Characterization of SRAP Domain DNA-Protein Crosslink Formation and Reversal with DNA Abasic Sites

By

Katherine Amidon Paulin

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
Biological Sciences
August 12, 2022
Nashville, Tennessee

Approved:

Todd Graham, Ph.D.
David Cortez, Ph.D.
Breann Brown, Ph.D.
Jared Nordman, Ph.D.
Brandt Eichman, Ph.D.

Copyright © 2022 Katherine Amidon Paulin
All Rights Reserved

DEDICATION

In loving memory of my grandfather Eugene “Gramps” Matyas. Thank you for inspiring my love of science from an early age and teaching me to be confident in my abilities.

ACKNOWLEDGMENTS

Brandt – Thank you for being such a supportive mentor and for fostering a positive and collegial lab environment. Thank you for giving me the freedom to pursue the research directions that interest me, and for helping me to stay on track. I am very thankful for the opportunities that you have provided for me and for your feedback and support to help me grow as a researcher.

Past and present Eichman lab members- I'm am so thankful to have such supportive and helpful colleagues. You all have made for a positive graduate experience.

Diana Chavez – Thank you for being my rotation mentor and getting me started on the right foot in the lab. I never expected my rotation project to turn into my thesis project and you helped me get it going. Thank you!

Petria Thompson – Thank you for being an amazing collaborator and friend! I miss working with you. You really got me through some stressful times and I always enjoy your company. Thank you for staying invested in this project and staying in touch. I couldn't have done this without you!

Jorge Rua Fernandez – Thank you for our fruitful discussions and keeping me in the loop about HMCES research. Your perspective has been very helpful for this project.

Dave Cortez – Thank you for enabling such a productive collaboration between our labs. Our weekly meetings when I first started helped me learn quickly and broaden my perspective when looking at data. I am very appreciative of your feedback and support.

My Committee – Thank you or your support and guidance throughout this project. Your questions and ideas have helped me grow as a scientist and greatly contributed to these discoveries.

Nordman Lab – Thank you for letting me use your ChemiDoc very frequently. I appreciate how you all are so friendly and accommodating. I enjoy catching up while waiting for my gels to image.

MBTP and CSB members – Thank you for creating such a positive community to encourage structural biology research. I enjoyed contributing to various committees and creating engaging events and I will miss the monthly gatherings!

My wonderful husband David - Thank you for always supporting in me and believing in me. You've been by my side cheering me on throughout my grad school journey and I couldn't have done it without your love and support. I love you so much!

Mom and Dad – Thank you for letting me move across the country to pursue my dreams. I am so appreciative of all your love and support. Thank you for fostering my love of science from a young age and always encouraging me and helping me through stressful times. I am very blessed to have you as parents.

Drew – Thank you for being the best brother! I always have fun when we hang out and I love your sense of humor. Thank you for sending me your music- it makes my day!

Kim – Thank you for being such a fun sister! I'm glad we can be friends and I love your energy! I'm also glad you can appreciate my science now.

Amidon family – Thank you to all my family for always being so supportive and encouraging and kind. I am so thankful to have such wonderful people in my life!

Paulin family – I'm so thankful to call you my family now! Thank you for being so loving and supportive of me.

My lifelong friends across the country – Rachel, Katie, Joyce, Katren – even though I don't get to see each of you often, I always love that we can pick back up right where we left off. Each of you always encourages me and I'm thankful to know you.

Core facilities were supported by the Vanderbilt-Ingram Cancer Center P30CA068485. Use of the Advanced Photon Source was supported by the U. S. 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).

The work in Chapters 2 and 3 was supported by NIH grants 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). The work in Chapters 4 was funded by grants from the National Institutes of Health (R35GM136401 to B.F.E., F31ES032334-01 to K.A.P, and R01ES030575 to D.C.).

The aoN-g probe was a gift from Yasuo Komatsu at National Institute of Advanced Industrial Science and Technology (AIST).

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

LIST OF ABBREVIATIONS

Δ: heat

5'-dRP: 5'-deoxyribose phosphate

AID: activation-induced deaminase

Alt-EJ: Alternative end joining

AP: apurinic/apyrimidinic/abasic site

AP-DNA: abasic site-containing DNA

APE1: AP endonuclease 1

APE2: AP endonuclease 2

APEX: AP endonuclease X

APOBEC3A: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

AP-ICL: abasic interstrand crosslink

BER: base excision repair

β-ME: β-mercaptoethanol

CMG: CDC45-MCM2-7-GINS helicase

CSR: class switch recombination

DDR: DNA damage response/repair

DNA: deoxyribonucleic acid

DPC: DNA protein crosslink

DpC: DNA peptide crosslink

DSB: double strand break

dsDNA: double stranded DNA

E. coli: *Escherichia coli*

EMSA: electrophoretic mobility shift assay

EndoIII: exonuclease 3 (*nei*)

EndoVIII: endonuclease 8 (*nth*)

FA: Fanconi anemia

Fpg: formamidopyrimidine DNA glycosylase

HGT: horizontal gene transfer

HMCES: 5-hydroxymethylcytosine binding, embryonic stem cell-specific

HPLC-MS: high-performance liquid chromatography tandem mass spectrometry

HR: homologous recombination

ICL: interstrand crosslink

Ig: immunoglobulin

IR: ionizing radiation

kDa: kilodalton

LINE-1: long interspersed element-1

MMR: mismatch repair

MMS: methylmethanesulfonate

NEIL3: endonuclease VIII-like 3

NEB: New England BioLabs

NER: nucleotide excision repair

NHEJ: non-homologous end joining

nt: nucleotide

ORF: open reading frame

P: 3′-phosphate

PAGE: polyacrylamide gel electrophoresis

PARP: poly(ADP-ribose) polymerase

PCNA: proliferating cell nuclear antigen

PDB: Protein Data Bank

PEG: polyethylene glycol

PIP : PCNA-interacting protein

PMSF: phenylmethylsulfonyl fluoride

POLβ: Polymerase beta

POLκ: Polymerase kappa

POLζ: Polymerase zeta

PUA: 3′-phospho-α,β-unsaturated aldehyde

RNA: ribonucleic acid

ROS: reactive oxygen species

RMSD: root-mean-square deviation

SD: standard deviation

SHM: somatic hypermutation

SPRTN: SprT-Like N-Terminal domain protein

SRAP: SOS response associated peptidase

SSB: single strand break

ssDNA: single stranded DNA

SUMO: small ubiquitin-like modifier

TCEP: tris(2-carboxyethyl)phosphine

TLS: translesion synthesis

UDG: uracil DNA glycosylase

UDGX: uracil DNA glycosylase X

UV: ultraviolet radiation

YedK: *E. coli* SRAP protein

CHAPTER 1

Introduction[1,2]

Abasic (AP) sites are abundant lesions that result from exposure to environmental toxins and during normal metabolism, and lead to cytotoxicity and mutagenicity if encountered during DNA replication. An alternative high-fidelity pathway for replication-associated AP site repair was recently discovered involving a unique protective DNA-protein crosslink (DPC) formed by SRAP proteins. This dissertation will focus on the mechanism by which SRAP proteins form crosslinks to AP sites, as well as the stability and resolution of these DPCs, laying a foundation for understanding how SRAP participates in AP site repair.

1.1 DNA damage

DNA is subject to damage from both endogenous and exogenous sources such as environmental toxins, ultraviolet radiation, and reactive oxygen species generated by cellular metabolism *(1)*. Chemical modifications of DNA can impair replication and transcription, disrupting normal cellular processes. Unrepaired DNA lesions can lead to mutations, genomic instability, cancer, or cell death, so it is important for cells to have functional DNA repair machinery *(2)*.

1.1.1 Abasic site formation

Apurinic/apyrimidinic (AP, or abasic) sites are one of the most common forms of DNA damage, occurring at rates of 10,000 to 30,000 per cell per day *(3-5)*. AP sites are generated directly by reactive oxygen species or ionizing radiation, and as intermediates during base excision repair of aberrant nucleotides. Frequent exposure of DNA to environmental toxins, UV radiation, and reactive cellular metabolites generates chemically modified nucleobases *(6-8)*. Common environmental toxins that induce oxidative stress leading to AP sites include agricultural chemicals, plastics, and detergents containing nitrosamines, as well as carbon tetrachloride and chloroform used as solvents in manufacturing *(9,10)*. Other pollutants such as heavy metals can

both lead to oxidative damage and inhibit some glycosylases and nucleases involved in AP site repair *(11,12)*. AP sites arise from either spontaneous or DNA glycosylase catalyzed hydrolysis of the *N*-glycosidic bond that links the modified base to the deoxyribose *(13,14)* (Fig. 1A). DNA glycosylases remove a large number of alkylated, oxidized, and deaminated bases as the first step of the base excision repair (BER) pathway and are thus primarily responsible for AP sites *(15,16)*. DNA glycosylases can be classified as monofunctional- removing the base leaving the intact AP site- or bifunctional exhibiting lyase activity that cleaves DNA 3′ the AP site to generate a 3′-phospho-α,β-unsaturated aldehyde (PUA) and a 5′-phosphorylated end *(14)*.

1.1.2 Abasic site outcomes

AP sites lead to numerous detrimental outcomes including blocked transcription or mutagenesis *(16,17)*, additionally, AP sites are unstable and reactive and can lead to DNA strand breaks, interstrand DNA crosslinks (ICLs), and DNA-protein crosslinks (DPCs) *(18-20)*. In solution, an AP site in DNA is at equilibrium between a ring-closed 2′-deoxy-D-erythro-pentofuranose and a ring-opened aldehyde (Fig. 1A). The AP site exists primarily in the cyclic furanose form as a mixture of α- and β-hemiacetals, with approximately 1% of the sugar in the ring-opened aldehyde form *(21,22)*. This electrophilic aldehyde is susceptible to base-catalyzed β-elimination of the 3′ phosphoryl group, generating a single-strand break *(23)* (Fig. 1B). AP sites can also react with exocyclic groups of nucleobases on the complimentary strand to generate ICLs *(19,24)* (Fig. 1C), and with primary amines in proteins to generate DPCs *(18)* (Fig. 1D). In addition to their reactivity, AP sites lead to stalled replication forks by inhibiting replicative polymerases *(25)*. Moreover, replication forks that encounter an AP site on the template strand can lead to a double-strand break (DSB) (Fig. 1E).

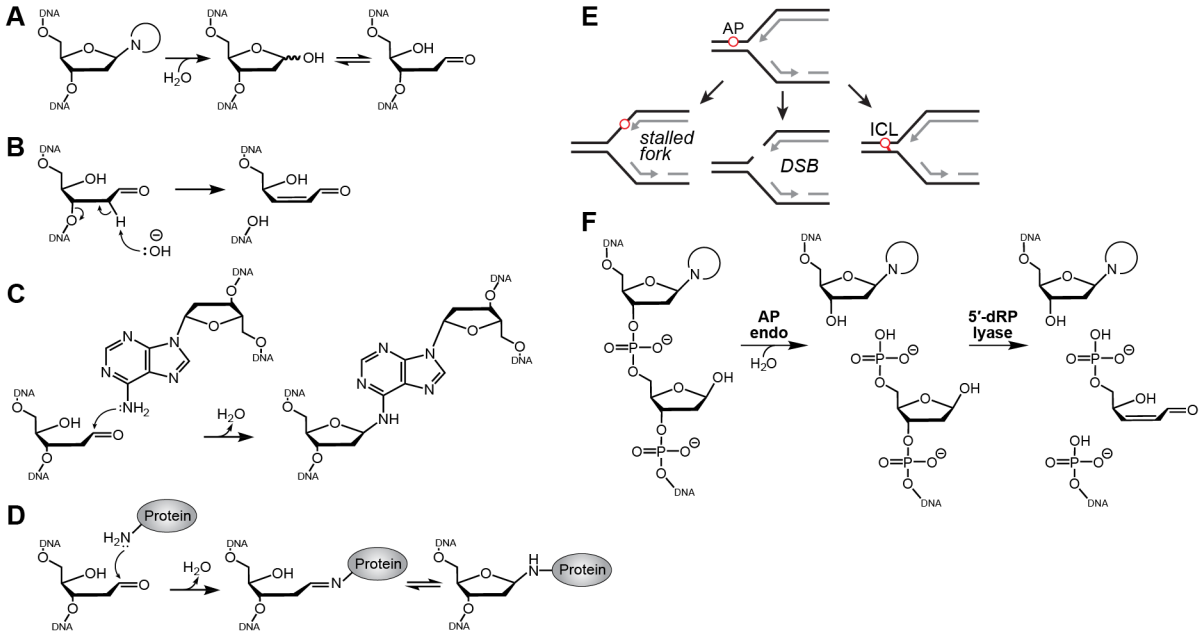

Fig. 1. Consequences of abasic sites. A. AP sites arise from enzymatic and spontaneous hydrolysis of the *N*-glycosidic bond and exist in either furanose or aldehyde forms. **B**. Base-catalyzed β-elimination of an AP site generates a strand break. **C,D**. Formation of an ICL (C) and a DPC (D) by nucleophilic attack of AP site C1′ by primary amines in DNA or proteins . **E**. Consequences of AP sites in the template strand during DNA replication. **F**. Incision of DNA by AP endonuclease and DNA lyases.

1.2 AP site repair and tolerance

1.2.1 Base excision repair (BER)

Despite the fact that AP sites form more readily in single-stranded DNA (ssDNA) *(13,26)*, AP site repair occurs within the context of double-stranded DNA (dsDNA) as part of BER (Fig. 2) *(14,27,28)*. In the second step of BER, AP endonuclease I (APE1) incises DNA 5′ to the AP nucleotide to generate a 3′-OH and 5′-deoxyribose phosphate (5′-dRP) residue (Fig. 1F) *(29)*. The 3′-OH serves as a substrate for long patch repair DNA synthesis by DNA pol δ or short-patch repair synthesis by DNA polymerase β (DNA pol β). DNA pol β also contains 5′-dRP lyase activity that cleaves the 3′ side of the AP site (Fig. 1F) *(30)*.

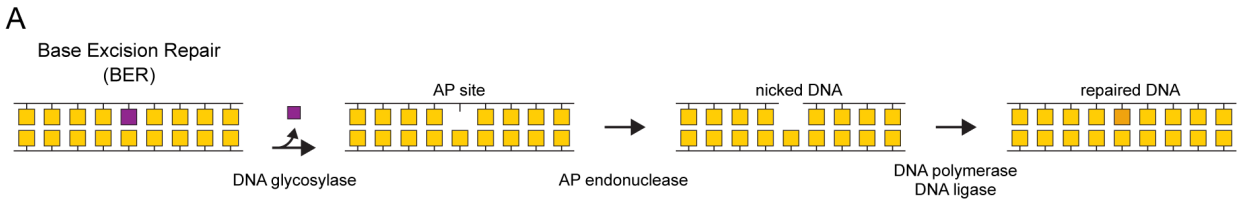

Fig. 2. *General base excision repair (BER) pathway* (A) Cartoon schematic of the base excision repair pathway. From left: a damage-specific DNA glycosylase catalyzes hydrolysis of the damaged base (purple) to generate an AP site. AP endonucleases incise the AP site 5′ to the lesion to create a gap. The nicked DNA is end-processed to remove the deoxyribose from the backbone, and DNA polymerase I synthesizes across the gap in an error-free manner and the gap is sealed by DNA ligase I.

1.2.1 Translesion synthesis (TLS)

AP sites that occur in ssDNA, such as those encountered during replication, are not removed by BER. Until recently, the only known fate of AP sites during DNA replication was lesion bypass by low-fidelity translesion synthesis (TLS) polymerases (Fig. 3) *(25,31)*. This error-prone damage-tolerance pathway allows replication to continue at the expense of potentially introducing mutations *(32,33)*. APE1 has much weaker activity for AP sites in ssDNA than in dsDNA *(34,35)*, and indeed, APE1 incision at an AP site in ssDNA would generate a strand break (Fig 3C).

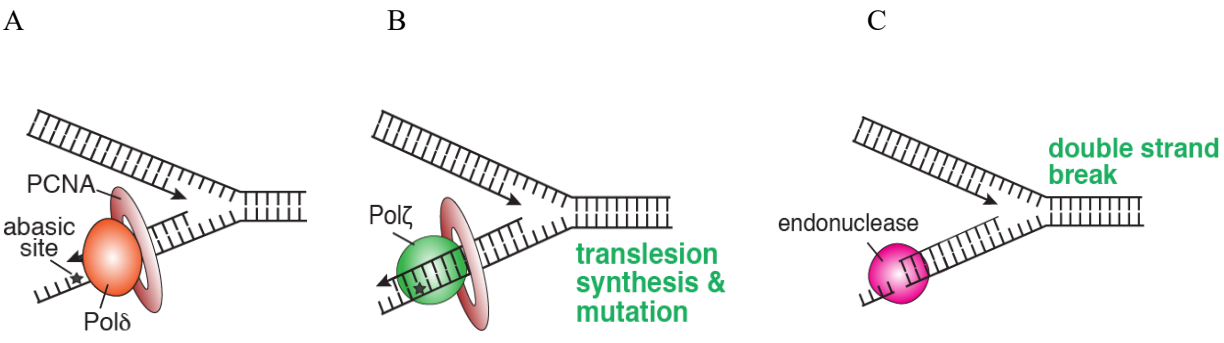

Fig. 3. Potential outcomes of AP sites encountered during replication.[3] (A) AP sites stall the replicative polymerase DNA pol δ. (B) The AP site can be bypassed by specialized translesion synthesis (TLS) polymerases which can insert a nucleotide across from the lesion in a low-fidelity process. (C) Endonucleases, such as APE1 can incise the AP site, but incision when the duplex is separated during replication can lead to a double strand break. [3] Figure adapted from: Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P. and Cortez, D. (2019) HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell*, **176**, 144-153.

1.3 SOS response-associated peptidase (SRAP) domain and identification of HMCES[2]

In 2013, the SOS Response Associated Peptidase (SRAP) domain was identified in a gene neighborhood analysis of bacterial SOS response operons *(36)*. The bacterial SOS response is a stress-induced network for DNA repair, adaptive mutagenesis, and cell cycle regulation, implying a role for SRAP proteins in DNA repair *(37)*. The SRAP domain is conserved across all domains of life, with a SRAP-containing protein found in every organism from bacteria to humans *(36)*. All SRAP domains contain an invariant cysteine at position 2 (Cys2) along with highly conserved glutamate and histidine residues. A number of SRAP structures determined by structural biology consortia were available in the Protein Data Bank (PDB) which revealed that the aforementioned residues lie at the center of a highly conserved positively charged channel (Fig. 4).

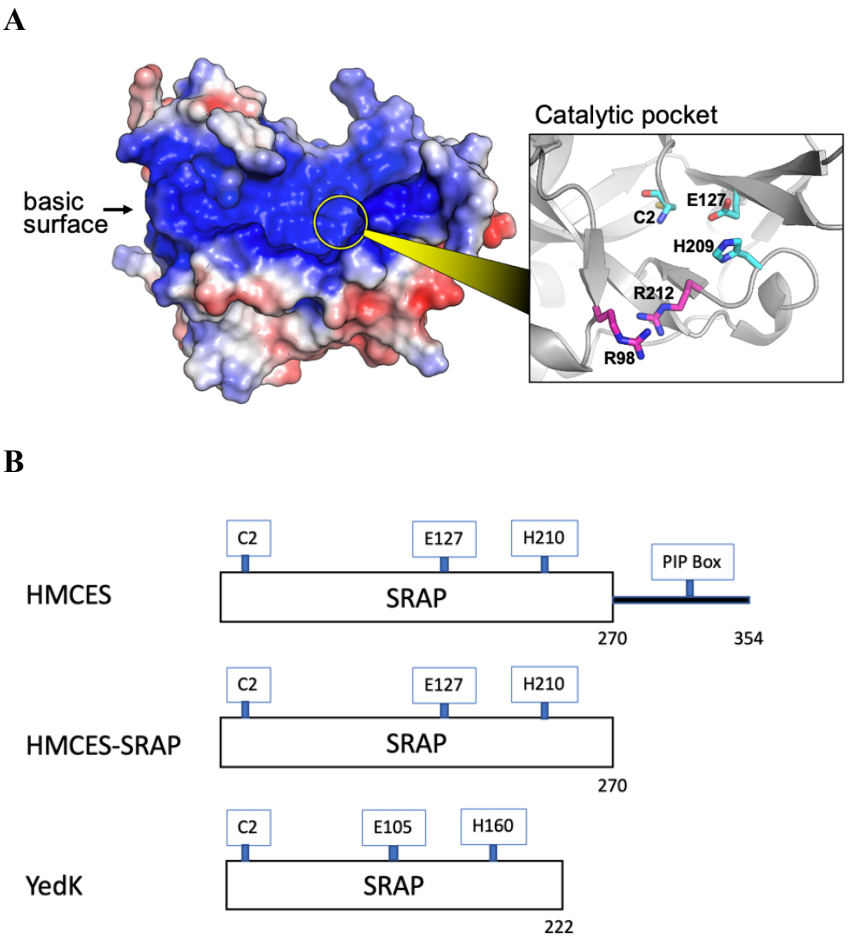

Fig. 4. SRAP protein structure. (A) *Structure of HMCES SRAP domain.* Crystal structure of HMCES SRAP domain (PDB ID 5KO9) Space filling model (left) of the overall structure colored by charge (blue positive, red negative, white neutral). Closer view of catalytic pocket with highly conserved residues shown as sticks (B) *Schematic of SRAP protein constructs utilized in this work.* Top to bottom: HMCES consists of a single SRAP domain and a C-terminal tail containing a PCNA-interacting protein (PIP) box. HMCES SRAP contains only the SRAP domain of human HMCES. YedK is the *E. Coli* ortholog of HMCES and contains a single SRAP domain.

The human SRAP protein 5-hydroxymethlycytosine (5hmC) binding, embryonic stem cell-specific (HMCES), and its *E. coli* ortholog, YedK, are similar in sequence (29% identity / 43% similarity) and structure. This high degree of conservation implies a crucial role for organism fitness, but surprisingly little is known about the function of these important proteins. HMCES was originally identified as a reader of 5-hmC in DNA *(38,39)* but demonstration of this function has not been replicated. HMCES deficient mice are viable and do not exhibit altered 5hmC levels *(40)*. Rather, HMCES has been shown to function in DNA repair *(41)*.

Our collaborators in the Cortez lab at Vanderbilt recently discovered a new high-fidelity pathway for processing replication-associated AP sites involving HMCES *(41)*. Importantly, HMCES-deficient cells exhibit increased genomic instability. Cells lacking HMCES accumulate AP sites and exhibit decreased viability and increased mutation frequency *(41)*. Additionally, HMCES is enriched at replication forks and directly interacts with the processivity factor for DNA polymerases, proliferating cell nuclear antigen (PCNA) *(41)*. In cells, abundance of HMCES DPC increases in response to AP-inducing agents, such as the environmental toxin potassium bromate or UV exposure. HMCES-DPCs are resolved over time by an unknown mechanism that may involve proteolysis *(41,42)*. HMCES is proposed to function in a novel repair pathway whereby it shields AP sites encountered during DNA replication from spontaneous strand breakage, endonucleases, or error-prone translesion synthesis.

1.4 SRAP DPC formation as a strategy for maintaining genome integrity

Intriguingly, the SRAP domain of HMCES natively forms covalent DNA-protein crosslinks (DPC) with AP sites in ssDNA, and this crosslinking ability is conserved in its *E. coli* ortholog, YedK *(41)*. SRAP exhibits a fascinating mechanism since it is one of only two known cases of a stable DNA-protein crosslink to AP sites without leading to strand cleavage *(43)*. SRAP

domains possess an invariant cysteine residue (Cys2) at the N terminus, which is required for DPC formation *in vivo* and *in vitro*. This activity is proposed to form the basis of a novel DNA repair pathway for AP sites encountered during replication, in which SRAP DPC protects AP sites from error-prone polymerases and nucleases *(41)*. A detailed mechanistic understanding of SRAP DPC formation and resolution crucial since SRAP proteins are unique in their ability to form natively occurring DPCs as well as their apparently protective role with AP sites. Additionally, most SRAP proteins remain uncharacterized. The high degree of SRAP conservation allows the mechanistic insights of this research to be informative across species, thereby increasing the potential for clinical utility.

SRAP-DPCs are relatively stable and resistant to boiling *(41,44)*. The high degree of stability of this crosslink is likely important for shielding the AP site from error-prone translesion synthesis and endonucleases. SRAP DPC abrogates spontaneous DNA strand breaks and also blocks nuclease activity of the enzyme AP endonuclease 1 (APE1) *(41,44)*. Indeed, even DPC proteolyzed to a peptide-DNA crosslink is resistant to cleavage by AP endonuclease 1 (APE1) *(41,44)*. Thus, SRAP is at the center of a fascinating albeit poorly understood pathway since it produces stable DNA-protein crosslinks to AP sites without leading to strand cleavage *(43)*.

1.5 SRAP structure-function approach

To investigate the nature of the SRAP DPC, I determined a crystal structure of YedK-DPC at atomic resolution (PDB 6NUA) revealing residues directly interacting with DNA as well as residues potentially involved in DPC formation and stability. This structure enabled a detailed model which forms the basis of my mechanistic hypothesis and design of mutants. Another structure of the human HMCES SRAP domain DPC was determined at the same time, albeit at lower resolution *(45)*. The DNA binding interfaces of the YedK and HMCES protein structures and are virtually identical, which allows YedK to be utilized to elucidate details of the SRAP domain pertinent to both bacteria and humans *(44,45)*.

The atomic resolution crystal structure revealed that the highly stable crosslink is formed from a thiazolidine linkage between the N-term Cys2 and the AP site *(46)*. I provide a brief overview of the structure here and Chapters 2 and 3 will discuss structural insights from the DPC structure in more detail. Cys2 is located at the N-terminus since Met1 is likely removed by an

aminopeptidase. In the DPC structure, continuous electron density from the Cys2 sulfhydryl and α-amino groups to the ring-opened AP site revealed the presence of a thiazolidine ring. The highly-conserved His160 hydrogen bonds to the hydroxyl of the ring-opened AP site and Glu105 hydrogen bonds with the phosphate 3′ to the AP site. The structure also revealed another conserved residue, Asp75, which hydrogen bonds to the carbonyl O and the backbone amide nitrogen of Cys2. Determining the mechanistic details of SRAP crosslinking to AP-DNA is critical since the process of how this DPC occurs as well as its role in DNA replication are not well understood.

The objective of this research is to determine how the SRAP domain utilizes a novel chemical mechanism to protect AP sites. To this end, I performed a structure-function analysis of the *E. coli* SRAP ortholog, YedK, which forms the basis of this dissertation. We were the first to determine structures of SRAP both covalently bound to AP-DNA and non-covalently bound to an AP site analog. Our structures also reveal the basis for SRAP's inability to bind to AP sites in duplex DNA and indicates a preference for AP sites at ssDNA-dsDNA junctions such as those formed by polymerase stalling at an AP site. Indeed, I show in Chapters 2 and 3 that the human (HMCES) and bacterial (YedK) SRAP proteins exhibit such a specificity for ssDNA-dsDNA junctions. Subsequent SRAP-DNA structures from other groups are consistent with our results and substantiate its conserved crosslinking function *(44,45,47)*.

The YedK DPC model enables mechanistic studies involving rational mutagenesis (Fig. 4). Structural and biochemical approaches elucidate the roles of conserved residues involved in SRAP-DPC formation, stability, and resolution. Chapter 2 will provide an analysis of published SRAP crystal structures, demonstrating the conservation of SRAP domain interaction with DNA and supporting YedK as an informative model for SRAP biochemistry across species. In Chapter 3, I discuss the structural basis of AP site protection by SRAP domains. The crystal structure of the YedK DPC explains the remarkable stability of the HMCES DPC, its resistance to strand cleavage, and the specificity for AP sites in ssDNA at junctions found when replicative polymerases encounter the AP lesion.

In Chapter 4, I review the collection of recently reported SRAP crystal structures from human HMCES and *E. coli* YedK, which provide a unified basis for SRAP specificity and a putative chemical mechanism of AP site crosslinking. I further discuss the structural and chemical basis for the stability of the SRAP DPC and how it differs from covalent DNA-protein intermediates in DNA lyase catalysis of strand scission. Finally, in Chapter 5, I present

unpublished preliminary data and discuss the future directions and implications of this project. Publication references in which I am a co/first-author are marked at the end of each chapter. All references are listed at the end of the dissertation.

[1]This work is adapted from my F31 proposal: NIH / NIEHS F31 ES032334-01 (Amidon) Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship "The molecular basis of SRAP domain DNA-protein crosslinking"
I wrote and edited the text and figures.

[2]This work is adapted from the introduction of: Amidon, K. M., and Eichman, B. F. (2020) Structural biology of DNA abasic site protection by SRAP proteins, *DNA Repair* ***94***, 102903.
I wrote and edited the manuscript and figures.

CHAPTER 2

Protection of abasic sites during DNA replication by a stable thiazolidine DNA-protein crosslink[4]

2.1 Introduction

Apurinic and apyrimidinic (abasic or AP) site repair via base excision repair (BER) depends on an intact DNA duplex*(48-50)*. While most AP sites form in double-stranded DNA (dsDNA), base loss is actually more rapid in single-stranded (ssDNA)*(51)*. 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 Polζ *(52)*. However, we recently discovered an alternative pathway dependent on the SRAP (SOS-Response Associated Peptidase) domain protein HMCES (5-Hydroxymethylcytosine Binding, ES Cell Specific) that improves cell viability and reduces mutation frequency*(53)*.

SRAP proteins are conserved in organisms from bacteria to humans, and in bacteria SRAP encoding genes are often spatially linked to DNA repair genes*(54)*. Human HMCES and *E. coli* YedK are similar in both sequence (29% identity and 43% similarity) and structure (C_α RMSD of 1.29 Å between PDB entries 5KO9 and 2ICU). Both HMCES and YedK preferentially bind ssDNA and efficiently form DNA-protein crosslinks (DPCs) to AP sites in ssDNA*(53)*. 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*(53)*. Despite the importance of the HMCES AP site DPC to this mechanism, the chemical nature of the crosslink 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 crosslink 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 crosslink 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.

2.2 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*(53)*. 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 crosslinked protein (Fig. 5a). We noticed while doing these experiments that boiling the DPC hydrolyzed the crosslink but incubation at 50°C did not (Fig. 5b). 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. 5c). Furthermore, extensive proteolysis of the DPC with proteinase K left a small peptide-DNA linkage that remains stable (Fig. 5d) and resistant to cleavage by APE1 (Fig. 5e). 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.

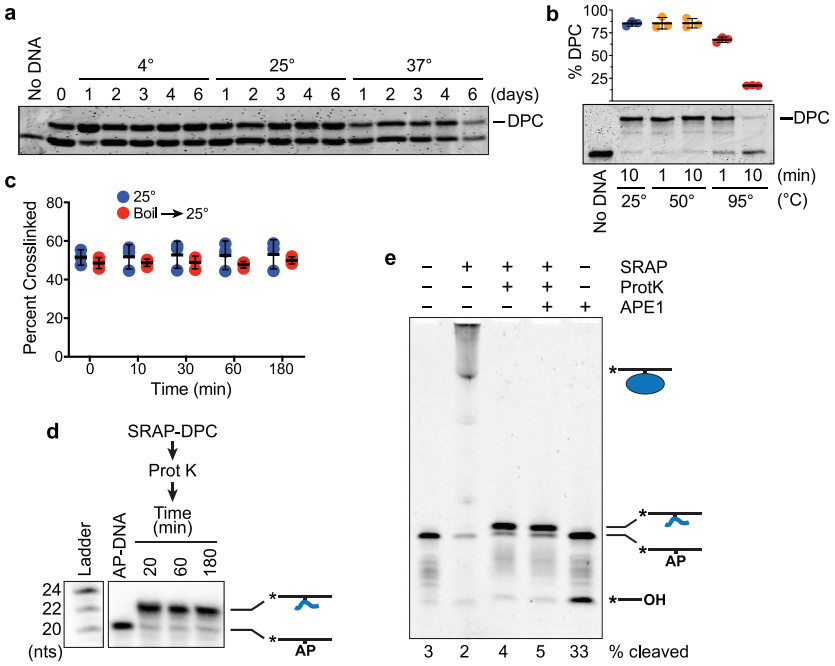

Fig. 5. Stability analysis of the human HMCES SRAP-abasic site DNA protein crosslink. **a,** HMCES SRAP DPC stability measured at the indicated temperatures. Free and DNA-crosslinked HMCES was detected by coomassie blue staining. The HMCES-DPC percentage in this experiment is approximately 50% because uncrosslinked 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 two minutes. **d,** HMCES SRAP domain was incubated with a 20-mer AP-site containing oligonucleotide to form a crosslink, digested with proteinase K followed by heat inactivation of the protease, and then incubated at 37°C for the times indicated. Electrophoresis and autoradiography was used to visualize the DNA. **e,** HMCES SRAP was incubated with 31-mer AP-DNA and digested with proteinase K, and the peptide DPC incubated with APE1 for 2 hours. Bands were visualized by Cy5 fluorescence.

To understand the molecular basis for the stability of the SRAP DPC, we determined a 1.6 Å crystal structure of *E. coli* YedK covalently crosslinked to 7-mer ssDNA (Table 6) containing an AP site. The entire DNA ligand is visible in the electron density (Fig. 6a). The protein does not undergo any appreciable conformational change upon binding DNA, with an RMSD of 1.16 Å for all atoms between unbound and DPC forms of YedK (Fig. 7). The core β-sheet forms an extended, positively charged channel that cradles the ssDNA phosphoribosyl backbone along one face of the protein (Fig. 6b-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 (Fig. 7). The hydrogen-bonding edges of

every nucleobase are exposed to solvent, and thus recognition of the AP site 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. 6b). 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*(53)*. The residues lining the DNA binding channel are the most highly conserved among SRAP domains (Figs. 6d-e, 7), 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 abrogate DNA binding for both proteins*(53)*.

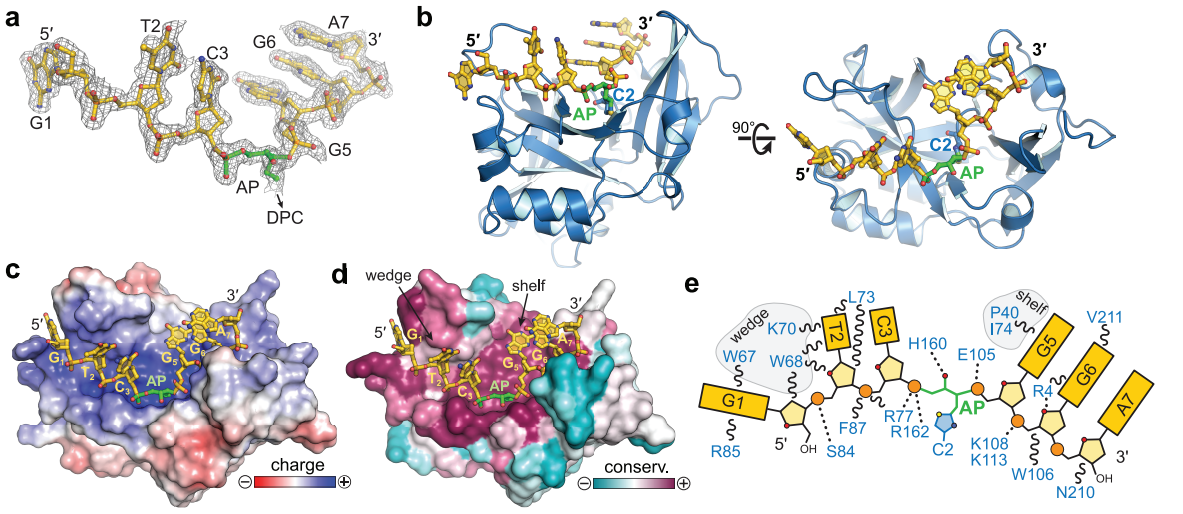

Fig. 6. YedK DPC crystal structure. **a,** DNA fit to 2Fo-Fc composite annealed omit electron density contoured at 1σ. **b,** Orthogonal views of *E. coli* YedK (blue) crosslinked to AP-DNA (gold). **c,d,** YedK solvent-accessible surface colored by electrostatic potential from −5 to +5 k_BT/e_C **(c)** and sequence conservation from 158 unique SRAP orthologs **(d)**. **e,** Schematic of DNA-protein interactions.

The AP site is positioned directly above Cys2, previously implicated in SRAP DPC formation*(53)*. 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 Cys2 side chain (Fig. 8a). The anomeric C1′ carbon of the AP site is covalently bonded to both the α-amino nitrogen and the side chain sulfur of Cys2 to form a thiazolidine ring (Fig. 8a). Such a linkage would be generated by nucleophilic

attack of the AP aldehyde C1′ carbon by Cys2 α-NH_2 to form a Schiff base intermediate, followed by subsequent attack of C1′ by the Cys2 sulfhydryl group (Fig. 8b)*(55)*. Consistent with crosslinking by Cys2, YedK DPC formation is abrogated by removal of the thiol in a C2A mutant*(53)*, and by a C2S mutant, which potentially forms an oxazolidine ring that would not be as stable as a thiazolidine (Fig. 8c,d)*(56,57)*.

Studies on the reaction of cysteine and aldehydes show that the equilibrium between Schiff base and thiazolidine greatly favors the latter*(58,59)*, 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 found in bifunctional DNA glycosylases that initiate BER (Fig. 8b)*(60-62)*. 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. 8c). This lyase activity was significantly reduced by performing the crosslinking reaction in the presence of sodium cyanoborohydride ($NaBH_3CN$), which acts as a reducing agent to stabilize the Schiff base intermediate (Fig. 8e)*(63)*. 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.

Cys2 belongs to a cluster of three conserved residues that includes Glu105 and His160 implicated in SRAP function*(54,64)*. These and several other evolutionarily conserved residues stabilize the DNA and protein sides of the thiazolidine linkage (Figs. 8f,g, 7). The AP site is stabilized by His160, which forms a hydrogen bond with the O4′ hydroxyl group (Fig. 8f). Similarly, Arg77 and Arg162, previously shown to be essential for DNA binding*(53)*, and Thr149, interact with the AP site 5′-phosphate (Fig. 8f,g). The Glu105 side chain fluctuates between two conformations at the crosslink (Figs. 8f, 9). 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.

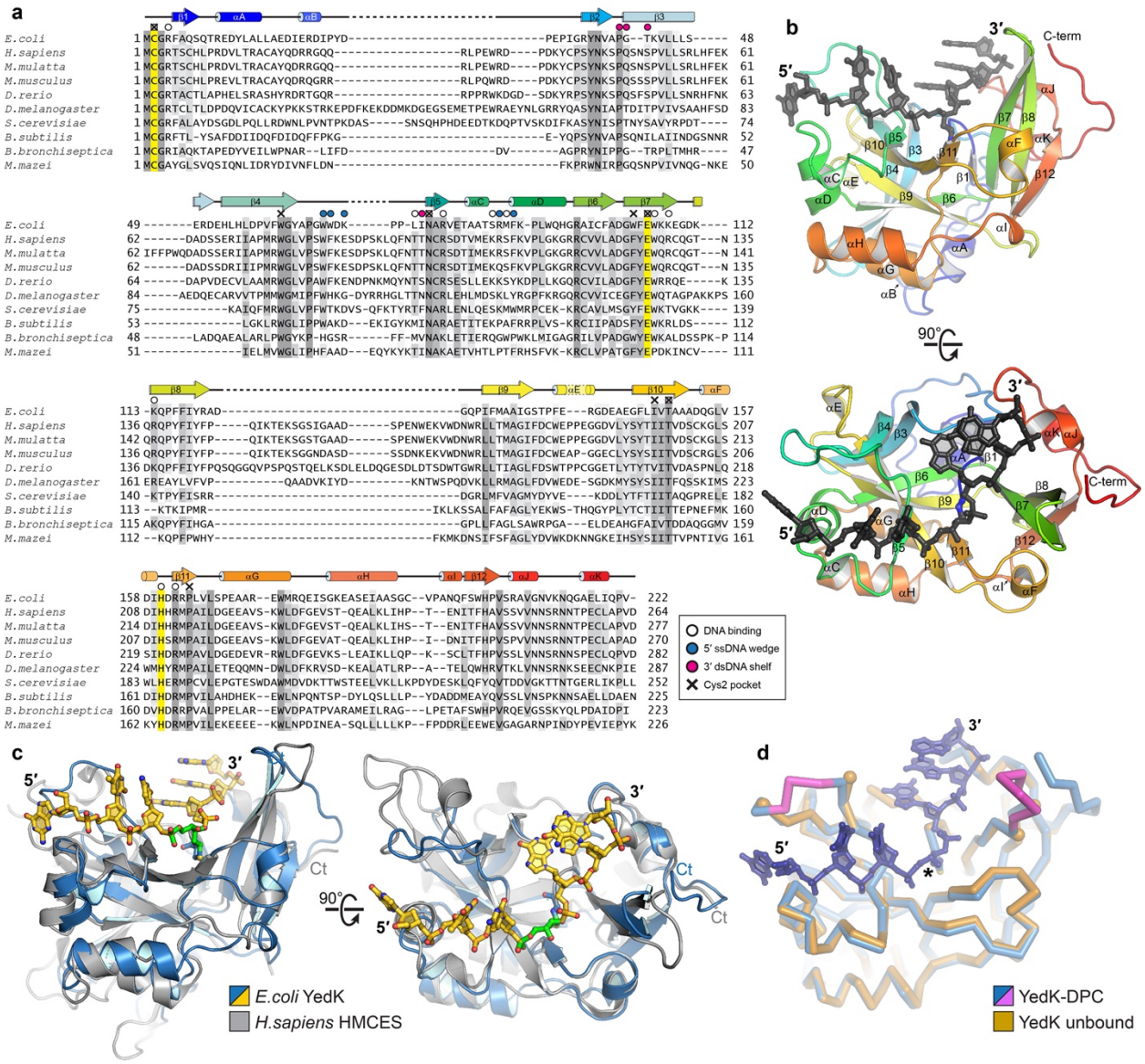

Fig. 7. SRAP conservation. a. Structure-based sequence alignment between *E. coli* YedK and human HMCES SRAP domain, together with a sequence alignment of SRAP domains from 8 additional species. Secondary structure from the YedK DPC structure is shown above the alignment. Symbols above specific residues denote those involved in binding DNA (circles), distorting DNA at the 5′ side of the AP site (wedge, blue circles), stacking against dsDNA immediately 3′ to the AP site (shelf, magenta circles), and stabilizing the thiazolidine crosslink (Cys2 pocket, "×"). **b.** Orthogonal views of the YedK DPC structure colored rainbow from N- (blue) to C-terminus (red). **c.** Superposition of *E. coli* YedK DPC (blue/gold) and human HMCES SRAP domain (PDB ID 5KO9, silver). The RMSD between the structures is 1.40 Å for all backbone atoms. d. Superposition of YedK DPC (blue/magenta) and free YedK (orange). Loops in YedK DPC that are disordered in the free protein are colored magenta. Proteins are shows as a Cα-backbone trace.

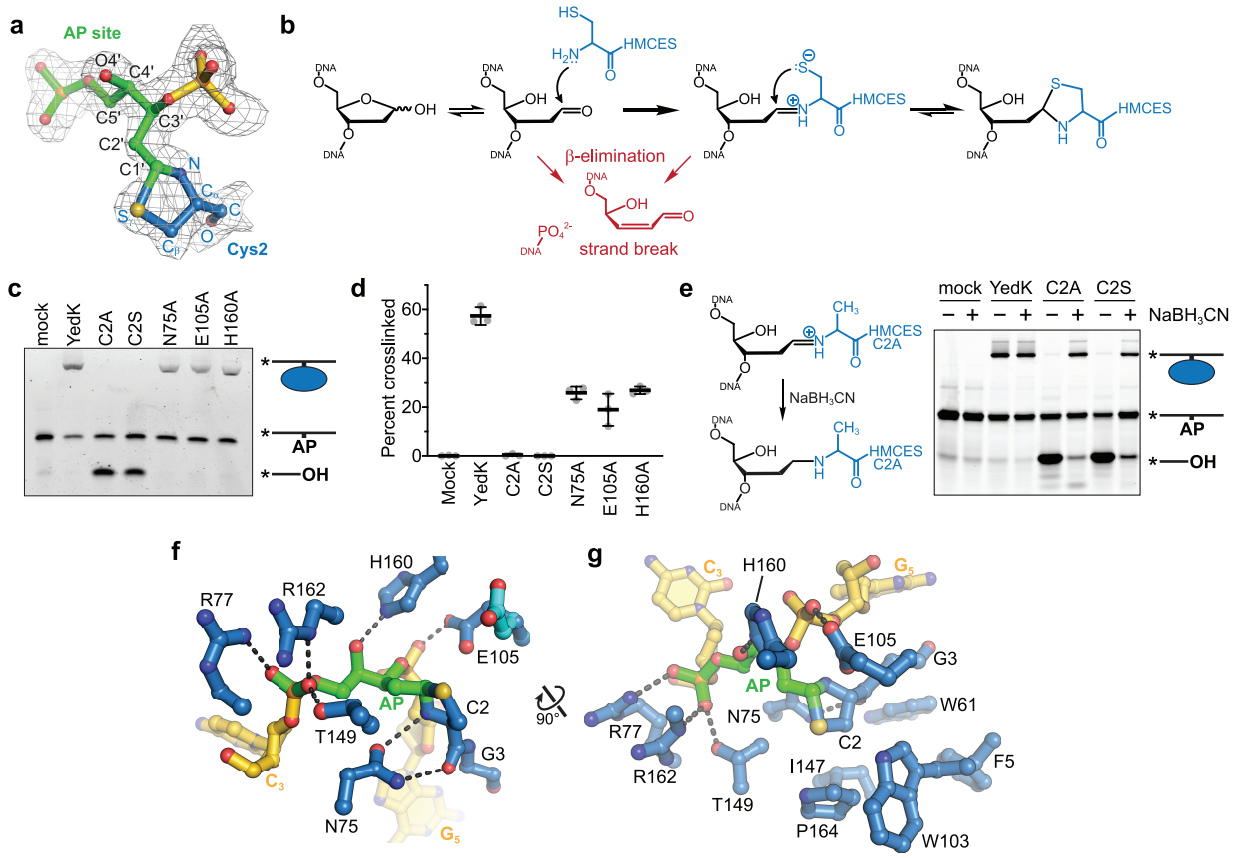

Fig. 8. The SRAP DPC forms a thiazolidine linkage stabilized by conserved residues. **a-h, a,** The DPC between the AP site (green) and Cys2 (blue) superimposed against 2Fo-Fc composite annealed omit electron density contoured at 1σ. **b,** Proposed chemical mechanism of the crosslinking reaction with competing lyase reactions in red. **c,** Representative denaturing PAGE gel showing crosslinking and lyase activity of YedK mutants. Bands were visualized by FAM fluorescence. **d,** Crosslinking efficiencies of YedK mutants (mean ± SD, n=3 independent measurements). **e,** $NaBH_3CN$ was added to crosslinking 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 Glu105 conformer is cyan. Dashed lines denote hydrogen bonds. **g,** Orthogonal view showing hydrophobic residues cradling Cys2. The second Glu105 conformer is not shown for clarity.

On the protein side of the crosslink, the carboxamide side chain from a highly conserved asparagine (Asn75) helps position the crosslinking nucleophile by forming two hydrogen bonds with the backbone amide nitrogen and carbonyl oxygen of Cys2 (Figs. 8f, 9). Consistent with their roles in stabilizing the crosslink, individual substitutions of Glu105, His160, or Asn75 with alanine reduced crosslinking efficiency (Fig. 8c,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 Cys2 from underneath (Figs. 8g, 7). Thus, the SRAP structure guides the AP site into a specific, solvent inaccessible environment suited for thiazolidine formation and protected from AP endonuclease cleavage.

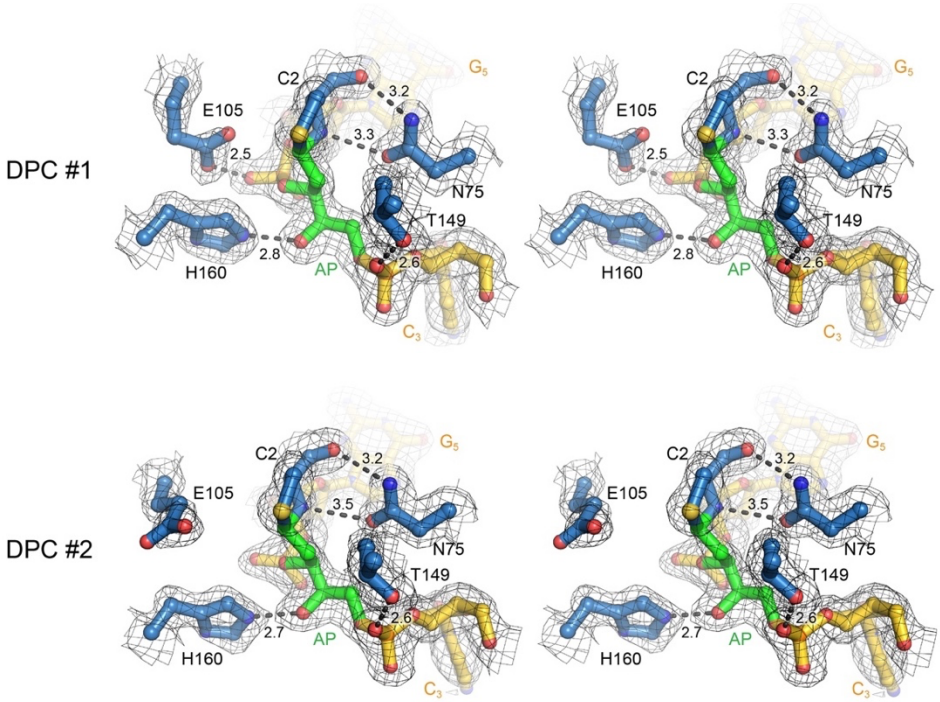

Fig. 9. Details of the SRAP active site. Stereo views of residues contacting the thiazolidine linkage in the two DNA-protein complexes in the asymmetric unit of the YedK DPC structure, superimposed with 1σ 2Fo-Fc electron density. Dashed lines and numbers indicate lengths of hydrogen bonds in Å.

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 (Fig. 10). The protein in the non-covalent complex is virtually identical to that of the DPC, except for modest repositioning of a β-hairpin (β 7-β8) that was disordered in the unbound YedK structure (PDB ID 2ICU) and that stabilizes the backbone of the DNA 3′ to the AP site in the DPC (Fig. 10a). 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 (Fig. 10b-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. 11a,b). Trp68 wedges the nucleobases of G1 and T2 apart, and G1 is stacked between Trp67 and Arg85 (Fig. 11b). 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 (Fig. 7).

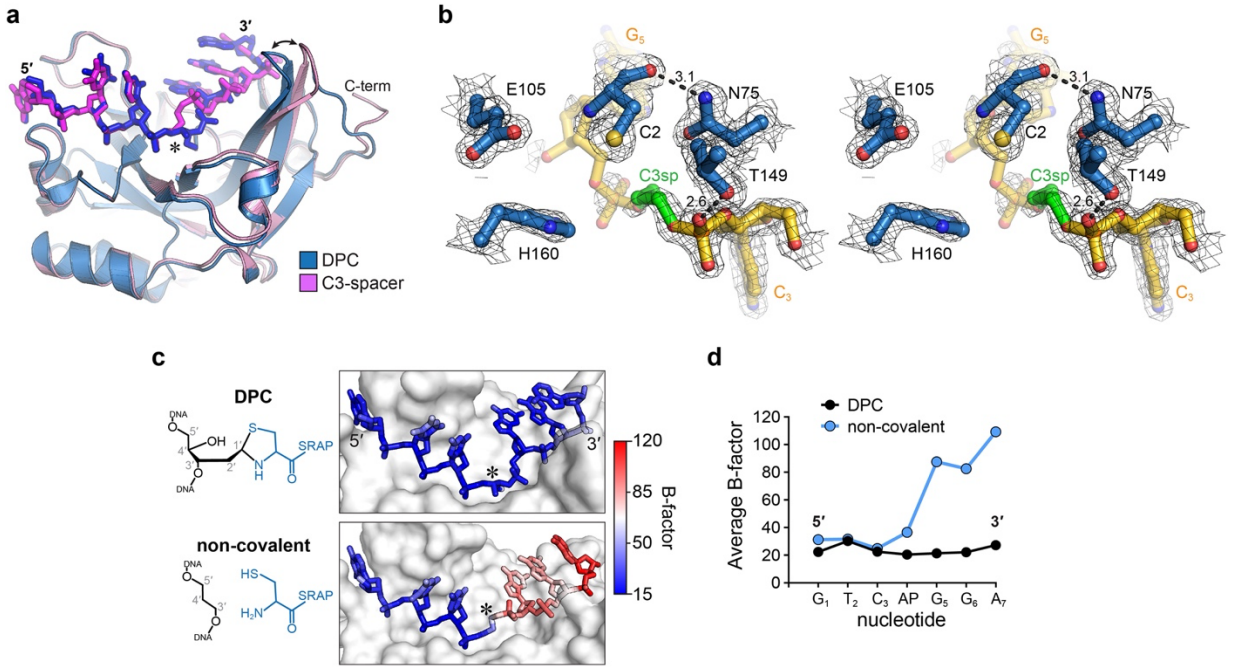

Fig. 10. Structural details of the non-covalent SRAP-DNA complex. a. Superposition of YedK covalently crosslinked to AP-DNA (blue) and non-covalently bound to C3-spacer DNA (magenta). The abasic site is marked with an asterisk. The double-headed arrow shows the most significant difference between the two structures—movement of β-hairpin (β7- β8) that stabilizes the backbone of the DNA 3′ to the AP site. b. Stereo view of the YedK active site in the YedK/C3-spacer-DNA structure, superimposed with 1σ 2Fo-Fc electron density. The C3-spacer is colored green and flanking nucleotides gold. Dashed lines and numbers indicate lengths of hydrogen bonds in Å. **c.** DNA in the DPC (top) and non-covalent C3-spacer (bottom) structures, colored by B-factor. **d.** Average B-factor of each nucleotide in the DPC (black) and non-covalent C3-spacer (blue) structures.

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. 11b). The residue adjacent to the AP site (guanine G5) stacks against Pro40 and Ile74 on the surface of the protein (Fig. 11b,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. Modelling shows that a complementary strand fits against the protein surface with no steric clashes (Fig. 11a-c). The 3′-end of the modeled strand stacks against Gly41 and Thr42, which together with Pro40 and Ile74 form a highly conserved "shelf" that would stabilize a base pair 3′ to the AP site (Fig. 11c, 7). 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′-overhaning template. Consistent with this prediction we found that HMCES is just as efficient at binding and crosslinking to an AP site immediately adjacent to the 3′ ssDNA-dsDNA junction as to ssDNA (Figs. 11d,e,f). In contrast, binding and crosslinking is less efficient when the dsDNA is present on the 5′-side of the AP site, consistent with the effect of the wedge motif.

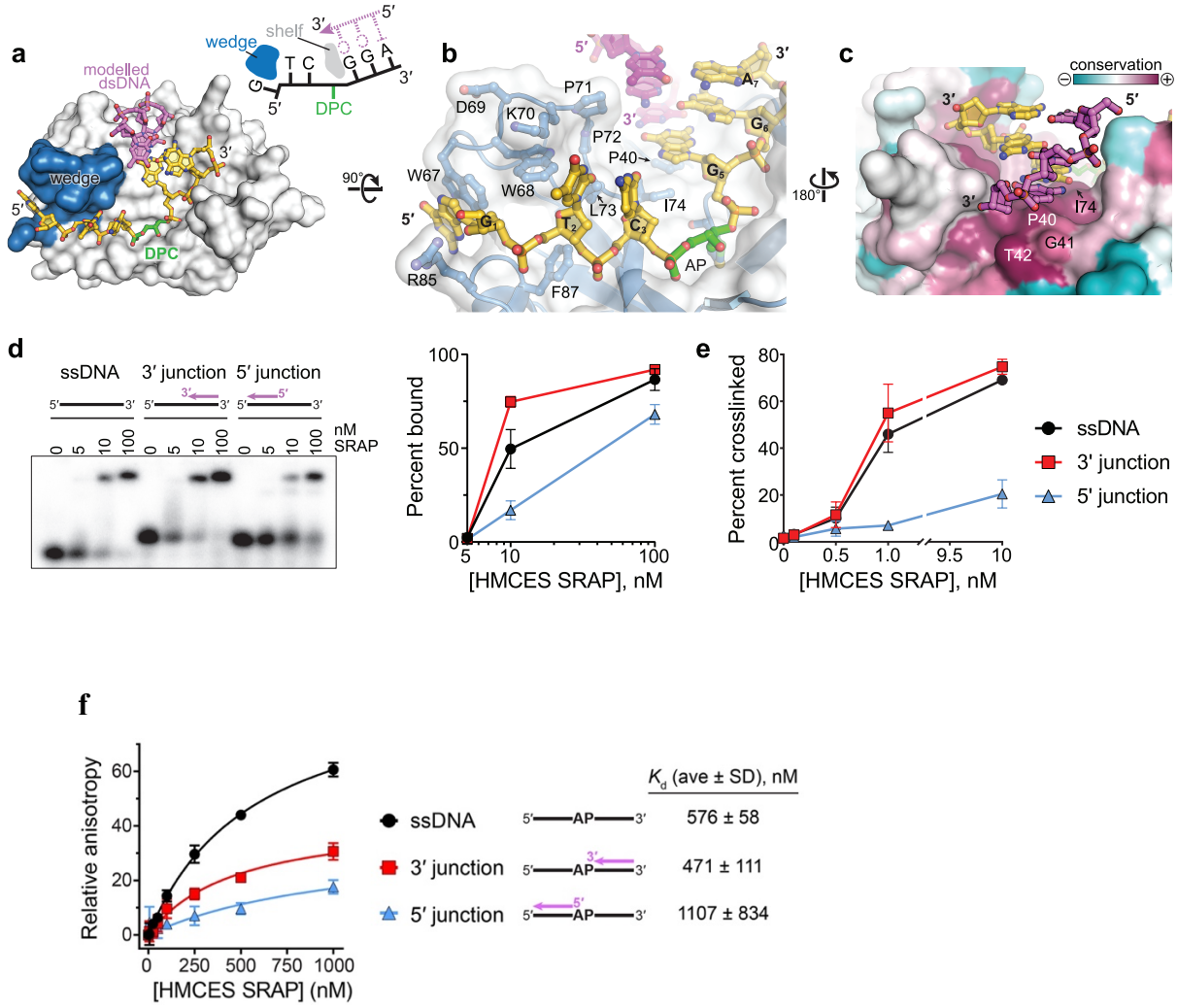

Fig. 11. SRAP can accommodate dsDNA 3′ to the AP site. a, Model of YedK DPC with a 3′ junction at AP site. The modeled complementary DNA strand is pink. The wedge domain blocking dsDNA access 5′ to the AP site is blue. **b,** Wedge-DNA interactions 5′ to the AP site. **c,** Sequence conservation of the DNA shelf that presumably stabilizes dsDNA 3′ to the AP site. **d,** 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. **e,** Percent of the indicated DNA substrates crosslinked to human HMCES SRAP domain (mean ± S.D., n=3 independent measurements). **f.** Binding of HMCES SRAP domain to ssDNA and ssDNA/dsDNA junctions containing a tetrahydrofuran (THF) abasic site analog. Binding was monitored by a change in fluorescence anisotropy as protein was titrated against DNAs that contained a FAM label at the 5′-end of the THF strand. The maximal change in anisotropy (amplitude of binding isotherm at saturation) is dependent on the tumbling rate of the FAM label, which is different for each of the three substrates. Dissociation constants (Kd) were derived from non-linear least squares fit of a two-state, single-site binding model to the data.

2.3 Discussion

The YedK-AP-DNA crosslink structure reveals how the unique DNA binding surface and N-terminal cysteine facilitates recognition and covalent crosslinking of HMCES and SRAP-containing proteins to AP sites in the context of ssDNA. Furthermore, the results explain the stability of this crosslink 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 the Schiff base, or (3) APE1 incision. This contrasts with unstable, transient DNA-protein Schiff base crosslinks that rapidly proceed to β-elimination as part of enzymatic strand cleavage reactions catalysed by bifunctional glycosylases and DNA polβ as part of the BER pathway*(60-62)*. Other proteins, including PARP-1, Histone H4, and Ribosomal protein uS3 can crosslink to AP sites, but in each case the DPC leads to strand scission*(65-68)*.

HMCES is named 5-Hydroxymethylcytosine (5hmC) Binding, ES Cell Specific 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*(69)*. Furthermore, the HMCES SRAP domain was shown to autoproteolyze itself and incise duplex DNA containing 5hmC*(64)*. 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 crosslink and unique thiazolidine DPC linkage supports the conclusion that these proteins act to maintain genome stability during DNA replication and thereby improve organism fitness.

---

[4] This work was originally published as: Thompson, P. S., Amidon, K. M., Mohni, K. N., Cortez, D., and Eichman, B. F. (2019) Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link, *Nature Structural & Molecular Biology* ***26***, 613-618.
I designed and completed experiments, interpreted results, and edited the manuscript.

CHAPTER 3

Structural biology of DNA abasic site protection by SRAP proteins[4]

Abasic (AP) sites are one of the most frequently occurring types of DNA damage. They lead to DNA strand breaks, interstrand DNA crosslinks, and block transcription and replication. Mutagenicity of AP sites arises from translesion synthesis (TLS) by error-prone bypass polymerases. Recently, a new cellular response to AP sites was discovered, in which the protein HMCES (5-hydroxymethlycytosine (5hmC) binding, embryonic stem cell-specific) forms a stable, covalent DNA-protein crosslink (DPC) to AP sites at stalled replication forks. The stability of the HMCES-DPC prevents strand cleavage by endonucleases and mutagenic bypass by TLS polymerases. Crosslinking is carried out by a unique, SRAP (SOS Response Associated Peptidase) domain conserved across all domains of life. Here, we review the collection of recently reported SRAP crystal structures from human HMCES and *E. coli* YedK, which provide a unified basis for SRAP specificity and a putative chemical mechanism of AP site crosslinking. We discuss the structural and chemical basis for the stability of the SRAP DPC and how it differs from covalent DNA-protein intermediates in DNA lyase catalysis of strand scission.

3.1 Introduction

An error-free pathway for repair of replication-associated AP sites was discovered recently that depends on the protein HMCES [5-hydroxymethlycytosine (5hmC) binding, embryonic stem cell-specific (originally C3Orf37)] *(41)*. Cells lacking HMCES exhibit elevated levels and delayed repair of AP sites, as well as increased double-strand breaks and mutation frequency from TLS. HMCES is recruited to replication forks via a direct interaction with proliferating cell nuclear antigen (PCNA), the processivity factor for replicative polymerases and a hub for replication-associated processes *(70)*. Importantly, HMCES forms a covalent DPC with AP sites present in ssDNA, but not dsDNA *(41)*. The HMCES DPC is highly stable and persists in cells *(41,46)*. Subsequent repair of the DPC is unknown, although the protein is eventually degraded via the

proteasome *(41)*. The current model is that the highly stable HMCES DPC protects AP sites from nuclease cleavage and mutagenic TLS polymerases *(41,71)*.

3.2 Function of the SRAP domain in AP site repair

HMCES cellular function and AP-site crosslinking activity depends on a highly conserved SRAP (SOS Response Associated Peptidase) domain, which constitutes the majority of the protein. The SRAP domain is conserved across all domains of life, with a SRAP-containing protein found in most bacteria and eukaryotes, some archaea, as well as viruses and bacteriophages *(36,41,46)*. HMCES is the only SRAP-containing protein present in humans. HMCES and other eukaryotic SRAP proteins have a C-terminal disordered region containing at least one non-canonical PCNA-interacting protein (PIP) box that mediates the direct interaction with PCNA *(41)*, whereas the SRAP domain constitutes the entirety of the bacterial proteins *(36)*. The bacterial SRAP proteins do not appear to have an interacting motif (QL(S/D)LF) for the DNA polymerase III β-subunit (β-clamp) *(72)*—the prokaryotic ortholog of PCNA *(73)*—and their potential to interact with the sliding β-clamp is unknown. Crosslinking to AP sites in ssDNA has been demonstrated for the SRAP domain of human HMCES and the *E. coli* ortholog YedK *(41)*.

SRAP was named for the proximity of the gene to other prokaryotic DNA repair (SOS response) genes, and for its highly conserved triad of cysteine, glutamate, and histidine residues (Cys2, Glu127, and His210 in HMCES) reminiscent of cysteine proteases *(36)*. Cys2 is invariant and required for HMCES function in cells and for the crosslinking activity of HMCES and YedK *(41)*. AP site crosslinking by Cys2 depends on removal of the N-terminal methionine by an aminopeptidase activity that exposes Cys2 at the N-terminus *(36,38,41)*. Cleavage of Met1 by HMCES autoproteolysis has been reported *(38)*, although all organisms contain essential methionyl aminopeptidases that are capable of removing this residue. Crosslinking activity also depends on natural AP sites, as a tetrahydrofuran (THF) analog lacking the hydroxyl group at C1′ does not react with SRAP *(41)*.

The native DPC formed between SRAP and AP sites is a unique DNA repair mechanism. Typically, proteins covalently conjugated to DNA occur either as deleterious lesions *(74-76)* or as a catalytic intermediate in DNA strand cleavage (lyase) reactions *(77-79)*. For example, the AP lyase activity of a number of DNA repair proteins depends on formation of a transient Schiff base

intermediate between protein amino and DNA carbonyl groups *(80,81)*. In contrast, SRAP DPCs are highly stable—on the order of hours in cells and days *in vitro* at physiological temperature, and resistant to boiling for up to 10 minutes *(41,46)*. Consequently, the SRAP DPC blocks nuclease activity by APE1, even when SRAP has been proteolyzed to leave a DNA-peptide crosslink *(41,46)*. Thus, the high degree of stability of the SRAP DPC is likely important for shielding AP sites from cleavage and explains the HMCES-dependent reduction of spontaneous DNA strand breaks in cells *(41)*.

3.3 Structural basis for SRAP interaction with DNA

Over the past 14 years, a number of SRAP protein structures have been deposited in the Protein Data Bank (PDB) (Table 1). The first of these were unpublished entries of bacterial SRAP proteins, including *E. coli* YedK, from structural genomics groups. More recently, structures of the HMCES SRAP domain (residues 1-270 and lacking the 84-residue C-terminal tail) appeared (Fig. 2). Despite the handful of structures and the high conservation of this protein domain, the function of SRAP was unknown at their time of deposition. In 2019, 12 crystal structures from HMCES SRAP and YedK were published, 11 of which were in complex with DNA (Table 1). This wealth of new structural information provided insights into the mechanism of these fascinating proteins, as well as a more detailed understanding of their interaction with DNA.

3.3.1 SRAP architecture

The HMCES SRAP domain and YedK, which share 29% sequence identity and 43% similarity, are highly similar in structure, with an RMSD of 1.2 Å for all C_α atoms (PDB IDs 5KO9 and 2ICU). The only noticeable differences between the HMCES and YedK structures are in the loop regions, which have weaker sequence similarity and in HMCES are noticeably longer. HMCES residues 149-159, found only in sequences from higher eukaryotes, are disordered in all available HMCES SRAP structures. According to the SCOP (Structural Classification of Proteins) database *(82)*, SRAP domains possess a unique BB1717-like fold containing a bifurcated β-sheet surrounded by several α-helices. The β-barrel core forms a positively charged DNA binding channel, with the Cys2 active site in the middle (Fig. 12A). Not surprisingly, this channel is the most highly-conserved region of the SRAP domain *(46,83)*. The channel cradles the phosphoribose

backbone in all available SRAP-DNA structures. Two highly conserved surface-exposed arginine residues, Arg98 and Arg212 in HMCES (Arg77 and Arg162 in YedK) interact with phosphates of the DNA backbone and are required for DNA binding *(41,45,46,83)*.

Despite the differences in the DNA ligands used in the HMCES and YedK structures (Table 1), the observed confirmations of DNA and the DNA-protein interactions were remarkably similar, even between non-covalent and DPC complexes, indicating that SRAP interaction with DNA is likely not sequence-dependent. Several of the YedK-DNA structures (PDB IDs 6NUA, 6NUH, 6KBS, 6KBZ, 6KCQ) contained a continuous ssDNA molecule bound across the entire channel, which showed the details for how the DNA backbone is highly kinked and twisted at the position of the active site (Fig. 12A) *(46,83)*. The HMCES SRAP structures (PDB IDs 6OEA, 6OEB, and 6OE7) contained duplex DNA with a 3′-ssDNA overhang tail that bound to one side of the channel up to the active site (Fig. 12B) *(45)*. The duplex portion of a symmetry-related molecule bound to the other side, creating a semi-continuous strand that overlays with the continuous ssDNA in the YedK structures (Fig. 12B,C). A more recent HMCES SRAP structure (PDB ID 6OOV) crystallized with a palindromic DNA containing 3′ overhangs on each end. The overhangs link two symmetry related protomers by binding to one side of their positive channels in the same manner as the previous HMCES structures, but neither protomer contains DNA on the other side of the channel *(40)*.

Importantly, all of the structures showed that a conserved "wedge" motif protrudes into one side of the positive channel and disrupts nucleobase stacking, precluding a complimentary second strand from pairing immediately adjacent to the 5′ side the AP site (Fig. 12A,B). This structural feature, together with the sharp kink in the DNA at the active site, explains SRAP's preference for ssDNA *(41,45,46,83)*. Comparison of DNA-bound and free forms of HMCES SRAP and YedK show that there is very little change in the SRAP DNA binding site upon binding DNA (C_α RMSD of 1.16 Å for YedK and 1.00 Å for HMCES SRAP), indicating that restructuring of the DNA in this particular way is a key feature of the protein. Furthermore, structures of YedK bound non-covalently to DNA containing either a full nucleotide or different types of abasic sites showed almost no difference in the DNA conformation, despite their different crystal packing arrangements (Fig. 12D,E).

Table 1. SRAP protein crystal structures

Protein	**DNA** [a]	**PDB ID**	**Reference**
H. sapiens HMCES SRAP domain	None	5KO9	*(45)*
H. sapiens HMCES SRAP domain	Non-covalent overhang 5′-CCAGACGTTG-3′ 3′-GGTCTG-5′	6OEA	*(45)*
H. sapiens HMCES SRAP domain	Non-covalent overhang 5′-CCAGACGTT-3′ 3′-GGTCTG-5′	6OEB	*(45)*
H. sapiens HMCES SRAP domain	Covalent overhang 5′-CCAGACGT(AP)-3′ 3′-GGTCTG-5′	6OE7	*(45)*
H. sapiens HMCES SRAP domain	Non-covalent palindromic 5′-CAACGTTGTTTT-3′ 3′-TTTTGTTGCAAC-5′	6OOV	*(40)*
E. coli YedK	Covalent ssDNA 5′-GTC(AP)GGA-3′	6NUA	*(46)*
E. coli YedK	Non-covalent ssDNA 5′-GTC(C3)GGA-3′	6NUH	*(46)*
E. coli YedK	Covalent ssDNA 5′-AAA(AP)AA-3′	6KCQ	*(83)*
E. coli YedK	Covalent ssDNA 5′-TTC(AP)-3′	6KIJ	*(83)*
E. coli YedK	Covalent ssDNA 5′-CGGT(AP)-3′	6KBX	*(83)*
E. coli YedK	Non-covalent ssDNA 5′-GGT(THF)GATTC-3′	6KBZ	*(83)*
E. coli YedK	Non-covalent ssDNA 5′-GGTCGATTC-3′	6KBS	*(83)*
E. coli YedK	None	6KBU	*(83)*
E. coli YedK	None	2ICU	SECSG, RSGI [b]
Bordetella bronchiseptica Q7WLM8	None	1ZN6	NESG [c]
Bordetella bronchiseptica BB2244	None	2BDV	NESG [c]
Agrobacterium tumefaciens Atu5096	None	2AEG	NESG [c]
Bacteroides thetaiotaomicron BT1218	None	2F20	NESG [c]

[a] **AP, abasic site; C3, C3-spacer, THF, tetrahydrofuran**
[b]**SECSG,** Southeast Collaboratory for Structural Genomics; **RSGI** RIKEN Structural Genomics/Proteomics Initiative
[c] NESG, Northeast Structural Genomics Consortium

3.3.2 SRAP accommodates 3′-junction structures

In contrast to the kinked DNA 5' to the AP site by the wedge motif, all of the structures revealed that the DNA on the other side of the active site adopts a B-form conformation, and that duplex DNA can be accommodated immediately adjacent to the 3′-side of the AP site *(45,46,83)*. Such a dsDNA-ssDNA junction would be formed by a stalled replicative polymerase during DNA synthesis. Four structures explicitly showed dsDNA bound to this side of the positively charged channel via crystal lattice interactions. Two HMCES SRAP structures (PDB IDs 6OE7, 6OEB) contained the dsDNA portion of a symmetry-related molecule bound to the channel such that the blunt end of the duplex stacked against a highly conserved protein surface immediately adjacent to the active site, which we refer to as the DNA "shelf" (Fig. 12F). Similarly, in two YedK structures crystalized with ssDNA (PDB IDs 6KBS, 6KBZ), the DNA from one complex partially hybridizes to the DNA from a symmetry-related molecule, effectively establishing a dsDNA-ssDNA junction on the 3′ side of the AP site (Fig. 12G). The kinking of the continuous ssDNA along the channel allows for the nested 3′-end of the duplex to stack against the DNA shelf. The structures of the dsDNA and nested 3′-ends in HMCES and YedK are remarkably similar (Fig. 12H). An additional YedK-ssDNA structure (PDB ID 6NUA) did not contain dsDNA, but the ssDNA 3′ to the AP site adopted a B-form conformation such that duplex DNA could be easily modeled to create a dsDNA-ssDNA junction with a nested 3′-end *(46)*. The structure of this modeled 3′-junction is virtually identical to both the human and bacterial SRAP structures that contain duplex from different crystal packing interactions, indicating that this interaction is an inherent property of SRAP proteins. Interestingly, ssDNA 3′ to the AP site in two additional YedK-DPC structures (PDB IDs 6KIJ, 6KBX) and in a non-covalent YedK-DNA complex (PDB ID 6NUH) either did not exhibit sufficient density to be modeled or showed high B-factors, suggesting that ssDNA 3′ to the AP site is relatively mobile in the absence of a base-paired strand to stabilize that region of DNA. The preferential binding to 3′-junctions were verified biochemically in both HMCES and YedK *(45,46,83)*.

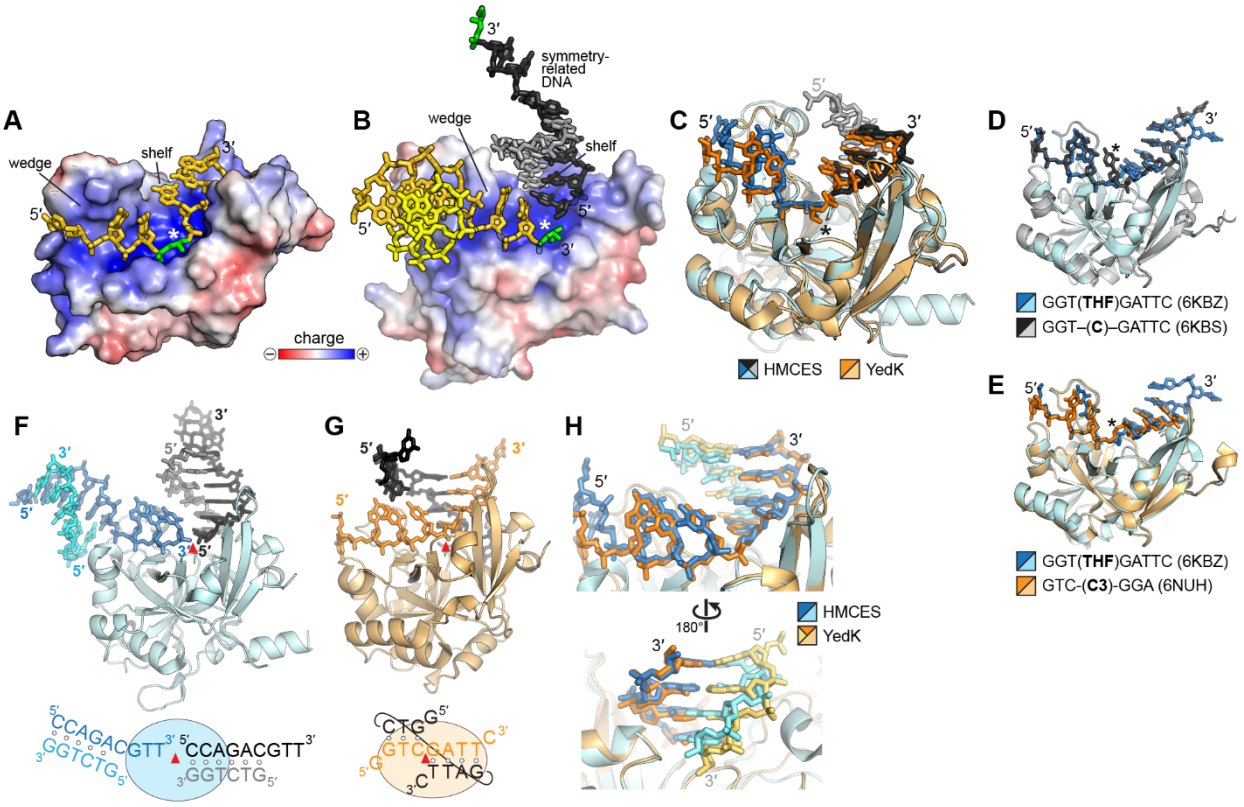

Fig. 12. SRAP-DNA structure. A,**B**. Electrostatic surface potential of (A) YedK-DPC (PDB ID 6NUA and (B) HMCES SRAP DPC (PDB ID 6OE7) structures. The AP sites are green, and the symmetry-related DNA molecule in HMCES is shown in black and grey. The white asterisk denotes the position of the active site. **C**. Superposition of YedK-DPC (PDB ID 6NUA) and HMCES SRAP DPC (PDB ID 6OE7). The DPC is marked with an asterisk. The ends of the DNA in the HMCES structure have been removed for clarity. **D**,**E**. Superposition of YedK bound non-covalently to ssDNA containing (D) a cytosine or an abasic site analog, and (E) two different abasic site analogs. The position of the nucleotide at the active site is marked with an asterisk. **F**,**G**. Similarity of dsDNA bound to (F) HMCES SRAP (PDB ID 6OEB) and (G) YedK (PDB ID 6KBS). The asymmetric unit is colored blue or orange, and symmetry related DNA is shown in black and grey. The red triangle marks the position of the active site. In the schematic at the bottom, base pairs are denoted by open circles. **H**. Superposition of the two structures in panels F and G. DNA is colored blue/cyan in HMCES and orange/gold in YedK.

3.4 The unique SRAP DNA-protein crosslink

The molecular basis for the stability of the SRAP DPC was elucidated by a 1.6-Å crystal structure of YedK crosslinked to AP-DNA (6NUA), which revealed a thiazolidine linkage between the ring-opened AP deoxyribose and the α-amino and sulfhydryl groups of the N-terminal Cys2 residue *(46)* (Fig. 13A). This specific DPC linkage was also observed in subsequent high-

resolution (1.2 Å) YedK-DPC structures *(83)* and in a 2.2-Å DPC structure of HMCES SRAP (PDB ID 6OE7, Fig. 13B), which revealed the conserved mode of SRAP crosslinking across domains of life *(45)*. Thiazolidine adducts between proteins and formaldehyde have been identified *(84,85)*, but to our knowledge a thiazolidine linkage between protein and DNA has not been described previously.

3.4.1 The SRAP active site

The residues responsible for DPC formation lie at the center of the DNA binding channel and include the invariant Cys2 and a highly conserved glutamate (Glu127 in HMCES; Glu105 in YedK), histidine (His210 in HMCES, His160 in YedK), and asparagine (Asn96 in HMCES; Asn75 in YedK). These residues are critical for crosslinking as substitution to alanine either abrogates or significantly reduces activity *(41,46,83)*. Their interactions with the AP site are conserved between HMCES and YedK structures (Fig. 3A,B). In the absence of DNA, the thiol side chain of Cys2 points down into the protein, with the N-terminal amine position varying (HMCES, PDB ID 5KO9; YedK, PDB ID 6KBU). In one structure (YedK, PDB ID 2ICU) Cys2 was not observed in the crystallographic data, suggesting a degree of disorder potentially due to the presence of an N-terminal tag. When either HMCES or YedK is bound to DNA containing a natural base (PDB IDs 60EA, 60EB, 600V, 6KBS), Cys2 is observed in multiple positions and the nucleobase is facing away from the active site *(45,83)*. In non-covalent complexes of YedK with an AP site analog, THF or C3-spacer (PDB IDs 6KBZ, 6NUH), Cys2 is rotated 180° relative to its buried position in the unbound structures, restructured into a pre-catalytic, crosslinkable position *(46,83)*. Notably, the backbone of the THF structure is twisted by 90º such that the THF is positioned towards the Cys2, similar to the DNA conformation seen in the DPC structures.

The histidine side chain hydrogen bonds to the O4′ hydroxyl of the ring-opened AP site in the DPC structure but not in the non-covalent THF structure. The structure of YedK noncovalently bound to DNA with a nucleobase in the active site position shows the nucleotide rotated such that His160 now forms a hydrogen bond with O3′ as opposed to O4′ in the DPC structure *(83)*. The structure of HMCES SRAP bound to palindromic DNA containing 3′ overhangs (PDB ID 6OOV) shows the 3′-OH of each overhang hydrogen bonded to His210 *(40)*. In the various DPC structures, the glutamate side chain is in close contact to the thiazolidine ring, anywhere from 2.6-4.2 Å observed in the various DPC structures. In the YedK DPC structures, a second glutamate

conformer is observed within hydrogen bonding distance of the phosphate 3′ to the AP site, and is thus likely protonated to avoid repulsion with the negative backbone *(46,83)*. On the other side of the thiazolidine crosslink, the asparagine hydrogen bonds with the carbonyl oxygen and the backbone amide nitrogen of Cys2.

3.4.2 Catalytic mechanism of SRAP DPC formation

Based on the SRAP-DNA structures and previous work on thiazolidine chemistry *(55,56,58,86)*, we propose the following mechanism for SRAP DPC formation at AP sites. Thiazolidine formation would proceed by nucleophilic attack of the AP site C1′ carbon by the N-terminal Cys2 α-amino group to form a Schiff base intermediate, followed by a second nucleophilic attack of the imino carbon by the Cys2 thiolate side chain (Figs. 13C, 14A). Because of the low abundance of the aldehyde form of an AP site *(21)*, it is possible that SRAP first catalyzes AP site ring opening, although the alternative in which SRAP traps a spontaneously formed aldehyde is also possible. Based on their positions in the crystal structures, Glu127 and His210 could catalyze ring opening by providing a general base to deprotonate the O1′ hydroxyl and a general acid to protonate O4′, both of which are necessary for aldehyde formation. The glutamate would then facilitate Schiff base formation by acid-base cycling to deprotonate Cys2 α-NH_2 and possibly protonate the AP site O1′. Glu127 would also be positioned to deprotonate the Cys2 sulfhydryl to create the thiolate nucleophile required to complete formation of the thiazolidine (Fig. 13C). A cycle of Glu127 ionization is supported by the two observed conformations—an ionized conformer contacting the thiazolidine nitrogen and a protonated conformer contacting the phosphate 3′ to the AP site *(46)*. The hydrophobic residues around the active site would raise the pK_a of the glutamate, increasing the likelihood that it acts as both a proton donor and acceptor in the acid-base cycle. Finally, based on its position in the structures, the highly conserved asparagine likely positions the N-terminal Cys2 for nucleophilic attack.

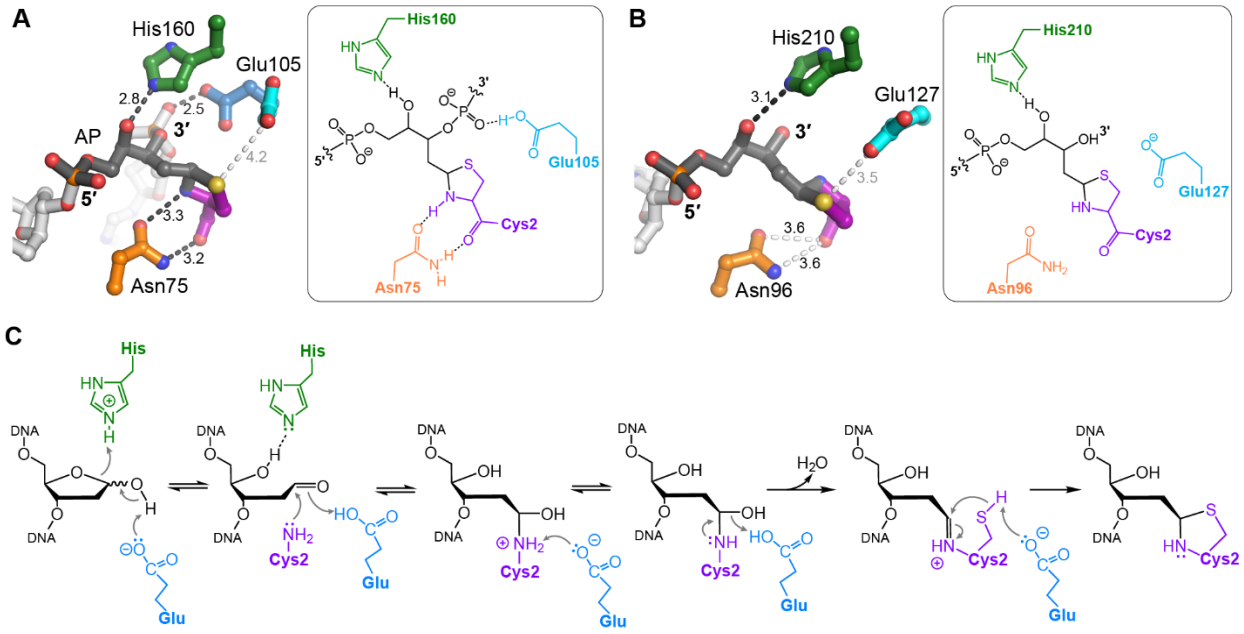

Fig. 13. SRAP active site and mechanism of crosslinking. A,**B**. Atomic details of the thiazolidine DPC in (A) YedK (PDB ID 6NUA) and (B) HMCES SRAP (PDB ID 6OE7). DNA is greyscale with AP site dark grey, and SRAP is colored by amino acid. Interatomic distances (Å) are labeled, with hydrogen bonds indicated by dark dashes and close contacts with light dashes. **C**. Proposed catalytic mechanism of crosslinking.

3.4.3 Thiazolidine stability

The Cys2 active site is surrounded by a pocket of strongly conserved hydrophobic residues. The solvent inaccessibility of this conserved DNA binding pocket likely provides an optimal environment for crosslink formation and helps to shield it from hydrolysis. However, shielding is not the sole basis for crosslink stability as proteolysis of the DPC down to a peptide-DNA crosslink is also persistent *(46)*. Indeed, the chemical nature of the thiazolidine linkage is inherently stable and its formation proceeds efficiently under acidic conditions, as revealed by reactions between free cysteine and various aldehydes *(55,58,87,88)*. The thiazolidine ring is stable up to at least 3M HCl, and reversal back to aldehyde and free cysteine is only observed under basic conditions greater than 1 M NaOH *(86)*. This stability has also been exploited in conjugation reactions to generate peptide-peptide linkages or to attach site-specific labels *(89,90)*.

The stability of the thiazolidine ring explains how SRAP protects AP sites from spontaneous strand breakage. Both AP sites and their Schiff base conjugates to proteins are susceptible to β-elimination, leading to cleavage of the DNA backbone (Figs 11B, 14).

Thiazolidines exist in equilibrium with the Schiff base, but the ring-closed thiazolidine is favored over the ring-opened iminium ion by 5 orders of magnitude *(55,56,58)*. Thus, the closed thiazolidine ring draws the equilibrium away from Schiff base degradation (Fig. 14A). Consistent with this premise, removal or substitution of the thiolate by a YedK C2A or C2S mutant not only abrogated crosslink formation but also led to cleavage of the DNA to form a product consistent with β-elimination at the AP site (Fig. 14B) *(46)*. Evidence for the Schiff base intermediate in SRAP was provided by borohydride trapping *(63)* of YedK C2A and C2S DPC intermediates (Fig. 14B), which dramatically reduced DNA strand cleavage *(46)*. Formation of a Schiff base also indicates that the N-terminal amine—not the Cys2 thiolate—initiates crosslink formation. Indeed, the hydrophobic environment of this residue would raise the pK_a of the thiol and lower the pK_a of the amine, favoring the amine as the nucleophile.

A C2S mutant could theoretically form an oxazolidine ring by the same mechanism as thiazolidine formation by Cys2. To our knowledge, however, there is no evidence of a SRAP C2S variant in nature. We did not observe any DPC formation in our biochemical experiments with a C2S mutant. An oxazolidine is less stable than a thiazolidine *(56,57)* likely because sulfur is 0.4 Å larger than oxygen, and can form longer bonds that could aid in the stability of the thiazolidine *(91,92)*. Moreover, the side chain of cysteine has a lower pK_a than that of serine, making cysteine a better nucleophile. Thus, an oxazolidine is unlikely to form, and if it did form, it would be unable to shift the equilibrium away from the competing β-elimination reaction.

3.4.4 Comparison to DPC formation in DNA lyases

The N-terminal cysteine in SRAP is unique among DNA repair proteins. A number of repair enzymes interact with AP sites via Schiff base intermediates, but typically do so to catalyze strand scission *(93-96)*. Bifunctional DNA glycosylases contain AP lyase activity that cleaves the DNA backbone 3′ to the AP site *(97)*. Pol β and Ku antigen contain 5′-deoxyribophosphate (5′-dRP) lyase activities important for single-strand break processing during BER and non-homologous end joining (NHEJ), respectively *(30,98,99)*. PARP1 and PARP2 have been shown to exhibit AP lyase and 5′-dRP lyase activity *in vitro (100,101)*. Similar to the SRAP crosslinking mechanism, DNA lyases use an amine nucleophile—either an internal lysine or an N-terminal proline or valine—to form a Schiff base intermediate that can be trapped by borohydride reduction *(63,80,102)*. Lyases lack a second nucleophile to stabilize the DPC, and instead the Schiff base

increases the basicity of C2′ to promote catalysis of β-elimination and cleavage of the DNA (Fig. 4C). Thus, the main distinction between AP lyases and SRAP is the ability of the SRAP crosslinking nucleophile to exhibit a second nucleophilic attack of the imino C1′ carbon to form a stable linkage, thereby acting as an AP-site sink to inhibit the competing elimination reaction. Wild type SRAP does not lead to, but rather *prevents* strand cleavage.

Interestingly, some AP lyases form stable DPCs with oxidized AP sites in mammalian cells, exhibiting half-lives of 15-60 min before proteasome dependent resolution *(76,103,104)*. These AP lyase DPCs were associated with an increase in double-strand breaks in cells, in contrast to the HMCES dependent reduction in double-strand breaks *(41,98)*. Thus, whereas HMCES DPCs play a protective role, the oxidative AP lyase DPCs are thought to be toxic byproducts of DNA repair proteins trapped in unproductive complexes *(76)*. In contrast, PARP1 has been shown to form a DPC with AP sites by reduction of the Schiff base via an intrinsic redox activity of the protein, similar to the reaction in borohydride trapping experiments. Like HMCES, the PARP1 DPC is proposed to protect AP sites and recruit BER factors for repair *(105)*.

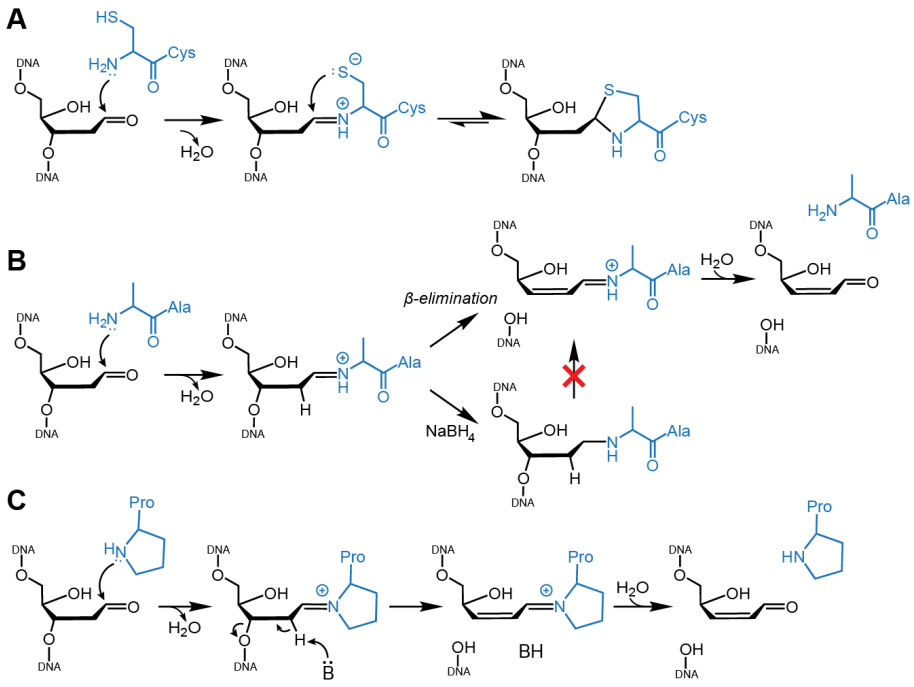

Fig. 14. Comparison of DPCs formed by SRAP and AP lyases. A. Wild-type SRAP forms a thiazolidine DPC via a Schiff base intermediate. **B**. The Schiff base formed by a SRAP C2A mutant is prone to β-elimination and can be reduced to a stable DPC via borohydride treatment. **C.** Catalysis of DNA lyase activity by bifunctional DNA glycosylases.

3.5 Discussion

This review has focused primarily on the mechanics of SRAP crosslinking to AP sites. However, in addition to its role in the replication stress response *(41,106)*, HMCES is implicated in other DNA related transactions across species, including viral replication *(107)*, aging *(108)*, alternative end-joining during B-cell class switch recombination *(40)*, and long interspersed element-1 (LINE-1) retrotransposition *(109-111)*. Moreover, HMCES expression is significantly down regulated in both B-cell lymphoma and multidrug resistant osteosarcoma cells *(112,113)*. HMCES was named for its identification in a screen for readers of modified bases in embryonic stem cells *(39)*, although this name is likely a misnomer as HMCES is expressed in all cell lineages *(114)* and its binding to oxidatively modified bases has not been reproduced. Moreover, global DNA methylation patterns are not significantly altered in HMCES knockouts *(40,41)*, and the catalytic SRAP domain is conserved in prokaryotes, which do not utilize 5hmC *(115)*.

An important open question relates to the fate of the HMCES DPC in cells. HMCES is ultimately targeted for ubiquitin-dependent destruction by the proteasome *(41)*. Is the resulting peptide-DNA crosslink repaired by nucleotide excision or some other repair pathway? In HMCES, several lysine residues located in the C-terminal tail and on the surface of the protein opposite the DNA binding channel have been found to be ubiquitylated via mass spectrometry *(116,117)*. The majority of these lysines are highly conserved in eukaryotic SRAP proteins. These lysines and other predicted phosphorylation and SUMOylation sites are mutated in some human tumors *(116,118)*. Other cancer associated mutations appear to be destabilizing mutations as they reside in the protein interior and likely result in improper folding and/or subsequent degradation. HMCES regulation and the fate of SRAP-AP DPCs in cells represent the next frontier in understanding this important class of protein.

---

[4] This work was originally published as: Amidon, K. M., and Eichman, B. F. (2020) Structural biology of DNA abasic site protection by SRAP proteins, *DNA Repair* ***94***, 102903.
I wrote and edited the manuscript

CHAPTER 4

Abasic site ring opening and DNA-protein crosslink reversal by the SRAP protein YedK[5]

4.1 Introduction

Apurinic/apyrimidinic (AP, or abasic) sites are one of the most ubiquitous DNA lesions. AP sites arise from either spontaneous or DNA glycosylase-catalyzed hydrolysis of the *N*-glycosidic bond that links the modified base to the deoxyribose *(13,14)*. Their impact on cellular processes results in large part from their instability and reactivity. In solution, an AP site exists as an equilibrium between a predominant cyclic furanose as a mixture of α- and β-hemiacetals and a ring-opened aldehyde form, the latter constituting approximately 1% of the total *(21,22,119)*. This electrophilic aldehyde can react with exocyclic groups of nucleobases on the complimentary strand to generate interstrand DNA crosslinks (ICLs) *(19,24)* and with primary amines in proteins to generate DNA-protein crosslinks (DPCs) *(18)*. The ring opened aldehyde form is also susceptible to base-catalyzed β-elimination of the 3′ phosphoryl group, generating a single-strand break *(23)*. AP sites occurring in ssDNA, such as those encountered during replication, can lead to stalled replication forks by inhibiting replicative polymerases *(14,16,17,25)*. Replication forks that encounter an AP site on the template strand can lead to a double-strand break (DSB) *(14,16,33,61)*.

AP sites in double-stranded (ds)DNA are repaired by the base excision repair (BER) pathway, but the fate of AP sites in ssDNA is not as well understood. During replication, AP sites can be bypassed by error-prone translesion synthesis (TLS) polymerases *(25,31-33)*. Recently, an alternative, higher fidelity pathway for repair of replication-associated AP sites was discovered that involves the protein HMCES *(41,106)*. Cells lacking HMCES exhibit elevated levels and delayed repair of AP sites, as well as increased double-strand breaks and mutation frequency from TLS *(41)*. Further supporting that HMCES responds to AP lesions, HMCES-deficient cells are hypersensitive to nuclear expression of APOBEC3A, which catalyzes deamination of cytosine to uracil in ssDNA that is converted to an AP site after removal by UNG *(71,120)*.

HMCES forms DPCs with AP sites in ssDNA but not dsDNA *(41)*, which led to a model in which the HMCES DPC protects AP sites from nuclease cleavage and mutagenic TLS

polymerases *(41,71)*. *In vitro,* both intact and proteolyzed HMCES DPCs are resistant to cleavage by AP endonuclease 1 (APE1) *(41,46)*. Typically, proteins covalently conjugated to DNA occur either as deleterious lesions *(74-76)* or as a catalytic intermediate in DNA strand cleavage (lyase) reactions *(77-79)*. By contrast, the HMCES DPC is highly stable and persists in cells on the order of hours, and has been shown to ultimately be resolved by a proteolytic-dependent mechanism under specific conditions *(41,46)*. In *Xenopus* extracts, HMCES DPCs form as intermediates in AP-ICL repair upon NEIL3 unhooking of the AP-ICL, and are substrates for SPRTN protease, which generates a DNA-peptide crosslink (DpC) *(121,122)*. However, the mechanism by which HMCES DPCs are resolved in mammalian cells remains to be determined.

HMCES contains a catalytic SRAP (SOS Response Associated Peptidase) domain that is conserved across all domains of life *(36,41,46)*. HMCES SRAP is similar in both sequence (29% identity / 43% similarity) and structure to *E. coli* YedK, with the highest degree of conservation within the DNA binding channel and at the active site *(45,46)*. An invariant cysteine at amino acid position 2 (Cys2) constitutes the extreme N-terminus after aminopeptidase removal of Met1 *(38,41)* and is required for DPC formation *in vivo* and *in vitro*. Crystal structures of HMCES SRAP and YedK crosslinked to AP-DNA revealed that Cys2 forms a highly stable thiazolidine linkage with the ring-opened aldehyde form of the AP site *(40,45,83,123)*, which helped explain the persistence of HMCES DPCs in cells. The SRAP active site contains highly conserved glutamate, histidine, and asparagine residues that contact the crosslinked AP site *(36,39,41)*. Mutation of these residues reduces crosslinking activity without disrupting DNA binding activity *in vitro (38,40,41,46,83)* and increases sensitivity to oxidative stress or ionizing radiation in cells *(41,106,124)*.

Despite the importance of HMCES in repair of AP sites, the mechanisms of DPC formation and resolution, and the roles of active site residues in these processes, are unknown. Here, we perform a biochemical and crystallographic analysis of the various steps involved in catalysis of DPC formation, using YedK as a model system. Our data provide evidence for AP site ring opening and Schiff base formation, both of which are necessary precursors to thiazolidine formation. The active site glutamate is involved in both processes, and the histidine contributes to ring opening. We find that YedK forms DPCs to cleaved DNA 3′-ends generated by DNA lyases. We also show that YedK catalyzes DPC reversal to reform a free AP site on the order of several hours *in vitro*, which has implications for resolution of the HMCES DPC in cells.

4.2 Results

4.2.1 Glu105 and His160 enable acid-base catalysis of DPC formation

The Cys2-linked, ring-opened AP site is stabilized by highly-conserved histidine, glutamate, and asparagine residues (Fig. 15A). In YedK, Asp75 hydrogen bonds to the carbonyl oxygen and the backbone amide nitrogen of Cys2, His160 hydrogen bonds to the hydroxyl (O4') of the ring-opened AP site, and Glu105 interacts with the thiazolidine ring and with His160 *(40,45,83,123)*. In our crosslinked YedK structures, a second conformer of Glu105 was observed in which the carboxylate hydrogen bonds with the phosphate 3′ to the AP site, strongly implying that Glu105 exists at least transiently in a fully protonated state *(46)*. Previous mutational analyses of SRAP active site residues involved only alanine substitution and were performed at a single time point *(41,46,83)*. To gain a more detailed understanding of the roles of the SRAP active site residues, we performed a kinetic analysis of variants that altered their hydrogen bonding or ionization potential. We verified by mass spectrometry that these mutants all lack the N-terminal methionine (Table 2). Thus, the active site residues do not play a role in N-terminal methionine removal from the bacterial protein, contrary to a previous report using mammalian HMCES expressed in HEK293T cells *(38)*.

Crosslinking kinetics were measured under single turnover conditions using a ssDNA oligo containing a centrally located AP site and a 5′-FAM label for visualization. In our assay, the rate of wild-type YedK DPC formation is a lower limit, as the reaction was nearly complete at our fastest time point. We first tested the kinetics of alanine point mutants (Fig. 15B,C). Surprisingly, Asn75, which was expected to position the N-terminal Cys2 for nucleophilic attack based on the structures, showed only a very modest decrease in crosslinking relative to wild-type YedK when mutated to alanine (Fig. 15C,F). In contrast, H160A exhibited at least 10-fold reduction in rate relative to wild-type. Alanine substitution of Glu105 had the largest effect of the three active site residues. The E105A crosslinking reaction was only 50% complete after 1 hour (Fig. 15B-D). These data are consistent with Glu105 and His160 as important for SRAP AP-site crosslinking, with Glu105 playing an essential role.

Table 2. ESI-MS intact mass values to verify removal of N-terminal methionine.[a, b, c]

		With Met1			Without Met1		
	Observed MW	**Calculated MW**	**Mass difference**	**% difference**	**Calculated MW**	**Mass difference**	**% difference**
WT	25572.5	25709.14	136.64	0.53	25577.95	5.45	0.02
N75A	25527.8	25666.11	138.31	0.54	25534.92	7.12	0.03
E105A	25511.9	25651.1	139.2	0.54	25519.91	8.01	0.03
E105D	25589.02	25695.11	106.09	0.41	25576.96	12.06	0.05
E105Q	25577.5	25708.15	130.65	0.51	25576.96	0.54	0.00
H160A	25529.6	25643.08	113.48	0.44	25511.88	17.72	0.07
H160E	25552.7	25701.11	148.41	0.58	25569.92	17.22	0.07
H160Q	25578.8	25700.13	121.33	0.47	25568.93	9.87	0.04
C2A	25540.1	25677.08	136.98	0.53	25545.89	5.79	0.02
C2A/E105Q	25520.8	25676.09	155.29	0.60	25544.9	24.1	0.09
C2A/H160Q	25527.8	25668.07	140.27	0.55	25536.87	9.07	0.04

[a]All masses reported in Da

[b]Mass difference = Observed MW- Calculated MW

[c] % difference = Mass difference / Observed MW

The proximity of Glu105 and His160 to the crosslink and to each other suggest that they participate in acid-base catalysis *(123)*. We therefore examined the crosslinking kinetics of a E105Q and H160Q mutants, which cannot participate in acid-base chemistry but retain the same hydrogen bonding potential as the wild-type enzyme (Figs. 15D-F, 16). As with the alanine mutant, E105Q severely impacted YedK activity (Fig. 15D,F; Table 3), strongly suggesting that ionization of the carboxylate is important for DPC formation. Consistently, an E105D mutant only modestly impacted catalysis. Both E105A and E105Q exhibited biphasic kinetics at pH 6 with short burst (k_{fast}) and prolonged slow (k_{slow}) phases, which are at least 2.5-fold and 1,000-fold slower than wild-type (Fig. 15D,F; Table 3). The H160Q substitution also reduced the crosslinking rate 10-fold (similar to H160A), whereas an H160E mutant exhibited only a 3-fold reduction in crosslinking rate compared to wild type (Fig. 15E,F; Table 3). Thus, both Glu105 and His160 likely participate in acid-base catalysis rather than merely stabilize the substrate or transition state via hydrogen-bond stabilization. Consistent with this, YedK exhibits a strong pH dependence on the crosslinking rate with maximal activity at lower pH (Fig. 15G; Fig. 16B). The apparent midpoint of 5.1 in the pH profile is consistent with Glu105 acting as a general acid. The

hydrophobic residues around the active site would raise the pK_a of the glutamate, increasing the likelihood that it acts as both a proton donor and acceptor.

Based on the mutational data and the configuration of active site residues around the AP site, we propose the following catalytic mechanism for DPC formation in three main phases: AP site ring opening, Schiff base formation, and thiazolidine formation (Fig. 15H). In the first phase, both Glu105 and His160 likely catalyze ring opening of the AP deoxyribose ring from the furan to aldehyde form, whereby H160 acts as a general acid to protonate O4' and Glu105 acts as a general base to deprotonate the hydroxide at C1'. In the second phase, Glu105 drives Schiff base formation by acting as both general acid and base to deprotonate Cys2 α-NH_2 and to hydrolyze the hydroxyl at C1'. In the final step, Glu105 deprotonates the Cys2 sulfhydryl group necessary to close the thiazolidine ring.

Table 3. YedK mutant crosslinking rates[a]

YedK mutant	k_1 (min^{-1})	k_2 (min^{-1})	Fold change (relative to WT)
WT [b]	10.3 ± 1.6		1
N75A	7.1 ± 1.5		0.7
E105A	4.1 ± 2.1	0.01 ± 0.01	0.4 | 0.001
E105D	6.3 ± 0.7		0.6
E105Q	1.2 ± 0.7	0.02 ± 0.02	0.1 | 0.002
H160A	1.1 ± 0.1		0.1
H160E	3.5 ± 1.1		0.3
H160Q	0.9 ± 0.1		0.1

[a] Reactions were carried out at pH 6.0, 25 ºC. Values are mean ± SD (n=3)
[b] Value for wild-type is a lower limit

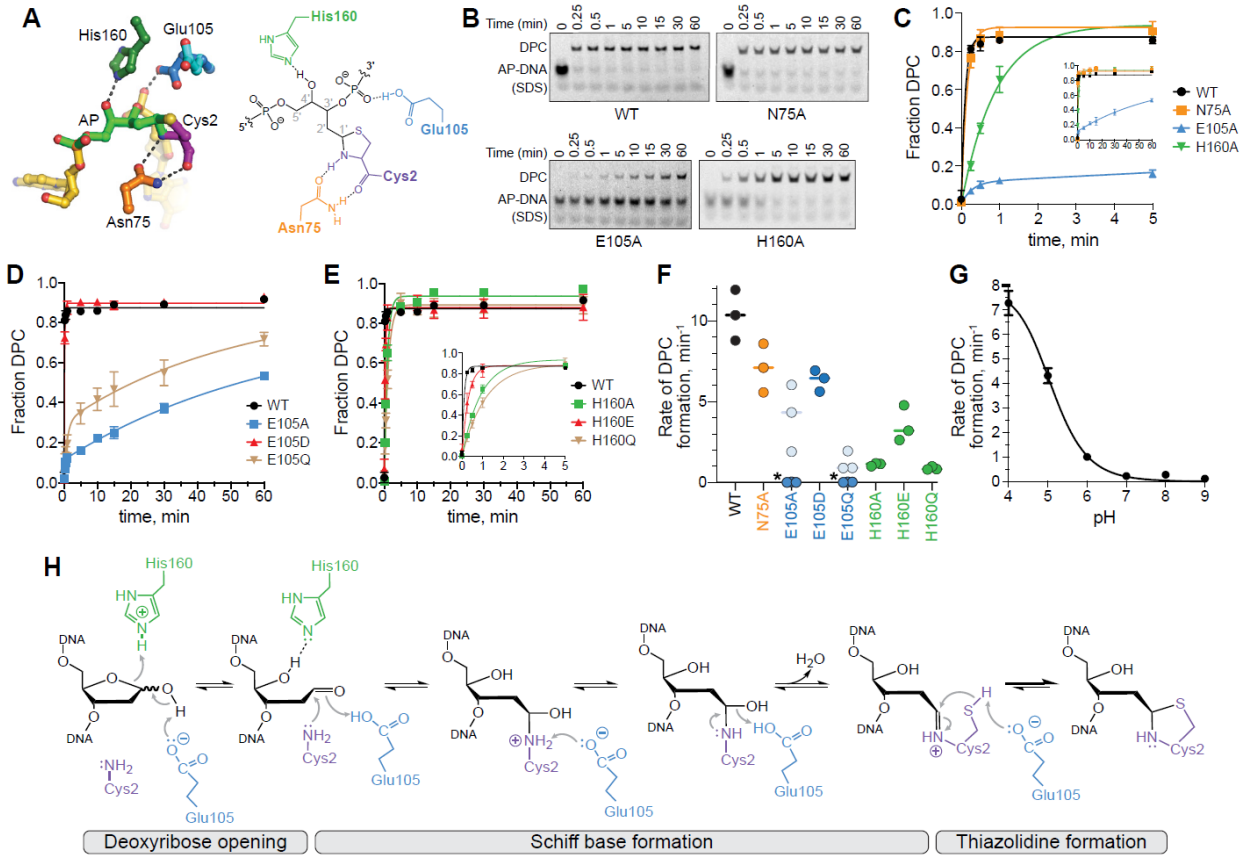

Figure 15. Glu105 and His160 enable acid-base catalysis of DPC formation. A. Active site of YedK DPC crystal structure (left) and schematic (right). **B**. Representative SDS-PAGE separation of uncrosslinked and crosslinked AP-DNA by wild-type and alanine mutant YedK. Crosslinking experiments were performed at 25 ºC and pH 6. DNA bands were visualized with by FAM fluorescence. **C,D,E**. Kinetics of DPC formation of active site alanine mutants (C), Glu105 mutants (D), and His160 mutants (E) at 25 ºC and pH 6 (mean ± SD, n=3). **F**. Rate constants derived from data in panels C-E. E105A and E105Q data were fit to a 2-phase exponential; k_{fast} is shown in blue and k_{slow} is cyan. Mean ± SD values are shown in Table S2. **G**. pH dependence of DPC formation for wild-type YedK at 18 ºC (mean ± SD, n=3). Kinetic traces are shown in Fig. 16. **H**. Proposed catalytic mechanism of YedK DPC formation.

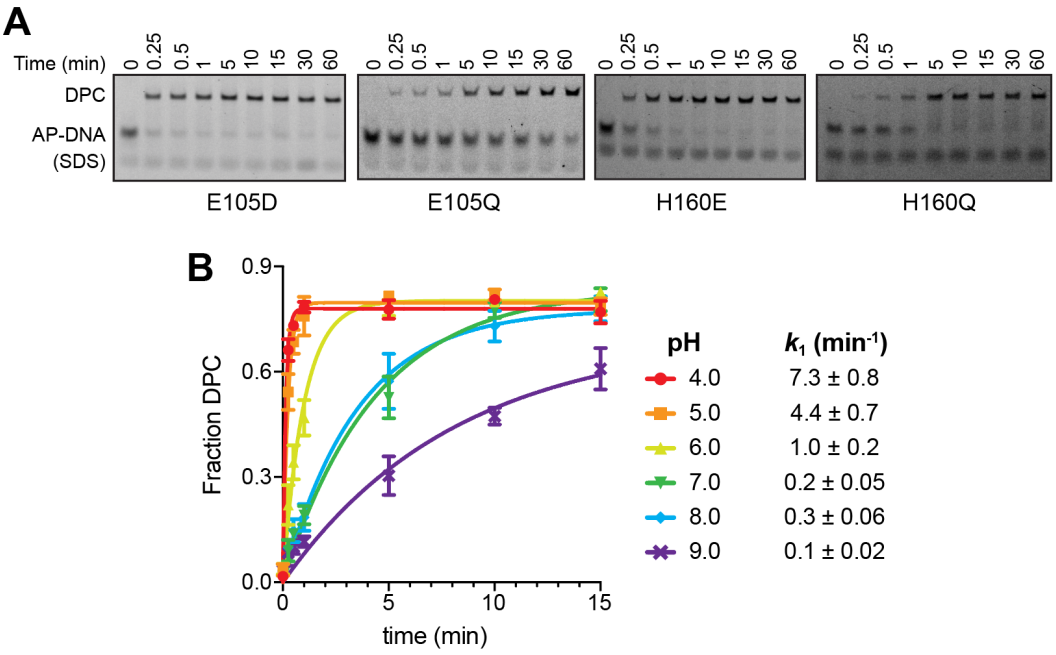

Figure 16. **YedK crosslinking kinetics. A**. Representative denaturing PAGE gels of AP-DNA crosslinking by YedK point mutants. Bands were visualized by FAM fluorescence. **B**. Kinetics of crosslinking at different pH (mean ± SD, n=3). Reactions were performed at 18 ºC. Rate constants derived from exponential fits to the curves are shown on the right and plotted in Fig. 15G.

4.2.2 YedK catalyzes AP site ring-opening

In solution, the AP site is at equilibrium between a ring-closed 2′-deoxy-D-erythro-pentofuranose and a ring-opened aldehyde. The AP site exists primarily in the cyclic furanose form with only 1% of the sugar in the more reactive ring-opened aldehyde form *(21,22)*. To investigate whether SRAP domains actively catalyze opening of the furan ring or simply capture a spontaneously formed aldehyde, we compared the rates of crosslink formation by two non-enzymatic probes to that of wild-type YedK under single-turnover conditions (Fig. 17A-C; Fig. 18A-C). The two non-enzymatic probes used were a YedK peptide consisting of the first 15 residues and including the N-terminal Cys2, and an aldehyde reactive probe, aoN-g, which reacts specifically to the aldehyde form of the AP site via an oxime linkage *(125)*. The rates of YedK peptide and aoN-g probe crosslinking were 0.09 ± 0.005 min^{-1} and 0.03 ± 0.001 min^{-1}, respectively, compared to 19.7 ± 2.4 min^{-1} for YedK. The 200-500-fold reduced rate in crosslinking by the two non-enzymatic probes suggests that YedK catalyzes AP site ring opening.

To test this further, we selectively blocked the N-terminal Cys2 with formaldehyde, which reacts more efficiently with cysteine than other amino acids to form a thiazolidine ring

(55,58,85,126) (Fig. 17D). Our proposed mechanism predicts that the α-NH_2 group of Cys2 initiates DPC formation after the first step of AP site ring-opening (Fig. 15H). Thus, formaldehyde blocking of the N-terminus renders Cys2 unreactive toward the AP site, allowing us to examine the effects of Glu105 and His160 on the ring-opening step. As expected, blocking Cys2 in wild-type YedK inhibited DPC formation and led to strand cleavage (Fig. 17E), consistent with spontaneous β-elimination of the AP aldehyde previously observed with a C2A mutant *(46)* (Fig. 17F). In contrast, strand cleavage was not observed in Cys2-blocked E105Q and H160Q proteins, indicating that these residues are essential for formation of the reactive AP aldehyde. We verified that the loss of β-elimination in the formaldehyde-treated mutants was not the result of reduced DNA binding (Fig. 18D). Combined with the reduced rates of crosslinking by non-enzymatic probes, these data are consistent with Glu105- and His160-catalyzed AP ring opening by SRAP.

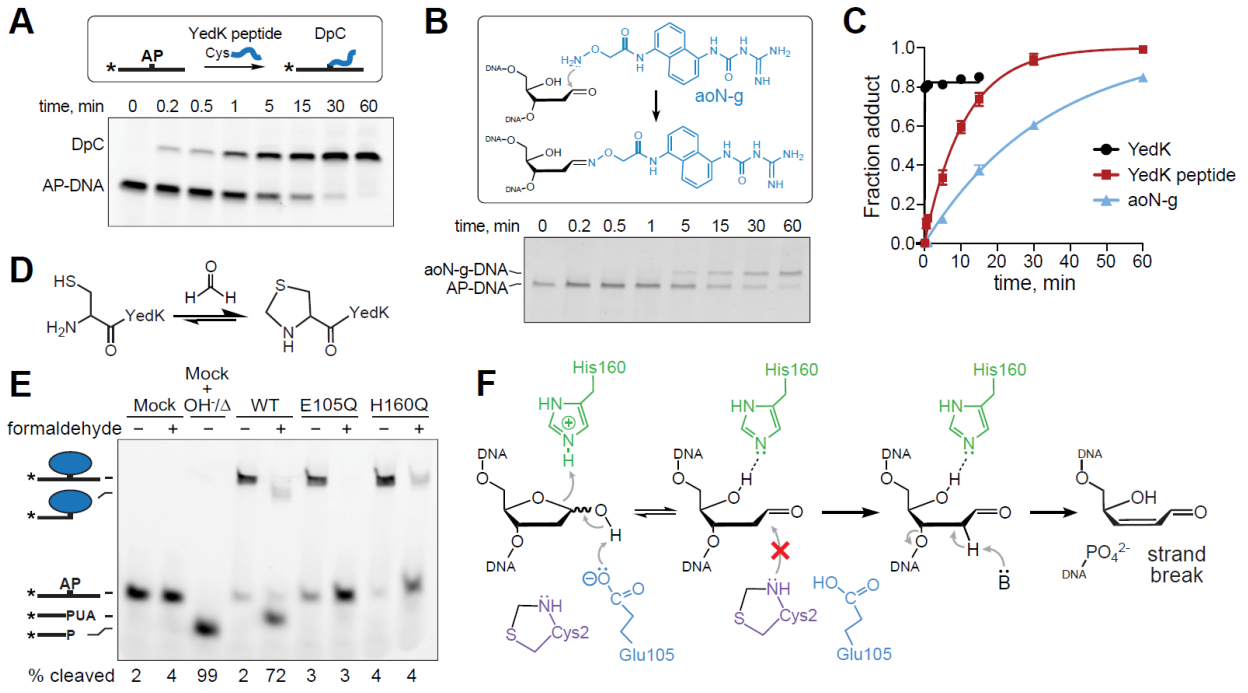

Figure 17. SRAP catalyzes AP site ring-opening. A. SDS-PAGE separation of AP-DNA crosslinked by YedK N-terminal peptide. **B**. Reaction of aldehyde reactive probe aoN-g to AP-DNA. **C**. Quantification of YedK peptide and aoN-g reaction with AP-DNA, compared to wild-type YedK (mean ± SD, n=3). **D**. Formaldehyde reacts with the YedK N-terminal Cys2 to form a thioazolidine. **E.** SDS PAGE of AP-DNA incubated with either buffer (mock) or native or formaldehyde-blocked YedK at 37 ºC for 1 hour. PUA, 3′-phospho-α,β-unsaturated aldehyde (β-elimination product); P, 3′-phosphate (β,δ-elimination product). **F**. Blocking Cys2 with formaldehyde prevents YedK crosslinking to the ring-opened AP-site, leading to strand breakage.

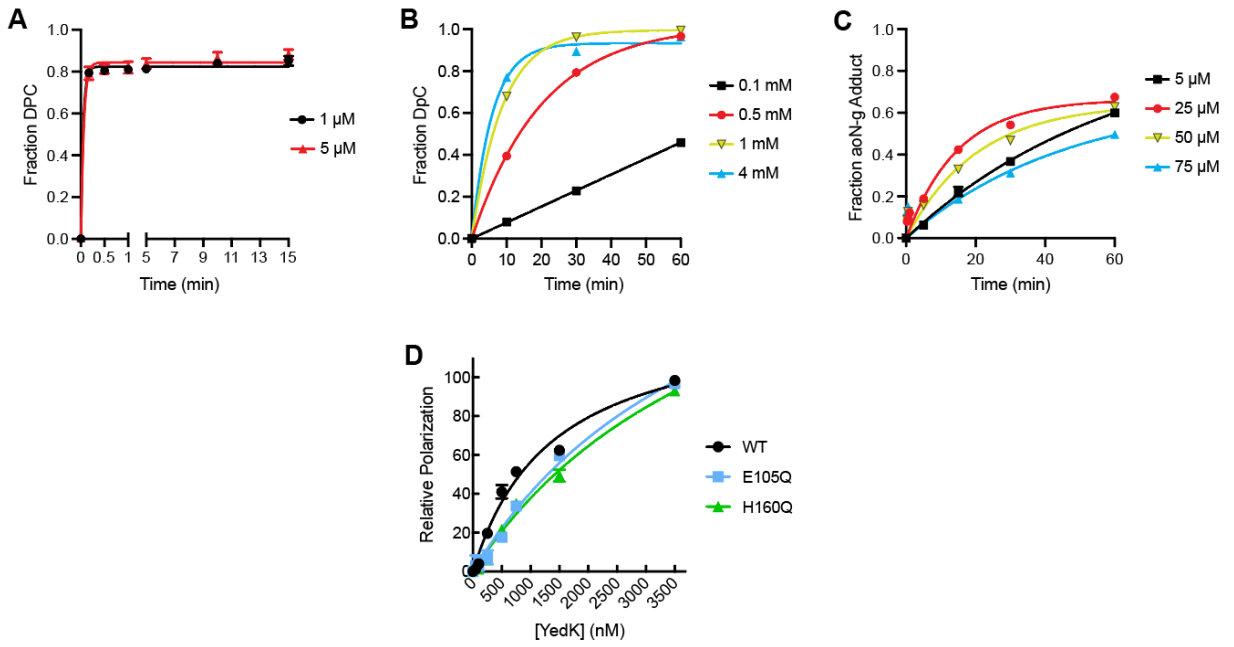

Figure 18. **Determination of saturating conditions for crosslinking. A-C.** Determination of saturating conditions for crosslink formation using increasing concentrations of (A) YedK, (B) YedK peptide, and (C) aoN-g probe. **D.** Binding of formaldehyde-blocked YedK to ssDNA containing a 5′-FAM label and a centrally located tetrahydrofuran (THF) abasic site analog. Binding was monitored by a change in fluorescence polarization as protein was titrated against DNA. Dissociation constants (K_d) derived from non-linear least squares fit of the data are 1.4 μM ± 0.007 (WT), 4.5 ±0.02 (E105Q), and 4.9 ± 0.02 (H160Q).

4.2.3 Glu105 catalyzes formation of the Schiff base intermediate

SRAP DPC formation likely proceeds through a Schiff base intermediate formed by nucleophilic attack of C1′ of the AP site by the α-amino group of Cys2 *(46,55,57,123)*. In the absence of the Cys2 thiolate side-chain, SRAP does not form DPCs *(41,46,83)* and instead generates DNA cleavage products indicative of DNA lyase activity (Fig. 19A). We previously provided evidence for the Schiff base intermediate by borohydride trapping *(63)* of DPCs in YedK C2A and C2S mutants *(46)*. To visualize this Schiff base intermediate, we determined a 1.82-Å crystal structure of YedK C2A with AP-DNA in the presence of borohydride. The electron density shows a linear linkage consistent with a reduced imine between the N-terminal alanine and the ring-opened AP site (Fig. 19B). This structure is highly similar to that of the wild-type YedK DPC (PDB ID 6NUA), with an RMSD of 0.71 Å for all C_α atoms. The DNA binding modality observed in the C2A DPC structure contains the same 90º kink and twist in the DNA backbone at the AP site observed in other SRAP-DNA structures *(40,41,83,123)* (Fig. 20).

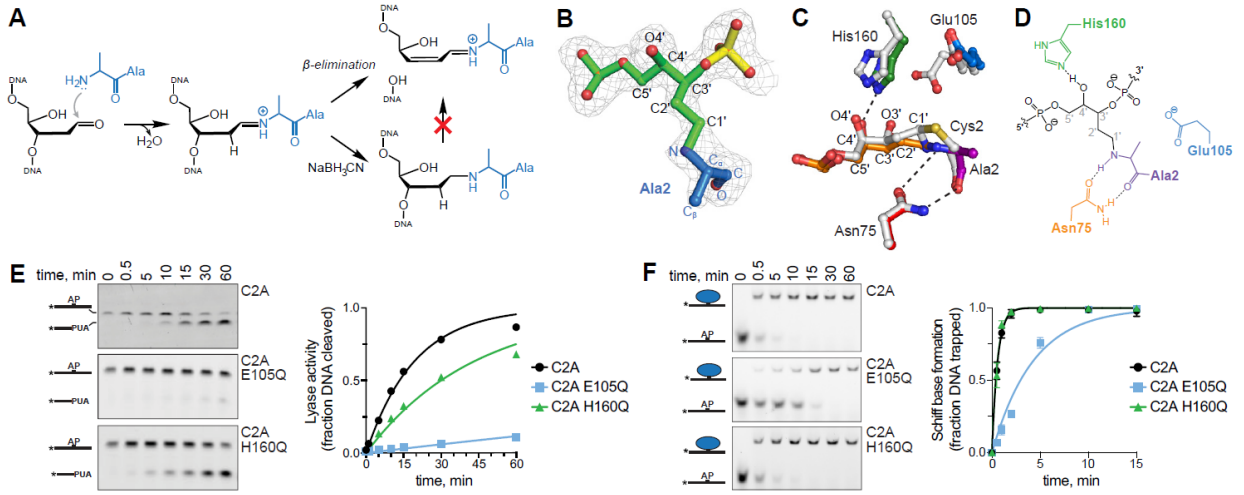

Figure 19. YedK DPC formation proceeds through a Schiff base intermediate. A. Borohydride reduction of the Schiff base formed between YedK C2A and AP-DNA prevents β-elimination. **B**. Crystal structure of the reduced YedK C2A DPC superimposed against $2F_o - F_c$ composite annealed omit electron density, contoured at 1σ. Ala2 is blue and the AP site is green. **C**. Superposition of wild-type (PDB ID 6NUA, white carbons) and C2A YedK (colored by residue) DPC structures. **D**. Schematic of the atomic interactions of the YedK C2A DPC. **E**. Kinetics of lyase activities of YedK C2A mutants in the absence of $NaBH_3CN$. Quantitation of data from three independent experiments is shown on the right (mean ± SD, n=3). Rate constants derived from exponential fits to the data are 0.05 ± 0.002 (C2A), 0.002 ± 0.0001 (C2A E105Q), and 0.02 ± 0.001 (C2A H160Q). **F**. Kinetics of Schiff base formation between YedK C2A mutants and AP-DNA in the presence of $NaBH_3CN$. Data from three experiments is quantified on the right (mean ± SD, n=3). Rate constants derived from exponential fits to the data are 1.7 ± 0.06 (C2A), 0.2 ± 0.02 (C2A E105Q), and 1.8 ± 0.09 (C2A H160Q).

The active site residues in the trapped Schiff base structure are positioned almost identically to those in the wild-type YedK DPC (Fig. 19C,D). The main notable difference in the C2A DPC structure is that Glu105 only exhibits one conformer; the interaction between carboxylate and DNA phosphate is not observed. In our proposed mechanism, Glu105 would catalyze Schiff base formation through deprotonation of Cys2 α-NH_2 and hydrolysis of C1′. To investigate the roles of the active site residues E105 and H160 in catalyzing Schiff base formation, we determined the kinetics of both lyase activity and borohydride-trapped crosslinking of C2A E105Q and C2A H160Q double mutants (Fig. 19E,F). The C2A E105Q double mutant severely reduced both activities relative to C2A alone, whereas C2A H160Q had a lesser effect. The rates of lyase activity in C2A E105Q and C2A H160Q were 25-fold and 2.5-fold slower than C2A (Fig.

19E), further supporting an important role for E105 in the steps prior to Schiff base formation. Similarly, C2A E105Q reduced the rate of DPC formation in the presence of borohydride by 10-fold, whereas C2A H160Q showed the same rate as C2A (Fig. 19F). These data indicate that E105, but not H160, is important for Schiff base formation, consistent with our model (Fig. 15H).

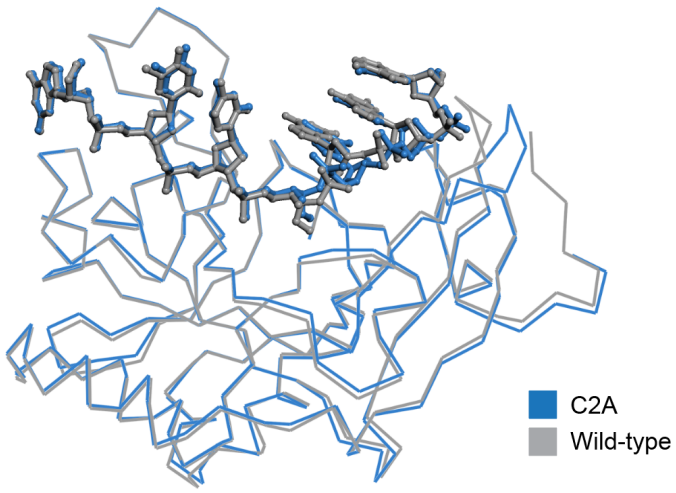

Figure 20. **Comparison of YedK C2A and wild-type DPC structures**. Superposition of the YedK C2A trapped Schiff base (this work) and the YedK DPC (PDB ID 6NUA) crystal structures.

4.2.4 YedK reacts with AP lyase products

AP sites are susceptible to spontaneous and DNA lyase catalyzed strand cleavage through β-elimination of the 3′ phosphoryl group, which generates a single-strand break with a 3′-phospho-α,β-unsaturated aldehyde (3′-PUA) on one strand and a 5′-phosphate on the other (Fig. 21A) *(23,97)*. The 3′-PUA may undergo further δ-elimination to liberate the ribose moiety, leaving a 3′-phosphate (3′-P). We tested the idea that SRAP could form a crosslink with the 3′-PUA by incubating AP-DNA with bifunctional DNA glycosylases endonuclease III (EndoIII/ Nth), endonuclease VIII (EndoVIII/ Nei), or YedK C2A, all of which cleave AP-DNA *(46,127,128)*, followed by incubation with wild-type YedK (Fig. 21B). In all cases, incubation with YedK resulted in the disappearance of the band corresponding to the lyase β-elimination product and the appearance of a corresponding DPC smaller in size to the YedK DPC formed with untreated AP-DNA. The amounts of the two DPCs in the three reactions were proportional to the amounts of uncleaved and cleaved AP-DNA from the lyase reaction, indicating that the lower molecular

weight DPC is formed from the 3′-PUA. Consistent with the requirement for a reactive aldehyde, YedK did not react with the δ-elimination product of EndoVIII. We also found that a preformed YedK DPC is refractory to DNA lyase cleavage by the glycosylases (Fig. 21B, right gel).

We next tested our panel of active site mutants against 3′-PUA DNA substrates generated by EndoIII. Interestingly, N75A, which had only a modest effect on crosslinking to an internal AP site (Fig. 1C,F), was unable to fully crosslink the 3′-PUA after 20 minutes (Fig. 21C). Most notably, E105A and E105Q were refractory to 3′-PUA crosslinking, further supporting the role of Glu105 in formation of the Schiff base intermediate. Finally, the H160 mutants had a milder effect on 3′-PUA crosslinking relative to an internal site, consistent with this residue playing a role in ring-opening but not Schiff base formation.

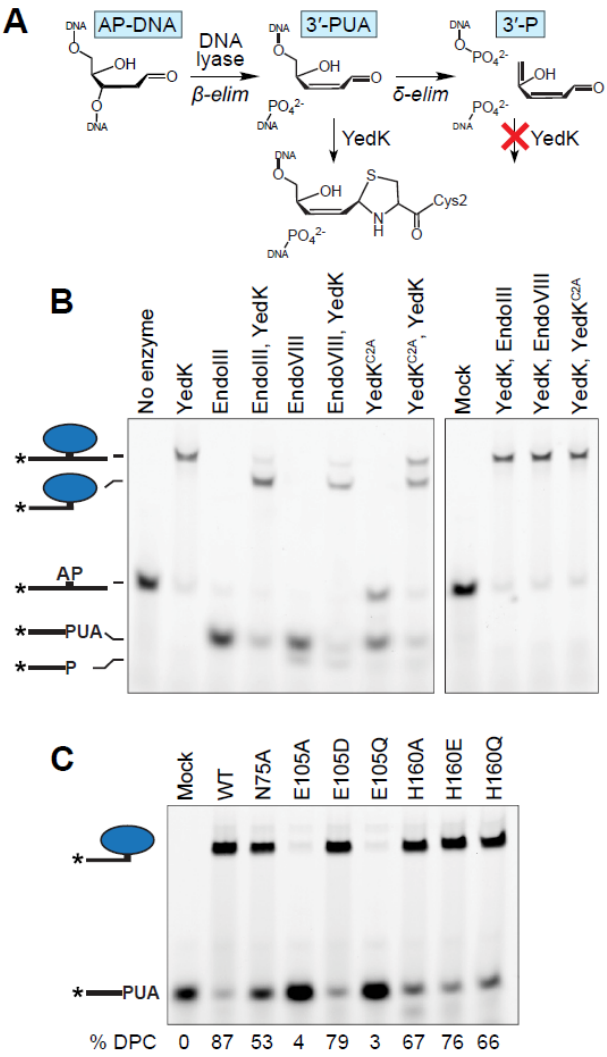

Figure 21. YedK reacts with AP lyase products. A. Reaction scheme for β-and δ-elimination by AP lyases. YedK can crosslink to the 3′-phospho-α,β-unsaturated aldehyde (3′-PUA) but not the 3′-phosphate (3′-P) of the δ-elimination product. **B**. YedK DPCs formed from the reaction products of AP lyases, EndoIII, EndoVIII, and YedK C2A. Enzymes are listed in order of addition. **C**. Crosslinking of YedK mutants with the 3′-PUA formed by EndoIII. Reactions were carried out for 20 m at 37 ºC.

4.2.5 The SRAP thiazolidine linkage is reversible

SRAP DPCs are highly stable—on the order of hours in cells, and days *in vitro* at physiological temperature *(41,46)*. The thiazolidine linkage itself is relatively stable and exists in an equilibrium with the Schiff base, with the thiazolidine favored by 5 orders of magnitude *(55,56,58,87,88)*. We previously demonstrated thermal hydrolysis of the SRAP DPC after 10 minutes at 90 ºC at pH 8.0 by following the protein component by SDS-PAGE *(46)*. To elucidate the chemical nature of the DNA component of the hydrolysis product, we examined the sizes and reactivity of DNA products liberated from thermally denatured DPCs (Fig. 22A). Heating the DPC at 90ºC for 10 minutes at pH 6.5 completely hydrolyzed the DPC to generate two DNA products consistent with intact and nicked AP-DNA, both of which would contain a reactive aldehyde. The nicking observed is the result of spontaneous AP site hydrolysis (i.e., not YedK dependent) since the amount of nicked AP-DNA in the mock reaction is the same as the thermally denatured DPC reaction (Fig. 22A). Addition of fresh YedK to the boiled DPC mixture generated two crosslinked species consistent with DPC formed from both DNA hydrolysis products. Thus, heat denaturation of the DPC leads to a direct reversal of the thiazolidine to regenerate a free, reactive AP site.

Since thiazolidine reversal and exchange with competing aldehydes is possible *(129)*, we next tested whether the crosslink is reversible in solution under physiological conditions. DPC was pre-formed with a 20-mer oligodeoxynucleotide containing an AP site (DPC-20), followed by addition of 4-fold excess of 40-mer AP-oligodeoxynucleotide to trap any hydrolyzed DPC-20 (Fig. 22B). We observed the appearance of a 40-mer DPC (DPC-40) and a disappearance of DPC-20, consistent with direct reversal of the original DPC and reformation of DPC with the longer AP-containing oligo trap. The half-time of the exchange reaction under our experimental conditions was 2-4 hours. We verified that the reverse reaction was enzyme catalyzed, as DPC exchange was not observed after 24 hours with YedK E105Q and was severely slowed with H160Q (Fig. 22C). We also observed spontaneous reversal of a DpC formed with the YedK N-terminal peptide (Fig. 22D). In this case we used the aoN-g probe as a trap to capture any hydrolyzed DpC. As with the YedK DPC, we observed the disappearance of DpC and appearance of aoN-g-DNA over time, consistent with direct reversal of the DpC. Reversal of the DpC was about 2-4 fold faster than that of the DPC.

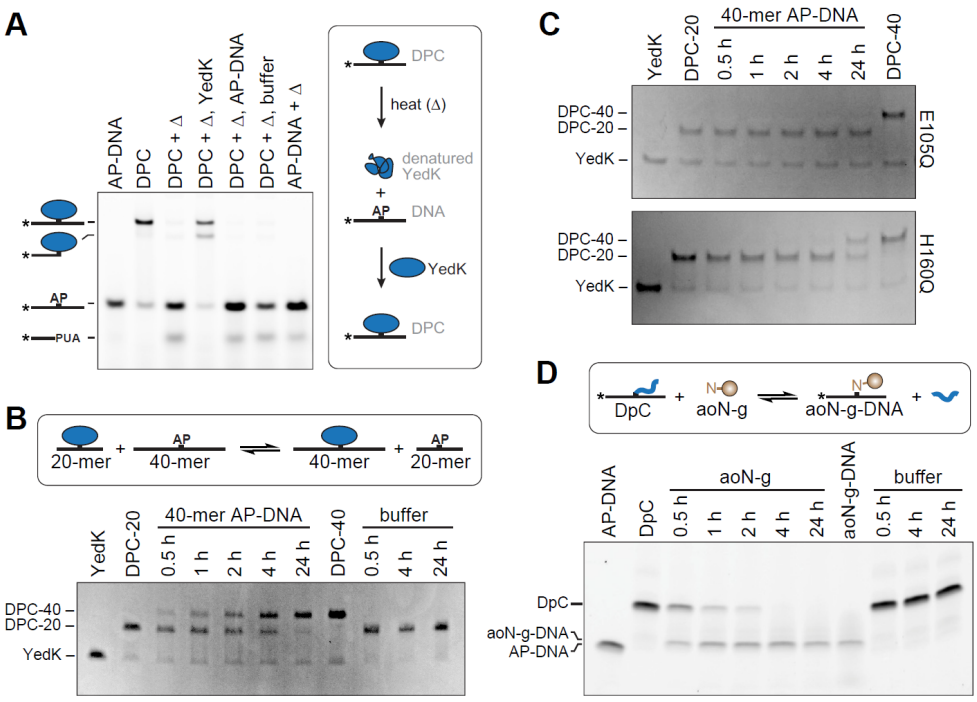

Figure 22. The SRAP DPC is reversible. A. Reaction of YedK DPC with heat (Δ), followed by either fresh YedK, AP-DNA, or buffer. **B**,**C**. Time course of YedK DPC reversibility at 37°C for wild-type (B) and E105Q and H160Q mutants (C). DPC between YedK and 20-mer AP-DNA (DPC-20) was incubated with four-fold excess 40-mer AP-DNA or buffer and analyzed at the indicated time points. Free protein and DPC were visualized via Coomassie stained SDS-PAGE. **D**. YedK DpC pre-formed with 20-mer ssDNA was incubated with four-fold excess aoN-g or buffer. DNA bands were visualized with FAM fluorescence.

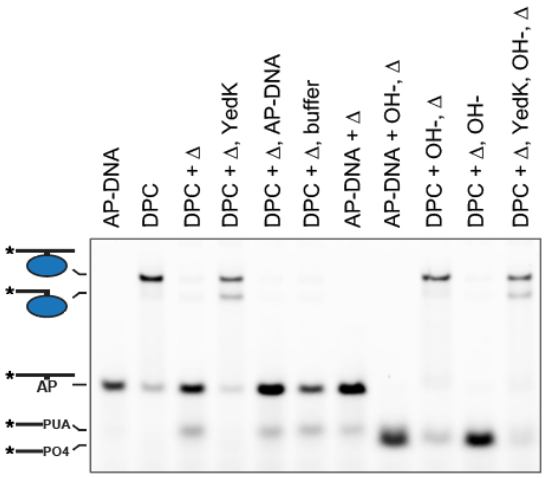

Figure 23. YedK DPC is refractory to strand breakage by alkaline pH. AP-DNA or DPC were reacted with hydroxide and heat (Δ) in the orders shown. Lanes 1-7 show that YedK DPC abolished after boiling can be re-formed with addition of fresh YedK but not DNA or buffer. Lanes 8-11 show that YedK DPC is refractory to cleavage by sodium hydroxide. DNA bands were visualized with FAM fluorescence. A cropped version of this gel (lanes 1-7) is shown in Fig. 22.

4.3 Discussion

4.3.1 Catalytic mechanism

Here, we provide a mechanistic basis for catalysis of AP site crosslinking by SRAP domains. Our data are consistent with Glu105- and His160-dependent, SRAP-catalyzed AP site ring opening to generate the reactive aldehyde necessary for attack by the Cys2 nucleophile. Three lines of evidence support this. First, wild-type YedK crosslinks to AP-DNA 2-3 orders of magnitude faster than the YedK peptide or the aldehyde reactive probe, aoN-g. These experiments were performed under saturating conditions to exclude diffusion rates from our interpretation. Secondly, by isolating the ring-opening step by formaldehyde blocking of Cys2 α-amino and thiolate side chain, we found that E105Q and H160Q mutants suppress spontaneous β-elimination of the AP site, suggesting that these residues are necessary to produce the reactive aldehyde form. Thirdly, our model predicts that His160 only plays a role in the ring-opening step. Consistently, YedK reactivity with a 3′-PUA, which effectively bypasses the requirement for the ring-opening step, were not as dependent on His160 as was the internal AP site crosslinking reaction.

Glu105 is by far the most important residue to DPC formation other than Cys2. In addition to facilitating ring-opening, likely by deprotonation of the hydroxide at C1′, Glu105 is also essential for Schiff base formation through its ability to deprotonate Cys2 α-NH_2 and to protonate of the water leaving group from C1′. Our YedK C2A crystal structure under reducing conditions confirms that the DPC reaction proceeds through a Schiff base intermediate, and that the N-terminal amine is the initial crosslinking nucleophile. Interestingly, we did not observe a Glu105 conformer in contact with the DNA phosphate in the C2A DPC structure. This is consistent with our proposed mechanism in which Glu105 likely deprotonates the Cys2 sulfhydryl for thiazolidine ring closure in the last step of the reaction. In the absence of Cys2, Glu105 would remain in its anionic form. Ionization of the Glu105 carboxylate is important for DPC formation since both E105A and E105Q had significant effects on catalysis. Both mutants exhibited biphasic kinetics at pH 6.0, with short burst (k_{fast}) and longer slow (k_{slow}) phases at least 2.5-fold and 1,000-fold slower than wild-type.

4.3.2 Reaction with DNA lyase products

HMCES reduces DNA double strand breaks (DSB) in cells, presumably by protection of the AP site from spontaneous or enzymatic cleavage at replication forks *(40,41,46,71,121)*. Our finding that YedK is capable of crosslinking to DNA containing a 3′-PUA suggests additional roles for SRAP in DNA repair. The 3′-PUA DPC is consistent with reactivity of cysteine with α-β-polyunsaturated aldehydes *(130)*. Specifically, we showed that YedK forms DPCs to the products of bacterial Nth and Nei glycosylases, raising the possibility that HMCES protects against the crosslinks to 3′-PUAs generated during abortive base excision repair, an activity that would contribute to HMCES-dependent reduction of spontaneous DNA strand breaks in cells. Moreover, it is interesting to speculate that in addition to protecting AP sites from strand cleavage, the SRAP domain may also mark broken DNA ends for subsequent repair.

An interesting aspect of the 3′-PUA crosslinking activity is how the enzyme gains access to the 3′-end generated by a glycosylase. SRAP has a strong preference for AP sites in ssDNA *(41,46,83)* and can accommodate dsDNA on the 3′-side of the AP site *(46,123)*. Lyase activity by a bifunctional DNA glycosylase would generate a substrate in which the duplex region is 5′ to the reactive PUA. This suggests that fraying must occur for SRAP to gain access to the end since SRAP cannot accommodate duplex DNA 5′ to the crosslink. Interestingly, the ssDNA 3′ to the AP site, which would not be present in a 3′-PUA substrate, was disordered in several crystal structures of both HMCES SRAP domain and YedK *(45,46,83,123)*, and thus the ssDNA is largely anchored to the protein by the interactions 5′ to the AP site. Binding of ssDNA containing an internal AP site fully extends across the active site to form a highly ordered DNA-protein complex that helps to orient the active site residues for catalysis *(123)*. Binding to ssDNA containing a 3′-PUA would not be as well ordered. Consistently, we found that Asn75, which likely helps stabilize and position Cys2 for nucleophilic attack on the AP site stabilize, was more important for YedK crosslinking to 3′-PUA than to an internal AP site.

4.3.3 DPC reversibility

The mechanism by which the HMCES DPC is ultimately resolved is unknown. In human cells, resolution may take hours and involves proteasomal degradation *(41)*. In *Xenopus* cell-free

extracts the DPC is converted to a DNA-peptide crosslink (DpC) by SPRTN protease *(121)*. How the DpC is resolved and whether the intact DPC is removed by alternative pathways are unclear. Here, we show that SRAP catalyzes direct reversal of its DPC back to a reactive AP site on a 2-4-hour timescale. We also observed a faster (1-2-hour) reversal of DpC, indicating that reversal also occurs spontaneously from exposure of the thiazolidine to solvent. The difference in rates of the enzymatic and spontaneous crosslink reversal is consistent with thiazolidine hydrolysis occurring through the rate-limiting step of ring-opening to form a Schiff base *(56)*, which is also susceptible to spontaneous hydrolysis. The relatively slow timescale of DPC reversal may be important to protect the AP-site during replication, which occurs over 7-8 hours *(131)*. The 2-4-hour reversal of the DPC may allow for transient protection of the AP site until replication is completed in a specific region, at which point the DPC would be reversed, placing the AP site in the context of a ss/dsDNA junction for subsequent repair.

Moreover, resolution of the DPC and DpC is likely cell-type or lesion-dependent (e.g. AP site vs AP-ICL). The HMCES DPC is ubiquitylated in cells and this may target the protein for proteolysis or serve to recruit other DNA repair factors to the lesion *(41,121)*. Additionally, there is evidence from *Xenopus* extracts that HMCES forms a DPC shortly after CMG helicase bypasses an unhooked AP-ICL, protecting the AP site from cleavage until TLS can proceed *(121)*. Since HMCES reduces mutation frequency in U2OS cells *(41,71)* and TLS is an error-prone process, the delay of TLS by HMCES DPC likely allows for recruitment of an error-free bypass mechanism, such as template switching, in certain cell types. In B cells, DPC formation by HMCES reduces deletions during somatic hypermutation (SHM), and it has been proposed that subsequent TLS may be an outcome in SHM *(42)*.

---

[5] This work is submitted for publication as: Paulin, K.A., Cortez, D., Eichman, B.F., (2022) Abasic site ring opening and DNA-protein crosslink reversal by the SRAP protein YedK, *J. Biol. Chem. Submitted*
This work has also been published as a preprint: Paulin, K.A., Cortez, D., Eichman, B.F., (2022) Abasic site ring opening and DNA-protein crosslink reversal by the SRAP protein YedK, *BioRxiv.*
I designed and performed the experiments and wrote the manuscript.

CHAPTER 5

Concluding discussions and future directions

Found across all domains of life, SRAP proteins represent an exciting area for future research. The high degree of conservation of the SRAP domain allows the structural and biochemical information presented in this work to be extrapolated to other proteins in this family, providing a strong foundation to support investigation of the roles of SRAP proteins in other organisms. The thiazolidine DPC formed between SRAP and AP-DNA has not yet been observed in any other protein. It is particularly notable that the SRAP DPC does not lead to strand breakage- an important distinction from other DPCs. Abasic (AP) sites are one of the most common DNA lesions that block replicative polymerases. HMCES recognizes and processes these lesions in the context of ssDNA, and HMCES has specificity for AP sites in ssDNA at junctions found when replicative polymerases encounter the AP lesion.

HMCES protects cells from strand breaks, inhibits mutagenic translesion synthesis, and participates in repair of interstrand DNA crosslinks derived from AP sites by forming a stable thiazolidine DNA-protein crosslink (DPC) to AP sites in single-stranded DNA (ssDNA). Despite the importance of HMCES to genome maintenance, the enzymatic mechanisms of DPC formation and resolution were unknown. Using the bacterial homolog YedK, I have shown that the enzyme catalyzes conversion of the AP site to its reactive, ring-opened aldehyde form, and proceeds through a Schiff base intermediate that forms prior to the more stable thiazolidine. The HMCES DNA-protein crosslink (DPC) intermediate is thought to shield the AP site from endonucleases and error-prone polymerases.

In this final section, I present preliminary data involving DPC mechanism insights, crystallography leads, and HMCES SRAP biochemistry. Additionally, since YedK reacts with polyunsaturated aldehydes at DNA 3′-ends generated by bifunctional DNA glycosylases and catalyzes direct reversal of the DPC to regenerate the AP site, I will discuss possible mechanisms by which HMCES DPCs are resolved in cells. I then explore potential implications in human health and disease. Finally, I discuss important future directions for this project from structural and biological perspectives.

5.2 Preliminary results and discussion

5.2.1 YedK E105A and E105Q exhibit biphasic kinetics.

We observed biphasic kinetics of DPC formation in E105A and E105Q mutants. Neither protein stability nor DNA binding differences were responsible since increased concentration of protein, or preincubating protein in reaction buffer before adding AP-DNA did not eliminate the biphasic kinetics of E105A, and thus we can rule out protein stability as a reason for this behavior. Elevated temperature did not eliminate the biphasic kinetics either. Interestingly, we did not observe this behavior at pH 4-5 (Fig. 24). DPC kinetics of E105Q at pH 4.0, 5.0, 9.0 exhibit one-phase association. At pH 9.0, the kinetics are likely two-phase, but first phase is likely not complete at the final 60 minute time point. The kinetics of pH 6.0, 7.0, 8.0 fit to two-phases.

This biphasic nature suggests that there are two forms of either the enzyme or the substrate at the onset of the reaction, one of which is primed for catalysis and bypasses the requirement for the enzyme in the initial step of the reaction. In the case of the enzyme, we speculate that two forms may exist that differ by the initial protonation state of the N-terminal α-amino group. Formation of a Schiff base indicates that the N-terminal amine, not the Cys2 thiolate, initiates crosslink formation. The hydrophobic environment of the Cys2 residue in SRAP would raise the pK_a of the thiol and lower the pK_a of the amine, favoring the amine as the nucleophile. If Glu105 catalyzes α-amino deprotonation, which is required for Schiff base formation, the initial burst would correspond to the population of the deprotonated N-terminal amine and the slow phase would correspond to the time required for spontaneous deprotonation of the N-terminal amine. Interestingly, the k_2 of both E105A and E105Q are similar to the rates of the uncatalyzed crosslinking reactions with the aoN-g probe or the YedK peptide.

E105Q does not display biphasic kinetics at pH below 6.0, consistent with the pH-dependence of Schiff base formation in which protonation of the amine under acidic conditions reduces its nucleophilicity *(81)*. At neutral pH, the rate determining step is acid-catalyzed dehydration of the carbinolamine addition product. Under acidic conditions, amine attack is rate-determining in Schiff base formation because conversion of the amine to its conjugate acid slows the rate of amine attack *(81)*.

Alternatively, the biphasic kinetics may be the result of the small (1%) population of AP site that exists as the reactive aldehyde in solution. In this case, the initial burst would correspond to AP-DNA in a spontaneous ring-opened state and k_{slow} would correspond to the population of AP-DNA in the more abundant, less reactive ring-closed state that requires Glu105 for activation *(22,119)*. The effect of pH on AP site ring-opening in DNA has not been investigated.

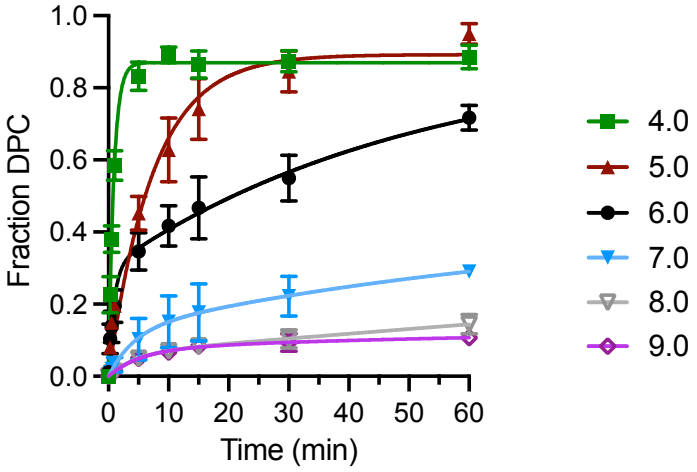

Figure 24. YedK E105Q DPC exhibits pH dependence. Kinetics of crosslinking at different pH (mean ± SEM, n=3). Reactions were performed at 25 ºC. pH 4.0, 5.0, 9.0 are fit to one-phase association. pH 6.0, 7.0, 8.0 are fit to two-phases.

Table 4. YedK E105Q pH dependence rates[a]

	1-phase	***2-phase***	
pH	**k_1 (min^{-1})**	**k_1 (min^{-1})**	**k_2 (min^{-1})**
4.0	1.1 ± 0.05	1.5 ± 0.5	0.3 ± 0.3
5.0	0.1 ± 0.01	1.5 ± 1.0	0.1 ± 0.02
6.0	0.1 ± 0.03	0.9 ± 0.3	0.01 ± 0.01
7.0	0.07 ± 0.02	0.2 ± 0.3	0.001 ± 0.08
8.0	0.07 ± 0.01	0.2 ± 0.2	0.001 ± 0.04
9.0	0.1 ± 0.02	0.2 ± 0.2	0.02 ± 0.1

[a] Reactions were carried out at 25 ºC. Values are mean ± SEM (n=3)

5.2.2 Thiazolidine stability

SRAP forms a stable thiazolidine linkage between a ring-opened AP site and the α-amino and sulfhydryl substituents of its N-terminal cysteine residue. The thiazolidine linkage explains the remarkable stability of the HMCES DPC and its resistance to strand cleavage. Proteolysis of the DPC down to a peptide-DNA crosslink is persistent on the order of hours *(46)*, but in the presence of an excess of AP-reactive trap, the crosslink can be reversed. Notably, spontaneous DPC reversal in the presence of excess AP site trap at 37 ºC does not lead to strand cleavage. A direct reversal of the thiazolidine would lead to an intact AP site which could then be reacted with fresh YedK to form a DPC. Thiazolidine hydrolysis occurs through the rate-limiting step of ring-opening to form a Schiff base *(56)*. The thiazolidine must be reversing back to the free AP site for the AP-reactive traps to act as a sink, as reported here. This is consistent with reports of thiazolidine spontaneous ring-opening at pH greater than 4.0 *(59)*. The thiazolidine ring formed between free cysteine and formaldehyde is stable under acidic conditions up to at least 3M HCl, and reversal back to aldehyde and free cysteine is only observed under basic conditions greater than 1 M NaOH *(86)*.

The thiazolidine ring formed by SRAP DPC is susceptible to milder conditions of reversal such as boiling at pH 6.0. However, SRAP DPC is resistant to reversal by 0.2 M NaOH at 70 ºC for 2 minutes (Fig. 23, Chapter 4), consistent with the stability of thiazolidines. Notably, the spontaneous DPC reversal in the presence of excess AP site trap at 37 ºC does not lead to strand cleavage. Several highly conserved resides that cradle the catalytic Cys2 also make the AP site inaccessible to solvent when sequestered in the active site, and this protection is lost when the protein is denatured by boiling, leaving the AP site susceptible to base catalyzed β-elimination. Since the thiazolidine is connected to the protein and within a hydrophobic active site, it is not surprising that the conditions of thiazolidine stability and reversal are different than that of free thiazolidine.

Since a borohydride-reduced Schiff base is known to be stable *(63)*, I compared the stability of the wild type thiazolidine linkage to that of the C2A reduced Schiff base. The C2A reduced Schiff base was resistant to boiling, whereas the wild type DPC was susceptible to hydrolysis (Fig. 25), consistent with results of DPC reversal under heat shown in Chapter 4. When borohydride was added to the wild type and AP DNA mixture at the same time, the resulting DPC was still susceptible to reversal by heat, indicating that the DPC was a thiazolidine linkage even in the

presence of the reducing agent. It is likely that thiazolidine formation proceeds more rapidly than Schiff base reduction. Rapid thiazolidine formation is likely important for efficient DPC formation and protection of AP sites in cells.

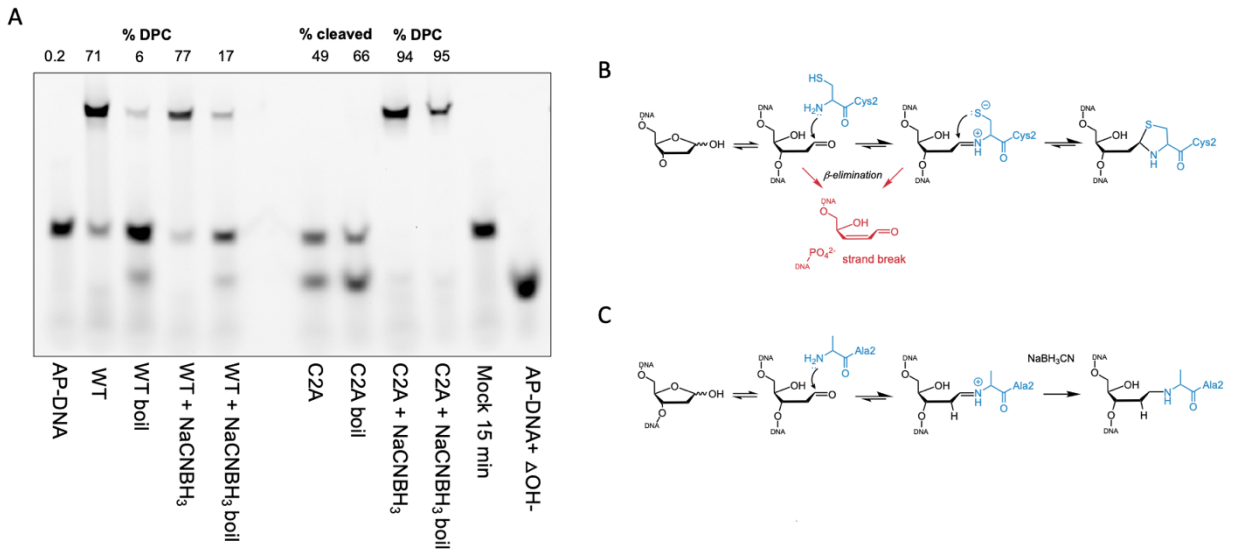

Figure 25. Reduced Schiff base is refractory to reversal by boiling. A. Reactions of AP-DNA with wild type YedK compared to C2A mutant YedK with and without borohydride reduction. Reactions were performed at 37 ºC. Boil indicates to 15 min at 90 ºC. **B.** General reaction of SRAP with AP DNA. **C.** Reaction ofC2A mutant with AP DNA and reduction by borohydride.

5.2.3 YedK ΔC2 mutants

To further investigate SRAP catalysis of AP site ring-opening, I sought to identify a point mutant which abrogates DPC formation, can still bind non-covalently, and does not lead to strand cleavage at the AP site. To this end, I generated a ΔC2 point mutant abrogating DPC formation and retaining non-covalent binding. To further probe other potential reaction intermediates, I generated double mutants with ΔC2 with an additional point mutation of either E105Q or H160Q. The aim was to utilize a non-crosslinking mutant to pursue a crystal structure of YedK bound non-covalently to a natural AP site, which is expected to reveal the interactions of the AP site with the other active site residues in a pre-catalytic complex. However, the ΔC2 mutants exhibited lyase activity and DPC could be trapped with borohydride. It is likely that the flexibility of the N-terminus allows for Schiff base formation between the alpha amino group of Gly3 and the AP site.

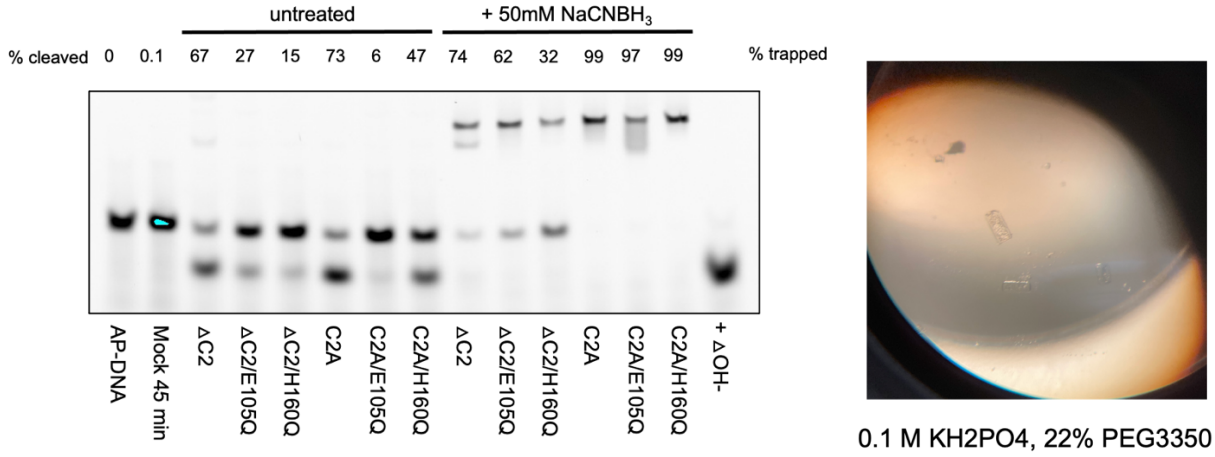

Figure 26. YedK ΔC2 mutants exhibit lyase activity and proceed through Schiff base intermediate. Left. Denaturing gel of reactions of YedK mutants with AP DNA at 37 ºC for 45 minutes. Sodium cyanoborohydride was added the reaction where indicated. DNA was visualized via fluorescence. **Right.** Disintegrating plate crystals of ΔC2/H160Q and AP-DNA.

YedK ΔC2/H160Q crystallography hits

After promising crystal screening results with ΔC2/H160Q and AP-DNA in the PEG-Ion screen (Hampton Research), I set up optimization trays and got crystals with ΔC2/H160Q and AP-DNA that dissolved shortly after forming (within 7 days). Crystals were plates that formed flush with the cover strip (hanging drop) so it was not possible to scoop them. The instability of these crystals is likely due to the lyase activity of the ΔC2 mutants (Fig. 26). Since the lyase activity is slow, it may be possible to optimize these crystals at a lower temperature.

5.2.4 YedK N75D

It is likely that Asn75 positions the Cys2 amino group for nucleophilic attack of the AP site since Asn75 hydrogen bonds to the carbonyl O and the backbone amide nitrogen of Cys2. Despite this positioning as well as the high degree of conservation of this residue, it was surprising that the N75A mutant behaved similarly to wild type. Due to the lack of a significant difference, N75A and other N75 mutants were not investigated further. However, the N75D mutant did not exhibit any appreciable crosslinking in a preliminary experiment (Fig. 27). I intend to repeat these assays via time course as described above. If N75D does indeed abrogate crosslinking, I hypothesize that this charge reversal would lead to a catalytically unproductive orientation of Cys2. Although the

carbonyl of the glutamate side chain could still position the Cys2 backbone amide N, the negative charge of the glutamate hydroxyl would not hydrogen bond to the carbonyl O of Cys2, but instead would repel, resulting in incorrect positioning of the crosslinking nucleophile for attack. If the preliminary crosslinking results are reproducible, N75D may be a crystallization target for non-covalently bound YedK with a natural AP site. Since mispositioning of key residues could make interpretation of N75D biochemical results difficult, crystals of N75D alone would be a useful first target to determine the impact of this mutation on the active site.

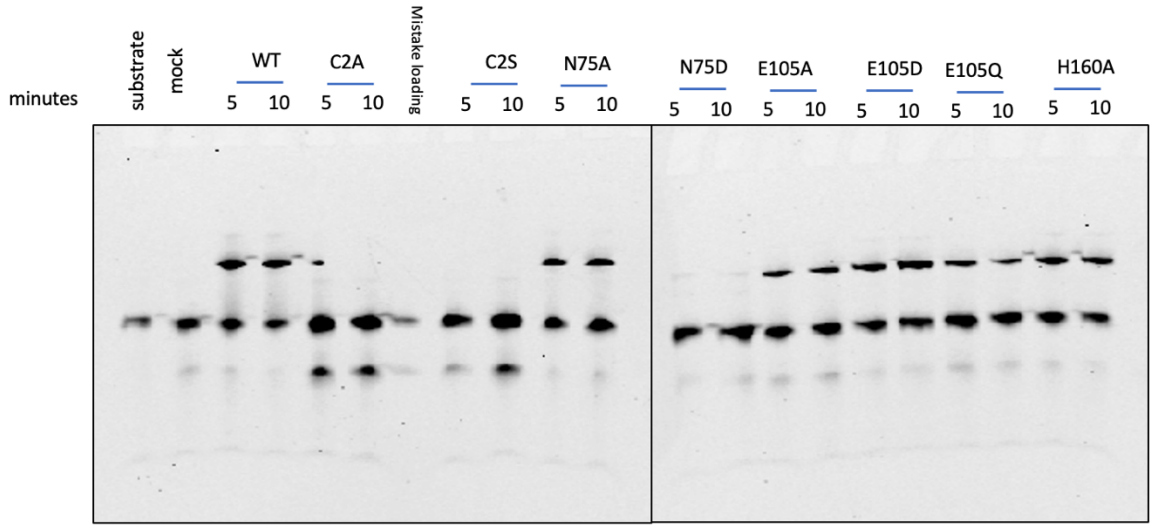

Figure 27. YedK N75D does not exhibit DPC formation or lyase activity after 10 min. Preliminary (2019) YedK mutant crosslinking panel showing N75D. Reactions were performed at 37 ºC. DNA was visualized via FAM fluorescence.

5.2.5 Verifying purification of free protein.

With the DPC being relatively stable, there was concern that instead of purifying free protein, I could be purifying DPC that formed prior to cell lysis or during purification. If DPC was the starting material for these assays, then the majority of my results would have been inconclusive. To verify that all of my proteins were not DPC, I ran all of my mutants on a gel. As a DPC positive control, I formed DPC by incubating wild type YedK with 7 nt AP-DNA (DPC-7), which is the DNA-binding footprint of the protein. I have already shown that nucleases likely cannot access DNA within the binding channel, so DPC being purified should run at the size of this positive control. I have also optimized conditions for separating DPC as small as 5 nt from free protein on SDS PAGE. Since the C2A mutant cannot form DPC, this was included as a negative control.

None of the proteins, wild type nor mutant, exhibited a shift except for the positive control DPC (Fig. 28). As such, it is likely that I have purified free protein as expected.

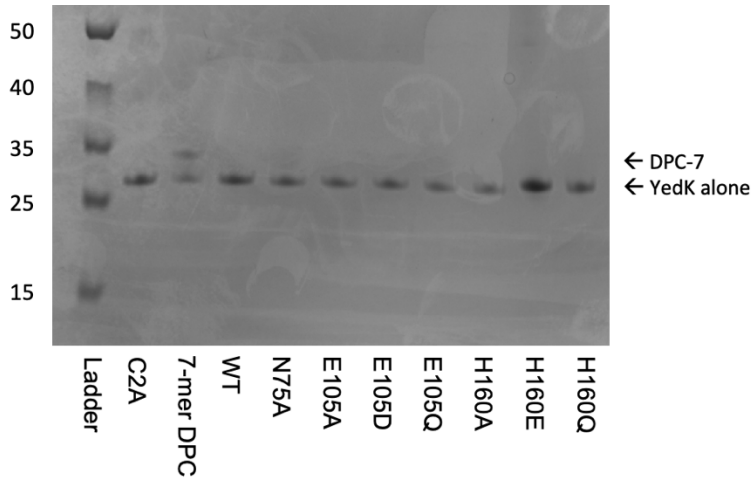

Figure 28. Purified protein is free YedK, not DPC. SDS PAGE of YedK mutant panel. Protein was visualized by Coomassie stain.

5.2.6 Preliminary crystallography.

HMCES

Broad crystallization screens of the 10-mer HMCES SRAP domain DPC identified a number of conditions which produced small crystals consistently reproducible in larger-scale optimizations (Figure 29). All hits shared a trend of PEG 3350 as the precipitant, and protein concentrations between 6-12 mg/mL. The DPC crystals were consistently small, despite many efforts at optimizing multiple parameters of the initial crystallization conditions. We first considered the possibility of C-terminal His-tag interfering with crystal contacts. To address this, we obtained a TEV-cleavable SRAP construct from Addgene. Concurrently, Petria Thompson cloned a PPS site into the original PBG101 SRAP construct so that the C-terminal His-tag could be cleaved. The PPS cleaved SRAP is indistinguishable from TEV-cleaved SRAP in both crosslinking activity and crystallization trials. Using the TEV-cleaved SRAP construct, I was able to reproduce SRAP crystals under conditions similar to those published in the PDB. However, the crystals in these optimization trials were too small to scoop.

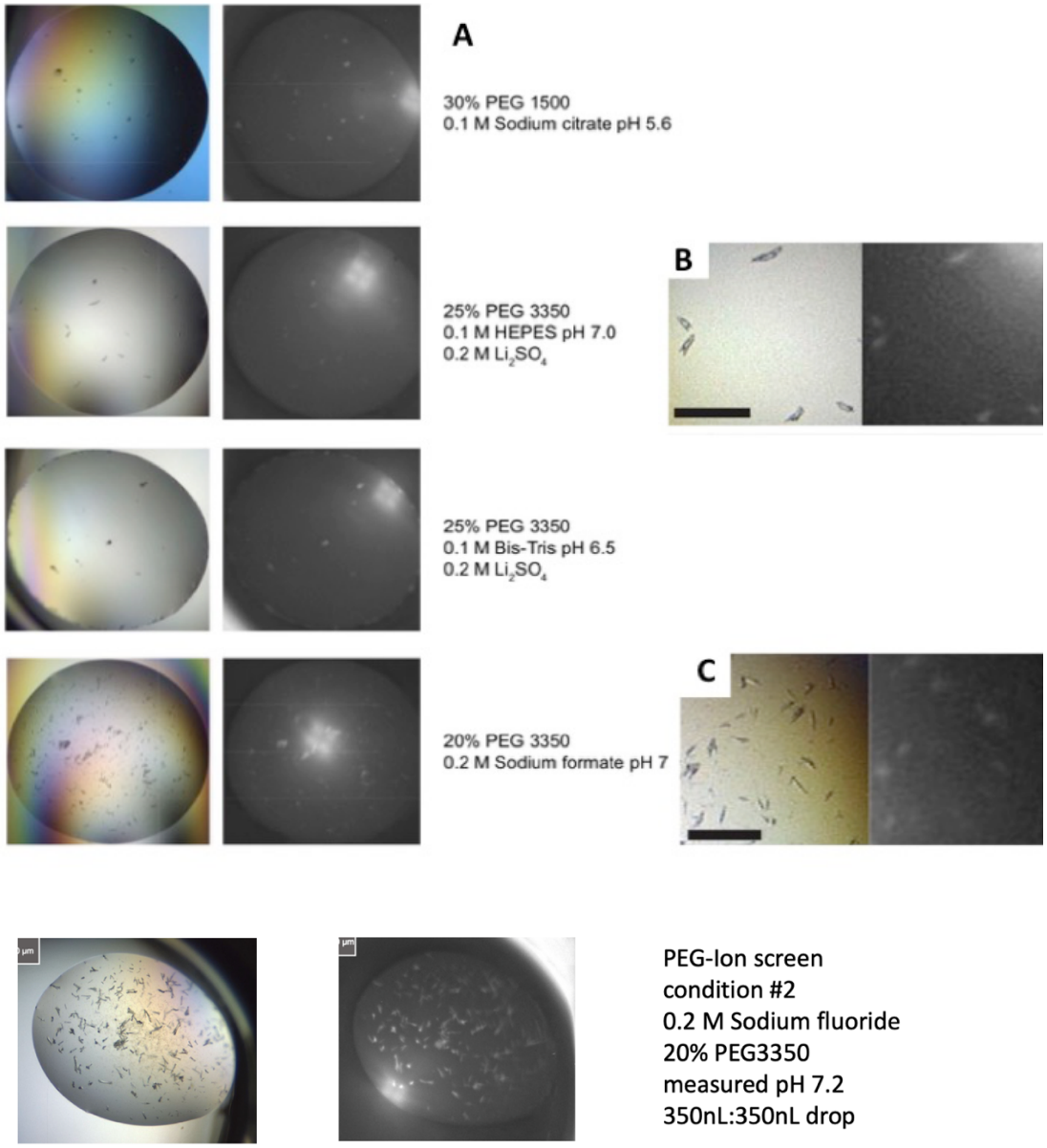

Figure 29. HMCES SRAP DPC crystal hits. A. UV-positive DNA-Protein crosslink (DPC) crystal hits of from broad screens set up on Mosquito robot: 8 mg/mL TEV-cleaved SRAP crosslinked to "H36" 10 nt ssDNA with AP site. **B, C.** Zoom of figures to the left and respective UV. Scale bar is 0.1 micron. **D.** SRAP crosslinked to 7mer.

Another approach tested was to shorten the bound DNA in the DPC via nuclease treatment of the complex. The crosslinked SRAP protects the DNA at the bound residues; so added nuclease should only cleave excess DNA ends that could interfere with crystallization. I have performed initial tests of nuclease digestion of DPC. Initial data shows that 1 U Benzonase nuclease added to

20 uL DPC at 10 uM is sufficient to produce a visible shift downward after 1 hour (Figure 30). A time course with a smaller amount of nuclease would be informative for exploring the potential for *in crystallo* digestion of the excess DNA.

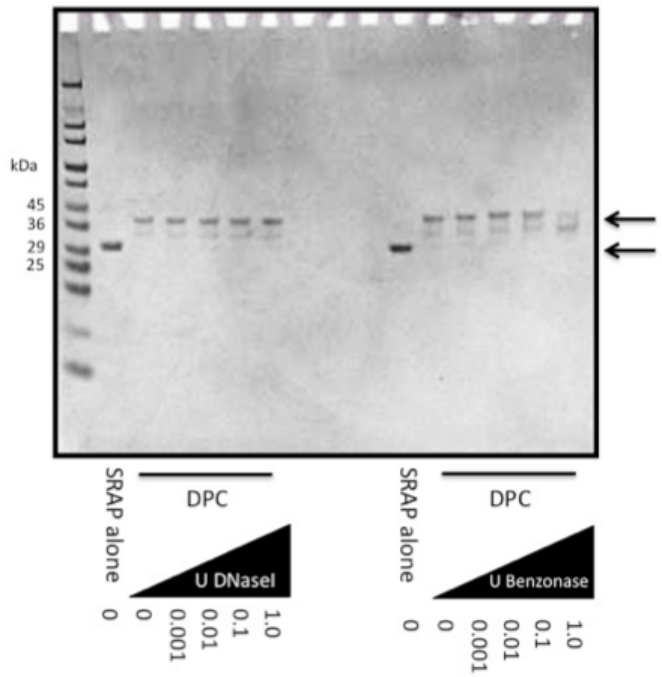

Figure 30. Nuclease digestion of SRAP DPC. SRAP was crosslinked to dual-labeled 29-mer AP-DNA then digested 1 hour with increasing concentrations of nuclease. Upper arrow indicates DPC, lower arrow indicates SRAP alone. Reactions were run on SDS-PAGE to visualize shift in band corresponding to degradation of DNA bound to the DPC.

YedK

The footprint of the protein is 7 nucleotides, and DNA ends stacking has been important for crystal packing with the structures I've determined thus far as well as other published structures (Chapter 3). I have needed to co-crystallize with either a C3 spacer or THF oligo to get hits. YedK crystals were pursued because crystals formed more readily and were large enough to scoop with YedK. Hanging drop was my preferred method of crystallization. Crystals formed readily in sitting drop, however the plate morphology of YedK crystals made these difficult to scoop since they stuck to the bottom of the tray. Scooping attempts resulted in crystals shattering. YedK tended to form plate clusters in hanging drops. The plates could easily be separated from each other by gentle manipulation with the loop before scooping.

YedK THF non-covalent complex crystal structure

I was able to determine a 1.2 Å resolution crystal structure with YedK and a 7 nt ssDNA containing a THF at the center. However, at such high resolution, I could see at least 3 different positions of a disordered loop. Since Phenix only let me place residues in one position, this led to poor statistics on the structure so it was not publishable. The structure is consistent with published a YedK THF structure PDB ID: 6KBZ *(83)*.

YedK C2A and AP-DNA

In an attempt to determine a non-covalent complex between YedK and a natural AP site, I set up several crystallization screens with YedK C2A with AP-DNA. However I was unable to get any hits, likely due to *in crystallo* lyase activity of C2A mutant.

YedK H160Q with C3 spacer

I am currently working towards crystallization of the YedK mutants E105Q and H160Q. This will inform on whether reduction of activity by these mutants is a result of rearrangement of the active site. I have set up crystallization screens with these two mutants. One condition out of several screens produced crystals for H160Q with C3 spacer ssDNA. I set up optimization trays and was able to grow crystals that look promising. These crystals will be scooped for testing diffraction. I did not get any hits in screens or optimizations with E205Q with and without ssDNA.

5.2.7 YedK knockout phenotypes in cells

Previous work by Petria Thompson in the Cortez lab investigated the impact of YedK knockouts in *E. coli* (personal communication). She did not observe any growth differences in the YedK knockout cells compared to wild type. Treatment of the YedK knockout cells with methyl methanesulfonate (MMS), which is known to increase AP sites, did not lead to a significant growth phenotype in YedK knockouts. It is possible that loss of YedK leads to increased TLS across AP lesions, which may increase mutation rate, but may not lead to an observable growth phenotype. Since loss of HMCES increases mutation frequency, it would be interesting to test whether loss of YedK increases mutation frequency in *E. coli.* One study reports that loss of YedK increases transformation efficiency in *E. coli* and the authors suggest that YedK may play a role in repressing plasmid replication. This may be due to AP sites accumulating on the plasmid prior to

transformation. Alternatively, testing a YedK knockout with overexpression of a monofunctional glycosylase such as AlkA in the presence of MMS would increase AP sites, and thus may lead to an observable phenotype. Testing a YedK knockout with overexpression of a bifunctional glycosylase could also be interesting and may lead to increased DNA breakage at AP sites catalyzed by the glycosylase which could impact the growth phenotype.

5.2.8 HMCES C2A biochemistry

YedK C2A exhibits lyase activity, but does C2A exhibit lyase activity in HMCES? A recent paper states that C2A lyase activity was not observed in HMCES SRAP *(42)*. It is unlikely that the lyase activity of YedK C2A is due to a contaminant since all YedK mutants were purified identically from the same *E. coli* cell line with the same equipment, but I only observe lyase activity in the C2A and C2S mutants. In my preliminary tests of HMCES SRAP C2A in 2018, I observed C2A lyase activity (Fig. 31). Lyase activity was observed after 15 min, with the reaction going to completion after 24 h. A small amount of DPC was observed with SRAP C2A with AP-DNA on SDS PAGE via both Coomassie stain and scanning of the same gel for Cy5 DNA. Interestingly, the DPC disappeared between 2 h and 24 h. It is important to note that those reactions underwent a mild heat workup of 55 ºC 5 min, so the complex may be due to tight binding. It would be worthwhile to repeat these HMCES SRAP C2A reactions but include the 70 ºC 1 min heat workup used in my present assays to determine whether the C2A DPC is a true result or an artifact of the workup when I was still optimizing the assay. The lack of C2A lyase activity observed by Wu et. al. may be due to the length of the reaction and there may by C2A lyase activity at later time points. It would be worthwhile to test HMCES SRAP C2A using my newer optimized crosslinking protocols.

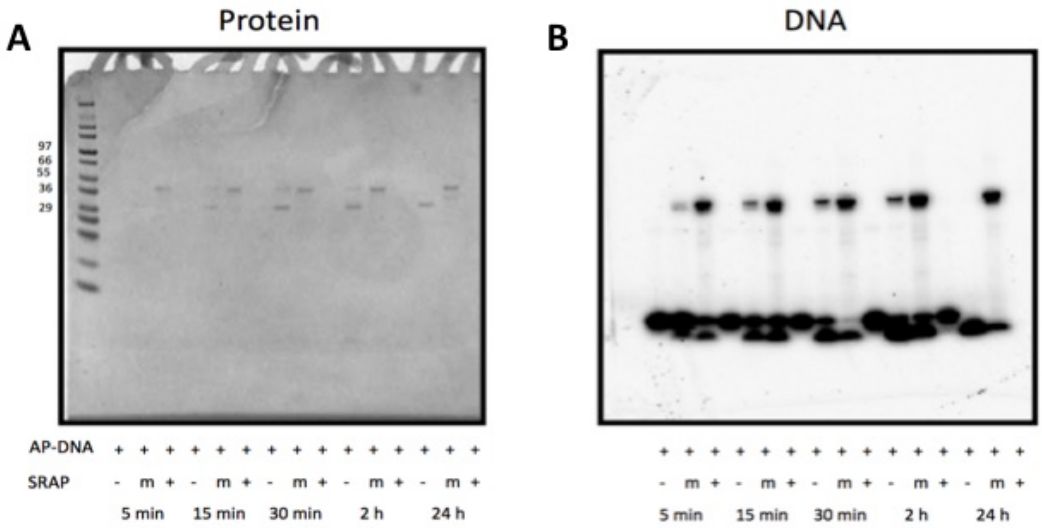

Figure 31. Time course of wild type and C2A mutant SRAP reactions with AP-DNA. A. Coomassie stain of 4-12% Bis-Tris gel to visualize protein. (-) buffer only; (m) C2A mutant SRAP; (+) wild type SRAP. **B.** Scan of 3′-Cy5 tag to visualize DNA of the gel shown in (A).

5.2.9 Structural analysis of post-translational modifications in HMCES

HMCES contains a number of post-translationally modified residues and several of these residues exhibit a high degree of conservation. It is likely that HMCES PTMs regulate recruitment of HMCES to AP lesions as well as mediating the recruitment of subsequent repair factors to the DPC. Since there are several HMCES SRAP crystal structures available (Chapter 3), and we have already mapped residue conservation onto the structure (Chapter 2), I sought to determine the positions on the crystal structure of residues that may be post-translationally modified in HMCES (Fig. 32). A number of these residues were on the C-terminal tail and therefore not in the crystal structure. However these C-terminal tail residues and PTMs may be important for PCNA interaction since HMCES C-terminal tail has been shown to mediate interaction with PCNA *(41)*.

I generated a list of PTM residues from the databases Phosphosite (phosphosite.org) and Biogrid (thebiogrid.org) and manually mapped these residues on the HMCES SRAP structure PDB ID: 6OE7. Conservation was determined from aligning HMCES to 100 of the closest homologs. I then cross referenced these residues which cases of HMCES mutations in cancer from The Cancer Genome Atlas (TCGA). Additionally, a disordered loop contains several conserved and post-translationally modified HMCES residues (Table 5, Fig. 33). It would be interesting to investigate whether these are important for recruiting subsequent repair mechanisms to HMCES DPC in cells.

Table 5. Post-translational modification sites in human HMCES.

H. sapeins HMCES residue[a]	Surrounding sequence	PTM[b]	% Conserved[c]	Location on *H. sapiens* HMCES[d]	Cases mutated in cancer[e]
K37	PEWRDPDkYCPSYNk	ub	N/A	back surface	
K44	kYCPSYNksPQSNsP	ub	43	back surface, conserved region	
S45	YCPSYNksPQSNsPV	p	N/A	back surface, conserved region	
S50	NksPQSNsPVLLSRL	p	45	back surface	
K61	LSRLHFEkDADSSER	ub	35	back surface	K61E
K83	GLVPSWFkEsDPskL	ub	92	back surface, conserved region	
S85	VPSWFkEsDPskLQF	p	N/A	back surface, conserved region	
S88	WFkEsDPskLQFNTT	p	57	back surface, conserved region	
K89	FkEsDPskLQFNTTN	ub, sm	68	back surface, conserved region	
K105	RSDTVMEkRSFkVPL	ub	98	DNA binding channel, may interact with phosphate of DNA	
K109	VMEkRSFkVPLGKGR	ub	57	5' end of DNA binding channel, may interact with phosphate of DNA	
K114	VPLGkGR	N/A	57	back surface	
K148	FIYFPQIkTEkSGSI	ub, sm	48	disordered loop	
K151	FPQIkTEkSGSIGAA	ub	18	disordered loop	
S160	GSIGAADsPENWEkV	p	N/A	disordered loop	
K166	DsPENWEkVWDNWRL	ub	N/A	back surface	K166Q
K204	VDSCKGLsDIHHRMP	N/A	51	back surface	
S207	VDSCkGLsDIHHRMP	p	43	back surface	
K225	DGEEAVSkWLDFGEV	ub	48	back surface	
T263	VNNSRNNtPECLAPV	p	21 (A/T)	back surface	T263S
K275	APVDLVVkkELRASG	ub, sm	49	C-term tail	
K276	PVDLVVkkELRASG	ub	55	C-term tail	K276R frameshift

S295	QWLATKsPKKEDsKt	p	23	C-term tail	
K297	QWLATKsPKKEDsKt	N/A	51	C-term tail	
K298	QWLATKsPKKEDsKt	N/A	26	C-term tail	K298K synonymous (G>A)
S301	KsPKKEDsKtPQkEE	p	28	C-term tail	S301* stop
T303	PKKEDsKtPQkEEsD	p	29 (S/T)	C-term tail	
K306	EDsKtPQkEEsDVPQ	ub	46	C-term tail	
S309	KtPQkEEsDVPQWsS	p	28	C-term tail	S309L
S315	EsDVPQWsSQFLQks	p	48	C-term tail	S315Y
K321	WsSQFLQksPLPTKR	ub	17	C-term tail	
S322	sSQFLQksPLPTKRG	p	45	C-term tail	
K327	PLPTkRGTAGLLEQW	ub	51	C-term tail	
T330	PLPTKRGtAGLLEQW	p	12	C-term tail	
K339	GLLEQWLkREkEEEP	ub, sm	29	C-term tail	
K342	EQWLkREkEEEPVAk	sm	14	C-term tail	
K349	kEEEPVAkRPYSQ	ub	51	C-term tail	

[a] Hyperlinks to phosphosite.org
[b] Human HMCES PTMs from phosphosite.org and thebiogrid.org
ub=ubiquitylation; p=phosphorylation; sm=sumoylation
[c] Conservation determined from HMCES aligned to 100 closest homologs "HMCES-100.aln"
[d] Location determined on PDB ID 6OE7 in Pymol with residue conservation overlay (Chapter 2)
[e] From The Cancer Genome Atlas (TCGA), accessed November 2021

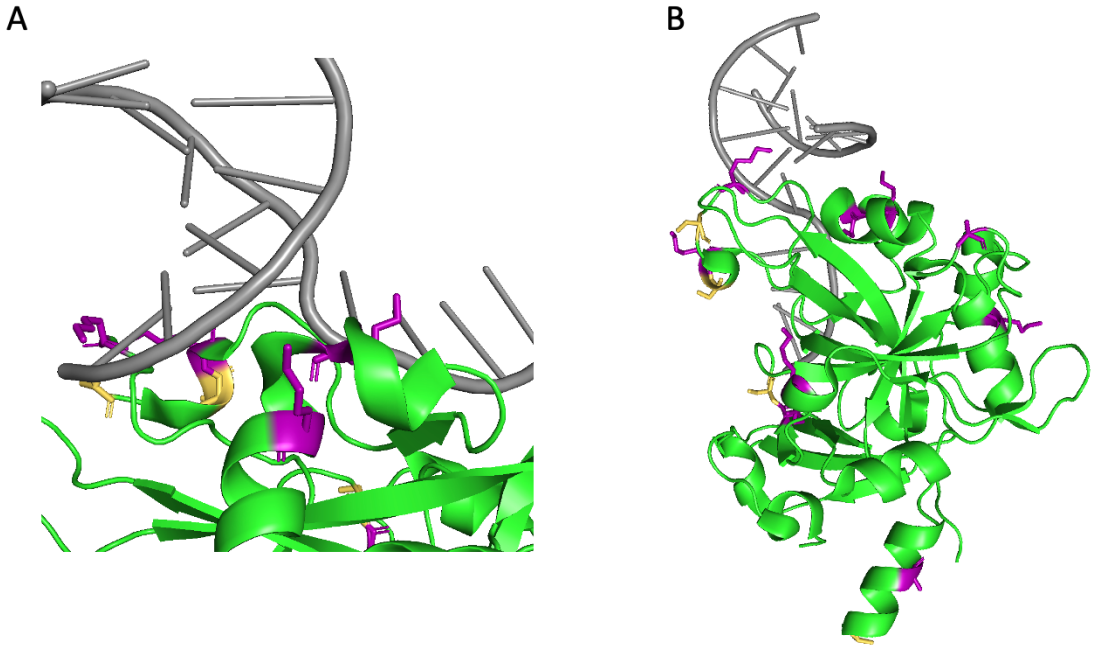

Figure 32. HMCES crystal structure with putative PTM sites identified. HMCES SRAP domain (PDB ID: 6OE7) shown in green with putatively ubiquitylated residues in purple and phosphorylated residues in yellow. DNA in grey. **A.** Conserved ubiquitylated lysines are located near DNA binding interface. **B.** Other post-translationally modified residues from Table 5 are located on the "back" of the protein, away from the DNA binding channel.

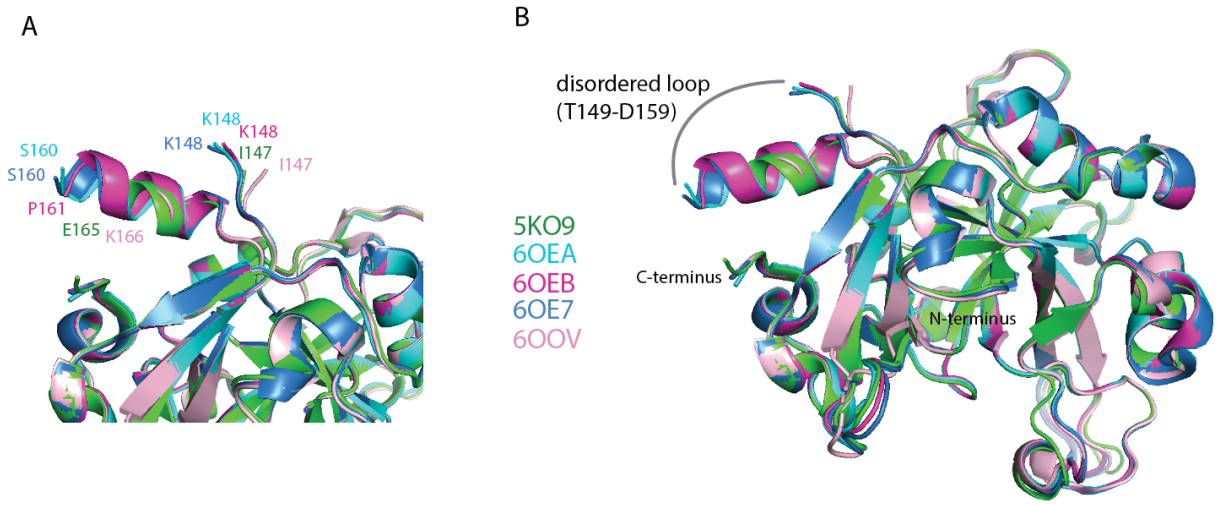

Figure 33. HMCES crystal structures overlay showing disordered loop. HMCES SRAP domain crystal structures available in the PDB exhibit a disordered loop region (T149-D159). **A.** Close up of disordered loop region with the residues before and after the loops labeled in colors corresponding to the PDB ID (center). **B.** Location of the disordered loop relative to the overall HMCES SRAP structure.

5.2.10 How do processivity factors modulate SRAP interaction with DNA?

In *E. coli*, the Pol III β subunit acts as the sliding clamp during replication *(132)*. YedK does not have C-terminal tail and it is unknown whether YedK interacts with the Pol III β subunit. It would be interesting to test the interaction between YedK and the sliding clamp in bacteria, the Pol III β-subunit, since YedK does not contain the C-terminal extension that contains the PIP box in HMCES. It is unlikely that YedK directly interacts with the Pol III β-subunit since β-interacting motifs are typically found within an N- or C-terminal tail *(133)*. However, this potential interaction could be investigated via Flag-YedK immunoprecipitation assays and subsequent immunoblotting with FLAG or Pol III β-subunit antibodies.

HMCES directly interacts with PCNA through a PIP motif on HMCES C-terminal tail *(41)*. PCNA travels with the replication fork and acts as a processivity factor for DNA polymerases. It remains unknown how interaction with PCNA impacts HMCES crosslinking ability. Since HMCES can interact with PCNA, what is the structural basis for HMCES interaction with PCNA? Since multi-protein complexes are not typically amenable to crystallography, structural information of HMCES-PCNA interaction could be investigated via cryo-electron microscopy (cryo-EM) to investigate the overall structure of PCNA in complex with HMCES crosslinked to a stalled fork structure. HMCES-DPC formation kinetics in the presence of PCNA and modified PCNA, along with DNA junction structure preference, could be used to determine ideal cryo-EM targets. Potential constructs for cryo-EM could be screened for complex rigidity via small-angle x-ray scattering (SAXS) to determine optimal constructs and DNA substrates for use as cryo-EM targets.

PCNA can undergo post-translational modifications at Lys164 which can subsequently impact recruitment of replication and repair machinery to DNA *(32,134)*. Single stranded DNA that forms at stalled forks is coated by replication protein A (RPA) and attracts the Rad6 E2 enzyme and Rad18 E3 ubiquitin ligase that monoubiquitylates Lys164 of PCNA *(135)*. PCNA monoubiquitylation favors interaction with translesion DNA polymerases whereas polyubiquitylation of PCNA is thought to promote error-free repair via fork regression, although the specific pathway of DNA damage tolerance remains an open question *(134,136-138)*. Conversely, SUMOylation at K164 can lead to homologous recombination *(136)*. Increased REV1 and REV3 (the catalytic subunit of PolZ, an error-prone translesion polymerase) are present at replication forks in HMCES-deficient cells *(41)*.

HMCES is proposed to block translesion polymerases from synthesis across the AP site. Translesion polymerases can be recruited through monoubiquitylated PCNA, so HMCES may compete with translesion polymerases by also being recruited through monoubiquitylated PCNA. As such, it would be interesting to investigate HMCES crosslinking in the presence of monoubiquitinated PCNA compared to that with unmodified PCNA. Full-length HMCES crosslinking rate in the presence and absence of PCNA was investigated *in vitro* by rotation student Menghan Mei. He did not observe a significant difference between the PCNA condition and no PCNA. It is likely that a cellular context is needed to investigate the effect of PCNA and both PTMS of HMCES and PCNA may impact DPC formation and reversal.

Interestingly, HMCES and E3 ubiquitin ligase SHPRH inactivation are synthetically lethal *(41)*. Additionally, SHPRH is enriched at replication forks in ΔHMCES cells *(41)*. SHPRH transfers ubiquitin to K164 of monoubiquitinated PCNA, generating K64-linked polyubiquitin chains on PCNA *(139)*. It is important to note that K48-linked ubiquitin chains target proteins for proteosomal degradation, whereas K64-linked chains are involved in DNA repair *(140)*. Increased PCNA ubiquitylation via inactivation of ubiquitin-specific protease 1 (USP1) inhibits HMCES recruitment to replication forks *(41)*. As such, it is possible that SHPRH polyubiquitination of PCNA recruits machinery for an alternative error-prone pathway in the absence of HMCES.

5.2.11 Biological implications and potential outcomes of SRAP DPC

Due to the structural and sequential conservation of SRAP domain, the biochemical activities presented here are likely to be consistent across species. A recent publication suggests that HMCES acts downstream of UDG *(42)*. As such, my experiential setup of forming AP sites from treatment of uracil-containing DNA with UDG is likely biologically relevant. HMCES plays a critical role in a range of processes important to human health, including cancer, viruses, and aging *(108,113,141-143)*. HMCES expressed in all tissues, with the highest expression in lymph nodes *(144)* suggesting its importance to the immune system. Moreover, HMCES expression is significantly down regulated in both B-cell lymphoma and multidrug resistant osteosarcoma cells *(113,145)*. In this final section I discuss potential biological roles of HMCES and YedK in the context of recent discoveries. Although most of the discussion below is theoretical, I hope it inspires future investigations of these fascinating proteins.

Since YedK DPC exhibits reversal over time, I wanted to test whether the HMCES SRAP domain is also reversible. DPC was pre-formed with a 20-mer oligodeoxynucleotide containing an AP site (DPC-20), followed by addition of 4-fold excess of 40-mer AP-oligodeoxynucleotide to trap any hydrolyzed DPC-20 (Fig. 31). We observed the appearance of a 40-mer DPC (DPC-40) and a disappearance of DPC-20, consistent with direct reversal of the original DPC and reformation of DPC with the longer AP-containing oligo trap. The half-time of the exchange reaction under our experimental conditions was greater than 4 hours. This is longer than that of YedK at pH 7.0, likely because the reversal with HMCES was conducted at pH 8.0 since the protein was unstable at pH 7.0. Since DPC formation of both YedK and HMCES is pH dependent, the difference in pH may account for the slower reversal of HMCES. (HMCES exhibits DPC pH dependence, personal communication with Petria Thompson).

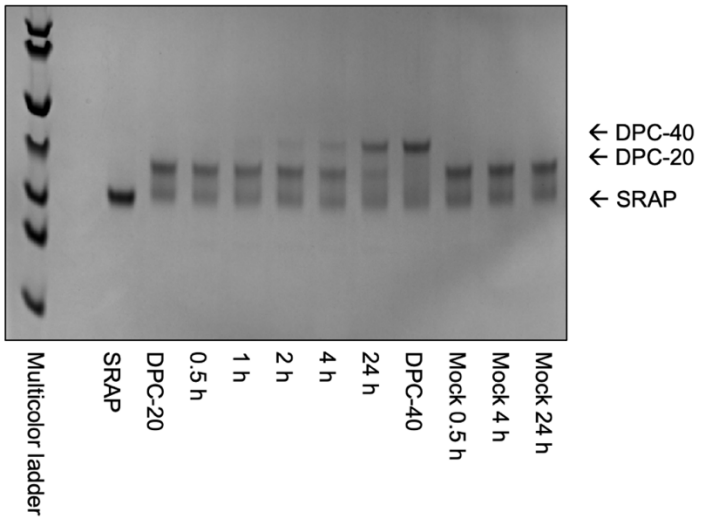

Figure 34. The HMCES SRAP DPC is reversible. Time course of HMCES SRAP DPC reversibility at 37°C for wild-type. DPC between YedK and 20-mer AP-DNA (DPC-20) was incubated with four-fold excess 40-mer AP-DNA or buffer and analyzed at the indicated time points. Free protein and DPC were visualized via Coomassie stained SDS-PAGE.

To enable a direct comparison between the reversal of YedK and HMCES, the assay could be repeated at the same pH where both proteins are stable, such as pH 6.0. The next step is to test DPC exchange after 24 hours with HMCES SRAP E127Q and H210Q. I am currently in the process of generating E127Q and H210Q HMCES SRAP mutant plasmids. Based on *in vitro*

results of E127Q and H210Q, a next step will be to test mutant DPC resolution in cells. H210Q in HMCES will be interesting to test in cells to see if it delays DPC resolution.

If in some cellular contexts DPC is indeed not proteolyzed and instead directly reversed to regenerate free enzyme, then it is not necessarily a suicide reaction and the enzyme can be reused. In contrast to other long-lived DPC intermediates, SRAP DPC reversal leaves an intact AP site (Chapter 4), rather than cleaving the AP site to generate a 3′-PUA *(146)*. Other 3′-PUA DPCs are formed by reaction with an internal AP site and are long-lived Schiff bases at equilibrium between the Schiff base and 3′-PUA, with the equilibrium favoring the Schiff base *(146)*. These Schiff base PUA-DPCs are proposed to protect the PUA until it can be repaired, ultimately placing the PUA back in context of duplex DNA where BER can continue.

YedK crosslinking to 3′-PUAs is a particularly interesting result and begs the question of whether this happens with HMCES in cells, and if so, what role does it serve? We showed that YedK forms DPCs to the products of bacterial Nth and Nei glycosylases, raising the possibility that both YedK and HMCES could protect against other protein crosslinks to 3′-PUAs generated during abortive base excision repair *(105,147)*. In addition to protecting AP sites from strand cleavage, the HMCES may also mark broken DNA ends for subsequent repair. This activity would contribute to HMCES-dependent reduction of spontaneous DNA strand breaks in cells. The PTMs of both HMCES and PCNA described in the section above may contribute to recruitment of subsequent repair factors. Future investigation of the stability and reversal of SRAP 3′-PUA DPC is warranted.

Interestingly, histone-PUA DPCs can be repaired directly by the nucleases APE1 and TREX1 *in vitro*, however this in the context of duplex DNA *(146)*. HMCES-deficient cells exhibit increased DSBs after APE1 treatment and both HMCES DPC and proteolyzed DpC block APE1 cleavage so it is unlikely that APE1 is involved in HMCES DPC resolution *(41,46)*. Inactivating both APE1 and APE2 simultaneously produced a significant increase in DSBs in HMCES-deficient cells *(71)* and HMCES and APE2 are synthetic lethal *(106)*. Taken together, this suggests that HMCES and AP endonucleases act in separate pathways and APE1 does not directly resolve HMCES DPC.

The YedK gene is in the gene neighborhood of the operon containing Ku *(36)*, and Ku binds DNA ends in non-homologous end joining (NHEJ) *(99)*. Ku prefers the same 3′ junction

DNA structure as SRAP proteins *(46,99)*. Given the ability of SRAP to form DPC with 3′-PUA DNA ends, one could speculate that SRAP DPC protects these ends after strand cleavage by Ku, possibly reducing deletion associated with NHEJ. Additionally, eukaryotic KU70/80 is involved in early NHEJ, whereas BRCA1 is involved in late-stage NHEJ *(148)*. Interestingly, loss of homologous recombination repair (HRR) factors BRCA1, and BRCA2 were HMCES synthetic lethal in HEK293T cells *(106)*. Conversely, no synthetic lethality was observed when BRCA1 or BRCA2 were inactivated in HMCES-deficient U2OS cells, however an increase in DSBs was observed *(41)*. It is possible that the difference in synthetic lethality is cell type dependent.

HMCES was recently found to facilitate DNA double-strand break repair through the microhomology-mediated alternative-end-joining (Alt-EJ) pathway *(149)*. Even though the Alt-EJ pathway with HMCES did not depend on Cys2, it has been hypothesized that HMCES DPCs with AP sites at 3′ junctions could facilitate end joining of DSBs, however this remains to be tested *(42)*. Additionally, it has been noted that AID introduces uracils at specific immunoglobulin loci and knockout of UDG disrupts antibody diversity, thus HMCES DPC with Ig-associated AP sites may also be important for antibody diversification *(42)*. HMCES reduces deletions in somatic hypermutation (SHM) and this activity is dependent on Cy2 *(42)*.

Presumably, the DPC prevents AP site cleavage in the ssDNA segments arising during the process of SHM, thereby preventing downstream deletions. The method of subsequent DPC resolution after AP site protection is unknown, but it is possible that direct reversal versus proteolysis of the DPC is context dependent. One outcome could be DPC proteolysis in an Ig context could allow for TLS gap filling past the AP site during SHM, as suggested by Wu and colleagues *(42)*. In other somatic cells, a different outcome could be direct reversal of the DPC after replication restart past the lesion is completed. Since SRAP cannot bind full duplex DNA, reversal or resolution of the DPC would likely place the AP site opposite a gap that could then be repaired in an error-free manner by homologous recombination *(150)*.

Table 6. Oligodeoxynucleotides used in these studies

Oligo Name [a]	Sequence (5′→3′) [b]	Notes
Quikchange mutagenic primers		
YedK N75A FWD	TGGGATAAACCGCCGCTGATTGCCGCCCGCGTAG	YedK N75A
yedK N75A REV	CTACGCGGGCGGCAATCAGCGGCGGTTTATCCCA	yedK N75A
yedK N75D FWD	ACGCGGGCGTCAATCAGCGGCGGTTTATCC	yedK N75D
yedK N75D REV	GGATAAACCGCCGCTGATTGACGCCCGCGT	yedK N75D
yedK E105Q FWD	CTTTTTTCCACTGAAACCAGCCATCGGCAAAACAG	yedK E105Q
Yedk E105Q REV	CTGTTTTGCCGATGGCTGGTTTCAGTGGAAAAAAG	Yedk E105Q
yedK E105A FWD	GCCTTCTTTTTTCCACGCAAACCAGCCATCGGC	yedK E105A
yedK E105A REV	GCCGATGGCTGGTTTGCGTGGAAAAAAGAAGGC	yedK E105A
Yedk E105D FWD	GTTTTGCCGATGGCTGGTTTGATTGGAAAAAAGAAGG	Yedk E105D
yedK E105D REV	CCTTCTTTTTTCCAATCAAACCAGCCATCGGCAAAAC	yedK E105D
yedK P164S FWD	CGACAGTACCAGTGAGCGGCGGTCATGAA	YedK P164S
yedK P164S REV	TTCATGACCGCCGCTCACTGGTACTGTCG	YedK P164S
YedK_C2A_For	GGCAAAGCGTCCAGCCATATGTCTCT	YedK C2A
YedK_C2A_Rev	AGAGACATATGGCTGGACGCTTTGCC	YedK C2A
YedK_C2S_For	CAAAGCGTCCACTCATATGTCTCTC	YedK C2S
YedK_C2S_Rev	GAGAGACATATGAGTGGACGCTTTGC	YedK C2S
YedK_H160A_For	GGCGGCGGTCAGCAATATCTACCAGACCTTGATCTG	YedK H160A
YedK_H160A_Rev	AGATCAAGGTCTGGTAGATATTGCTGACCGCCGCCC	YedK H160A
YedK_H160D_For	GCGGCGGTCATCAATATCTACCAGACCTTGATCT	Did not work
YedK_H160D_Rev	GATCAAGGTCTGGTAGATATTGATGACCGCCGCC	Did not work
H160N_101819_F	GCGGCGGTCATTAATATCTACCAGACCTTGATCT	Not tested since Q5 worked
H160N_101819_R	GATCAAGGTCTGGTAGATATTAATGACCGCCGCC	Not tested since Q5 worked
E127Q_F_qc	GACAGCGCTGCCACTGGTAAAAACCGTCCGCAAG	HMCES E127Q
E127Q_R_qc	CTTGCGGACGGTTTTTACCAGTGGCAGCGCTGTC	HMCES E127Q
H210Q_F_qc	CGGGCATGCGGTGCTGAATGTCGCTCAAC	HMCES H210Q
H210Q_R_qc	GTTGAGCGACATTCAGCACCGCATGCCCG	HMCES H210Q
Q5 mutagenic primers		
H160D_101819_F	GGTAGATATTGATGACCGCCGCC	YedK H160D
H160DQ_101819_R	AGACCTTGATCTGCCGCAG	YedK H160D, YedK H160Q reverse primer
H160Q_101819_F	GGTAGATATTCAAGACCGCCGCC	YedK H160Q
H160E_110920_F	GGTAGATATTGAAGACCGCCGC	YedK H160E

H160E_110920_R	AGACCTTGATCTGCCGCA	YedK H160E
delC_010620_F	GGACGCTTTGCCCAATCC	YedK △C2
delC_010620_R	CATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC	YedK △C2
HISdelC_010620_F	GAAAACCTGTATTTTGAGGGACGCTTTGCCCAATCC	Incorrect primer (no N-terminal methionine)
HISdelC _010620_R	GTGGTGGTGATGATGATGCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC	Incorrect primer (no N-terminal methionine)
H160N_103119_F	GGTAGATATTAATGACCGCCGCC	YedK H160N
H160N_103119_R	AGACCTTGATCTGCCGCA	YedK H160N
NTEV_F	CACGAAAACCTGTATTTTCAGGGACGCTTTGCCCAATCC	YedK △C2 with TEV-cleavable N-terminal His tag (not needed)
NTEV_R	ATGATGGTGATGATGATGTCCCATATGTCTCTCTCTCTCGCG	YedK △C2 with TEV-cleavable N-terminal His tag (not needed)
delCtermHis_F	TGAGCGGCCGCGAGAGAG	To remove C-terminal 6XHIS (not needed)
delCtermHis_R	TGGAACAGAACTTCCAGAACAGGTTG	To remove C-terminal 6XHIS (not needed)
Neil3V2C_2/4/2021_F	TTAACTGATGTGTGAAGGTCCAGGC	Neil3 V2C
Neil3V2C_2//2021_R	TCTCCTTCTGTACCAGTTTAC	Neil3 V2C
E127Q_2022Q5_F	TGGATTCTATCAATGGCAGCGATG	HMCES E127Q
E127Q_2022Q5_R	TCTGCTAAAACGACACAG	HMCES E127Q
H210Q_2022Q5_F	GAGTGACATCCAACACAGGATGC	HMCES H210Q
H210Q_2022Q5_R	AAGCCTTTGCAGGAATCC	HMCES H210Q
Crosslinking Assays		
FAM_U_Cy5	FAM-CGGGCGGCGGCAUAGGGCGCGGGCCTTTTTT-Cy5	Crosslinking kinetics
FAM_U_20	FAM-TCTTCTGGTCUGGATGGTAGT	DPC reversal over time
40_U	GGAATCTGACTCTTCTGGTCUGGATGGTAGTTAAGTCTTGT	DPC reversal over time
FAM_U_35	FAM-ATGACTCTTCTGGTCUGGATGGTAGTTAAGTCTTGT	DPC reversal
Crystallography		
H36	GTCUGGA	YedK DPC structures
HMCES 58	GTC/idSp/GGA	YedK and HMCES SRAP THF structures
HMCES 59	GTC/iSpC3/GGA	YedK C3-spacer structures
3 mer	TCC	HMCES SRAP junction crystallography
dT_THF	TTTTT/idSp/TTTT	HMCES SRAP THF crystallography

dT_U	TTTTT/ideoxyU/TTTT	HMCES SRAP DPC crystallography
10dT	TTTTTTTTTT	HMCES SRAP non-covalent crystallography
7dT_U	TTT/ideoxyU/TTT	HMCES SRAP DPC crystallography
5dT_U	TT/ideoxyU/TT	HMCES SRAP DPC crystallography
DNA binding		
KA_FAM	FAM-ATGACTCTTCTGGTC/idSp/GGATGGTAGTTAAGT	YedK ss-dsDNA junction anisotropy
KA_1	GACCAGAAGAGTCAT	YedK ss-dsDNA junction anisotropy
KA_2	ACTTAACTACCATCC	YedK ss-dsDNA junction anisotropy
FAM_THF_15	FAM-TCTGGTC[THF]GGATGGT	YedK anisotropy
FAM_THF	FAM-CGGGCGGCGGCA/idSp/AGGGCGCGGGCCTTTTT	YedK ss-dsDNA junction anisotropy
FA_5_junction	TGCCGCCGCCCG	YedK ss-dsDNA junction anisotropy
FA_3_junction	AAAAAGGCCCGCGCCCT	YedK ss-dsDNA junction anisotropy

Table 7. Plasmids used in these studies[67]

		(bp)	(bp)	(bp)	(5'/3' sites)	
HMCES						
addgene-HMCES-SRAPd-5KO9	H. sapiens HMCES SRAP domain, CT 6X HIS, TEV cleavable tag	5228	896	6124	XbaI/ BamHI	Rosetta 16°C, 0.3mM IPTG
pBG101-HMCES-SRAP (PPS)	H. sapiens HMCES SRAP domain, CT 6xHis-GST PPS cleavable tag	5207	961	6168	NdeI/ Not1	BL-21(RIL) 16°C, 0.5mM IPTG
pBG101-HMCES-SRAP C2A	H. sapiens HMCES SRAP domain C2A, CT 6xHis-GST PPS cleavable tag	5207	961	6168	NdeI/ Not1	BL-21(RIL) 16°C, 0.5mM IPTG
pBG101-HMCES-SRAP C2S	H. sapiens HMCES SRAP domain C2S, CT 6xHis-GST PPS cleavable tag	5207	961	6168	NdeI/ Not1	BL-21(RIL) 16°C, 0.5mM IPTG
ppt5 FL HMCES WT codon optimized	H. sapiens HMCES (full length) CT 6X HIS tag (not cleavable) in modified pBG101	5335	1064	6399	NdeI/ Not1	BL-21(RIL) 16°C, 0.5mM IPTG
pKM294 pBG101-GST-C3orf37 2-354	H. sapiens HMCES (full length) NT 6X-His-GST tag, PPS-cleavable, leaves scar, reduced crosslinking activity	6032	1069	7101	BamHI/ Not1	BL-21(RIL) 16°C, 0.5mM IPTG
YedK						
pBG101-YedK (same as ppt11)	YedK, CT 6xHis- PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-C2A	YedK C2A, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-C2S	YedK C2S, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-N75A-1	YedK N75A, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-N75D-2	YedK N75D, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-E105A-1	YedK E105A, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-E105D-1	YedK E105D, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-E105Q	YedK E105Q, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG

pBG101-YedK-H160A-1	YedK H160A, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-H160D-2	YedK H160D, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-H160N-1	YedK H160N, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-H160Q	YedK H160Q, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-H160E-2	YedK H160Q, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-P164S-1	YedK P164S, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-deltaC-2	YedK deltaC, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	711	6018	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-deltaC-2-E105A	YedK deltaC, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	711	6018	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-deltaC2-H160A	YedK deltaC, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	711	6018	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-C2A-E105Q	YedK C2A, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-C2A-H160Q	YedK C2A, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-deltaC2-E105Q	YedK deltaC, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	711	6018	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-deltaC2-H160Q	YedK deltaC, CT 6xHis-PPS cleavable tag in modified pBG101*	5307	711	6018	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-E105A-H160A	YedK E105A/H160A, CT 6xHis- PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG
pBG101-YedK-E105A-H160A	YedK E106Q/H160Q, CT 6xHis- PPS cleavable tag in modified pBG101*	5307	714	6021	NdeI/ Not1	BL-21 16°C, 0.3 mM IPTG

[6] All plasmids sequence-verified

[7] All plasmids used in these studies have kanamycin antibiotic resistance

*no N-terminal 6X-HIS tag

REAGENTS.

Reagents were purchased or obtained from the following suppliers or labs and were of the highest purity available. Oligonucleotides were purchased from Integrated DNA Technologies (IDT) (Table 6). Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich, and all enzymes were purchased from New England Biolabs (NEB). The aoN-g probe was a gift from Yasuo Komatsu at National Institute of Advanced Industrial Science and Technology (AIST). Plasmids for HMCES, HMCES SRAP domain, and pBG101 YedK WT were obtained from Petria Thompson in the Cortez lab (2019). HMCES 5KO9 plasmid was from Addgene.

PROTEIN EXPRESSION AND PURIFICATION.

Site-directed mutagenesis.

YedK mutants (C2A, C2A/E105Q, N75A, E105A, E105D, E105Q, H160A) were generated using the QuikChange Site-Directed Mutagenesis Kit (Agilent). The forward and reverse mutagenic extension reactions were performed separately to improve primer annealing, and the corresponding single stranded copies of the plasmid were combined. YedK point mutants (ΔC2, H160E, H160Q, C2A/H160Q) were generated using the Q5 Site-Directed Mutagenesis Kit (NEB). HMCES SRAP mutants C2A and C2S were a gift from Petria Thompson. Mutant plasmids were sequence verified (GenHunter). Plasmids are listed in Table 6.

YedK expression and purification.

Escherichia 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 Luria broth (LB) containing 15 ng/mL kanamycin at 37 °C to 0.8 OD_{600}, and YedK overexpression was induced at 16 °C for 16 hr 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 RPM for 30 min and passed through a 22-gauge needle prior to loading onto a 5 mL Ni-NTA column. The column was washed

with 6-column volumes lysis buffer with 20 mM imidazole, and bound proteins were eluted with lysis buffer with 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 with Amicon MWCO 10 kDa centrifugal filters. Protein aliquots were flash-frozen in liquid nitrogen and stored at -80 °C.

Mutant YedK expression and purification.

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. YedK was exchanged into S200 buffer (20 mM TRIS-HCl pH 8.0, 100 mM NaCl, 10% glycerol, 2 mM TCEP), concentrated, flash-frozen in liquid nitrogen, and stored at -80 °C. Wild-type YedK for biochemical experiments shown in figures 12A-B, 14B, and 15 was further purified by gel filtration on a 16/600 Superdex 200 column (GE Healthcare) in S200 buffer.

ESI-MS intact protein analysis.

The absence of the N-terminal methionine in HMCES SRAP domain, YedK, and YedK mutants was verified via intact mass electrospray ionization (ESI) mass spectrometry (Vanderbilt Mass Spectrometry Core). In brief, purified protein was subjected to electrospray ionization, which produces peaks of multiply charged intact ions of the protein of interest. The multiple peaks are due to adduction of ions and the adduct mass can be estimated by the mass differences of the adducted ion. Each peak is an independent measure of the mass of the parent species so the signal can be averaged using a single spectrum. The deconvolution algorithm developed by Fenn *et. al.* transforms multiple peaks into a single peak, allowing determination of the intact mass of the protein of interest *(151)*. To evaluate the presence or absence of an N-terminal methionine, I calculated the expected intact masses of each protein from their amino acid sequence (ExPASy) with and without the N-terminal methionine. The N-terminal methionine would add an additional 131 Da. All experimental masses were within 20 Da of the expected mass without N-terminal methionine. Percent error was calculated (|observed MW-theoretical MW|/|theoretical MW|)*100

HMCES SRAP expression and purification.

Human HMCES SRAP domain (amino acids 1-270) utilized in Chapter 2 was expressed in a modified pBG101 vector containing a Rhinovirus 3C (PreScission) protease cleavable hexahistidine tag. HMCES SRAP was purified similar 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 prior to S200 size exclusion chromatography in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol, and 10 mM DTT. Mutant SRAP C2A plasmid and C2S plasmid were obtained from the David Cortez lab at Vanderbilt and purified the same as above.

Human HMCES SRAP domain with a TEV cleavage site was obtained from Addgene for early crystallography trials (chapter 5). The target protein was over-expressed in Rosetta II *E. coli* in Terrific Broth medium with 30 ug/mL kanamycin. When the OD_{600} of the culture reached 1.7, the flasks were placed on ice 30 min and the culture was induced with 0.3 mM final IPTG concentration then returned to shaking at 16°C for 16 hours. The cells were harvested, homogenized, and bound to Ni-NTA resin as described above. The column was washed with 6x column volume lysis buffer/20 mM imidazole. Bound proteins were eluted using 6x column volume lysis buffer/300 mM imidazole. The N-terminal His-tag was removed by overnight incubation with TEV protease (1:30 w/w) at 4 °C during dialysis (50 mM Tris-HCl pH8.0, 300 mM NaCl, 2 mM TCEP). Uncut proteins and TEV protease were removed by passing the solution through 2mL Ni-NTA beads. After repass over the Ni-NTA column, HMCES SRAP was purified via anion exchange via a HiTrap Q column prior to S200 size exclusion chromatography in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol, and 10 mM DTT.

DNA BINDING.

HMCES SRAP binding to DNA junctions.

Sequences of oligonucleotides used in the biochemical assays are listed in Table 5. Relative binding affinity was measured by EMSA using ^{32}P-labeled DNAs containing a deoxyuracil. 1 nM DNA 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 BSA) at 37 °C for 1 hr. Ficoll was added to a final concentration of 1.25% and the samples were resolved on a 10%

polyacrylamide gel in 1X 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.

Blocked YedK binding to ssDNA.

Relative binding affinities of formaldehyde-blocked YedK were measured by fluorescence anisotropy using ssDNA (FAM_THF_15) 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 Buffer B in a 384-well plate for 20 min at 4°C. Binding was monitored by a change in fluorescence polarization as described above.

CROSSLINKING ASSAYS

Preparation of AP-DNA.

AP-DNA was prepared by incubating 1 μM uracil-containing oligonucleotide with 0.6 U UDG (New England Biolabs) *(152)* in UDG Buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT) at 37 °C for 15 m. AP-DNA was prepared fresh for each reaction. Sequences of oligonucleotides used in biochemical assays are listed in Table 5.

DPC stability over time and temperature.

For the experiments shown in Figs. 5a-c, AP-DNA was prepared by incubating 50 μM uracil-containing oligonucleotides with 25 units of uracil DNA glycosylase*(152)* (UDG, New England Biolabs) in Buffer X1 (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 X1 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. 5c, DPCs were formed at 37 °C for 12 hr and

treated with either no heat or 95 °C for 2 min prior to incubation at 25 °C. Free and DNA-crosslinked HMCES were separated on 10% polyacrylamide Tris-glycine gels.

Proteolyzed DpC stability.

Reaction products were separated on 15% polyacrylamide urea gels in 1X TBE buffer. In Fig. 5d, AP-DNA was prepared by incubating 100 nM uracil-containing ssDNA with 1 unit of UDG in Buffer X1, crosslinks 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 hr, followed by proteinase K (Sigma Aldrich) digestion for 5 min. In Fig. 5e, 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 hr. 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.

Schiff base trapping.

E. coli YedK DPCs 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 hr. Schiff base intermediates 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 hr.

HMCES SRAP crosslinking to DNA junctions.

DNA binding reactions with ssDNA-dsDNA junctions were carried out with 10 nM DNA and increasing concentrations of HMCES SRAP at 37 °C for 1 hr in Buffer X1. Crosslinking reactions with ssDNA-dsDNA junctions were carried out with 1 nM AP-DNA and increasing concentration of HMCES SRAP at 37 °C for 1 hr in Buffer X1.

YedK DNA-protein crosslinking kinetics.

YedK DPCs were formed by incubation of 1 μM protein and 35 nM 5′-FAM-labeled AP oligonucleotide (FAM_U_Cy5) at 25 °C in Buffer A (20 mM Tris-HCl pH 6.0, 10 mM NaCl, 1 mM EDTA, and 5 mM DTT). Reactions were stopped at various time points by adding an equal volume of SDS Buffer (100 mM Tris-HCl pH 6.9, 16% glycerol, 3.2% SDS, 6% formamide, 0.5% β-mercaptoethanol) and incubating on ice. To confirm generation of AP sites, 35 nM 5′-FAM-labeled AP oligonucleotide in Buffer A was treated with 0.2 M NaOH for 3 m at 70°C. For Schiff base trapping**,** $NaCNBH_3$ was added to the YedK and AP-DNA mixture to a final concentration of 50 mM. All samples were heated 70°C for 1 m prior to loading the gel. DPC and AP-DNA were separated on 4-12% Bis-Tris gels (Invitrogen) pre-run with MES SDS Running buffer (Invitrogen). 5′-FAM-labeled AP oligonucleotide was visualized on a Typhoon Trio (GE Healthcare) using excitation and emission wavelengths of 532 and 575 nm. Band intensities was quantified using GelAnalyzer 19.1 (www.gelanalyzer.com).

For the WT pH dependence experiments, YedK DPCs were formed by incubation of 500 nM protein and 35 nM 5′-FAM-labeled AP oligonucleotide (FAM_U_Cy5) in Buffer B (20 mM Tris-HCl, 15 mM sodium citrate, 5 mM citric acid, 5 mM NaCl, 1 mM EDTA, 5 mM DTT, pH adjusted with HCl/NaOH and 0.22 μm filtered). AP-DNA was preincubated in reaction buffer at 18 °C for 10 m prior to addition of YedK and incubation at 18 °C at given time points over the course of 30 m. For the E105Q pH dependence experiments, YedK DPCs were formed by incubation of 1 uM protein and 35 nM 5′-FAM-labeled AP oligonucleotide (FAM_U_Cy5) in Buffer B 25 °C over the course of 60 m. Reactions were quenched, analyzed by SDS-PAGE, and visualized by FAM fluorescence.

Peptide and aoN-g crosslinking.

Aldehyde reactive probe analog, aoN-g *(125)*, was a gift from Yasuo Komatsu at National Institute of Advanced Industrial Science and Technology (AIST). YedK peptide consisting of the amino acids 2-16 (CGRFAQSQTREDYLA) was synthesized by Genscript. 50 nM 5′-FAM-labeled AP-DNA (FAM_U_20) was incubated at 25 ^{O}C with saturating concentrations of aoN-g (25 μM), YedK peptide (1 mM), or YedK (1 μM) in Buffer C (20 mM HEPES pH 7.0, 10 mM

NaCl, 1 mM EDTA, 5 mM DTT). DNA-probe and DNA-peptide reactions were quenched by adding 8 µL reaction to 8 µL Stop Buffer (40 mM EDTA-Na2, 8M urea, 20 µM glutaraldehyde) and 4 µL of Loading buffer. YedK DPC reactions were stopped by adding an equal volume of SDS Buffer. All reactions were heated to 70 °C 1 m prior to loading gel. DNA-probe or DNA-peptide adducts were separated from AP-DNA on 15% polyacrylamide urea gels prerun in 0.5 X TBE buffer (50 mM Tris-HCl pH 8, 45 mM boric acid, 1 mM EDTA). DPCs were resolved via 4-12% Bis-Tris gels (Invitrogen) prerun in MES SDS running buffer (Invitrogen), and FAM-DNA visualized by fluorescence.

N-terminal blocking.

YedK and mutants stored in S200 buffer were thawed, spun 20,000 x g for 10 minutes, and diluted 10 µM in Buffer B at pH 7.0. Formaldehyde was added to a final percentage of 1% and reactions were incubated at 25 ºC for 2 h before quenching by addition of 125 mM glycine. Reactions were buffer exchanged into fresh Buffer B pH 7.0 using G-25 desalting columns (Cytiva). 3.5 µM of formaldehyde-treated protein was incubated with 35 nM AP-DNA (FAM_U_35) in Buffer B pH 7.0 for 60 m at 37 ºC. Reaction products were resolved via 4-12% Bis-Tris gels in MES running buffer and FAM-DNA visualized by fluorescence.

Lyase kinetics and Schiff base trapping over time

Experiments to measure C2A lyase kinetics were performed the same as DNA-protein crosslinking experiments, with C2A mutants used in place of wild-type YedK with the following modifications: reactions were incubated at 37 ºC and stopped at given time points with equal volumes of Loading buffer (80% w/v formamide, 10 mM EDTA, 3 µg/µL Blue dextran) and reaction products were resolved via 10% polyacrylamide urea gels prerun in 0.5 X TBE buffer (50 mM Tris-HCl pH 8, 45 mM boric acid, 1 mM EDTA). Experiments to measure C2A Schiff base trapping kinetics were performed the same as DNA-protein crosslinking experiments, with C2A mutants used in place of wild-type YedK with the following modification: 50 mM $NaBH_3CN$ was added to the reaction mixture prior to taking the first time point.

YedK reaction with lyase products.

500 nM of YedK, EndoIII, EndoVIII, or YedK C2A was incubated with 35 nM 5′-FAM-labeled AP-DNA (FAM_U_35) in Buffer B pH 7.0 at 37 ºC for 60 m. Aliquots were removed from each reaction, quenched by addition of an equal volume of SDS Buffer, and placed on ice. To each of the remaining reactions, fresh YedK or buffer as indicated was added to a final concentration of 500 nM and incubated at 37 ºC for another 20 m. Reactions were stopped by adding 10 µL DPC reaction to 10 µL SDS Buffer and placed on ice. Reaction products were resolved on 4-12% Bis-Tris gels (Invitrogen) and FAM-DNA was visualized using excitation and emission wavelengths of 495 and 519 nm on a Chemidoc (BioRad).

3′-PUA reaction with YedK mutants.

3′-PUA-DNA was generated by incubation of 1 µM 5′-FAM-labeled AP-DNA (FAM_U_35) with 20 U EndoIII (NEB) in Buffer B pH 7.0 at 37 ºC for 60 m. 500 nM YedK was incubated with 35 nM 3′-PUA-DNA in Buffer B pH 7.0 at 37 ºC for 60 m. Reactions were stopped with an equal volume of SDS Buffer. Reaction products were resolved on 4-12% Bis-Tris gels (Invitrogen) and visualized by FAM fluorescence.

Crosslink reversal assays.

For thermal DPC denaturation experiments, YedK DPC was formed by incubation of 35 nM 5′-FAM-labeled AP-DNA (FAM_U_35) with 500 nM YedK in Buffer C pH 6.5 for 60 m at 37 ºC. The reaction was heated at 90 ºC for 10 m to hydrolyze DPC, followed by addition of either buffer, fresh AP-DNA, or fresh YedK to the hydrolyzed DPC mixture and incubated for 10 m at 37 ºC. Reactions were performed in HEPES rather than Tris buffer to avoid amines in the buffer being a confounding factor leading to strand cleavage *(153)*. Reactions were stopped by adding an equal volume (10 µL) of SDS Buffer, products resolved on 4-12% Bis-Tris gels and visualized by FAM fluorescence.

Reversal trapping experiments were performed by incubating 10 µM 20-mer AP-DNA (FAM_U_20) and 2 µM YedK in Buffer C pH 7.0 for 18 h at 37 ºC, which led to >90% DPC-20 formation. DPC-20 was incubated with a 4-fold excess (40 µM) of 40-mer AP-DNA (40_U) to trap any hydrolyzed DPC-20. Reactions were quenched with equal volumes of SDS Buffer. Each

time point was initiated in reverse so that all reactions were quenched for the same length of time. Reaction products were resolved on 4-12% Bis-Tris gels and Coomassie stained for protein *(46)*.

DNA-peptide crosslink reversal over time.

25 nM 20-mer AP-DNA (FAM_U_20) and 0.5 mM YedK peptide were incubated in Buffer C pH 7.0 for 1 h at 37 ºC to form 100% DNA-peptide crosslink (DpC). 2 mM aoN-g probe was added and reactions were incubated at 37 ºC. Reactions were quenched by mixing 8 µL reaction, 8 µL 2X glutaraldehyde Stop Buffer, and 4 µL Loading buffer. Time points were initiated in reverse to maintain equal quenching times. Products were resolved on 15% polyacrylamide urea gels prerun in 0.5 X TBE buffer (50 mM Tris-HCl pH 8, 45 mM boric acid, 1 mM EDTA) and DNA was visualized via FAM fluorescence.

HMCES SRAP DNA-protein crosslinking and lyase assays.

The DNA sequence used was FAM-d(CGG GCG GCG GCA UAG GCG CGG GCC TTT TTT)-Cy5 with 6-carboxyfluorescein (FAM) and Cy5 dyes covalently attached at the 5′ and 3′ ends, respectively. UDG (1.5 µM) was incubated with 50 µM DNA for 2 hours at 37 °C in UDG reaction buffer (20 mM TRIS-HCl, 1 mM TCEP, 1 mM EDTA, pH 7.0) to generate abasic sites (AP-DNA) at the position of the dU. The AP-DNA was diluted to a final concentration of 2 mM and incubated with 1 mM SRAP for 30 minutes (or other time where indicated). Reactions were run on 15% TBE-urea gels and imaged on a Typhoon Trio with 526 nm and 650 nm lasers to visualize FAM- and Cy5-labeled DNA. The same reactions were run concurrently on 4-12% Bis-Tris gels and Coomassie stained to visualize protein. All dilutions were made in UDG reaction buffer. Where indicated, samples were heated to 70°C for 1 minute before loading gel. DNA band volumes were quantified on GelAnalyzer 2010 (Lazar Software). For quantification of the fraction SRAP DPC, only heat (70 °C) denatured samples were analyzed.

Statistics and Reproducibility

Experiments were completed at least three times unless otherwise indicated.
Structures were deposited in the Protein Data Bank under accession codes 6NUA (DPC), 6NUH (C3-spacer), 8D2M (C2A trapped Schiff base).

X-ray Crystallography

Preparation of SRAP DPC for crystallization.

Abasic (AP) DNA was generated by treating 10-mer ssDNA, d(TGGTCUGGAT), was incubated with 5 μM UDG at 37 °C for 30 minutes to make AP sites at the position of the dU. SRAP stored in S200 buffer at 95% purity after anion exchange was then thawed and incubated with AP-DNA for 1 hour at 37 °C in the presence of UDG reaction buffer (NEB) and 1 mM DTT. Crosslinking reactions were then buffer exchanged into 50 mM NaCl MonoQ buffer (50 mM NaCl, 1 M TRIS pH 8.0, 2 M MgCl2, 2 mM DTT). SRAP DPC was enriched and separated from unbound protein via anion exchange on a MonoQ 5/50 GL column and eluted with a NaCl gradient. Fractions containing greater that 60% DPC were pooled, concentrated, and buffer exchanged for setup into crystallization trays on the same day.

SRAP DPC with 10-mer ssDNA was buffer exchanged into SRAP Setup Buffer (80 mM NaCl, 20 mM TRIS pH 8.0, 2 mM TCEP, 0.5 mM EDTA). Crystal trays were set up with DPC both directly after MonoQ step and after subsequent size exclusion step. Initial crystallization broad screens were set up in 96-well trays (Axygen) with 250 nL protein and 250 nL mother liquor on the Moquito robot (PAC-Van) and imaged via RockImager (Formulatrix). UV-positive hits from these screens were further optimized in 24-well sitting-drop and hanging-drop trays with variations of initial hit conditions. Small crystals were obtained from variations on 20% PEG 3350, 0.2 M Sodium formate pH 7.02, 8 mg/mL DPC. Limited nucleolysis of bound DNA was investigated *in crystallo* by adding 0.1 U Benzonase (Millipore) to DPC reactions for 1 hour at 37°C before being placed direcly into 96-well sitting drop trays. SRAP alone after initial anion exchange step was crystallized in 26% PEG3350, 0.05 M BTP; 2% Tacsimate pH 6.0, 18 mg/mL (TEV-cleaved SRAP); 1 μL:1 μL drop, 21°C.

X-ray crystallography of YedK DPC and YedK noncovalent complex.

AP-DNA was prepared by incubating 50 μM 7-mer d(GTCUGGA) ssDNA with 2.5 units of uracil DNA glycosylase (UDG, New England Biolabs) in Buffer X1 at 37 °C for 30 min. YedK DPC was generated by incubation of 20 μM YedK with 25 μM AP-DNA for 1 hr 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 ºC by mixing equal volumes of 3 mg/mL 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 days 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 7-mer 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 7-mer C3-spacer ssDNA at 4 °C for 30 min. Crystals were grown by hanging drop vapor diffusion at 21 ºC from drops containing 2 μL DNA-protein 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 days 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 *(154)*. Data collection statistics are provided in Table 8. Phasing and refinement was carried out using the PHENIX suite of programs *(155)*. Phasing of the DPC structure was carried out by molecular replacement of a previously determined structure of YedK alone (PDB accession 2ICU). The protein was subjected to simulated annealing, atomic coordinate, temperature factor, and TLS refinement prior to building the DNA model. The entirety of the 7-mer ssDNA and the Cys2-DNA crosslink was readily apparent in the density maps. All seven nucleotides and the Cys2-AP crosslink were manually built in Coot *(156)*, 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 structure of PDB ID 5FF2, and the stereochemistry of AP-site and Cys2 ring substituents verified my manual inspection of the electron density prior to model building. The DNA-protein 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 prior to 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, TLS refinement, the C3-spacer and the 3′-end of the DNA was 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 Cys2 removed from the structure factor calculation were used to guide placement and refinement of the crosslink 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 DataBank under accession codes 6NUA (DPC) and 6NUH (C3-spacer).

All structural biology software was curated by SBGrid *(157)*. Structure images were created in PyMOL (https://pymol.org). Sequence conservation was mapped onto the structure using the Consurf Server *(158)*. 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.

Table 8. X-ray data collection and refinement statistics for YedK DPC and non-covalent complex.[a]

	SRAP DPC (YedK-AP-DNA)	**Non-covalent complex (YedK/C3spacer-DNA)**
Data collection		
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)	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)
Avg. $I/\sigma I$	14.8 (1.9)	21.3 (2.6)
Completeness (%)	97.4 (95.2)	98.5 (91.1)
Redundancy	4.4 (2.9)	4.1 (4.0)
Wilson *B*-factor (Å^2)	18.0	12.3
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}	0.171 (0.255)	0.143 (0.169)
R_{free}[b]	0.222 (0.292)	0.177 (0.205)
No. atoms[c,d]		
Protein	3,627	1,801
DNA	268	131
Water	280	209
Other	0	14
Avg. *B*-factors[c,d,e] (Å^2)		
Protein	26.0	16.9
DNA	28.7	63.9
Water	29.3	24.4
Other	-	38.0
R.m.s. deviations		
Bond lengths (Å)	0.010	0.008
Bond angles (°)	1.035	0.963

[a] Statistics for the highest resolution shell are shown in parentheses.
[b] R_{free} was determined from the 5% of reflections excluded from refinement.
[c] Riding hydrogen atoms were not included in no. atoms or avg. *B*-factors.
[d] Other: bis-tris buffer (C3).
[e] Equivalent isotropic *B*-factors were calculated in conjunction with TLS-derived anisotropic *B*-factors

X-ray crystallography of YedK C2A borohydride-trapped DPC.

AP-DNA was prepared by incubating 50 μM 7-mer ssDNA [d(GTCUGGA]] with 10 U of uracil DNA glycosylase (UDG, New England Biolabs) in UDG Buffer at 37 °C for 1.5 h. YedK C2A protein was buffer exchanged into Buffer C. The Schiff base intermediate was trapped by incubating 24 μM YedK C2A with 25 μM AP-DNA at 25 °C for 5 m, adding $NaBH_3CN$ to a final concentration of 50 mM, and incubating at 25 °C for 18 h. DPC was further purified by gel filtration on a 16/300 Superdex 200 column (GE Healthcare). Fractions containing >90% DPC were pooled and buffer exchanged into 80 mM NaCl, 20 mM TRIS-HCl pH 8.0, 1 mM TCEP, 0.5 mM EDTA for crystallization experiments. DPC was crystallized by hanging drop vapor diffusion at 21 ºC by mixing equal volumes of 2 mg/mL protein and reservoir solution containing 25% (wt/vol) PEG 3350 and 0.2 M NaH_2PO_4. Crystals were harvested 22 days after setting the drops, cryoprotected in 30% PEG 3350, 0.2 M NaH_2PO_4, and 10% (vol/vol) glycerol, and flash-cooled in liquid nitrogen.

X-ray diffraction data were collected at the Advanced Photon Source beamline 21-ID-F at Argonne National Laboratory and processed with HKL2000 *(154)*. Data collection statistics are shown in Table 9. Phasing and refinement were carried out using the PHENIX suite of programs *(155)*. Phases were determined by molecular replacement using the protein model from the YedK DPC structure (PDB accession 6NUA) followed by simulated annealing to eliminate model bias prior to further refinement. After refinement of atomic coordinates, temperature factors, and TLS-derived anisotropic B-factors, DNA was manually built in Coot, guided by $2mF_o$-DF_c and mF_o-DF_c electron density maps. The Ala2-DNA crosslink as well as the entirety of the 7-mer ssDNA was readily apparent in the density maps. To minimize model bias, annealed mF_o-DF_c omit maps were calculated by removing the Ala2 and the AP-site of the DNA. Geometry restraints for the linkage were generated from idealized coordinates of a reduced Schiff base (ChemDraw). The DNA-protein model was iteratively refined by energy minimization and visual inspection of the electron density maps. The final YedK-C2A-DNA model was validated using the wwPDB Validation Service and contained no residues in the disallowed regions of the Ramachandran plots. Refinement and validation statistics are presented in Table S4. All structural biology software was curated by SBGrid *(157)*. Structure images were created in PyMOL (https://pymol.org). The structure was deposited in the Protein Data Bank under accession code 8D2M.

Table 9. X-ray data collection and refinement statistics for covalent Schiff base complex of YedK C2A and AP-DNA.[a]

Data collection	Covalent Schiff base complex of YedK C2A and AP-DNA
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	60.94 41.47 82.58
α, *β*, *γ* (°)	90.00, 95.65, 90.00
Resolution (Å)	50.00 - 1.83 (1.90 - 1.83)
R_{sym}	0.076 (0.667)
R_{meas}	0.101 (0.891)
Avg. $I/\sigma I$	11.9 (1.4)
Completeness (%)	99.53 (96.06)
Redundancy	2.1 (2.1)
Wilson *B*-factor (Å^2)	19.8
Refinement	
Resolution (Å)	37.03 - 1.82 (1.87 - 1.82)
No. reflections	37,024 (3,583)
R_{work}	0.178 (0.254)
R_{free}[b]	0.226 (0.307)
No. atoms[c]	
Protein	3,500
DNA	246
Water	260
Other	0
Avg. *B*-factors[c,d] (Å^2)	
Protein	26.6
DNA	29.4
Water	27.8
Other	-
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.109

[a] Statistics for the highest resolution shell are shown in parentheses.
[b] R_{free} was determined from the 5% of reflections excluded from refinement.
[c] Riding hydrogen atoms were not included in no. atoms or avg. *B*-factors.
[d] Equivalent isotropic *B*-factors were calculated in conjunction with TLS-derived anisotropic *B*-factors

References

1. Jackson, S.P. and Bartek, J. (2009) The DNA-damage response in human biology and disease. *Nature*, **461**, 1071.
2. Friedberg, E.C. (2005) Suffering in silence: the tolerance of DNA damage. *Nature reviews Molecular cell biology*, **6**, 943.
3. Nakamura, J., Walker, V.E., Upton, P.B., Chiang, S.-Y., Kow, Y.W. and Swenberg, J.A. (1998) Highly sensitive apurinic/apyrimidinic site assay can detect spontaneous and chemically induced depurination under physiological conditions. *Cancer research*, **58**, 222-225.
4. Nakamura, J. and Swenberg, J.A. (1999) Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer research*, **59**, 2522-2526.
5. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *nature*, **362**, 709-715.
6. Barnes, D.E. and Lindahl, T. (2004) Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells. *Annual Review of Genetics*, **38**, 445-476.
7. Cadet, J. and Wagner, J.R. (2013) DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb Perspect Biol*, **5**.
8. De Bont, R. and van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, **19**, 169-185.
9. Campbell, N. and McCulloch, A. (1998) The climate change implications of manufacturing refrigerants: a calculation of 'production'energy contents of some common refrigerants. *Process safety and environmental protection*, **76**, 239-244.
10. Halliwell, B. (1999) Oxygen and nitrogen are pro-carcinogens. Damage to DNA by reactive oxygen, chlorine and nitrogen species: measurement, mechanism and the effects of nutrition. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, **443**, 37-52.
11. Zharkov, D.O. and Rosenquist, T.A. (2002) Inactivation of mammalian 8-oxoguanine-DNA glycosylase by cadmium (II): implications for cadmium genotoxicity. *DNA repair*, **1**, 661-670.
12. McNeill, D.R., Wong, H.K., Narayana, A. and Wilson III, D.M. (2007) Lead promotes abasic site accumulation and co-mutagenesis in mammalian cells by inhibiting the major abasic endonuclease Ape1. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, **46**, 91-99.
13. Lindahl, T. and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry*, **11**, 3610-3618.
14. Krokan, H.E. and Bjørås, M. (2013) Base excision repair. *Cold Spring Harbor perspectives in biology*, **5**, a012583.
15. Mullins, E.A., Rodriguez, A.A., Bradley, N.P. and Eichman, B.F. (2019) Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends in biochemical sciences*.
16. Hitomi, K., Iwai, S. and Tainer, J.A. (2007) The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA repair*, **6**, 410-428.

17. Tsutakawa, S.E., Lafrance-Vanasse, J. and Tainer, J.A. (2014) The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA repair*, **19**, 95-107.
18. Nakamura, J. and Nakamura, M. (2020) DNA-protein crosslink formation by endogenous aldehydes and AP sites. *DNA repair*, 102806.
19. Yang, Z., Price, N.E., Johnson, K.M., Wang, Y. and Gates, K.S. (2017) Interstrand cross-links arising from strand breaks at true abasic sites in duplex DNA. *Nucleic acids research*, **45**, 6275-6283.
20. Lindahl, T. and Andersson, A. (1972) Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry*, **11**, 3618-3623.
21. Overend, W.G. (1950) 533. Deoxy-sugars. Part XIII. Some observations on the Feulgen nucleal reaction. *Journal of the Chemical Society (Resumed)*, 2769-2774.
22. Manoharan, M., Ransom, S.C., Mazumder, A., Gerlt, J.A., Wilde, J.A., Withka, J.A. and Bolton, P.H. (1988) The characterization of abasic sites in DNA heteroduplexes by site specific labeling with carbon-13. *Journal of the American Chemical Society*, **110**, 1620-1622.
23. Lhomme, J., Constant, J.F. and Demeunynck, M. (1999) Abasic DNA structure, reactivity, and recognition. *Biopolymers: Original Research on Biomolecules*, **52**, 65-83.
24. Price, N.E., Johnson, K.M., Wang, J., Fekry, M.I., Wang, Y. and Gates, K.S. (2014) Interstrand DNA–DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *Journal of the American Chemical Society*, **136**, 3483-3490.
25. Haracska, L., Unk, I., Johnson, R.E., Johansson, E., Burgers, P.M., Prakash, S. and Prakash, L. (2001) Roles of yeast DNA polymerases δ and ζ and of Rev1 in the bypass of abasic sites. *Genes & development*, **15**, 945-954.
26. Kavli, B., Otterlei, M., Slupphaug, G. and Krokan, H.E. (2007) Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA repair*, **6**, 505-516.
27. Srivastava, D.K., Berg, B.J.V., Prasad, R., Molina, J.T., Beard, W.A., Tomkinson, A.E. and Wilson, S.H. (1998) Mammalian abasic site base excision repair Identification of the reaction sequence and rate-determining steps. *Journal of Biological Chemistry*, **273**, 21203-21209.
28. Krokan, H.E., Standal, R. and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *Biochemical Journal*, **325**, 1-16.
29. Spiering, A. and Deutsch, W. (1986) Drosophila apurinic/apyrimidinic DNA endonucleases. Characterization of mechanism of action and demonstration of a novel type of enzyme activity. *Journal of Biological Chemistry*, **261**, 3222-3228.
30. Allinson, S.L., Dianova, I.I. and Dianov, G.L. (2001) DNA polymerase β is the major dRP lyase involved in repair of oxidative base lesions in DNA by mammalian cell extracts. *The EMBO journal*, **20**, 6919-6926.
31. Schaaper, R.M., Kunkel, T.A. and Loeb, L.A. (1983) Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proceedings of the National Academy of Sciences*, **80**, 487-491.
32. Andersen, P.L., Xu, F. and Xiao, W. (2008) Eukaryotic DNA damage tolerance and translesion synthesis through covalent modifications of PCNA. *Cell research*, **18**, 162.
33. Powers, K.T. and Washington, M.T. (2018) Eukaryotic translesion synthesis: Choosing the right tool for the job. *DNA repair*, **71**, 127-134.

34. Erzberger, J.P., Barsky, D., Schärer, O.D., Colvin, M.E. and Wilson III, D.M. (1998) Elements in abasic site recognition by the major human and Escherichia coli apurinic/apyrimidinic endonucleases. *Nucleic acids research*, **26**, 2771-2778.
35. Wilson, D.M., Takeshita, M., Grollman, A.P. and Demple, B. (1995) Incision activity of human apurinic endonuclease (Ape) at abasic site analogs in DNA. *Journal of Biological Chemistry*, **270**, 16002-16007.
36. Aravind, L., Anand, S. and Iyer, L.M. (2013) Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biology direct*, **8**, 20.
37. Janion, C. (2008) Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli. *International journal of biological sciences*, **4**, 338.
38. Kweon, S.-M., Zhu, B., Chen, Y., Aravind, L., Xu, S.-Y. and Feldman, D.E. (2017) Erasure of Tet-Oxidized 5-Methylcytosine by a SRAP Nuclease. *Cell reports*, **21**, 482-494.
39. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Münzel, M., Wagner, M., Müller, M. and Khan, F. (2013) Dynamic readers for 5-(hydroxy) methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146-1159.
40. Shukla, V., Halabelian, L., Balagere, S., Samaniego-Castruita, D., Feldman, D.E., Arrowsmith, C.H., Rao, A. and Aravind, L. (2019) HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. *Molecular Cell*, **77**, 384-394.
41. Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P. and Cortez, D. (2019) HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell*, **176**, 144-153. e113.
42. Wu, L., Shukla, V., Yadavalli, A.D., Dinesh, R.K., Xu, D., Rao, A. and Schatz, D.G. (2022) HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes Dev*, **36**, 433-450.
43. Ahn, W.-C., Aroli, S., Kim, J.-H., Moon, J.H., Lee, G.S., Lee, M.-H., Sang, P.B., Oh, B.-H., Varshney, U. and Woo, E.-J. (2019) Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. *Nature chemical biology*, **15**, 607.
44. Thompson, P.S., Amidon, K.M., Mohni, K.N., Cortez, D. and Eichman, B.F. (2019) Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nature Structural & Molecular Biology*, 1.
45. Halabelian, L., Ravichandran, M., Li, Y., Zeng, H., Rao, A., Aravind, L. and Arrowsmith, C.H. (2019) Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nature Structural & Molecular Biology*, 1.
46. Thompson, P.S., Amidon, K.M., Mohni, K.N., Cortez, D. and Eichman, B.F. (2019) Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nature Structural & Molecular Biology*, **26**, 613-618.
47. Wang, N., Bao, H., Chen, L., Liu, Y., Li, Y., Wu, B. and Huang, H. (2019) Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of E. coli yedK. *Nucleic Acids Research*.
48. Krokan, H.E. and Bjoras, M. (2013) Base excision repair. *Cold Spring Harbor perspectives in biology*, **5**, a012583.

49. Hitomi, K., Iwai, S. and Tainer, J.A. (2007) The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst)*, **6**, 410-428.
50. Tsutakawa, S.E., Lafrance-Vanasse, J. and Tainer, J.A. (2014) 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.
51. Kavli, B., Slupphaug, G., Mol, C.D., Arvai, A.S., Peterson, S.B., Tainer, J.A. and Krokan, H.E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *The EMBO journal*, **15**, 3442-3447.
52. Schaaper, R.M., Kunkel, T.A. and Loeb, L.A. (1983) Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proc Natl Acad Sci U S A*, **80**, 487-491.
53. Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P.M. and Cortez, D. (2019) HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell*, **176**, 144-153 e113.
54. Aravind, L., Anand, S. and Iyer, L.M. (2013) Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biology direct*, **8**, 20.
55. Kallen, R.G. (1971) Mechanism of reactions involving Schiff base intermediates. Thiazolidine formation from L-cysteine and formaldehyde. *Journal of the American Chemical Society*, **93**, 6236-6248.
56. Canle, M., Lawley, A., McManus, E.C. and O'Ferrall, R.A.M. (1996), *Pure and Applied Chemistry*, Vol. 68, pp. 813.
57. Just, G., Chung, B.Y., Kim, S., Rosebery, G. and Rossy, P. (1976) Reactions of oxygen and sulphur anions with oxazolidine and thiazolidine derivatives of 2-mesyloxymethylglyceraldehyde acetonide. *Canadian Journal of Chemistry*, **54**, 2089-2093.
58. Ratner, S. and Clarke, H.T. (1937) The Action of Formaldehyde upon Cysteine. *Journal of the American Chemical Society*, **59**, 200-206.
59. Fife, T.H., Natarajan, R., Shen, C.C. and Bembi, R. (1991) Mechanism of thiazolidine hydrolysis. Ring opening and hydrolysis of 1,3-thiazolidine derivatives of p-(dimethylamino)cinnamaldehyde. *Journal of the American Chemical Society*, **113**, 3071-3079.
60. Brooks, S.C., Adhikary, S., Rubinson, E.H. and Eichman, B.F. (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1834**, 247-271.
61. Fromme, J.C. and Verdine, G.L. (2004), *Advances in Protein Chemistry*. Academic Press, Vol. 69, pp. 1-41.
62. Beard, W.A. and Wilson, S.H. (2006) Structure and Mechanism of DNA Polymerase β. *Chemical Reviews*, **106**, 361-382.
63. Billman, J.H. and Diesing, A.C. (1957) Reduction of Schiff bases with sodium borohydride. *The Journal of Organic Chemistry*, **22**, 1068-1070.
64. Kweon, S.M., Zhu, B., Chen, Y., Aravind, L., Xu, S.Y. and Feldman, D.E. (2017) Erasure of Tet-Oxidized 5-Methylcytosine by a SRAP Nuclease. *Cell Rep*, **21**, 482-494.
65. Prasad, R., Horton, J.K., Dai, D.P. and Wilson, S.H. (2019) Repair pathway for PARP-1 DNA-protein crosslinks. *DNA Repair (Amst)*, **73**, 71-77.

66. Prasad, R., Horton, J.K., Chastain, P.D., 2nd, Gassman, N.R., Freudenthal, B.D., Hou, E.W. and Wilson, S.H. (2014) Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. *Nucleic acids research*, **42**, 6337-6351.
67. Sczepanski, J.T., Wong, R.S., McKnight, J.N., Bowman, G.D. and Greenberg, M.M. (2010) Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proc Natl Acad Sci U S A*, **107**, 22475-22480.
68. Grosheva, A.S., Zharkov, D.O., Stahl, J., Gopanenko, A.V., Tupikin, A.E., Kabilov, M.R., Graifer, D.M. and Karpova, G.G. (2017) Recognition but no repair of abasic site in single-stranded DNA by human ribosomal uS3 protein residing within intact 40S subunit. *Nucleic acids research*, **45**, 3833-3843.
69. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146-1159.
70. Boehm, E.M., Gildenberg, M.S. and Washington, M.T. (2016) The Many Roles of PCNA in Eukaryotic DNA Replication. *Enzymes*, **39**, 231-254.
71. Mehta, K.P.M., Lovejoy, C.A., Zhao, R., Heintzman, D.R. and Cortez, D. (2020) HMCES Maintains Replication Fork Progression and Prevents Double-Strand Breaks in Response to APOBEC Deamination and Abasic Site Formation. *Cell Reports*, **31**.
72. Duzen, J.M., Walker, G.C. and Sutton, M.D. (2004) Identification of specific amino acid residues in the E. coli beta processivity clamp involved in interactions with DNA polymerase III, UmuD and UmuD'. *DNA Repair (Amst)*, **3**, 301-312.
73. Tomer, G., Reuven, N.B. and Livneh, Z. (1998) The beta subunit sliding DNA clamp is responsible for unassisted mutagenic translesion replication by DNA polymerase III holoenzyme. *Proc Natl Acad Sci U S A*, **95**, 14106-14111.
74. Covey, J.M., Jaxel, C., Kohn, K.W. and Pommier, Y. (1989) Protein-linked DNA strand breaks induced in mammalian cells by camptothecin, an inhibitor of topoisomerase I. *Cancer Res*, **49**, 5016-5022.
75. Ide, H., Nakano, T., Salem, A.M.H. and Shoulkamy, M.I. (2018) DNA-protein cross-links: Formidable challenges to maintaining genome integrity. *DNA Repair (Amst)*, **71**, 190-197.
76. Quiñones, J.L., Thapar, U., Wilson, S.H., Ramsden, D.A. and Demple, B. (2020) Oxidative DNA-protein Crosslinks Formed in Mammalian Cells by Abasic Site Lyases Involved in DNA Repair. *DNA Repair*, 102773.
77. Zharkov, D.O., Rieger, R.A., Iden, C.R. and Grollman, A.P. (1997) NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg) protein. *Journal of Biological Chemistry*, **272**, 5335-5341.
78. Gilboa, R., Zharkov, D.O., Golan, G., Fernandes, A.S., Gerchman, S.E., Matz, E., Kycia, J.H., Grollman, A.P. and Shoham, G. (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *Journal of Biological Chemistry*, **277**, 19811-19816.
79. Pommier, Y., Pourquier, P., Urasaki, Y., Wu, J. and Laco, G.S. (1999) Topoisomerase I inhibitors: selectivity and cellular resistance. *Drug Resistance Updates*, **2**, 307-318.
80. McCullough, A.K., Sanchez, A., Dodson, M., Marapaka, P., Taylor, J.-S. and Lloyd, R.S. (2001) The reaction mechanism of DNA glycosylase/AP lyases at abasic sites. *Biochemistry*, **40**, 561-568.

81. Cordes, E.H. and Jencks, W.P. (1962) On the Mechanism of Schiff Base Formation and Hydrolysis. *Journal of the American Chemical Society*, **84**, 832-837.
82. Andreeva, A., Kulesha, E., Gough, J. and Murzin, A.G. (2019) The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Research*, **48**, D376-D382.
83. Wang, N., Bao, H., Chen, L., Liu, Y., Li, Y., Wu, B. and Huang, H. (2019) Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of E. coli yedK. *Nucleic Acids Research*, **47**, 10388-10399.
84. Higgins, K.A. and Giedroc, D. (2014) Insights into Protein Allostery in the CsoR/RcnR Family of Transcriptional Repressors. *Chem Lett*, **43**, 20-25.
85. Metz, B., Kersten, G.F., Hoogerhout, P., Brugghe, H.F., Timmermans, H.A., de Jong, A., Meiring, H., ten Hove, J., Hennink, W.E., Crommelin, D.J. *et al.* (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem*, **279**, 6235-6243.
86. Pesek, J. and Frost, J. (1975) Decomposition of thiazolidines in acidic and basic solution: Spectroscopic evidence for Schiff base intermediates. *Tetrahedron*, **31**, 907-913.
87. Schubert, M.P. (1936) COMPOUNDS OF THIOL ACIDS WITH ALDEHYDES. *Journal of Biological Chemistry*, **114**, 341-350.
88. Cook, A.H. and Heilbron, I.M. (1949) In Clarke, H. T., Johnson, J. R. and Robinson, R. (eds.), *Chemistry of Penicillin*. Princeton University Press, pp. 921-972.
89. Gentle, I.E., De Souza, D.P. and Baca, M. (2004) Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjugate chemistry*, **15**, 658-663.
90. Zhang, L. and Tam, J.P. (1996) Thiazolidine formation as a general and site-specific conjugation method for synthetic peptides and proteins. *Analytical biochemistry*, **233**, 87-93.
91. Bondi, A.v. (1964) van der Waals volumes and radii. *The Journal of physical chemistry*, **68**, 441-451.
92. Eramian, H., Tian, Y.-H., Fox, Z., Beneberu, H.Z. and Kertesz, M. (2013) On the anisotropy of van der Waals atomic radii of O, S, Se, F, Cl, Br, and I. *The Journal of Physical Chemistry A*, **117**, 14184-14190.
93. Piersen, C.E., Prasad, R., Wilson, S.H. and Lloyd, R.S. (1996) Evidence for an imino intermediate in the DNA polymerase β deoxyribose phosphate excision reaction. *Journal of Biological Chemistry*, **271**, 17811-17815.
94. Bailly, V., Verly, W., O'Connor, T. and Laval, J. (1989) Mechanism of DNA strand nicking at apurinic/apyrimidinic sites by Escherichia coli [formamidopyrimidine] DNA glycosylase. *Biochemical Journal*, **262**, 581-589.
95. Sun, B., Latham, K.A., Dodson, M. and Lloyd, R.S. (1995) Studies on the Catalytic Mechanism of Five DNA Glycosylases PROBING FOR ENZYME-DNA IMINO INTERMEDIATES. *Journal of Biological Chemistry*, **270**, 19501-19508.
96. Golan, G., Zharkov, D.O., Grollman, A.P., Dodson, M., McCullough, A.K., Lloyd, R.S. and Shoham, G. (2006) Structure of T4 pyrimidine dimer glycosylase in a reduced imine covalent complex with abasic site-containing DNA. *Journal of molecular biology*, **362**, 241-258.
97. Purmal, A.A., Rabow, L.E., Lampman, G.W., Cunningham, R.P. and Kow, Y.W. (1996) A common mechanism of action for the N-glycosylase activity of DNA N-glycosylase/AP lyases from E. coli and T4. *Mutat Res*, **364**, 193-207.

98. Roberts, S.A., Strande, N., Burkhalter, M.D., Strom, C., Havener, J.M., Hasty, P. and Ramsden, D.A. (2010) Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature*, **464**, 1214-1217.
99. Strande, N., Roberts, S.A., Oh, S., Hendrickson, E.A. and Ramsden, D.A. (2012) Specificity of the dRP/AP lyase of Ku promotes nonhomologous end joining (NHEJ) fidelity at damaged ends. *Journal of Biological Chemistry*, **287**, 13686-13693.
100. Kutuzov, M.M., Khodyreva, S.N., Ilina, E.S., Sukhanova, M.V., Amé, J.C. and Lavrik, O.I. (2015) Interaction of PARP-2 with AP site containing DNA. *Biochimie*, **112**, 10-19.
101. Khodyreva, S., Prasad, R., Ilina, E., Sukhanova, M., Kutuzov, M., Liu, Y., Hou, E., Wilson, S. and Lavrik, O. (2010) Apurinic/apyrimidinic (AP) site recognition by the 5′-dRP/AP lyase in poly (ADP-ribose) polymerase-1 (PARP-1). *Proceedings of the National Academy of Sciences*, **107**, 22090-22095.
102. Prakash, A., Eckenroth, B.E., Averill, A.M., Imamura, K., Wallace, S.S. and Doublié, S. (2013) Structural investigation of a viral ortholog of human NEIL2/3 DNA glycosylases. *DNA Repair (Amst)*, **12**, 1062-1071.
103. Quiñones, J.L., Thapar, U., Yu, K., Fang, Q., Sobol, R.W. and Demple, B. (2015) Enzyme mechanism-based, oxidative DNA–protein cross-links formed with DNA polymerase β in vivo. *Proceedings of the National Academy of Sciences*, **112**, 8602-8607.
104. DeMott, M.S., Beyret, E., Wong, D., Bales, B.C., Hwang, J.-T., Greenberg, M.M. and Demple, B. (2002) Covalent trapping of human DNA polymerase β by the oxidative DNA lesion 2-deoxyribonolactone. *Journal of Biological Chemistry*, **277**, 7637-7640.
105. Prasad, R., Horton, J.K. and Wilson, S.H. (2020) Requirements for PARP-1 covalent crosslinking to DNA (PARP-1 DPC). *DNA Repair (Amst)*, **89**, 102824.
106. Srivastava, M., Su, D., Zhang, H., Chen, Z., Tang, M., Nie, L. and Chen, J. (2020) HMCES safeguards replication from oxidative stress and ensures error-free repair. *EMBO Rep*, e49123.
107. Viktorovskaya, O.V., Greco, T.M., Cristea, I.M. and Thompson, S.R. (2016) Identification of RNA Binding Proteins Associated with Dengue Virus RNA in Infected Cells Reveals Temporally Distinct Host Factor Requirements. *PLoS Negl Trop Dis*, **10**, e0004921.
108. Horikoshi, M., Day, F.R., Akiyama, M., Hirata, M., Kamatani, Y., Matsuda, K., Ishigaki, K., Kanai, M., Wright, H. and Toro, C.A. (2018) Elucidating the genetic architecture of reproductive ageing in the Japanese population. *Nature communications*, **9**, 1977.
109. Taylor, M.S., LaCava, J., Mita, P., Molloy, K.R., Huang, C.R.L., Li, D., Adney, E.M., Jiang, H., Burns, K.H. and Chait, B.T. (2013) Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell*, **155**, 1034-1048.
110. Taylor, M.S., Altukhov, I., Molloy, K.R., Mita, P., Jiang, H., Adney, E.M., Wudzinska, A., Badri, S., Ischenko, D. and Eng, G. (2018) Dissection of affinity captured LINE-1 macromolecular complexes. *Elife*, **7**, e30094.
111. Miyoshi, T., Makino, T. and Moran, J.V. (2019) Poly (ADP-Ribose) Polymerase 2 Recruits Replication Protein A to Sites of LINE-1 Integration to Facilitate Retrotransposition. *Molecular cell*, **75**, 1286-1298. e1212.
112. Wang, Y., Zeng, L., Liang, C., Zan, R., Ji, W., Zhang, Z., Wei, Y., Tu, S. and Dong, Y. (2019) Integrated analysis of transcriptome-wide m6A methylome of osteosarcoma stem cells enriched by chemotherapy. *Epigenomics*, **11**, 1693-1715.

113. Gao, H.-X., Nuerlan, A., Abulajiang, G., Cui, W.-L., Xue, J., Sang, W., Li, S.-J., Niu, J., Ma, Z.-P. and Zhang, W. (2019) Quantitative proteomics analysis of differentially expressed proteins in activated B-cell-like diffuse large B-cell lymphoma using quantitative proteomics. *Pathology-Research and Practice*, **215**, 152528.
114. Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A. and Edlund, K. (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics*, **13**, 397-406.
115. Pfeifer, G.P., Szabó, P.E. and Song, J. (2019) Protein Interactions at Oxidized 5-Methylcytosine Bases. *Journal of molecular biology*.
116. Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. and Zhang, B. (2004) PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, **4**, 1551-1561.
117. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G. and McAdam, R. (2019) The BioGRID interaction database: 2019 update. *Nucleic acids research*, **47**, D529-D541.
118. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M. and Network, C.G.A.R. (2013) The cancer genome atlas pan-cancer analysis project. *Nature genetics*, **45**, 1113.
119. Wilde, J.A., Bolton, P.H., Mazumder, A., Manoharan, M. and Gerlt, J.A. (1989) Characterization of the equilibrating forms of the aldehydic abasic site in duplex DNA by oxygen-17 NMR. *Journal of the American Chemical Society*, **111**, 1894-1896.
120. Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J. and Harris, R.S. (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol*, **17**, 222-229.
121. Semlow, D.R., MacKrell, V.A. and Walter, J.C. (2022) The HMCES DNA-protein cross-link functions as an intermediate in DNA interstrand cross-link repair. *Nat Struct Mol Biol*.
122. Wu, R.A., Semlow, D.R., Kamimae-Lanning, A.N., Kochenova, O.V., Chistol, G., Hodskinson, M.R., Amunugama, R., Sparks, J.L., Wang, M., Deng, L. *et al.* (2019) TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature*, **567**, 267-272.
123. Amidon, K.M. and Eichman, B.F. (2020) Structural biology of DNA abasic site protection by SRAP proteins. *DNA Repair*, **94**, 102903.
124. Pan, Y., Zuo, H., Wen, F., Huang, F., Zhu, Y., Cao, L., Sha, Q.Q., Li, Y., Zhang, H., Shi, M. *et al.* (2022) HMCES safeguards genome integrity and long-term self-renewal of hematopoietic stem cells during stress responses. *Leukemia*.
125. Kojima, N., Takebayashi, T., Mikami, A., Ohtsuka, E. and Komatsu, Y. (2009) Construction of highly reactive probes for abasic site detection by introduction of an aromatic and a guanidine residue into an aminooxy group. *J Am Chem Soc*, **131**, 13208-13209.
126. Kamps, J.J.A.G., Hopkinson, R.J., Schofield, C.J. and Claridge, T.D.W. (2019) How formaldehyde reacts with amino acids. *Communications Chemistry*, **2**, 126.
127. Melamede, R.J., Hatahet, Z., Kow, Y.W., Ide, H. and Wallace, S.S. (1994) Isolation and characterization of endonuclease VIII from Escherichia coli. *Biochemistry*, **33**, 1255-1264.

128. Bailly, V. and Verly, W.G. (1987) Escherichia coli endonuclease III is not an endonuclease but a beta-elimination catalyst. *Biochem J*, **242**, 565-572.
129. Saiz, C., Wipf, P., Manta, E. and Mahler, G. (2009) Reversible thiazolidine exchange: a new reaction suitable for dynamic combinatorial chemistry. *Org Lett*, **11**, 3170-3173.
130. Esterbauer, H., Ertl, A. and Scholz, N. (1976) The reaction of cysteine with α,β-unsaturated aldehydes. *Tetrahedron*, **32**, 285-289.
131. Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N. and Oda, M. (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem*, **79**, 89-130.
132. Kong, X.-P., Onrust, R., O'Donnell, M. and Kuriyan, J. (1992) Three-dimensional structure of the β subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, **69**, 425-437.
133. Patoli, A.A., Winter, J.A. and Bunting, K.A. (2013) The UmuC subunit of the E. coli DNA polymerase V shows a unique interaction with the β-clamp processivity factor. *BMC structural biology*, **13**, 12.
134. Hendel, A., Krijger, P.H., Diamant, N., Goren, Z., Langerak, P., Kim, J., Reißner, T., Lee, K.-y., Geacintov, N.E. and Carell, T. (2011) PCNA ubiquitination is important, but not essential for translesion DNA synthesis in mammalian cells. *PLoS genetics*, **7**, e1002262.
135. Marians, K.J. (2018) Lesion bypass and the reactivation of stalled replication forks. *Annual review of biochemistry*, **87**, 217-238.
136. Li, X., Stith, C.M., Burgers, P.M. and Heyer, W.-D. (2009) PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase δ. *Molecular cell*, **36**, 704-713.
137. Motegi, A., Liaw, H.-J., Lee, K.-Y., Roest, H.P., Maas, A., Wu, X., Moinova, H., Markowitz, S.D., Ding, H. and Hoeijmakers, J.H. (2008) Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. *Proceedings of the National Academy of Sciences*, **105**, 12411-12416.
138. Motegi, A., Sood, R., Moinova, H., Markowitz, S.D., Liu, P.P. and Myung, K. (2006) Human SHPRH suppresses genomic instability through proliferating cell nuclear antigen polyubiquitination. *J Cell Biol*, **175**, 703-708.
139. Unk, I., Hajdú, I., Fátyol, K., Szakál, B., Blastyák, A., Bermudez, V., Hurwitz, J., Prakash, L., Prakash, S. and Haracska, L. (2006) Human SHPRH is a ubiquitin ligase for Mms2–Ubc13-dependent polyubiquitylation of proliferating cell nuclear antigen. *Proceedings of the National Academy of Sciences*, **103**, 18107-18112.
140. Sun, L. and Chen, Z.J. (2004) The novel functions of ubiquitination in signaling. *Current opinion in cell biology*, **16**, 119-126.
141. Rosenbaum, J.C., Bonilla, B., Hengel, S.R., Mertz, T.M., Herken, B.W., Kazemier, H.G., Pressimone, C.A., Ratterman, T.C., MacNary, E. and De Magis, A. (2019) The Rad51 paralogs facilitate a novel DNA strand specific damage tolerance pathway. *Nature communications*, **10**, 1-11.
142. Scaturro, P., Kastner, A.L. and Pichlmair, A. (2019) Chasing Intracellular Zika Virus Using Proteomics. *Viruses*, **11**, 878.
143. Bidet, K. and Garcia-Blanco, M.A. (2018), *Dengue and Zika: Control and Antiviral Treatment Strategies*. Springer, pp. 45-62.

144. Fagerberg, L., Hallstrom, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K. *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*, **13**, 397-406.
145. Wang, Y., Zeng, L., Liang, C., Zan, R., Ji, W., Zhang, Z., Wei, Y., Tu, S. and Dong, Y. (2019) Integrated analysis of transcriptome-wide m6A methylome of osteosarcoma stem cells enriched by chemotherapy. *Epigenomics*.
146. Wei, X., Wang, Z., Hinson, C. and Yang, K. (2022) Human TDP1, APE1 and TREX1 repair 3'-DNA-peptide/protein cross-links arising from abasic sites in vitro. *Nucleic Acids Res*, **50**, 3638-3657.
147. Quiñones, J.L. and Demple, B. (2016) When DNA repair goes wrong: BER-generated DNA-protein crosslinks to oxidative lesions. *DNA repair*, **44**, 103-109.
148. Kieffer, S.R. and Lowndes, N.F. (2022) Immediate-Early, Early, and Late Responses to DNA Double Stranded Breaks. *Front Genet*, **13**, 793884.
149. Shukla, V., Halabelian, L., Balagere, S., Samaniego-Castruita, D., Feldman, D.E., Arrowsmith, C.H., Rao, A. and Aravind, L. (2019) HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. *Molecular Cell*.
150. Adar, S., Izhar, L., Hendel, A., Geacintov, N. and Livneh, Z. (2009) Repair of gaps opposite lesions by homologous recombination in mammalian cells. *Nucleic Acids Res*, **37**, 5737-5748.
151. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*, **246**, 64-71.
152. Krokan, H.E., Saetrom, P., Aas, P.A., Pettersen, H.S., Kavli, B. and Slupphaug, G. (2014) Error-free versus mutagenic processing of genomic uracil--relevance to cancer. *DNA Repair (Amst)*, **19**, 38-47.
153. Haldar, T., Jha, J.S., Yang, Z., Nel, C., Housh, K., Cassidy, O.J. and Gates, K.S. (2022) Unexpected Complexity in the Products Arising from NaOH-, Heat-, Amine-, and Glycosylase-Induced Strand Cleavage at an Abasic Site in DNA. *Chemical Research in Toxicology*.
154. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307-326.
155. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J. and Grosse-Kunstleve, R.W. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213-221.
156. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126-2132.
157. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M. and Sliz, P. (2013) Collaboration gets the most out of software. *eLife*, **2**, e01456.
158. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, **44**, W344-350.