ANCIENT SRAP DOMAIN SHIELDS ABASIC SITES TO PROMOTE GENOME INTEGRITY

By

Petria Salewa Thompson

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

Biochemistry

May 31st, 2020

Nashville, Tennessee

Approved:

F. Peter Guengerich, Ph.D.

David Cortez, Ph.D.

Brandt Eichman, Ph.D.

Kathleen L. Gould, Ph.D.

Scott W. Hiebert, Ph.D.

To my parents, Peter and Georgia

ACKNOWLEDGEMENTS

It's difficult to articulate what the gravity of earning a PhD in Biochemistry means to me. My grandparents only received an elementary school education. My mother, Georgia, grew up in a two-room house in a rural village. With her three siblings, she supported her family by helping my Grandma, Becca, sell food at a local market. My father, Peter, grew up in Kingston, Jamaica, in a neighborhood plagued by gang violence and political unrest. His mother worked as a nanny to raise her eight children. When they were growing up, education was a gatekeeper to opportunity, and the chance to learn was limited by finances and circumstances. My parents, the first teachers in my life, instilled in me the value of not just an education, but of learning. They taught me to continuously approach life with curiosity and to remain humble, to work diligently and pursue opportunities to learn more. My parents' sacrifices, perseverance, relentless drive to help their daughters thrive is at the root of my success. Because of them, I am in the first in my family to complete a doctoral degree. This is a win not only for me, but also for us.

So much of my success is owed to the mentors in my life. To David Cortez, my thesis advisor, I am incredibly grateful for the training environment you provided during graduate school. Dave's enthusiasm for science and love for discovery is genuinely infectious. Thank you for teaching me how to think critically, to challenge existing ideas, and how to write stellar papers and grants. You are an exceptional mentor. Thank you to my thesis dissertation committee—Drs. F. Peter Guengerich, Brandt Eichman, Kathy Gould, and Scott Hiebert-—the hours of discussion, criticism, and encouragement was

truly a pivotal experience in my career as a scientist. I am grateful for the time and energy you spent increasing my scientific rigor. To my scientific mentors across my career, Drs. Dinah S. Singer, Edward T. Morgan, and Chanelle Case-Borden—thank you for investing your time and energy in training me to become a better scientist. And more importantly, thank you for believing that I could (and can) achieve anything. All those seeds of opportunity and hope have now sprouted with the completion of this part of my training.

For me, learning is the journey to uncover the unknown and further understand the complexities of the world around me. I joined the Vanderbilt MSTP program with the goal to become an excellent physician-scientist and cancer researcher. What I didn't realize is how much I would value the educational programming. I know that because of this program, I have become an excellent teacher, mentor, leader, and advocate and quite frankly, a better human. To the leadership team, thank you for supporting every single one of my varied interests. To my MSTPals, thank you for being excellent soundboards, for engaging in deep and candid conversations, and for every much-needed breaks (including rock climbing trips, gym days, dinner and dance parties, Monday night blues) and just everything in between.

To all the members of the Cortez lab, I could not have been successful without your insight and support. Thank you for challenging me to think critically, asking me thoughtful questions during lab meetings, and offering timely experimental suggestions. To Jorge and Archana, I appreciate your friendship and encouragement. To my Biochemistry graduate school friends: Sarah Arcos, Katie Rothamel, Samantha Lisy, and Monica Bomber—thank you. To be surrounded by such smart, fierce, driven female

scientists has been a highlight of my graduate school experience. I look forward to continuing our friendship in the years to come. This work would not have been possible without the help and support of Katherine Amidon. Her expertise in X-ray crystallography was critical, and I am so glad that she is not only a great scientific collaborator but an amazing friend.

I began writing my thesis at the beginning of the COVID-19 pandemic. My life, our lives, have fundamentally changed. While there have been some personal losses and professional set-backs, there has also been moments of joy. To Allison Arcos, thank you for your tremendous support during this time. Your company, encouragement, and ability to enjoy life even in the midst of a crisis is inspirational. I am so grateful to have you in my life. To my sisters, Vanecia and Timberley, thank you for late night calls, funny texts, and the constant stream of funny memes and videos. It is a true joy to have two wonderful, smart, talented, and driven black women in my corner at all times.

This work could not have been done without generous support from the National Institute of Health and National Cancer Institute.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

4NQO	4-nitroquinoline 1-oxide
5' dRP	5’ terminal deoxyribose phosphate
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
8-oxo-G	7,8-dihydro-8-oxoguanine
ACRC/GCNA-1	acid repeat containing/germ cell nuclear antigen
AID	activation-induced deaminase
Alt-EJ	microhomology-mediated alternative-end joining
AP	Abasic/Apurinic/apyrimidinic
APE1	AP endonuclease 1
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ATR	ataxia-telangiectasia and Rad3-related
BER	Base Excision Repair
CDG	cytosine DNA glycosylase
CPD	cyclobutane pyrimidine dimers
CSR	class switch recombination
dL	2-deoxyribonolactone
DLBCL	diffuse large B-cell lymphoma
DPC	DNA-protein crosslink
DSB	double strand breaks
dsDNA	double-stranded DNA
DTT	dithiothreitol
EMSA	electrophoretic mobility shift analysis
FapyG	2,6-diamino-4-hydroxy-5-formamidopyrimidine
FEN1	flap endonuclease 1
GG-NER	global genome nucleotide excision repair
HMCES	5-hydroxymethylcytosine embryonic stem cells specific
HR	homologous recombination
ICL	intra- or inter-strand crosslinks
Ig	immunoglobulin
IR	ionizing radiation
MBD4	methyl-CpG-binding domain protein 4
MCM	minichromosome maintenance protein complex
MMS	methyl methanesulfonate

MPG	N-methyl-purine DNA glycosylase
NEIL	endonuclease VIII-like
NER	Nucleotide Excision Repair
NHEJ	non-homologous end joining
NTHL1	endonuclease III-like
OGG1	8-oxoguanine glycosylase
PARP	poly(ADP-ribose) polymerase 1
PIP	PCNA interacting motif
RECQL	RecQ like helicase
RNAPII	RNA polymerase II
SHM	somatic hypermutation
Shu2	SWIM domain containing protein
SMARCAL1	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin
SPRTN	Spartan, C1orf124
SRAP	SOS response peptidase
SSB	single-stranded DNA binding protein
ssDNA	single-stranded DNA
TC-NER	transcription-coupled nucleotide excision repair
TDG	thymine DNA glycosylase
THF	tetrahydrofuran
TLS	Translesion synthesis
TRAIP	TRAF interacting protein
UNG	uracil DNA glycosylase
UV	ultraviolet
XRCC1	X-ray repair cross-complementing 1
ZRANB3	Zinc Finger RANBP2-Type Containing 3

CHAPTER I

Introduction

DNA is under constant threat by endogenous and environmental DNA damaging agents. Apurinic/apyrimidinic (AP or abasic) are the most frequent DNA lesions. AP sites result from cleavage of the N-glycosylic bond between the nitrogenous base and the deoxyribose sugar, leaving an intact phosphodiester backbone. Base loss can happen because of spontaneous hydrolysis, base damage leading to destabilization of N-glycosyl bond, or the action of specialized DNA glycosylases. There are numerous evolutionarily conserved repair pathways for AP sites reflecting their physiological importance and mutagenic potential. AP site lesions that persist are roadblocks to transcription and DNA replication. Furthermore, these non-coding lesions are cytotoxic and a threat to genomic integrity since they can cause strand-breaks, interstrand DNA crosslinks, mutations, and DNA-protein crosslinks. Therefore, understanding mechanisms of formation and repair of AP sites are critically important. In this chapter, I will review how AP sites are formed, how they are processed by different repair and tolerance mechanisms depending on context, and their consequences for genome stability.

AP Site Formation

AP sites are the most frequent lesion in cells with estimates of ~10,000-20,000 per human cell per day (1). They form spontaneously, as a consequence of destabilization of

the N-glycosyl bond by other types of DNA damage, and enzymatically by glycosylases (Figure 1.1A).

Spontaneous Base Loss

Although DNA is a relatively stable biomolecule, the N-glycosyl bond is prone to hydrolysis. Spontaneous depurination events occur at a rate of ~3 x 10^{-11} nucleotides per second *in vitro* under physiological conditions in duplex DNA (2, 3). In *E. coli*, 0.5 depurinations occur per cell per generation, and in mammalian cells, approximately 10,000 purines are lost from the genome per day. Guanines are 1.5-times more likely to undergo depurination compared to adenines. Although the mechanism of base loss is similar for purines and pyrimidines, the rate of depyrimidination is 1/20^{th} the rate for depurination since the N-glycosyl bond of pyrimidines are more stable (4). Furthermore, abasic sites preferentially form at sites of DNA replication (5). The stretches of single-stranded DNA (ssDNA) on the lagging strand are more vulnerable to chemical attack and spontaneous bass loss. In fact, the rate of depurination is accelerated more than four times in ssDNA versus double-stranded DNA (dsDNA) (2).

Base Damage and Destabilization of N-Glycosylic Bond

Several environmental and cancer therapeutic genotoxins including alkylating agents, oxidizing agents, ionizing radiation, and ultraviolet radiation can cause nucleobase loss and the formation of AP sites. Base modification by alkylation of nucleobases weakens the N-glycosyl bond. More specifically, base modification generates an unstable positive charge on the ring base. This protonation on the ring is

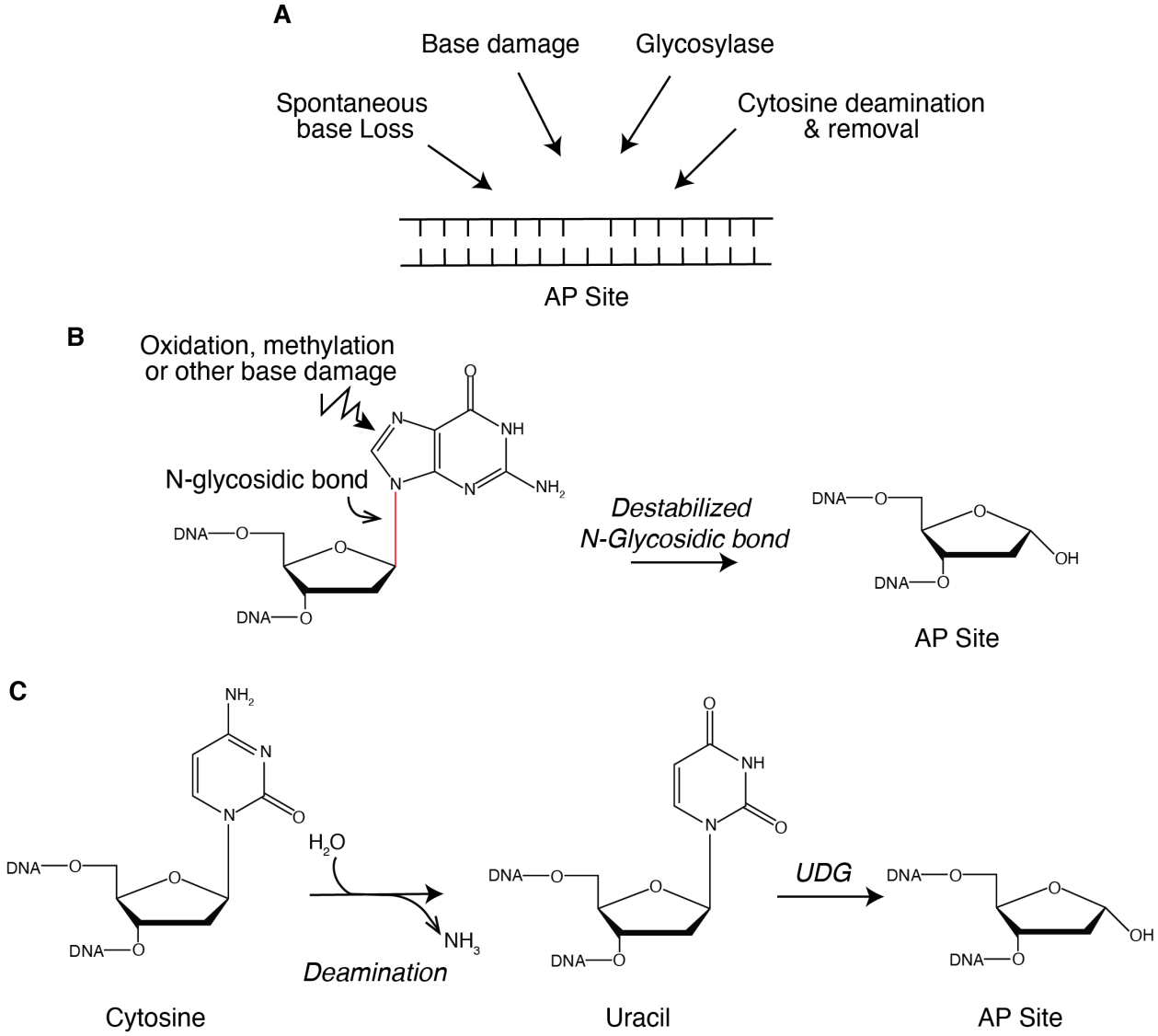

Figure 1.1: Mechanisms of AP site formation. (A) Potential sources of AP site formation. (B) Destabilization of N-glycosidic bond after base damage generates AP site. (C) Spontaneous or catalyzed deamination of cytosine generates uracil. Subsequent action of UDG produces an AP site.

stabilized via electron resonance leading to N-glycosyl bond cleavage (6). Nitrogenous bases are predominately protonated on N^3 and N^7 of purines and O^2 of pyrimidines (7, 8). The most common alkylating agents used in anticancer therapies including methanesulfonate esters, nitrosoureas, and nitrogen mustards generate AP sites (9).

Oxidative damage is a significant burden to cells. There are more than 100 types of oxidized base modification present in mammalian DNA (10). Reactive oxygen species such as superoxide, hydrogen peroxide, and hydroxyl radicals are by-products of cellular metabolism that can modify all four DNA bases (11). Because of the low redox potential of guanine residues, 7,8-dihydro-8-oxoguanine (8-oxoG) is the most common oxidized lesion and is present in large quantities (12, 13). In addition, the formamidopyrimidine lesions 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and 4,6-diamino-5-hydroxy-5-formamidopyrimidine are other highly mutagenic oxidative lesions caused by fragmentation of the purine imidazole ring. Similar oxidative lesions are seen on adenine. Finally, free radicals tend to attack thymine and cytosines residues at the 5,6-double bond leading to various oxidized base products (reviewed in (11, 12). These oxidized base derivates cause AP sites directly through destabilization of the N-glycosylic bond (Figure 1.1B) or are recognized by DNA glycosylases leading to AP site formation.

Ionization radiation (IR) generates a broad spectrum of DNA damage products with double strand breaks the predominant lesion causing cellular lethality. However, IR-induced base damage can also generate AP sites. IR produces hydroxyl radicals that directly attack the N-glycosyl bond or react with the 5,6-double bond in nucleobases saturating the ring and decreasing N-glycosyl bond stability (14, 15). Other ring-saturated

derivates of IR, such as thymine glycol, are recognized by DNA glycosylases that then generate AP sites (16, 17).

Cyclobutane pyrimidine dimers (CPD) are the predominant lesion formed by ultraviolet (UV) radiation. However, other UV induced photoproducts can indirectly produce abasic site lesions (18). The addition of water across the 5,6-double bond of cytosine generates cytosine hydrate. Cytosine hydrate undergoes deamination to uracil hydrate which is subsequently dehydrated to uracil (19). This UV-induced product forms more readily in ssDNA than duplex DNA, and subsequent action of uracil DNA glycosylase generate AP sites (1). Other notable UV induced photoproducts that can lead to abasic sites include thymine glycol, alkali labile purine lesions, and other pyrimidine hydrates (13, 20).

Action of Specialized DNA Glycosylases

DNA glycosylases produce abasic sites as an intermediate step during the repair of damaged DNA by the base excision repair pathway (discussed below). These specialized enzymes are generally well-conserved through all domains of life. DNA glycosylases catalyze the hydrolysis of the N-glycosidic bond between DNA base and sugar-phosphate, releasing the base. DNA glycosylases can be classified as monofunctional or bifunctional. Monofunctional glycosylases remove the base leaving only the intact AP site whereas bifunctional glycosylases have lyase activity that cleaves DNA 3’ of the AP site to generate a 3’ unsaturated aldehydic and a 5’-phosphorylated ends (13).

Some DNA glycosylases are specific for base-pair mismatches. Thymine DNA glycosylase (TDG) and methyl-CpG-binding domain protein 4 (MBD4) recognize G: T mismatch and G:U or G:T mismatch bases, respectively. Other DNA glycosylases are specialized for a variety of damaged bases, including 8-oxoguanine glycosylase (OGG1) that recognizes and removes 8-oxoG and FapyG lesions and N-methyl-purine DNA glycosylase (MPG) that repairs 3meA, 7meG, and 3meG lesions. Most single glycosylase knockouts are not lethal, suggesting overlapping substrate recognition amongst DNA glycosylases (13). Most bifunctional glycosylases such as NEIL1, NEIL2, and NEIL3 (endonuclease VIII-like) and NTHL1 (endonuclease III-like) are specialized for oxidized pyrimidines and ring-opened purines (13).

Uracil DNA glycosylase is perhaps the most conserved among yeast, bacteria, and mammalian cells. All mammalian glycosylases that remove uracil are monofunctional. Mammalian uracil DNA glycosylase (UNG) travels with the replisome generating abasic sites. Uracils are incorporated into DNA via numerous mechanisms. DNA polymerases can incorporate dUMP instead of dTMP during replication (21). While polymerases are highly selective, the genome size and much higher concentration of rNTPs than dNTPs in cells mean that more than 13,000 ribonucleotides are incorporated during each round of DNA replication in budding yeast and orders of magnitude larger numbers in the human genome (22). Depletion of cellular dNTP pools, which can happen in response to drug treatments that interfere with ribonucleotide reductase and replication timing dysregulation caused by oncogenes, likely increases uracil incorporation (23).

Another prominent source of uracil incorporation in DNA is cytosine deamination (Figure 1.1C). Spontaneous deamination of cytosine introduces ~100-500 uracils per day

in human cells (24). Cytosines are more prone to heat-induced degradation compared to other DNA bases. This deamination occurs spontaneously at neutral pH. The rate of cytosine deamination is higher in ssDNA than dsDNA (2, 18, 25) , and approximately 1-2% of DNA in proliferating mammalian cells is in the single-stranded form at any given time, especially during the processes of replication and transcription (18). Subsequent uracil removal by UNG generates significant numbers of AP sites. Furthermore, the additive action of enzymes, such as cytosine deaminases, can exacerbate the number of abasic sites in the genome. For example, the AID/APOBEC (activation-induced deaminase/apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of DNA cytosine deaminases specifically act on ssDNA (26). While APOBEC enzymes are usually retained in the cytoplasm and act to restrict invading nucleic acids from pathogens, cancer cells often exhibit elevated APOBEC expression and nuclear localization. Thus, cytosine deamination by APOBEC enzymes generates two mutational signatures in cancer. One of the signatures is the result of misincorporations across from abasic sites (27, 28). This mutagenesis is biased towards the lagging strand since that is where the majority of ssDNA is present to be targeted by the APOBEC enzymes during replication (29, 30). Thus, generation of uracils in DNA is a notable source of genome instability that drives tumorigenesis in part through production of abasic sites.

Consequences of AP Sites for Genome Stability

AP sites have multiple deleterious consequences (Figure 1.2). They interfere with the natural information content function of the DNA and are obstacles to RNA

polymerases. They are reactive intermediates in the generation of interstrand and DNA protein crosslinks. When encountered during DNA replication, they can generate mutations. Finally, they are potent blocks to DNA polymerases, which can threaten the completion of DNA replication.

AP Sites are Intermediates in the Formation of Interstrand and DNA-Protein Crosslinks

Abasic sites exist as a sugar anomer in an equilibrating mixture of a closed-ring furanose (99%) and an open-ring aldehyde (1%) (31). The open-ring aldehyde form is highly reactive (32). AP sites can be converted into strand-breaks via a β-elimination reaction where the 3' phosphodiester bond of the aldehyde form is hydrolyzed to generate a 3'-terminal unsaturated sugar and a terminal 5'-phosphate. The presence of nucleophilic molecules, including thiols, amines, polyamines, and basic proteins, further stimulates this reaction (reviewed in (33)). Furthermore, AP sites can generate DNA intra- and inter-strand crosslinks (ICLs) (34, 35) . The AP aldehyde can react with the exocyclic amino group on nucleobases, especially guanine (34, 36). Interestingly, these abasic site ICLs are cleaved by the DNA glycosylase NEIL3. This NEIL3 mediated "unhooking" prevents DSB formation and curiously generates another AP site intermediate (37).

AP sites are also prone to forming crosslinks with multiple proteins. Several repair enzymes, including PARP1/2 (38–40), Pol β (41, 42), and KU (43, 44) can form DPCs with AP site lesions or BER intermediates (reviewed here (45, 46)) (Figure 1.2). However, these DPCs are often thought to be suicidal, deleterious, or transient intermediates. PARP1/2, Pol β, and KU all form transient Schiff-base intermediates that can be trapped

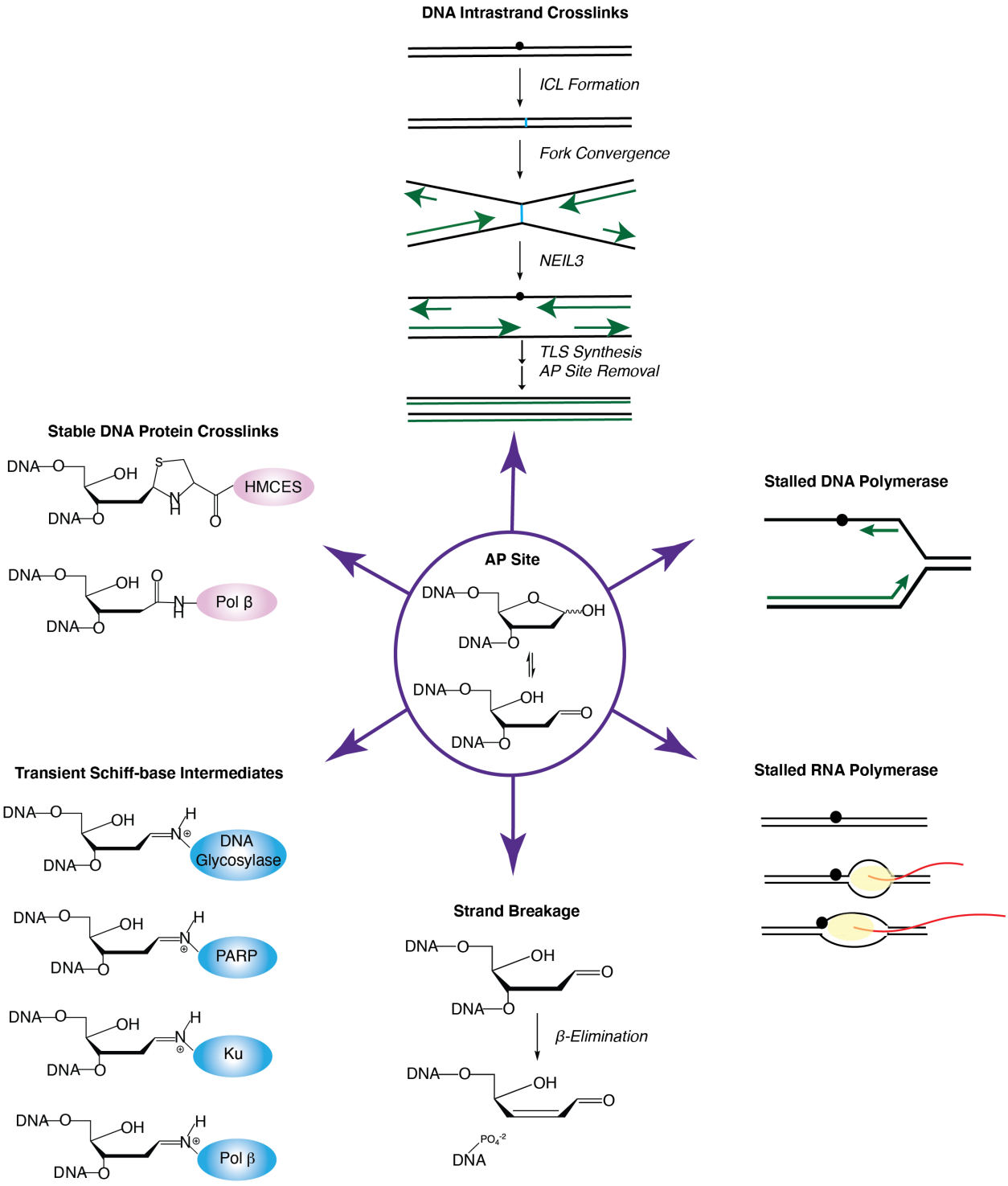

Figure 1.2: Consequences of unrepaired AP sites. AP sites (in center) react to form stable (pink) or transient (blue) DNA-protein crosslinks (DPCs) (left), generate intrastrand crosslinks (ICLs) (top), stall DNA polymerases and RNA Polymerases (right), and cause strand breakage via β-elimination (bottom).

in the presence of $NaBH_4$. Most Schiff-base intermediates are unstable and are quickly resolved by β-elimination to release the enzyme. Interestingly, Pol β also forms a dead-end suicide DPC with 2-deoxyribonolactone (dL), an oxidized AP site derivative (41). This Pol β DPC occurs in the absence of $NaBH_4$, forms in vivo, and is repaired in a proteasome-dependent mechanism (42). Structures of other DNA enzymes that form DPCs such as topoisomerase and DNA methyltransferases are in the literature. However, these DPCs are often covalently trapped using mechanism-based compounds. Furthermore, at least in the case of topoisomerase cleavage complexes, their repair is mediated by a specific set of enzymes (ie-TDP1/TDP2) (47),(48).

AP Sites are a Strong Block to Replicative Polymerases

AP sites preferentially form at sites of DNA replication due to the vulnerability of ssDNA to base damage (5). Although AP sites can retain B-DNA structure (49), they are a potent block to replicative polymerases (Pol α, δ, and ε) (50–52). AP sites are highly mutagenic, and their mutational capacity varies depending on the organism and system used(53). In *Escherichia coli (E. coli)* cells, replication past the AP site is characterized by the preferential incorporation of adenosine (A) opposite AP sites (54). This “A-rule” predicts that depurinations should produce transversions, and depyrimindations should produce transitions. In transfection experiments in which single-stranded bacteriophage DNA was introduced into *E. coli* and recovered from cells, this was the exact mutational pattern observed (50, 55). Furthermore, this rule is observed in systems using purified DNA polymerases and constructs containing abasic sites (52, 56, 57). In *Saccharomyces cerevisiae*, using a dsDNA plasmid system, dAMP was preferentially inserted opposite

AP sites (58). However, other nucleotides have been reported to be preferentially inserted opposite abasic sites in other studies. Gibbs and colleagues observed a dCMP preference in a gapped duplex shuttle vector with a single abasic site located within a single-stranded region (59). Similarly, Otsuka and colleagues using a short oligonucleotide transformation observed a preference for dCMP insertion opposite abasic sites (60). In another study, using spontaneous mutagenesis of the SUP4-o gene in APN1 deficient yeast dGMP was preferential inserted (61). Dissimilar results have also been obtained in eukaryotic cells suggesting that the A-rule may not always be operable (62–64). In any case, the preference for nucleotide insertion across form AP sites is dependent on a multiplicity of factors including sequence context and cell type. Furthermore, peculiarities dependent on the assay such as aberrant DNA synthesis of transfected DNA vs. chromosomal replication may also play a role in observed differences in nucleotide preference.

Repair of AP sites in dsDNA

Cells must cope with the high levels of AP site lesions to avoid deleterious consequences. In dsDNA, AP sites can be repaired using the high-fidelity methods of Base Excision Repair (BER) and Nucleotide Excision Repair (NER) since an intact DNA template remains to provide coding information for repair synthesis. These repair systems are fast, accurate, and likely conduct the majority of AP site repair (13, 65, 66).

Base Excision Repair

BER protects the genome from base damage by endogenous and exogenous sources. BER is a series of sequential reactions coordinated by multiple enzymes to remove small damaged or distorted DNA bases, AP sites, and AP site intermediates (reviewed here (13, 67–70)). Briefly, the five major steps are: recognition and removal of incorrect or damaged DNA base, excision of AP site by AP endonuclease or AP lyase, cleaning up of DNA ends, repair synthesis by DNA polymerase, and sealing of the nick by DNA ligase. BER can be broken into two subcategories--short-patch repair for single-nucleotide lesions and long-patch repair for multiple-nucleotide lesions. In general, short-patch repair is more prevalent and utilizes APE1 and Pol β (71).

For short patch repair, the modified base is removed by a monofunctional DNA glycosylase. AP endonuclease incises the DNA 5' to the AP lesion, leaving behind 3' hydroxyl and 5' terminal deoxyribose phosphate (5' dRP) DNA ends. The resulting one-nucleotide gap in duplex DNA is then repaired by Pol βor Pol λ with the help of accessory factors poly(ADP-ribose) polymerase 1 (PARP) and X-ray repair cross-complementing 1 (XRCC1). Finally, DNA ligase I or 3 ligates the DNA ends to seal the nick. For long patch repair, Pol α, ε, or β perform strand displacement synthesis to remove 2-10 nucleotides. A 5' flap is generated and then removed by flap endonuclease 1 (FEN1). This generates the DNA end required for ligation. Short and long-patch BER occur in parallel. Changes in cell type, cell cycle, the initiating glycosylase, or the ATP concentration all have been implicated in BER pathway choice (72, 73).

Nucleotide Excision Repair of AP Sites

In addition to BER, nucleotide excision repair (NER) can also remove AP sites from DNA. Two types of NER are possible – global genome NER (GG-NER) or transcription-coupled NER (TC-NER) (reviewed here (65, 66)) . In GG-NER, large distortions in the DNA helix, such as pyrimidine dimers, promote the recruitment of NER machinery. NER enzymes generate single-strand nicks in the DNA 5' and 3' to damage to generate a ~25-30 nucleotide single-strand fragment (74). The subsequent gap is repaired by the actions of DNA polymerase and DNA ligase. Generally, small base damage and AP sites are not thought to generate the needed helical distortion for recognition by GG-NER. However, in *E. coli* the central NER protein uvrABC can generate 3' and 5' nicks to an AP site *in vitro* (75, 76). Similarly, eukaryotic NER proteins can recognize and process AP sites *in vitro* (77). In yeast, NER inactivation increases sensitivity of AP endonuclease-deficient cells to MMS. Furthermore, the rate of AP site repair is decreased and the MMS-induced mutation frequency is increased in BER- and NER-deficient cells (78). Other studies have also found an overlap or crosstalk between BER and NER pathways suggesting that NER and BER have the ability to recognize and repair similar lesions (79).

TC-NER is a major repair mechanism for AP sites in cells. AP sites are a potent block to T7 RNAP and RNA polymerase II (RNAPII) during transcription when present on the actively transcribed strand (80). The AP lesion structurally interferes with nucleotide incorporation, and subsequent nucleotide extension past the AP site is slow (81). The stalling of transcription at AP sites is highly mutagenic (82), and AMP is preferentially inserted opposite AP sites by RNAPII (81). Furthermore, highly transcribed sequences are more susceptible to AP site accumulation and are a significant source of transcription-

associated mutagenesis (83–86). NER deficiencies cause increased transcriptional mutagenesis at the AP site (83, 87). The paused RNAPII at lesions signals the recruitment of the NER machinery. Overall, the final steps of excision, repair, and ligation remain the same as GG-NER.

AP Site Repair in ssDNA Context

AP sites in ssDNA present a different threat to genome stability that generally cannot be solved by excision repair. ssDNA AP sites can form wherever ssDNA exists such as at telomeres or in transcription bubbles. However, they are most prevalent during DNA replication due to either the unwinding of dsDNA containing an unrepaired AP site or the generation of a new AP site in the ssDNA on the lagging strand. Thus, most ssDNA abasic site repair or tolerance mechanisms operate in the context of replication. BER enzymes such as APE1 do have activity on AP sites in ssDNA, although AP-ssDNA processing by APE1 is ~20-fold less compared to AP-dsDNA (88). Many DNA glycosylases also have activity against AP-ssDNA substrates (89–91). There may be a function for BER repair in ssDNA at sites of transcription or replication (89). However, since BER cleaves the DNA backbone, it generally would not be useful for AP sites in ssDNA since there would not be a template for repair. Translesion synthesis (TLS) polymerases provide one error-prone mechanism of AP site tolerance. Homologous recombination or fork reversal combined with template switching can provide an error-free mechanism. Finally, two additional initiating mechanisms for repair or tolerance of ssDNA-AP sites were recently

described, which are mediated by Rad51 paralogs (92) or the evolutionarily conserved AP basic site shielding protein HMCES (93).

TLS Bypass of AP Sites

TLS polymerases replicate past lesions that block replicative polymerases, thus allowing cells to tolerate DNA damage (Figure 1.3A). These polymerases have a larger active site compared to replicative polymerase allowing accommodation of damaged or distorted bases. However, the low fidelity and lack of exonuclease proofreading activity of TLS polymerases combined with the lack of coding information on the template strand make bypass of DNA lesions error prone. Here we will focus on the abasic site specificity of TLS polymerases. The specificity of bypass for other lesions has been reviewed previously (94, 95).

TLS polymerases bypass often follows a two-step process where one polymerase inserts a nucleotide across from the lesion and another polymerase extends past the lesion. Following bypass and extension by TLS, further extension is resumed by high fidelity polymerases. TLS polymerases involved in abasic site bypass include the Y family of polymerases including Pol eta (Rad30) and Rev1 in *S. cerevisiae*, DNA Polη, Polι, and Polκ, and Rev1 in humans, and DNA Pol IV (DinB) and Pol V (UmuC) in *E. coli*. The Y-family polymerases generally act as insertors across from the AP site lesions. B-family polymerases such as *E. coli* DNA pol II and eukaryotic Polζ usually perform extension synthesis from the inserted nucleobase opposite the abasic site. Other polymerases such as Polλ and Polμ are generally not thought to be involved in AP site bypass.

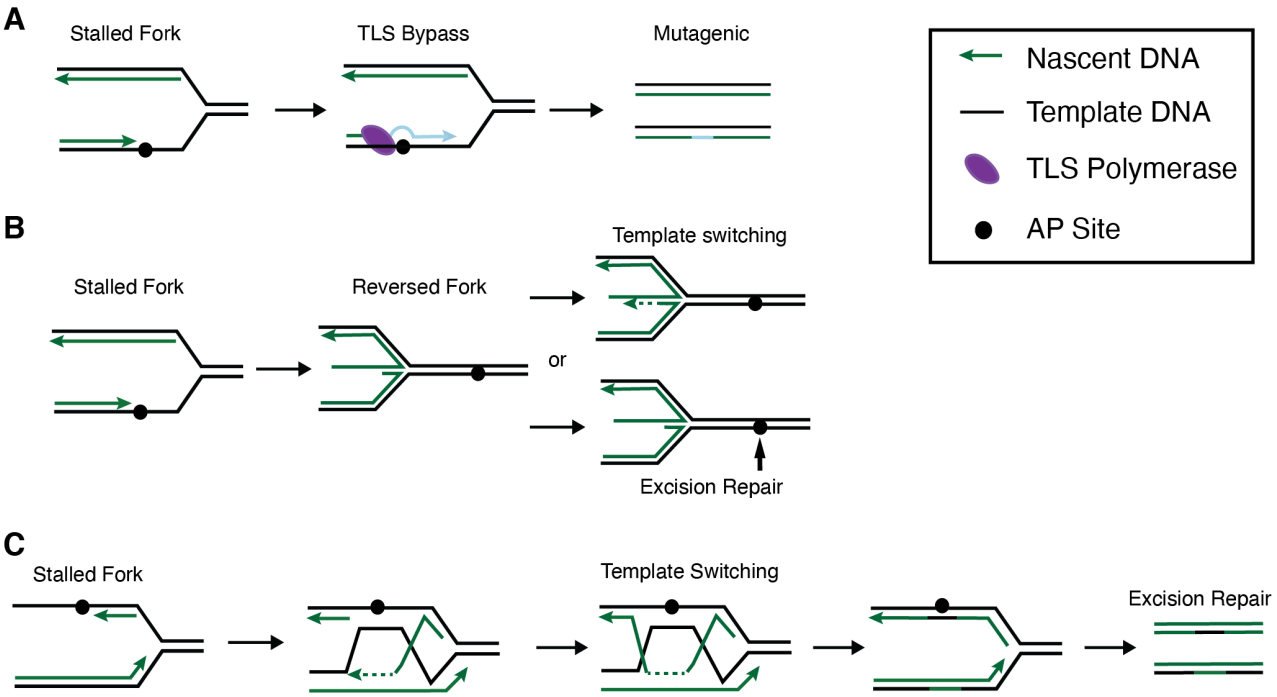

Figure 1.3: AP site repair ssDNA. (A) Translesion synthesis (TLS) polymerase bypass an AP lesion to promote error-prone damage tolerance. (B) Fork reversal generates a four-way chicken foot structure that can facilitate template switching or excision repair. (C) Template switching using a strand invasion step can place a ssDNA AP lesion in the context of dsDNA for error-free bypass and repair.

In vertebrates, much of the *in vivo* evidence for TLS bypass of AP sites comes from studies of immunoglobulin (Ig) somatic hypermutation (SHM). During B-cell development, B cells undergo a second round of antibody diversification involving the Ig locus. In this process, the rearranged IgG genes are hypermutated by activation-induced deaminase (AID), a member of the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of proteins. AID intentionally generates deoxyuridines in ssDNA. A portion of these uracils are subsequently processed by uracil DNA glycosylase to generate abasic sites that are then bypassed by TLS polymerases. Data suggests a role for Polη, Polι, Rev1, and Pol ζ in SHM although the modest effects of genetic inactivation of Polη and Polι suggest there is significant redundancy or compensation.

Polη is adept at bypassing UV photoproducts. However, its overall contribution to AP site bypass is controversial. In yeast, pol eta (Rad30) deficiency is dispensable for AP site bypass (96). In patients with a variant form of xeroderma pigmentosum (XPV), the frequency of SHM is normal (97). Similarly, in Polη-deficient mice the frequency of mutations remains overall the same compared to wild-type (98). However, the Ig mutational profile in both the XPV patients and Polη-deficient mice suggest that Polη contributes to mutations at A and T (97, 98). *In vitro*, Polη can bypass AP sites but its efficiency is dependent on sequence context and experimental conditions (99–102)

Polι may also have limited activity in AP site bypass. The overall frequency of SHM is decreased in BL2 Burkitt's Lymphoma cells deficient for Polι (103). However, mice with a natural nonsense mutation in the *POLI* gene have a normal frequency of Ig hypermutation (104). Inactivation of Polι with simultaneous loss of Polη in mice does not alter the mutation pattern seen by Polη alone suggesting a marginal role for Polι in AP site bypass

(105). Polι can incorporate nucleotides opposite an abasic site lesion with preference for dGTP and dTTP over dATP and dCTP (106, 107) but it cannot do extension synthesis after nucleotide incorporation (108).

The REV1 Y family DNA polymerase, interacts with REV7 and inserts dCMP opposite AP sites due to its intrinsic dCMP transferase activity (109, 110). Mice with a *rev1* deficiency have a normal frequency of SHM but a decrease in C to G and C to G transversions suggesting a role for *Rev1* in AP site bypass (111). Similarly, disruption of the REV1 gene in DT40 cells causes defects in SHM (112) .Using a subtelomeric ssDNA reporter system, Chan and colleagues were able to assess the *in vivo* specificity of AP site bypass by different TLS polymerases in yeast (96). By overexpressing APOBEC3G, they were able to selectively generate AP sites in ssDNA and show that Rev1 and Rev3 (the catalytic subunit of Pol*ζ*) were necessary for TLS bypass of AP sites (96). Additionally, Rev1-deficient yeast exhibit a marked decrease in mutation frequency of AP-ssDNA (96). Interestingly, catalytic dead Rev1 expressing cells have wild-type mutation frequencies suggesting it has a non-catalytic, scaffolding function for other TLS polymerases (58, 96, 113). However, the composition of mutations in the Rev1-catalytic dead cells is markedly different (96). Thus, Rev1 may be the predominant polymerase that inserts dCMP across from AP sites, which is consistent with other studies (59, 111, 112, 114, 115).

To date there does not seem to be a significant role for Pol*κ* (116, 117) or Pol θ in SHM, although Polθ can bypass AP sites *in vitro* (118, 119).The role of Polβ in SHM is controversial and its effects in B-cell development may be due other DNA repair functions (120–122). After insertion of a base across from the AP, site Pol*ζ* is involved in the

majority of extension polymerization prior to switching to high fidelity polymerases (113). Polζis a complex of the catalytic subunit REV3 and an accessory subunit REV7. Error-prone synthesis by Polζ can continue for several hundred nucleotides (123). In yeast, Rev3 is required for AP site induced mutagenesis (96, 113). Studies based on transfection of plasmids with a single AP site show that Rev3 is necessary for the majority of mutagenic TLS bypass (58, 114). Mice expressing *rev3l* anti-sense RNA have reduced mutation frequencies at all base pairs in the IgV_H genes and impaired affinity maturation of memory B cells indicating a significant function for Polζ in SHM (124). Thus, Polζ is critical for TLS-mediated bypass of AP sites.

In *E. coli,* SOS induction increases replication-dependent AP site tolerance by increasing expression of TLS polymerases. The SOS-induced polymerases umuC and umuD (Pol V) and dinB (pol IV) can bypass AP sites. Pol V is thought to be the major lesion bypass DNA polymerase for AP sites *in vivo* and contributes to the A-rule by preferentially incorporating dAMP (125–127). Deletion of umuDC operon causes a 3.5-fold decrease in spontaneous mutation frequency in BER deficient *E. coli* (128). The UmuC-UmuD' complex in conjunction with RecA and SSB (single-stranded DNA binding protein) act to rescue replicative polymerase stalled at AP sites (129, 130). In the absence of UmuD' and UmuC AP sites are skipped resulting in mostly single-nucleotide deletions (130). Pol IV (dinB) is 5-10-fold more accurate than Pol V (umuCD), but it is inefficient at AP site bypass (125, 127), and DinB does not seem to play a major role in AP site bypass.

Recombination and Template-Switching Repair of AP Sites

Homologous recombination (HR)-mediated repair and template switching have also been suggested as an error free mechanism of AP site repair. Certainly if AP sites are processed into DSBs or if AP sites are part of complex lesions such as those generated by radiation or free radicals, then then homology-directed repair of the break including resection and strand invasion would be useful (1, 131). HR-mediated repair may be particularly important during DNA replication. When a replication fork encounters an AP site, the helicase unwinds the DNA but the polymerase stalls leading to an accumulation of ssDNA. This increase of ssDNA could invoke HR and template switching pathways.

Using an engineered thymine DNA glycosylase to generate chromosomal AP sites, Otterle and colleagues demonstrated that AP site cytotoxic effects are enhanced in cells deficient in BER, TLS, and recombination repair (132). Low level induction of AP sites decreases viability of HR-deficient cells. Furthermore, chromosomal AP sites are highly mutagenic in *E. coli* strains deficient in recombination or BER (133). In *S. cerevisiae*, BER-deficient strains have an increased rate of HR (79). When HR is diminished in the setting of BER deficiency, a stronger mutator phenotype results compared to loss of BER or HR alone. Likewise, when TLS and BER are inactivated, there is a synergistic increase in the recombination rate indicating that multiple, partly compensatory pathways can converge to accomplish AP site repair (79). In mammalian cells, HR was shown to repair DNA gaps opposite an abasic site in a plasmid-based assay (134).

During replication, template switching can promote error-free repair of AP sites. In template switching the undamaged information on the sister chromatid is utilized to generate duplex DNA that can then be processed via one of the dsDNA AP site repair

mechanisms. Template switching could occur via a fork reversal process in which a four-way junction is generated by annealing the two nascent DNA strands (Figure 1.3B), or a strand invasion mechanism dependent on HR proteins like RAD51 or RAD52 (135) (Figure 1.3C). In both cases, DNA synthesis followed by a resolution step allows bypass of the lesion without introduction of a mutation (136). The AP site can then be repaired via BER at a later time. Alternatively, the lesion could be repaired while the fork is reversed to generate an intact template for continued DNA synthesis. In *E. coli*, repair of a gapped plasmid containing an abasic site was repaired 20% of the time by template switching (137). Fork reversal is thought to be a frequent event in human cells (136), but whether it actually facilitates error-free AP site tolerance is unknown (138) .

Shu Complex Promotes Damage Tolerance of AP Sites

The *S. cerevisiae* Shu complex is a heterotetrametric complex containing Shu2 (SWIM domain containing protein) and the RAD51 paralogs Csm2, Pys3, and Shu1. Based on subunit homologies to RAD51, it was originally characterized in HR repair of DSBs (139). More recently, an activity in controlling ssDNA AP site processing was discovered (Figure 1.4A). Shu mutant yeast are sensitive to MMS (methyl methanesulfonate)-induced replication blocking lesions and have an elevated mutation frequency consistent with an activity in promoting error-free lesion tolerance (140). Indeed, Shu promotes error-free bypass of MMS-induced alkylation damage and abasic sites specifically when they are present on the lagging strand template during replication (140). Csm2-Psy3 of the Shu complex binds double-flap DNA substrates with AP sites *in vitro* and accumulates on AP site associated chromatin *in vivo*. This binding of abasic

sites blocks AP endonuclease activity thereby preventing DSB formation(92). The Shu complex may both protect the ssDNA AP site from cleavage and promote RAD51-dependent template switching to prevent mutations (92). A Shu2 protein (SWS1) is found in human cells (141). SWS1 binds a RAD51 paralog called SWSAP1 (142). Inactivation of these proteins increases cellular sensitivity to DNA damaging agents that stall forks and reduces fork restart (143). However, it is not yet known if it has an ssDNA AP-site tolerance activity like the yeast Shu complex.

HMCES Initiates a New Mechanism for AP-ssDNA Recognition and Repair

The evolutionarily conserved protein HMCES (5-hydroxymethylcytosine embryonic stem cells specific) was recently discovered to initiate a pathway for recognition and repair of AP sites in ssDNA (93) (Figure 1.4B). HMCES was initially identified as a possible reader of 5hmC in embryonic stem cell extracts using a dsDNA molecule containing 5hmC as bait (144). The Cortez lab initial detected HMCES in both a whole genome siRNA screen and a proteomic screen as a replication fork associate protein required for genomic maintenance (93, 145, 146). HMCES-deficient cells are hypersensitive to ATR (ataxia-telangiectasia and Rad3-related) protein kinase inhibition. ATR is a master regulator of the replication stress response. Loss of DNA repair genes causes an increase in replication fork instability and the requirement for ATR (147). HMCES is also enriched at replication forks at levels similar to the minichromosome maintenance protein complex (MCM), a known replicative helicase required for DNA replication (93, 145). These observations suggest that HMCES is an evolutionarily conserved replication fork protein that promotes DNA damage tolerance or repair.

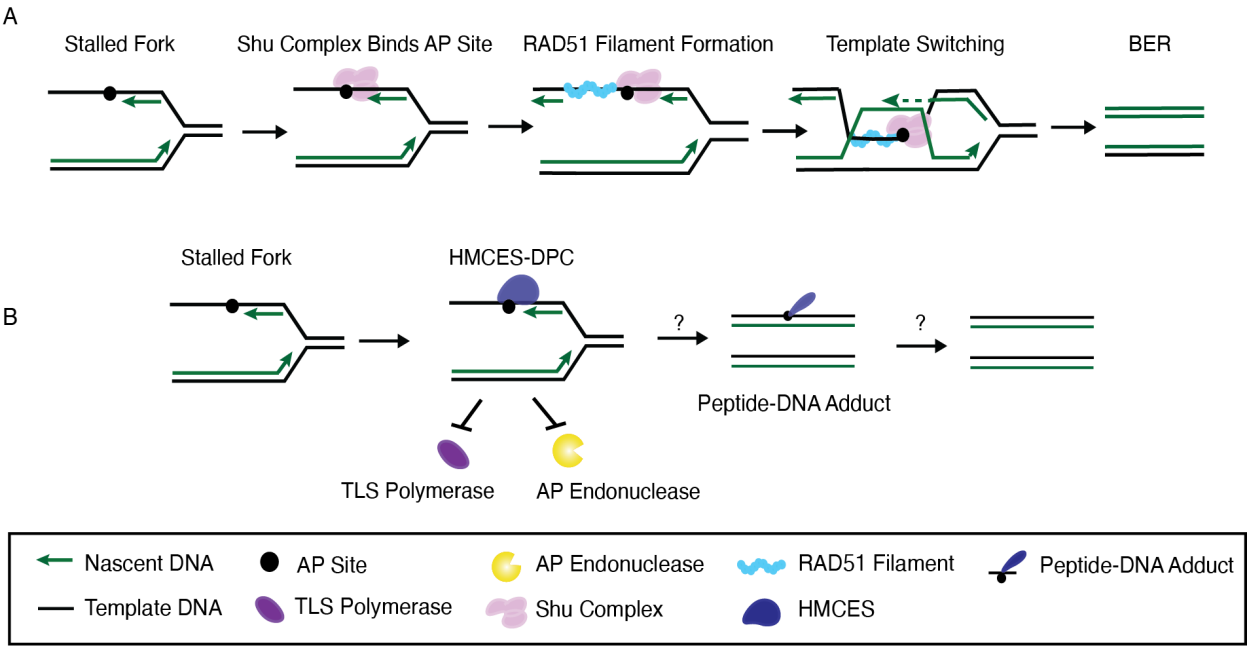

Figure 1.4: New mechanisms of AP site recognition and repair in ssDNA. (A) Model of Shu-mediated error-free damage tolerance of AP site. Shu complex binding to an AP site promotes RAD51 filament formation and template switching using the sister chromatid. BER is used to repair AP site post-replicatively. (B) HMCES initiates repair of AP sites by forming a DNA-protein crosslink to AP site in ssDNA especially in the context of DNA replication. The HMCES-AP site DPC prevents endonuclease cleavage and TLS bypass. The mechanisms by which the HMCES-DPC is ultimately repaired remain unknown.

HMCES contains an SOS response peptidase (SRAP) domain. SRAP proteins are genetically linked to SOS genes (the bacterial DNA damage response) like the TLS polymerases (148). SRAP domains have a proposed thiol autopeptidase activity dependent on a conserved triad of residues (cysteine, glutamate, and histidine) (148). This peptidase activity was suggested to be an autoproteolytic switch. (149). However, the only residue that would be removed by this proposed activity would be the N-terminal methionine since the cysteine is almost universally encoded by residue two, and amino-peptidases should typically remove the methionine. SRAP has also been reported to act as a nuclease towards 5hmc-containting DNA (146). However, HMCES-deficient or overexpressing human cells did not exhibit changes in the levels of 5mc or 5hmC, and there are very few changes in gene expression in HMCES-deficient cells. Furthermore, HMCES deficiency also does not cause a global change 5hmc in bone marrow cells (149) or adult mice (150), although some change was reported in embryonic stem cells (150). We found that HMCES covalently crosslinks to AP sites in ssDNA, and this crosslinking requires the N-terminal invariant cysteine and DNA binding residues within the SRAP domain (93).

HMCES-deficient cells exhibit marked sensitivity to AP site inducing DNA damage agents, including IR, UV, MMS, and KBrO3, and have a defect in AP site resolution (93). HMCES is recruited to chromatin in response to these agents selectively in S-phase and directly interacts with PCNA (93). Furthermore, HMCES deficient cells are hypersensitive to the nuclear expression of APOBEC3A, further supporting that HMCES is responding to AP sites in cells (151). The HMCES-DPC forms in cells and is resolved at least in part via a ubiquitin-dependent mechanism (93). HMCES-DPC formation reduces mutation

frequencies as HMCES-deficient cells have elevated Pol ζ levels at replication forks, and TLS activity generates the slow elongation kinetics seen in cells expressing nuclear APOBEC (93, 151). Additionally, the HMCES-DPC prevents cleavage of AP sites by AP endonucleases and thus the formation of double-strand breaks. Unifying these observations, the model is that HMCES acts at DNA replication forks to shield AP sites in ssDNA by forming a novel DNA-protein crosslink (DPC). HMCES-DPC prevents AP site repair via other pathways that would generate DSB or mutations, and thus promotes genetic stability (Figure 1.4B).

Understanding the Mechanism of HMCES-DPC Formation

The precise chemical mechanism of HMCES covalent attachment to AP site ssDNA DNA was unknown. Here, I will reveal the molecular mechanism of HMCES-DPC formation using biochemical assays and structural biology. In Chapter III, I demonstrate that HMCES and yedK are ssDNA binding proteins that crosslink to AP-ssDNA. This AP-ssDNA is the preferred substrate for HMCES, and HMCES does not have a preference for 5mC and its oxidized derivatives. SRAP DNA binding is dependent on conserved arginine within the basic groove channel, and the crosslinking activity is dependent on an invariant N-terminal cysteine. The SRAP-DPC is stable, resistant to AP endonucleases, and is enhanced under reduced and acidic conditions. Given the remarkable stability of the SRAP-DPCs, I became curious about the chemical nature of the DPC. In Chapter IV, a high-resolution crystal structure of yedK and AP-ssDNA reveals that the SRAP-DPC is formed via a thiazolidine linkage. The crystal structure also supports the biological context

of HMCES function, shielding AP sites at a stall replication fork. In Chapter V, I will describe how these findings offer insight into a new AP-ssDNA repair pathway. Importantly, I will discuss how this work will impact understanding of a new evolutionarily conserved DNA repair pathway, DPC repair, and the potential role for HMCES in B-cell development.

CHAPTER II

Materials and Methods

Protein Purification

E. coli YedK was expressed in a modified pBG101 vector containing a Rhinovirus 3C (PreScission) protease cleavable hexahistidine tag. *E. coli* BL21 (DE3) cells were grown in 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.

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

Human HMCES-SRAP domain (amino acids 1-270) was purified 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.

Codon optimized C-terminally tagged human SRAP domain (1-270)-HIS, yedK-HIS, N-terminal HIS-GST-SRAP (1-270), and K61-HIS-HMCES fusion proteins were induced with 1mM IPTG for 6 hr at 30° C in BL21 bacterial cells. Cells were resuspended in HIS binding buffer (50mM Tris pH 8, 100mM NaCl, 10mM imidazole, 0.1mM EDTA,5% glycerol, 1mMPMSF, 1 mg/mL aprotinin, and 1 mg/mL leupeptin), treated with 1mg/mL lysozyme for 30 min on ice, and sonicated. Soluble lysates were combined with Ni-NTA agarose for 3 hr and bound proteins were washed three times with HIS wash buffer (50mM Tris pH 8, 300mM NaCl, 20mM imidazole, 0.1mM EDTA, 5% glycerol, 1mM PMSF, 1 mg/mL aprotinin, and 1 mg/mL leupeptin), and eluted with HIS elution buffer (50mM Tris pH 8, 300mM NaCl, 300mM imidazole, 0.1mM EDTA, 5% glycerol, 1mM

PMSF, 1 mg/mL aprotinin, and 1 mg/mL leupeptin). Samples were then separated using a S200 size exclusion column in 50mM Tris pH 8,150mM NaCl, 5% glycerol, and 10mM DTT. Fractions containing SRAP or yedK were pooled. Where indicated, tags were cleaved overnight at 4°C with PreScission protease and the free GST and the PreScission protease (which has a GST tag) were removed with glutathione Sepharose beads.

EMSA for DNA Binding

Sequences of oligonucleotides used in the biochemical assays are listed in Table 2.1. Relative binding affinity was measured by EMSA using ^{32}P-labeled DNAs containing a deoxyuracil. DNA (1 nM) was incubated with the indicated concentration of HMCES-SRAP, GST-HMCES, or yedK-HIS 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% (v/v) 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.

For denaturing electrophoresis, formamide was added to a final concentration of 33%, samples were boiled for 5 min, and resolved on 10% polyacrylamide gels containing 8M urea in 1X TBE (100mM Tris, 90mM boric acid, 2mM EDTA) at 40V for 180 min at room temperature.

DNA-protein Crosslinking Assays

For the experiments shown in Figure 3.6A-C, abasic (AP) DNA was prepared by incubating 50 µM uracil-containing oligonucleotides with 25 units of uracil DNA

glycosylase (UDG, New England Biolabs) in Buffer X (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT) at 37 °C for 30 min. Human HMCES-SRAP was incubated with AP-DNA in Buffer X at the following concentrations: 20.8 μM protein + 25 μM DNA (Figure 3.6A) and 0.75 μM protein + 1.5 μM DNA (Figure 3.6B,C). For the experiment shown Figure 3C, 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.

For the experiments shown in Figures 3.6D-E, 4.4C-E, and 4.5E, reaction products were separated on 15% polyacrylamide urea gels in 1X TBE buffer. In Figure 3.6D, 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 Figure 3.6E, 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.

E. coli YedK DPCs (Figure 4.4B-D) were formed from incubation of 1 μM protein and 10 nM 5′-FAM-labeled oligonucleotide in 20 mM Tris-HCl pH 6.0, 1 mM EDTA, and 5 mM DTT at 37 °C for 1 hr. Schiff base intermediates (Figure 4.4D) 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.

Crosslinking reactions with ssDNA-dsDNA junctions (Figure 4.5E) were carried out with 10 nM AP-DNA and increasing concentration of HMCES-SRAP at 37 °C for 1 hr in Buffer X.

Biotin-DNA Pull Down Assay

Dynabeads T1 (Life Technologies) were washed twice in TE buffer and bound to biotinylated DNA substrates at 4 ° C for 1 hr. Beads were washed twice again in TE buffer followed with two washes in binding buffer (10mM Tris pH 7.9, 50mM NaCl, 10mM $MgCl_2$, 2mM DTT, 0.1mg/mL BSA in RNase/DNase free water). 1 ul of beads with 5 picomoles of bound DNA was resuspended in binding buffer. Approximately 2.5 pmol purified K61-HIS-HMCES was added to the mix and rotated at 4 °C for 30min. The supernatant discarded and the beads were washed 2X in wash buffer (10mM Tris pH 7.0, 50 mM NaCl, 10 mM MgCL2, 2mM DTT, 0.1mg/ml BSA 0.1% NP-40 in RNase/DNase free water). The samples were boiled in 2X sample buffer for 5min. Captures were analyzed by immunoblotting for HMCES (rabbit anti-HMCES; Sigma, HPA044968). The DNA oligos used for pulldown assays are listed in Table 2.1.

X-ray Crystallography

Abasic (AP) DNA was prepared by incubating 50 μM 7-mer d(GTCUGG) 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 protein-DNA solution, 2 μL reservoir containing 0.1 M Bis-Tris pH 5.4 and 23% (w/v) PEG 3350, and 0.5 μL DPC microcrystal seed stock stored in 30% PEG 3350 and 0.2 M KH_2PO_4. Crystals were harvested after 16 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 (152). Data collection statistics are provided in Table 4.1. Phasing and refinement was carried out using the PHENIX suite of programs (153). 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 (154), 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 protein-DNA model was iteratively refined by energy minimization and visual inspection of the electron density maps. The C3-spacer structure was phased by molecular replacement using the protein from the DPC structure, followed by simulated annealing to eliminate model bias 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 (155). Structure images were created in PyMOL (https://pymol.org). Sequence conservation was mapped onto the

structure using the Consurf Server (156). 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. Structures were deposited in the Protein DataBank under accession codes 6NUA (DPC) and 6NUH (C3-spacer).

E. coli Cellular Assays

Genetic knockouts of *yedK_17 and yedK_18* were obtain from the *E. coli* Keio knockout collection (Dharmacon, GE Healthcare) that contained a kanamycin resistance cassette in place of the endogenous gene. The *yedK_1* was generated in the lab by Nancy Zhao and Kareem Mohini. To generate E. coli growth curves, overnight cultures were diluted to OD600≈0.01 in LB in a 96-well flat-bottom plate. The plate was incubated at 30 °C with shaking for 12 hr and cell density was measured at 600 nm every 20 minutes using a Bio-Tek Synergy microplate reader. For E. coli survival after DNA damage agents, a saturated overnight culture from a single colony was diluted to OD600≈0.01 in fresh LB media and grown to OD600≈0.4 at 37 °C. Serial dilutions were plated on LB plates containing 0.05% w/v MMS or LB plates irradiated with 50 J/m2 UV. For E. coli cells treated with 4NQO, cells were plated on LB plates with 10 μM 4NQO, incubated at 37 °C, and colonies were counted the next morning. The CFU/mL was then calculated.

For the genetic experiments with APOBEC3A, APOBEC3A was expressed from a pSF_OBX1 (Kan^R) plasmid that allows for constitutive gene expression at low levels in *Δyedk* cells where the kanamycin cassette was removed. The ORF of *yedK* WT or C2A was cloned in tandem with A3A in the psF_OBX1_A3A plasmid. Protein expression of

YedK was not confirmed as there was no working YedK antibody. Individual colonies were grown overnight at 37 °C in LB + kanamycin. Overnight cultures were diluted to OD600≈ 0.01, 100 uL of 10^{-7} dilution of OD600≈0.4 cells were plated on LB + kanamycin plates. Plates were incubated at 37 °C and colonies were counted the following morning. The CFU/ug DNA plasmid was then calculated.

Gene Expression by Quantitative RT-PCR

Saturated E. coli cultures from a single colony were diluted to OD600≈ 0.01 in LB media and grown at 37 °C to an OD600≈ 0.5, after which 0.05% w/v MMS was added and cultures were grown at 37 °C for an additional hour. Cells were collected (1 mL), pelleted, and frozen at -80 °C. Cultures were lysed and the RNA extracted with the Aurum Total RNA Mini Kit (Bio-Rad) according to the manufacture's protocol. Residual genomic DNA was removed from the RNA by treatment with DNaseI. The extracted RNA was verified by running on 1% agarose gel and quantified absorbance at 260 and 280 nm using a NanoDrop UV-Vis spectrophotometer. The specificity of the primers was verified by agarose gel analysis of PCR products. cDNA synthesis and qPCR were performed in a single step reaction using the iTaq Universal SYBR Green One-Step Kit (Bio-Rad) on an Applied Biosystems StepOne Plus PCR machine. The housekeeping gene used was *gapA*. The results from the qPCR experiment were performed in three technical replicates. The fold expression change was calculated using the formula: (fold expression change) = $2^{-\Delta\Delta Ct}$, where Ct is the cycle threshold for amplification above baseline, ΔCt = Ct (gene of interest) – Ct (housekeeping gene), and ΔΔCt = ΔCt (treated sample) – ΔCt (untreated sample).

Table 3.2 Sequence of DNA Ligands Used (5' → 3')

#	Sequence	Length (nt)
1	GGATGATGACTCTTCTGGTCTGGATGGTAGTTAAGTGTTGAG	42
2	CTCAACACTTAACTACCATCCGGACCAGAAGAGTCATCATCC	42
3	ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCACCTGCAGGTTCACCC	60
4	TCGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAGGCACTGGTAGAATTCGGCAGCGT	60
5	/5BiosG/GGATGATGACTCTTCTGGTCCGGATGGTAGTTAAGTGTTGAG	42
6	GGATGATGACTCTTCTGGTC/iMed-C/GGATGGTAGTTAAGTGTTGAG	42
7	GGATGATGACTCTTCTGGTC/i5HydMe-dC/GGATGGTAGTTAAGTGTTGAG	42
8	CTCAACACTTAACTACCATC/iMe-dC/GGACCAGAAGAGTCATCATCC	42
9	CTCAACACTTAACTACCATC/i5HydMe-dC/GGACCAGAAGAGTCATCATCC	42
10	/5BiosG/GGATGATGACTCTTCTGGTC/iMed-C/GGATGGTAGTTAAGTGTTGAG	42
11	/5BiosG/GGATGATGACTCTTCTGGTC/i5HydMe-dC/GGATGGTAGTTAAGTGTTGAG	42
12	GGATGATGACTCTTCTGGTC/idSp/GGATGGTAGTTAAGTGTTGAG	42
13	GGATGATGACTCTTCTGGTC/ideoxyU/GGATGGTAGTTAAGTGTTGAG	42
14	TGGTC/ideoxyU/GGAT	10
15	FAM-CGGGCGGCGGCA/ideoxyU/AGGGCGCGGGCCTTTTTT-Cy5	31
16	GGGTGAACCTGCAGGTGGGCAAAGATGTCC	30
17	TAGCAAGGCACTGGTAGAATTCGGCAGCGT	30
18	CTTCTGGTC/ideoxyU/GGATGGTAGT	20

Table 3.2 Schematic of DNA Ligands Prepared in Each Figure.

Figure	Sub-Figure	DNA Structure	Annealed Oligonucleotides
3.2	A	ssDNA	1
		dsDNA	1+2
	B	Splayed arm	3+4
		ssDNA	1
	C	ssDNA	5
		dsDNA	5+2
3.3	A	ssDNA	1
		ssDNA + 5mC	6
		ssDNA +5hmC	7
	B	dsDNA dC/dC	1+2
		dsDNA 5mC/5mC	6+8
		dsDNA 5hmC/5hmC	7+9
	C	ssDNA dC	5
		ssDNA 5mC	10
		ssDNA 5hmC	11
		dsDNA dC/dC	4+2
		dsDNA 5mC/dC	11+2
		dsDNA 5hmC/dC	7+2
3.4	B	ssDNA dT	1
		ssDNA THF	12
		ssDNA + abasic site	13+ UDG
	C	ssDNA	1
3.5	A	ssDNA + abasic site	13+UDG
	B	ssDNA + abasic site	14+UDG
3.6	A	ssDNA + abasic site	14+UDG
	B	ssDNA + abasic site	13+UDG
	C	ssDNA + abasic site	13+UDG
3.7	A	ssDNA + abasic site	18+ UDG
	B	ssDNA + abasic site	15+UDG
3.8		ssDNA + abasic site	14+UDG
4.5	D	ssDNA	1
		3’ ssDNA/dsDNA	1+16
		5’ ssDNA/dsDNA	1+17
	E	ssDNA	13+ UDG
		3’ ssDNA/dsDNA	13+16
		5’ ssDNA/dsDNA	13+17

Table 3.3 List of qPCR Primers

Oligo Name	Sequence (5'→ 3')	Description
yedK_F5	TTCATCTGGATCCGGTTTTC	HMCES E. coli ortholog
yedK_R5	GGTTGTCCATCAGCACGATA	
LexA_FP	CTGTTGCAGGAAGAGGAAGAA	SOS Response
LexA_RP	GGAAGGATCGACCTGATAATGAC	
GapA_FP	CGGTACCGTTGAAGTGAAAGA	Housekeeping control
GapA_RP	ACTTCGTCCCATTTCAGGTTAG	
Tag_FP	GAGTCAGGACCCGCTTTATATT	Negative Control
Tag_RP	GGACGGTGATCCACGATAAT	

CHAPTER III

SRAP Forms DNA-Protein Crosslink with Abasic Sites In ssDNA*

Introduction

Human HMCES contains one structured domain, the SOS Response Associated Peptidase (SRAP) domain. Nearly every organism, in all three kingdoms of life, encodes for a single SRAP domain protein. The SRAP domain is linked to several genes within the SOS response (DNA damage response in bacteria), including TLS polymerases such as DinB (148). Human HMCES and the *E. coli* ortholog YedK are similar in sequence identity (29%) and similarity (43%). Furthermore, the human and bacterial SRAP domains are structurally similar with an RMSD of 1.2 Å for all backbone atoms. The SRAP domain contains an invariant cysteine at amino acid position two as well as a conserved histidine and glutamic acid that sits within a putative catalytic pocket (148). Adjacent to this catalytic pocket is a positive groove surface (Figure 3.1).

The biological function of HMCES is unclear. Aravind and colleagues speculated that SRAP was an SOS response associated protein with thiol peptidase activity, but this was never verified experimentally (148). Another group identified HMCES as a reader of 5-methylcytosine (5mC) and its oxidized derivatives, but the function of HMCES as an epigenetic reader was not tested (144). Kweon and colleagues described SRAP as an autoproteolytic endonuclease with

* Parts of this chapter were published in Mohni KN, Wessel SR, Zhao R, Wojciechowski AC, Luzwick JW, Layden H, Eichman BF, Thompson PS, Mehta KPM, and Cortez, D. (2019) HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell*. **176**, 144-153, and Thompson PS+, Amidon KM+, Mohni KN, Cortez D, Eichman BF. (2019) Protection of Abasic Sites During DNA Replication by A Stable Thiazolidine Protein-DNA Crosslink. *Nat. Struct. Mol. Biol*. 2019 Jul;26(7):613-618. +Equal contribution

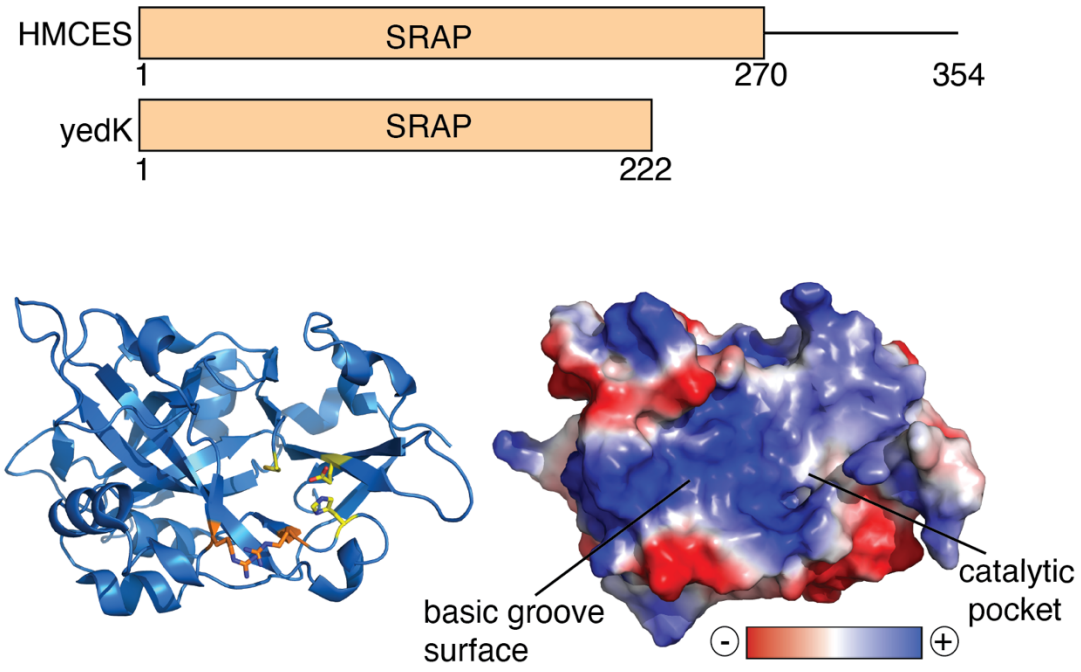

Figure 3.1: Structure of SRAP domain (PDB: 5KO9). (A) Cartoon schematic of HMCES-SRAP and YedK proteins. (B) Ribbon structure with the predicted catalytic residues (Cys, Glu, His) highlighted in yellow and surface exposed arginines are shown in orange. (C) Electrostatic map of human SRAP domain.

selective activity for TET-oxidized derivatives of 5mC (149). Thus, HMCES has three proposed functions: peptidase, epigenetic reader, and endonuclease.

The Cortez lab identified HMCES as a replication fork associated protein required for efficient replication stress response (93, 145, 157). HMCES-deficient cells exhibit marked sensitivity to DNA damaging agents that cause distinct DNA lesions, including short-wavelength ultraviolet radiation (UV), ionizing radiation (IR), methyl methanesulfonate (MMS), and potassium bromate ($KBrO_3$). These DNA lesions are predominately repaired by nucleotide excision repair (NER), double-strand break (DSB) repair, and base excision repair (BER), respectively. Interestingly, all these agents can also generate abasic sites through the destabilization of the N-glycosyl bond (1). The DNA lesion that HMCES recognized was not defined, and the significance for HMCES DNA binding was unclear. In this chapter, I will describe the identification of the preferred substrate for HMCES, an AP site in single-strand DNA. I will show that HMCES recognition of AP sites generates an unusual DNA protein crosslink (DPC). Also, this DPC formation is enhanced by dithiothreitol (DTT) and low pH. I will also show the structural determinants of the SRAP domain that are required for DNA binding and crosslink formation. In particular, the invariant cysteine at amino acid position two is critical for DPC formation. Finally, I will show that the SRAP-DPC is stable and resistant to cleavage by AP endonucleases.

Results

HMCES is a ssDNA Binding Protein

We hypothesized that the large positive groove track might be a DNA binding surface, based on the available unpublished structure of the SRAP domain (Figure 3.1). We tested the DNA binding activity of recombinant HMCES-SRAP, GST-HMCES, and yedK-HIS using electrophoretic mobility shift assays.

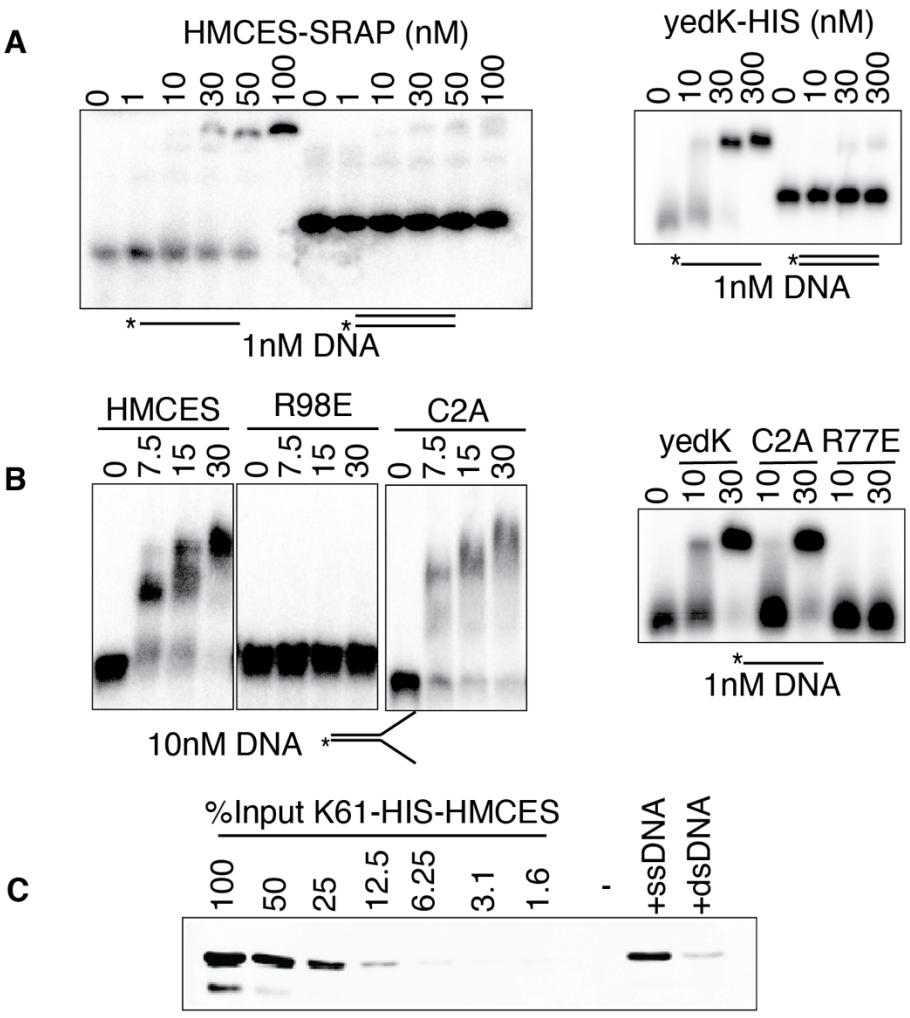

Figure 3.2: HMCES is a ssDNA binding protein. (A) Electrophoretic mobility shift analysis (EMSA) of 1 nM ssDNA or dsDNA incubated with HMCES-SRAP (*left*) or YedK-HIS (*right*) at the indicated concentrations. (B) EMSA of 1 nM ssDNA incubated with WT, C2A, or R98E GST-HMCES with GST tag cleaved (*left*) or WT, C2A, R77E YedK-HIS (*right*). (C) DNA pull-down assay with purified K61-HIS-HMCES and ssDNA or dsDNA coupled to magnetic beads.

HMCES-SRAP and yedK-HIS bind ssDNA at subnanomolar affinity but not dsDNA (Figure 3.2A). Furthermore, point mutations of two surface-exposed arginines in GST-HMCES (R98 and R212) and yedK-HIS (R77 and R162) abrogate DNA binding activity (Figure 3.2B). The preference for ssDNA over dsDNA was further verified using a biotin-DNA pull-down assay with purified recombinant HMCES protein (Figure 3.2C).

HMCES Does Not Have A Preference for 5mc or 5hmc

HMCES was identified as a possible reader of 5-hydroxymethylcytosine with endonuclease activity towards DNA substrates containing 5mC and its oxidized derivatives (144, 149). However, we did not observe any changes in the total amounts of 5mC or 5hmC in HMCESΔ or HMCES overexpressing cells (93). To test if 5mC and its oxidized derivate 5hmC could be potential substrates for HMCES, we determined the DNA binding activity of HMCES to ssDNA and dsDNA substrates with C, 5mC, or 5hmC in a CpG context *in vitro*. Based on the electrophoretic mobility shift assay, HMCES does not have a preference for DNA ligands with C, 5mC, or 5hmC in ssDNA in the sequence context tested (Figure 3.3A). Furthermore, there was no preference for dsDNA substrates with C, 5mC, or 5hmC located opposite of each other on each strand (Figure 3.3B). A biotin-DNA pull-down assay with purified recombinant HMCES protein was also used to test the preference for DNA ligands with C, 5mc, or 5hmC. HMCES did not have a preference for ssDNA with C or 5mc with a modest increase in binding to 5hmc. This is more than likely due to experimental variability. Finally, there was some residual binding to duplex DNA with the oxidized cytosine derivatives on one strand (Figure 3.3C).

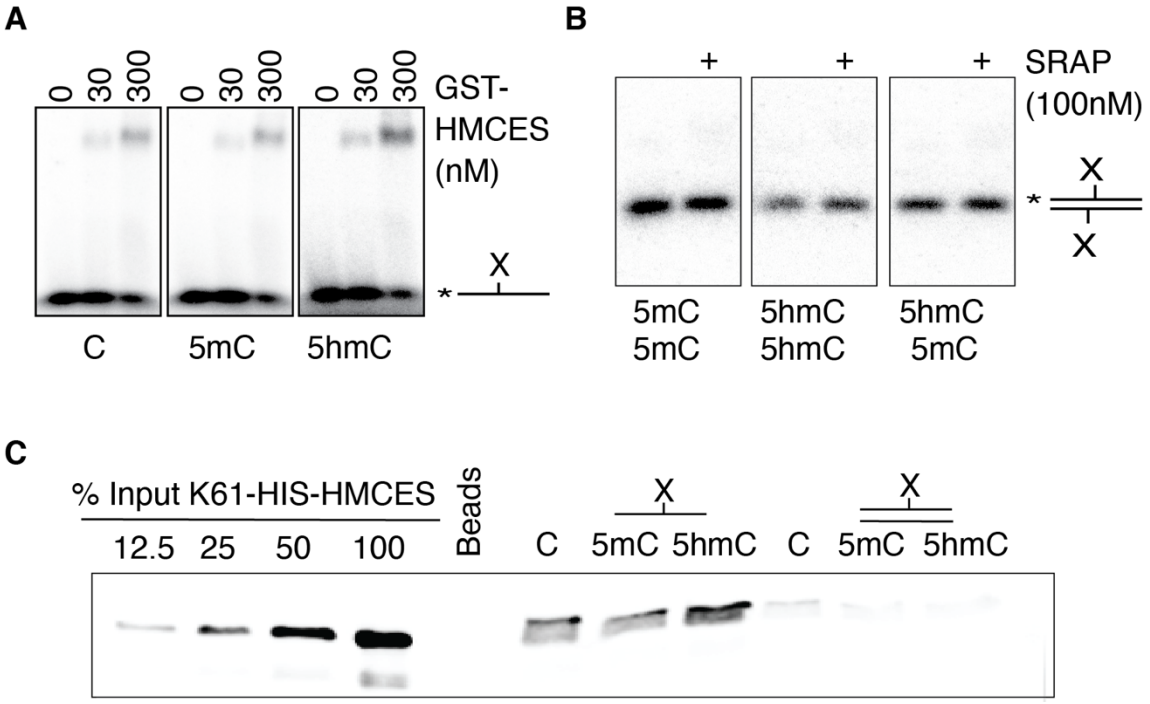

Figure 3.3 HMCES does not have a preference for 5mC or 5hmC. Electrophoretic mobility shift analysis of GST-HMCES (A) or HMCES-SRAP (B) with ssDNA 10 nM (A) or dsDNA(B) ligands with modified cytosine in a CpG context. (C) DNA pull-down assay using purified K61-HIS-HMCES and ssDNA or dsDNA with modified cytosines coupled to magnetic beads. Figures A & B produced by K. Mohni and figure C was produced by P. Thompson.

SRAP Proteins Bind Natural Abasic Sites

HMCES deficient cells are hypersensitive to IR, MMS, and shortwave UV radiation. These distinct DNA damaging agents generate AP sites through oxidation of the DNA base and destabilization of the N-glycosyl bond. We hypothesized that AP sites could be the common DNA lesion that underlies the hypersensitivity of HMCESΔ cells. The SRAP domain of HMCES and yedK both bind to ssDNA containing an AP site. Surprisingly, this binding persisted even under denaturing conditions, suggesting a covalent interaction between SRAP and AP site containing ssDNA (Figure 3.4B).

AP sites exist in equilibrium between the ring-closed furanose form and the open-chain aldehyde form (31, 158). Interestingly, the SRAP proteins did not covalently crosslink to ssDNA containing a tetrahydrofuran (THF) abasic site mimic, which cannot undergo ring opening to form an aldehyde. Because SRAP did not crosslink to THF, we concluded that the SRAP proteins covalently crosslink to the open-ring furanose AP site and not the ring-closed furanose. SRAP crosslinking activity is dependent on the invariant second amino acid cysteine and the positive DNA binding surface (Figure 3.4C). Also, treatment of the SRAP-DPC with protease restored the DNA gel migration to that of full-length ssDNA, which suggests that SRAP does not contain endonuclease activity (Figure 3.4C). Finally, SRAP is unable to crosslink to AP sites in duplex DNA (93).

SRAP-DPC Formation is Enhanced Under Reduced and Acidic Conditions

To further understand the chemical requirements for SRAP-DPC formation, I incubated abasic site containing ssDNA with HMCES-SRAP in the presence or absence of the reducing agent, dithiothreitol (DTT). Interestingly, HMCES-SRAP binds a dT unmodified ssDNA substrate more efficiently in the presence of DTT (Figure 3.5A). The presence of multiple DNA-protein complexes may be due to multiple protein binding events as the 42-nucleotide ssDNA ligand

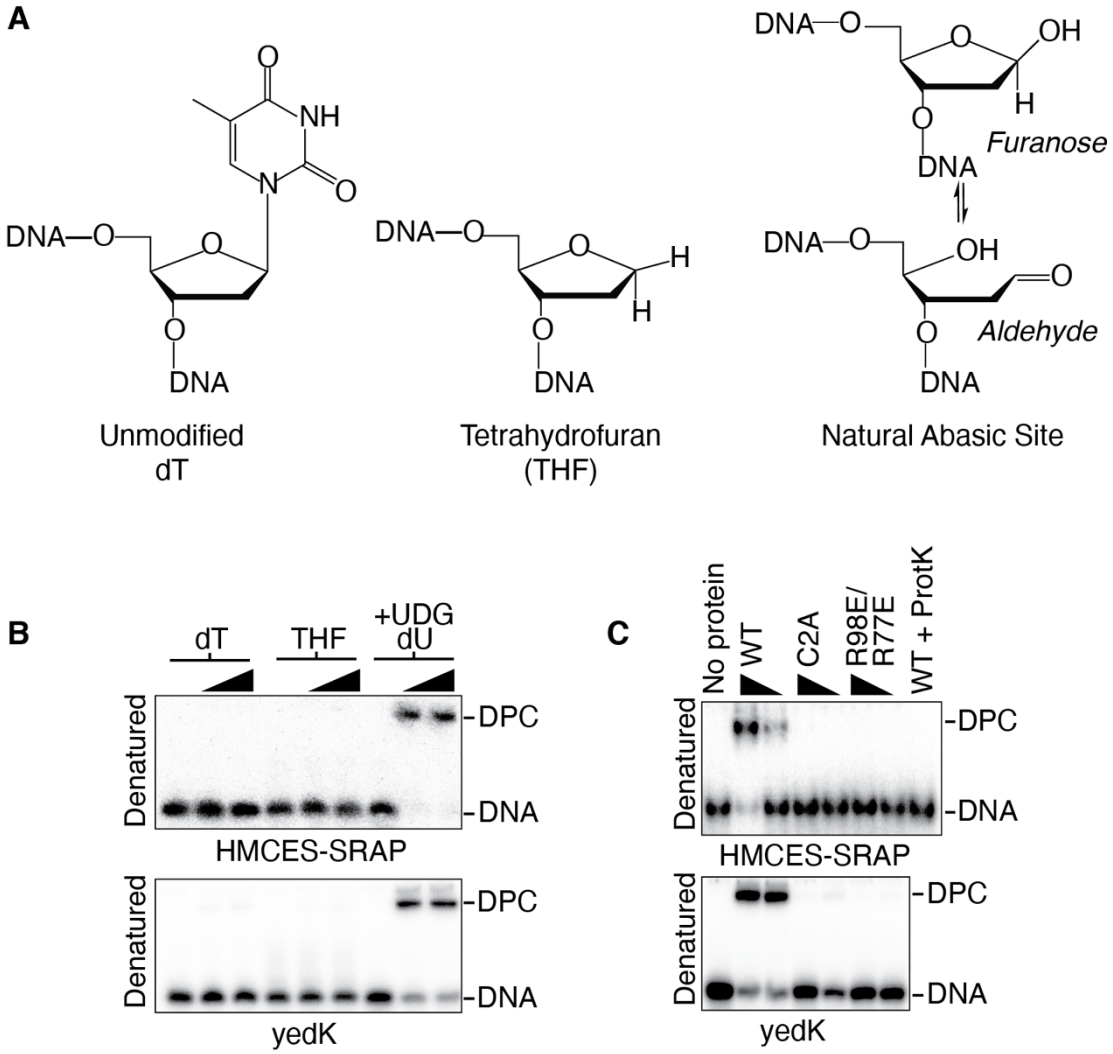

Figure 3.4 HMCES covalently crosslinks to abasic sites. (A) Schematic DNA oligos used unmodified (dT), tetrahydrofuran (THF), and UDG treated dU to generate natural abasic site. (B) Denaturing EMSA of 1 nM of dT, THF, or abasic site (dU + UDG) ssDNA incubated with wild type HMCES-SRAP and YedK (3, 10 nM). (C) Denaturing EMSA of 1 nM of dT, THF, or abasic site (dU +UDG) ssDNA incubated with wild type or mutant HMCES-SRAP (10, 1 nM) or YedK (30, 10 nM). Proteinase K was used to digest the HMCES-SRAP-DPC. HMCES-SRAP experiments were completed by K. Mohni and YedK experiments were produced by A. Wojciechowski under the direction of P. Thompson.

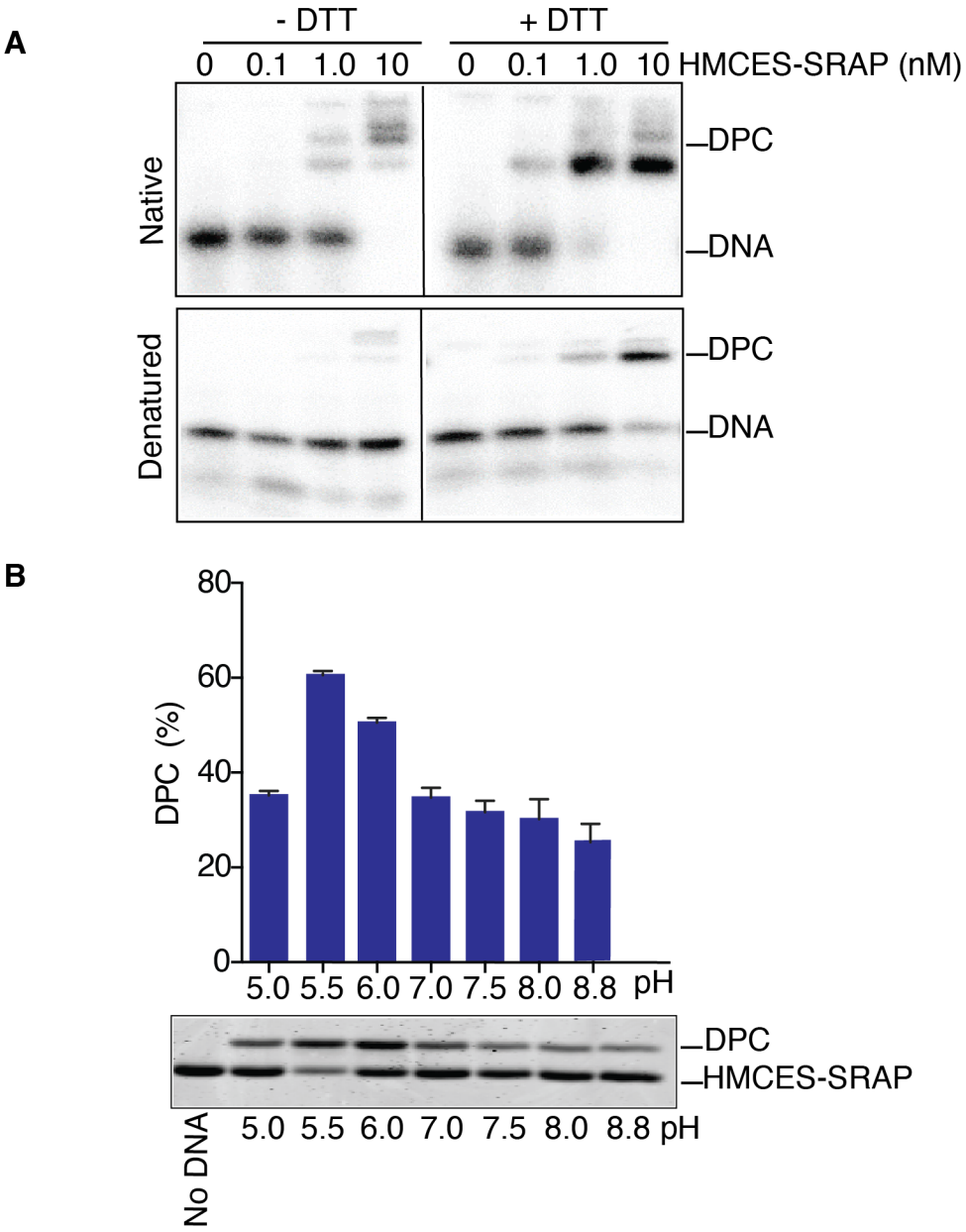

Figure 3.5: HMCES-SRAP-DPC formation is enhanced under reducing and acidic conditons. (A) Native (top) and denaturing (bottom) EMSA of 1 nM of abasic site (dU +UDG) ssDNA incubated with HMCES-SRAP (0.1,1.0,10 nM) in the presence or absence of 1 mM dithiothreitol (DTT). (B) HMCES-SRAP-DPC measured at the indicated pH levels. Free and DNA-crosslinked HMCES-SRAP were detected via Coomassie blue staining (mean +/- s.d., n=3 independent measurements).

could accommodate more than one HMCES-SRAP molecule. Interestingly, HMCES-SRAP forms a covalent linkage to abasic site containing ssDNA only in the presence of DTT. Other requirements for HMCES-DPC formation were also tested. Using a range of pH buffers (5.0 to 8.8), I observed that the HMCES-SRAP-DPC forms more readily under acidic conditions when incubated with abasic site containing ssDNA for one hour (Figure 3.5B). This data suggests that the chemical mechanism for HMCES-DPC formation occurs more readily under reduced and acidic conditions.

SRAP-DPC Formation with AP-ssDNA is Stable

The ease of detecting a HMCES-DPC in cells hints that it may be a stable chemical linkage (93). Furthermore, the persistence of SRAP-DPC, even under denaturing conditions, suggested that SRAP-DPC is very stable. To test the stability of HMCES-DPC, I performed several experiments (Figure 3.6A-C). Incubating the human HMCES-SRAP domain DPC at 4, 25, 37 °C for up to six days did not change the percentage of crosslinked protein (Figure 3.6A). Thus, the HMCES-DPC remains stable for several days and under physiological conditions. I noticed while doing these experiments that boiling the DPC hydrolyzed the crosslink, but that incubation at 50 °C did not (Figure 3.6B). Protein denaturation is insufficient 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 (Figure 3.6C). Therefore, the SRAP-DPC is remarkably stable.

HMCES-DPC is resolved by a proteasome-dependent manner that presumably leaves behind a peptide-DNA adduct (93). To test the stability of this peptide-DNA adduct, I proteolyzed the DPC with proteinase K. Extensive proteolysis of the DPC with proteinase K left a small peptide-DNA linkage that remains stable (Figure 3.7A). Furthermore, the peptide-DNA linkage is resistant to cleavage by AP endonuclease 1 (APE1) (Figure 3.7B). The SRAP-DPC is readily purified by size exclusion chromatography (Figure 3.8). This protein-DNA interaction persists

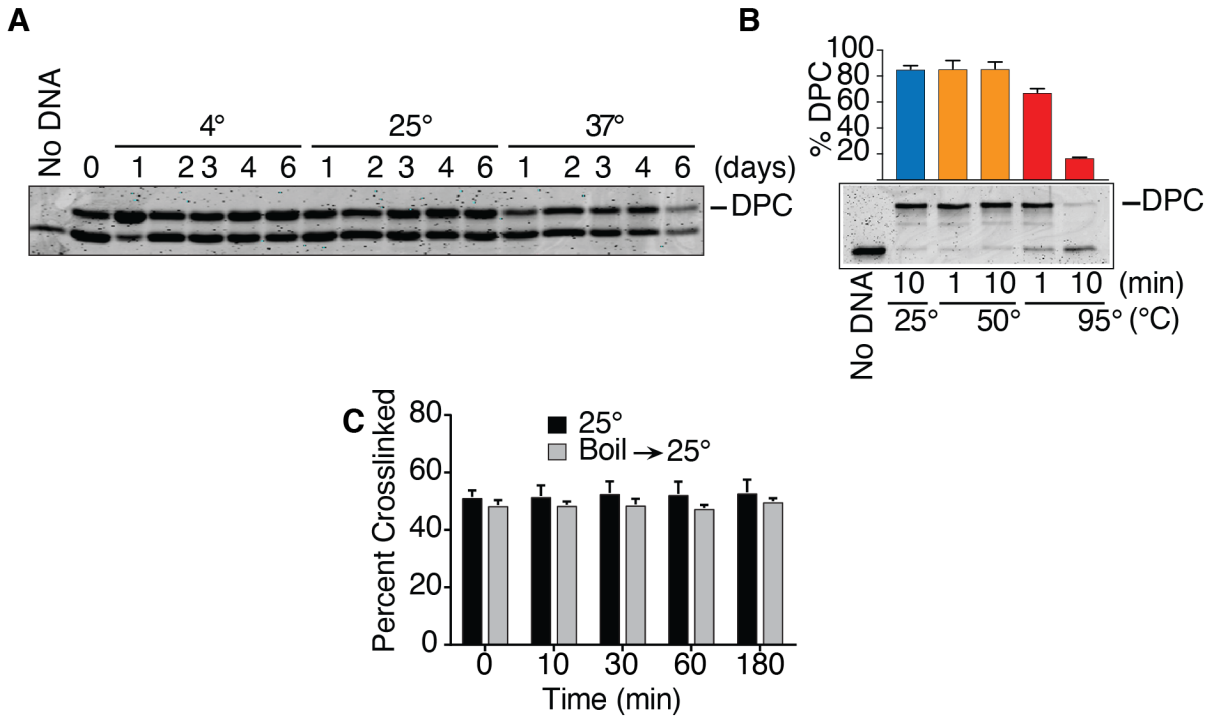

Figure 3.6: Stability analysis of human HMCES-SRAP abasic site DNA-protein crosslink. (A) HMCES-SRAP-DPC stability measured at the indicated temperatures. Free and DNA-crosslinked HMCES were detected by Coomassie blue staining. The HMCES-DPC percentage in this experiment is approximately 50% because uncross-linked DNA was removed by dialysis after a short reaction time. (B) Boiling the HMCES-DPC causes hydrolysis (mean +/- s.d., n=3 independent experiments).(C) HMCES-DPC stability measured before or after denaturation by boiling for 2 min.

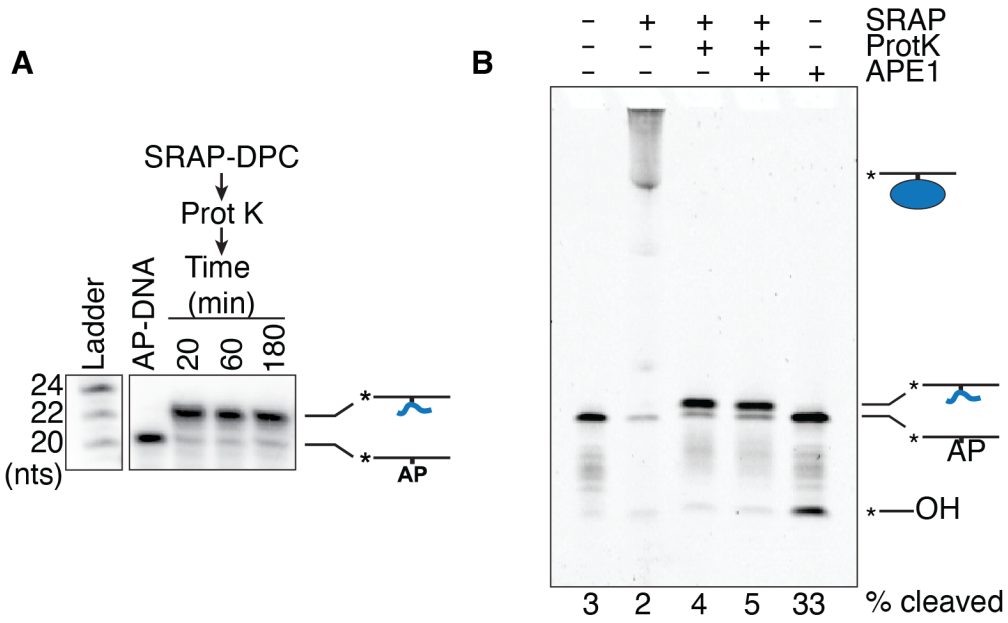

Figure 3.7 HMCES-SRAP peptide-DNA adduct is stable and resistant to AP endonuclease cleavage. (A) HMCES-SRAP domain was incubated with 20-mer AP site containing oligonucleotide to form a crosslink, digested with proteinase K followed by heat inactivation of the protease, and then incubated at 37°C for the times indicated. Electrophoresis and autoradiography were used to visualize DNA. (B) HMCES-SRAP was incubated with 31-mer AP DNA and digested with proteinase K, and the peptide DPC was intubated with APE1 for 2h. Bands were visualized by Cy5 fluorescence. Figure 3.7A produced by P. Thompson and Figure 3.7B produced by K. Amidon.

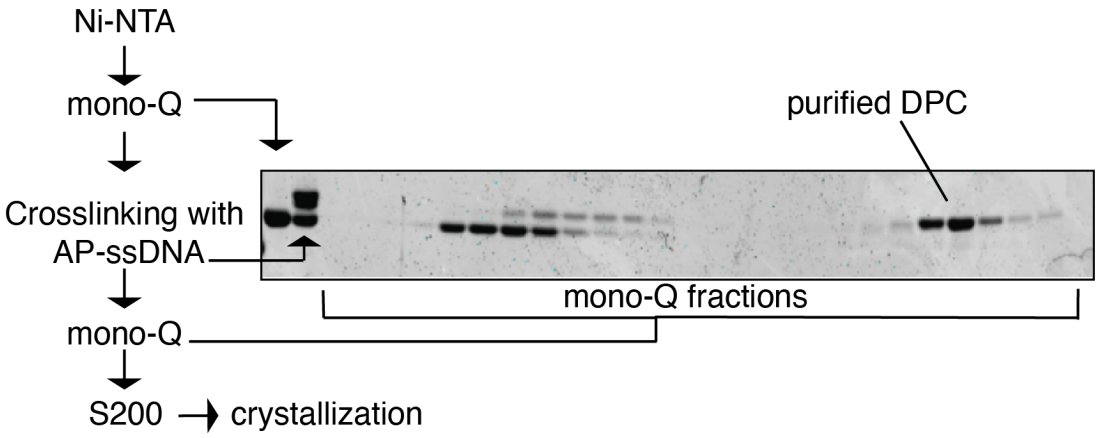

Figure 3.8 SRAP-DPC purification. Schematic of the purification for HMCES-SRAP domain covalently linked to a 10-mer AP site ssDNA. Coomassie-stained SDS-PAGE gel of fractions from the second mono-Q step is shown.

even through gel filtration. Thus, the HMCES-DPC is unlikely to be reversible in physiological conditions, and its resolution almost certainly requires proteolysis followed by either an unidentified enzymatic action to remove the linkage or nucleotide excision.

Discussion

HMCES is a Misnomer

HMCES derives its name from a proteomics screen identifying proteins that bind 5hmC and its oxidized derivates (144). The duplex DNA bait used in this screen contained eight modified cytosine bases and therefore was not B-form DNA but a distorted DNA structure (144). In fact, other DNA repair proteins such as SMARCAL1 (SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin) and RECQL (RecQ like helicase) with no known role in recognition of 5mC or its oxidized derivates were also identified in this study (144). This supports the idea that the lesion recognized by HMCES was not purely 5mC or 5hmC. Here, I show that HMCES does not have a preference for 5mC or the oxidized 5hmC in ssDNA or dsDNA context. DNA cytosine methylation is a well-characterized epigenetic modification with essential roles in regulating gene expression, maintenance of epigenetic memory, and cell fate. In mammals, these methylation events predominantly occur at CpG dinucleotides (159). Of the ~20 genes that were differentially regulated with HMCES loss, most were associated with the p53 damage response (93). In addition, there was no difference in the global levels of 5mC or 5hmC in HMCESΔ cells vs. wild type (93). Therefore, loss of HMCES contributes to an activation of the DNA damage response and not necessarily to global changes in gene regulation. The name HMCES is a misnomer.

HMCES was also described as an autoproteolytic endonuclease. However, we have not observed any protease activity in any experiment. Furthermore, my data show that digesting the SRAP-DPC restores migrating ssDNA to that of full-length ssDNA (Figure 3.4C). An enzyme with dual enzymatic activity seems highly unusual. One key difference between my work versus Kweon and colleagues is that my purification of SRAP proteins involved multiple purification steps versus one affinity purification step (149). The nuclease observed by Kweon and colleagues may be due to a contaminating protein.

Invariant N-terminal Cysteine is Critical for SRAP Function

The data suggest that the invariant cysteine at amino acid position two is critical for SRAP-DPC formation. Notably, of the 3508 SRAP domain sequences listed in protein databases, 3472 (99%) have the putative catalytic cysteine at amino acid position two. Methionine followed by a second amino acid cysteine is not a common occurrence. Of the 20,394 proteins in the UniProt human protein database, only 170 protein (0.83%) have a methionine followed by a second amino acid cysteine, which is less than the expected ~650. When comparing all known proteins on ScanProsite, 2285/559077 (0.41%) have a cytosine as the second amino acid. The mass spectrometry data indicate that the first amino acid methionine is removed from SRAP (not shown). Furthermore, Edman degradation analysis indicate that the methionine is removed making the cysteine the N-terminus. Methionine removal could be due to the predicted autopeptidase activity of SRAP (148). The SRAP predicted autopeptidase activity was proposed to act as a regulatory switch akin to autopeptidase activity found in LexA (148). However, N-terminal methionine excision is also catalyzed by methionyl-aminopeptidases (MAP) in bacteria, yeast, and humans (160). Excision is influenced by the length of the penultimate amino acid with smaller length amino acids (G, A, P, S, T, V, C), producing more efficient removal of N-methionine

(161). The N-terminal amino acid for SRAP is a cysteine, but it is unclear whether the N-terminal methionine is removed via SRAP or MAP.

Interestingly, N-terminally tagged HMCES cannot rescue multiple HMCESΔ cellular phenotypes (unpublished data K. Mohni). In addition, IR sensitivity could be rescued by expression of WT protein but not catalytic (C2A), DNA binding (R98E), or PIP (WL/AA) box mutants (93). HMCESΔ cells complemented with C2A mutant have a defect in DPC formation *in vivo* (93). In vitro, N-terminally tagged SRAP proteins are defective in forming SRAP-DPC. Even a four amino acid N-terminal scar reduces SRAP-DPC formation by ~65% (93). Therefore, having cysteine at the N-terminus is critical for SRAP biology and SRAP-DPC formation.

The N-terminus of HMCES sits within its catalytic pocket (Figure 3.1). The requirement for DTT and acidic pH for efficient DPC formation suggests that a reducing and acidic environment promotes the chemical formation of HMCES-DPC. DTT is often used to reduce disulfide bonds in proteins; thus, preventing intramolecular or intermolecular bonds. During the initial purifications of SRAP proteins, we noticed two distinct molecular weight peaks on size exclusion chromatography. These peaks coalesced in the presence of DTT. This suggests that HMCES forms higher-order protein complexes. The increase in DPC formation under acidic conditions is surprising given that the pKa of the -SH of cytosine ~8.6. At acidic pH, the -SH group should be protonated while at basic pH, the thiol group would be deprotonated. The pKa of N-terminal α-amino groups can vary significantly ~6.8-9.1(162). The N-terminal cysteine of SRAP sits deep within a catalytic pocket. Therefore, the molecular environment surrounding the cysteine may change the local pKa of this N-terminal cysteine to promote DPC formation. The chemical mechanism of how Cys2 functions in DPC formation is discussed in Chapter IV.

SRAP-DPC is Remarkably Stable

The SRAP-DPC is stable for days at different temperatures and pH levels. Anecdotally, SRAP-DPC left at 4 °C for months is still intact when observed by Coomassie blue staining. Coupled with this remarkable stability in vitro, in cells, the HMCES-DPC is easily detected and long-lived. SRAP-DPC resolution is on the order of hours (93). Covalent DPCs are often toxic to cells. Therefore, the stability of the HMCES-DPC crosslink should be problematic for cell viability. The mechanism for efficient and timely removal of the HMCES-DPC is unknown. HMCES-DPC removal seems to be mediated by a ubiquitin-protease dependent mechanism, but it is also possible that an unknown enzyme removes the SRAP-DPC directly (93). Other DPCs described in the literature, such as trapped topoisomerase cleavage complexes, are removed by distinct enzymes (TDP1, TDP2)(163). Furthermore, metalloproteases, such as SPRTN, have been shown to remove covalent DNA-protein crosslinks directly (164–166). The SRAP peptide-DNA adduct is also stable and resistant to cleavage by the BER enzyme, APE1. Thus, BER is not the next step in HMCES-DPC repair. Nucleotide excisions repair or recombination DNA repair are other possibilities for the next step in peptide-DNA adduct resolution. The potential mechanisms for HMCES-DPC resolution will be discussed in more detail in Chapter V.

CHAPTER IV

SRAP Forms a Stable Thiazolidine Linkage with AP-ssDNA*

Introduction

HMCES shields abasic sites from AP endonucleases preventing double-strand breaks and promotes a less mutagenic pathway for ssDNA AP site repair. The HMCES-DPC is found in cells, increases in abundance in response to AP-site inducing agents, and is resolved over time by an unknown mechanism that is at least partially proteasome-dependent. Furthermore, HMCES deficient cells have an increase in AP sites and a delay in AP site resolution. Both the HMCES and YedK SRAP domains bind DNA with a strong preference for ssDNA over dsDNA. More importantly, both proteins efficiently form a DNA protein crosslink (DPC) to an AP site containing ssDNA. This DPC formation requires DNA binding residues as well as a cysteine that is always at the second amino acid in SRAP proteins. Therefore, HMCES-DPC formation is critical for cellular response to AP sites.

The chemical mechanism of SRAP-DPC formation and how the SRAP domain detects AP sites were unknown. In Chapter III, I discussed how the N-terminal cysteine is critical for DPC formation and HMCES biology. Interestingly, N-terminal cysteines are used in native chemical ligation chemistry to generate site-specific labels and peptide-peptide linkages. In the presence of aldehydes, N-terminal cysteines can readily form a

thiazolidine moiety (167–169), and this reaction occurs efficiently under acidic conditions (170). We hypothesized that SRAP forms a stable thiazolidine moiety. To understand the chemical mechanism of SRAP-DPC formation, I used a structural biology approach in collaboration with Katherine Amidon in the Brandt Eichman lab.

In this chapter, I will show that the SRAP-DPC is formed by a thiazolidine linkage between the N-terminal cysteine of the SRAP and the aldehyde AP site. The crystal structure of SRAP explains why the crosslink shields AP sites from endonuclease. Furthermore, the SRAP-DPC structure explains the specificity for ssDNA and gives insight into how SRAP is uniquely poised to generate a crosslink with an AP site containing ssDNA. HMCES-SRAP can accommodate a 3' truncated ssDNA-dsDNA junction as might be expected when a replicative polymerase stalls at an AP site. Evolutionarily conserved residues act in the formation and protection of the thiazolidine linkage and explain the substrate preferences of SRAP. These evolutionarily conserved features highlight the importance of the SRAP domain.

Results

YedK DPC crystal structure

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 containing an AP site (Table 4.1). The entire DNA ligand is visible in the electron density (Figure 4.1A). A core β-sheet forms an extended, positively charged channel that cradles the ssDNA (Figure 4.1B-C). The phosphoribosyl backbone of the DNA hydrogen

bonds to basic residues within the positively charged channel, and the solvent exposed hydrogen bonding-edges of the nucleobases point away from the surface of the protein (Figure 4.1B-E). Thus, the recognition of the AP site would not be sequence-dependent. Furthermore, residues lining the DNA binding channel are the most highly conserved among SRAP domains suggesting conservation of DNA binding modality (Figure 4.1D; Figure 4.2A-B). In fact, both YedK and HMCES have similar preferences to binding ssDNA and mutations of conserved amino acids within the positive channel abrogate DNA binding (93). The protein does not undergo any appreciable conformational change upon binding DNA, with an r.m.s.d of 1.16 Å for all atoms between unbound and DPC forms of YedK (Figure 4.2C). 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 (Figure 4.1). This sharp distortion would preclude pairing of a complementary DNA strand in the vicinity of the AP site, and may explain why SRAP disfavors binding to dsDNA (93). The AP site sits within the putative catalytic pocket positioned directly above Cys2 (Figure 4.1B-D). There is no electron density for a N-terminal methionine proximal to the cysteine. In addition, the calculated molecular weight via mass spectrometry of purified SRAP corresponds to the expected SRAP protein without the methionine (data not shown). Therefore, the N-terminus of the protein is a cysteine as the methionine is likely removed by aminopeptidases.

Table 4.1 Data collection and refinement statistics.

	YedK/AP-DNA covalent DPC (PDB 6NUA)	**YedK C3-spacer-DNA non-covalent complex** (PDB 6NUH)
Data Collection[a]		
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, b, c (Å)	61.26, 41.89, 81.42	47.54, 44.13, 55.09
α, β,γ (°)	90.00, 95.79, 90.00	90.00, 102.34, 90.00
Resolution (Å)	50.00-1.64 (1.67-1.64)[b]	100.00-1.60 (1.66-1.60)
R_{sym}	0.098 (0.500)	0.075 (0.397)
R_{meas}	0.110 (0.595)	0.086 (0.455)
Avg. I/σI	14.8 (1.0)	21.3 (2.6)
$CC_{1/2}$	0.989 (0.823)	0.990 (0.869)
Completeness (%)	97.4 (95.2)	98.5 (91.1)
Redundancy	4.4 (2.9)	4.1 (4.0)
Refinement		
Resolution (Å)	40.50-1.64 (1.67-1.64)	39.60-1.59 (1.65-1.59)
No. Reflections	49,681 (2,331)	29,612 (2,391)
R_{work}/R_{free}	0.171/ 0.22	0.143/ 0.177
No. Atoms		

Protein	3,627	1,802
DNA	268	131
Bis-Tris	-	14
Water	280	210
B factors		
Protein	26.0	17.1
DNA	28.7	64.3
Bis-Tris	-	38.3
Water	29.3	24.4
R.m.s deviations		
Bond lengths (Å)	0.010	0.008
Bond angles (°)	1.035	0.959

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

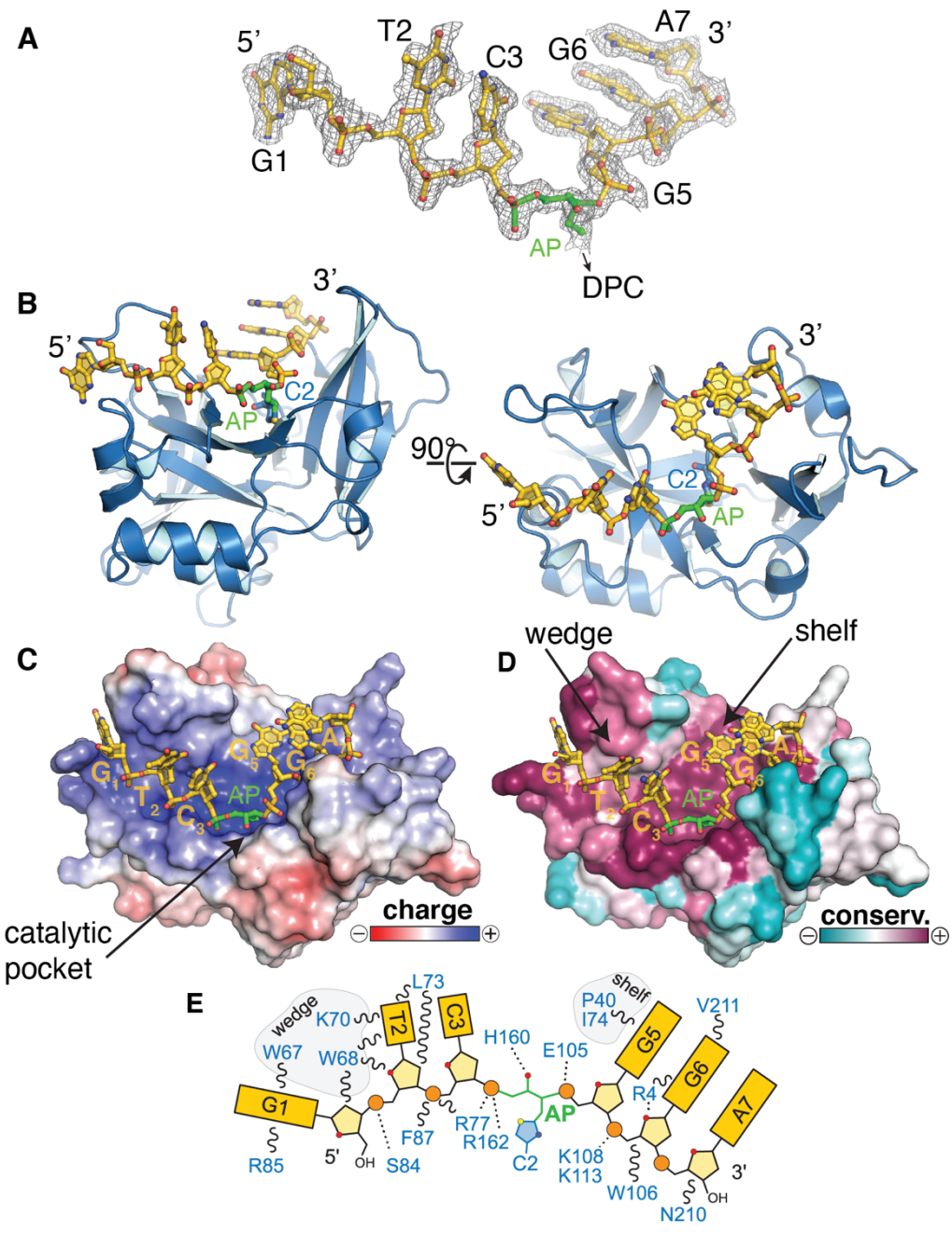

Figure 4.1: YedK DPC crystal structure. (A) DNA ligand (7-mer) with AP site at position 4 surrounded by electron density map. (B) Orthogonal views of *E. coli* YedK (blue) crosslinked to AP DNA (gold). (C, D) YedK solvent-accessible surface colored by electrostatic potential from -5 to +5 k_BT/e_C. (C) and sequence conservation from 158 unique SRAP orthologs (D). (E) Schematic of protein-DNA interactions.

Thiazolidine Linkage is Surrounded by Conserved Residues

A closer look at the nature of the crosslink reveals that the AP site is covalently crosslinked to the Cys2 of SRAP. The electron density clearly shows that the AP site in the ring-opened form, with continuous density between C3′ and the Cys2 side chain (Figure 4.2A). Importantly, 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 five-membered thiazolidine ring (Figure 4.2A).

Many of the residues that interact with the DNA ligand and the AP site are well conserved among the SRAP domains suggesting that ssDNA binding and AP site recognition is a conserved process (Figure 4.3A, B). The invariant Cys2 belongs to a cluster of three conserved residues that includes Glu105 and His160 implicated in SRAP function (148, 149). These and several other evolutionarily conserved residues stabilize the DNA and protein sides of the thiazolidine linkage (Figure 4.2B, C). The AP site is stabilized by His160, which forms a hydrogen bond with the O4′ hydroxyl group (Figure 4.2B). Similarly, Arg77 and Arg162, previously shown to be essential for DNA binding (93) and Thr149, interact with the AP site 5′-phosphate (Figure 4.2B). The Glu105 side chain fluctuates between two conformations at the crosslink (Figure 4.2B). One conformer (cyan) places one carboxylate oxygen 3.5 Å from the thiazolidine C1′. The second conformer (blue) is within hydrogen bonding distance to the phosphate 3′ to the AP site, strongly implying that the carboxylate is protonated to avoid electrostatic repulsion with the DNA. The second conformer points back toward the core of the protein and sits further away from the thiazolidine ring. On the protein side of the crosslink, the carboxamide side

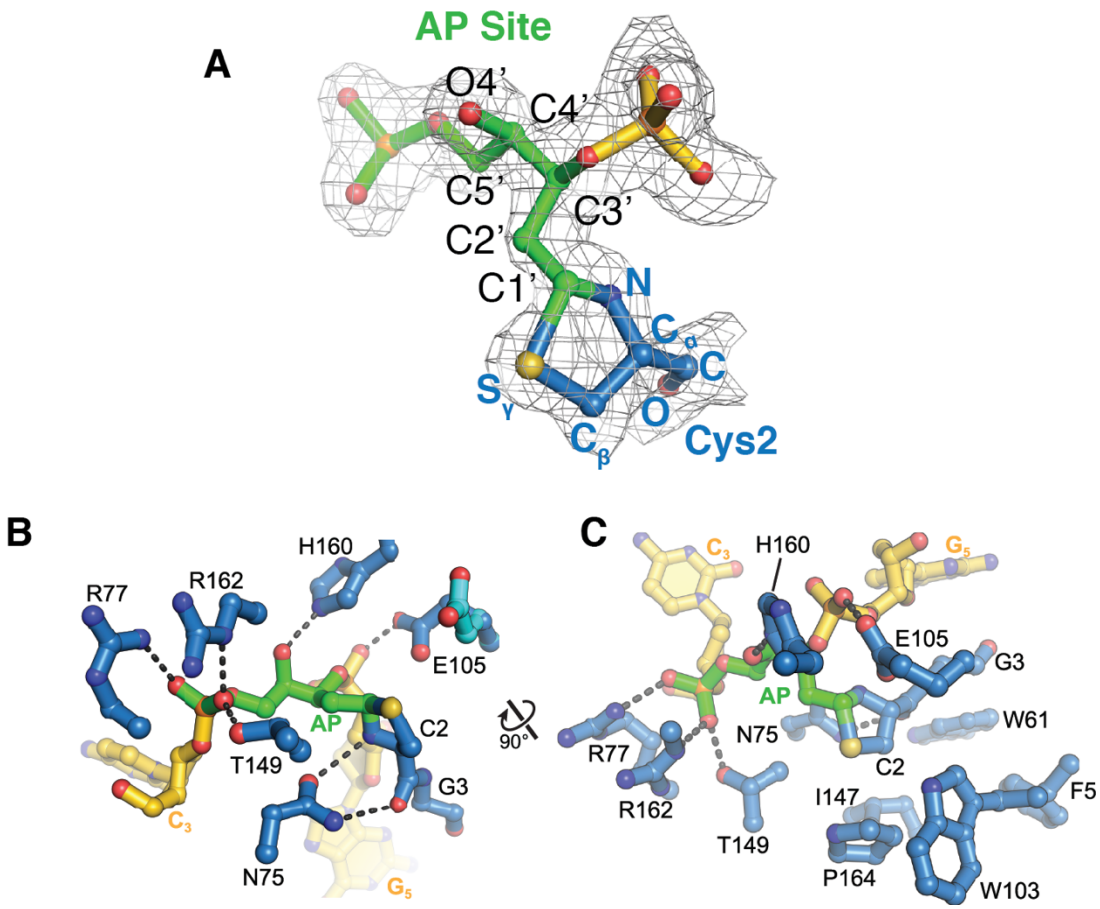

Figure 4.2: Thiazolidine linkage formed between SRAP and AP site is surrounded by conserved residues. (A) The DPC between the AP site (green) and Cys2 (blue) is superimposed against $2F_0$-F_c composite annealed omit electron density, contoured to 1$\boldsymbol{\sigma}$. (B) Residues contacting the DPC (DNA, gold; AP site, green; protein, blue). The alternate Glu105 conformers in cyan pointing away from the protein core and blue pointing towards the protein core. (C) Orthogonal view showing hydrophobic residues cradling Cys2. Only one Glu105 conformer is shown for clarity.

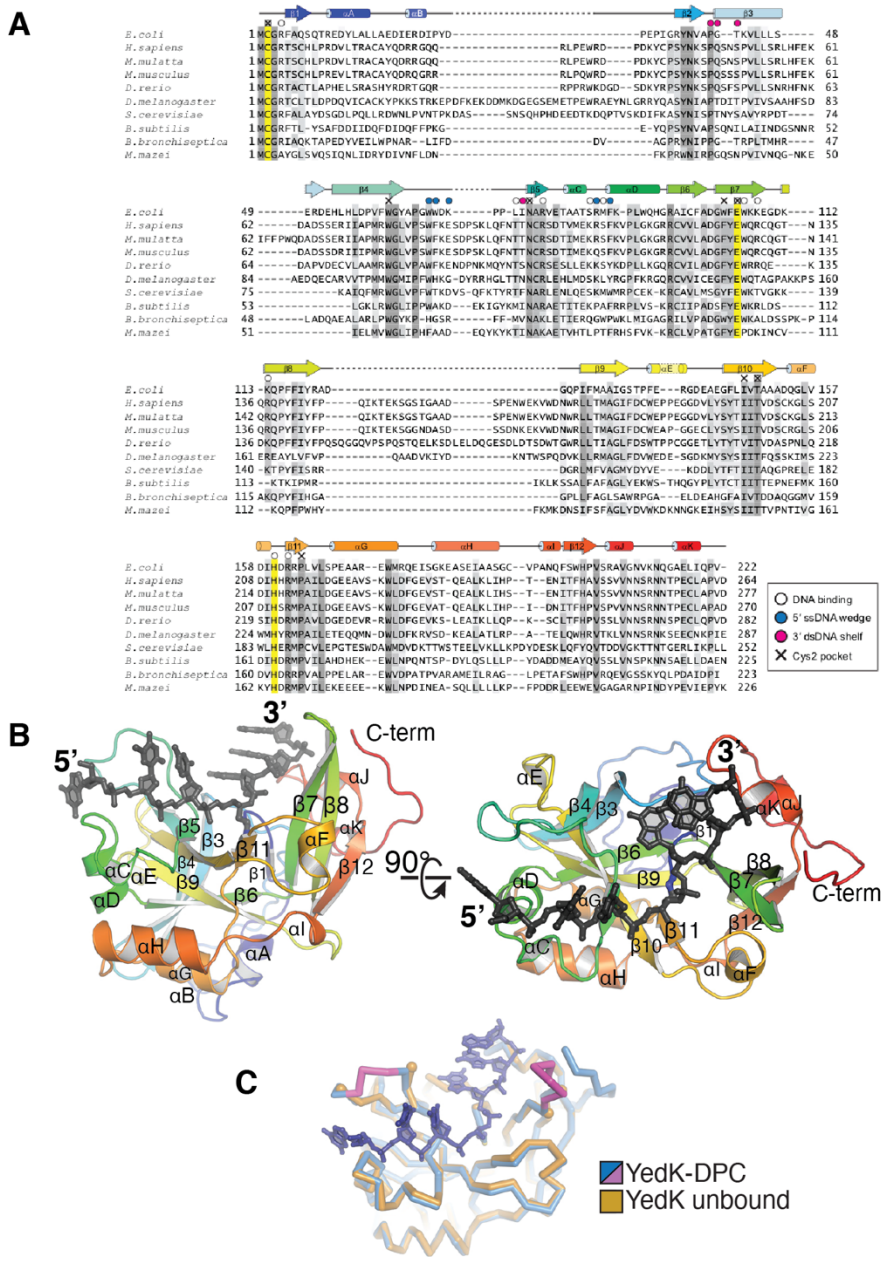

Figure 4.3: Conservation of SRAP domain. (A) Structure-based sequence alignment between E. coli YedK and human HMCES-SRAP domain, together with a sequence alignment of SRAP domains from 8 additional species. Secondary structure for YedK DPC structure is shown above alignment. (B) Orthogonal views of YedK DPC structure colored rainbow from N-(blue) to C-terminus (red). (C) Superposition of YedK DPC (blue/magenta) and free YedK orange. Loops in YedK DPC that are disordered in the free proteins are colored magenta. Proteins are shown as a Cα-backbone trace.

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 (Figure 4.2B). In addition, there are several highly conserved residues (W103, P164, I147, W61, F5) that create a hydrophobic pocket to cradle Cys2 from underneath (Figure 4.2C). Thus, the SRAP structure guides the AP site into a specific, solvent inaccessible environment suited for thiazolidine formation and protected from AP endonuclease cleavage.

SRAP-DPC Formed via a Schiff-Base Intermediate

The cysteine thiols react with aldehyde to form thiazolidine product under acidic conditions. This reaction proceeds via a reversible Schiff-base (170), (171). Therefore, we hypothesized that the thiazolidine 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 (Figure 4.4A).

We mutated the catalytic cysteine (C2A, C2S), other conserved catalytic residues (E105A, H150A), and N75 that hydrogen bonds with the backbone amide nitrogen and the carbonyl oxygen of Cys2. YedK DPC formation is abrogated by removal of the thiol in a C2A mutant (93) (Figure 4.4B). We predicted that the C2S mutant would potentially form an oxazolidine ring. However, oxazolidines are not as stable as thiazolidines (172, 173). Therefore, the C2S mutant does not form a stable DPC (Figure 4.4B). Other mutant E105A, H160A, and N75A proteins also have reduce crosslinking efficiency when compared to WT (Figure 4.4B).

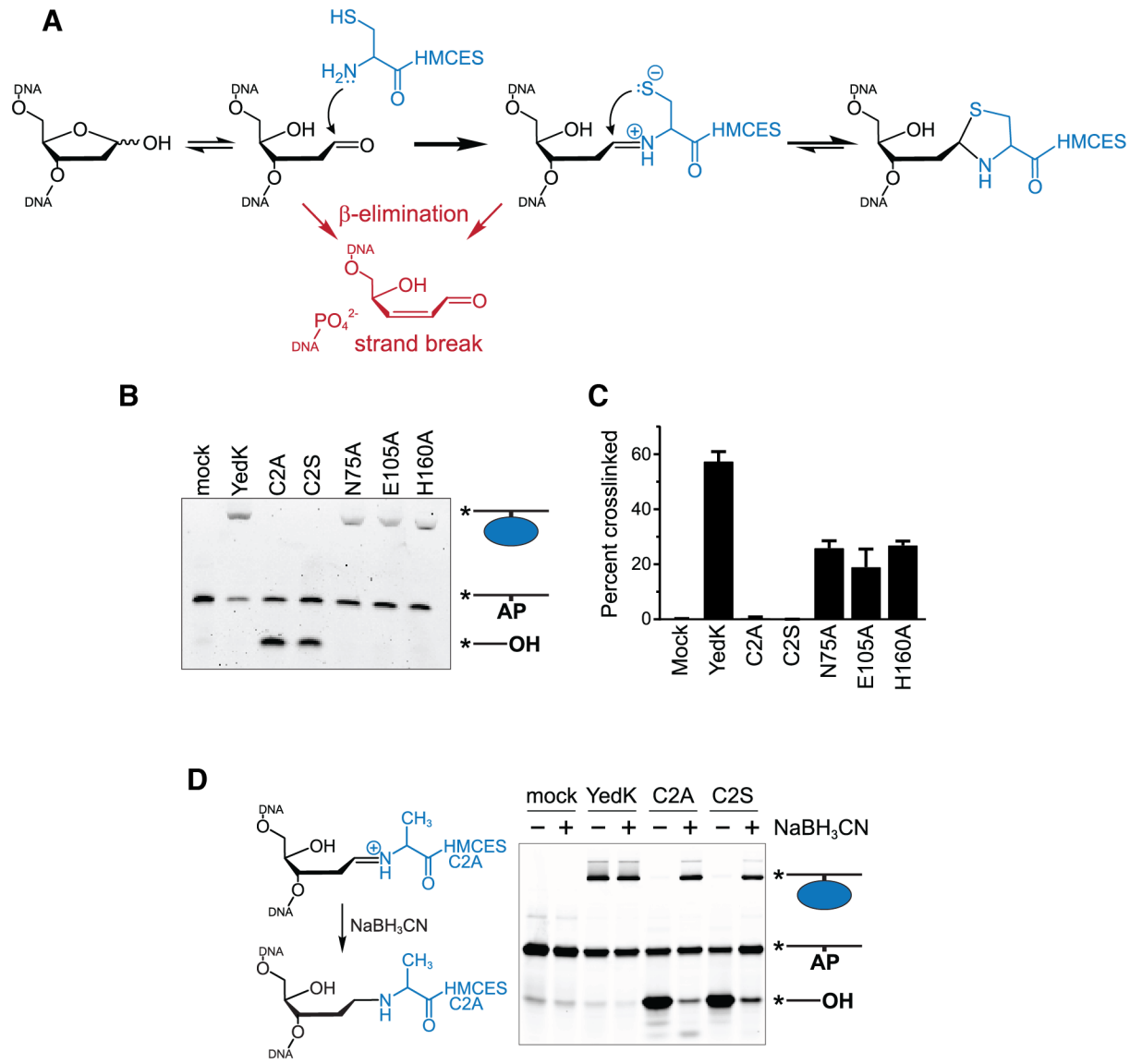

Figure 4.4: SRAP-DPC is formed via Schiff-base intermediate. (A) Proposed chemical mechanism of cross-linking reaction with competing lyase reaction in red. (B) Denaturing EMSA showing crosslinking and lyase activity of YedK mutant. Bands were visualized by FAM fluorescence. (C) Crosslinking efficiencies of YedK mutants (mean +/- s.d., n=3 independent measurements). (D) $NaBH_3CN$ was added to crosslinking reaction to trap Schiff-base intermediates of YedK C2A and C2S mutants. The NaBH3CN-reduced Schiff-base is resistant to β-elimination. Bands were visualized by FAM fluorescence. Figures B-D were produced by K. Amidon.

Studies on the reaction of cysteine and aldehydes show that the equilibrium between Schiff-base and thiazolidine greatly favors the latter (170, 174). With wild-type protein, 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. (Figure 4.4B-D) (13, 175). In contrast, both the C2A and C2S YedK mutants exhibited DNA lyase activity when incubated with ssDNA containing an AP site (Figure 4.4B, D). This lyase activity was significantly reduced by performing the crosslinking reaction in the presence of sodium cyanoborohydride ($NaBH_3CN$). The Schiff-base intermediate is reductively trapped in the presence of $NaBH_3CN$. (Figure 4.4D) (176). These results 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.

SRAP Accommodates dsDNA 3' to the AP Site

Both the DPC and non-covalent yedK-ssDNA 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 (Figure 4.5A, B). Trp68 wedges the nucleobases of G1 and T2 apart, and G1 is stacked between Trp67 and Arg85 (Figure 4.5B). Such a distortion would prevent duplex formation with DNA 5′ to the AP site.

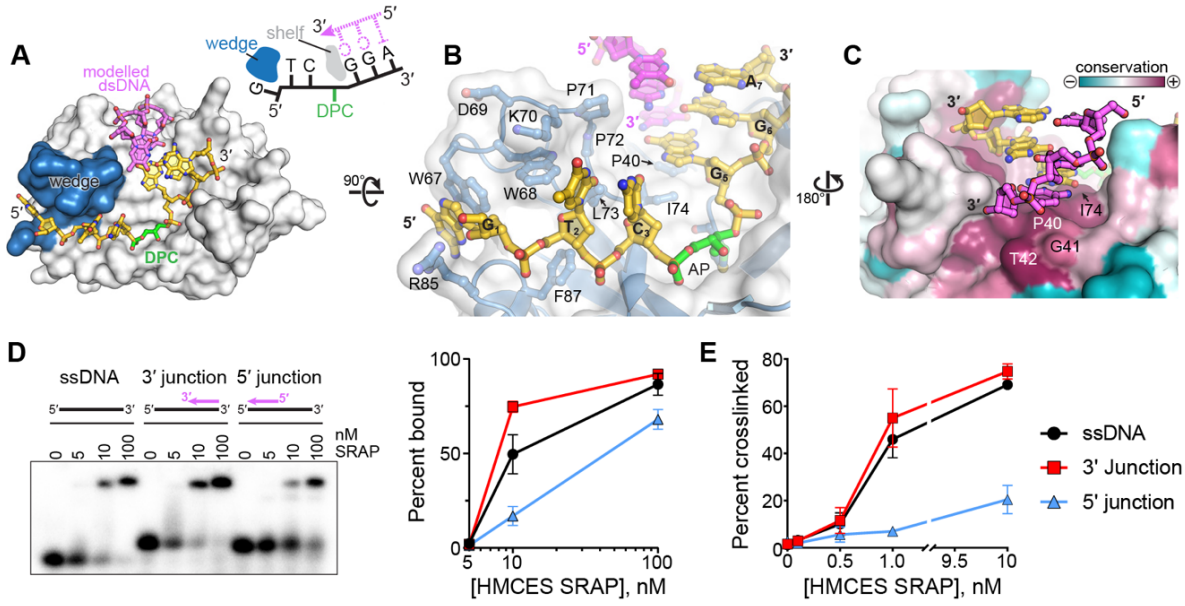

Figure 4.5: SRAP can accommodate dsDNA 3' to the AP site. (A) Model of YedK DPC with a 3' junction at AP site. The modeled complementary DNA strand is pink. The wedge domain blocking dsDNA access 5' to the AP site is blue. (B) Wedge-DNA interactions 5' to the AP site. (C) Sequence conservation of the 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 the mean +/- s.d., n=3 independent measurements. (E) Crosslinking efficiency of human HMCES-SRAP domain to the indicated DNA ligands (mean +/- s.d., n=3).

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 (Figure 4.5B). The residue adjacent to the AP site (guanine G5) stacks against Pro40 and Ile74 on the surface of the protein (Figure 4.4B, C). The exposure of the hydrogen bonding faces of the G5, G6, and A7 nucleobases 3′ to the AP site would allow for base pairing of a second strand up to the 3′-side of the AP site. Modeling shows that a complementary strand fits against the protein surface with no steric clashes (Figure 4.5A-C). The 3′-end of the modeled strand stacks against Gly41 and Thr42, which form a highly conserved “shelf” together with Pro40 and Ile74 (Figure 4.5C). 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 (Figure 4.5D, E). 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.

Non-Covalent SRAP DNA Complex

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 (Figure 4.6). The protein in the non-covalent complex is virtually identical to that of the DPC, except for modest

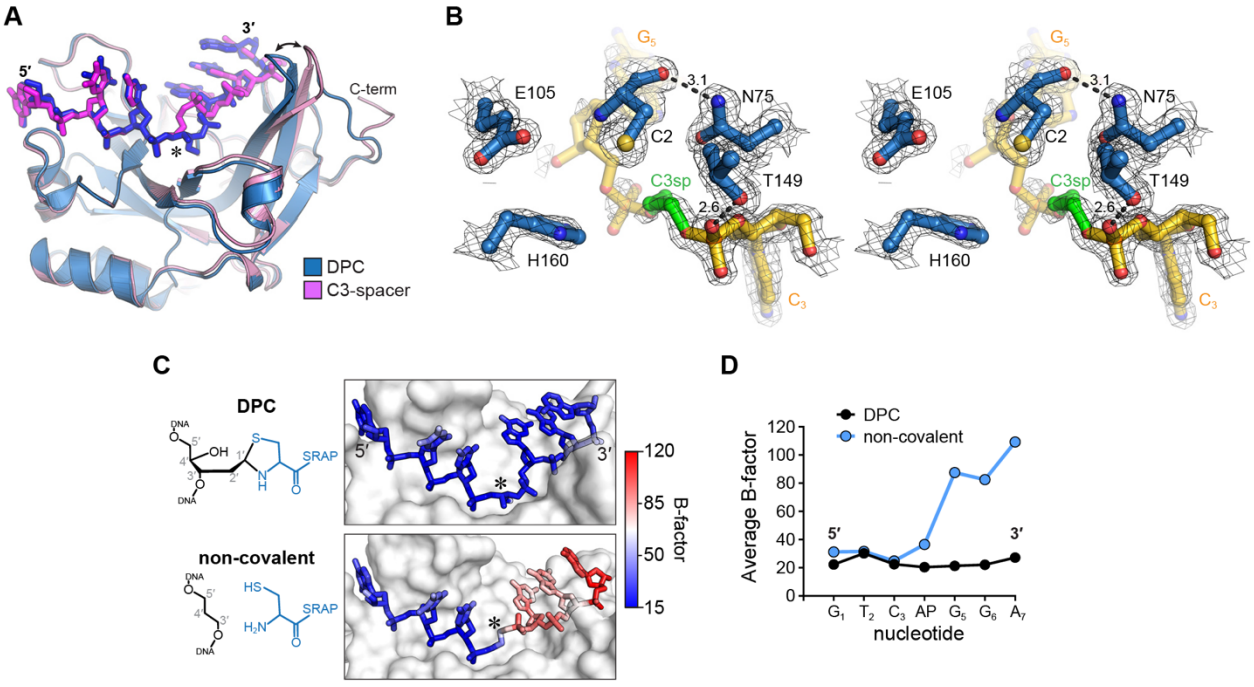

Figure 4.6: Structural details of the non-covalent SRAP-DNA complex. (A) Superposition of YedK covalently crosslinked to AP-DNA (blue) and non-covalently bound to C3-spacer DNA (magenta). The abasic site is marked with an asterisk. The double-headed arrow shows the most significant difference between the two structures—movement of the 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-spac-er-DNA structure, superimposed with 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.

repositioning of a β-hairpin (β7-β8) that stabilizes the backbone of the DNA 3′ to the AP site (Figure 4.6A). In this 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 (Figure 4.6B-D). The destabilized 3′-DNA end resulted in a crystal packing difference between the two complexes.

Discussion

To better understand the unusual mechanism of SRAP mediated DNA repair, we examined the nature of the SRAP DNA-protein crosslink using structural biology. The SRAP protein forms a stable thiazolidine linkage using highly conserved residues, and this linkage is safeguarded by a hydrophobic environment.

Crystal Structure Supports SRAP Biology

The SRAP-DPC structure contributes to the overall knowledge of SRAP biology. Critical residues involved in formation and stabilization of the thiazolidine linkage are conserved through more than 150+ plus SRAP proteins (Figure 4.2ab), suggesting that SRAP domains are poised for the protection of ssDNA AP sites. The ssDNA is kinked at a 90° angle on top of the catalytic cysteine suggesting that SRAP guides the AP site towards the active site. Moreover, the SRAP structure corroborates the preference for ssDNA vs dsDNA. The highly conserved wedge motif occludes dsDNA 5’ to the AP site by interrupting nucleobase stacking. Yet, dsDNA can be modelled 3’ to the AP site without

steric hinderance (Figure 4.5). We show that SRAP has a preference for 3'ssDNA/dsDNA junction structure. This is the exact structure expected when a replicative polymerase stalls at an AP site. Crystal structures from two other groups confirm that HMCES and yedK can accommodate duplex DNA 3' to the AP site (177, 178). Therefore, SRAP preference for the 3' ssDNA/dsDNA structure is likely a biologically relevant interaction and not a crystallographic artefact. More importantly, the SRAP-DPC structures support a model where HMCES acts at a stalled replication fork (93).

Function of Predicted Catalytic Residues in Thiazolidine Formation

SRAP contains a putative catalytic triad of the invariant cysteine and highly conserved histidine and glutamate residues. Several groups have speculated that SRAP has functions as a thiol peptidase (148, 149). The invariant cysteine is absolutely required for DPC formation (93). In my hands, mutating other conserved residues in the predicted catalytic triad (H160 and G105) diminished DPC formation. However, it is unclear if these three residues (Cys2, His160, Glu105) participate in an acid-base-nucleophile triad commonly seen with cysteine proteases (179). Typically, alterations of any component of a catalytic triad results in a decrease in enzyme activity. However, we observed a decrease in crosslinking efficiency for the H160A mutant protein while Wang et. al observed an increase crosslinking efficiency (178). These differences may be due changes pH of buffers used, 6.0 vs 8.0 respectively. The orientation of the base and nucleophile in a catalytic triad determines the mode of catalysis. In cysteine proteases, the thiol group is deprotonated by an adjacent basic chain amino acid, usually histidine (179). This histidine-activated cysteine functions as a nucleophile. In the SRAP alone and

SRAP-ssDNA structures the highly conserved histidine is not in hydrogen bonding distance to thiol and would not act as a general base catalyst. Instead, H160 hydrogen bonds with O4' and could potentially participate in ring-opening of the AP site. The Glu105 may participate general base to deprotonate both the Cys2 amino group and side chain. A similar mechanism is used by bifunctional glycosylases where the amino group of a lysine side chain is deprotonated by aspartate residue. Subsequent nucleophilic attack on the N-glyosidic bond forms a Schiff-base intermediate and cleavage of the DNA backbone 3' to the lesion (180). The C2A mutant has similar DNA binding activity compared to WT (93). Of note, we did not test if mutating the conserved His160, Glu105, Asn75 residues changed DNA binding affinity. So, the decrease we observed in DPC formation may be due to a decrease in DNA binding. Interestingly, while we observed AP lyase activity with the yedK C2A mutants similar lyase activity was not observed with HMCES-SRAP. This may be due to a difference in experimental conditions. Both SRAP and bifunctional glycosylases use an amino group nucleophile to generate a Schiff-base intermediate. Therefore, it is unlikely that SRAP functions as a thiol peptidase. Nevertheless, further understanding the potential catalytic activity and the kinetics of SRAP-DPC formation will be important for future studies.

Thiazolidine Formation via A Schiff-Base Intermediate

Here, we show thiazolidine formation proceeds via a Schiff-base intermediate. One consideration is the observed thiazolidine ring is a crystallographic artifact. However, the electron density is consistent with a thiazolidine ring. Furthermore, other SRAP-DPC DNA structures support a thiazolidine linkage between Cys2 of SRAP and AP site containing

DNA (177, 178). Because these observed structures with HMCES and yedK proteins have different crystal packing patterns, it unlikely that this linkage is not a thiazolidine ring. Finally, Katherine Amidon determined a crystal structure of $NaBH_3CN$ reduced C2A mutant with AP DNA and the reduced Schiff-base intermediate is visible (unpublished data).

CHAPTER V

Summary and Future Directions

Summary

AP sites are one of the most common type of DNA damage. Thousands of AP sites form each day in our cells by spontaneous depurination and depyrimidination, base damage via exogenous and endogenous agents, and cytosine deamination. Unrepaired these AP sites have deleterious consequences and threaten genome integrity. AP sites present as a block to transcription and DNA replication and can give rise to double-strand breaks, intra-strand crosslinks, and DNA protein crosslinks. Although AP sites occur more frequently in ssDNA, there was not a dedicated DNA repair pathway for AP-ssDNA lesions. Canonical mechanisms of AP site repair, such as BER and TLS, would lead to DSBs or mutagenic bypass. Here, we identified a new mechanism for AP site repair in the context of ssDNA. HMCES, an evolutionarily conserved protein, crosslinks to abasic sites in vitro and in cells. This crosslink formation is the first initial step in a novel DNA repair pathway that shields abasic sites from DSB formation and shunts repair to a less mutagenic pathway.

In chapter III, I described the preferred substrate for HMCES, an AP site containing ssDNA. I challenged the misidentified role for HMCES in reading 5mC and its oxidized derivates. I also showed that the HMCES-DPC is stable and resistant to cleavage by AP endonucleases. An invariant N-terminal cysteine is critical for DPC formation, and disrupting this cysteine through mutations (C2A, C2S) or the addition of N-terminal amino

acids causes DPC formation defects. In chapter IV, I described how several highly conserved residues help stabilize the ssDNA, guide the AP site to predicted catalytic pocket, and protects the thiazolidine linkage. Mainly, the SRAP-DPC structure and accompanying biochemical analysis further defined the chemical mechanism for DPC formation. SRAP-DPCs are formed via a Schiff-base intermediate, and the covalent thiazolidine ring is formed by the α-amino group of Cys2 and the ring-opened AP site. SRAP-DPC has a preference for AP-ssDNA and AP sites at 3' ssDNA/dsDNA junctions. This is the structure expected when a replicative polymerase stalls at an AP site. In this final chapter, I will discuss limitations of this new AP-ssDNA repair pathway, mainly is this pathway evolutionarily conserved and how is the HMCES-DPC resolved. Finally, I will discuss implications for HMCES in B-cell development.

Future Directions

Loss of function Phenotypes of E. coli YedK

SRAP domains are present in all domains of life. However, most of the work determining the biological function of SRAP has been with human or mouse HMCES (93, 149, 150). The HMCES-SRAP domain is similar in both amino acid sequence and structure to the *E. coli* ortholog, yedK. In bacteria, SRAP-containing proteins are near operons that contain SOS (the bacterial DNA damage response) genes such as TLS polymerases (148). Furthermore, a bioinformatic analysis found that *yedK* is functionally linked to *dinB* (181). In this work, I showed that YedK forms a stable DPC with AP-ssDNA and that many of the structural motifs and residues critical for DPC formation are well conserved among the thousands of SRAP proteins. Little is known about the biological

function of YedK. A significant unanswered question is whether SRAP function in AP site repair is evolutionarily conserved. Other questions that examine the role of yedK in the DNA damage response include: 1) Is *yedK* an SOS response gene? 2) Are *ΔyedK* cells sensitive to DNA damaging agents? 3) Are *ΔyedK* cells sensitive to an increase in abasic sites? 3) Does YedK form DPCs in bacterial cells similar to HMCES in human cells?

Activation of the SOS response occurs when DNA damage causes an accumulation of RecA coated ssDNA. The RecA-ssDNA filament triggers autoproteolytic cleavage of the LexA repressor and thus the expression of SOS responsive genes. Most SOS responsive genes have a ~20 bp consensus "SOS-box" or LexA binding site near their promoters (182, 183). Using statistical modeling Lewis et al. defined a heterology index (HI) for LexA binding sites (184). HI indicates the deviation of an SOS-box sequence from the consensus LexA binding sequence. DNA sequences that have a low HI bind LexA tightly while sequences that have a high HI are not expected to bind LexA. HI indices of ≥15 indicate that the LexA repressor does not bind the DNA sequence (185). Based on my preliminary data, the HI index for yedK may be high, but this will have to be determined using more specific techniques. YedK contains a predicted LexA binding sequence upstream of its promoter (186). In a large microarray analysis of genes differentially expressed in response to UV, *yedK* was moderately decreased (187). Additionally, my preliminary results show that in response to 0.1% MMS the relative expression of yedK, measured by qRT-PCR, is diminished when compare with untreated controls (Figure 5.1). The SOS response upregulates ~50 genes, yedK is not on this list (186). YedK may be

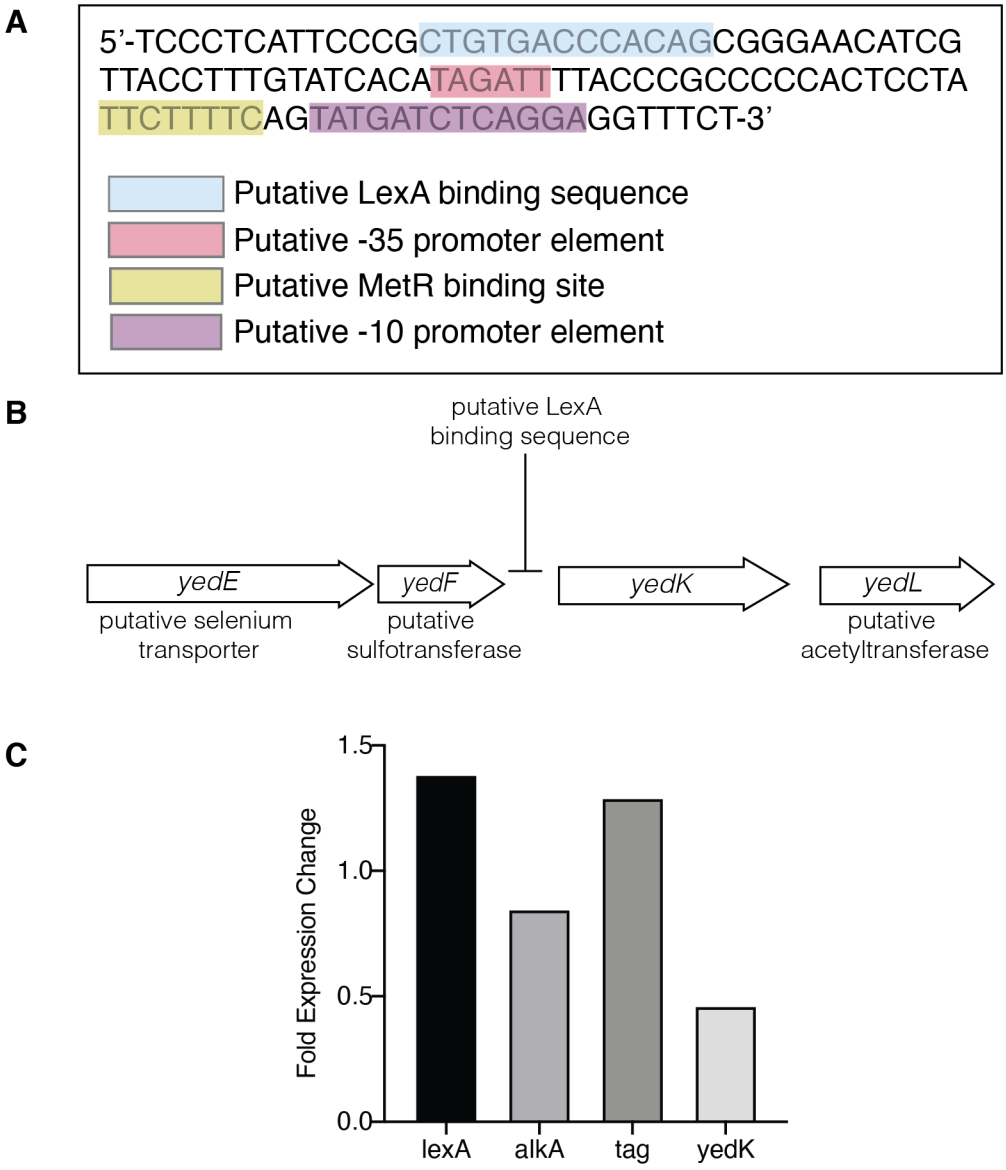

Figure 5.1: yedK is a potential SOS response gene. (A) Sequence 5' to yedK transcription start site (TSS) with predicted LexA binding site sequence, -35 and -10 promoter elements, and MetR binding site highlighted. (B) Schematic of gene cluster coding for *yedK*. (C) qRT-PCR results of DNA repair genes after treatment of *E. coli* with 0.1% w/v MMS for 1 hr (n=1).

negatively regulated by the SOS response or indirectly downregulated by another SOS response gene, but the functional interplay between the SOS response and *yedK* is unclear. Experiments identifying *yedK* expression in response to other DNA damaging agents and EMSAs using the predicted LexA-binding sequence and purified LexA will be required to fully assess if *yedK* regulated in an SOS-dependent manner. In addition, observing the phenotypes *ΔyedK* plus knockouts of known SOS response genes, in particular *dinB* (i.e., *ΔyedK ΔdinB)*, would be useful in identifying functional significance. Identification of a genetic or biochemical association between *yedK* and other genes in the SOS response is a worthwhile venture.

HMCES-deficient cells exhibit marked sensitivity to DNA damaging agents (MMS, UV, IR, and $KBrO_3$) that can generate AP sites. To determine if similar phenotypes are observed, *E. coli* K-12 *ΔyedK* strains were made (188). *E. coli yedk* knockouts are viable and do not exhibit any obvious growth defects in rich media. Furthermore, these knockouts are not sensitive to UV, MMS, or 4NQO (4-nitroquinoline 1-oxide) (Figure 5.2). When APOBEC3G, a cytosine deaminase that generates uracil in ssDNA, is expressed in *ΔyedK* cells there is a moderate decrease cell viability when compared to WT. However, this decrease in cell viability is not rescued with expression of WT or C2A YEDK (Figure 5.3). So far, it seems that the robust and often redundant DNA damage response in *E. coli* is unperturbed by *yedK* loss. *E. coli* use multiple DNA repair pathways (BER, recombination, and TLS) to repair chromosomal AP sites (133). Generating double knockouts of *yedK* plus known DNA response genes (*dinB*, *uvrA*, *umuCD*) may be useful in sensitizing *E. coli* to abasic sites. Finally, other tools for AP

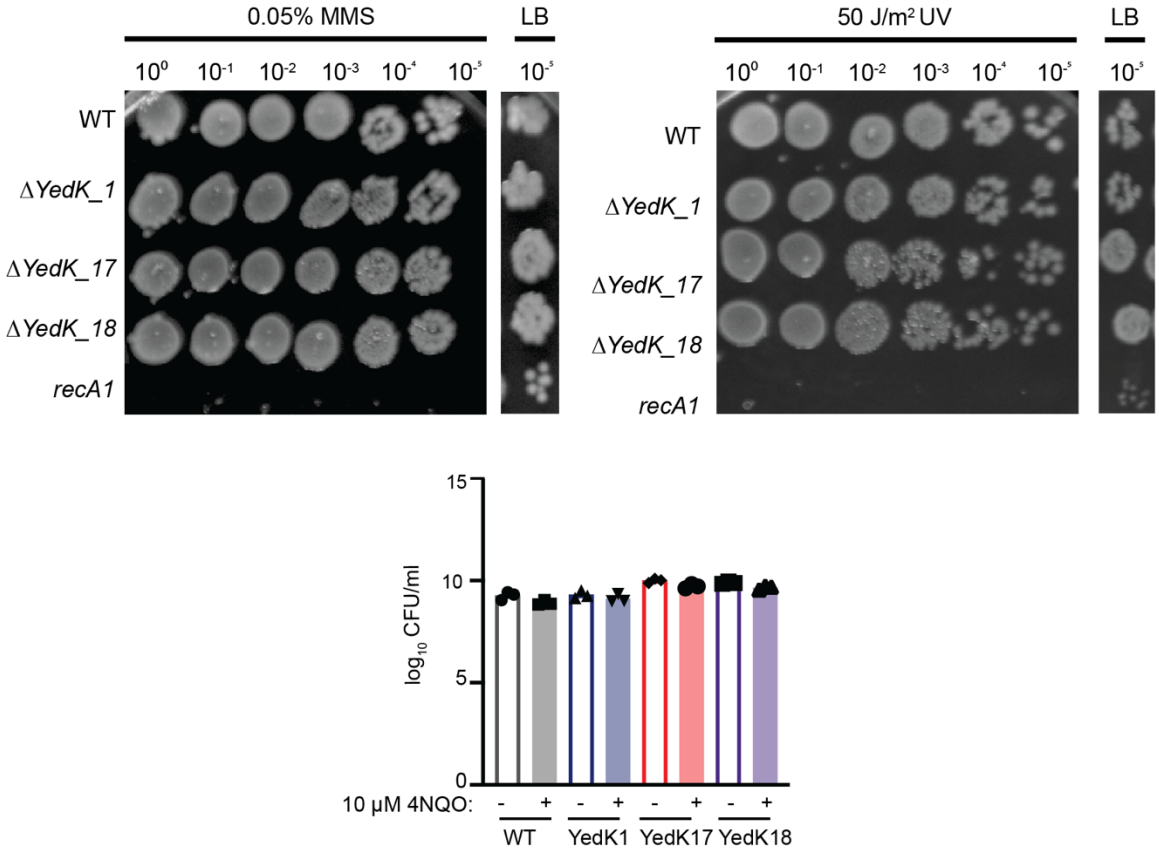

Figure 5.2: Sensitivity of ΔyedK *E. coli* mutants to DNA damaging agents (A, B) *E. coli* cells were grown to OD600≈0.4 in LB medium at 37 °C and 5 µL of serial dilutions (10^0 to 10^{-5}) were spotted on LB plates containing (A) 0.05% w/v MMS or (B) LB plates irradiated with UV 50 J/m^2. The 10^{-5} dilution from the untreated LB plate is shown for reference. Images are representative of three independent experiments. (C) *E. coli* cells were grown at 37 °C to OD600≈0.4 and 50 µL of cells were plated on LB only (-) or 10 uM 4NQO (+) plates. The 1, 17, 18 represent three different *ΔyedK* mutant clones and *recA1* is an inactivated form of RecA that is sensitive to DNA damage.

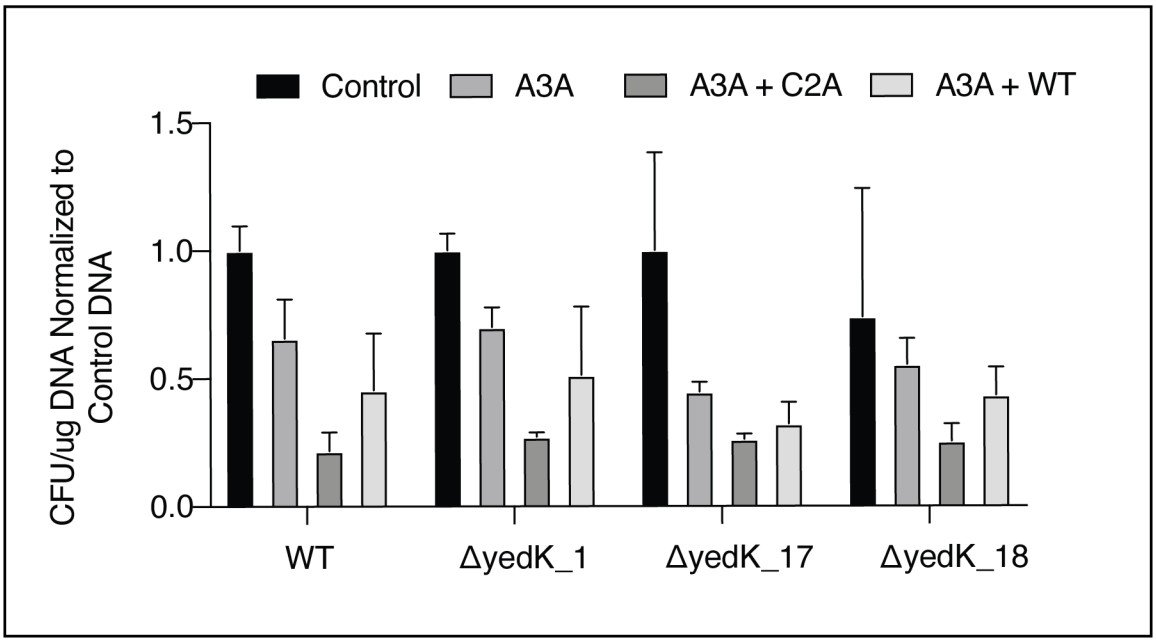

Figure 5.3: Sensitivity of ΔyedK *E. coli* to the expression of APOBEC3A. *E. coli* cells were transformed with 0.1 µg of pSF_OBXI, pSF_OBXI_A3A, pSF_OBXI_A3A_YEDKWT, or pSF_OBXI_A3A_YEDKC2A plasmids. Individual colonies were grown overnight at 37 °C, diluted to OD600≈ 0.01 and grown to OD600≈0.5 before plating 100 µL of 10^{-7} dilution on LB + kanamycin plates. Plates were then incubated at 37 °C overnight and total colonies were counted. CFU=colony forming units.

site generation such as engineered DNA glycosylases (TDG and CDG) should be considered (189).

HMCES-DPCs are readily formed *in vivo* and are removed at least in part via a proteasome-dependent process (93). This DPC formation is the first critical step in a DNA repair pathway that shield abasic sites from mutagenic repair. It is unknown if YedK forms DPCs in vivo. To date, the tools needed to address this question are not available. Mainly, an antibody to endogenous YedK and a rapid approach to DNA adducts recovery (RADAR) assay adapted to DPCs in bacterial cells (190). For human HMCES, expression of N-terminal and C-terminal tagged proteins did not rescue HMCESΔ phenotypes. Therefore, expressing a N-terminal or C-terminal tagged YedK may interfere with yedK function in bacterial cells. Instead, an internally tagged YedK protein could be used, but functionality would have to be determined. Recently, Aldred and colleagues developed a RADAR-based assay to isolate DPCs in bacterial cells. Optimizing this assay for YedK-DPCs analysis will be useful (191). More tools will be required to understand if SRAP-mediated AP site repair is conserved in bacteria.

Degradation of HMCES-DPC

My current model is that HMCES-DPC shields abasic sites at replication forks to promote error-free repair of abasic sites and to prevent cleavage by AP endonucleases. Mutations that abrogate HMCES crosslinking activity cause reduced cell proliferation, increased sensitivity to DNA damaging agents, and elevated mutation frequencies (93). Therefore, HMCES-DPC formation is critical for genome integrity.

Formation of DPCs, however, can stall or inhibit DNA replication and transcription and can be toxic to cells (192). It seems counterintuitive that an AP site shield, like HMCES, could also function as a “weapon” against genome integrity. However, AP sites are more cytotoxic to cells than uracils alone, and unrepaired AP sites can cause DNA strand breaks and mutations. The HMCES-DPC DNA repair intermediate could be more toxic than the initial AP site lesion. On the other hand, shielding AP sites for later repair, when the appropriate machinery is available, is a timely strategy to maintain genome integrity. The more curious question is how are HMCES-DPCs resolved (Figure 5.4).

In general, replication coupled DPCs are resolved by proteolysis or by direct crosslink hydrolysis (TDP1/TDP2)(163, 166, 193). After DNA damage, HMCES-DPCs are ubiquitinated, accumulate in cells, and are then resolved over time in a proteasome-dependent manner (93). Accordingly, DPCs are repaired via replication-coupled proteolysis that is either proteasome-dependent or proteasome-independent (165, 194–196). In *Xenopus* extracts, proteasome recruitment to DPCs is at least partially dependent on the replisome-coupled ubiquitination of the DPC (194, 195). Protein ubiquitination is initiated by the E1, E2, E3 enzymatic cascade that culminates in the covalent attachment of ubiquitin to a lysine residue of the target protein (197). Currently, we do not know which E3 ligase ubiquitinates HMCES-DPCs. The lab is currently working on identifying the HMCES-DPC E3 ligase. Using a list of known replication-fork associated E3 ligase, we are investigating which E3 ligase that when inactivated leads to an abundance of HMCES-DPCs after DNA damage (145). Recently, TRAIP, a replication-fork associated E3 ligase, has been proposed to ubiquitinate DPCs in *Xenopus* egg extracts (194). Proteasome-dependent polyubiquitination is partially dependent on TRAIP, as TRAIP

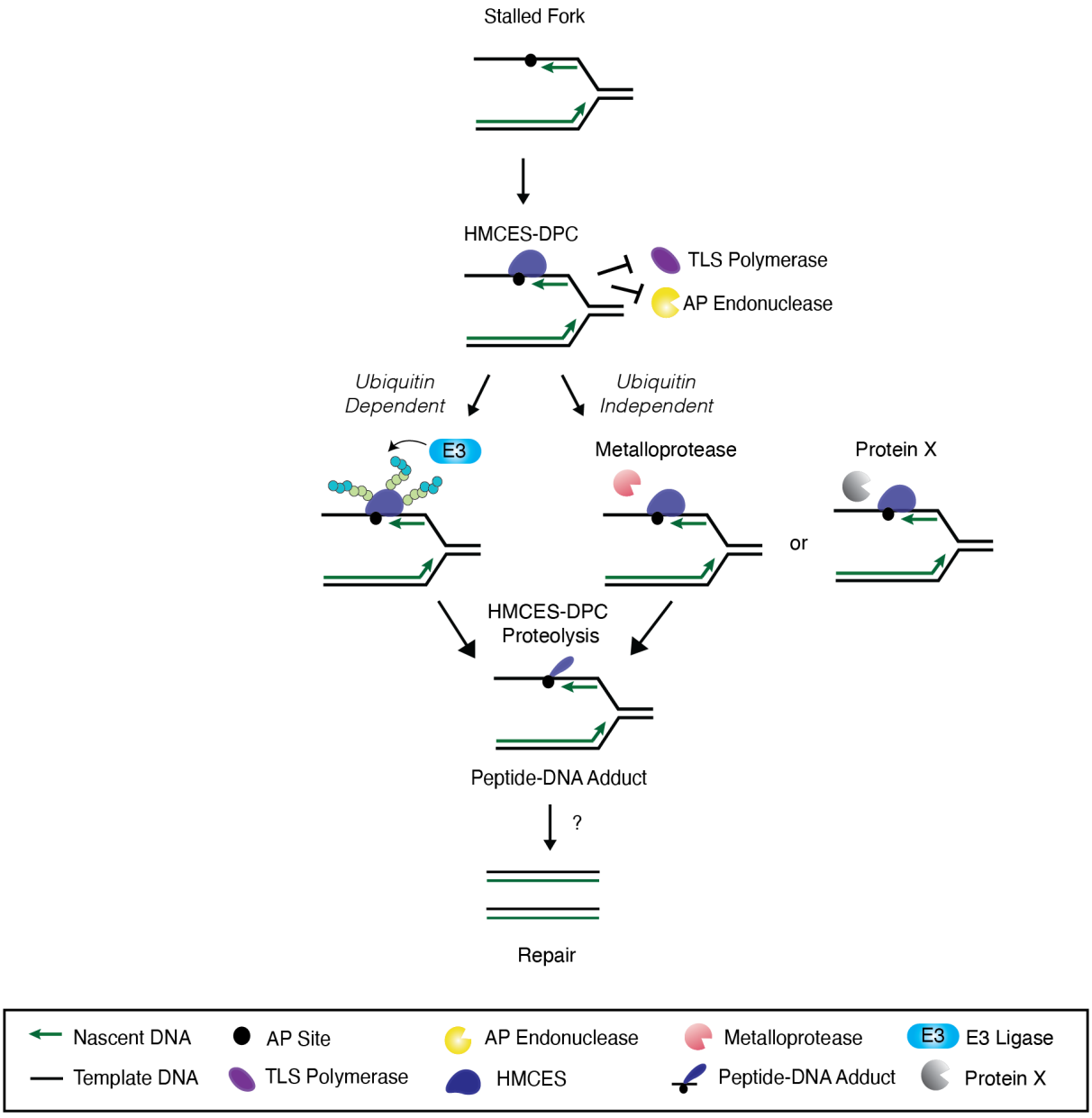

Figure 5.4: Model for HMCES-DPC degradation. HMCES-DPC forms at AP site stalled replication fork. DPC formation prevents DSB by AP endonucleases and mutagenic bypass by TLS polymerase. HMCES-DPC degradation proceeds via a ubiquitin-dependent proteolysis triggered by unknown E3 ligase(s) or via a ubiquitin-independent proteolysis triggered by a metalloprotease or unknown protein X. The resulting peptide-DNA adduct is resolved by an unknown mechanism.

depletion did not affect the ubiquitylation of DPCs in ssDNA (194). A second unknown E3 ligase may be responsible for signaling final degradation (194). Unsurprisingly, initial experiments inactivating TRAIP showed no difference in HMCES-DPC levels after damage (unpublished data). While this biased approach may be appropriate in narrowing down the list of over 500+ E3 ligases in human cells, it will be challenging to identify one HMCES-DPC specific E3 ligase (197). Indirect effects of perturbing ubiquitin homeostasis will confound the results. Furthermore, it is highly probable that multiple E3 ligases are responsible for HMCES-DPCs ubiquitination. As other DPC have been shown to be ubiquitinated by several E3 ligases. Discovering the E3 ligases that contribute to HMCES-DPC degradation will be crucial in understanding how HMCES-DPCs are resolution.

Recently, Spartan (SPRTN), a replication-specific metalloprotease, was identified as a central player in DPC repair. SPRTN is cell cycle regulated, increases in activity during the S phase, and functions in replication-coupled DPC proteolysis (165, 196). SPRTN directly cleaves the proteinaceous part of DPCs in ssDNA (165, 196). SPRTN-mediated DPC degradation does not require polyubiquitination, but instead depends on nascent strand extension near the DPC (194). One possibility is SPRTN directly degrades HMCES-DPCs. Preliminary studies indicate that knockdown of SPRTN in U2OS cells leads to a modest increase in HMCES-DPCs after UV damage (unpublished data). Further experiments are needed to determine if SPRTN plays a role in HMCES-DPC resolution. Testing the ability of SPRTN to degrade HMCES-DPCs *in vitro* would corroborate any *in vivo* findings. Indeed, collaborative activity between SPRTN and an E3 ligase for HMCES-DPC removal is also a possibility.

Other proteases such as ACRC/GCNA (acid repeat containing/germ cell nuclear antigen) and Ddi1 should also be considered for HMCES-DPC removal. ACRC/GCNA-1 is a SprT domain-containing metalloprotease that is phylogenetically similar to SPRTN (198). Unlike SPRTN, ACRC recruitment to DPCs is SUMO dependent (199). ACRC/GCNA-1 is critical for response to DPC forming agents in *Caenorhabditis elegans* (199). Likewise, Ddi1, a newly identified protein contributes to DPC repair in yeast (200, 201). While it is unclear if Ddi1-mediated proteolysis is dependent on the proteasome, investigating if Ddi1 homologs contribute to HMCES-DPC proteolysis may be useful (200, 201). It will be interesting to see which protease mediates HMCES-DPC removal.

Another consideration is that an unknown protein directly cleaves the HMCES-DPC. A dedicated protein for specific DPC cleavage is not unusual. For example, topoisomerases form irreversible protein tyrosyl-DNA complexes. These topoisomerase DPCs are cleaved by tyrosyl-DNA-phosphodiesterase (TDP1 and TDP2) (163). Identifying an unknown protein that cleaves the HMCES-DPC would be novel. Initial experiments identifying HMCES interacting proteins via mass spectrometry used an N-terminal FLAG-HMCES (93). PCNA was identified as a direct interactor with HMCES via the C-terminal PCNA-interacting motif on HMCES (93). However, N-terminal tagged HMCES does not form DPCs readily (93). Therefore, we may have missed any HMCES-DPCs specific interacting proteins. I propose using biotinylated DNA coupled with covalently linked HMCES to identify potential HMCES-DPCs in nuclear extracts. Such an experiment would need to be optimized to avoid degradation of the biotinylated DNA and inhibition of the proteasome to avoid degradation of HMCES. Immunoprecipitation (IP) of the biotinylated HMCES-DPC complex, followed by mass spectrometry (MS), could

identify novel HMCES-DPC interacting proteins. One pitfall is that the unidentified protein and HMCES-DPC interaction will be transient or weak, and thus visualization by IP-MS would not be possible.

Resolution of HMCES-DPCs

DPCs that are encountered by the replisome are degraded to short peptides via two proteolysis-dependent mechanisms that are proteasome-independent (SPRTN) or proteasome-dependent (TRAIP and/or another E3 ligase) (165, 194, 196). These short peptides are then bypassed by the translesion polymerase complex REV1-Pol ζ (202). DPC repair is inherently mutagenic. Mohni et al. proposed that HMCES shields AP sites to promote error-free repair (93). This proposal was based on the following: 1) the TLS polymerases REV1 and REV3 are increased at HMCESΔ replication forks, 2) HMCES-deficient cells have elevated mutation frequencies, and 3) the increased mutation frequency in HMCES-deficient cells is dependent on REV3 (93). The increased mutation frequencies were based on a UV-induced mutagenesis assay (203). Inherently, another UV-induced lesion could cause an increase in mutation frequency in HMCES-deficient cells. Furthermore, the data also indicate that TLS polymerases and HMCES act in different pathways. The error-free portion of the HMCES-mediated AP site repair model needs to be fully evaluated. Fluctuation assays in *ΔyedK* vs WT *E.coli* would be an ideal place to begin (204). Nevertheless, if the error-free repair model of HCMES were to be true, an error-free mechanism to resolve the HMCES-DPC must be explored.

Template switching is an attractive mechanism for error-free repair of HMCES-DPCs. A template-switching mechanism would allow for error-free DNA synthesis by

utilizing the undamaged sister chromatid as a template (135, 136). Once in duplex DNA, the HMCES-DPC could be processed by excision repair. Fork reversal could be another mechanism for placing the HMCES-DPC back into the context of duplex DNA. Exploring the role of fork reversal enzymes such as SMARCAL1, ZRANB3, and HTLF in HMCES-DPC resolution could be useful. However, the experimental design to answer this question (fork reversal enzymes for HMCES-DPC resolution) is challenging.

The temporal resolution of HMCES-DPC proteolysis coupled to template switching should be also be explored. One model is that template switching could occur after DPC ubiquitination and proteolysis but before TLS bypass of the peptide-DNA. This would place the smaller peptide-DNA adduct in the context of duplex DNA for error-free DNA synthesis. How this process would be regulated is not clear, and testing this hypothesis in mammalian system may be challenging. Perhaps, an in vitro assay with model substrates could be used. Alternatively, developing a system in Xenopus egg extracts is attractive. Although, to date there is limited evidence of template switching occurring in this system. Ultimately, understanding the intricacies of HMCES-DPC resolution will help to culminate the final steps in this new DNA repair pathway. Overall, the elucidating the steps for HMCES-DPC resolution will also contribute to the field of DPC repair.

Potential Role for HMCES in B-Cell Development

During an immune response, germinal centers B-cells proliferate, differentiate, and can undergo somatic hypermutation (SHM) or class switch recombination (CSR) to generate antibody diversity. Activated B-cells undergo two additional genomic alterations for antibody diversification, somatic hypermutation (SHM) and class switch recombination

(CSR). Activation-induced cytidine deaminase (AID) initiates SHM and CSR by converting cytosines to uracils in ssDNA (205, 206). These uracils are then converted to abasic sites via BER. In SHM, U: G lesion mismatches and abasic sites generate point mutations in Ig variable regions, which underlies the selection for increased antigen affinity (205). In CSR, these complex abasic site lesions cause single-strand breaks that promote the generation of DSB (206). Recombination repair of DSBs during CSR drives isotype switching from IgM/IgD to IgG, IgA, IgE (206). Abasic sites are present during SHM and CSR, indicating a potential role of HMCES, an abasic site shielding protein, in B-cell development.

Degradation of the FOXO1 transcription factor in a diffuse large B-cell lymphoma (DLBCL) cell line leads to a rapid downregulation of transcription at select germinal center B-cell genes, including AID and curiously HMCES (unpublished data Hiebert lab). While the transcriptional regulation of HMCES has been unexplored, this data at least suggests that HMCES may play a role in B-cell development. Recently, Shukla and colleagues using CRISPR/cas9 generated *hmces* knockout mice observed that B-cell activation leads to an increase in *Hmces* mRNA expression(150). There were no observed defects in overall hematopoiesis, but *hmces* deficiency did lead to a defect in CSR. The authors attributed this HMCES-mediated defect in CSR to a new function of HMCES in DSB repair. Specifically, they conclude that HMCES functions in microhomology-mediated alternative-end joining (Alt-EJ), and this HMCES Alt-EJ function was dependent on DNA binding but not crosslinking activity. Complementation of ΔHMCES CH12 B-cells with C2A mutants but not DNA binding mutants rescued the CSR defect. They, however, did not complement their plasmid-based Alt-EJ assays with C2A or DNA binding mutants so

the conclusion made that a HMCES-mediated defect in Alt-EJ is due to DNA binding and not the crosslinking activity of HMCES is unsupported. Furthermore, all of their complementation assays used N-terminal tagged HA proteins. N-terminal tagged HMCES does not rescue HMCESΔ phenotypes such as IR sensitivity (93). Finally, HMCESΔ U2OS cells expressing untagged HMCES crosslinking mutant (C2A) are still sensitive to IR. Therefore, Alt-EJ is unlikely to explain the IR sensitivity of HMCES-deficient cells.

A HMCES-mediated defect in CSR is probable. However, the mechanism of HMCES involvement is unclear. One possibility is that downregulation of HMCES mRNA in B-cells depletes the HMCES pool able to bind abasic sites, shielding them from AID. Recent studies in the lab suggest that HMCES protects cells from APOBEC3A-induced abasic sites lending evidence to a possible mechanism in CSR (93). Further studies, however, are required to understand the relationship of HMCES abasic site shielding activity in SHM and CSR. HMCES-deficiency does sensitize cells to IR and causes an increase in γH2AX (93), (150). HMCES may be an attractive drug target for germinal cell derived B-cell lymphomas. HMCES inhibition could potentially act as a radiosensitizer in a subset of these cancers. Furthermore, HMCES inhibition could increase the mutational load in a subset of these cancers. HMCES inhibition would lead to increase mutations and perhaps the presence of neoantigens, thus increasing the immunogenicity of these cancers. Therefore, generating a HMCES inhibitor could be another potential avenue of investigation.

Conclusion

The prevalence, mutagenicity, and potential for cytotoxicity of AP sites explains why there are so many evolutionarily conserved mechanisms for AP site repair and tolerance. The context of when and where an AP site lesion occurs helps to determine which repair pathway is used, but additional regulation that remains poorly understood must also be important. Given the long history of AP site repair studies, the recent identification of a new mechanism mediated HMCES underscores how much we do not know. In this thesis, I show that HMCES forms a stable thiazolidine linkage with AP-ssDNA and the chemical mechanism of DPC formation is conserved, at least in the human and E. coli SRAP proteins. Further studies will undoubtedly generate new surprises for this new model of AP site repair and may reveal new mechanisms for DPC repair. Finally, these studies have health relevance. In the context of cancer, exploiting AP site repair may be advantageous especially in the context of human B-cell lymphoma where there is a higher expression of AID and high AP site levels.

References

1. Friedberg, E. C., Walker, G. C., Wolfram, S., Wood, R. D., Schulltz, R. A., and Ellenberger, T. (2006) *DNA Repair and Mutagenesis, Second Edition*, ASM Press, Washington, DC
2. Lindahl, T., and Nyberg, B. (1972) Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry*. **11**, 3610–3618
3. Barnes, D. E., and Lindahl, T. (2004) Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells. *Annu. Rev. Genet*. **38**, 445–76
4. Lindahl, T., and Karlström, O. (1973) Heat-Induced Depyrimidination of Deoxyribonucleic Acid in Neutral Solution. *Biochemistry*. **12**, 5151–5154
5. Chastain, P. D., Nakamura, J., Rao, S., Chu, H., Ibrahim, J. G., Swenberg, J. A., and Kaufman, D. G. (2010) Abasic sites preferentially form at regions undergoing DNA replication. *FASEB J.* **24**, 3674–3680
6. Loeb, L. A., and Preston, B. D. (1986) Mutagenesis by Apurinic/Apyrimidinic Sites. *Annu. Rev. Genet.* **20**, 201–230
7. Lawley, P. D., and Brookes, P. (1963) Further Studies on the Alkylation of Nucleic Acids and Their Constituent Nucleotides. *Biochem. J.* **89**, 127–138
8. Singer, B. (1976) All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature*. **264**, 333–339
9. McNeill, D. R., Lam, W., DeWeese, T. L., Cheng, Y. C., and Wilson, D. M. (2009) Impairment of APE1 function enhances cellular sensitivity to clinically relevant alkylators and antimetabolites. *Mol. Cancer Res.* **7**, 897–906
10. Croteau, D. L., and Bohr, V. A. (1997) Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells. *J. Biol. Chem.* **272**, 25409–25412
11. Cadet, J., and Richard Wagner, J. (2013) DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb. Perspect. Biol.* **5**, a012559
12. Whitaker, A. M., Schaich, M. A., Smith, M. S., Flynn, T. S., and Freudenthal, B. D. (2017) Base excision repair of oxidative DNA damage: From mechanism to disease. *Front. Biosci. - Landmark*. **22**, 1493–1522
13. Krokan, H. E., and Bjørås, M. (2013) Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, 1–22
14. Téoule, R. (1987) Radiation-induced DNA Damage and Its Repair. *Int. J. Radiat. Biol. Relat. Stud. Phys.* **51**, 573–589
15. Dunlap, B., and Cerutti, P. (1975) Apyrimidinic sites in gamma-irradiated DNA. *FEBS Lett.* **51**, 188–190
16. 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

17. Higgins, S. A., Frenkel, K., Cummings, A., and Teebor, G. W. (1987) Definitive characterization of human thymine glycol N-glycosylase activity. *Biochemistry*. **26**, 1683–1688
18. Lindahl, T. (1979) DNA Glycosylases, Endonucleases for Apurinic/Apyrimidinic Sites, and Base Excision-Repair. in *Progress in Nucleic Acid Research and Molecular Biology*, pp. 135–192, Academic Press, Cambridge, **22**, 135–192
19. Boorstein, R. J., Hilbert, T. P., Teebor, G. W., and Cunningham, R. P. (1990) Formation and Stability of Repairable Pyrimidine Photohydrates in DNA. *Biochemistry*. **29**, 10455–10460
20. Ganguly, T., and Duker, N. J. (1991) Stability of DNA thymine hydrates. *Nucleic Acids Res.* **19**, 3319–3323
21. Joyce, C. M. (1997) Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1619–1622
22. Williams, J. S., Lujan, S. A., and Kunkel, T. A. (2016) Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* **17**, 350–363
23. Tomilin, N. V., and Aprelikova, O. N. (1989) Uracil-DNA Glycosylases and DNA Uracil Repair. *Int. Rev. Cytol.* **114**, 125–179
24. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*. **362**, 709–715
25. Bhagwat, A. S., Hao, W., Townes, J. P., Lee, H., Tang, H., and Foster, P. L. (2016) Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2176–2181
26. Conticello, S. G., Langlois, M. A., Yang, Z., and Neuberger, M. S. (2007) DNA Deamination in Immunity: AID in the Context of Its APOBEC Relatives. *Adv. Immunol.* **94**, 37–73
27. Rebhandl, S., Huemer, M., Greil, R., and Geisberger, R. (2015) AID/APOBEC deaminases and cancer. *Oncoscience*. **2**, 320–333
28. Burns, M. B., Temiz, N. A., and Harris, R. S. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **45**, 977–983
29. Hoopes, J. I., Cortez, L. M., Mertz, T. M., Malc, E. P., Mieczkowski, P. A., and Roberts, S. A. (2016) APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep.* **14**, 1273–1282
30. Seplyarskiy, V. B., Soldatov, R. A., Popadin, K. Y., Antonarakis, S. E., Bazykin, G. A., and Nikolaev, S. I. (2016) APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res.* **26**, 174–182
31. 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 17O NMR. *J. Am. Chem. Soc.* **111**, 1894–1896
32. Lhomme, J., Constant, J. F., and Demeunynck, M. (1999) Abasic DNA structure, reactivity, and recognition. *Biopolymers*. **52**, 65–83
33. Talpaert-Borlè, M. (1987) Formation, detection and repair of AP sites. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **181**, 45–56

34. Dutta, S., Chowdhury, G., and Gates, K. S. (2007) Interstrand cross-links generated by abasic sites in duplex DNA. *J. Am. Chem. Soc.* **129**, 1852–1853
35. Sczepanski, J. T., Jacobs, A. C., and Greenberg, M. M. (2008) Self-promoted DNA interstrand cross-link formation by an abasie site. *J. Am. Chem. Soc.* **130**, 9646–9647
36. 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. *J. Am. Chem. Soc.* **136**, 3483–3490
37. Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C., and Walter, J. C. (2016) Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell*. **167**, 498-511.e14
38. Khodyreva, S. N., Prasad, R., Ilina, E. S., Sukhanova, M. V., Kutuzov, M. M., Liu, Y., Hou, E. W., Wilson, S. H., and Lavrik, O. I. (2010) Apurinic/apyrimidinic (AP) site recognition by the 5′-dRP/AP lyase in poly(ADP-ribose) polymerase-1 (PARP-1). *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22090–22095
39. Prasad, R., Horton, J. K., Chastain, P. D., 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 Res.* **42**, 6337–51
40. 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
41. 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. *J. Biol. Chem.* **277**, 7637–7640
42. 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. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8602–8607
43. 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
44. 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. *J. Biol. Chem.* **287**, 13686–13693
45. 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 (Amst).* **87**, 102773
46. Quiñones, J. L., and Demple, B. (2016) When DNA repair goes wrong: BER-generated DNA-protein crosslinks to oxidative lesions. *DNA Repair (Amst).* **44**, 103–109
47. Yang, S. W., Burgin, A. B., Huizenga, B. N., Robertson, C. A., Yao, K. C., and Nash, H. A. (1996) A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. *Proc. Natl. Acad. Sci.* **93**, 11534–11539
48. Ledesma, F. C., El Khamisy, S. F., Zuma, M. C., Osborn, K., and Caldecott, K. W.

(2009) A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature*. **461**, 674–678
49. Cuniasse, P., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Lebret, M., Guschlbauer, W., and Fazakerley, G. V. (1987) An abasic site in DNA. Solution conformation determined by proton NMR and molecular mechanics calculations. *Nucleic Acids Res.* **15**, 8003–8022
50. 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
51. Lockhart, M. L., Deutsch, J. F., Yamaura, I., Cavalieri, L. F., and Rosenberg, B. H. (1982) Termination of DNA synthesis in vitro at apurinic sites but not at ethyl adducts on the template. *Chem. Biol. Interact.* **42**, 85–95
52. Sagher, D., and Strauss, B. (1983) Insertion of Nucleotides Opposite Apurinic/Apyrimidinic Sites in Deoxyribonucleic Acid during in Vitro Synthesis: Uniqueness of Adenine Nucleotides. *Biochemistry*. **22**, 4518–4526
53. Loeb, L. A., and Preston, B. D. (1986) Mutagenesis by Apurinic/Apyrimidinic Sites. *Annu. Rev. Genet.* **20**, 201–230
54. Strauss, B. S. (2002) The “A” rule revisited: Polymerases as determinants of mutational specificity. *DNA Repair (Amst).* **1**, 125–135
55. Kunkel, T. A., Schaaper, R. M., and Loeb, L. A. (1983) Depurination-induced infidelity of DNA synthesis with purified DNA replication proteins in vitro. *Biochemistry*. **22**, 2378–2384
56. Randall, S. K., Eritja, R., Kaplan, B. E., Petruska, J., and Goodman, M. F. (1987) Nucleotide insertion kinetics opposite abasic lesions in DNA. *J. Biol. Chem.* **262**, 6864–6870
57. Boiteux, S., and Laval, J. (1982) Coding Properties of Poly(deoxycytidylic acid) Templates Containing Uracil or Apyrimidinic Sites: In Vitro Modulation of Mutagenesis by Deoxyribonucleic Acid Repair Enzymes. *Biochemistry*. **21**, 6746–6751
58. Pagès, V., Johnson, R. E., Prakash, L., and Prakash, S. (2008) Mutational specificity and genetic control of replicative bypass of an abasic site in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1170–1175
59. Gibbs, P. E. M., and Lawrence, C. W. (1995) Novel mutagenic properties of abasic sites in saccharomyces cerevisiae. *J. Mol. Biol.* **251**, 229–236
60. Otsuka, C., Sanadai, S., Hata, Y., Okuto, H., Noskov, V. N., Loakes, D., and Negishi, K. (2002) Difference between deoxyribose- and tetrahydrofuran-type abasic sites in the in vivo mutagenic responses in yeast. *Nucleic Acids Res.* **30**, 5129–5135
61. Kunz, B. A., Henson, E. S., Roche, H., Ramotar, D., Nunoshiba, T., and Demple, B. (1994) Specificity of the mutator caused by deletion of the yeast structural gene (APN1) for the major apurinic endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 8165–8169
62. Gentil, A., Renault, G., Madzak, C., Margot, A., Cabral-Neto, J. B., Vasseur, J. J., Rayner, B., Imbach, J. L., and Sarasin, A. (1990) Mutagenic properties of a unique abasic site in mammalian cells. *Biochem. Biophys. Res. Commun.* **173**,

704–710
63. Klinedinst, D. K., and Drinkwater, N. R. (1992) Mutagenesis by apurinic sites in normal and ataxia telangiectasia human lymphoblastoid cells. *Mol. Carcinog.* **6**, 32–42
64. Cabral Neto, J. B., Caseira Cabral, R. E., Margot, A., Le Page, F., Sarasin, A., and Gentil, A. (1994) Coding properties of a unique Apurinic/Apyrimidinic site replicated in mammalian cells. *J. Mol. Biol.* **240**, 416–420
65. Shuck, S. C., Short, E. A., and Turchi, J. J. (2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res.* **18**, 64–72
66. Spivak, G. (2015) Nucleotide excision repair in humans. *DNA Repair (Amst).* **36**, 13–18
67. David, S. S., O'Shea, V. L., and Kundu, S. (2007) Base-excision repair of oxidative DNA damage. *Nature*. **447**, 941–950
68. Wallace, S. S. (2014) Base excision repair: A critical player in many games. *DNA Repair (Amst).* **19**, 14–26
69. Beard, W. A., Horton, J. K., Prasad, R., and Wilson, S. H. (2019) Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annu. Rev. Biochem.* **88**, 137–162
70. Kim, Y.-J., and M. Wilson III, D. (2011) Overview of Base Excision Repair Biochemistry. *Curr. Mol. Pharmacol.* **5**, 3–13
71. Sukhanova, M. V, Khodyreva, S. N., Lebedeva, N. A., Prasad, R., Wilson, S. H., and Lavrik, O. I. (2005) Human base excision repair enzymes apurinic/apyrimidinic endonuclease1 (APE1), DNA polymerase beta and poly(ADP-ribose) polymerase 1: interplay between strand-displacement DNA synthesis and proofreading exonuclease activity. *Nucleic Acids Res.* **33**, 1222–9
72. Petermann, E., Ziegler, M., and Oei, S. L. (2003) ATP-dependent selection between single nucleotide and long patch base excision repair. *DNA Repair (Amst).* **2**, 1101–14
73. Fortini, P., and Dogliotti, E. (2007) Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst).* **6**, 398–409
74. Sancar, A. (1996) DNA Excision Repair. *Annu. Rev. Biochem.* **65**, 43–81
75. Snowden, A., Van Houten, B., Kow, Y. W., Van Houten, B., and Van Houten, B. (1990) Damage Repertoire of the Escherichia coli UvrABC Nuclease Complex Includes Abasic Sites, Base-Damage Analogues, and Lesions Containing Adjacent 5′ or 3′ Nicks. *Biochemistry*. **29**, 7251–7259
76. Lin, J. J., and Sancar, A. (1989) A New Mechanism for Repairing Oxidative Damage to DNA: (A)BC Excinuclease Removes AP Sites and Thymine Glycols from DNA. *Biochemistry*. **28**, 7979–7984
77. Reardon, J. T., Bessho, T., Kung, H. C., Bolton, P. H., and Sancar, A. (1997) In vitro repair of oxidative DNA damage by human nucleotide excision repair system: Possible explanation for neurodegeneration in Xeroderma pigmentosum patients. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9463–9468
78. Torres-Ramos, C. A., Johnson, R. E., Prakash, L., and Prakash, S. (2000)

Evidence for the Involvement of Nucleotide Excision Repair in the Removal of Abasic Sites in Yeast. *Mol. Cell. Biol.* **20**, 3522–3528
79. Swanson, R. L., Morey, N. J., Doetsch, P. W., and Jinks-Robertson, S. (1999) Overlapping Specificities of Base Excision Repair, Nucleotide Excision Repair, Recombination, and Translesion Synthesis Pathways for DNA Base Damage in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **19**, 2929–2935
80. Tornaletti, S., Maeda, L. S., and Hanawalt, P. C. (2006) Transcription Arrest at an Abasic Site in the Transcribed Strand of Template DNA. *Chem. Res. Toxicol.* **19**, 1215–1220
81. Wang, W., Walmacq, C., Chong, J., Kashlev, M., and Wang, D. (2018) Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2538–E2545
82. Yu, S.-L., Lee, S.-K., Johnson, R. E., Prakash, L., and Prakash, S. (2003) The Stalling of Transcription at Abasic Sites Is Highly Mutagenic. *Mol. Cell. Biol.* **23**, 382–388
83. Kim, N., and Jinks-Robertson, S. (2010) Abasic Sites in the Transcribed Strand of Yeast DNA Are Removed by Transcription-Coupled Nucleotide Excision Repair. *Mol. Cell. Biol.* **30**, 3206–3215
84. Kim, N., and Jinks-Robertson, S. (2009) dUTP incorporation into genomic DNA is linked to transcription in yeast. *Nature*. **459**, 1150–1153
85. Sweder, K. S., and Hanawalt, P. C. (1992) Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10696–10700
86. Clauson, C. L., Oestreich, K. J., Austin, J. W., and Doetsch, P. W. (2010) Abasic sites and strand breaks in DNA cause transcriptional mutagenesis in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3657–3662
87. Kitsera, N., Rodriguez-Alvarez, M., Emmert, S., Carell, T., and Khobta, A. (2019) Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Res.* **47**, 8537–8547
88. Marenstein, D. R., Wilson, D. M., and Teebor, G. W. (2004) Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded DNA. *DNA Repair (Amst).* **3**, 527–533
89. Dou, H., Mitra, S., and Hazra, T. K. (2003) Repair of Oxidized Bases in DNA Bubble Structures by Human DNA Glycosylases NEIL1 and NEIL2. *J. Biol. Chem.* **278**, 49679–49684
90. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., and Slupphaug, G. (2002) hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J. Biol. Chem.* **277**, 39926–39936
91. Schormann, N., Ricciardi, R., and Chattopadhyay, D. (2014) Uracil-DNA glycosylases - Structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.* **23**, 1667–1685
92. 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., De Magis, A., Kwon, Y., Godin, S. K., Van Houten, B., Normolle, D. P., Sung, P., Das, S. R., Paeschke, K., Roberts, S. A., Vandemark, A. P., and Bernstein, K. A. (2019) The Rad51 paralogs facilitate a novel DNA strand specific damage tolerance pathway. *Nat. Commun.* **10**, 3515

93. 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.e13
94. Prakash, S., Johnson, R. E., and Prakash, L. (2005) Eukaryotic Translesion Synthesis DNA Polymerases: Specificity of Structure and Function. *Annu. Rev. Biochem.* **74**, 317–353
95. Waters, L. S., Minesinger, B. K., Wiltrout, M. E., D'Souza, S., Woodruff, R. V., and Walker, G. C. (2009) Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiol. Mol. Biol. Rev.* **73**, 134–154
96. Chan, K., Resnick, M. A., and Gordenin, D. A. (2013) The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA Repair (Amst).* **12**, 878–889
97. Zeng, X., Winter, D. B., Kasmer, C., Kraemer, K. H., Lehmann, A. R., and Gearhart, P. J. (2001) DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat. Immunol.* **2**, 537–541
98. Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J. C., and Reynaud, C. A. (2005) Contribution of DNA polymerase η to immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.* **201**, 1191–1196
99. Masutani, C., Kusumoto, R., Iwai, S., and Hanaoka, F. (2000) Mechanisms of accurate translesion synthesis by human DNA polymerase eta. *EMBO J.* **19**, 3100–9
100. Zhao, B., Xie, Z., Shen, H., and Wang, Z. (2004) Role of DNA polymerase η in the bypass of abasic sites in yeast cells. *Nucleic Acids Res.* **32**, 3984–3994
101. Haracska, L., Washington, M. T., Prakash, S., and Prakash, L. (2001) Inefficient Bypass of an Abasic Site by DNA Polymerase η. *J. Biol. Chem.* **276**, 6861–6866
102. Patra, A., Zhang, Q., Lei, L., Su, Y., Egli, M., and Guengerich, F. P. (2015) Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase. *J. Biol. Chem.* **290**, 8028–8038
103. Faili, A., Aoufouchi, S., Flatter, E., Guéranger, Q., Reynaud, C. A., and Weill, J. C. (2002) Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. *Nature*. **419**, 944–947
104. McDonald, J. P., Frank, E. G., Plosky, B. S., Rogozin, I. B., Masutani, C., Hanaoka, F., Woodgate, R., and Gearhart, P. J. (2003) 129-Derived strains of mice are deficient in DNA polymerase ι and have normal immunoglobulin hypermutation. *J. Exp. Med.* **198**, 635–643
105. Martomo, S. A., Yang, W. W., Vaisman, A., Maas, A., Yokoi, M., Hoeijmakers, J. H., Hanaoka, F., Woodgate, R., and Gearhart, P. J. (2006) Normal hypermutation

in antibody genes from congenic mice defective for DNA polymerase ι. *DNA Repair (Amst).* **5**, 392–398
106. Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009) DNA Synthesis across an Abasic Lesion by Human DNA Polymerase ι. *Structure*. **17**, 530–537
107. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S., and Prakash, L. (2000) Eukaryotic polymerases [iota] and [zeta] act sequentially to bypass DNA lesions. *Nature*. **406**, 1015–1019
108. Choi, J. Y., Lim, S., Kim, E. J., Jo, A., and Guengerich, F. P. (2010) Translesion Synthesis across Abasic Lesions by Human B-Family and Y-Family DNA Polymerases α, δ, η, ι, κ, and REV1. *J. Mol. Biol.* **404**, 34–44
109. Lin, W., Xin, H., Zhang, Y., Wu, X., Yuan, F., and Wang, Z. (1999) The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Res.* **27**, 4468–4475
110. Nelson, J. R., Lawrence, C. W., and Hinkle, D. C. (1996) Deoxycytidyl transferase activity of yeast REV1 protein. *Nature*. **382**, 729–731
111. Jansen, J. G., Langerak, P., Tsaalbi-Shtylik, A., Van Den Berk, P., Jacobs, H., and De Wind, N. (2006) Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J. Exp. Med.* **203**, 319–323
112. Simpson, L. J. (2003) Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *EMBO J.* **22**, 1654–1664
113. Haracska, L., Unk, I., Johnson, R. E., Johansson, E., Burgers, P. M. J., Prakash, S., and Prakash, L. (2001) Roles of yeast DNA polymerases δ and ζ of Rev 1 in the bypass of abasic sites. *Genes Dev.* **15**, 945–954
114. Gibbs, P. E. M., McDonald, J., Woodgate, R., and Lawrence, C. W. (2005) The relative roles in vivo of Saccharomyces cerevisiae Pol η, Pol ζ, Rev1 protein and Pol32 in the bypass and mutation induction of an abasic site, T-T (6-4) photoadduct and T-T cis-syn cyclobutane dimer. *Genetics*. **169**, 575–582
115. Kim, N., Mudrak, S. V., and Jinks-Robertson, S. (2011) The dCMP transferase activity of yeast Rev1 is biologically relevant during the bypass of endogenously generated AP sites. *DNA Repair (Amst).* **10**, 1262–1271
116. Schenten, D., Gerlach, V. L., Guo, C., Velasco-Miguel, S., Hladik, C. L., White, C. L., Friedberg, E. C., Rajewsky, K., and Esposito, G. (2002) DNA polymerase K deficiency does not affect somatic hypermutation in mice. *Eur. J. Immunol.* **32**, 3152–3160
117. Shimizu, T., Shinkai, Y., Ogi, T., Ohmori, H., and Azuma, T. (2003) The absence of DNA polymerase kappa does not affect somatic hypermutation of the mouse immunoglobulin heavy chain gene. *Immunol. Lett.* **86**, 265–70
118. Seki, M., Masutani, C., Yang, L. W., Schuffert, A., Iwai, S., Bahar, I., and Wood, R. D. (2004) High-efficiency bypass of DNA damage by human DNA polymerase Q. *EMBO J.* **23**, 4484–4494
119. Hogg, M., Seki, M., Wood, R. D., Doublié, S., and Wallace, S. S. (2011) Lesion bypass activity of DNA polymerase θ (POLQ) is an intrinsic property of the pol

domain and depends on unique sequence inserts. *J. Mol. Biol.* **405**, 642–652
120. Poltoratsky, V., Prasad, R., Horton, J. K., and Wilson, S. H. (2007) Down-regulation of DNA polymerase β accompanies somatic hypermutation in human BL2 cell lines. *DNA Repair (Amst).* **6**, 244–253
121. Esposito, G., Texido, G., Betz, U. A., Gu, H., Müller, W., Klein, U., and Rajewsky, K. (2000) Mice reconstituted with DNA polymerase beta-deficient fetal liver cells are able to mount a T cell-dependent immune response and mutate their Ig genes normally. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1166–1171
122. Wu, X., and Stavnezer, J. (2007) DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J. Exp. Med.* **204**, 1677–1689
123. Kochenova, O. V., Daee, D. L., Mertz, T. M., and Shcherbakova, P. V. (2015) DNA Polymerase ζ-Dependent Lesion Bypass in Saccharomyces cerevisiae Is Accompanied by Error-Prone Copying of Long Stretches of Adjacent DNA. *PLoS Genet.* **11**, e1005110
124. Diaz, M., Verkoczy, L. K., Flajnik, M. F., and R. Klinman, N. (2001) Decreased Frequency of Somatic Hypermutation and Impaired Affinity Maturation but Intact Germinal Center Formation in Mice Expressing Antisense RNA to DNA Polymerase ζ. *J. Immunol.* **167**, 327–335
125. Tang, M., Pham, P., Shen, X., Taylor, J. S., O'Donnell, M., Woodgate, R., and Goodman, M. F. (2000) Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature*. **404**, 1014–1018
126. Reuven, N. B., Arad, G., Maor-Shoshani, A., and Livneh, Z. (1999) The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication. *J. Biol. Chem.* **274**, 31763–31766
127. Maor-Shoshani, A., Hayashi, K., Ohmori, H., and Livneh, Z. (2003) Analysis of translesion replication across an abasic site by DNA polymerase IV of Escherichia coli. *DNA Repair (Amst).* **2**, 1227–1238
128. Janion, C., Sikora, A., Nowosielska, A., and Grzesiuk, E. (2003) E. coli BW535, a triple mutant for the DNA repair genes xth, nth, and nfo, chronically induces the SOS response. *Environ. Mol. Mutagen.* **41**, 237–242
129. Rajagopalan, M., Lu, C., Woodgate, R., O'Donnell, M., Goodman, M. F., and Echols, H. (1992) Activity of the purified mutagenesis proteins UmuC, UmuD', and RecA in replicative bypass of an abasic DNA lesion by DNA polymerase III. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10777–10781
130. Reuven, N. B., Tomer, G., and Livneh, Z. (1998) The mutagenesis proteins UmuD′ and UmuC prevent lethal frameshifts while increasing base substitution mutations. *Mol. Cell*. **2**, 191–199
131. Ma, W., Westmoreland, J. W., Gordenin, D. A., and Resnick, M. A. (2011) Alkylation base damage is converted into repairable double-strand breaks and complex intermediates in g2 cells lacking ap endonuclease. *PLoS Genet.* **7**, e1002059
132. Otterlei, M., Warbrick, E., Nagelhus, T. A., Haug, T., Slupphaug, G., Akbari, M., Aas, A., Steinsbekk, K., Bakke, O., and Krokan, H. E. (1999) Post-replicative base

excision repair in replication foci. *EMBO J.* **18**, 3834–3844
133. Otterlei, M., Kavli, B., Standal, R., Skjelbred, C., Bharati, S., and Krokan, H. E. (2000) Repair of chromosomal abasic sites in vivo involves at least three different repair pathways. *EMBO J.* **19**, 5542–5551
134. 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
135. Cortez, D. (2019) Replication-Coupled DNA Repair. *Mol. Cell*. **74**, 866–876
136. Bhat, K. P., and Cortez, D. (2018) RPA and RAD51: Fork reversal, fork protection, and genome stability. *Nat. Struct. Mol. Biol.* **25**, 446–453
137. Izhar, L., Goldsmith, M., Dahan, R., Geacintov, N., Lloyd, R. G., and Livneh, Z. (2008) Analysis of Strand Transfer and Template Switching Mechanisms of DNA Gap Repair by Homologous Recombination in Escherichia coli: Predominance of Strand Transfer. *J. Mol. Biol.* **381**, 803–809
138. Neelsen, K. J., and Lopes, M. (2015) Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nat. Rev. Mol. Cell Biol.* **16**, 207–220
139. Martino, J., and Bernstein, K. A. (2016) The Shu complex is a conserved regulator of homologous recombination. *FEMS Yeast Res.* **16**, fow073
140. Godin, S. K., Zhang, Z., Herken, B. W., Westmoreland, J. W., Lee, A. G., Mihalevic, M. J., Yu, Z., Sobol, R. W., Resnick, M. A., and Bernstein, K. A. (2016) The Shu complex promotes error-free tolerance of alkylation-induced base excision repair products. *Nucleic Acids Res.* **44**, 8199–8215
141. Martín, V., Chahwan, C., Gao, H., Blais, V., Wohlschlegel, J., Yates 3rd, J. R., McGowan, C. H., and Russell, P. (2006) Sws1 is a conserved regulator of homologous recombination in eukaryotic cells. *EMBO J.* **25**, 2564–2574
142. Liu, T., Wan, L., Wu, Y., Chen, J., and Huang, J. (2011) hSWS1·SWSAP1 is an evolutionarily conserved complex required for efficient homologous recombination repair. *J. Biol. Chem.* **286**, 41758–66
143. Martino, J., Brunette, G. J., Barroso-González, J., Moiseeva, T. N., Smith, C. M., Bakkenist, C. J., O'Sullivan, R. J., and Bernstein, K. A. (2019) The human Shu complex functions with PDS5B and SPIDR to promote homologous recombination. *Nucleic Acids Res.* **47**, 10151–10165
144. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., Van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*. **152**, 1146–1159
145. Wessel, S. R., Mohni, K. N., Luzwick, J. W., Dungrawala, H., and Cortez, D. (2019) Functional Analysis of the Replication Fork Proteome Identifies BET Proteins as PCNA Regulators. *Cell Rep.* **28**, 3497-3509.e4
146. Dungrawala, H., and Cortez, D. (2015) Purification of proteins on newly synthesized DNA using iPOND. *Methods Mol. Biol.* **1228**, 123–131
147. Saldivar, J. C., Cortez, D., and Cimprich, K. A. (2017) The essential kinase ATR: ensuring faithful duplication of a challenging genome. *Nat. Rev. Mol. Cell Biol.* **18**, 622–636

148. 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. *Biol. Direct*. **8**, 20
149. 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
150. Shukla, V., Halabelian, L., Balagere, S., Samaniego-Castruita, D., Feldman, D. E., Arrowsmith, C. H., Rao, A., and Aravind, L. (2020) HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. *Mol. Cell*. **77**, 384-394.e4
151. 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. *Under Rev.*
152. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
153. 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., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221
154. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132
155. 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–e01456
156. 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 Res.* **44**, W344–W350
157. Dungrawala, H., Rose, K. L., Bhat, K. P., Mohni, K. N., Glick, G. G., Couch, F. B., and Cortez, D. (2015) The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. *Mol. Cell*. **59**, 998–1010
158. Beger, R. D., and Bolton, P. H. (1998) Structures of apurinic and apyrimidinic sites in duplex DNAs. *J. Biol. Chem.* **273**, 15565–73
159. Wu, H., and Zhang, Y. (2011) Mechanisms and functions of Tet proteinmediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452
160. Wingfield, P. T. (2017) N-Terminal Methionine Processing. *Curr. Protoc. protein Sci.* **88**, 6.14.1-6.14.3
161. Hirel, P. H., Schmitter, J. M., Dessen, P., Fayat, G., and Blanquet, S. (1989) Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8247–8251
162. Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2009) A summary of the

measured pK values of the ionizable groups in folded proteins. *Protein Sci.* **18**, 247–251
163. Pommier, Y., Huang, S. N., Gao, R., Das, B. B., Murai, J., and Marchand, C. (2014) Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). *DNA Repair (Amst).* **19**, 114–129
164. Stingele, J., Schwarz, M. S., Bloemeke, N., Wolf, P. G., and Jentsch, S. (2014) A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell*. **158**, 327–338
165. Stingele, J., Bellelli, R., Alte, F., Hewitt, G., Sarek, G., Maslen, S. L., Tsutakawa, S. E., Borg, A., Kjær, S., Tainer, J. A., Skehel, J. M., Groll, M., and Boulton, S. J. (2016) Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol. Cell*. **64**, 688–703
166. Stingele, J., Bellelli, R., and Boulton, S. J. (2017) Mechanisms of DNA-protein crosslink repair. *Nat. Rev. Mol. Cell Biol.* **18**, 563–573
167. Gentle, I. E., De Souza, D. P., Baca, M., Ian E. Gentle, David P. De Souza, A., and Baca, M. (2004) Direct Production of Proteins with N-Terminal Cysteine for Site-Specific Conjugation. *Bioconjug. Chem.* **15**, 658–663
168. Zhang, L., and Tam, J. P. (1996) Thiazolidine Formation as a General and Site-Specific Conjugation Method for Synthetic Peptides and Proteins. *Anal. Biochem.* **233**, 87–93
169. Bermejo-Velasco, D., Nawale, G. N., Oommen, O. P., Hilborn, J., and Varghese, O. P. (2018) Thiazolidine chemistry revisited: A fast, efficient and stable click-type reaction at physiological pH. *Chem. Commun.* **54**, 12507–12510
170. Ratner, S., and Clarke, H. T. (1937) The Action of Formaldehyde upon Cysteine. *J. Am. Chem. Soc.* **59**, 200–206
171. Liu, C.-F., and Tam, J. P. (1994) Chemical Ligation Approach To Form a Peptide Bond between Unprotected Peptide Segments. Concept and Model Study. *J. Am. Chem. Soc.* **116**, 4149–4153
172. Canle, M., Lawley, A., McManus, E. C., and O'Ferrall, R. A. M. (1996) Rate and equilibrium constants for oxazolidine and thiazolidine ring-opening reactions. *Pure Appl. Chem.* **68**, 813–818
173. Just, G., Young Chukg, B., Kim, S., Rosebery, G., Phillip, A., O U, B. Y., Chung, G., Gerald, K. R., and Ross, P. (1976) Reactions of oxygen and sulphur anions with oxazolidine and thiazolidine derivatives of 2-mesyloxymethylglyceraldehyde acetonide. *Can. J. Chem.* **54**, 2089
174. 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. *J. Am. Chem. Soc.* **113**, 3071–3079
175. Brooks, S. C., Adhikary, S., Rubinson, E. H., and Eichman, B. F. (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta - Proteins Proteomics*. **1834**, 247–271
176. Billman, J. H., and Diesing, A. C. (1957) Reduction of Schiff Bases with Sodium Borohydride. *J. Org. Chem.* **22**, 1068–1070
177. 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. *Nat. Struct. Mol. Biol.* **26**, 607–612
178. 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 Res.* **47**, 10388–10399
179. Buller, A. R., and Townsend, C. A. (2013) Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc. Natl. Acad. Sci.* **110**, E653 LP-E661
180. Jacobs, A. L., and Schär, P. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*. **121**, 1–20
181. Hu, P., Janga, S. C., Babu, M., Díaz-Mejía, J. J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., Chandran, S., Christopoulos, C., Nazarians-Armavil, A., Nasseri, N. K., Musso, G., Ali, M., Nazemof, N., Eroukova, V., Golshani, A., Paccanaro, A., Greenblatt, J. F., Moreno-Hagelsieb, G., and Emili, A. (2009) Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. *PLoS Biol.* **7**, e96–e96
182. Walker, G. C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli. *Microbiol. Rev.* **48**, 60–93
183. Fernández De Henestrosa, A. R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H., and Woodgate, R. (2000) Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol. Microbiol.* **35**, 1560–1572
184. Lewis, L. K., Harlow, G. R., Gregg-Jolly, L. A., and Mount, D. W. (1994) Identification of High Affinity Binding Sites for LexA which Define New DNA Damage-inducible Genes in Escherichia coli. *J. Mol. Biol.* **241**, 507–523
185. Fernández de Henestrosa, A. R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H., and Woodgate, R. (2000) Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol. Microbiol.* **35**, 1560–1572
186. Simmons, L. A., Foti, J. J., Cohen, S. E., and Walker, G. C. (2008) The SOS Regulatory Network. *EcoSal Plus*. **2008**, 10.1128/ecosalplus.5.4.3
187. Courcelle, J., Khodursky, A., Peter, B., Brown, P. O., and Hanawalt, P. C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics*. **158**, 41–64
188. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* 10.1038/msb4100050
189. 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. *EMBO J.* **15**, 3442–3447
190. Kiianitsa, K., and Maizels, N. (2013) A rapid and sensitive assay for DNA-protein covalent complexes in living cells. *Nucleic Acids Res.* **41**, e104
191. Aldred, K. J., Payne, A., and Voegerl, O. (2019) A RADAR-Based Assay to Isolate Covalent DNA Complexes in Bacteria. *Antibiot. (Basel, Switzerland)*. **8**, 17
192. Vaz, B., Popovic, M., and Ramadan, K. (2017) DNA–Protein Crosslink Proteolysis Repair. *Trends Biochem. Sci.* **42**, 483–495
193. 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., Mimoso, C. A., Low, E., Patel, K. J., and Walter, J. C. (2019) TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature*. **567**, 267–272
194. Larsen, N. B., Gao, A. O., Sparks, J. L., R€, M., Walter, J. C., and Duxin Correspondence, J. P. (2019) Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts Replication-coupled DPC Proteolysis SPRTN-mediated Proteasome-mediated Ubiquitin chains SPRTN Proteasome Polymerase approach CMG helicase CMG helicase ssDNA DNA polymerase DNA polymerase. *Mol. Cell*. **73**, 574-588.e7
195. Duxin, J. P., Dewar, J. M., Yardimci, H., and Walter, J. C. (2014) Repair of a DNA-Protein Crosslink by Replication-Coupled Proteolysis. *Cell*. **159**, 346–357
196. Vaz, B., Popovic, M., Newman, J. A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A. N., Vendrell, I., Fischer, R., Torrecilla, I., Drobnitzky, N., Freire, R., Amor, D. J., Lockhart, P. J., Kessler, B. M., McKenna, G. W., Gileadi, O., and Ramadan, K. (2016) Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. *Mol. Cell*. **64**, 704–719
197. Hershko, A., and Ciechanover, A. (1998) The Ubiquitin System. *Annu. Rev. Biochem.* **67**, 425–479
198. Fielden, J., Ruggiano, A., Popović, M., and Ramadan, K. (2018) DNA protein crosslink proteolysis repair: From yeast to premature ageing and cancer in humans. *DNA Repair (Amst).* **71**, 198–204
199. Borgermann, N., Ackermann, L., Schwertman, P., Hendriks, I. A., Thijssen, K., Liu, J. C., Lans, H., Nielsen, M. L., and Mailand, N. (2019) SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J.* **38**, e101496
200. Serbyn, N., Noireterre, A., Bagdiul, I., Plank, M., Michel, A. H., Loewith, R., Kornmann, B., and Stutz, F. (2020) The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. *Mol. Cell*. **77**, 1066-1079.e9
201. Yip, M. C. J., Bodnar, N. O., and Rapoport, T. A. (2020) Ddi1 is a ubiquitin-dependent protease. *Proc. Natl. Acad. Sci.* **117**, 7776 LP – 7781
202. Duxin, J. P., Dewar, J. M., Yardimci, H., and Walter, J. C. (2014) Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell*. **159**, 346–357
203. Seetharam, S., Waters, H. L., Seidman, M. M., and Kraemer, K. H. (1989) Ultraviolet Mutagenesis in a Plasmid Vector Replicated in Lymphoid Cells from a Patient with the Melanoma-prone Disorder Dysplastic Nevus Syndrome. *Cancer Res.* **49**, 5918–5921
204. Luria, S. E., and Delbrück, M. (1943) Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*. **28**, 491–511
205. Di Noia, J. M., and Neuberger, M. S. (2007) Molecular Mechanisms of Antibody Somatic Hypermutation. *Annu. Rev. Biochem.* **76**, 1–22
206. Stavnezer, J., and Schrader, C. E. (2014) IgH Chain Class Switch Recombination: Mechanism and Regulation. *J. Immunol.* **193**, 5370–5378