humans, and in bacteria SRAP-encoding genes are often spatially linked to DNA repair genes[7]. Human HMCES and *E. coli* YedK are similar in both sequence (29% identity and 43% similarity) and structure (C_α r.m.s. deviations of 1.29 Å between PDB entries 5KO9 and 2ICU). Both HMCES and YedK preferentially bind ssDNA and efficiently form DPCs to AP sites in ssDNA[6]. DPC formation requires conserved DNA-binding residues and an invariant cysteine that is almost always encoded as the second amino acid in SRAP proteins. The HMCES DPC is also formed in cells, increases in abundance in response to AP site-inducing agents and is resolved over time by a mechanism that is at least partially proteasome-dependent[6]. Despite the importance of the HMCES AP site DPC to this mechanism, the chemical nature of the cross-link and how the SRAP domain detects the AP site are unknown.

To better understand this unusual mechanism of DNA repair, we examined the nature of the HMCES-DNA interaction. Our results indicate that SRAP proteins cross-link to AP sites via a stable thiazolidine DNA-protein linkage formed with the N-terminal cysteine and the aldehyde form of the AP deoxyribose. This linkage and its solvent inaccessibility explain why the cross-link shields the AP site from endonucleases and likely necessitates a proteolysis-dependent mechanism for resolution. Furthermore, the structure of the SRAP DPC explains the ssDNA specificity, but suggests HMCES could accommodate a dsDNA-ssDNA 3′ junction, as might be expected when a replicative polymerase stalls at the AP site. As predicted, we show that HMCES has a preference for exactly this type of DNA structure.

Results

The SRAP domains of both human HMCES and *E. coli* YedK form covalent linkages to AP sites in ssDNA, but the nature of the DPC is unknown. The ease of detecting a HMCES DPC in cells suggests it may be a stable chemical linkage[6]. Indeed, incubating the human HMCES SRAP domain DPC at 4, 25 or 37 °C for up to six days did not change the percentage of cross-linked protein (Fig. 1a). We noticed while doing these experiments that boiling the DPC hydrolyzed the cross-link, but that incubation at 50 °C did not (Fig. 1b). Protein denaturation is not sufficient for hydrolysis since the DPC amount does not change over time when it is incubated at room temperature after denaturing the protein by boiling for a short time (Fig. 1c). Furthermore, extensive proteolysis of the DPC with proteinase K left a small peptide-DNA linkage that remains stable (Fig. 1d) and resistant to cleavage by AP endonuclease 1 (APE1) (Fig. 1e). Thus, the HMCES AP DPC is unlikely to be reversible in physiological conditions and resolution almost certainly requires proteolysis followed by either an unidentified enzymatic action to remove the linkage or nucleotide excision repair.

To understand the molecular basis for the stability of the SRAP DPC, we determined a 1.6-Å resolution crystal structure of *E. coli* YedK covalently cross-linked to a heptamer ssDNA containing an AP site (Table 1). The entire DNA ligand is visible in the electron density (Fig. 2a). The protein does not undergo any appreciable conformational change on binding DNA, with an r.m.s.d. of 1.16 Å for all atoms between unbound and DPC forms of YedK

[1]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN, USA. [2]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA. [3]These authors contributed equally: Petria S. Thompson, Katherine M. Amidon. *e-mail: david.cortez@vanderbilt.edu; brandt.eichman@vanderbilt.edu

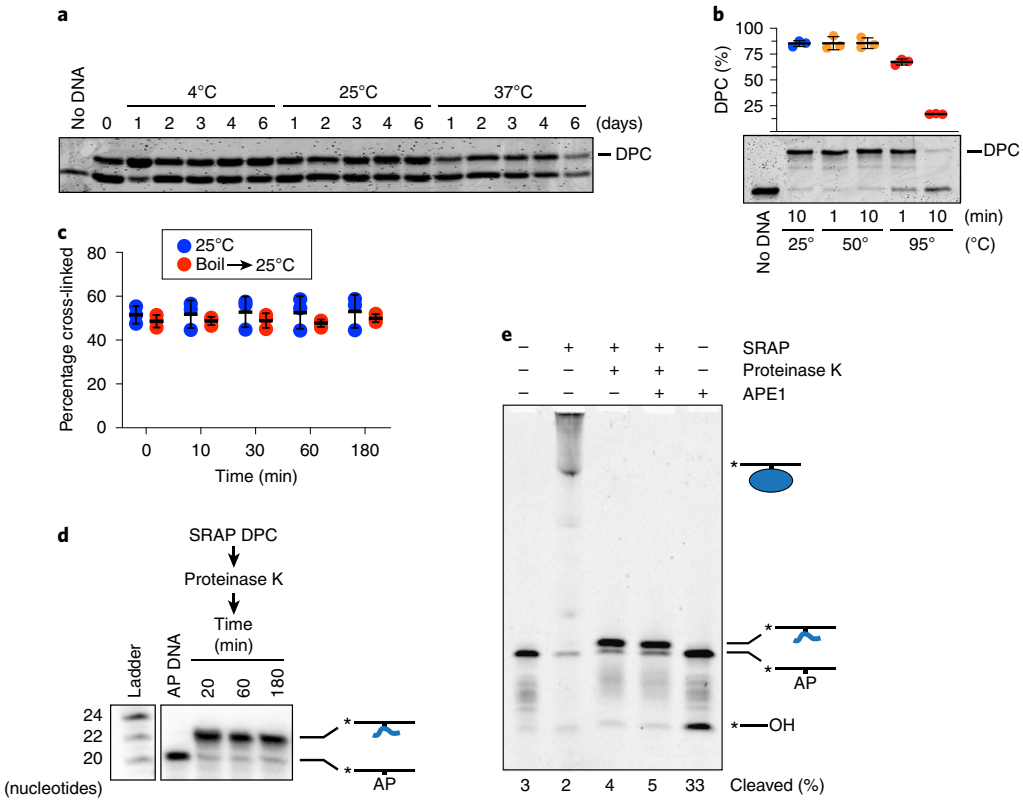

Fig. 1 | Stability analysis of the human HMCES SRAP-abasic site DNA-protein cross-link. a, HMCES SRAP DPC stability measured at the indicated temperatures. Free and DNA-cross-linked HMCES were detected by Coomassie blue staining. The HMCES DPC percentage in this experiment is approximately 50% because uncross-linked DNA was removed by dialysis after a short reaction time. **b**, Boiling the HMCES DPC causes hydrolysis (mean ± s.d., $n = 3$ independent measurements). **c**, HMCES DPC stability measured before or after denaturation by boiling for 2 min. **d**, HMCES SRAP domain was incubated with a 20-mer AP site containing oligonucleotide to form a cross-link, digested with proteinase K followed by heat inactivation of the protease and then incubated at 37 °C for the times indicated. Electrophoresis and autoradiography were used to visualize the DNA. **e**, HMCES SRAP was incubated with a 31-mer AP DNA and digested with proteinase K, and the peptide DPC was incubated with APE1 for 2 h. Bands were visualized by Cy5 fluorescence. Uncropped gel images are shown in Supplementary Dataset 1. Source data for **b** and **c** are available online.

(Supplementary Fig. 1d). The core β-sheet forms an extended, positively charged channel that cradles the ssDNA phosphoribosyl backbone along one face of the protein (Fig. 2b–d). The conformation of the DNA is further constrained by nucleobase π-stacking and van der Waals interactions from random coil and α-helical motifs at each end of the binding channel that were disordered in the unbound structure (Supplementary Fig. 1d). The hydrogen-bonding edges of every nucleobase are exposed to solvent, and thus recognition of the DNA would not be sequence-dependent. Most strikingly, the DNA backbone is severely kinked and twisted by 90° at the AP site, placing the nucleobases of each flanking trinucleotide orthogonal to one another (Fig. 2b). This sharp distortion precludes pairing of a complementary DNA strand in the vicinity of the AP site, and explains why SRAP disfavors binding to dsDNA[6]. The residues lining the DNA-binding channel are the most highly conserved among SRAP domains (Fig. 2d,e and Supplementary Fig. 1), suggesting conservation of DNA-binding modality. Indeed, both YedK and HMCES have similar preferences to bind ssDNA and mutation of conserved amino acids in the channel abrogates DNA binding for both proteins[6].

The AP site is positioned directly above Cys 2, previously implicated in SRAP DPC formation[6]. This cysteine is at the N terminus of the protein since the methionine is likely removed by aminopeptidases. The electron density clearly shows the AP site in the ring-opened form, with continuous density between C3′ and the Cys 2 side chain (Fig. 3a). The anomeric C1′ carbon of the AP site is covalently bonded to both the α-amino nitrogen and the side chain sulfur of Cys 2 to form a thiazolidine ring (Fig. 3a). Such a linkage would be generated by nucleophilic attack of the AP aldehyde C1′ carbon by Cys 2 α-NH_2 to form a Schiff base intermediate, followed by subsequent attack of C1′ by the Cys 2 sulfhydryl group (Fig. 3b)[8]. Consistent with cross-linking by Cys 2, YedK DPC formation is abrogated by removal of the thiol in a C2A mutant[6], and by a C2S mutant, which potentially forms an oxazolidine ring that would not be as stable as a thiazolidine (Fig. 3c,d)[9,10].

Studies on the reaction of cysteine and aldehydes show that the equilibrium between Schiff base and thiazolidine greatly favors the latter[11,12], explaining why we do not see any evidence for DNA lyase activity that can result from β-elimination of the Schiff base intermediate, such as is found in bifunctional DNA glycosylases that initiate BER (Fig. 3b)[13–15]. In contrast to the wild-type protein, both the C2A and C2S mutant exhibited DNA lyase activity when incubated with ssDNA containing an AP site (Fig. 3c). This lyase activity was significantly reduced by performing the cross-linking reaction in the presence of sodium cyanoborohydride ($NaBH_3CN$), which acts as a reducing agent to stabilize the Schiff base intermediate (Fig. 3e)[16]. These results further support a reaction mechanism that includes capture of the Schiff base intermediate by nucleophilic attack of the cysteine thiol and explains why this residue is invariant in all SRAP proteins.

Cys 2 belongs to a cluster of three conserved residues that includes Glu 105 and His 160 and is implicated in SRAP function[7,17].

Table 1 | Data collection and refinement statistics

	YedK/AP-DNA covalent DPC (PDB 6NUA)	YedK/C3-spacer-DNA non-covalent complex (PDB 6NUH)
Data collection[a]		
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	61.26, 41.89, 81.42	47.54, 44.13, 55.09
α, β, γ (°)	90.00, 95.79, 90.00	90.00, 102.34, 90.00
Resolution (Å)	50.00–1.64 (1.67–1.64)[b]	100.00–1.60 (1.66–1.60)
R_{sym}	0.098 (0.500)	0.075 (0.397)
R_{meas}	0.110 (0.595)	0.086 (0.455)
$I/\sigma(I)$	14.8 (1.9)	21.3 (2.6)
$CC_{1/2}$	0.989 (0.823)	0.990 (0.869)
Completeness (%)	97.4 (95.2)	98.5 (91.1)
Redundancy	4.4 (2.9)	4.1 (4.0)
Refinement		
Resolution (Å)	40.50–1.64 (1.67–1.64)	39.60–1.59 (1.65–1.59)
No. reflections	49,681 (2,331)	29,612 (2,391)
R_{work} / R_{free}	0.171 / 0.222	0.143 / 0.177
No. atoms		
Protein	3,627	1,802
DNA	268	131
Bis-Tris	-	14
Water	280	210
B factors		
Protein	26.0	17.1
DNA	28.7	64.3
Bis-Tris	-	38.3
Water	29.3	24.4
R.m.s. deviations		
Bond lengths (Å)	0.010	0.008
Bond angles (°)	1.035	0.959

[a]Data for each structure were generated from a single crystal. [b]Values in parentheses are for highest-resolution shell.

These and several other evolutionarily conserved residues stabilize the DNA and protein sides of the thiazolidine linkage (Fig. 3f,g and Supplementary Fig. 1). The AP site is stabilized by His 160, which forms a hydrogen bond with the O4′ hydroxyl group (Fig. 3f). Similarly, Arg 77 and Arg 162, previously shown to be essential for DNA binding[6], and Thr 149, interact with the AP site 5′-phosphate (Fig. 3f,g). The Glu 105 side chain fluctuates between two conformations at the cross-link (Fig. 3f and Supplementary Fig. 2). One conformer places one carboxylate oxygen 3.5 Å from the thiazolidine C1′ and the second within hydrogen-bonding distance to the phosphate 3′ to the AP site, strongly implying that the carboxylate is protonated to avoid electrostatic repulsion with the DNA. The second conformer points back toward the core of the protein and sits further away from the thiazolidine ring. On the protein side of the cross-link, the carboxamide side chain from a highly conserved asparagine (Asn 75) helps position the cross-linking nucleophile by forming two hydrogen bonds with the backbone amide nitrogen and carbonyl oxygen of Cys 2 (Fig. 3f). Consistent with their roles in stabilizing the cross-link, individual substitutions of Glu 105, His 160 or Asn 75 with alanine reduced cross-linking efficiency (Fig. 3c,d). In addition to the direct contacts to the DNA and the thiazolidine linkage, there are several highly conserved residues that create a hydrophobic pocket to cradle Cys 2 from underneath (Fig. 3g and Supplementary Fig. 1). Thus, the SRAP structure guides the AP site into a specific, solvent inaccessible environment suited for thiazolidine formation and protected from AP endonuclease cleavage.

We also determined a crystal structure of YedK bound non-covalently to a ssDNA oligomer containing a C3-spacer in place of the AP site (Supplementary Fig. 3). The protein in the non-covalent complex is virtually identical to that of the DPC, except for modest repositioning of a β-hairpin (β7–β8) that was disordered in the unbound YedK structure (PDB: 2ICU) and that stabilizes the backbone of the DNA 3′ to the AP site in the DPC (Supplementary Fig. 3a). In the non-covalent complex, the DNA at the 5′ end is positioned as in the DPC structure. However, the 3′ end of the DNA in the non-covalent complex is more mobile, as evidenced by weaker electron density and higher *B* factors for the 3′ nucleotides and including the C3-spacer (Supplementary Fig. 3b–d). The destabilized 3′ DNA end resulted in a crystal packing difference between the two complexes.

Both the DPC and non-covalent complex structures suggest that the SRAP domain can accommodate dsDNA on the 3′ side of the AP site, but would disfavor duplex formation on the 5′ side. The DNA backbone on the 5′ side of the AP site is kinked 90° by a wedge motif (residues 65–73 and 84–87), which stacks against the second and third nucleotides (G1 and T2) from the AP site (Fig. 4a,b). Trp 68 wedges the nucleobases of G1 and T2 apart, and G1 is stacked between Trp 67 and Arg 85 (Fig. 4b). Such a distortion would prevent duplex formation with DNA 5′ to the AP site. The importance of the wedge motif to HMCES function is underscored by the strong conservation of these residues among SRAP domains (Supplementary Fig. 1).

In contrast to the distorted 5′ side of the DPC, all three nucleobases on the 3′ side of the AP site are stacked in a B-DNA conformation (Fig. 4b). The residue adjacent to the AP site (guanine, G5) stacks against Pro 40 and Ile 74 on the surface of the protein (Fig. 4b,c). The exposure of the hydrogen-bonding faces of the G5, G6 and A7 nucleobases 3′ to the AP site would allow for base pairing of a second strand up to the 3′ side of the AP site. Modeling shows that a complementary strand fits against the protein surface with no steric clashes (Fig. 4a–c). The 3′ end of the modeled strand stacks against Gly 41 and Thr 42, which together with Pro 40 and Ile 74 form a highly conserved 'shelf' that would stabilize a base pair 3′ to the AP site (Fig. 4c and Supplementary Fig. 1). Conservation of this shelf region implies that binding to AP sites in the context of a 3′-truncated ssDNA-dsDNA junction is an important feature. This is the exact context in which SRAP proteins should operate at a stalled replication fork, since DNA polymerase stalling at an AP site leaves a 3′-truncated nascent strand with a 5′-overhanging template. Consistent with this prediction we found that HMCES is just as efficient at binding and cross-linking to an AP site immediately adjacent to the 3′ ssDNA-dsDNA junction as to ssDNA (Fig. 4d,e and Supplementary Fig. 4). In contrast, binding and cross-linking is less efficient when the dsDNA is present on the 5′ side of the AP site, consistent with the effect of the wedge motif.

Discussion

The YedK-AP-DNA cross-link structure reveals how the unique DNA-binding surface and N-terminal cysteine facilitate recognition and covalent cross-linking of HMCES and SRAP-containing proteins to AP sites in the context of ssDNA. Furthermore, the results explain the stability of this cross-link and the substrate preferences that correspond to DNA structures formed when polymerases stall at abasic sites.

The thiazolidine linkage acts as a sink for abasic sites and prevents strand breaks resulting from (1) non-enzymatic β-elimination at C2′, (2) lyase activity from enzyme-catalyzed β-elimination of

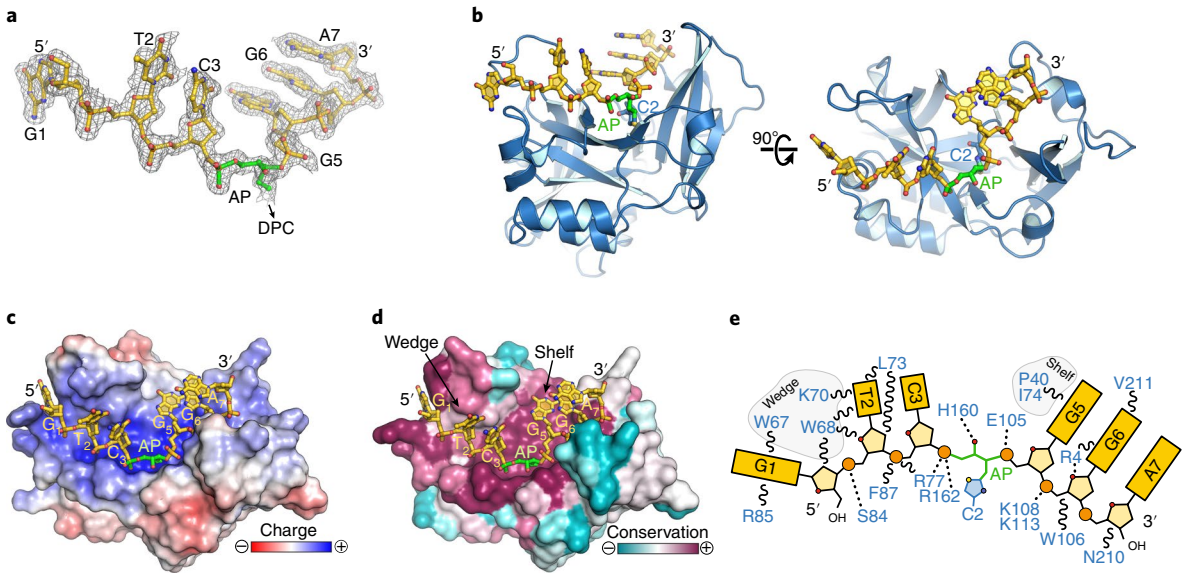

Fig. 2 | YedK DPC crystal structure. a, DNA fit to $2F_o - F_c$ composite annealed omit electron density, contoured at 1σ. **b**, Orthogonal views of *E. coli* YedK (blue) cross-linked to AP DNA (gold). **c**,**d**, YedK solvent-accessible surface colored by electrostatic potential from −5 to +5 k_BT/e_C (**c**) and sequence conservation from 158 unique SRAP orthologs (**d**). **e**, Schematic of protein-DNA interactions.

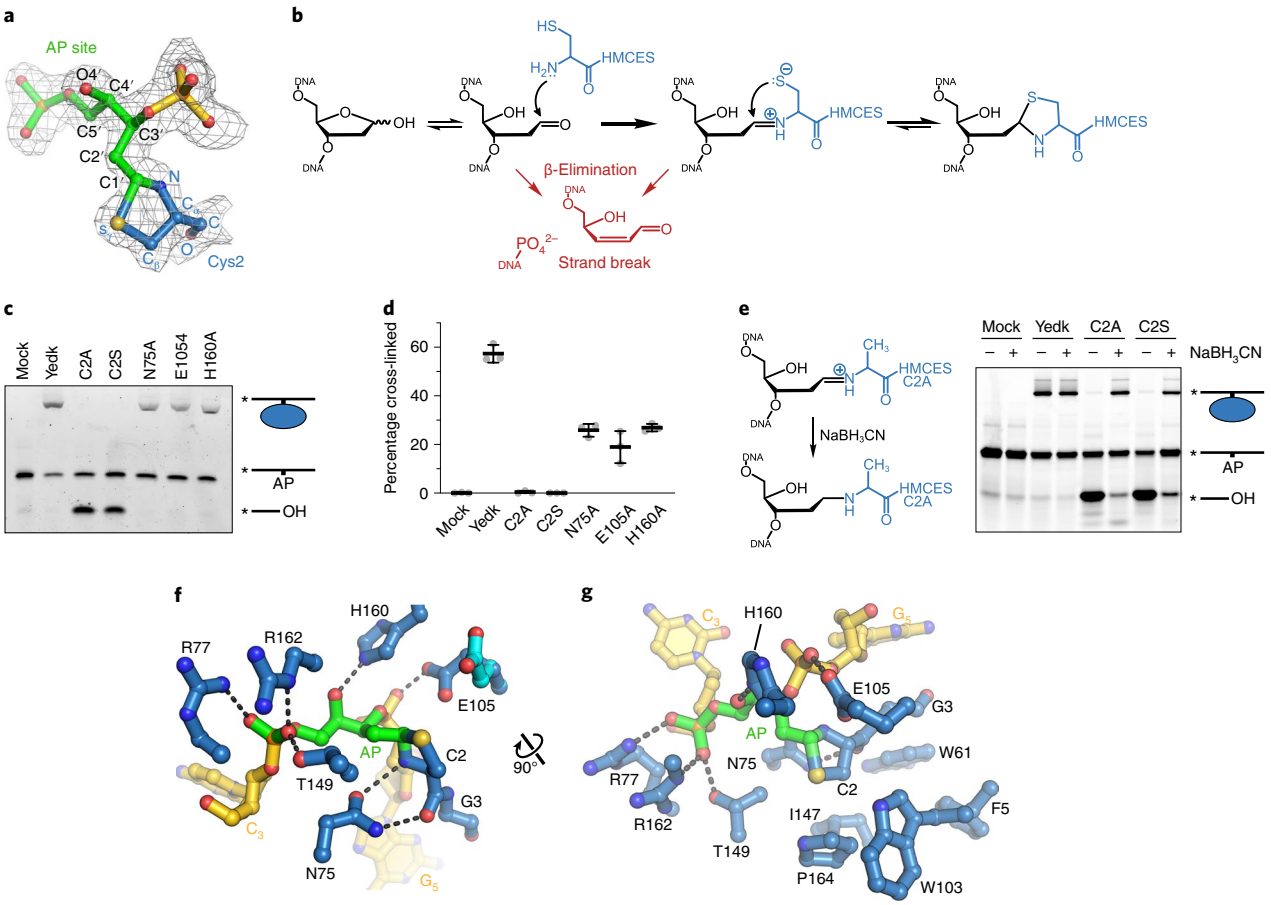

Fig. 3 | The SRAP DPC forms a thiazolidine linkage stabilized by conserved residues. a, The DPC between the AP site (green) and Cys 2 (blue) superimposed against $2F_o - F_c$ composite annealed omit electron density, contoured at 1σ. **b**, Proposed chemical mechanism of the cross-linking reaction with competing lyase reactions in red. **c**, Representative denaturing PAGE gel showing cross-linking and lyase activity of YedK mutants. Bands were visualized by FAM fluorescence. **d**, Cross-linking efficiencies of YedK mutants (mean ± s.d., $n = 3$ independent measurements). **e**, $NaBH_3CN$ was added to cross-linking reactions to trap the Schiff base intermediates of YedK C2A and C2S mutants. The $NaBH_3CN$-reduced Schiff base is refractory to β-elimination. Bands were visualized by FAM fluorescence. **f**, Residues contacting the DPC (DNA, gold; AP site, green; protein, blue). The alternate Glu 105 conformer is cyan. Dashed lines denote hydrogen bonds. **g**, Orthogonal view showing hydrophobic residues cradling Cys 2. The second Glu 105 conformer is not shown for clarity. Uncropped gel images are shown in Supplementary Dataset 1. Source data for **d** are available online.

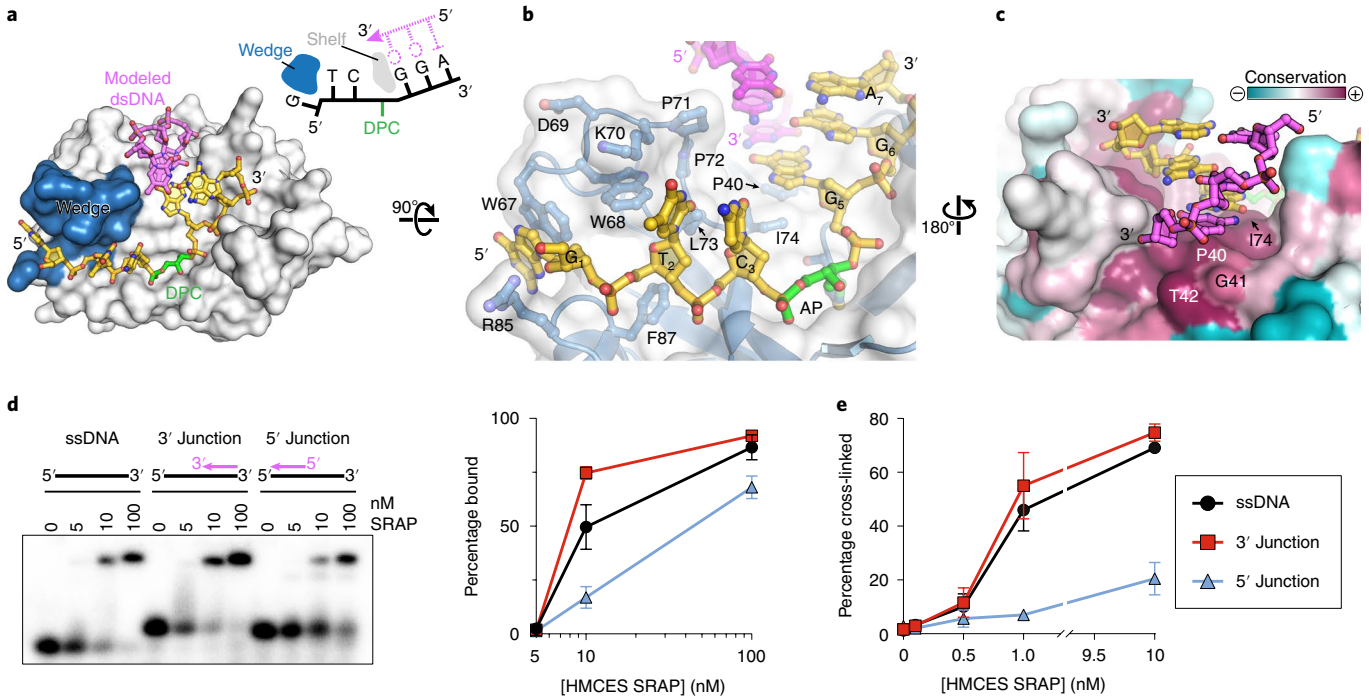

Fig. 4 | SRAP can accommodate dsDNA 3′ to the AP site. a, Model of YedK DPC with a 3′ junction at the AP site. The modeled complementary DNA strand is pink. The wedge domain blocking dsDNA access 5′ to the AP site is blue. **b**, Wedge-DNA interactions 5′ to the AP site. **c**, Sequence conservation of the DNA shelf that presumably stabilizes dsDNA 3′ to the AP site. **d**, Electrophoretic mobility shift assay (EMSA) showing binding of human HMCES SRAP domain to the indicated DNA ligands. The plot shows mean ± s.e.m. from $n = 3$ independent measurements. Uncropped gel image is shown in Supplementary Dataset 1. **e**, Percentage of the indicated DNA substrates cross-linked to human HMCES SRAP domain (mean ± s.d., $n = 3$ independent measurements). Source data for **d** and **e** are available online.

the Schiff base or (3) APE1 incision. This contrasts with unstable, transient protein-DNA Schiff base cross-links that rapidly proceed to β-elimination as part of enzymatic strand cleavage reactions catalyzed by bifunctional glycosylases and DNA polymerase β as part of the BER pathway[13–15]. Other proteins, including PARP-1, Histone H4 and Ribosomal protein uS3 can cross-link to AP sites, but in each case the DPC leads to strand scission[18–21].

HMCES is named 5-hydroxymethylcytosine (5hmC) binding, embryonic stem cell-specific protein because it was identified in a proteomics experiment using duplex DNA containing multiple 5hmC residues as a bait to purify proteins from embryonic cell lysates[22]. Furthermore, the HMCES SRAP domain was shown to autoproteolyze itself and incise duplex DNA containing 5hmC[17]. The DNA-bound SRAP structure suggests SRAP is unlikely to recognize 5hmC in the context of duplex DNA and we have not observed either the proteolysis or duplex DNA incision activity reported.

A single SRAP domain protein exists in organisms in all three domains of life, indicating a critical function even though knockouts in human, yeast and bacterial cells are viable. The stability of the SRAP-AP-DNA cross-link and unique thiazolidine DPC linkage supports the conclusion that these proteins act to maintain genome stability during DNA replication and thereby improve organism fitness.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41594-019-0255-5.

Received: 8 February 2019; Accepted: 17 May 2019;
Published online: 24 June 2019

References

1. Krokan, H. E. & Bjoras, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).
2. Hitomi, K., Iwai, S. & Tainer, J. A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst.)* **6**, 410–428 (2007).
3. Tsutakawa, S. E., Lafrance-Vanasse, J. & Tainer, J. A. The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst.)* **19**, 95–107 (2014).
4. Kavli, B. et al. Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J.* **15**, 3442–3447 (1996).
5. Schaaper, R. M., Kunkel, T. A. & Loeb, L. A. Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proc. Natl Acad. Sci. USA* **80**, 487–491 (1983).
6. Mohni, K. N. et al. HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. *Cell* **176**, 144–153 e13 (2019).
7. Aravind, L., Anand, S. & Iyer, L. M. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol. Direct* **8**, 20 (2013).
8. Kallen, R. G. Mechanism of reactions involving Schiff base intermediates. Thiazolidine formation from L-cysteine and formaldehyde. *J. Am. Chem. Soc.* **93**, 6236–6248 (1971).
9. Canle, M., Lawley, A., McManus, E. C. & O'Ferrall, R. A. M. Rate and equilibrium constants for oxazolidine and thiazolidine ring-opening reactions. *Pure Appl. Chem.* **68**, 813 (1996).
10. Just, G., Chung, B. Y., Kim, S., Rosebery, G. & Rossy, P. Reactions of oxygen and sulphur anions with oxazolidine and thiazolidine derivatives of 2-mesyloxymethylglyceraldehyde acetonide. *Can. J. Chem.* **54**, 2089–2093 (1976).
11. Ratner, S. & Clarke, H. T. The action of formaldehyde upon cysteine. *J. Am. Chem. Soc.* **59**, 200–206 (1937).
12. Fife, T. H., Natarajan, R., Shen, C. C. & Bembi, R. Mechanism of thiazolidine hydrolysis. Ring opening and hydrolysis of 1,3-thiazolidine derivatives of *p*-(dimethylamino)cinnamaldehyde. *J. Am. Chem. Soc.* **113**, 3071–3079 (1991).
13. Brooks, S. C., Adhikary, S., Rubinson, E. H. & Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica Biophys. Acta Proteins Proteom.* **1834**, 247–271 (2013).

14. Fromme, J. C. & Verdine, G. L. Base excision repair. *Adv. Protein Chem.* **69**, 1–41 (2004).
15. Beard, W. A. & Wilson, S. H. Structure and mechanism of DNA polymerase β. *Chem. Rev.* **106**, 361–382 (2006).
16. Billman, J. H. & Diesing, A. C. Reduction of Schiff bases with sodium borohydride. *J. Org. Chem.* **22**, 1068–1070 (1957).
17. Kweon, S. M. et al. Erasure of Tet-oxidized 5-methylcytosine by a SRAP nuclease. *Cell Rep.* **21**, 482–494 (2017).
18. Prasad, R., Horton, J. K., Dai, D. P. & Wilson, S. H. Repair pathway for PARP-1 DNA-protein crosslinks. *DNA Repair (Amst.)* **73**, 71–77 (2019).
19. Prasad, R. et al. Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. *Nucleic Acids Res.* **42**, 6337–6351 (2014).
20. Sczepanski, J. T., Wong, R. S., McKnight, J. N., Bowman, G. D. & Greenberg, M. M. Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proc. Natl Acad. Sci. USA* **107**, 22475–22480 (2010).
21. Grosheva, A. S. et al. Recognition but no repair of abasic site in single-stranded DNA by human ribosomal uS3 protein residing within intact 40S subunit. *Nucleic Acids Res.* **45**, 3833–3843 (2017).
22. Spruijt, C. G. et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).

Acknowledgements

This work was supported by NIH grants nos. R01ES030575 (D.C.), R01GM117299 (B.F.E.) and P01CA092584 (D.C. and B.F.E.). P.S.T. was supported by F30CA228242, K.M.A. was supported by the Vanderbilt Molecular Biophysics Training Program (T32GM08320) and K.N.M. was supported by T32CA009582. Core facilities were supported by the Vanderbilt-Ingram Cancer Center P30CA068485. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of LS-CAT Sector 21 beamline was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817).

Author contributions

P.S.T., K.M.A. and K.N.M designed and completed the experiments, interpreted results and edited the manuscript. D.C. and B.F.E. supervised and funded the project, designed experiments and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0255-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.C. or B.F.E.

Peer review information: Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Protein purification. *E. coli* YedK was expressed in a modified pBG101 vector containing a rhinovirus 3C (PreScission) protease cleavable hexahistidine tag. *E. coli* BL21 (DE3) cells were grown in LB medium containing 15 ng ml^{-1} kanamycin at 37 °C to 0.8 optical density OD_{600}, and YedK overexpression was induced at 16 °C for 16 h after addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were collected by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 8.0 at 4 °C, 500 mM NaCl, 10% glycerol, 10 mM imidazole) with 1 mM each of leupeptin, pepstatin and aprotinin. The lysate was homogenized using dounce and pressure homogenizers (Avestin Emulsiflex), centrifuged at 20,500 r.p.m. for 30 min and passed through a 22-gauge needle before loading onto a 5 ml Ni-NTA column. The column was washed with lysis buffer containing 20 mM imidazole, and bound proteins were eluted with lysis buffer containing 300 mM imidazole. The N-terminal His-tag was removed by overnight incubation with PreScission protease (1:30 w/w) at 4 °C during dialysis (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM TCEP). The solution was passed over 2 mL Ni-NTA resin, and the flow-through further purified using gel filtration on a 16/300 Superdex 200 column (GE Healthcare) in S200 buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10% glycerol, 2 mM TCEP). YedK-containing fractions were concentrated to 4 mg ml^{-1} with Amicon MWCO 10 kDa centrifugal filters. Protein aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

YedK point mutants were generated using the QuikChange Site-Directed Mutagenesis Kit (Agilent), in which forward and reverse PCR reactions were performed separately to improve mutagenic primer annealing, and the corresponding single-stranded copies of the plasmid combined. Mutant plasmids were sequence verified. Mutant proteins were overexpressed and purified the same as wild type without the size-exclusion step. Mutant YedK was buffer exchanged in S200 buffer, flash-frozen in liquid nitrogen and stored at −80 °C.

Human HMCES SRAP domain (amino acids 1–270) was purified similarly to YedK with the following modifications. After repass over the Ni-NTA column, HMCES SRAP was purified via anion exchange via a HiTrap Q column before S200 size-exclusion chromatography in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol and 10 mM DTT.

DNA binding. Sequences of oligonucleotides used in the biochemical assays are listed in Supplementary Table 1. Relative binding affinity was measured by an electrophoretic mobility shift assay using ^{32}P-labeled DNAs containing a deoxyuracil. DNA (1 nM) was incubated with the indicated concentration of HMCES SRAP protein in reaction buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT, 100 μg ml^{-1} BSA) at 37 °C for 1 h. Ficoll was added to a final concentration of 2.5% and the samples were resolved on a 10% polyacrylamide gel in 1× TBE buffer (100 mM Tris-HCl pH 8, 90 mM boric acid, 2 mM EDTA) at 40 V for 180 min at 4 °C. Fluorescence anisotropy was used to measure binding of HMCES SRAP to ssDNA-dsDNA junctions containing a tetrahydrofuran (THF) abasic site analog. The THF strand contained 6-carboxfluorescein (FAM) at the 5′ end. Protein was titrated against 25 nM DNA in binding buffer (20 mM Tris-HCl pH 8.0, 100 nM NaCl, 10 mM $MgCl_2$, 5 mM DTT) in a 384-well plate for 20 min at 4 °C. Fluorescence was measured using a BioTek Synergy H1 Hybrid Reader with a filter cube containing 485/20 nm excitation and 528/20 nm emission filters.

DNA-protein cross-linking assays. For the experiments shown in Fig. 1a–c, AP-DNA was prepared by incubating 50 μM uracil-containing oligonucleotides with 25 units of uracil DNA glycosylase[23] (UDG, New England Biolabs) in Buffer X (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT) at 37 °C for 30 min. Human HMCES SRAP was incubated with AP-DNA in Buffer X at the following concentrations: 20.8 μM protein + 25 μM DNA (Fig. 1a) and 0.75 μM protein + 1.5 μM DNA (Fig. 1b,c). For the experiment shown Fig. 1c, DPCs were formed at 37 °C for 12 h and treated with either no heat or 95 °C for 2 min before incubation at 25 °C. Free and DNA-cross-linked HMCES were separated on 10% polyacrylamide Tris-glycine gels.

For the experiments shown in Figs. 1d,e, 3c–e and 4d, reaction products were separated on 15% polyacrylamide urea gels in 1× TBE buffer. In Fig. 1d, AP-DNA was prepared by incubating 100 nM uracil-containing ssDNA with 1 unit of UDG in Buffer X, cross-links formed with 10 nM AP-DNA and 100 nM SRAP in 20 mM Tris-acetate pH 8.0, 50 mM potassium acetate, 10 mM magnesium acetate and 5 mM DTT at 37 °C for 1 h, followed by proteinase K (Sigma Aldrich) digestion for 5 min. In Fig. 1e, DPC was formed using 1 μM human HMCES SRAP and 10 nM 3′-Cy5-labeled oligonucleotide in 20 mM Tris-HCl pH 6.0, 50 mM NaCl, 10 mM $MgCl_2$ and 5 mM DTT at 37 °C for 1 h. DPC was then digested with proteinase K at 37 °C for 5 min. APE1 (NEB) was added where indicated and incubated at 37 °C for 120 min.

E. coli YedK DPCs (Fig. 3c,d) were formed from incubation of 1 μM protein and 10 nM 5′-FAM-labeled oligonucleotide in 20 mM Tris-HCl pH 6.0, 1 mM EDTA and 5 mM DTT at 37 °C for 1 h. Schiff base intermediates (Fig. 3e) were trapped by incubating 2 μM YedK with 6 μM 5′-FAM-labeled oligonucleotide in 20 mM HEPES-NaOH pH 7.0, 100 mM NaCl, 1 mM DTT at 25 °C for 5 min, after which $NaCNBH_3$ was added to a final concentration of 50 mM and reactions incubated at 25 °C for 18 h.

DNA-binding reactions with ssDNA-dsDNA junctions (Fig. 4d) were carried out with 10 nM DNA and increasing concentrations of HMCES SRAP at 37 °C for 1 h in Buffer X. Cross-linking reactions with ssDNA-dsDNA junctions (Fig. 4e) were carried out with 1 nM AP-DNA and increasing concentration of HMCES SRAP at 37 °C for 1 h in Buffer X.

X-ray crystallography. AP-DNA was prepared by incubating 50 μM heptamer d(GTCUGGA) ssDNA with 2.5 units of uracil DNA glycosylase (UDG, New England Biolabs) in Buffer X at 37 °C for 30 min. YedK DPC was generated by incubation of 20 μM YedK with 25 μM AP-DNA for 1 h at 37 °C in MES pH 5.5, 50 mM NaCl, 10 mM $MgCl_2$ and 5 mM DTT. YedK DPC was purified via cation exchange on a MonoS 5/50 GL column, concentrated and buffer exchanged into 20 mM Tris pH 8.0, 80 mM NaCl, 2 mM TCEP and 0.5 mM EDTA. YedK DPC was crystallized by hanging-drop vapor diffusion at 21 °C by mixing equal volumes of 3 mg ml^{-1} YedK DPC and reservoir solution containing 16% (w/v) PEG 3350 and 0.2 M KH_2PO_4. Diffraction-quality crystals were grown from drops that were seeded with microcrystals produced in the same condition and that had been stabilized in 30% PEG 3350 and 0.2 M KH_2PO_4. Crystals were harvested 7 d after setting the drops and cryoprotected in 10% (v/v) glycerol, 30% PEG 3350 and 0.2 M KH_2PO_4 and flash-frozen in liquid nitrogen.

The non-covalent YedK–DNA complex was crystallized using the same heptamer DNA sequence as in the DPC, but with a C3-spacer (Integrated DNA Technologies) in place of the AP site. The YedK–DNA complex was formed by incubating 80 μM YedK with 96 μM heptamer C3-spacer ssDNA at 4 °C for 30 min. Crystals were grown by hanging-drop vapor diffusion at 21 °C from drops containing 2 μl protein-DNA solution, 2 μl reservoir containing 0.1 M Bis-Tris pH 5.4 and 23% (w/v) PEG 3350 and 0.5 μL DPC microcrystal seed stock stored in 30% PEG 3350 and 0.2 M KH_2PO_4. Crystals were harvested after 16 d into 0.1 M Bis-Tris pH 5.4, 30% PEG 3350 and 10% (v/v) glycerol, and flash-frozen in liquid nitrogen.

X-ray diffraction data were collected at the Advanced Photon Source beamlines 21-ID-D (DPC) and 21-ID-F (C3-spacer) at Argonne National Laboratory and processed with HKL2000 (ref. [24]). Data collection statistics are provided in Table 1. Phasing and refinement was carried out using the PHENIX suite of programs[25]. The DPC structure was phased by molecular replacement of a previously determined structure of YedK alone (PDB 2ICU). The protein was subjected to simulated annealing, atomic coordinate, temperature factor, and TLS refinement before building the DNA model. The entirety of the heptamer ssDNA and the Cys 2-DNA cross-link was readily apparent in the density maps. All seven nucleotides and the Cys 2-AP cross-link were manually built in Coot[26], guided by $2mF_o - DF_c$ and $mF_o - DF_c$ electron density maps. Geometry restraints for the thiazolidine linkage were generated from idealized coordinates of (2R,4R)−1,3-thiazolidine-2,4-dicarboxylic acid (ligand 5XB) from the 1.47-Å resolution structure of PDB 5FF2, and the stereochemistry of AP site and Cys 2 ring substituents verified by manual inspection of the electron density before model building. The protein-DNA model was iteratively refined by energy minimization and visual inspection of the electron density maps. The C3-spacer structure was phased by molecular replacement using the protein from the DPC structure, followed by simulated annealing to eliminate model bias before further refinement. The three nucleotides at the 5′ end of the DNA were readily apparent in the residual electron density. After several rounds of coordinate, *B* factor, and TLS refinement, the C3-spacer and the 3′ end of the DNA were visible, albeit with much weaker electron density. To minimize model bias in either structure, $2mF_o - DF_c$ composite omit and $mF_o - DF_c$ annealed omit electron density maps with AP or C3-spacer and Cys 2 removed from the structure factor calculation were used to guide placement and refinement of the cross-link or the C3-spacer. The final YedK–DNA models were validated using the wwPDB Validation Service and contained no residues in the disallowed regions of the Ramachandran plots. Structures were deposited in the Protein Data Bank under accession codes 6NUA (DPC) and 6NUH (C3-spacer).

All structural biology software was curated by SBGrid[27]. Structure images were created in PyMOL (https://pymol.org). Sequence conservation was mapped onto the structure using the Consurf Server[28]. YedK DPC containing a ssDNA-dsDNA junction was modeled by superposition of ideal B-DNA with the sequence d(GGA/TCC) onto the three d(GGA) nucleotides at the 3′ end of the ssDNA in the YedK DPC crystal structure.

Statistics and reproducibility. All experiments were completed at least three times unless otherwise indicated.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Structures have been deposited in the Protein Data Bank under accession codes 6NUA (DPC) and 6NUH (C3-spacer). Source data for Figs. 1b,c, 3d, 4d,e and Supplementary Fig. 4 are available with the paper online as Source Data for Figs. 1, 3, and 4. All other data are available upon request.

References

23. Krokan, H. E. et al. Error-free versus mutagenic processing of genomic uracil–relevance to cancer. *DNA Repair (Amst.)* **19**, 38–47 (2014).
24. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
25. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
26. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
27. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
28. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

nature research

Corresponding author(s): Brandt F. Eichman; David Cortez

Last updated by author(s): May 10, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	X-ray diffraction data were collected at the Advanced Photon Source beamlines 21-ID-F and 21-ID-D at Argonne National Laboratory and processed with HKL2000.
Data analysis	Phasing and refinement was carried out using the PHENIX suite of programs. Autoradiograms were quantified by ImageQuant software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

PDB ID codes 6NUA and 6NUH

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n/a
Data exclusions	no data were excluded
Replication	all biochemical experiments were performed at least three times. We confirm all data was reproducible.
Randomization	n/a
Blinding	n/a

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging