UNDERSTANDING THE MECHANISMS OF HMCES DNA-PROTEIN CROSSLINK REMOVAL AND ITS IMPORTANCE IN HUMAN CELLS

By

Jorge Rua Fernandez

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
In partial fulfillment of the requirements
For the degree of

DOCTOR IN PHILOSOPHY

In Biochemistry

May 10, 2024
Nashville, Tennessee

Approved

Dr. Scott Hiebert
Dr. David Cortez
Dr. Brandt Eichman
Dr. James Dewar
Dr. Ian Macara

To Flor, Olimpio, Omar, Oliver, and Jorge

Maybe you don't know, but we've been working on this together.

ACKNOWLEDGMENTS

It is not easy to think about all the people who helped to reach this important moment in my academic life, so I will start with my mentors. I am grateful to my advisor David Cortez. Your guidance, support, and enthusiasm for science have boosted my motivation to do research through all these years. Thanks for teaching me how to think critically and challenge my ideas in every lab meeting. Dave wants all his students to be successful, and I knew that since I joined his lab. He made all the accommodations to make the laboratory accessible for someone with a physical disability like me. He provided me with a very collaborative and safe environment where I only had to worry about doing science. I know that reaching the end of my PhD journey is when things got even more difficult and not only for me. Thus, I need to make a special Thank you, Dave, for your support outside of the lab. Your visits at the hospital, your updates in the paper, for bringing me a pumpkin for Halloween, and remind me that life does not need to change that much even if it gets tougher. It's been a long road and I'm happy I did it under your guidance Dave, you are a good mentor.

I also want to thank my thesis dissertation committee: Dr. Hiebert, Dr. Macara, Dr. Eichman, and Dr. Dewar. All your comments, suggestions, and ideas helped me prioritize experiments and taught me the importance of generating clean and robust data. I want to give a special shout-out to Ian Macara, who welcomed me to Vanderbilt in his lab before I was even thinking about doing grad school. Ian showed me all the potential of doing research at Vanderbilt and I've been very happy to have him during my PhD work as well. To Dr. Hinton, who mentored me in a small group where I could explore other academic topics related to mental health, diversity, and inclusion. Back in Peru, I want to thank Dr. Michael Talledo, my mentor and friend who gave me my first research experience. Michael, you showed me that having a disability is just one more feature in myself and that I can do everything I put my mind on. Thanks for letting me be myself in your lab in Lima, I will never forget the scientific discussions, the singing, us playing music on the lab birthdays, and hiking with you is an experience that I always carry with me.

To all the members of the Cortez lab, you have helped me in unique ways during all these years that helped me to improve my work substantially. Some special thanks to the people who welcomed me in the lab: Petria, my HMCES partner and friend who always encouraged me to enjoy life while doing research. Archana – Thank you for all your support and reminding me that

I was worthy even when science did not help. Taha – I enjoyed my Arabic classes while working on experiments that made us stay in the lab very late. Kavi and Wenpeng, you guys are the smartest scientists and also the nicest people in town. To the current people in the lab: Gaby – your light shines along to your intelligence, thanks for your constant support and being a friend in and outside the lab. Celeste – we were meant to meet. Thanks for your immense support during the great times and the difficult ones. Rahul – I have enjoyed the singing sessions at the TC and I promise to watch that movie. Madison and Matthew – thanks for your ideas and for helping me with protocols while dealing with purified proteins. Sander – thanks for letting me mentor you, you are an intelligent nice person. Nancy, thanks for keeping the lab together, and Yasmine, I look forward to reading your work in the future.

To my family, my mom Flor and dad Olimpio, who have been always there for me. Mom – I'm inspired by you. You managed to succeed in your scientific career in Peru and still be a lovely mom to your three kids. There is nothing you cannot do and I reflect myself on you. Thanks for always reminding me of my worth. Dad – you always support my ideas, my goals, my music, and everything that makes me happy. I am so grateful to both of you, who taught me to walk when I was a baby, and helped me to walk again as a teenager after my surgery, and now, while I'm writing this thesis, you are still here, willing to keep walking next to me. To my brother, Oliver. Since I was a kid, you've been always taking care of me, until now, constantly checking if I'm eating, sleeping, taking my medicine, and asking about the discoveries I'm making. Your curiosity and passion for your work are a constant inspiration. To my brother Omar – thank you for making me feel at home whenever I talk to you and for reminding me that there is life also happening outside the lab. Living in Peru, you guys gave me a home full of love, support, and laughs. I know you always have my back.

Coming to study in the USA allowed me to gain a family of friends. I want to thank Lihn, for her incredible support. Thanks for singing with me. To Andrea, my role model who taught me to celebrate the small lab victories. To Caye, Mariana, Samika, Becky, Alisson, Yelena, Diego, Angel, Angelo, Julissa. You guys have made my Grad school life so full of joy. To Andrew Peters, for his constant support and sushi dinners. To Cindy Young for the funny friend chicken meals. To Justin Danyluck, for being a great friend from the distance. Everyone in Nashville who was willing to give me a hand to make my life easier, thanks.

I started doing science in Peru and I got the chance to come to the USA thanks to an internship program for Peruvian undergrads, the REPU program. I want to thank Abel, Kenyi, and Sofia, who allowed me to pursue science outside my country. They created this beautiful program that I hope to continue to help in the future. To my friends from the Talledo lab in Lima: Fran, Javier, Romina, Andres, thanks for the wonderful experience of doing research with friends. To Juanita, Arturo, Johan, and Alfredo, thanks for always believing in me.

Finally, I want to thank myself. It is a long journey and sometimes I feel I do not give myself too much credit. Life is not easy, it is full of everything: failed experiments, lost relatives, and spine surgeries, but it is also full of karaoke nights, group dinners, dancing parties, promising data, great research talks, and a published paper. I want to thank myself for giving me the chance to go through all those experiences and be here, grateful, for all what I accomplished.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

5' dRP	5’ 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
AMP	adenosine monophosphate
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
CPT	camptothecin
DPC	DNA-protein crosslink
DSB	double-strand breaks
dsDNA	double-stranded DNA
DTT	dithiothreitol
EMSA	electrophoretic mobility shift analysis
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
LMA	low melting point agarose
MBD4	methyl-CpG-binding domain protein 4
MCM	minichromosome maintenance protein complex

MMS	methyl methanesulfonate
MetAPs	methionine aminopeptidase
MPG	N-methyl-purine DNA glycosylase
NEIL	endonuclease VIII-like
NER	Nucleotide Excision Repair
NHEJ	non-homologous end joining
NTHL1	endonuclease III-like
NLS	nuclear localization sequence
OGG1	8-oxoguanine glycosylase
PARP	poly(ADP-ribose) polymerase 1
PIKK	Phosphatidylinositol 3-kinase-related kinases
PIP	PCNA interacting motif
PNKP	polynucleotide kinase phosphatase adenosine monophosphate
RNA polymerases	RNAPs
RNAPII	RNA polymerase II
RT	room temperature
SB	sample buffer
SHM	somatic hypermutation
Shu2	SWIM domain-containing protein
SPRTN	Spartan, C1orf124
SRAP	SOS response peptidase
ssDNA	single-stranded DNA
TC-NER	transcription-coupled nucleotide excision repair
TEC	transcription elongation complex
TDG	thymine DNA glycosylase
THF	tetrahydrofuran
TLS	Translesion synthesis
TRAIP	TRAF interacting protein
UBA E1	Ubiquitin-like modifier-activating enzyme 1
UDG/UNG	uracil DNA glycosylase
UGI	uracil glycosylase inhibitor
UV	ultraviolet

CHAPTER I

INTRODUCTION

The human genome contains billions of DNA base pairs that must undergo a complete and accurate replication in each cell cycle. However, cellular DNA is constantly damaged by endogenous and exogenous lesions, which accumulation can lead to neurodegenerative diseases, premature aging, and cancer. A prevalent lesion in mammalian DNA is apurinic/apyrimidinic (AP, or abasic) sites, which is the loss of a single base in a nucleotide sequence. Approximately, ~10,000 -20,000 AP sites are formed in each human cell, every day[1], becoming the most frequent lesion in mammalian cells. Abasic sites can occur spontaneously or by enzymatic activity. In either case, the consequences of this lesion can go from the stall of DNA polymerases to forming more complex lesions such as DNA-interstrand crosslinks (ICLs) or DNA-protein crosslinks (DPCs)[2]. For that reason, it is important to know more about their formation, consequences, and how to resolve AP sites while maintaining the accuracy of DNA replication.

Formation of abasic sites

Thousands of abasic sites are formed in each human cell every day. There are different potential sources for AP site formation: They can occur spontaneously or by base modification which can lead to its removal by different mechanisms.

Spontaneous formation of abasic sites.

Abasic sites can be formed by spontaneous hydrolysis of the N-glycosidic bond. Purines are lost on a higher frequency than pyrimidines[3] (Fig 1.1A). Additionally, base loss is four times more frequent in single-stranded DNA (ssDNA) than in double-stranded DNA (dsDNA)[4]. Thus, sites of DNA replication are prone to abasic site formation.

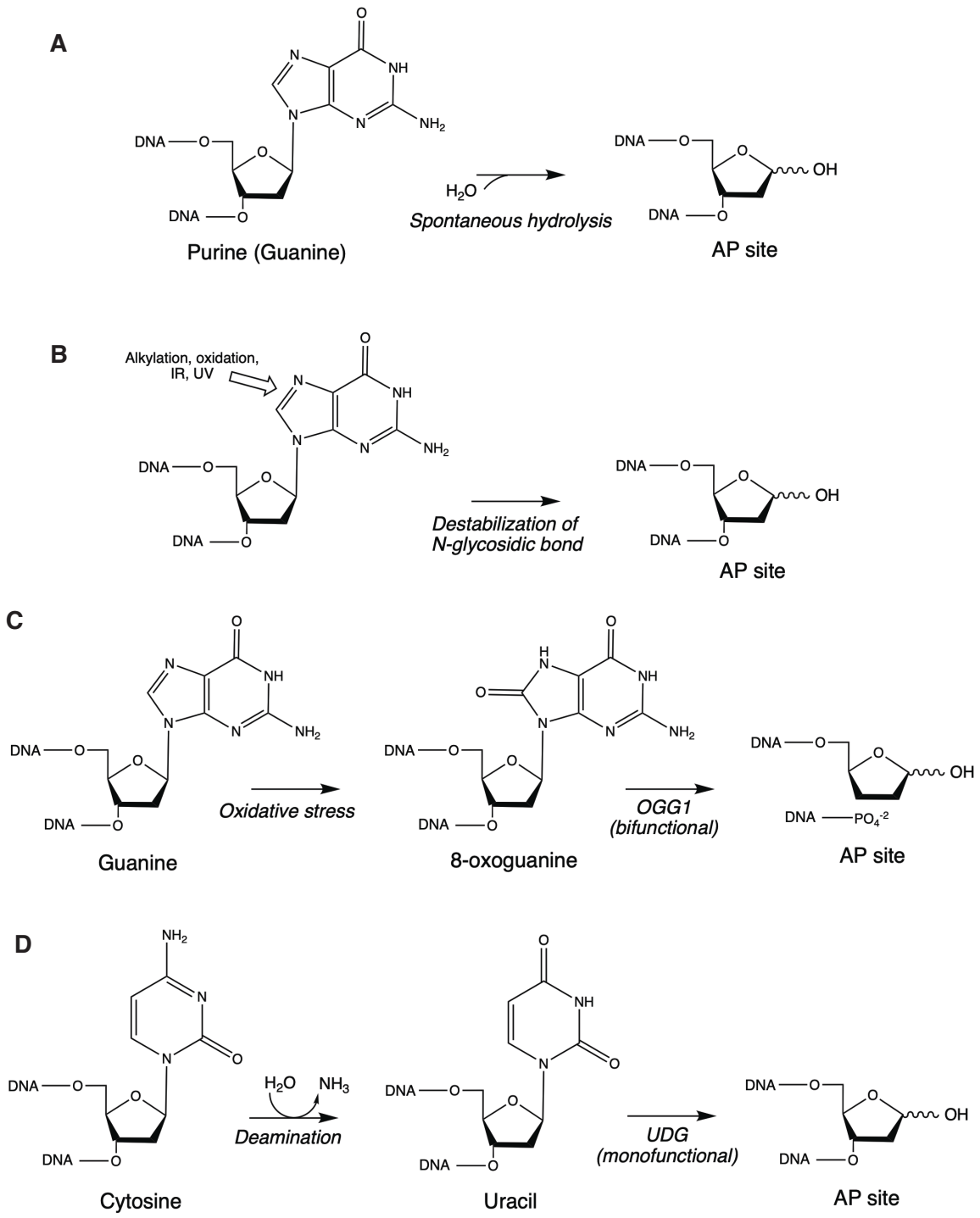

Figure 1.1 Formation of AP sites.

(A) Spontaneous loss of the base generates AP site. (B) Base modifications induce destabilization of the N-glycosidic bond and produce an AP site. (C) Enzymatic activity of bifunctional glycosylase OGG1 that removes oxidized bases and cleaves the 3' end of the AP site. (D) Cytosine deamination generates uracil which is removed by monofunctional glycosylase UDG and forms an AP site.

Base modification.

Alkylating agents are chemical compounds that can transfer alkyl groups onto the high nucleophilic nitrogen or oxygen atoms of DNA bases[5]. These can be endogenous alkylating compounds, as the suggested methyl donor S-adenosylmethionine (SAM)[6], or exogenous agents. There are five major classes of alkylating agents: alkyl sulfonates, triazene, nitrosoureas, nitrogen mustards, and aziridines, which differ in the number of reactive sites, the type of alkyl groups transfer, or the particular chemical reactivity[5]. Base alkylation can weaken the N-glycosidic bound by generating a positive charge in the ring. This protonation is stabilized via electron resonance, leading to the bond cleavage and the generation of an AP site[7] (Fig 1.1B). Moreover, alkylation chemotherapy is a major therapeutic modality for different types of tumors. Some anticancer therapies include methanesulfonate esters, nitrosoureas, and nitrogen mustards, which generate AP sites[8]. For that reason, it is highly important to understand more about its repair pathways in cancer cells.

Reactive oxygen species (ROS) such as hydrogen peroxide, superoxide, and hydroxyl radicals are by-products of cellular metabolism that can modify all four DNA bases in mammalian cells. The most common oxidation product is 7,8-dihydro-8-oxoguanine (8-oxoG)[9] because of the low redox potential of guanines. Free radicals can also attack pyrimidines thymine and cytosine to generate oxidized base products that can destabilize the N-glycosidic bond or be recognized by DNA glycosylases that then can form abasic sites. Ionizing radiation (IR) produces hydroxyl radicals that directly attack and destabilize the N-glycosidic bond. Ultraviolet (UV) radiation can induce photoproducts such as thymine glycol, alkali labile purine lesions, and other pyrimidine hydrates that can lead to AP site formation[10].

Enzymatic activity.

DNA glycosylases are DNA enzymes that recognize and excise modified bases as an initial step on the BER pathway. Monofunctional DNA glycosylases, sever the *N*-glycosidic bond between the damaged base and the phosphodiester DNA backbone, creating an AP site. For example, N-methyl-purine DNA glycosylase (MPG) can recognize and excise 3meA, 7meG, and 3meG modifications. Meanwhile, bifunctional DNA glycosylases perform base damage removal and in addition, they incise the backbone 3' to the lesion to create a single-nucleotide gap. For example,

8-OxoG DNA glycosylase 1 (OGG1) removes 8-oxoG and cleaves the 3'-end of the AP site[11,12] (Fig 1.1C). NEIL1, NEIL2, and NEIL3 recognize oxidized pyrimidines and ring-opened purines[13].

Mammalian uracil DNA glycosylase (UNG) is monofunctional and removes uracil and 5-Fluorouracil (5-FU) residues in ssDNA and dsDNA[11]. Uracil is a pyrimidine frequently found in RNA; however, it can be integrated into DNA due to the misincorporation of dUMP instead of dTTP during DNA replication or as a result of cytosine deamination. Uracil misincorporation happens because of nucleotide pool imbalance, due to the much higher concentration of rNTPs than dNTPs in cells. Ribonucleotide incorporation into DNA can occur due to enzymatic errors during DNA synthesis by DNA polymerase α (Pol α), delta (Pol δ), and epsilon (Pol ε) in human cells[14]. Cytosine deamination is another source of uracil incorporation in DNA. This deamination can occur spontaneously in human cells or by the activity of cytosine deaminases. The AID/APOBEC (activation-induced deaminase/apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of DNA cytosine deaminases specifically act on ssDNA[15]. These enzymes can create a mutation signature in cancer cells which are strongly enriched on the lagging strand during DNA replication[16]. Thus, the presence of uracil in the lagging strand by cytosine deamination can lead to later removal by UNG and the formation of abasic sites (Fig 1.1D). Taken together, there are many ways AP sites can be formed and their presence during DNA replication can have multiple outcomes.

Consequences of abasic sites

Increase in strand breaks

AP sites increase the frequency of strand breaks because of their chemical structure. The absence of the base generates a sugar anomer in an equilibrium between a closed-ring furanose (99%) and an open-ring aldehyde (1%)[17] (Fig 1.2A). The aldehyde form confers high sensitivity to alkaline conditions that can induce strand breaks. Mild alkaline conditions make the abasic site undergo β-elimination where the 3' phosphodiester bound of the aldehyde form is hydrolyzed to generate a 3'-terminal unsaturated sugar and a terminal 5'-phosphate (Fig 1.2B). In a more drastic alkaline medium, a second elimination reaction (δ-elimination) occurs[17]. Other molecules such as thiols, primary and secondary amines, polyamines, basic proteins, or aldehyde reagents can promote β-elimination[18].

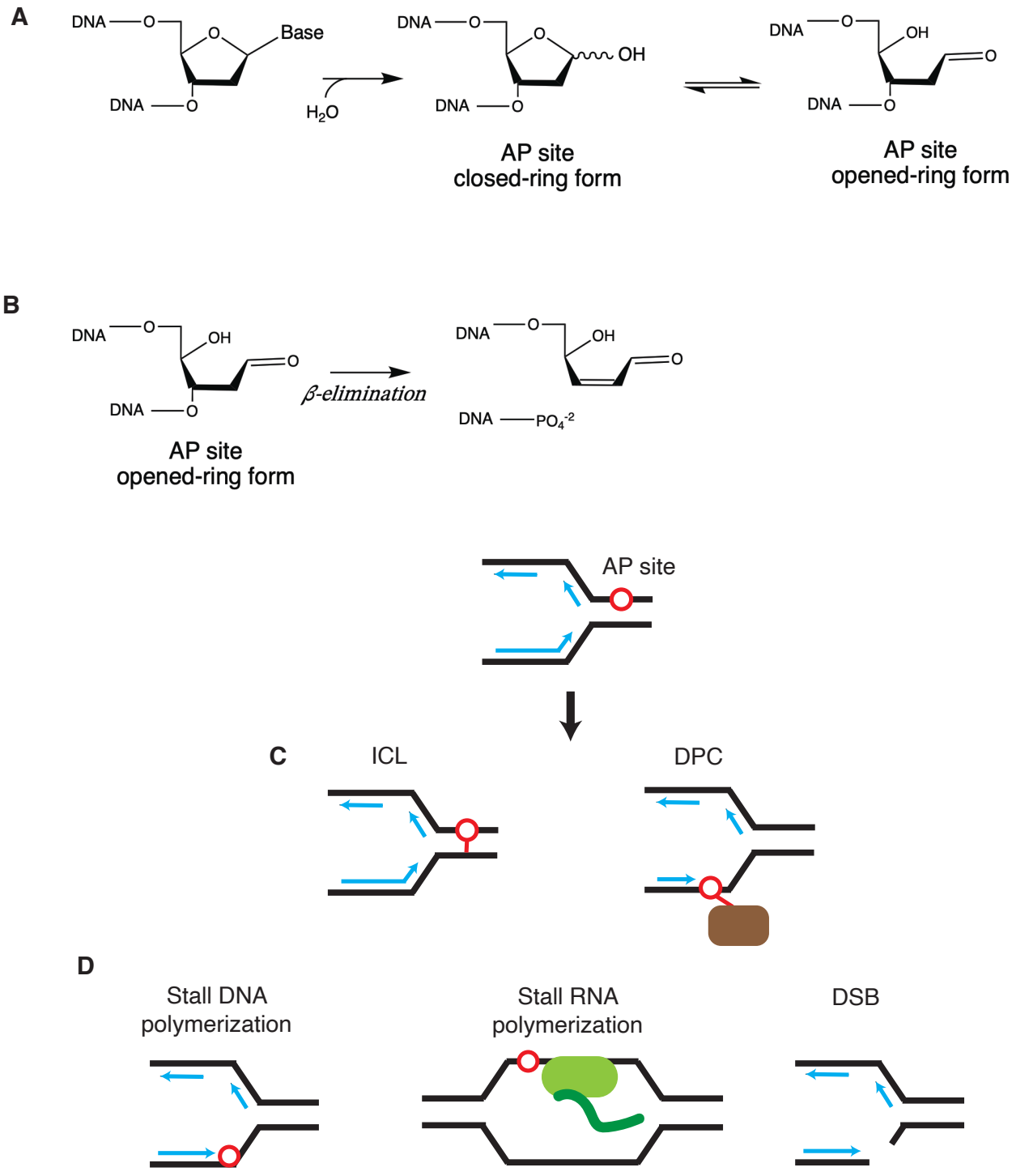

Figure 1.2. Consequences of AP sites.

(A) The AP site stays in equilibrium between a closed-ring furanose and an open-ring aldehyde. (B) Alkaline conditions induce β-elimination of an AP site and generate a strand break. (C) AP sites can form ICL and DPC (brown box) which can stall replication. (D) AP sites at the replication fork can stall DNA and RNA polymerization (green circle) and induce DSBs.

Formation of interstrand and DNA-protein crosslinks.

Endogenous processes, anticancer drugs, and toxins can generate interstrand DNA-DNA crosslinks. The common feature of an ICL is that two nucleotides on opposite strands are covalently joined, which becomes a major barrier to DNA replication (Fig 1.2C). In the presence of abasic sites, the AP aldehyde reacts with the exocyclic amino group of nucleobases such as guanine to form an AP-ICL[19]. Even if it is not DNA sequence-specific, AT-rich regions in the surrounding AP site lead to higher rates of AP-ICL formation[20]. Glycosylase NEIL3 is known to unhook this crosslink, which generates another AP site intermediate for further repair[21].

The open-ring form of the AP site can also react with primary amines in proteins to generate DPCs[22] (Fig 1,2C). Several repair enzymes such as PARP1[23], Polymerase β[24], and KU[25] can form DPCs with AP sites due to the formation of a Schiff-base intermediate, which can be resolved by β-elimination but that can be trapped in the presence of $NaBH_4$. Other well-documented DPCs formed in DNA are the ones formed by topoisomerase; however, they do not require the presence of AP sites and they are trapped using mechanism-based compounds. In general, the presence of DPCs in the cell is highly deleterious because they represent bulky lesions that can stop replication, and transcription, and induce strand breaks.

Stall of DNA polymerases and mutagenesis.

Abasic sites can appear as very small lesions, however, they can block replicative (Pol α, δ, and ε) and transcription polymerases[26,27]. Moreover, if the replication fork encounters an AP site on the template strand, that can lead to DSB due to the activity of AP endonucleases such as APE1 (Fig 1.2D). AP sites can completely stop or just slow down polymerization. RNA polymerases (RNAPs) from bacteria and eukaryotes can bypass AP sites with moderate transcriptional pause because the AP site is not properly loaded into the +1 site of the active center, preventing full translocation of the transcription elongation complex (TEC)[28]. During DNA replication, in *E. coli* and human cells, AP sites are severe blocks that can stall polymerases and promote mutagenesis due to predominately incorporating an adenine nucleotide (A) opposite to the lesion[2,29,30]. This "A-rule", predicts that depurinations and depyrimidinations will produce transversions and transitions, respectively. During transcription, RNA polymerases also introduce an A opposite to the AP sites[28].

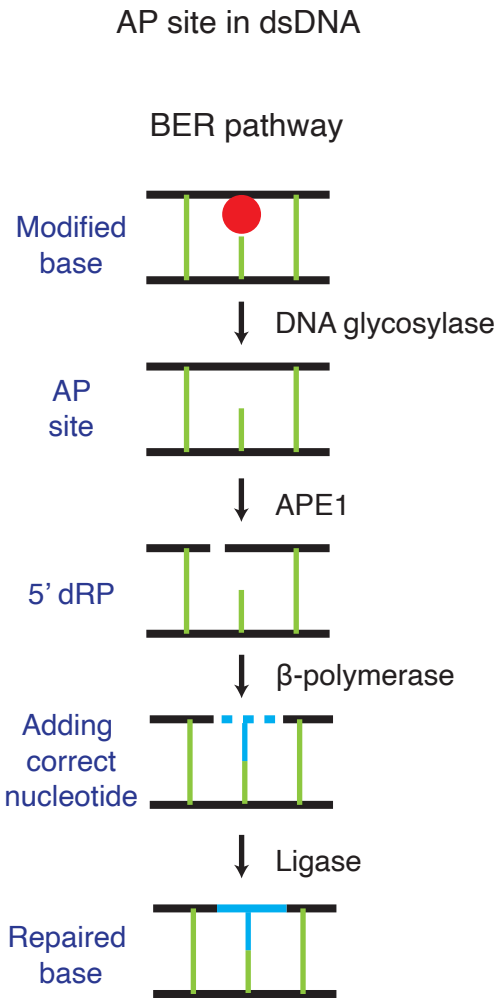

Figure 1.3. Removing AP sites located in dsDNA.

A diagram of the short-patch BER showing a modified base (red circle) that is removed by a DNA glycosylase, which generates an AP site. Endonuclease APE1 cleaves 5' to the AP site generating a 5'dRP, which then is removed by Polβ which also introduces a nucleotide (blue line) using the complementary strand as a template. This is an error-free repair pathway to remove AP sites.

The cells bypass AP sites by the activity of Translesion DNA synthesis (TLS) polymerases, which are low-fidelity polymerases that introduce mutations while overcoming abasic sites, so they represent an error-prone mechanism of damage-tolerance[31]. TLS polymerase activity involves a 2-step process: 1. A polymerase inserts a nucleotide across the lesion and 2. Another polymerase extends past the lesion[32]. Sometimes, depending on the lesion, some TLS polymerases can do both[33]. The human Y-family of TLS polymerases are generally insertors across from the AP site lesions and include pols η, ι, κ, and REV1[34]. Catalytic efficiency experiments showed that pol η inserted T and A opposite to the AP site for later full length extension and Pol ι and REV1 readily incorporated one base opposite AP sites but then stopped[29]. B-family polymerases such as eukaryotic Polζ usually perform extension synthesis from the inserted nucleobase opposite to the AP site[32]. However, it can also act as both the inserter and extender, for example when replicating over thymine glycol lesions generated by free-radicals[33]. Other TLS polymerases such as Polλ and Polμ are not generally involved in overcoming AP sites. Taken together, the stall of DNA polymerases can lead to increased mutagenesis.

Pathways to repair AP sites

Repair of AP sites located in dsDNA

Base excision repair. The BER pathway repairs single modified bases located in dsDNA. A damage-specific DNA glycosylase recognizes and removes the damaged base, leaving an abasic site. Next, the major AP-endonuclease in mammalian cells, APE1 cleaves 5' to the AP site, forming a 5'-deoxyribosephosphate (5'-dRP) moiety, which is removed by the dRP lyase activity of Polβ that simultaneously fills the gap using the second strand as a template[35] (Fig 1.3). When BER is initiated by NEIL DNA glycosylases, they possess β,δ-elimination activity which creates a single-nucleotide gap containing 3' and 5' -phosphate ends. The 3' -phosphate is removed by the polynucleotide kinase phosphatase (PNKP) before nucleotide insertion by Pol β occurs[36]. Then, the remaining nick in the phosphodiester backbone is sealed by XRCC1-Lig IIIα to complete a short-patch BER. A long-patch BER occurs when the 5'-dRP moiety is resistant to Pol β activity and switches to Pol δ/ε. These DNA polymerases add 2 to 8 nucleotides into the single nucleotide gap, creating a 5'-flap that can be recognized and excised by the endonuclease activity of FEN-1, in a PCNA-dependent manner[35,37]. Finally, the remaining nick in the DNA backbone is then repaired by DNA ligase I, to complete long-patch BER[35].

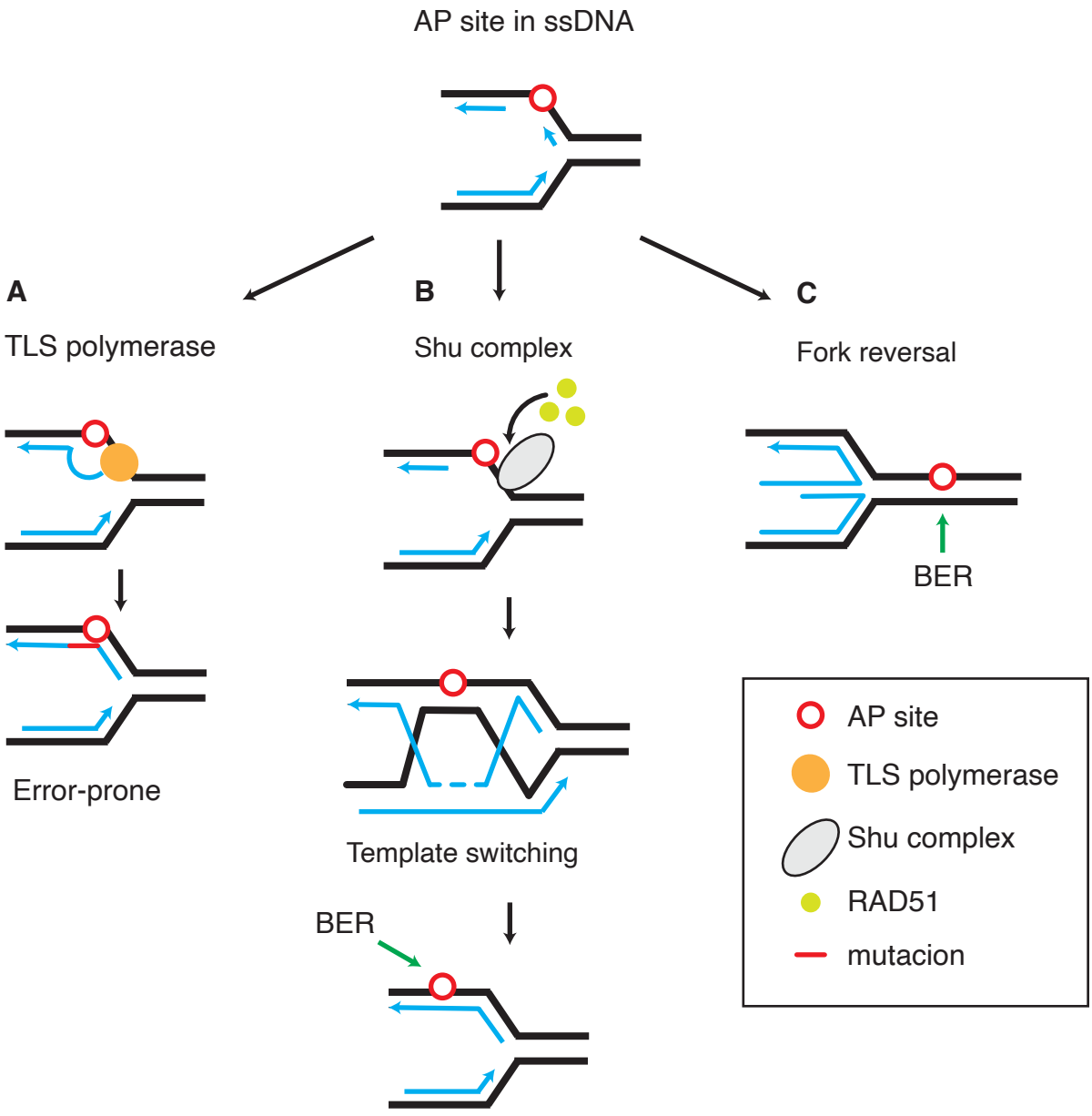

Figure 1.4. Tolerance of AP sites located in ssDNA.

(A)TLS polymerase can overcome the ssDNA AP site by introducing mutations during polymerization. (B) The Shu complex binds the AP site and by recruiting RAD51 induces template switching to localize the lesion in the context of dsDNA. This allows the BER pathway to promote an error-free mechanism. (C) Fork reversal can localize the AP site in dsDNA to promote repair by BER.

Nucleotide excision repair. AP sites in dsDNA can also be repaired by the nucleotide excision repair pathway (NER). There are two mechanisms of DNA damage detection by NER: the global genome NER (GG-NER) or the transcription-coupled NER (TC-NER). The first one detects the lesion by helix distortions (disturbed base pairing) due to structural changes of nucleotides, and the second one is activated when RNAS polymerase II (RNA Pol II) is stalled during transcript elongation by a lesion in the template strand. GG-NER enzymes generate single-strand nicks in the DNA 5' and 3' to damage to generate a 25 to 30 single-strand nucleotide fragment. This generates a gap that is repaired by the action of DNA polymerase and DNA ligase[38]. The TC-NER is a major repair mechanism for AP sites in cells because they stall RNA Pol II when they are present in the actively transcribed strand[39]. This generates transcriptional pauses and the preferred incorporation of adenosine monophosphate (AMP) opposite to the AP site[28], possibly generating transcript errors. Furthermore, highly transcribed sequences are more susceptible to AP site accumulation and are the source of transcription-associated mutagenesis[40]. The final steps of excision, repair, and ligation remain the same as GG-NER.

Tolerance of AP sites located in ssDNA

TLS bypasses AP sites. Since an AP site can stall replication polymerases, they can switch to TLS polymerases to overcome AP sites located in ssDNA. However, they can introduce mutations because they lack the intrinsic 3'-to-5' exonuclease domain of replicative DNA polymerases that function to proofread the newly synthesized strand[41]. In vertebrates, TLS bypass of AP sites is very frequent during somatic hypermutation (SHM) in B cells, a process that increases the variability of immunoglobulin genes in these cells by inducing the formation of AP sites. TLS polymerases can be detrimental to DNA replication fidelity, nevertheless, mutagenesis increases variability in the Ig locus, which is beneficial for the immune response[42]. Thus, inaccurate synthesis due to AP sites is beneficial for genetic diversity and evolutionary fitness.

The Shu Complex. This is a recently described pathway mediated by a yeast heterotetrametric complex containing the protein Shu2 and the RAD51 paralogs Csm2, Pys3, and Shu1. Rosenbaum and collaborators reported that Csm2 is recruited to chromatin when abasic sites accumulate[43]. Moreover, Csm2-Psy3 suppresses lagging strand AP site mutation and can bind to abasic site analogs (tetrahydrofuran, THF) in double-flap DNA, more tightly when THF is located at the 5' oligonucleotide, at the junction[43]. Thus, they suggested a model where right after the fork encounters an AP site, the Shu complex binds the AP site to promote Rad51 filaments

formation and template switching, which uses the undamaged information on one sister chromatid to generate duplex DNA with the strand that contains the AP site (Fig 1.4B). This requires a strand invasion step dependent on homology recombination (HR) proteins such as RAD51 or RAD52[44], that can place the AP lesion in the context of dsDNA for BER. Additionally, Shu mutant yeast is sensitive to MMS-induced replication-blocking lesions and presents an elevated mutation frequency, exhibiting substitution patterns consistent with TLS activity bypassing AP sites and three-methyl cytosine (3meC)[45]. Thus, the Shu complex seems to promote error-free lesion tolerance for single-strand-specific alkylation lesions that can promote AP sites.

Fork reversal. In the presence of replication stress, fork reversal promotes replication-coupled DNA repair or bypass of DNA damage[44]. During this process, fork reversal is regulated by multiple proteins that promote the reannealing of the nascent DNA strands to convert a three-way junction to a four-way junction (chicken foot), which can place a lesion in the context of dsDNA for repair via BER at a later time (Fig 1.4C). Fork reversal is thought to be frequent in humans but it is still unknown if it is promoting an error-free pathway to repair ssDNA AP sites.

HMCES, a new mechanism to process AP sites

Discovering HMCES

Human HMCES (5-hydroxymethylcytosine, embryonic ES-cell-specific), previously known as C3Orf37, is a protein present in all domains of life, from bacteria to humans[46,47]. HMCES contains a single domain called SOS Response Associated Peptidase (SRAP) domain[46,48]. Upon DNA damage, the bacterial SOS response is activated and involves high-fidelity repair mechanisms, followed by low-fidelity, damage tolerance pathways, which include error-prone translesion DNA polymerases PolII (polB), PolIV (dinB), and PolV (umuC, umuD). Thus, the bacterial SOS pathway is an inducible DNA damage repair system that is essential for bacterial adaptation, pathogenesis, and diversification[49]. Arvind and colleagues performed a sequence and structural analysis of bacterial genomes deposited in the Genbank database and identified a putative thiol peptidase that was linked to several genes involved in the SOS response (taking its name from there). This SRAP domain contains very conserved residues predicted to catalyze auto-proteolysis, however, this thiol peptidase activity was not verified experimentally[50].

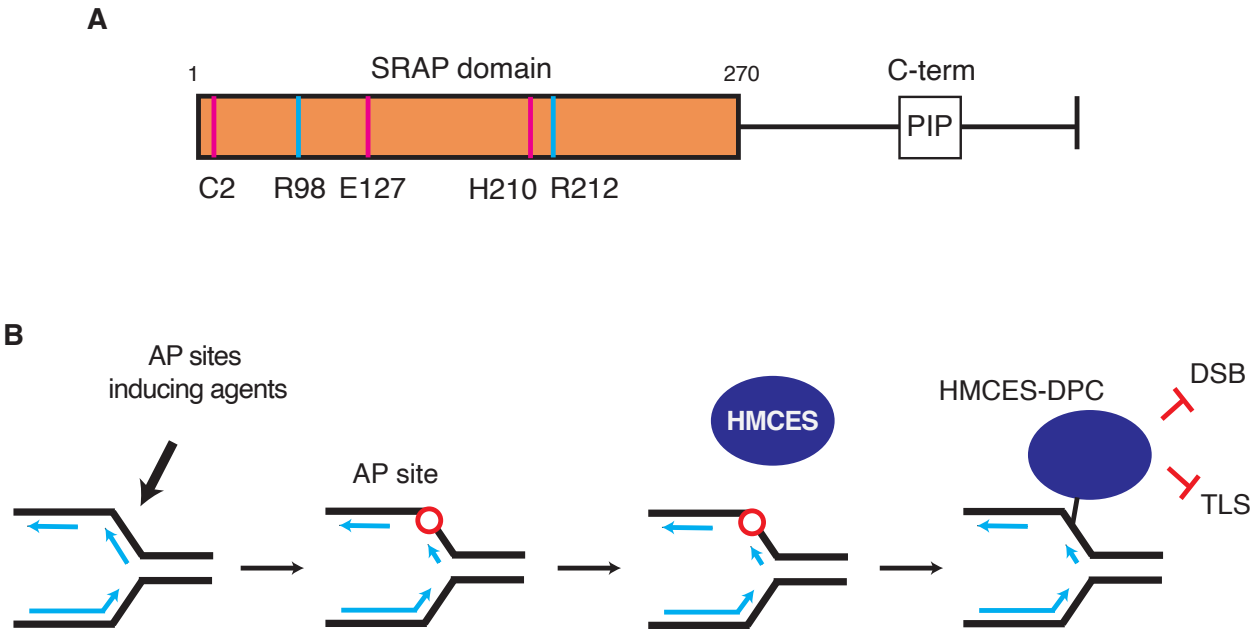

Figure 1.5. HMCES shields ssDNA AP sites.

(A) Diagram of HMCES's SRAP domain depicting highly conserved residues important for DNA binding (blue lines) and for the protein crosslink (magenta lines). The C-terminal region contains a PIP box for PCNA interaction. (B) Proposed model for HMCES-DPC formation at the replication fork while preventing DNA damage by DSBs and TLS-induced mutagenesis.

Later on, HMCES was identified as a reader for 5-hydroxy-methylcytosine (5hmC) and its oxidized derivatives in mouse embryonic stem (ES) cells[51], but it has not been directly tested. Kewon and collaborators also showed that the SRAP domain promotes autoproteolytic cleavage, shows endonuclease activity, and its expression reduced the induced 5hmC levels in ES cells[48]. Taken together, initial findings suggested that HMCES acts as an epigenetic reader of Tet-oxidized derivatives of 5mC and its SRAP domain functions as a peptidase and endonuclease.

The Cortez lab, using the isolation of proteins on nascent DNA (iPOND)[52] technique, identified HMCES as part of the replication fork in a proteomic screen in human cells[47,53]. HMCES interacts with the proliferating cell nuclear antigen (PCNA) protein and it is recruited to chromatin during the S-phase, suggesting a role during DNA synthesis. Contrary to previously reported, HMCES-deficient cells did not show changes in the levels of 5mC or 5hmC with very few changes in gene expression[47,54]. We did not observe any endonuclease activity either[47]. These observations led our lab to search for other functions of HMCES. Mohni and collaborators noticed that HMCES-deficient cells are hypersensitive to ATR (ataxia-telangiectasia and Rad3-related) inhibition. ATR is a kinase that safeguards the genome because it regulates DNA damage signaling and coordinates DNA repair[55,56]. Additionally, gene products that are required to survive ATR inhibition often function in DNA replication or repair[57], suggesting that HMCES is involved in these processes. Furthermore, Inactivating HMCES through RNA interference or gene disruption causes hypersensitivity to DNA damage agents such as IR, UV light, methyl methanesulfonate (MMS), and $KBrO_3$[47]. Interestingly, despite the different types of DNA damage these agents can cause, all of them produce a common lesion: AP site. These findings in our lab led us to search for other HMCES functions, one at the replication fork involved in maintaining genome integrity depending on the presence of AP sites.

HMCES crosslinks to AP site

HMCES is the only SRAP-containing protein present in humans and the SRAP domain constitutes the majority of the protein, meanwhile, it represents the entirety of the protein in its bacterial ortholog Yedk[46,58]. HMCES also has a C-terminal disorder region that contains a PCNA-interacting protein-box (PIP-box) that allows its interaction with PCNA (Fig 1.5A). The SRAP domain contains a highly conserved triad formed by cysteine, histidine, and glutamate (Cys2, His210, and Glu127 in HMCES)[46]. At the Cortez lab, we observed that Cys2 of HMCES allows

the crosslink of the protein with AP sites only if they are located in ssDNA[47]. A mechanism that is also present in the HMCES's bacterial ortholog YedK[58,59]. This crosslink requires prior binding to DNA, which is mediated by two highly conserved basic residues (Arg98 and Arg212). Formation of the HMCES DNA-protein crosslink (HMCES-DPC) shields the AP site from the activity of endonucleases such as human apurinic/apyrimidinic endonuclease (APE1) and the formation of DSBs, maintaining genome integrity[47] (Fig 1.5B). HMCES crosslinking to the AP site depends on removing the methionine at the N-terminal region (Met1). Removal of Met1 by HMCES autoproteolysis was previously reported[48], but possibly Met1 is also removed by essential methionine aminopeptidase (MetAPs) that are present in bacteria and humans and remove this N-terminal residue from newly synthesized peptides and proteins[60]. Either way, HMCES will crosslink only to natural AP sites because it needs the aldehyde group that is lacking in AP site analogs such as THF[47]. Since we could detect HMCES-DPC formation *In vitro* and cells, we decided that the best way to understand the molecular mechanism of this crosslink was by analyzing its structure.

Structure of HMCES crosslink to AP site

HMCES binds to ssDNA

HMCES SRAP and YedK share 29% sequence identity and 43% similarity, which makes them highly similar in structure (PDB IDs 5KO9[61] and 2ICU). In both cases, the SRAP domain crosslink to the AP site requires initial interaction with DNA through its SRAP domain. Crystallography data show that the SRAP domain contains a highly-conserved region, a positively charged cleft, and the active site where the crosslink takes place (Fig 1.6A). The basic cleft bends ssDNA directing it toward the Cys2 and positioning the AP site directly above the active site residues for the crosslink[58,61,62] (Fig 1.6B). The ssDNA-binding cleft contains surface-exposed residues Arg98 and Arg212 in HMCES, and mutations of these residues disrupt DNA binding and HMCES-DPC formation[47]. Moreover, hydrogen-bonding edges of all nucleobases are exposed to solvent, thus HMCES binding to the DNA is sequence-independent. There is a wedge motif blocking dsDNA access 5' to the AP site and the ssDNA backbone is severely kinked and twisted by 90° at the AP, both preventing the pairing of a complementary DNA strand around it, which disfavors SRAP binding to dsDNA[58]. Strikingly, structure data also shows that DNA adopts a B-form conformation after the active site which allows duplex DNA to be accommodated immediately adjacent to the 3'-side of the AP site. Thus HMCES-DPC can be present in ss/dsDNA junctions[58,63].

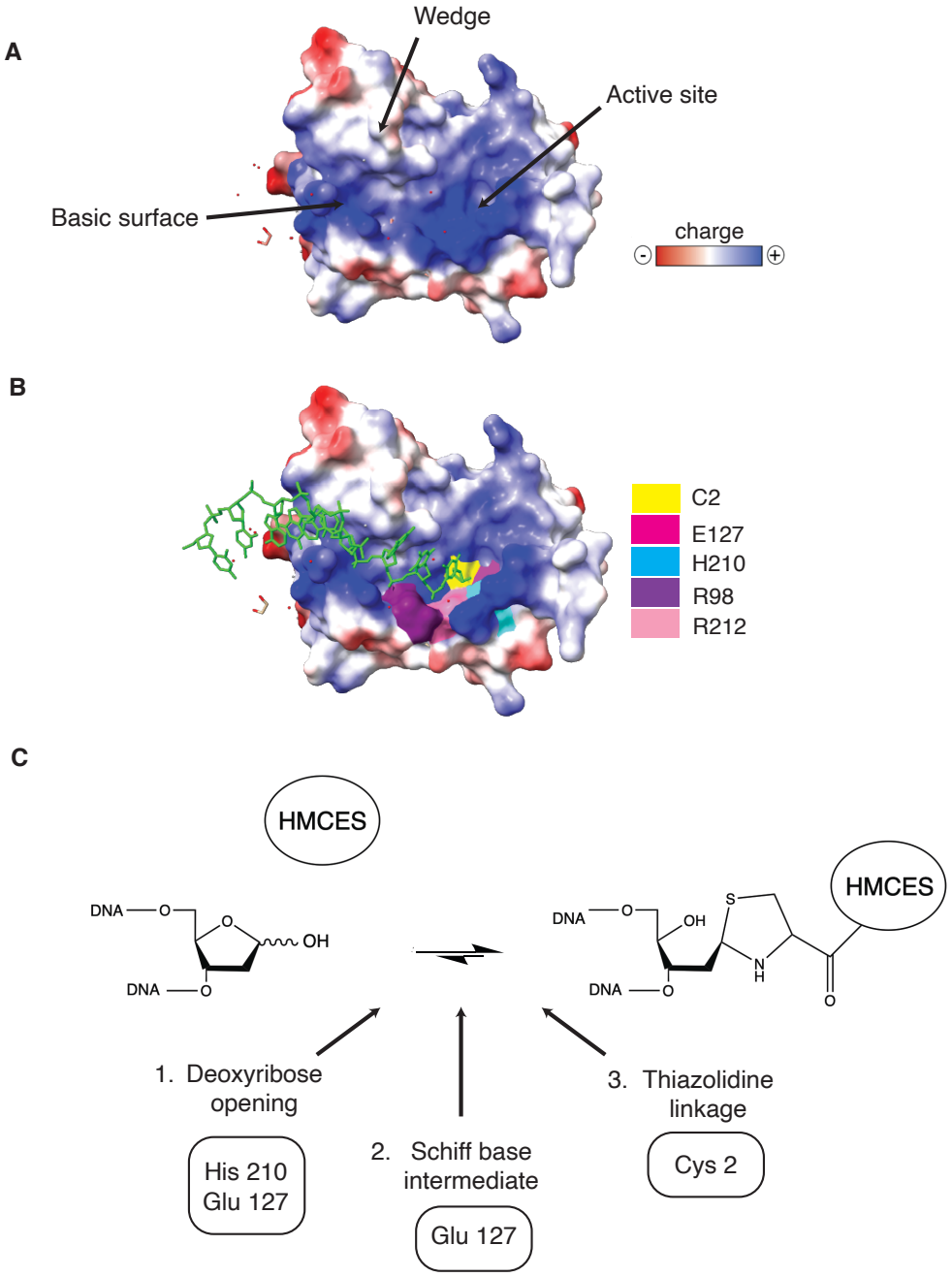

Figure 1.6 HMCES-DPC formation.

(A) Surface charge of HMCES SRAP domain showing its basic cleft, the wedge motif, and the active site (PDB: 6OEB[61]). (B) Structure of HMCES SRAP domain showing how the basic cleft bends ssDNA (green) towards the active site to locate the AP site above Cys2. Basic residues R98 (purple) and R212 (pink) provide binding to ssDNA and Cys2 (yellow), E127 (magenta), and His210 (Cyan) will mediate the crosslink (PDB: 6OEB[61]) (C) Three-step mechanism to form the thiazolidine linkage. Each step is mediated for the residues present at the active site (boxes) and maintains the equilibrium that favors HMCES-DPC formation.

Formation of the Thiazolidine linkage

At the active site of the SRAP domain, we find Cys2, His210, and Glu127 which mediate a thiazolidine ring formation with the AP site in a 3-step process (Fig 1.6C). First, the AP site ring opening requires His to act as an acid to protonate deoxyribose O4 on the AP site and the Glu acts as a general base to deprotonate the hydroxide at C1. Second, the same Glu drives Schiff base formation by acting as both, general acid and base to deprotonate Cys2 α-NH2 and to hydrolyze the hydroxyl at C1. Third, the thiazolidine formation is accomplished by a first nucleophilic attack of the Cys α-NH2 to C1 and a second one of the thiol group[59]. The equilibrium between Schiff base and thiazolidine favors the latter, preventing the β-elimination process of the Schiff base intermediate that leads to strand break[58,59]. A detailed diagram of this proposed catalytic mechanism was published by Paulin and collaborators to explain YedK DPC formation[59]. The thiazolidine linkage is greatly stable at 37 C and mutation of residue Cys2 (C2A) prevents HMCES-DPC formation[47]. Cells expressing HMCES C2A are as hypersensitive to DNA damage agents as HMCES KO cells, suggesting that HMCES-DPC crosslink is relevant for cell viability during DNA stress.

HMCES prevents the detrimental effects of AP sites

HMCES-DPC protects cells against DNA damage agents

Unlike damaging DPC formation with other proteins, the HMCES-DPC is thought to be protective and beneficial to the cell. HMCES-DPC prevents AP site cleavage thereby reducing DSBs[47,58] and decreasing mutation frequency[47,64]. As mentioned before, HMCES-deficient cells are hypersensitive to IR, MMS, and UV radiation. The re-expression of wild-type HMCES could largely rescue the IR hypersensitivity but not the catalytic (C2A) or DNA binding (R98E) mutants[47], suggesting that HMCES-DPC formation is important to prevent this type of DNA damage. Furthermore, cells depend on HMCES-DPC formation for their resistance against oxidative stress. We already noticed that HMCES null cells are hypersensitive to $KBrO_3$[47]. Cells expressing the C2A or the E127A mutant are also most sensitive to ROS such as H_2O_2. In 2020, the Durocher lab presented a genetic map from a CRISPR screen against DNA-damage agents, where genes whose mutation leads to their depletion from the cell population after genotoxin exposure have a negative NormZ (gene-level normalized Z-scores) value. They set a threshold with a NormZ less than -3 (with a false discovery rate lower than 15%) for mutated genes that cause sensitization to the genotoxin. HMCES shows a NormZ of -9,44 for $KBrO_3$, suggesting that HMCES is highly

required to protect the cells against oxidative damage[65]. Interestingly enough, in this study, HMCES is also clustered with other genes involved in the non-homologous-end joining (NHEJ) pathway, but its connection is still not completely understood.

HMCES protects cells from APOBEC-mediated AP sites

Genetically disrupted HMCES (HMCES knockout, KO) cells are hypersensitive to alkylating and oxidative reagents that induce base modification, which are removed by glycosylases and generate AP sites. Thus HMCES-lacking cells are sensitized to reagents that induce AP site formation[47,64]. APOBEC is a cytosine deaminase that also creates AP sites by the activity of UNG. APOBEC3A (A3A) is a major driver of mutagenesis in human cancers[66]. Inducible nuclear expression of APOBEC3A promotes HMCES localization to the chromatin where HMCES protects the cell from A3A-induced DSBs[67]. HMCES KO or depletion of the protein by shRNA is synthetically lethal on non-small cell lung cancer cells expressing APOBEC3A. Similarly, HMCES KO cells show hypersensitivity to the inhibition of Phosphatidylinositol 3-kinase-related kinases (PIKKs), which are common treatments for tumor therapy[68]. Collectively, inhibition of HMCES might be a possible strategy to treat A3A-expressed tumors, maybe by enhancing the effectiveness of small molecule inhibitors for DNA damage signaling kinases.

HMCES counteracts TLS activity.

Until recently, the only known fate for AP sites in DNA replication was their bypass by TLS polymerases. However, HMCES may promote an error-free mechanism repair because its presence at the replication fork counteracts the recruitment of REV1 and REV3 (catalytic subunit of POLζ), which are TLS polymerases[47]. As mentioned above, TLS polymerization introduces mutations while overcoming AP sites. HMCES KO cells are synthetically lethal with TLS factors POLH, REV1[64], and REV3[47]. Additionally, HMCES depletion slows replication forks after induction of AP sites, and this effect can be reversed by inhibition of REV1-Polζ or depletion of Polκ[67]. These results suggest that HMCES antagonizes AP site bypass by TLS polymerases and promotes a more rapid, error-free mechanism.

HMCES function in blood cells.

HMCES mRNA expression is strongly induced in activated B cells. These cells undergo rapid proliferation and acquire mutations at their immunoglobin (Ig) loci through cytosine deamination activity of the enzyme AID (activation-induced cytidine deaminase). Mutations driven by AID result in somatic hypermutation (SHM), an increase of single nucleotide changes, and the generation of DSBs to induce class switch recombination (CSR). The SHM is a programmed process in which point mutation accumulates in the antibody V-region genes of both the heavy and the light chains. This process occurs at the germinal centers of B-cells as part of antibody affinity maturation which is a key mechanism for antibody diversity and effective immune response[69]. SHM produces point mutations when uracils produced by AID are processed in a mutagenic manner, which creates DNA strand breaks and is susceptible to deleterious deletions. Specifically, AID-mediated cytosine deamination targets “hotspots” described as the consensus motif DGYW, which are abundant at the tandem repeat units of S regions of exons involved in CSR[70]. HMCES deficiency leads to increased deletions without significantly affecting other parameters of SHM in mouse and human B cells. HMCES function in SHM depends on its N-terminal Cys and ssDNA binding residues, suggesting that HMCES-DPC formation is required[71]. Additionally, the function of HMCES in SHM also depends on the activity of UNG, which forms AP sites after removing uracils created by AIDs[71]. This supports the idea that HMCES has a protective role in controlling AP site processing to prevent further damage during the maturation of B cells.

Hematopoietic stress drives quiescent hematopoietic stem cells (HSCs) to proliferate, generating ROS that promotes oxidative damage and the formation of AP sites. Even though HMCES is dispensable for steady-state hematopoiesis, long-term (LT)-HSCs show a compromised self-renewal capacity during hematopoietic stress in the absence of HMCES, as well as an increase in DNA lesions[72]. Interestingly, elevated expression of HMCES occurs frequently in acute lymphocytic leukemia (ALL) and is associated with poor clinical outcomes[72]. This suggests that HMCES might be a potential therapeutic target against ALL while maintaining normal hematopoiesis.

In summary, HMCES’s function heavily relies on its crosslink capacity to control excessive levels of AP sites and provide cellular fitness (Fig 1.7A).

Functions of HMCES that do not require crosslink to AP sites

HMCES functions in the Alternative end-joining pathway

Coming back to B cells, during CSR, some DSBs repair through the microhomology-mediated alternative-end joining (Alt-EJ) pathway, which requires strand resection, generation of ssDNA overhangs, and use of microhomologies at the joined ends[73]. HMCES-deficient mice show normal hematopoiesis but deficient CSR in primary B cells. This deficiency comes from a reduction in the Alt-EJ pathway on HMCES KO cells[74]. Strikingly, the C2A mutant does not alter the ability of HMCES to mediate CSR, contrary to its DNA-binding mutant, which seems to be more important to HMCES function in Alt-EJ[74]. The possible connection between HMCES and the Alt-EJ comes from the capacity of HMCES binding switching regions and protecting ssDNA overhangs during CSR (Fig 1.7B). Additionally, Shukla and collaborators supported this model by the resolved structure of two YedK molecules in complex with two ssDNAs (PDB: 6KBS)[62] on a conformation that resembles the structure expected for HMCES-mediated, microhomology-directed DNA end joins during CSR[74].

HMCES function during development.

Transforming growth factor β (TGF-β) superfamily of proteins have indispensable roles in early embryo development, immunity, tissue homeostasis, etc. The function of (TGF-β) largely relies on transcriptional regulation by SMAD proteins. In mouse embryonic stem cells (mESCs), nodal/activin or BMP4 ligands induce TGF-β signaling activation by the receptor-activated SMAD proteins (R-SMADs), the transcription factors (TFs) of the pathway. R-SMADs target gene loci important for neural differentiation. HMCES binds to R-SMAD proteins and chromatin immunoprecipitation sequencing (ChIP-seq) shows HMCES-SMADs co-occupancy at active chromatin, suggesting a role in TGF-β signaling. Moreover, HMCES KO alters the transcriptome of the TGF-β family signaling and impairs *Xenopus laevis* embryo development[54]. In this study, the authors do not test if HMCES's ability to crosslink nor bind to DNA is required in this process (Fig 1.7B). Even though, it was previously reported that the absence of HMCES can affect embryo development in frogs[48], it is possible that the AP site-related processes mentioned before are also being affected due to the disruption of HMCES and impairing embryo development.

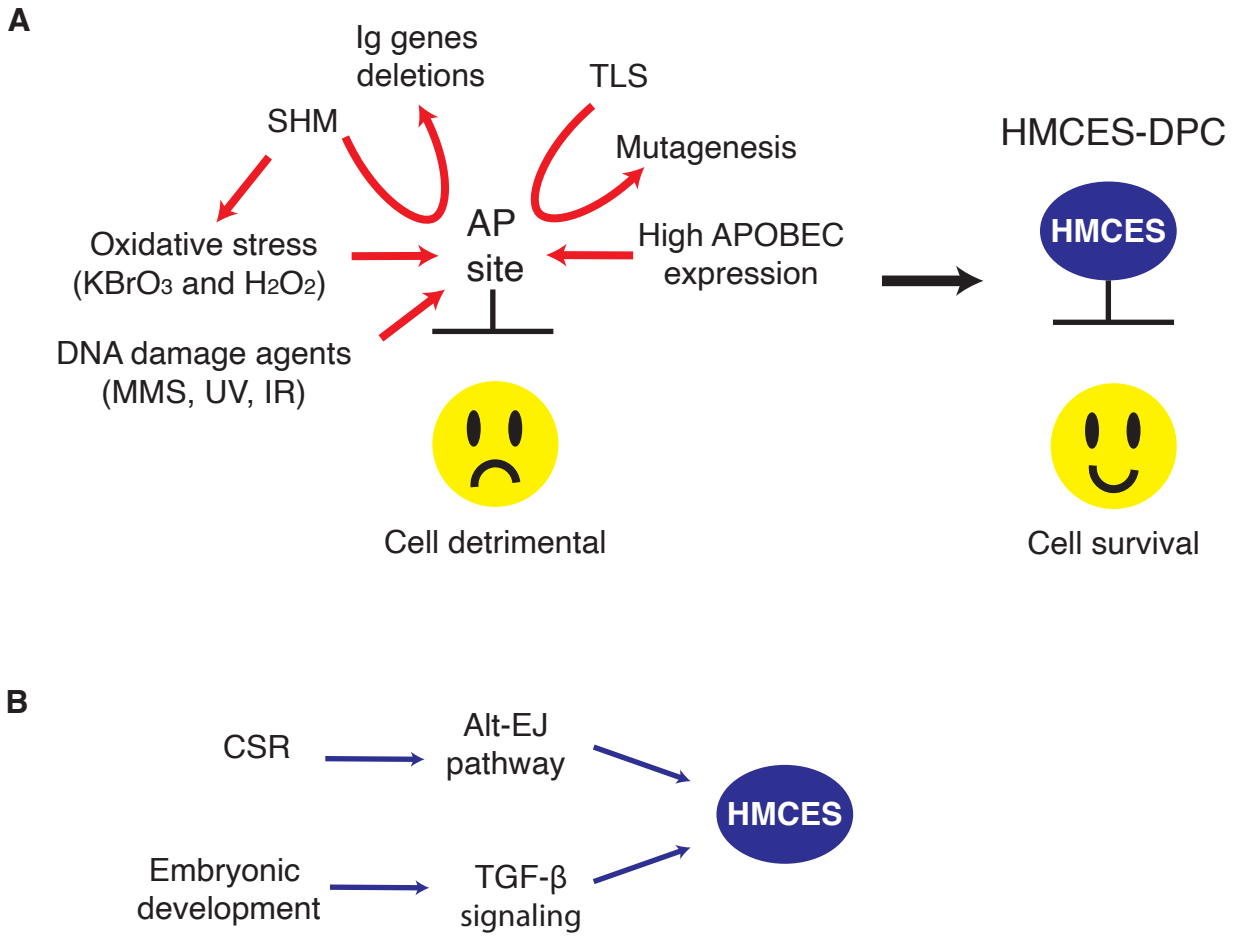

Figure 1.7. Importance of HMCES-DPC formation in cells.

(A) Different sources of AP site formation can be deleterious for cells in the absence of HMCES. Specifically, HMCES-DPC formation is required to maintain cell survival under these circumstances. (B) Proposed functions for HMCES that do not require its crosslink to AP sites involve a step in the Alt-EJ pathway and a role in TGF-β signaling during embryonic development.

Removal of DPCs

HMCES-DPC protects cell viability against DNA damage agents, mostly the ones that generate AP sites. However, HMCES-DPC still represents a bulky lesion that can prevent basic DNA processes such as replication and transcription. For that reason, HMCES-DPC has to be removed and there is more than one possible way to do it. There are different mechanisms to remove DPCs but I'm focusing on the ones that have been reported to act during DNA replication.

DNA replication-coupled DPC removal

DPCs can be formed by different endogenous and exogenous agents, and the type of crosslink and DNA substrate will define how the DPC is removed[75]. Proteolysis is the major pathway for DPC removal where enzymes degrade the adducted proteins. In bacteria and eukaryotic cells, only DPC that are smaller than 11 kDa can be removed by NER[76,77]. Furthermore, DPCs need to be specifically recognized for degradation among different proteins that interact with DNA for regular processes. For that reason, post-translational modifications (PTMs) are used to target DPCs for removal. Generally, replication-coupled DPC targeting requires ubiquitylation, meanwhile, replication-independent DPC targeting is mediated by SUMOylation[78]. Other PTMs such as PARylation can impede proteasomal-degradation by inducing DPC deubiquitylation[79].

Ubiquitin E3 ligases. In the *Xenopus* system, Johannes Walter's research group noticed that DPCs are degraded by proteolysis in a replication-coupled manner, where DPC is removed complementary by the proteasome and SPRTN[80,81]. In his model, the DPC is the target for polyubiquitylation partially dependent on the E3 ligase TRAIP[82,83], which allows the bypass of the replicative Cdc45-Mcm-GINS (CMG) helicase[84]. Then, additional ssDNA-dependent ubiquitylation of the DPC induces proteasomal degradation, while the complementary pathway is mediated by SPRTN, an essential protease that removes DPCs at replication forks[85,86]. Polymerase-blocking DNA lesions which include cyclobutene pyrimidine dimer (CPD), ICL, and DPC can promote the activity of another E3 ligase called RFWD3, which can be recruited to the replication fork by replication protein A (RPA) for ICL repair[87]. RFWD3 regulates DNA damage-induced ubiquitylation of PCNA and other proteins located in ssDNA, which later stimulates gap-filling DNA synthesis by TLS polymerases[88]. Additionally, the majority of E3 ligases belong to the RING family and within that family, Cullin-RING E3 ligases (CRLs) are the largest superfamily, with over 200 members in cells. CUL3 and CUL4B are scaffold proteins from different CRLs whose downregulation not only increased the amount of TOP1-DPC but also reversed

camptothecin (CPT) resistance[89,90]. $CRL2^{Lrr1}$ promotes CMG unloading mediated by p97 but has not been associated with DPC removal yet. In summary, ubiquitin E3 ligases are important regulators of DPC, some of them directly target DPC for proteasomal degradation and others might provide the signal for other proteases.

Proteases. Proteasome-independent degradation can also remove DPCs. The aspartic protease Ddi1 assists in the processing of Top1-DNA covalent complexes and it functions independently of proteasome activity[91]. However, Ddi1 contains ubiquitin-like (UBL) and ubiquitin-associated (UBA) domains[92], and some *in vitro* evidence shows that Ddi1 can act on polyubiquitylated substrates[93]. FAM111A is another protease that interacts with PCNA, and DNA and requires a trypsin-like protease domain to ensure fork progression[94]. Lack of FAM111A causes trapping-dependent PARPi sensitization and TOP1cc accumulation, suggesting a role in removing these crosslinked proteins. Moreover, FAM111A contains 4 domains, two of which are ubiquitin-like domains, but their exact function is unclear[95]. The ACRC/GCNA is a Sprt-like protease that removes DPCs that are SUMOylated in a replication-independent manner[96], however, recent discoveries showed that SUMOylation of DPC is also important to remove DPC during DNA replication[97]. Ddi1, FAM111A, and ACRC/GCNA are proteases involved in DPC removal, but among human enzymes, SPRTN is the only essential gene in a variety of cancer cell lines.

SPRTN. DNA-dependent metalloprotease SPRTN (the functional homolog of Wss1 in metazoans), degrades DPC during replication. In the last decade, this protein has been reported to remove different types of DPC and tightly associated DNA binding proteins in cells and *in vitro*: M.HpaII, TOP1/2, formaldehyde-induced DPCs, PARP1, CHK1, and USP1[85,86,98–101]. DPC processing by SPRTN requires a ubiquitin switch regulating its chromatin accessibility and a negative feedback loop based on autocatalytic cleavage. Moreover, it shows full activity when the bulky lesions are located at ss/dsDNA junctions[102,103]. SPRTN's substrates do not require ubiquitylation, however, in human cells, SPRTN interacts with ubiquitin- and SUMO-modified proteins and SUMOylated DPCs accumulate in SPRTN-depleted cells[104]. Mass spectrometry analysis of DPCs in mammalian cells was used in genetically engineered cells expressing SPRTN variants that are associated with premature aging and early-onset liver cancer in Ruijs-Aalfs syndrome. This study found that SPRTN requires SUMO-targeted ubiquitylation of protein adducts and can occur in addition to proteasomal DPC degradation[105]

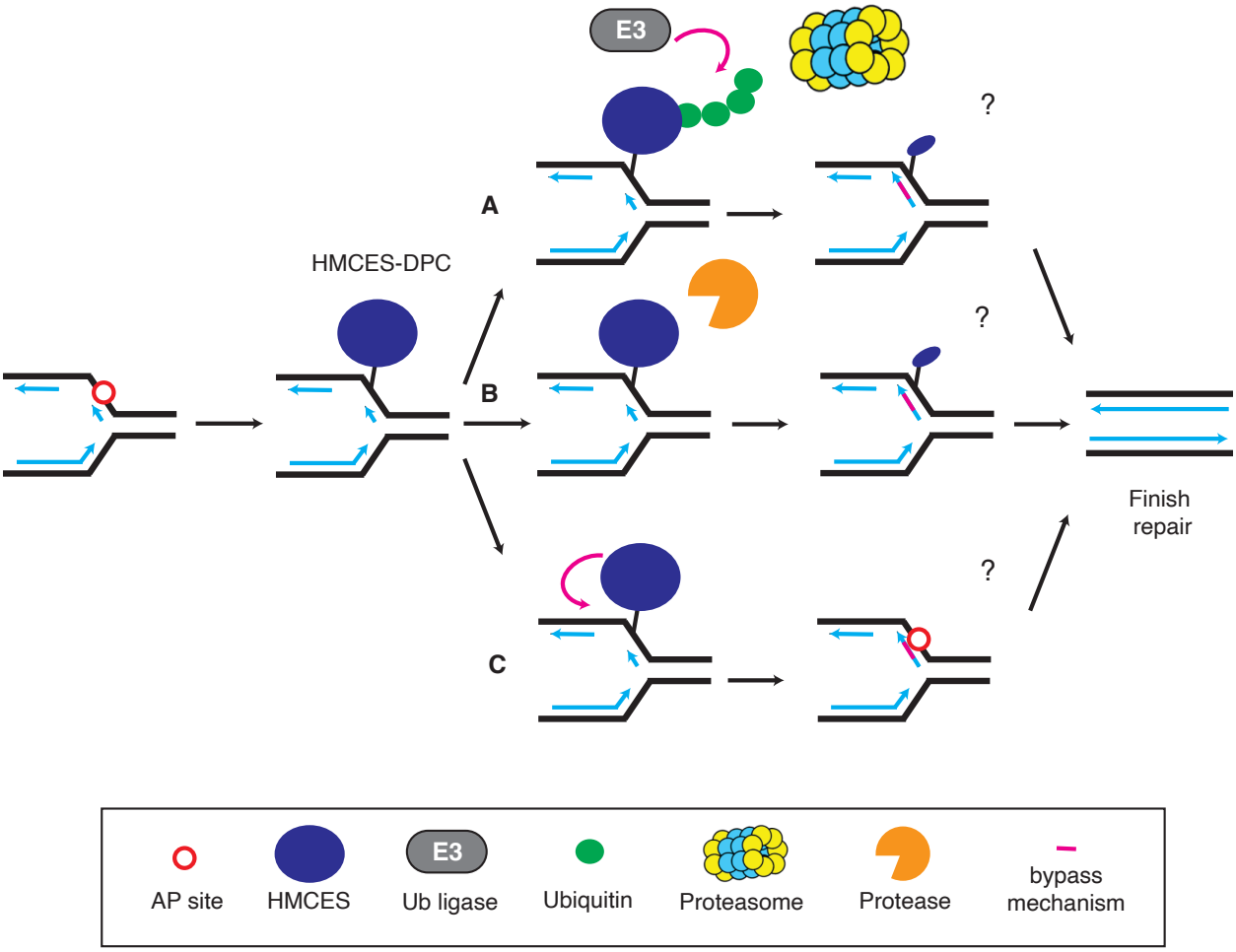

Figure 1.8. Possible mechanisms for HMCES-DPC removal

(A) A proteasome-dependent pathway requires the ubiquitination of HMCES-DPC mediated by an E3 ligase (TRAIP, RFWD3, CRLs). Then, it will be targeted for proteasome degradation (B) Some proteases (SPRTN, FAM111A, Ddi1, ACRC/GCNA), do not require previous target ubiquitylation to degrade the DPC present at replication forks. (C) Self-reversal is a newly described mechanism where the DPC catalyzes its removal (magenta arrow). In all cases, DNA adducts or newly exposed AP sites would need a bypass mechanism (TLS, template switching, etc.) to finish repair.

Removal of proteins crosslink to AP sites

As HMCES, other types of DPCs are formed during DNA replication and rely on the presence of AP sites. The DNA repair factor poly(ADP-ribose) polymerase 1 (PARP1) can crosslink to AP sites located in ssDNA. PARP1 induces strand cleavage by β-elimination and a Schiff base intermediate, which results in PARP1 crosslinked to the O3' side of the β-elimination-induced nick in DNA. Some cellular data show that PARP1-DPC levels increase after exposure to MMS and are modestly higher with simultaneous proteasome inhibitor (MG132) treatment. This suggests that PARP1-DPC is regulated by the proteasome[106,107]. Another protein with similar activity is Polymerase β (Polβ), which can crosslink to AP sites in their attempt to repair AP sites formed during oxidative stress. Polβ also crosslinks to the 5' end of a DNA nick and cells treated with 1,10-copper-ortho-phenanthroline [$Cu(OP)_2$] increase Polβ-DPC levels. Treatment with MG132 after DNA stress increases Polβ-DPC levels, suggesting proteasome-dependent degradation[24,107].

Mechanisms of HMCES-DPC removal

HMCES-DPC is an intermediate step in the ICL repair pathway.

AP site can react with exocyclic amines on the opposite strand and form ICLs. These AP-ICLs are highly toxic lesions that can block DNA replication and transcription[108]. In the *Xenopus* extracts, AP-ICL is repaired in a replication-coupled mechanism that requires the activity of NEIL3 to unhook the crosslink, which exposes the original AP site to HMCES and forms an HMCES-DPC. HMCES crosslink is subsequently degraded by SPRTN, generating a peptide adduct that can be passed by TLS polymerase[109]. Furthermore, sequencing of AP site-containing plasmids in the absence or presence of HMCES, showed that HMCES-DPC promotes dG insertion during TLS activity[110]. This suggests a thermodynamic constraint on TLS polymerase while passing the DPC, similar to the "A rule" during the insertion opposite to an unprotected AP site, or that the SRAP domain evolved to suppress mutagenesis from cytosine deamination or other modifications that can lead to its removal and the generation of AP sites. *In vitro,* data shows that DNA helicase FANCJ unfolds HMCES-DPC for later degradation by SPRTN[111]. FANCJ is also required to support CMG bypass in the absence of RTEL1 and promote translesion synthesis past stable DPCs, which explains how TLS still occurs, albeit slowly, when a DPC cannot be degraded[84,111].

In summary, HMCES-DPC can be an intermediate step in the AP-ICL repair mechanism, and later processing after its degradation or unfolding, might involve TLS or template switching.

Pathways for HMCES-DPC removal during DNA replication

My thesis project aims to understand how HMCES-DPC is removed in human cells. During these years working on my research, some other groups have explored different mechanisms to explain HMCES-DPC removal. One model proposes that HMCES-DPC is resolved by proteolysis, which can involve HMCES-DPC ubiquitylation by a specific E3 ligases[88] for later proteasome degradation (Fig 1.8A) or the direct activity of proteases[94,110] that might or not require PTMs to target the DPC (Fig 1.8B). Another novel model proposes that HMCES mediates its auto-release of the crosslink and its regulation might depend on the type of the DNA substrate where is crosslinked[59,112,113] (Fig 1.8C). All these possible pathways would generate a DNA-adduct or a newly available AP site that would require a bypass mechanism (TLS, template switching, etc.) to complete repair. Importantly, all these models are supported mostly by biochemical data, but they still have not been tested in human cells or their effect in case these pathways are disrupted. For that reason, in my thesis project, I'm testing all these possible mechanisms using a system to track HMCES-DPC removal in human cells.

CHAPTER II

MATERIALS AND METHODS

Generation of stable cell lines

U2OS HMCESΔ cells were generated previously[47]. Stable cell lines expressing WT-HMCES, E127Q, HMCES without C-terminus (ΔC), and empty vector (EV) were generated by transduction of U2OS HMCESΔ cells with a pLPG backbone lentivirus. Stable cell lines overexpressing WT HMCES, E127Q, C2A, C2A/EQ, or ΔC, were generated by transduction of U2OS HMCESΔ cells with pLPCX retrovirus containing a CMV promoter. Protein expression was corroborated by Western blot after three days of puromycin selection. Cell lines and plasmids used are described in Table 2.1

Cell Transfections

siRNA transfections were performed using a pool of Lipofectamine RNAimax (individually described in Table 2.2) according to manufacturer's instructions and depending on the cell line. Plasmid transfections to produce lentivirus or retrovirus are also described in Table 2.3

Plasmids

HMCES E127Q plasmid was generated by Gibson assembly of gene block containing the point mutation into the pLPG plasmid with gD promoter. Over-expression plasmids were created by cloning HMCES WT, E127Q, HMCES ΔC, C2A, or C2A/EQ cDNA into pLPCX plasmid containing CMV promoter. Plasmids were corroborated by sequencing. pUGI-NLS UDG Inhibitor (UGI) plasmid was purchased from Addgene (Cat#101091).

Cell line name	Promoter	Backbone	Plasmid	Characteristic
WT HMCES	gD	pLPG	pKM354	Wild type HMCES at endogenous levels
HMCES EQ	gD	pLPG	pJRF10	HMCES E127Q at endogenous levels
HMCES ΔC	gD	pLPG	pKM355	HMCES with no C-term at endogenous levels
HMCES OE WT	CMV	pLPCX	pJRF01	Overexpression of WT HMCES
HMCES OE EQ	CMV	pLPCX	pJRF16	Overexpression of E127Q HMCES
HMCES OE ΔC	CMV	pLPCX	pJRF04	Overexpression of HMCES with no C-term
HMCES OE C2A	CMV	pLPCX	pDC1351	Overexpression of HMCES C2A
HMCES OE C2A/EQ	CMV	pLPCX	pDC1352	Overexpression of double mutant C2A/E127Q
HMCES Bio-ID2	CMV	MCS-BioID2-HA pBabe-puro	pJRF07	Expression of HMCES fused to BioID2 through its c-term region
EV	CMV	pLPCX	-	No expressing HMCES

Table 2.1 *List of stable cell lines generated for the project.*

All these cell lines were generated using U2OS HMCESΔ cells[47].

Analysis of parental ssDNA with native BrdU staining

Cells were plated in a 96-well glass-bottom, poly-L-lysine coated plate adding 100µl to each well and incubating at room temperature for 1 h. Cells were pulsed with 20µM BrdU for 18 hours. BrdU was washed off for two hours before drug treatment. Cells were treated with CD437 (5µM, diluted in DMSO) (Tocris Bioscience, cat#1549) or HU (0.3mM, 3mM) for 30 min and subsequently fixed with 3% paraformaldehyde, 2% sucrose (in PBS) for 10 min. Cells were permeabilized in PBT (0.5% triton X-100 in PBS) for 10 min, blocked for 1 hour in 10% normal goat serum, and probed with anti-BrdU antibody (1:250 dilution, AbCam Cat#ab6326), followed by Goat anti-Rat IgG (1:250 dilution, Thermo Cat# A11007). The plate was imaged and analyzed directly using a Molecular Devices ImageXpress high-content imager.

AP site detection

Genomic DNA was purified as described in the RADAR method, quantified, and diluted to 100 ng/µl in dH_20. Abasic sites were labeled by incubation of 2.5 µg DNA with 5 mM biotinylated aldehyde reactive probe (ARP; Dojindo Laboratories, A305) for 1 hour at 37°C. The DNA was ethanol precipitated, washed twice with 70% ethanol, resuspended in dH_20, and quantified. For the loading control, 50 ng DNA was diluted in 6X SSC, denatured, and dot-blotted onto a nylon membrane. The membrane was treated with 1.5 M NaCl, 0.5 N NaOH for 10 min, followed by 1 M NaCl, 0.5M Tris-HCl pH 7.0 for 10 min, and DNA was crosslinked to the membrane using a Stratalinker UV crosslinker (1200 J/m^2). The membrane was blocked with 5% milk in TBST immunoblotted for ssDNA (Millipore Sigma, MAB3034). For ARP detection, 500 ng or 250 ng DNA was diluted in 6X SSC, denatured, and applied to a nylon membrane, as above. The membrane was blocked with 5% bovine serum albumin in TBST and biotin was detected with Streptavidin-HRP (ThermoFisher) using a 1:10000 dilution.

Immunoblotting

Cell lysates were extracted using Igepal lysis buffer (1% Igepal, 150mM NaCl, 50mM Tris pH 7.4) enriched with a protease inhibitor cocktail (Roche). Proteins were analyzed by SDS-PAGE and immunoblotting using antibodies described in Table 2.4. For loading control, I used the stain-free imaging option from Bio-Rad, which images total protein levels as a control. The precast gels are Criterion TGX 12+2 well stain-free precast gels. Images were obtained by chemiluminescence (Femto reagent). For analyzing expression cells, ECL reagent is enough.

siRNA SPRTN (C1orf124)		
J-015442-19 C1orf124	CAUCAAAGUCAAAAGCGAA	Dharmacon (Horizon)
J-015442-20 C1orf124	GGUUCUGCCUGUCGUGUAU	Dharmacon (Horizon)
J-015442-21 C1orf124	GGCUAUGUCAAACGAGCUA	Dharmacon (Horizon)
J-015442-22 C1orf124	AAGUGGAGCGUGCGAAUGA	Dharmacon (Horizon)

Table 2.2 SPRTN siRNA sequences

	Cell line	Forward/ Reverse	Number of cells	Protocol
siRNA transfection	U2OS	Reverse	$3x10^5$ cells in 60mm dish	6.4µl of Dharmafect + 40pmoles of siRNA; 500µl optimem each --Incubate 20 minutes at room temperature and add dropwise to cells
	HCT116	Reverse	$1x10^6$ cells in 35mm dish	5µl RNAimax + 20pmoles siRNA in 200µl of optimem each --Incubate 15 minutes at room temperature and add dropwise to cells
	virus	Cell line	Number of cells	Protocol
Transfections for making viruses	Retrovirus (pLPCX)	GP2-293	$2x10^6$ cells in 60mm dish	Mix: - Optimem: 35ul, - Vector (1.4 ug), - pSVG: 2ug - PEI: 1ug/ml 15 min incubation at RT Add dropwise to cells Puromycin selection
	Lentivirus (pLPG)	293FT	$2x10^6$ cells in 60mm dish	Mix: - Optimem: 35ul, - Vector (1.4 ug), - pPAX2: 3ug - pSVG: 2ug - PEI: 1ug/ml 15 min incubation at RT Add dropwise to cells Puromycin selection

Table 2.3 Summary of transfection protocols

DNA combing

Cells were labeled for 20 min with 20 μM CldU (Sigma, C6891) followed by 40 min with 100 μM IdU (Sigma, I7125) and approximately 300,000 cells were collected by trypsinization and embedded in agarose plugs. DNA combing was performed according to the manufacturer's instructions (Genomic Vision) with minor modifications using a combing machine. DNA-combed coverslips were baked for 2 hours at 65°C and stored at -20°C. The DNA-coated coverslips were denatured for 10 min with freshly prepared 0.5M NaOH, 1M NaCl solution, washed with PBS, and dehydrated consecutively in 70%, 90%, and 100% ethanol before air drying. Coverslips were blocked with 10% goat serum, 0.1% Triton X-100 in 1X PBS and immunostaining was performed with antibodies that recognize CldU (1:25 dilution, Abcam, ab6326) and IdU (1:5 dilution, BD Biosciences, 347580) for 1 hour at room temperature. Coverslips were then washed in PBS, probed with secondary antibodies for 30 min at room temperature, washed with PBS and mounted using ProLong Gold (ThermoFisher). Images were captured using a 40X oil objective (Nikon Eclipse Ti) and fiber length analysis was performed using Nikon Elements software.

RADAR assay

Cells were synchronized with 2mM thymidine overnight. Then, thymidine was removed and cells recovered in normal growth medium for 2h before treatment with 5μM CD437 for 30 min. Cells were washed twice with PBS and recovered in normal growth medium. Cells were lysed in RADAR buffer: RLT plus buffer (Qiagen Cat#:1053393) supplemented with 1% Sarkosyl. Genomic DNA was ethanol precipitated by the addition of ½ volume 100% ethanol and incubation at -20°C for 5 min. After full-speed centrifugation for 20 min, the DNA pellet was washed twice with 70% ethanol. Next, remove the ethanol remaining and resuspend in 8mM NaOH at 65°C at 800rpm for 3 hr. DNA concentration was determined by spectrophotometry. DNA sample (20μg) was digested with Pierce universal nuclease (125U/ml) in 1X TBS with 2mM $MgCl_2$ at 37°C at 300rpm for 1 hr. Samples were boiled for 5 min and applied to a nitrocellulose membrane with a slot blot apparatus. The membrane was blocked for 1 hr with 5% non-fat dry milk in TBST and immunoblotted for HMCES. For the DNA blot, DNA sample (1μg) was added in 1ml of 6X SSC buffer. The sample was boiled and then on ice for 10 min each and added to a nylon membrane with a slot apparatus. The membrane was placed face up on Whatman paper soaked with solution A (1.5M NaCl, 0.5M NaOH) for 10 min and in solution B (1.5M NaCl, 0.5M Tris pH7.5) for 5 min.

	Antibody	Dilution	Species	Company	Catalog number
Primary antibody	HMCES	1:3000	Rabbit	Sigma	HPA044968
	SPRTN (DVC1)	1:3000	Mouse	John Rouse lab	N/A
	Nedd8	1:1000	Rabbit	abcam	Ab81254
	Ubiquitin	1:3000	Mouse	Calbiochem	ST1200
	ssDNA	1:1000	Mouse	DSHB	AB_10805144
	H3	1:10000	Mouse	abcam	Ab10799
Secondary antibody	IgG (H+L)	1:10000	Rabbit	Fischer scientific	111-035-144
	StarBright Blue 700 IgG	1:10000	Mouse	Biio-Rad	12004158

Table 2.4 List of antibodies used.

After air drying, the membrane is crosslinked with UV 1200J/m^2 and blocked for 1 hr with 5% non-fat dry milk in TBST and immunoblotted for ssDNA.

Preparation of AP-DNA

Sequences of oligonucleotides used in the biochemical assays are listed in Supplementary Table 2.5, AP-DNA was prepared by incubation of 100μM uracil-containing oligonucleotide (700μM of the trap oligo) and 8 U UDG in NEB UDG Buffer, supplemented with 1mM DTT in a total volume of 20μl, at 37°C for 20 min. AP-DNA was prepared fresh for each reaction.

Cross-link reversal assay

WT and E127Q HMCES were purified as previously described[59]. HMCES-DPC was formed by incubation of 10μM 20mer AP-DNA with 2μM HMCES (WT or E127Q) in DPC buffer (20μM HEPES, 10mM NaCl, 1mM EDT, pH 8) overnight at 37°C. DPC-20 was incubated with a 50-fold excess of 40mer AP-DNA to trap any reversed HMCES. Reactions were stopped by adding an equal volume of 2X SDS buffer. Each time point was initiated in reverse so that all reactions were quenched for the same length of time. Reaction products were resolved on 4 to 12% Bis-Tris gels and Coomassie stained for detection.

Reversal trapping in duplex DNA required incubation of 10μM 40mer AP-DNA with 2μM WT HMCES in DPC buffer overnight at 37°C. DPC-40 was incubated with a 1.4-fold excess of non-complementary oligo (40mer control) or the complementary oligo (c40mer). Simultaneously, a 50-fold excess of 20mer AP-DNA was added to trap any reversed HMCES. Each time point was initiated in reverse so all reactions were quenched for the same length of time and by adding an equal volume of 2X SDS buffer. Reactions products were resolved on 4 to 12% Bis-tris gels, 1X MES running buffer, and coomassie stained for protein. The reversal percentage was calculated as the percentage of DPC-20 compared to the total at each time point. Sequences of DNA oligos are described in Table 2.5.

Viability assays

$5x10^3$ cells/well in a 96-well dish were seeded for all the different cell lines. Cells were incubated in a DMEM growth medium for 5 days. For drug hypersensitivity assays, cells were treated with

CD437 for 24 hr, or $KBrO_3$ for 48 h and then recovered in normal growth medium for 4 and 3 days, respectively. AlamarBlue readout was performed using a BioTek multimode reader. All viability measurements are presented as a percentage of the untreated control. Overall cell growth of over-expression cell lines and EV was done by methylene blue staining. Cells were seeded and incubated at 37°C for 5 days.

Immunofluorescence

Seed 10000 U2OS cells per well in a 96-well dish. The next day, make aliquots of EdU 10μM, CD437 5μM, MG132 10μM, and TAK243 10μM in pre-calibrated growth media. (Add and remove volumes using a micropipette to avoid using the vacuum). Do the treatment starting from the last time point and finishing with the only drug incubation samples (T0 or 0h). Treatment: Add 200μl of EdU for 25 min at 37 C. Remove EdU and add 200μl of CD437 for 30 min at 37 C. Remove CD437 and wash twice with 1X PBS (carefully). Add normal growth media and incubate. After completing all the time points, fix cells with 3% paraformaldehyde/2% sucrose for 10 min at RT. Then wash twice with 1X PBS and the plate can be stored at 4C to continue the staining part.

For the staining part. Wash cells with 1X PBS. Permeabilize with Triton X-100 solution for 10 min on ice (or at 4 degrees). Wash 3x with PBS (thaw 488 Azide) and block with goat serum/BSA 0.1% triton (blocking buffer) for 30 min at RT. Then perform the click reaction: 1) 1X PBS, 2) 488 azide, 3) NaAscorbate (20mg/ml) and 4) CuSO4 100mM (order is important). Apply 100 μl of click reaction per well and incubate in the dark for 30 min at RT. Wash cells with 1x PBS. Dilute HMCES Ab 1:100 in the blocking buffer and add 80ul for 1 hr at RT (covered) in each well, except for the No primary control. Wash 3 times with 1X PBS. Dilute secondary Ab 1:350 (Alexa Fluor 594 anti-Rb) in blocking buffer and add 80μl for 20 min at RT (covered) in each well, except for the No secondary control. Wash 3x with PBS. Finally, add PBS+DAPI (1:10000 in PBS) for 2 min. Rinse with 1X PBS before imaging.

Name	Sequence	Company
20mer	d(TCTTCTGGTCUGGATGGTAGT)	IDT
40mer	d(GGAATCTGACTCTTCTGGTCUGGATGGTAGTTAAGTCTTGT)	IDT
C40mer	d(ACAAGACTTAACTACCATCCAGACCAGAAGAGTCAGATTCC)	IDT
T40mer	d(GGAATCTGACTCTTCTGGTCTGGATGGTAGTTAAGTCTTGT)	IDT

Table 2.5 Sequence of oligonucleotides used in the biochemical assays.

Neutral comet assay

Seed 2x10^5 synchronized U2OS cells (2mM thymidine overnight) in each 6cm dish. Next day, remove thymidine and recover cells in normal media (have a 1μM CPT for 1 h, as a control). Trypzine cells and resuspend them in a total volume of 1 ml. Chill the lysis solution at 4 C for at least 20 mins before use. Melt LMA (low melting point agarose) above 65 C and cool in 37 C for at least 20 mins. Add 5ul of resuspended cells in 100ul of LMA and immediately pipette all 100ul into each of the two-well comet slides. Place slides at 4 C in the dark for 15 mins. Immerse slides in prechilled lysis solution (50ml) at 4C for 1 hr. Rinse slides in pre-chilled TAE 1X and keep it for 30 mins in 1X TAE at 4C. Perform neutral electrophoresis for 1 h at 21V in 850ml of TAE buffer. Immerse slides in DNA precipitation solution for 30 min at RT. Immerse slides in 70% ethanol for 30 min at RT. Dry samples at 45 c for 15 min. Store cells overnight at RT. Place 100ul of 1X SYBR Gold in each sample (1:10000) for 30 min and remove excess (doing it 1:20000 might help to reduce background noise). Rinse with water for 5 min and remove excess. Let dry for 30 mins. Image at the microscope at 20X (get at least 20 images per well).

Chromatin fractionation

U2OS cells (2x10^6 cells per each 10 cm plate) are trypsinized, centrifuged, and washed with 1X PBS for possible storage at -80. To get the soluble fraction: resuspend cell pellets in 200ul buffer A (100mM NaCl, 300mM sucrose, 3mM MgCl2,10mM Pipes pH 6.8, 1mM EGTA, 0.2% Triton X-100, 1mM DTT and protease inhibitor cocktail) and leave on ice for 5 minutes (avoid bubble formation). Spin at 1300g for 4 min. Remove supernatant and collect as “soluble fraction” (can add 40ul of 6X SB and store at -20). To collect the chromatin fraction: resuspend the pellet with 1ml of buffer A. Spin at 1300g for 4 min and discard the supernatant. Lyse nuclei in 200ul of buffer B (50mM Tris-HCl pH 7.5, 150mM NaCl, 5mM EDTA, 1% Triton X-100, 1mM DTT, protease inhibitor cocktail, and 2mM MgCl2) for 10 mins. Add pierce universal nuclease (1μl in each sample) and incubate overnight at 4 C. Next day, add 40ul of 6X SB, boil and spin at 10,000g for 1 min. Remove the supernatant and collect this as “chromatin fraction”. Both soluble and chromatin fractions can be used for SDS-PAGE in a 4-15% gel.

Quantification and statistical analysis

Statistical analyses were completed with PRISM v.9. Descriptions of statistical tests can be found in the figure legends.

CHAPTER III

DEVELOPING A SYSTEM TO STUDY HMCES-DPC REMOVAL IN HUMAN CELLS

Introduction

There are different techniques to detect DPCs in cells, each of them with its benefits and drawbacks. The rapid approach to DNA adduct recovery (RADAR) is one of the most widely used methods for the biochemical isolation of DPCs. It requires cell lysis under stringent conditions with strong detergents and chaotropic agents to later isolate DNA only maintaining covalently attached proteins. Next, a process of DNA digestion releases the proteins for their detection[114,115]. The RADAR assay has a simple execution and is very specific. However, low-specificity antibodies and potential contamination with RNA-protein crosslinks are some drawbacks.

I used the RADAR assay to analyze HMCES-DPC levels in human cells. Low basal levels of HMCES-DPC and the assay's sensitivity require the induction of HMCES-DPC formation for their detection. DNA-damaging agents such as KBrO3 generate abasic sites, but it yields a modest increase in HMCES-DPC[47] which is hard to detect by RADAR. Additionally, $KBrO_3$-induced HMCES-DPCs keep forming 3 hours after treatment and removal ends 3 hours later[47]. The time length of these processes makes it unlikely that HMCES-DPC is formed and removed along DNA replication. This also suggests that these processes might be overlapping, which would make it difficult to analyze HMCES-DPC removal independently from its synthesis. Another disadvantage of using ROS agents to induce AP sites is that base modification is unrestricted to ssDNA. I tried to overcome that by using APOBECA3A inducible cells[67] to promote AP site formation specifically in the lagging strand, but the amount of HMCES-DPC formed seemed to not be enough to be detected by the RADAR assay.

Ultraviolet (UV) radiation can generate a 10-fold increase in HMCES-DPC levels[47], which makes it a more attractive reagent. UV light predominantly forms cyclobutene pyrimidine dimers (CPD), but other induced photoproducts can indirectly produce AP sites[116]. However, UV light crosslinks different proteins surrounding DNA despite the presence of AP sites, suggesting that UV light might not be the best AP site-inducing agent to analyze HMCES-DPC removal in cells.

For that reason, I needed a better tool to analyze HMCES-DPC resolution in cells. Surprisingly, Dr. Kavi Mehta and Dr. Courtney Lovejoy, in our lab, encountered CD437, which is a polymerase alpha inhibitor, as a reagent that induces an HMCES-DPC signal that improves the signal-background noise ratio problem from other reagents. Additionally, removing CD437 allows for immediate tracking of HMCES-DPC resolution over time. I used preliminary data from other members in the lab about the dose and timing for the activity of CD437 in cells and I developed a system using CD437 to analyze HMCES-DPC removal in human cells.

Results

CD437 increases levels of AP sites in ssDNA for HMCES-DPC detection

AP sites are preferentially formed at DNA replication sites, and the depurination rate is accelerated four times in ssDNA compared to dsDNA[2,3]. Moreover, the stretches of ssDNA on the lagging strand are more vulnerable to chemical attack and spontaneous base loss. CD437 is a direct Polymerase alpha inhibitor that prevents DNA replication *In vivo*. It strongly binds the Polα subunit POLA1 and inhibits lagging strand synthesis[117]. Thus, we hypothesized that CD437 would expose more parental ssDNA in the lagging strand and produce AP sites.

We noticed that U2OS cells treated with 5μM of CD437 for 30 mins rapidly generate large amounts of ssDNA compared to other replication-stalling agents like hydroxyurea (HU)[118] (Fig 3.1A). CD437 treatment dramatically slows DNA synthesis and approximately half of the ongoing replication forks do not restart after its removal (Fig 3.1B). ssDNA is more vulnerable to chemical attack and spontaneous depurination, which leads to AP site formation. Indeed, we could detect AP sites with an aldehyde reactive probe (ARP) after the addition of CD437 (Fig 3.1C). Since HMCES crosslinks specifically to AP sites located in ssDNA[47], we hypothesized that CD437 would induce HMCES-DPC formation. As expected, cells treated with 5μM CD437 for 30 min showed a large increase in the HMCES-DPC signal using the RADAR assay (Fig 3.1D). The HMCES-DPCs generated by CD437 were greatly reduced in cells transduced with a plasmid expressing the uracil DNA glycosylase inhibitor (UGI), almost taking it to basal levels (Fig 3.1E), suggesting that HMCES reacted with AP sites largely created by glycosylase activity in these conditions.

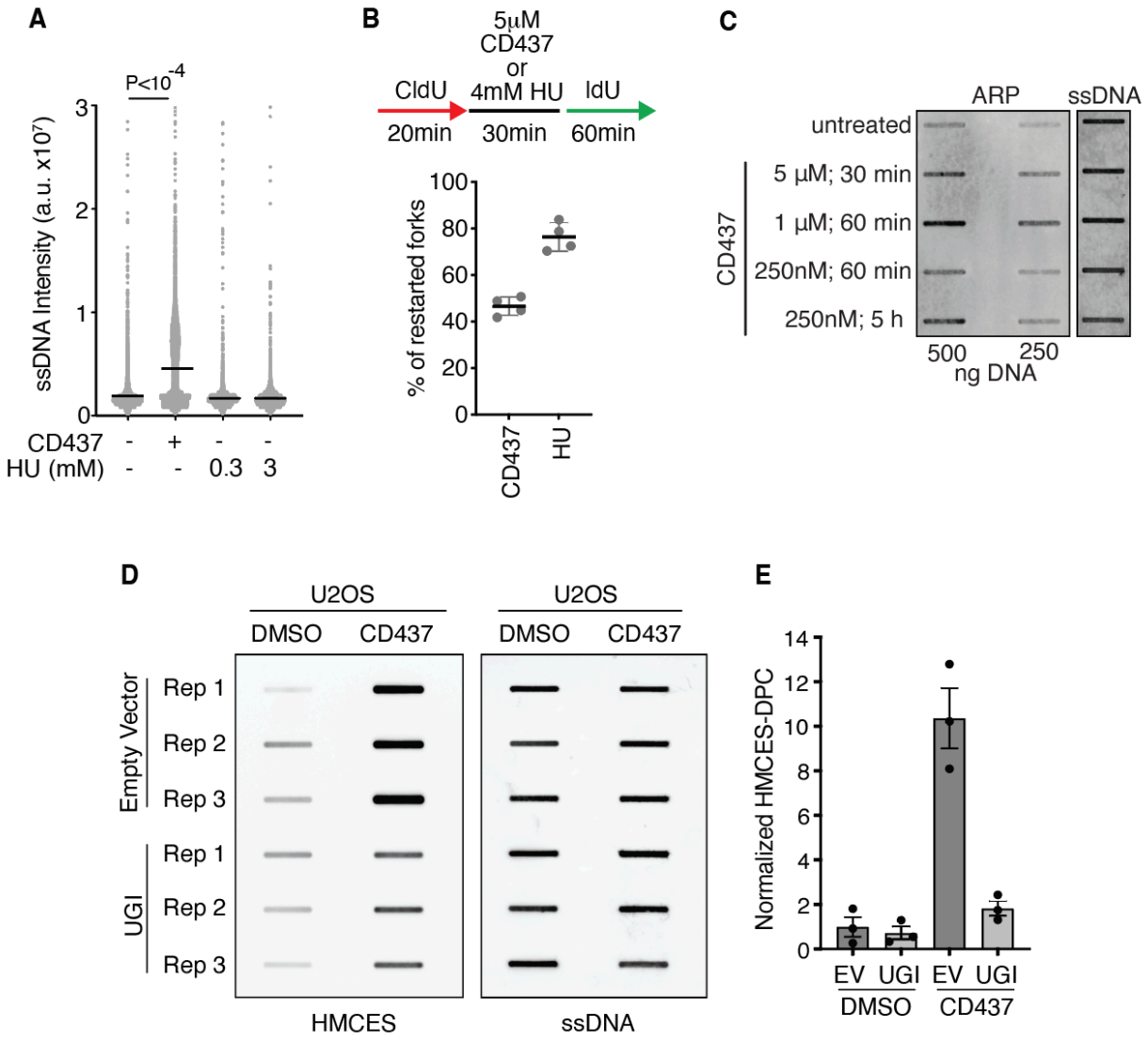

Figure 3.1 CD437 induces AP sites in ssDNA and HMCES-DPC formation

(A) Levels of ssDNA in HCT116 cells after treatment with 5 μM of CD437 or 0.3 or 3mM HU for 30 min measured using the native BrdU staining assay. The intensity of individual nuclei and mean is shown in; the Kruskal-Wallis test. (B) U2OS cells were labeled and treated as indicated. DNA combing was performed, and the percentage of restarted forks (red + green DNA tracks) compared to the total (red and green only) was measured (Mean ± SEM, n=4). (C) DNA was purified from cells treated as indicated and reacted with the aldehyde reactive probe (ARP) to measure AP sites. ARP-reacted DNA was blotted on membranes and probed with streptavidin-HRP to measure AP sites or with ssDNA antibody to measure DNA loading. (D) RADAR assay of U2OS cells expressing or not UGI and treated with 5 μM of CD437 or DMSO for 30 min (shown in triplicates). Separate membranes were blotted using HMCES or ssDNA antibody. (E) Quantification of D, using ssDNA blot to normalize HMCES-DPC values (Mean ± SEM, n=3). Experiment A was performed by Dr. Kavi Mehta. Dr. Celeste Giansanti performed experiment B. Dr. Courtney Lovejoy performed experiments C and D.

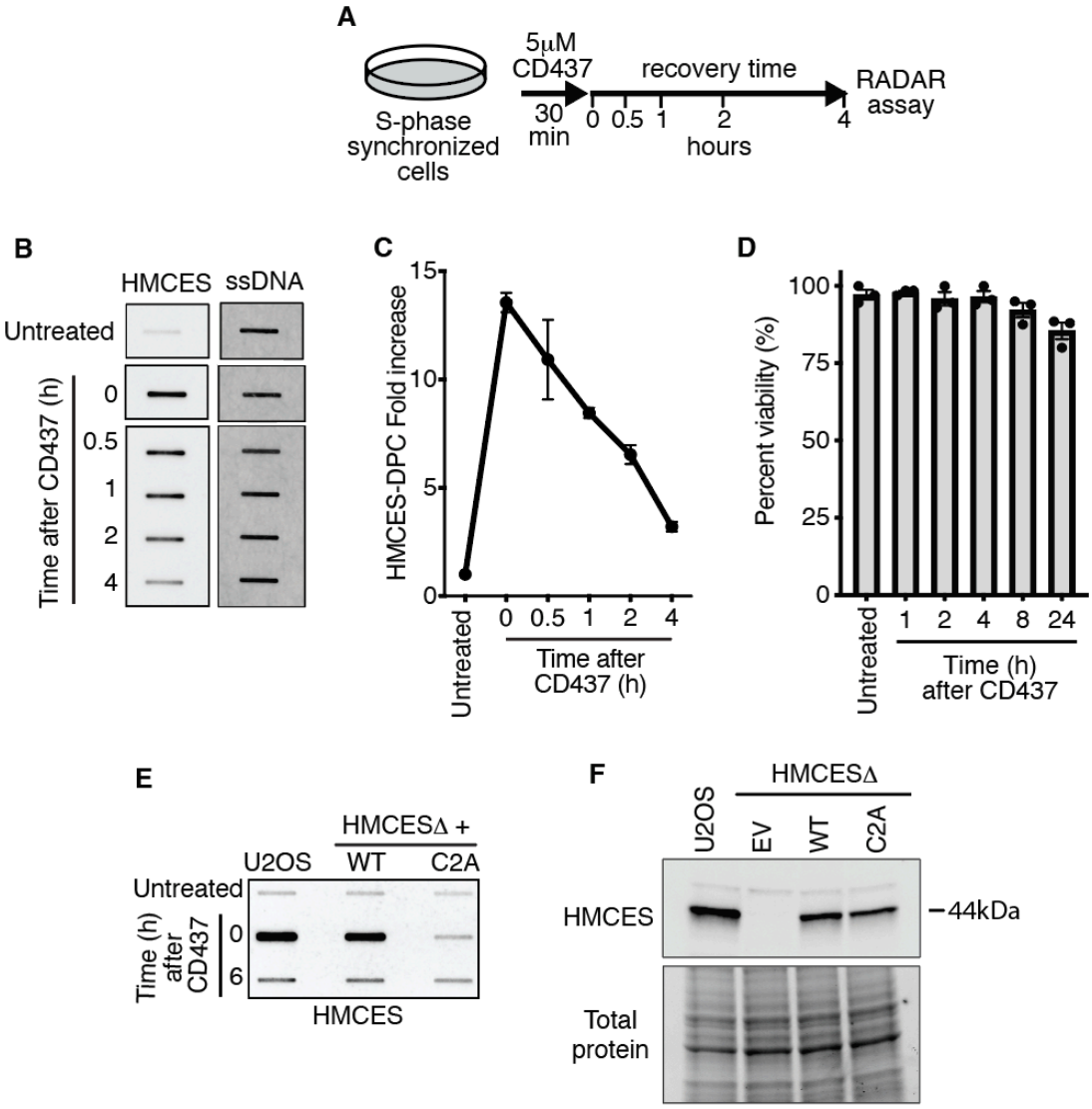

Figure 3.2. CD437-induced HMCES-DPC is removed over time.

(A) Schematic of detection of HMCES-DPC removal in cells after CD437 treatment. Cells were synchronized and released from thymidine for 2 h before treatment with 5 μM CD437 for 30 min. Samples were taken immediately after CD437 treatment (0 h) and at the indicated time points. (B) Representative image of HMCES-DPC resolution assay. HMCES-DPC was detected with an HMCES antibody. The samples were probed on the same blot, but the image was cut and reordered for the figure. The amount of DNA in each sample was quantified by blotting with a ssDNA antibody. (C) Quantification of B (Mean ± SEM, n = 3). (D) Cells were treated with 5μM CD437 for 30 minutes and then incubated in fresh media. Cell viability was measured at the indicated times using trypan blue staining. (E) Representative HMCES-DPC RADAR assay of parental U2OS or HMCESΔ cells complemented with WT or C2A HMCES and treated with 5μM CD437 or DMSO (untreated) for 30 min. Cells were immediately harvested (T = 0) or the CD437 was removed and cells were allowed to recover for 6 h before harvesting. (F) Western blot of cells used in E. Total protein loading was obtained by Free stain imaging.

CD437 system allows tracking of HMCES-DPC removal over time

HMCES-DPCs produced by $KBrO_3$ are resolved over time[47]. To evaluate if CD437-induced HMCES-DPCs are resolved similarly and to measure the rate of DPC resolution, S-phase synchronized cells were treated with CD437 for 30 min and then they were allowed to recover in normal growth media and harvested at varying time points to analyze HMCES-DPC levels (Fig 2.2A). I observed a strong HMCES-DPC signal immediately after CD437 treatment that declined rapidly during recovery (Fig 2.2B). Quantification shows an HMCES-DPC half-life between 1 and 2 h, and approximately 80% of the HMCES-DPC is removed after 4 h (Fig 2.2C). To verify that the HMCES-DPC signal does not go down because of cell death, I checked cell viability in cells exposed to the same dose of CD437 and noticed that cells remained viable for at least 8 h after exposure (Fig 2.2D). To confirm that CD437-induced HMCES-DPCs are formed via a C2 linkage, I tested DPC formation in cells in which endogenous HMCES was deleted by gene editing (HMCESΔ), and either wild-type (WT) or an HMCES C2A mutant protein that is unable to crosslink to the abasic site was expressed by retroviral integration. As expected, no HMCES-DPCs were detected in cells expressing HMCES C2A after CD437 treatment, whereas cells complemented with WT HMCES had an equivalent DPC level to the parental U2OS cells (Fig 2.2E). The HMCES C2A protein was expressed at similar levels to WT (Fig 2.2F). Thus, we conclude that CD437 promotes HMCES-DPC formation in cells by increasing AP sites in ssDNA and provides a quantifiable system to analyze HMCES-DPC removal.

Discussion

CD437 is a good reagent to promote HMCES-DPC formation. This drug acts in ssDNA by inactivating lagging strand synthesis. It produces HMCES-DPC levels that are easily detected by the RADAR assay in a short amount of time (30 min). Additionally, the removal of HMCES-DPC starts right after the drug is removed, which allows the time course analysis, clearly separated from the synthesis process. Another advantage of using CD437 is that the HMCES-DPC signal relies on the catalytic activity of Cys2, suggesting that is crosslink-dependent which follows the model of HMCES crosslinking to the open form of an AP site. However, it is still unknown if HMCES can crosslink to other substrates besides AP sites. The removal of HMCES-DPC seems to have a half-life of 1-2 h, which gives a time frame to test different pathways that might be directing this process (proteolysis, self-removal, etc.) and see if inhibiting them can stop or delay HMCES-DPC resolution.

How is CD437 generating AP sites?

CD437 is a good agent for analyzing HMCES-DPC removal. Strikingly, this drug produces high HMCES-DPC levels, which can only be formed by the high presence of AP sites in ssDNA. This can be explained due to the inhibition of Polα by CD437 which enables strand uncoupling during DNA replication. This increases the amount of ssDNA due to the continuous unwinding of the double helix mediated by CMG[118]. We also noticed that there is a reduction but not complete abolition in fork restart, which seems possible because CD437 does not inhibit leading strand synthesis[118].

Most of the HMCES-DPC formed by CD437 were reduced by UGI expression, suggesting that the AP sites are formed mostly from glycosylase activity. Consistently, previous studies on the role of HMCES in SHM showed that HMCES operates downstream of Family I UDGs, called UNG[71]. The UDG superfamily is classified into six families based on their substrate specificity and UNGs are highly specific for uracil[119]. Other UDGs like the MUG family can excise thymine from G-T mismatches and other kinds of substrates such as 5-FU, 5-hydroxymethyluracil (5-hmU), and 3,N4-ethyenocytosine[120]. It is still unclear if UGI is specific to UNG or is preventing the removal of other substrates besides uracil that might be present in the parental lagging strand.

Another possibility is that uracil is the predominant modification in ssDNA after CD437 treatment due to cytosine deamination. In E*. coli*, there is more cytosine deamination at replication forks in the lagging-strand template than in the leading-strand template[121]. In human cells, APOBEC3A- and APOBEC3B-induced mutagenesis primarily results from the deamination of the lagging strand during DNA replication[122]. More detailed analysis is necessary to understand the high levels of AP sites upon CD437 treatment, but so far it seems to be mostly related to uracil removal.

Considerations on CD437

I used CD437 to prevent lagging strand synthesis and accumulate ssDNA. However, Polα initiates DNA synthesis on both, the leading and the lagging strands by synthesizing an RNA/DNA hybrid primer that is used by Polε and Polδ, respectively, to extend DNA during replication[123]. Ercilla and collaborators reported that a low dose of CD437 in a short amount of time does not affect overall DNA synthesis, allowing ssDNA accumulation by strand uncoupling[118]. Thus, my experimental

approach is mostly affecting lagging strand synthesis. Additionally, it is important to acknowledge that CD437 is not naturally present in human cells, and temporal Polα inhibition does not seem a viable regular process in either. Nevertheless, it can be used as a model to study cellular DNA lesions such as AP sites or DNA gaps. Additionally, deficiency in the polymerase alpha primase activity has been associated with some diseases[124].

CHAPTER IV

TESTING IF PROTEOLYSIS IS A MECHANISM TO PROMOTE HMCES-DPC REMOVAL

Introduction

The HMCES-DPC is ubiquitylated in cells and treatment with MG132 delays DPC removal after potassium bromate treatment[47], suggesting a proteasome-dependent degradation. In the *Xenopus* system, Johannes Walter's research group suggested that DPCs can be removed in a replication-coupled manner, where the DPC is removed by proteolysis in a complementary manner by the proteasome and SPRTN[80]. In his model, DPC is a target of polyubiquitylation, which is partially dependent on the E3 ligase TRAIP[82]. Later on, the Duxin lab showed in the same system that E3 ligase RFWD3 targets PCNA and HMCES-DPC for ubiquitylation[88]. They also provide evidence that HMCES is SUMOylated. Accordingly, a mass spectrometry analysis of formaldehyde-induced SUMOylation changes showed C3orf37 as a SUMOylation target[96], but there is still no evidence that is required for HMCES-DPC removal.

In 2021, Daniel Semlow and collaborators proposed that HMCES-DPC is degraded by SPRTN as an intermediate step in the AP-ICL repair pathway. In this model, fork convergence at an AP-ICL activates NEIL3-dependent unhooking, exposing the AP site at a ss/dsDNA junction, where it is crosslinked by HMCES and promoting HMCES-DPC hydrolysis by SPRTN[110]. *In vitro* data, shows that FANCJ unfolds HMCES-DPC and exposes the DNA underlying, which might allow SPRTN to bind and promote DPC degradation[111]. HMCES-DPC degradation by SPRTN as an intermediate step in the AP-ICL repair pathway does not involve proteasome activity. Additionally, protease FAM111A is predicted to interact with HMCES, however, there is no direct evidence it induces HMCES-DPC removal[125].

In this Chapter, I use the CD437-mediated HMCES-DPC removal system to study if proteolysis is mediating HMCES-DPC resolution. Here I find that the proteasome and SPRTN have a modest and null effect, respectively, on HMCES-DPC removal. HMCES-DPC is ubiquitylated, however, it does not seem required for its crosslink nor removal.

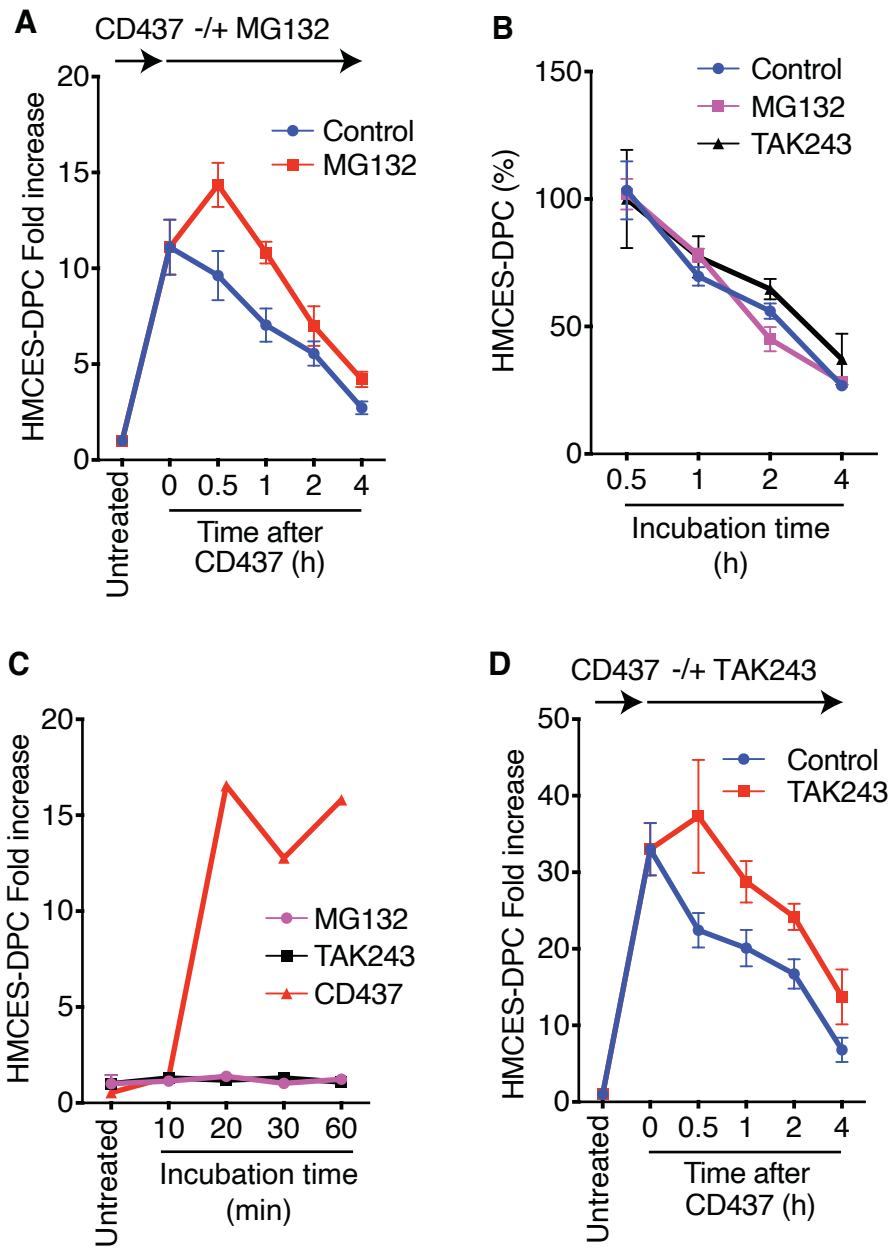

Figure 4.1. Proteasome inhibition does not prevent HMCES-DPC removal.

(A) RADAR assay of HMCES-DPC levels. A dose of 10μM MG132 was added immediately after CD327 treatment (Mean ± SEM, n = 5). (B) Percentages remaining of HMCES-DPC with HMCES-DPC levels at 0.5h set to 100%. RADAR assay was performed as A. but also testing TAK243 (C) RADAR assay of HMCES-DPC formation during incubation with CD437, MG132, or TAK243 for the indicated times. (D) RADAR assay of HMCES-DPC levels. A dose of 10μM TAK243 was added immediately after CD327 treatment (Mean ± SEM, n = 4)

Results

Testing the proteasome activity in HMCES-DPC removal

Incubation with MG132 after $KBrO_3$ treatment delays HMCES-DPC resolution, suggesting a proteasome-dependent degradation[47]. To test the activity of the proteasome in HMCES-DPC removal, I utilized our CD437 system to track DPC levels in the absence or presence of MG132. Treating cells with MG132 increased the total amount of HMCES-DPC formed after CD437 incubation by approximately 25% 30 min after removal (Fig 4.1A). However, MG132 did not prevent the resolution of HMCES-DPCs, which proceeded at least as quickly after the 30 min time point as vehicle-treated cells (Fig 4.1B). Incubating cells with MG132 alone in the absence of CD437 does not induce HMCES-DPC formation (Fig 4.1C); therefore, the increase within the first 30 min of release from CD437 required ssDNA formation. I also observed similar results by inhibiting the Ubiquitin-like modifier-activating enzyme 1 (UBA1 E1) enzyme with TAK243[126], which blocks protein ubiquitylation (Fig 4.1D). These results suggest that ubiquitylation and the proteasome are not essential for resolving the HMCES-DPC, although they may affect the amount of DPC formed or retained at early time points after CD437 treatment.

Testing if ubiquitylation is required for HMCES-DPC removal

Since MG132 and TAK243 are expected to increase and decrease the ubiquitylation of HMCES-DPC respectively, I decided to verify the activity of the drugs by doing an SDS-PAGE electrophoresis analysis of the samples used in the RADAR assays to detect any PTM present on the HMCES-DPC. The results show that MG132 increases HMCES-DPC levels with strong smearing only during the first 30 min and then rapidly goes down, meanwhile, TAK243 samples show also an increase of HMCES-DPC signal compared to the Control but reduced smearing during the early time points suggesting that HMCES-DPC is ubiquitylated (Fig 4.2A top membrane). A ubiquitin blot verifies the high and low signals in the samples from the MG132 and TAK243, respectively (Fig 4.2A bottom membrane). Additionally, Courtney Lovejoy and Atherva Garje in our lab, developed stable cell lines that express HMCES fused with a deubiquitinate enzyme (DUB) and its inactive form (DUBi) which prevent and allow ubiquitination, respectively (Fig 4.2B). The deubiquitinate enzyme was cloned from a virus catalytic site which was previously tested in yeast[127]. Cell lines express the fused proteins accordingly (Fig 4.2C), however, the activity of the DUB enzyme blocking HMCES-DPC ubiquitylation has not been verified yet. As a

consequence of this limitation, I would consider any further results as preliminary data. Next, I used these cell lines to ask whether inhibiting direct HMCES-DPC ubiquitylation would prevent its removal. The results show that HMCES-DPC DUB and HMCES-DPC DUBi are removed similar to WT HMCES-DPC (Fig 4.2D). Accordingly, to what was previously shown using TAK243, where inhibiting overall ubiquitylation did not interfere with HMCES-DPC removal. Collectively, additional experiments are required to determine if ubiquitylation is important for HMCES crosslink or removal.

Testing if Neddylation is promoting HMCES-DPC removal

Neddylation is a posttranslational modification that promotes the activation of Nedd8, a ubiquitin-like molecule that depends on an E1, E2, and E3 enzyme to be activated and modify its substrate[128]. Nedd8 activates the largest ubiquitin E3 ligase family, the Cullin-RING ligases. There are approximately 600 Ub E3 ligases in eukaryotic cells. If HMCES-DPC is degraded by the proteasome, testing the neddylation pathway would narrow the candidates that might mediate HMCES-DPC ubiquitination. The small molecule MLN4924 is a very efficient Nedd8 inhibitor[129]. It binds NEDD8 Activating Enzyme E1 Subunit 1 (NAE1) enzyme, preventing Nedd8 activation and, consequently, Cullin-dependent E3 ligases activation. Therefore, I am using this inhibitor to test if inactivation of Cullin-dependent E3 ligases, can impair HMCES-DPC removal (Fig 4.3A).

I treated synchronized HCT116 cells with CD437 (5μM) for 30min, cells were incubated with MLN4924 (10μM) and analyzed in a time course for 4 hours. HMCES-DPC were stabilized mostly during the first 30 min after treatment (Fig 4.3B) and then rapidly removed, showing a similar curve as TAK243. Likewise, incubation of cells with MLN4924 and no previous CD437 treatment did not induce HMCES-DPC formation even after 1 hour (Fig 4.3C). To test if Neddylation is mediating the PTM previously observed, I performed an SDS-PAGE with RADAR samples of cells incubated with MG132, TAK243, or MLN4924 after CD437 treatment. Results show the same smearing is reduced when inhibiting ubiquitination or neddylation (Fig 4.3D top), suggesting that HMCES-DPC is ubiquitylated during those first 30 min after removing CD437 and this may be mediated by Cullin-RING ligases. Interestingly, TAK243 also reduces Nedd8 active levels (Fig 4.3D bottom). UBA1 can activate Nedd8 when the last one is in excess, and conjugate it to a ubiquitin-specific E2 enzyme and then be part of a mixed ubiquitin-nedd8 chain[130]. Thus, the inactivation of UBA1 by TAK243 may also be affecting Nedd8 activation.

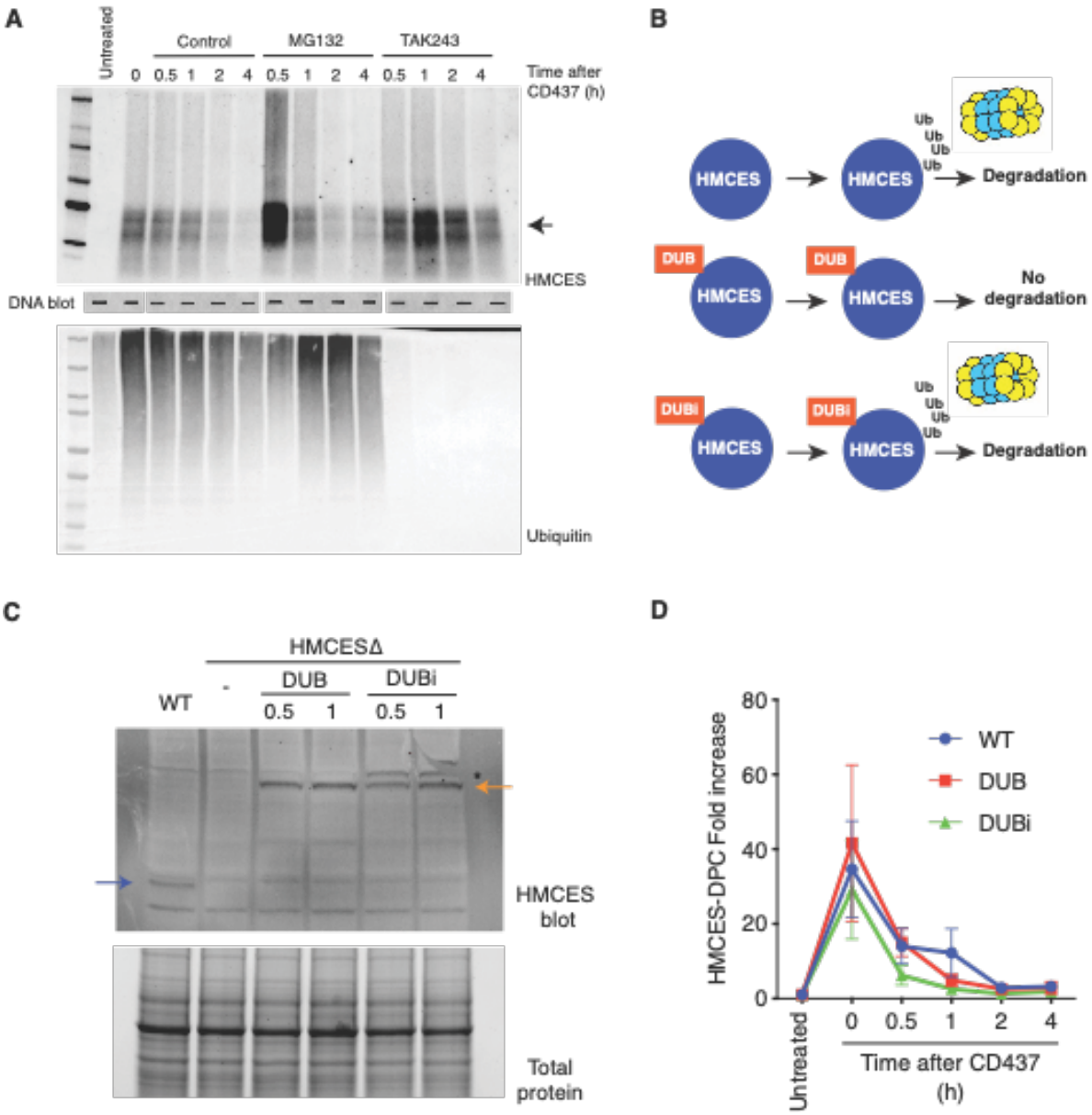

Figure 4.2. Ubiquitylation of HMCES-DPC is not required for its removal.

(A) SDS-PAGE using RADAR samples of cells treated as described. The top membrane is blotting against HMCES. The black arrow points to HMCES-DPC. The DNA blot below shows the amount of DNA that was loaded for each sample. The bottom membrane is blotting against Ubiquitin. (B) Diagram showing HMCES WT, HMCES DUB, and HMCES DUBi and the expectations for each protein degradation. (C) Western blot of cells expressing wild-type HMCES (blue arrow) and the HMCES-fused forms: DUB and DUBi (orange arrow). Virus infection dilutions (0.5 and 1) are used for making cell lines. (*) unspecific band. (D) RADAR assay showing HMCES-DPC resolution after a 5μM CD437 treatment. Experiment C was performed by Atherva Garje.

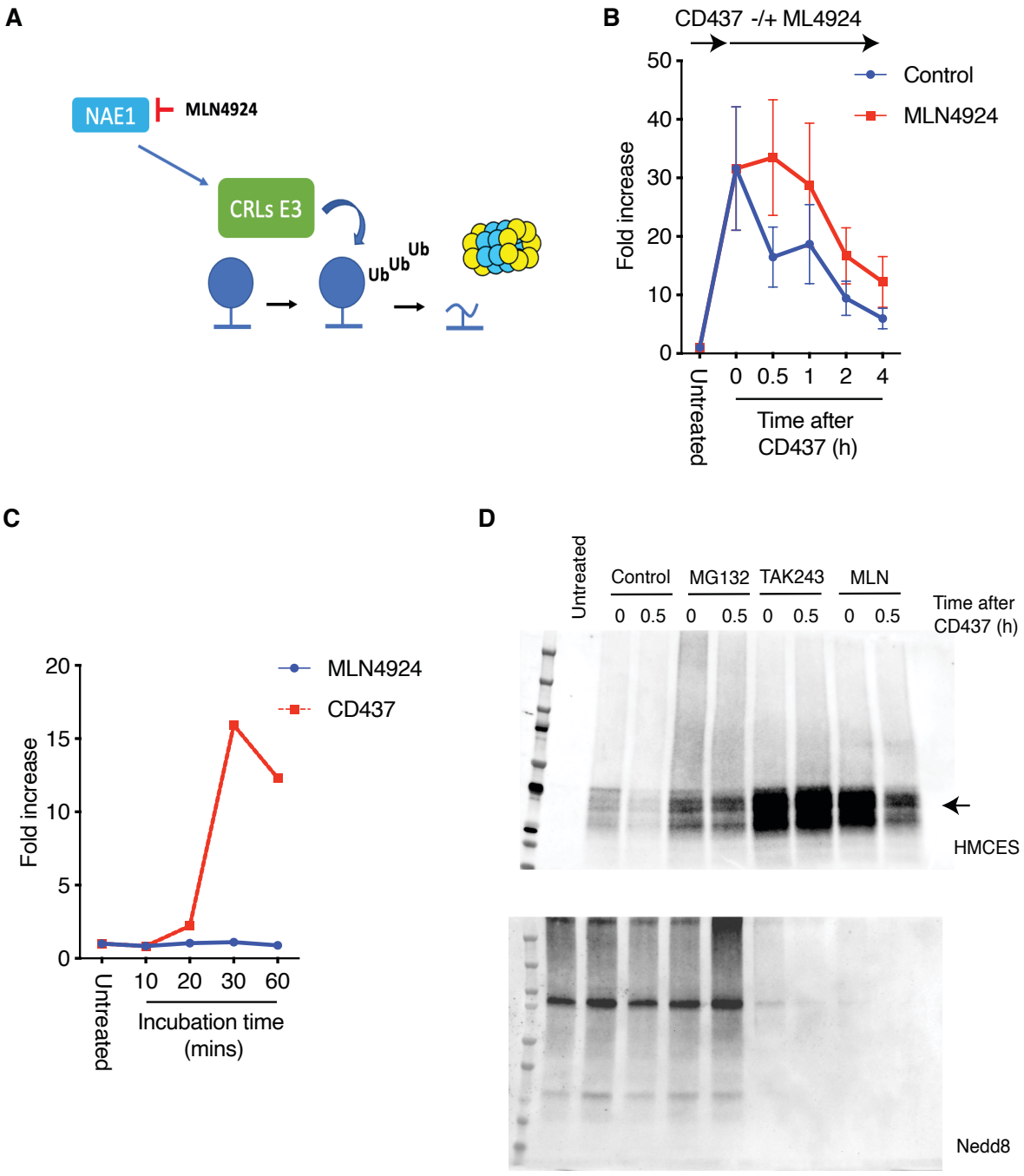

Figure 4.3. Neddylation is not necessary for removing HMCES-DPC.

(A) Diagram of MLN4924 activity on preventing proteasome-dependent degradation. (B) RADAR assay of HMCES-DPC levels. A dose of 10μM MLN4924 was added immediately after CD327 treatment (Mean ± SEM, n = 5). (C) RADAR assay of HMCES-DPC formation during incubation with CD437, or MLN4924 for the indicated times. (D) SDS-PAGE using RADAR samples of cells treated as described. The top membrane is blotting against HMCES. The black arrow points to HMCES-DPC. The bottom membrane is blotting against Nedd8.

In summary, neddylation might be inducing ubiquitylation of HMCES-DPC but is not completely required for its removal.

Testing SPRTN activity on HMCES-DPC removal

SPRTN can remove HMCES-DPC during ICL repair in the *Xenopus* extracts[109], so I tested whether the SPRTN protease is important for HMCES-DPC resolution in human cells. SPRTN is essential for mammalian cell survival[131]; therefore, I used small interfering RNA (siRNA) to deplete SPRTN acutely and test its activity. Knockdown efficiency was confirmed by immunoblotting (Fig 4.4A) and we also verified that SPRTN was functionally inactivated by measuring replication fork speed, which was previously shown to be slowed by SPRTN inactivation[99] (Fig 4.4B). Depletion of SPRTN did not prevent CD437-induced HMCES-DPC removal (Fig 4.4C). Additionally, I tested a possible redundancy between SPRTN and the proteasome; however, MG132 did not prevent HMCES-DPC removal in SPRTN knockdown cells (Figure 4.4D). Therefore, our data suggest that neither the proteasome nor SPRTN activity is critical to removing HMCES-DPC induced by Polα inhibition.

Discussion

The proteasome is not required for HMCES-DPC removal

Proteasome inhibition cannot prevent HMCES-DPC removal in the CD437 system. However, it clearly shows an increase in HMCES-DPC levels during the early times after CD437 removal. I do not know if MG132 itself is inducing ssDNA or AP sites that can lead to HMCES-DPC formation during those 30 mins, but it seems unlikely since the other drugs used had a similar effect. We tried to test this by the ARP assay but the results were not conclusive. It is also possible that the proteasome mediates HMCES-DPC degradation during that small fraction of time and then a complementary pathway takes place. Larsen and collaborators suggested that DPC degradation at the replication fork is achieved by the activity of the proteasome and SPRTN[132]. However, I did not see an additive effect when using MG132 in SPRTN knockdown cells (Fig 4.4D). Besides SPRTN, other proteases known to target DPCs at the replication fork might complement the proteasome, for example, FAM111A[94], ACRC[96], and Ddi1[91]. From those, I tested FAM111A with similar results (data not shown). However, HMCES-DPC levels in siFAM111A cells right after CD437 treatment were higher than siNT cells, suggesting a possible regulation during those 30

mins of incubation with the drug. Unifying these observations, a complementary pathway to remove HMCES-DPC along with the proteasome would require a deeper analysis.

Posttranslational modification of HMCES-DPC

There is evidence that HMCES undergoes PTMs[47,88,109]. Semlow and collaborators showed that during the ICL repair pathway, proteasome inhibition increased HMCES-DPC ubiquitylation but it did not prevent its removal[109]. Similarly, in my results, proteasome inactivation also increased HMCES-DPC PTMs, which were reduced during ubiquitin inhibition and it did not prevent crosslink removal either (Fig 4.2A). Furthermore, Inhibition of neddylation also showed a decrease in PTMs (Fig 4.3D), suggesting that it is mediated by Cullin-dependent E3 ligases, which are activated by Neddylation. CRLs are classified according to their cullin domain, which will bind adaptor and receptor proteins that allow targeting specific substrates[133]. In an attempt to keep studying a proteasome-dependent mechanism, I individually inactivated cullins by siRNA. After achieving decent knockdown efficiency, I did not see a delay in HMCES-DPC removal in U2OS cells transfected individually with siCUL1, siCUl2, siCUL3, or siCUL4A (data not shown). One possible explanation is that the cullin proteins remaining are inducing degradation (CUL4B, CUL5, or CUL7) or that they are working cooperatively, supplying the activity of the absent cullin and making this analysis more challenging. Additionally, it is still elusive if other PTMs such as SUMOylation and PARylation promote HMCES-DPC removal since these pathways are also involved in DPC degradation, as described in an important review by Leng and Duxin[78].

HMCES is ubiquitylated but is also possible that ubiquitylation is not targeting degradation. Duxin's research group showed that HMCES-DPC is ubiquitylated by RFWD3[88] which is known to mediate ubiquitylation of RPA and PCNA at replication forks. However, RPA ubiquitylation is not a K48 linkage that leads to proteasome degradation[134]. Alternatively, ubiquitylation could be required for HMCES crosslink, to recruit other proteins to the HMCES-DPC, or might be relevant after HMCES is removed from the DNA (in case is not by proteolysis). To test these possibilities, I used the HMCES-DUB cells, which did not show any impairment in crosslink nor removing HMCES (Fig 4.2C), suggesting that ubiquitylation was not required for these processes. Strikingly, I could detect some differences between soluble HMCES-DUB and HMCES-DUBi after CD437 treatment, meaning that ubiquitylation might be regulating soluble HMCES instead (Fig 6.1B). However, the activity of these cell lines preventing direct HMCES ubiquitylation has not been verified yet, which makes it difficult to make any conclusion.

Gallina and collaborators also provided evidence that HMCES is SUMOylated[88]. DPC can be targeted for SUMOylation which can lead to the recruitment of Ub E3 ligases[135] or proteases such as SPRTN[97], ACRC[136], and Wss1[137]. In my results, I can still detect some upper smearing on HMCES-DPC even if ubiquitylation was depleted by TAK243 (Fig 4.2A), suggesting another PTM is present. HMCES is predicted to contain around 14 residues on its surface that are susceptible to ubiquitylation, another 14 residues for phosphorylation, and 3 for acetylation (PhosphositePlus.org). To study the importance of ubiquitylation, a previous postdoc in the lab, Kareem Mohni, mutated all the Lys at the HMCES surface, but that protein cannot even be expressed in cells. It will be important to understand more about the HMCES PTMs since they might provide other functions besides crosslinking to the AP sites.

SPRTN is not required for HMCES-DPC removal

The lack of a strong requirement for SPRTN in HMCES-DPC removal in the CD437-treated human cells could be because I did not achieve sufficient SPRTN inactivation with siRNA. Using a degron system[138] would be a good alternative approach. Another possibility is that CD437 is not generating an appropriate substrate for SPRTN, which removes DPC efficiently when they are located in ss/dsDNA junctions[139]. Since many forks do not restart, these CD437-generated ssDNA AP sites may not be present in ss/dsDNA junctions. Thus, SPRTN would not be able to degrade HMCES because it has weak binding to ssDNA and is inactive on that type of DNA substrate. Finally, different proteases involved in DPC degradation (FAM111A, ACRC, Ddi1, etc.) may be complementing the absence of SPRTN, then simultaneous inactivation would be required to study if different proteolysis pathways are removing HMCES-DPC.

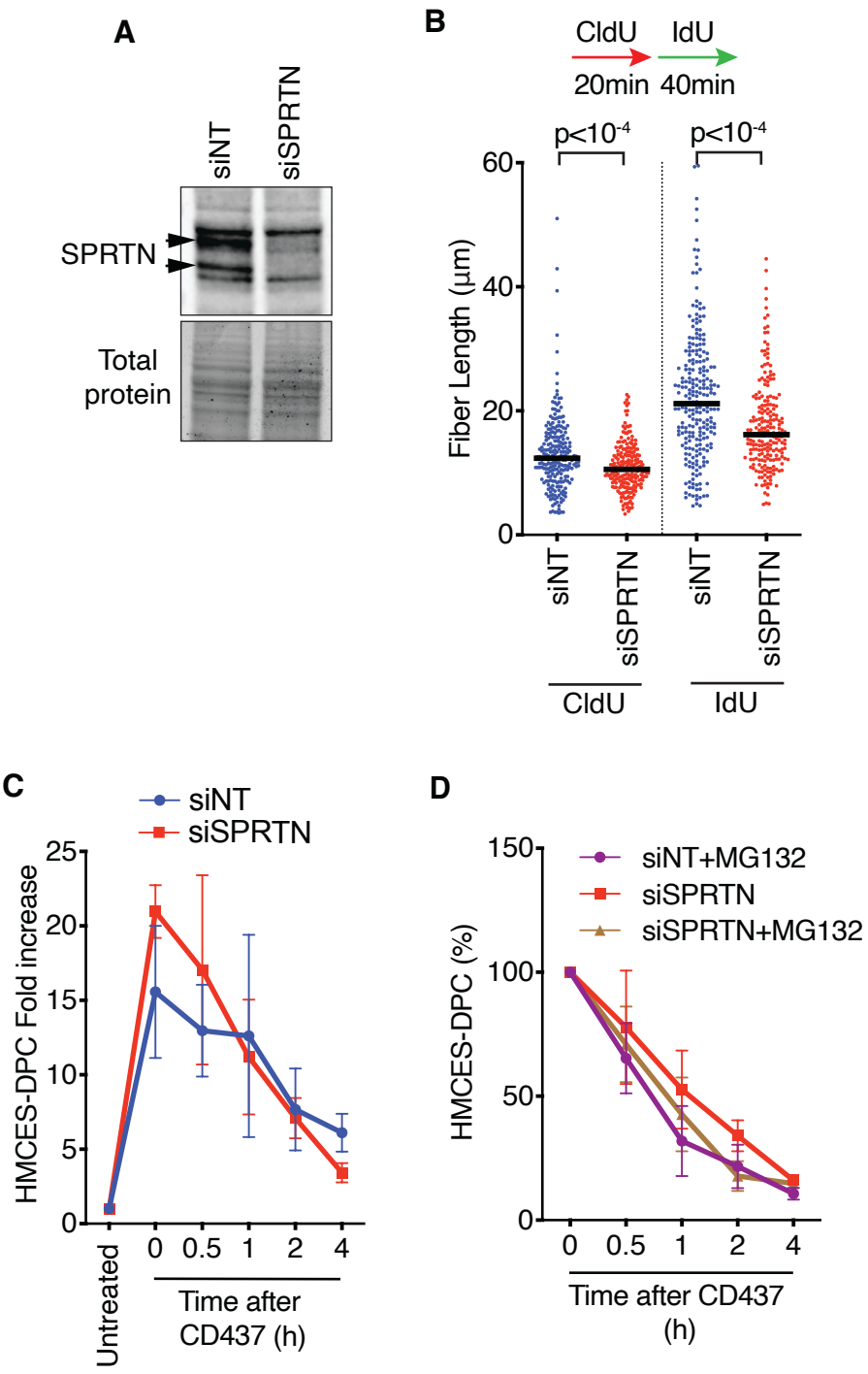

Figure 4.4. SPRTN inhibition does not prevent HMCES-DPC removal.

(A) Immunoblot of cells transfected with non-targeting siRNA (siNT) or siRNA against SPRTN (siSPRTN). (B) DNA combing assay analysis of replication fork speed in cells transfected with siNT or siSPRTN. Bar represents the median, and p values were derived from Mann-Whitney test. (C) RADAR assay of HMCES-DPC levels in cells transfected with indicated siRNAs (Mean ± SEM, n = 3).

CHAPTER V

SELF-REVERSAL FACILITATES THE RESOLUTION OF HMCES-DPC IN CELLS

Introduction

The reaction between the active site of the SRAP domain and the opened-ring form of the AP site favors the Thiazolidine linkage formation. The Eichman lab has recently uncovered the mechanisms of YedK DPC stabilization, where the SRAP domain catalyzes ring opening of abasic sites and reversal of the crosslink[59]. The latter one is delayed by Glu105 and His160 point mutations. This self-reversal mechanism is also present in human HMCES[59] and mutation of its equivalent Glu105 (Glu127) impairs HMCES-DPC reversal *in vitro* as well[140]. The release of HMCES-DPC seems to be determined by DNA context since the crosslink is more stable in ss/dsDNA junctions than in ssDNA, and the presence of dsDNA favors the reversed state of the protein[140,141]. This biochemical evidence suggests that HMCES-DPC reversal allows for transient protection of the AP site until replication is completed in a specific region and places the AP site in a more suitable DNA context for repair, however, there was not yet any cellular evidence of this pathway or its importance for cellular fitness. In this chapter, I find that HMCES Glu127 mediates HMCES-DPC reversal in human cells. Furthermore, I show that disruption of this auto-release mechanism has a deleterious effect on cell viability in the presence or absence of DNA stress. Here I provide evidence that supports a self-reversal mechanism for HMCES crosslink removal.

Results

HMCES Glu127 mediates the self-reversal reaction

Although the HMCES-DPC thiazolidine linkage appears stable and resistant to repair enzymes like AP endonucleases, biochemical experiments using a second ssDNA containing an AP site as a trap showed that it is reversible[59]. In the HMCES bacterial ortholog YedK, the reversal reaction was disrupted in E105Q and H160Q mutants with the former having a stronger effect. To test if the equivalent glutamic acid residue is also important for the self-reversal of human HMCES, I took a similar experimental approach. The HMCES-DPC was formed by incubating the SRAP domain with a 20-nucleotide DNA oligo containing an AP site (DPC-20). Next, I added a 40

nucleotide AP DNA oligo in 50-fold excess to trap any self-reversed protein during incubation at 37 degrees (Fig 5.1A). The DPC-20 slowly decreased over time as a DPC-40 formed accordingly with self-reversal regenerating an intact and active HMCES protein capable of crosslinking again to another available ssDNA AP site (Fig 5.1B). In contrast, the E127Q HMCES-DPC is completely unable to reverse even after 24 hours of incubation (Fig 5.1C). This result indicates that E127 is necessary for human HMCES-SRAP crosslink self-reversal.

HMCES reversal is stimulated by a duplex-forming oligonucleotide

The half-life of the HMCES-DPC self-reversal *in vitro* is longer than 4 hours (Fig 5.1C). This is much greater than the half-life of the DPC seen after CD437 treatment in cells (Fig 3.2C). Moving from one AP-ssDNA site to another requires not only the reversal of the thiazolidine linkage but also the disengagement of DNA binding. Our previous studies showed that even without crosslinking, HMCES has a very high affinity for ssDNA[47]. In contrast, HMCES cannot bind dsDNA. Thus, I reasoned that if a complementary oligonucleotide capable of forming a duplex with the ssDNA to which HMCES is crosslinked was included in the reversal reaction, we may be able to increase the speed at which I could observe the reversal. Indeed, this is the case. After generating HMCES-DPC-40, I added a complemented oligo (c40mer), which shifted the DPC-40 band in an SDS-PAGE gel (dsDPC-40). Next, I added a 50-fold excess of 20mer containing an AP site to trap any HMCES that undergoes self-reversal (Fig 5.2A). As expected, the control sample containing a non-duplex forming 40mer oligonucleotide (T40mer control) showed a reduction of DPC-40 along with the formation of DPC-20. However, the reversal of DPC-40 in the presence of the duplex-forming oligonucleotide is faster compared to the control (Fig 5.2B). After 1 hour of incubation, the dsDNA structure yielded more than 60% reversal, compared to HMCES-DPC in ssDNA, which reached a little more than 20% (Fig 5.2C). This result is consistent with two recent reports that found that DNA duplex formation caused an apparent accelerated self-reversal rate[140,141] and indicates that the intrinsic rate of self-reversal is comparable with the reversal rate observed in cells.

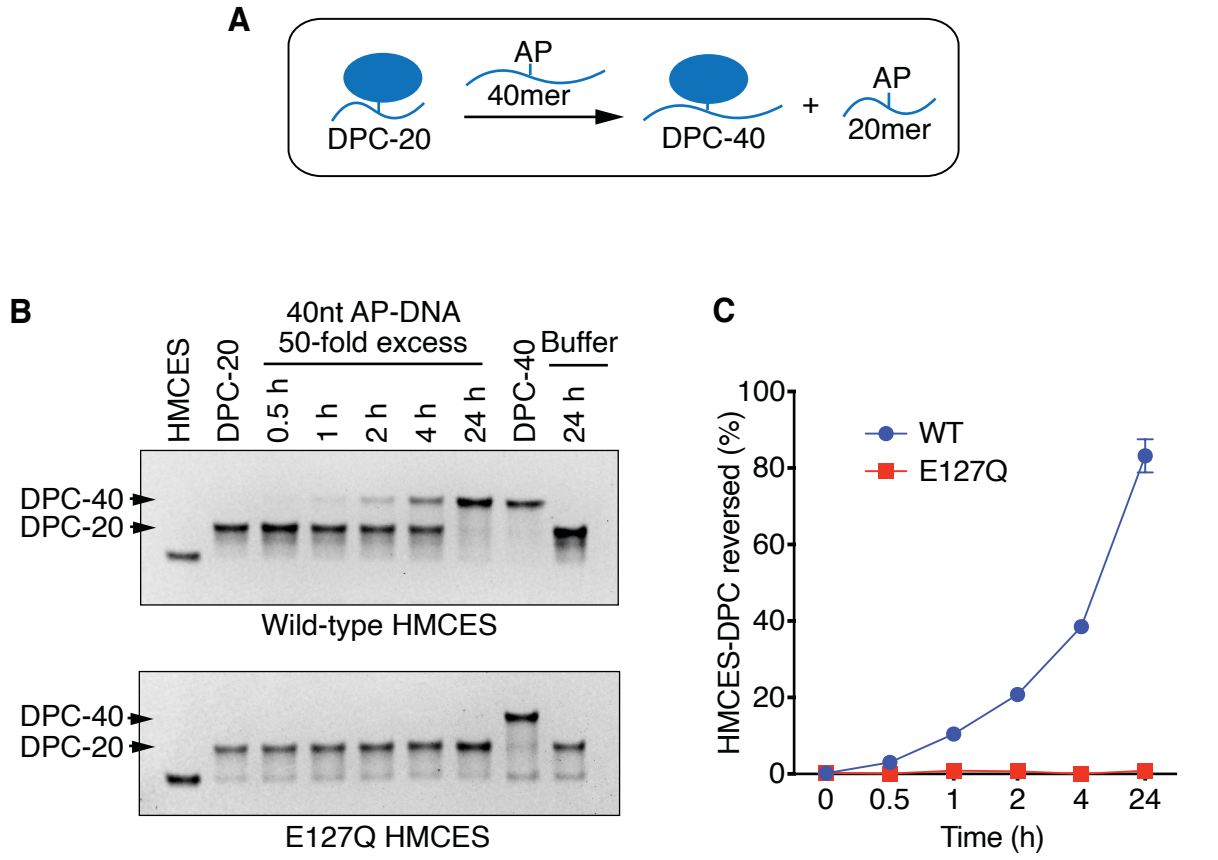

Figure 5.1. Glu127 catalyzes HMCES crosslink self-reversal.

(A) Schematic of HMCES-DPC biochemical self-reversal assay. The AP site containing 40-nucleotide ssDNA oligonucleotide (40-mer) was added at 50X excess. (B) Representative time course experiment of HMCES-DPC reversal with purified WT or E127Q HMCES. The first lane is HMCES protein without any DNA. The free protein and DPC-20 (DPC with 20-nucleotide oligonucleotide) or DPC-40 (DPC with 40-nucleotide oligonucleotide) were visualized by Coomassie-stained SDS-PAGE. (C) Quantification of HMCES-DPC reversal. Percent reversed is the amount of DPC-40 product compared to the total. (Mean ± SEM, n = 3). Dr Katherine Paulin from the Eichman lab trained me to perform the biochemical assays.

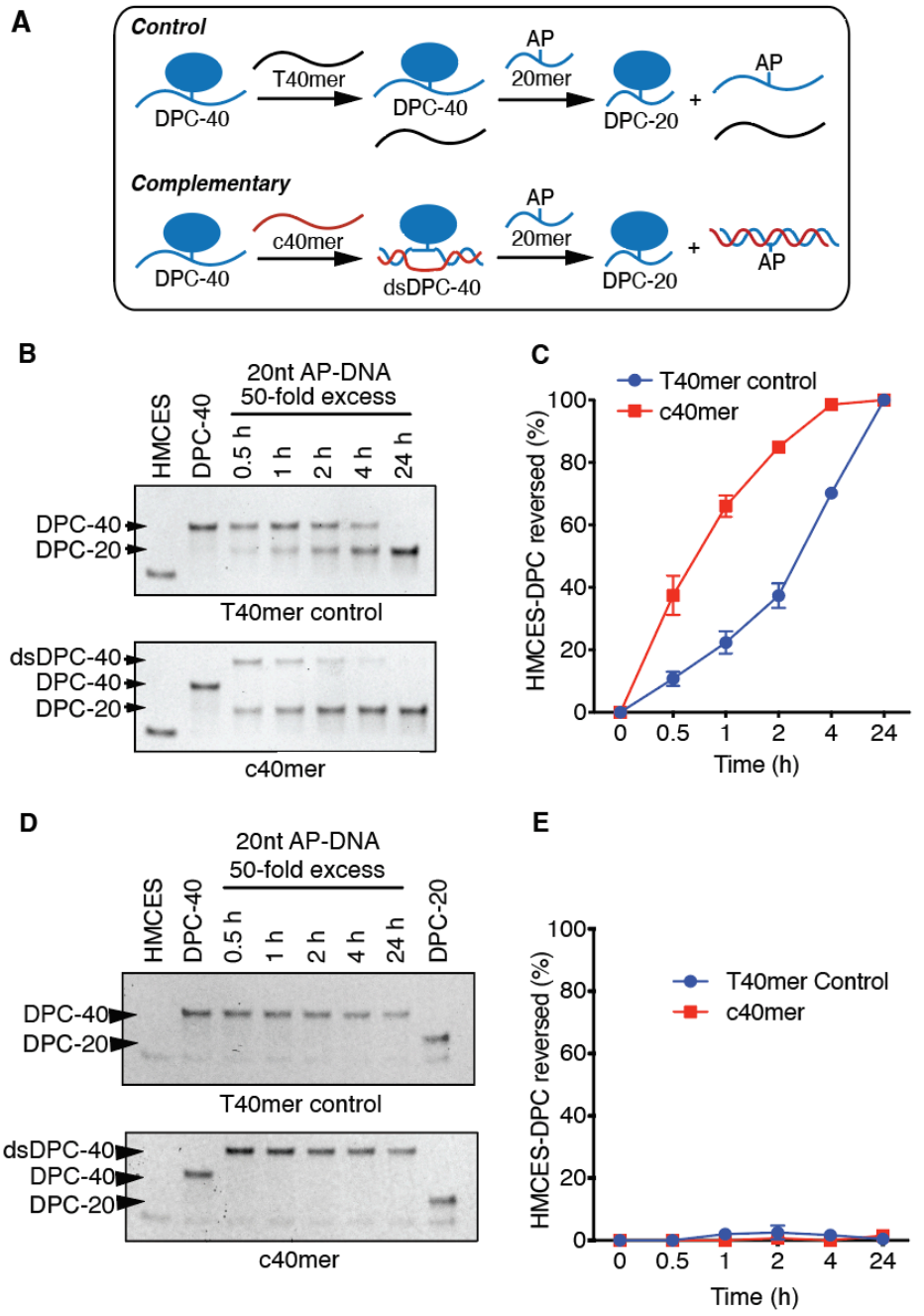

Figure 5.2. Duplex DNA promotes faster reversal of HMCES WT but not E127Q.

(A) Schematic of HMCES-DPC reversal assay comparing ssDNA vs. duplex DNA. The HMCES DPC-40 was incubated with a complementary (c40mer) or non-complementary (T40mer control) oligonucleotide with a 20-nucleotide ssDNA trap containing an AP site. (B) Representative time course of WT HMCES-DPC reversal. The higher migrating band is the product of the hybridization of c40mer with ssDNA HMCES-DPC (dsDPC-40). (C) Quantification of WT HMCES-DPC reversal from ssDNA and dsDNA. Percent reversed is the amount of the DPC-20 product compared to the total. (Mean ± SEM, n = 3). (D) Representative time course of HMCES-DPC reversal of E127Q mutant with the same size expected products than B. (E) Quantification of E127Q HMCES-DPC reversal from ssDNA and dsDNA. (Mean ± SEM, n = 3).

HMCES E127Q cannot reverse its crosslink. Since duplex DNA formation prevents the re-crosslink of HMCES to the AP site and accelerates the detection of the self-reversal process, I anticipated that the E127Q mutant crosslink would not be affected by changing its DNA substrate. Results showed that I was correct. There is no DPC-20 band formation either in the control or in the samples where the DPC was incubated with the complementary oligo (Fig 5.2D). I can still detect the band shift of dsDPC-40, which confirms that the duplex DNA was formed but it did not promote any self-reversal process in the E127Q mutant even after 24 h of incubation (Fig 5.2E). Stingele's research group showed similar results on HMCES E127A mutant[113], supporting the role of Glu127 mediating self-reversal.

HMCES Glu127 mediates HMCES-DPC removal in cells

To test if HMCES-DPC auto-reversal is an important resolution pathway in cells, I made use of the E127Q HMCES protein. First, I complemented the HMCESΔ cells to create stable cell lines expressing only the E127Q or WT HMCES (Fig 5.3A). Analysis of HMCES-DPC levels in untreated, synchronized cells showed more DPC in E127Q cells compared to WT cells (Fig 5.3B), suggesting that the crosslinked state is increased in cells expressing the mutant even in the absence of added genotoxic stress. I next treated the E127Q or WT HMCES expressing cells with CD437 to induce HMCES-DPC formation and tracked DPC resolution over time. The E127Q HMCES-DPC formed a similar total level of DPC as wild-type cells 30 minutes after CD437 treatment (Fig 5.3C), although the fold increase when compared to the untreated cell control was less since it started at a higher basal level. Strikingly, the resolution kinetics of the E127Q was significantly delayed by at least one hour compared to wild-type HMCES (Fig 5.3D and 5.3E), suggesting that E127-dependent self-reversal is an important process in cells. After 2 hours of recovery, the DPC level was reduced almost to approximately the same amount in both cell lines, suggesting alternative mechanisms to complete removal in the absence of self-reversal.

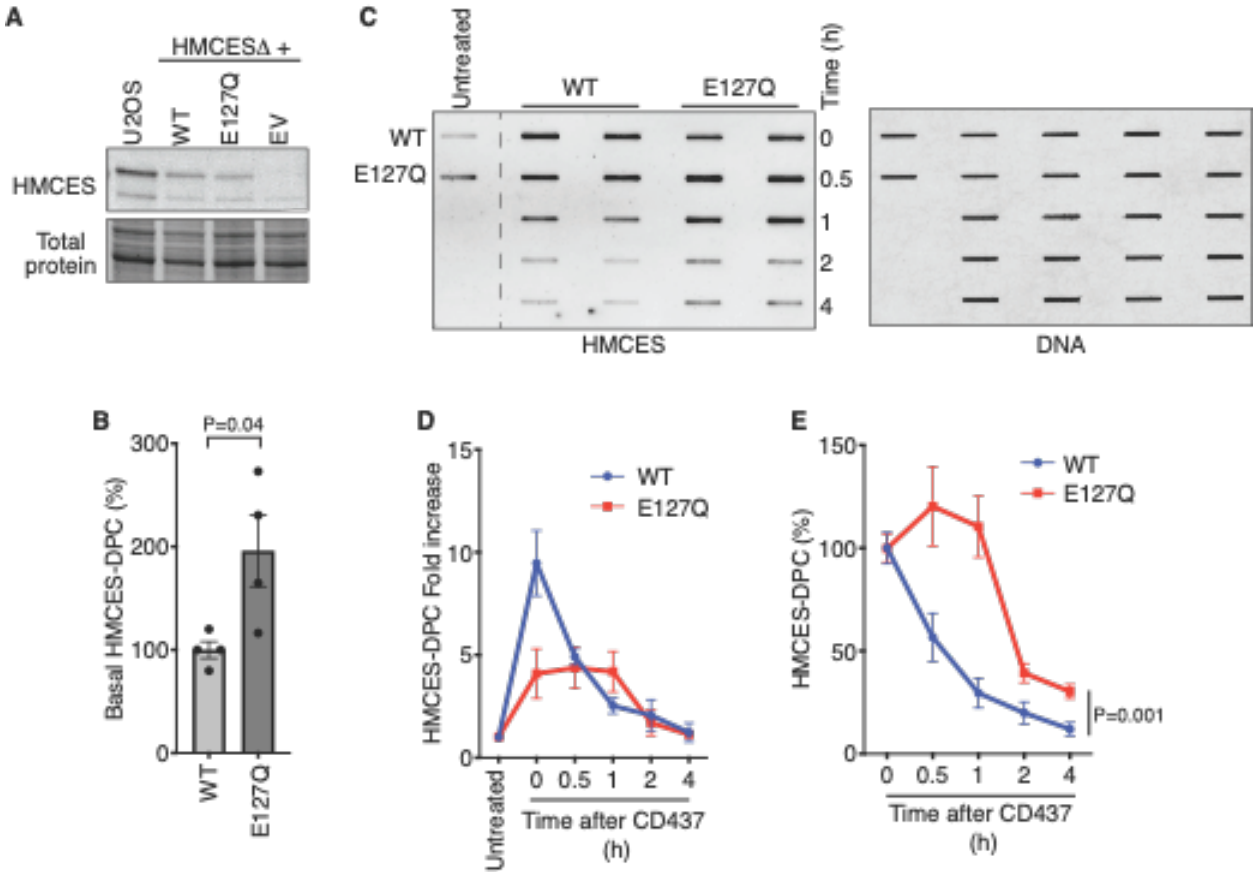

Figure 5.3 Inactivating HMCES self-reversal delays DPC resolution in cells.

(A) Immunoblot of U2OS or HMCESΔ cells complemented with WT HMCES, E127Q HMCES, or EV. (B) Quantification of HMCES-DPC levels in untreated cells that express WT or E127Q HMCES. RADAR assay was performed with samples from S-phase synchronized cells. Mean ± SEM, n = 4, two-tailed t-test. (C) Representative image of RADAR assay to detect HMCES-DPC from WT- or E127Q-HMCES-expressing cells. (D) RADAR assay of HMCES-DPC levels from WT- or E127Q-HMCES-expressing cells. Mean ± SEM, n = 4. (E) Quantification of HMCES-DPC levels with time zero immediately after CD437 set at 100%. (Mean ± SEM, n = 4, two-way ANOVA).

SPRTN affects levels of E127Q HMCES

The E127Q HMCES-DPC starts to get removed 1h after recovering from CD437 treatment, suggesting that there might be another pathway taking place to remove HMCES-DPC when self-reversal is impaired. To test whether this other pathway is proteasome-dependent, I performed the RADAR assay with the E127Q cells but added TAK243 to inhibit ubiquitylation. I detected a small increase in DPC levels during the first hour of incubation but then rapidly came down (Fig 5.4A). To test the activity of SPRTN, I used siRNA to deplete the protease in the E127Q cells. The results showed bigger DPC levels during the first hour compared to TAK243 incubation (Fig 5.4B), suggesting a more active role of SPRTN in removing HMCES when unable to self-reverse. In both cases, E127Q HMCES-DPC is not completely stabilized along the entire time course. Possibly, a combination of both pathways is required or a different protease is involved. Later on, it got my attention that siSPRTN had a bigger effect compared to its previous analysis on HMCES-DPC removal in wild-type U2OS cells (Chapter IV, Fig 4.4). Thus, I asked whether SPRTN knockdown was affecting total HMCES levels in the absence of damage. Strikingly, SPRTN depletion increased total E127Q HMCES levels compared to WT cells by more than a 2-fold increase (Fig 5.4C and D). These results indicate that SPRTN might regulate E127Q HMCES levels and partially promote their removal when inducing DPC formation with CD437.

Effects of expressing the self-reversal-deficient E127Q HMCES protein

HMCESΔ cells accumulate DSBs[47]. Expressing endogenous levels of the WT or E127Q HMCES protein reduced the DSBs as measured by a neutral comet assay, indicating that the E127Q protein can at least partly protect abasic sites from cleavage (Fig 5.5A).

DPC accumulation is detrimental to cells. Hence, I analyzed how the expression of the E127Q HMCES protein affects cell fitness. Without the addition of exogenous DNA-damaging agents, cells expressing near-endogenous levels of E127Q HMCES exhibited 30% less growth/viability as measured by an alamarBlue assay than HMCESΔ cells containing an empty vector (EV) or HMCESΔ cells expressing WT HMCES (Fig 5.5B). I noticed that the growth defect of the E127Q-expressing cells diminished over time as they were maintained in culture, suggesting an adaptation mechanism.

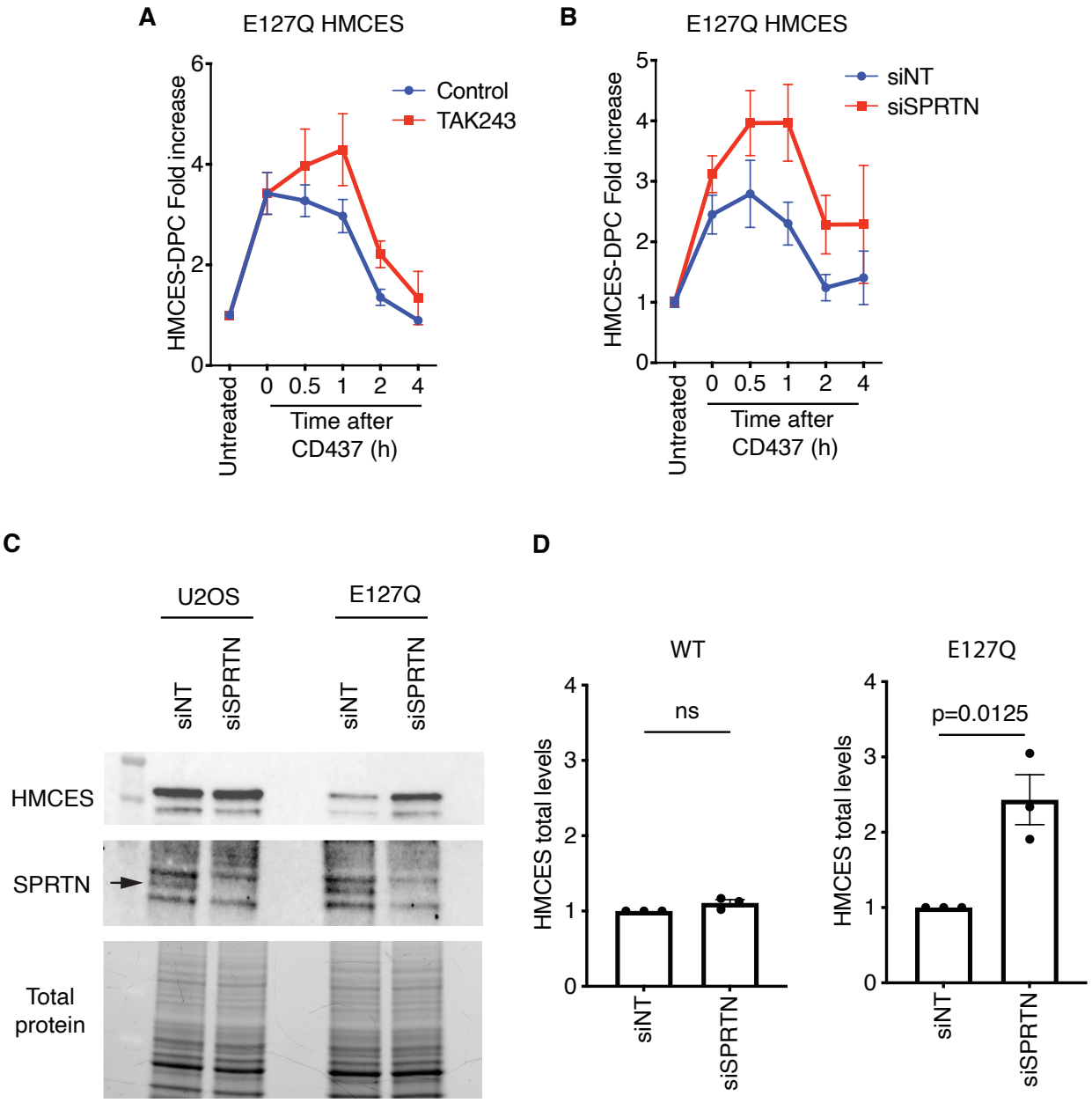

Figure 5.4 SPRTN inactivation increases HMCES E127Q levels.

(A) RADAR assay of E127Q HMCES-DPC levels in cells either that were mock treated or treated with TAK243. TAK243 was added immediately after CD437 treatment. Mean ± SEM, n=3. (B) RADAR assay of E127Q HMCES-DPC levels in cells transfected with non-targeting (siNT) or SPRTN (siSPRTN) siRNAs. Mean ± SEM, n=4. (C) Representative immunoblot of wild type and E127Q U2OS cells transfected with siNT and siSPRTN. The black arrow points out the SPRTN protein. (D) Quantification of D. on total HMCES protein levels (Mean ± SEM, n=3, two-tailed, t-test).

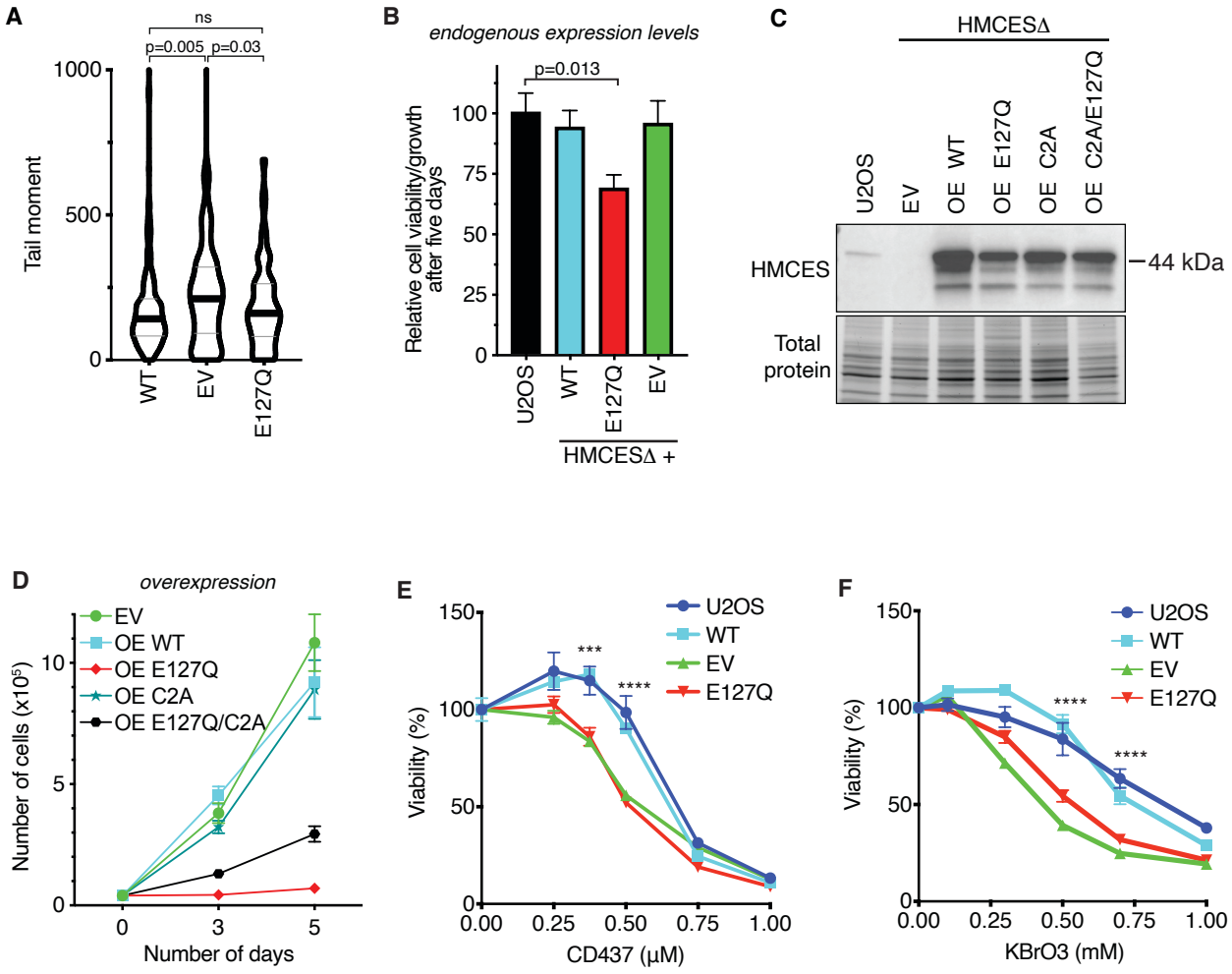

Figure 5.5 HMCES self-reversal is important for cell fitness and responses to DNA damage.

(A) Neutral comet assay was used to measure DSBs in HMCESD cells containing an empty vector (EV) or expressing near endogenous levels of WT or E127Q HMCES. P values were calculated by a one-way ANOVA with a Dunnett posttest. (B) U2OS cells and HMCESΔ cells expressing an EV or near-endogenous levels of WT or E127Q HMCES were plated at equal cell numbers. Cell proliferation/viability was measured using alamarBlue 5 days later. Mean ± SEM, n = 6, one-way ANOVA with Dunnett post-test. (C) Immunoblot of U2OS or HMCESD cells containing an empty vector (EV) or overexpressing WT, E127Q, C2A, or C2A/E127Q HMCES proteins. (D) HMCESΔ cells were infected with retroviruses to overexpress the indicated HMCES WT and mutant proteins. Cells were selected for 3 days with puromycin and then plated at equal cell numbers. Viable cells were counted using trypan blue staining at each time point. Mean ± SEM, n = 3. (E) Percentage of the viability of the indicated cells treated with CD437 measured using alamarBlue 4 days after a 24 h exposure to drug. Two-way ANOVA, n = 3. WT vs. E127Q, ***p < 0.0004, ****p < 0.0001. (F) Percentage of viability of the indicated cells treated with

$KBrO_3$ as measured using alamarBlue 3 days after a 48 h exposure to drug. Two-way ANOVA, n = 3. WT vs. E127Q, ****p < 0.0001.

I next infected HMCESΔ cells with a retrovirus containing HMCES WT and mutant cDNAs expressed from a strong promoter to cause overexpression (Fig 5.5C). Overexpressing E127Q HMCES from a strong promoter (OE E127Q) had a more deleterious effect leading to substantially slower cell growth (Fig 5.5D). In contrast, cells overexpressing WT HMCES (OE WT) or HMCESΔ cells infected with an EV proliferated normally (Fig 5.5D). Combining a C2A mutation with the E127Q mutation partially mitigated the growth defect of the E127Q overexpression, consistent with the idea that persistent thiazolidine linkages in the E127Q-HMCES-overexpressing cells contribute to the reduced cell fitness (Fig 5.5D).

HMCESΔ cells and cells that express crosslink-deficient HMCES mutants are hypersensitive to DNA-damaging agents that generate AP sites[47,67,68]. To test whether inactivating self-reversal also causes hypersensitivity, we exposed cells expressing near-endogenous levels of WT HMCES or the E127Q mutant to increasing doses of CD437 and let them recover in normal growth media before measuring viability. As expected, HMCESΔ cells transduced with an EV had reduced viability in response to CD437 treatment. E127Q cells also showed the same detrimental effect; meanwhile, WT HMCES recapitulated the viability of control U2OS cells (Fig 5.5E). Similarly, E127Q cells were hypersensitive to $KBrO_3$ (Fig 5.5F). These results further support the idea that HMCES self-reversal is an important mechanism for resolving the HMCES-DPC.

Discussion

Abasic sites are frequent DNA lesions that threaten genome stability, especially when they are present in ssDNA, where BER cannot be used for repair. HMCES provides an evolutionarily conserved mechanism to recognize and shield these ssDNA lesions from inappropriate processing that can generate DSBs. However, repair requires the removal of the HMCES-DPC. My results demonstrate that there is a self-reversal mechanism in human cells. The reversibility of the crosslink depends on E127, which is positioned adjacent to the thiazolidine linkage created by the N-terminal cysteine residue[58,61,62]. Mutation of E127 largely prevents the reversal reaction without preventing crosslink formation in biochemical assays and delays the resolution of the HMCES-DPC in cells. Furthermore, cells expressing only the E127Q HMCES protein proliferate

slowly, accumulate HMCES-DPCs even in the absence of added genotoxic stress, and are hypersensitive to agents that increase AP site formation.

Is E127Q only affecting self-reversal?

Some of the effects of E127Q HMCES in cells could be because of a reduction in the rate of crosslink formation or other effects of the mutation other than lack of self-reversal. Although not essential to form the DPC, the E127 residue promotes ring opening of the AP site deoxyribose from the furan to aldehyde form to facilitate crosslink formation[59].

The E127Q HMCES-DPC levels do not increase as much as the WT HMCES-DPC in response to CD437. This difference could be due to the increase in the steady-state level of the E127Q HMCES-DPC compared to WT in the untreated cells making the fold change look smaller, but it also likely reflects a reduced amount of crosslink formation. Nonetheless, the E127Q HMCES protein can at least partly protect AP sites from forming DSBs (Fig 5.5A), and E127A HMCES was also previously shown to protect AP sites during SHM[71].

Endogenous levels of E127Q expression modestly reduce cell growth, but this effect disappears over increasing passages, suggesting the cells adapt. Overexpression of E127Q from a strong promoter causes severe toxicity. The overexpression toxicity is substantially, but not completely, mitigated by combining E127Q with a C2A mutation, which blocks thiazolidine DPC formation. The lack of complete rescue suggests the E127Q mutation also has another effect other than preventing the reversal of the thiazolidine DPC linkage. Mutation of E127 or the equivalent site in YedK is reported to increase its ssDNA binding affinity compared to WT[62,113]. In addition, the C2A HMCES protein can form an unstable Schiff base with the AP site and promote DNA strand cleavage via a β-elimination reaction[58,59]. These other effects of these mutations (or a combination of them in the double mutant) may contribute to the cellular phenotypes generated by expressing these proteins.

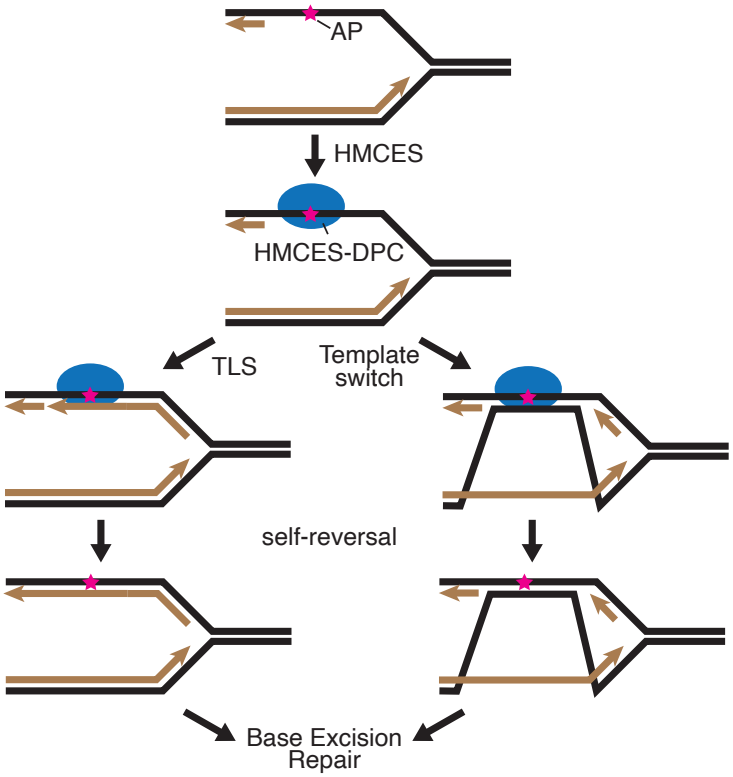

Figure 5.6 Model of CD437-induced HMCES-DPC reversal and repair

HMCES crosslinks to the ssDNA AP site at the replication fork. HMCES-DPC shields the AP site because the activity of the BER pathway at this time can induce strand breakage. Meanwhile, a DNA duplex can be formed by a different mechanism such as TLS or template switching. A dsDNA substrate will prevent re-crosslink of HMCES and promote its self-reversal, so now the exposed AP site is localized in dsDNA, where the BER pathway can promote an error-free repair.

Complementary pathways to remove E127Q HMCES-DPC

My data are consistent with recent biochemical studies showing that the thiazolidine linkage is reversible[59,113,141]. This linkage was initially thought to be highly stable because it appeared unchanged in biochemical reactions even days after formation, could be observed by crystallography, and blocked the action of AP endonucleases[58]. However, the addition of a second ssDNA-AP oligonucleotide trap revealed that HMCES could move to another substrate[59], indicating the DPC formation is reversible and regenerates an intact HMCES protein. Inhibiting HMCES-DPC self-reversal delayed, but did not prevent DPC removal. Previous studies on DPCs in *Xenopus* egg extracts found two major repair mechanisms involving the SPRTN protease and the proteasome[80]. Inhibiting the ubiquitin-dependent or SPRTN-dependent pathways yielded a modestly higher level of the E127Q HMCES-DPC after CD437 treatment (Fig 5.4B), suggesting they may act in parallel or as backup pathways to the self-resolution mechanism. Nevertheless, SPRTN is not active in removing DPCs present in dsDNA because its ZBD domain engages with unpaired bases[103]. Alternatively, E127Q HMCES-DPC may generate a bubble structure since the protein is binding only one DNA strand through its basic cleft[59] and the strands might anneal on both sides of the DPC. Then, SPRTN can remove DPCs in bubble structures[103]. Additionally, FANCJ can also help to unfold the DPC[111] and provide a more suitable target for SPRTN to remove the persistent crosslink. Further studies are required to be sure how the E127Q HMCES-DPC is finally removed. An increase in the amount of the HMCES-DPC at early time points in these circumstances could also be due to unknown indirect effects on DPC formation. SPRTN removes DPCs at replication forks[85,86], showing full activity when these bulky lesions are located at ssDNA/dsDNA junctions[102,103]. Furthermore, SPRTN can remove HMCES-DPCs formed as an intermediate in ICL repair in *Xenopus* egg extracts[110].

dsDNA formation to promote crosslink reversal

The DNA-binding cleft in HMCES and YedK only accommodates ssDNA on one side of the AP site[58,61,62,142]. This creates specificity for either AP sites in ssDNA or the ones that exist at a dsDNA/ssDNA junction such as what would form if a POL stalls at the lesion. The apparent rate of self-reversal is greatly increased when a complementary oligonucleotide is added to the HMCES-DPC to generate duplex DNA. This increased reversal rate is likely due to the inability of HMCES to rebind the duplex DNA that forms as HMCES releases from the ssDNA. Coupling DPC resolution to the generation of duplex DNA in cells would allow HMCES to shield the AP site until it can be properly repaired by BER (Figure 5.6). Duplex DNA formation in cells could be generated

by TLS synthesis across from the HMCES-DPC. Recent studies showed that TLS across from the DPC could be facilitated either by HMCES proteolysis[109,112] or by the action of FANCJ on the intact DPC[111]. In either case, the outcome of TLS would include mutations. However, HMCES-deficient cells have increased mutation rates and increased recruitment of TLS POLs to replication forks and exhibit synthetic lethality with TLS POL inactivation, suggesting that another mechanism might operate normally[47,64,67]. Alternatively, template switching could be utilized to generate the duplex DNA, providing an error-free repair mechanism (Figure 5.6). Further studies will be needed to determine if duplex DNA formation in cells is important and which of these mechanisms is preferred.

CHAPTER VI

FUTURE DIRECTIONS AND CONCLUSIONS

Summary of dissertation work

Apurinic/apyrimidinic (AP) sites are one of the most frequent DNA lesions in mammalian cells. HMCES is an evolutionarily conserved protein found at replication forks that covalently crosslinks to AP sites in ssDNA, forming a DNA-protein crosslink (DPC). The HMCES-DPC shields ssDNA AP sites from endonucleases and TLS polymerase activity to help maintain genome integrity. However, how the HMCES-DPC is resolved in human cells and produces better outcomes for genome stability and organism fitness remains unknown. My thesis project examines the possible mechanisms involved in HMCES-DPC removal. In Chapter III, I outline and test a system utilizing DNA polymerase alpha inhibitor CD437 to induce the accumulation of ssDNA and AP sites during DNA synthesis. Hence, I used CD437 to track and quantify the formation and resolution of HMCES-DPCs in human cells. In Chapter IV, I used the CD437-induced HMCES-DPC removal system to test the requirement of the proteasome and SPRTN in this process. In Chapter V, I used biochemical and cellular experiments to analyze the self-removal pathway. I identified HMCES's residue Glu127 as a major mediator of the self-reversal mechanism. Additionally, I described the detrimental repercussions of preventing crosslink reversal in cells. Overall, my thesis project has provided a new system to study HMCES-DPC removal, proving that HMCES crosslink reversal is important to maintain cellular fitness.

Limitations of the study

The first limitations in this project were related to the lack of agents that could induce detectable levels of HMCES-DPC whose resolution could be also detected over time. CD437 is a useful tool to analyze HMCES-DPC removal in cells. Nevertheless, it is still an unusual situation in the cell to completely stop Polα activity for a specific amount of time. Moreover, cell viability starts to reduce 24 hrs after release from the drug. As an alternative approach, I tried to use the inducible-ABOBEC cell line developed in the lab to generate AP sites in cells specifically through cytosine deamination[67], but the increase of HMCES-DPC levels could not be detected by the RADAR assay.

HMCES-DPC resolution was observed in HCT116 and U2OS cells, but crosslink resolution seems to be a bit faster in the latter one. Possible differences throughout different cell lines should be considered since all the HMCES mutants have been developed from an HMCES null background in U2OS cells. Additionally, HMCES DUB cell line's ability to prevent HMCES ubiquitylation has not been directly proven yet. Thus, it is difficult to make strong conclusions from those experiments. Neither was able to prove direct ubiquitylation of HMCES-DPC using the Ni-NTA kit previously used in the lab[47]. Ubiquitylation of HMCES-DPC has been addressed indirectly and more experiments are required to know if ubiquitylation of HMCES has any role prior or after crosslink reversal.

Finally, HMCES-DPC resolution mechanisms may depend on the context or cell type. The E127Q HMCES mutation reduces DPC formation efficiency and may affect DNA binding or other HMCES functions in addition to preventing self-reversal. These changes could contribute to the phenotypes observed in cells expressing this mutant protein. Finally, further studies will be needed to determine if duplex DNA formation in cells is important to the resolution mechanism and how the duplex DNA is generated.

Future directions

What happens with HMCES after reversal?

My model suggests that HMCES crosslink is reversed once duplex DNA is formed, which localizes an exposed AP site in dsDNA, available for an error-free repair by BER. Consequently, HMCES is released as an intact soluble protein able to crosslink again to a new available ssDNA AP site, as I could detect from the self-reversal biochemical assays. Then, if soluble HMCES is recycled to protect future AP sites, its nuclear levels should never change. Nevertheless, some preliminary data suggest something different. Analysis of nuclear HMCES by IF, 2h after CD437 treatment, shows a reduction in total HMCES levels. However, those levels were completely rescued with MG132 or TAK243 treatment (Fig 6.1A), suggesting a proteasome-dependent degradation. HMCES nuclear-integrated intensity by IF cannot differentiate the signal that comes from the soluble protein and the one from HMCES-DPC. However, according to the RADAR assay, HMCES-DPC levels are not rescued by MG132 or TAK243 when analyzed 2hr after CD437 treatment (Fig 4.1A and B). Thus, the proteasome-dependent degradation seems to be restricted to the soluble HMCES. In an attempt to keep testing this hypothesis, I performed chromatin fractionation using the HMCES-DUB and HMCES-DUBi cells. I asked whether the soluble fraction of HMCES was reduced after CD437 treatment and if that could be prevented by inhibiting ubiquitylation. Indeed, preliminary results show that soluble HMCES levels are reduced 1 h after CD437 treatment and they are mostly rescued by TAK243. HMCES-DUB levels instead, do not decrease as much, and incubation with TAK243 has no effect. Finally, the inactivated deubiquitinate HEMCES-DUBi almost completely recapitulates WT HMCES (Fig 6.1B). The fraction missing in all samples even with TAK243 might be present at the chromatin fraction. Still acknowledging that the functionality of these HMCES fused cell lines has not been verified yet, these preliminary data suggest that degradation of soluble HMCES requires ubiquitylation. RADAR data showed no difference in HMCES-DPC removal within these cell lines (Fig 4.2C), supporting the hypothesis that ubiquitylation does not regulate HMCES degradation until is removed from the DNA.

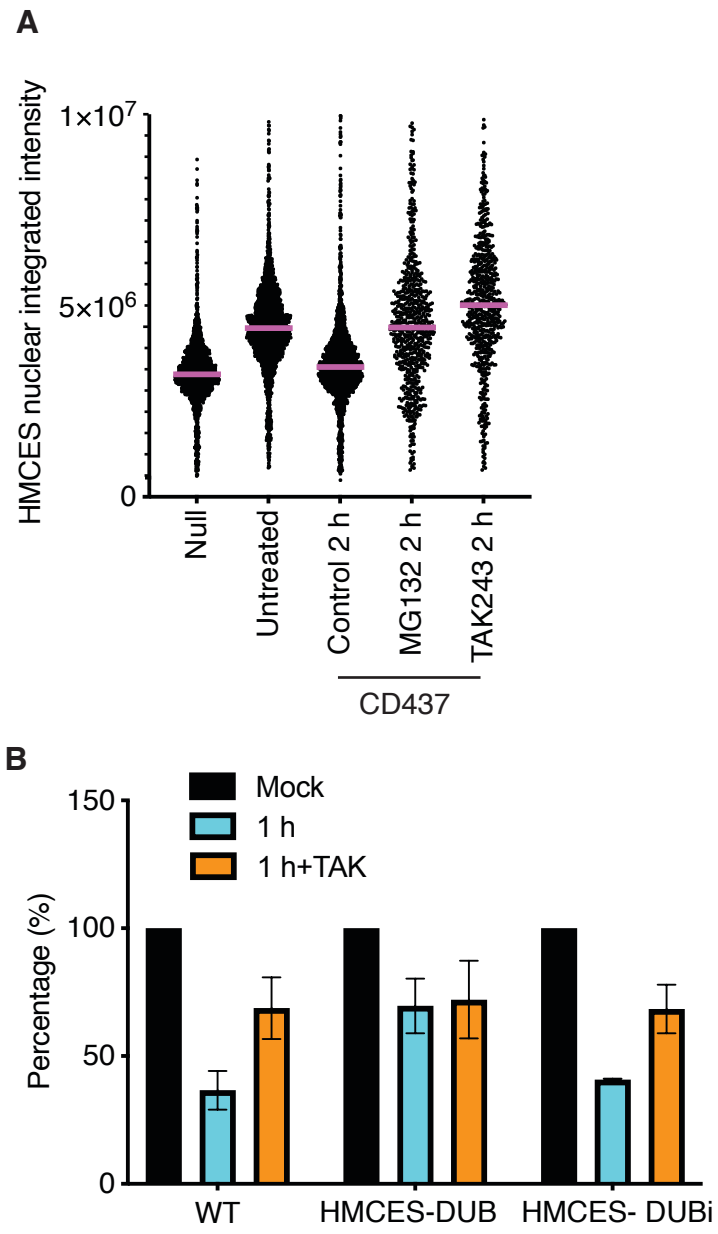

Figure 6.1 Total and soluble HMCES levels after CD437 and their regulation by ubiquitylation.

(A) Quantification of total HMCES nuclear signal in EdU positive cells (n=2) (B) Quantification of the soluble fraction of HMCES in the described cell lines, 1 h after CD437 treatment (5μM for 30 min) in the presence or not of TAK243 (n=2).

My current model to explain the fate of HMCES after reversal proposes that HMCES crosslinks to the AP site and then is ubiquitylated by an unknown E3 ligase that possibly depends on neddylation. Recovering 1 hr after CD437 gives enough time for DNA duplex formation which promotes self-reversal of HMCES-DPC, already in a ubiquitylated state that allows its degradation by the proteasome (Fig 6.2A). Proteasome (Fig 6.2B) and ubiquitin (Fig 6.2C) inhibition accumulate the soluble reversed HMCES but do not prevent HMCES-DPC removal at this point.

Further investigation is required to test this model. A more depicted time course of total and insoluble HMCES levels would help to understand what happens with HMCES after crosslink reversal. Testing levels of the C2A mutant after CD437 treatment should answer if crosslinking is previously required for HMCES degradation. Likewise, I would use E127Q HMCES cells to test whether or not reversal is also required, doing the analysis 1 h after CD437 treatment. Other interesting questions encompass: Which is the E3 ligase mediating HMCES ubiquitylation? Why would be important to degrade it when is no longer blocking the DNA? Additionally, knowing more about the regulation of soluble HMCES levels may help to explain the small increase in HMCES-DPC levels I detected by RADAR 30 min after CD437 while inhibiting the proteasome (Fig 4.1A and D).

Exploring new functions of HMCES

HMCES might have other functions besides protecting AP sites. The Alt-EJ refers to the repair events that are independent of the canonical NHEJ (c-NHEJ) because the DSB requires extensive end processing before ligation[143]. Hence, the Alt-EJ pathway often uses short stretches of homology (microhomology) to bridge the break during repair[143]. Shukla and collaborators reported that HMCES depletion reduces the CSR events in B cells by affecting the Alt-EJ and not the c-NHEJ pathway[74]. Strikingly, the Cys catalytic activity is not as required for this process as its DNA binding capacity[74], suggesting a function besides AP site formation. They claim that HMCES is part of the Alt-EJ pathway by binding and protecting the microhomology regions. This hypothesis is supported by the co-crystal structure of two HMCES SRAP domains in complex with palindromic DNA flanked by 3 and 4 nucleotide overhangs that are covered by each SRAP domain[144] (PDB: 6OOV[61]). A similar structure was resolved with two YedK molecules in a complex with two ssDNA (PDB:6KBS[62]). However, microhomology is not essential for Alt-EJ and[145,146]. Furthermore, c-NHEJ can use microhomology during repair, but only 1-2 base pairs of microhomology between the two DNA termini[147,148]. Moreover, A CRISPR screen against DNA-

damage agents also presented HMCES in a cluster with genes involved in the NHEJ pathway[65]. This contradicts the hypothesis of HMCES being part of the Alt-EJ pathway and opens a new possibility: that HMCES may be recruited to help decide the repair fate of a DSB, according to the levels of DNA ends resection. Many steps in this mechanism are still unknown. Although, this could be a complementary activity of HMCES in B cells, besides its role in SHM where crosslink formation is required[71]. HMCES protein expression is high in cytotoxic T-lymphocytes, B-lymphocytes, and lymph nodes (genecards.org as HMCES and proteomicsdb.org as C3orf37). Somatic hypermutation in B cells occurs when they reach the germinal center of secondary lymphoid tissues, such as lymph nodes[149]. More functional analysis in these tissues would expand our knowledge on this protein: Is HMCES activity in the Alt-EJ pathway only required during B cell maturation and cell type-specific? Does it require HMCES to go under any posttranslational modification? If crosslink to AP sites is not required, and DNA binding is not sequence dependent, how does it recognize the DNA overhangs? Using B lymphocyte cell lines such as CH12F3 allows easier genetic and biochemical manipulation than primary B cells[150]. Additionally, identifying new HMCES interactors would provide more functional information.

Another function recently explored about HMCES is its activity regulating the transcriptome of the TGF-β signaling in mESCs. HMCES and R-SMAD proteins co-occupy at transcriptionally active chromatin to regulate gene expression. Thus, in the absence of HMCES, those genes are upregulated and causing embryonic defects [54]. Since HMCES Chip-seq revealed enrichment of SMAD2 and SMAD4 motifs, it would be interesting to know how HMCES reaches those specific DNA sequences. Is recruited due to specific epigenetic marks? Is there a protein recruiting HMCES to active chromatin sites? HMCES was previously reported to impact methylation distribution at a specific embryonic state[48]. Presumably, this goes again to functions related to specific cell tissues and at specific times in development. Moreover, during particular periods of development, the embryo is more or less susceptible to oxidative stress[151]. Thus, it would be interesting to know if the HMCES crosslink capacity is relevant during embryogenesis.

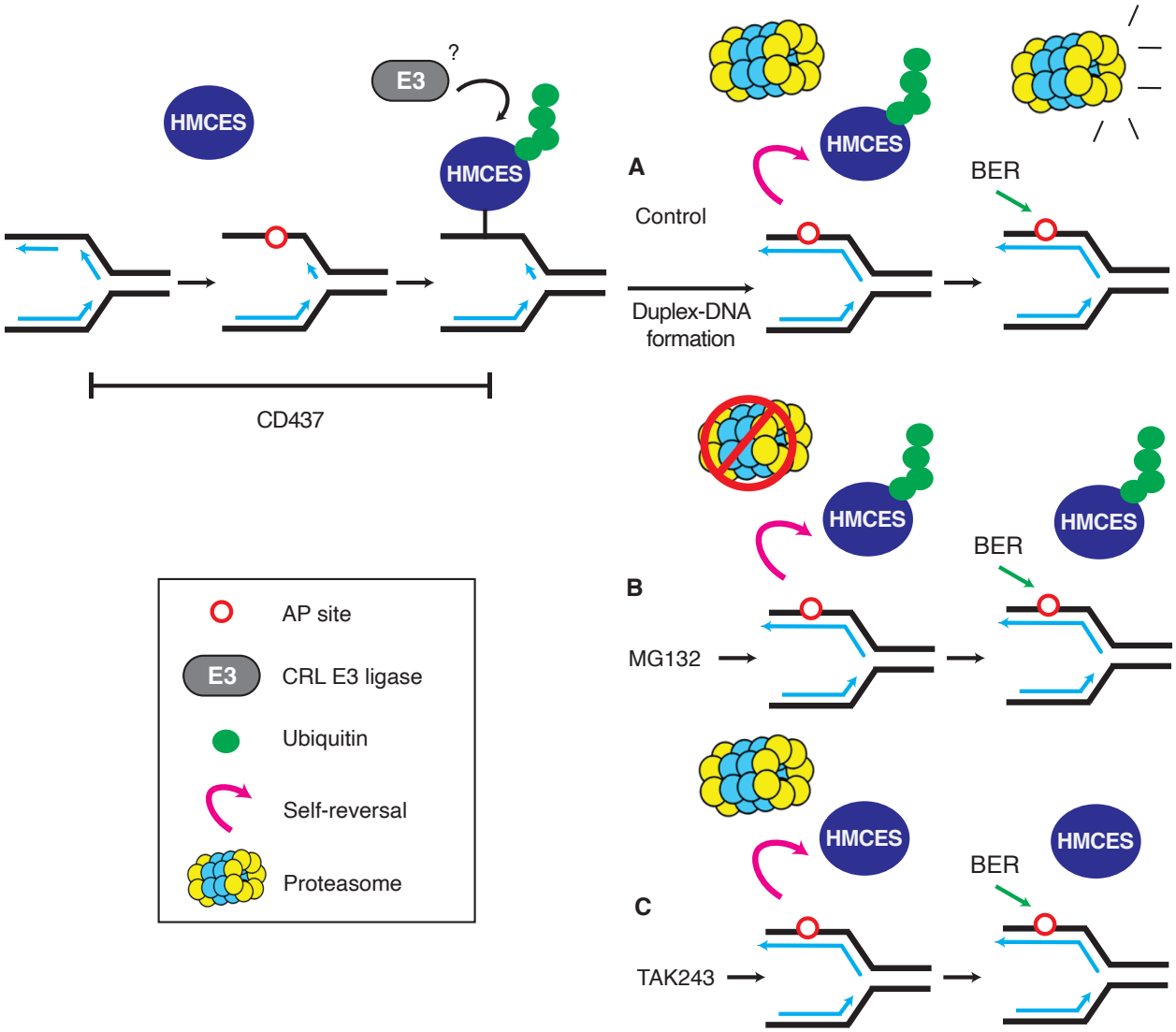

Figure 6.2 Possible proteasome-dependent degradation of HMCES after reversal.

Diagram of HMCES-DPC formation after CD437 treatment and ubiquitylation by a CRL E3 ligase (activated by neddylation) (A) Recovery in normal growth media promotes HMCES-DPC self-reversal and soluble HMCES-DPC will be degraded by the proteasome. (B) MG132 incubation inhibits the proteasome and accumulates reversed soluble HMCES (C) TAK243 incubation inactivates ubiquitylation without affecting crosslink reversal and accumulating soluble HMCES.

HMCES as a therapeutic target for cancer cells.

There is an increase ROS production in cancer cells that activates pro-tumorigenic signaling, enhances cell survival, and drives DNA damage and genetic instability[152]. Nevertheless, tumor cells develop a mechanism to adjust to high ROS by expressing elevated levels of antioxidant proteins to detoxify by maintaining pro-tumorigenic signaling and preventing apoptosis[152]. HMCES protect leukemia cells from oxidative DNA damage, and high expression levels of HMCES in ALL cells are associated with poor prognosis[72]. Moreover, the depletion of HMCES in human leukemia cell lines resulted in decreased proliferation, increased apoptosis, and elevated strand breaks[72]. Thus, inhibiting HMCES as a therapeutic strategy for ALL is another possible future direction. HMCES is also essential in lungs and head and cancer cell lines due to the higher APOBEC expression[153,154]. The TCGA portal does not show yet any somatic mutations in HMCES related to a specific type of cancer. Nevertheless, developing small molecules to inhibit its crosslink activity or its ssDNA binding region would be a new approach with possible clinical outcomes.

New interactions of HMCES

HMCES interact with PCNA through its C-terminal region[47]. HMCES can be ubiquitylated by RFWD3, suggesting a transient interaction, but the exact mechanism and the modified residues are unknown. HMCES is also a target for SUMOylation[88,96] but the SUMO E3 ligase is still elusive. An unbiased approach to finding HMCES interactors during CD437 treatment or another AP site inducer would help to discover the proteins involved in HMCES crosslink and removal, if any. Thus, I attempted to use the BioID2 system, where a promiscuous biotin ligase detects protein-protein associations as well as proximate proteins in living cells[155]. I cloned BioID2 to the C-terminal region of HMCES and generated a stable cell line expressing the fused protein in an HMCES null background. Protein expression was verified by immunoblot; Unfortunately, HMCES protein was mostly expressed compared to its fused version (Fig 6.3). This meant that probably the BioID2 enzyme was cleaved from the protein or maybe the fused protein was negatively selected during cell passages. No biotinylation was detected in these cells (data not shown), suggesting that expression levels of the fused protein were not enough. Probably fusing the enzyme at the C-terminus is affecting some functions in HMCES and needs to be added in a different region. Still, it would be important to explore HMCES interactions, also in the protein soluble state since HMCES might have functions besides its crosslinking activity[54,144].

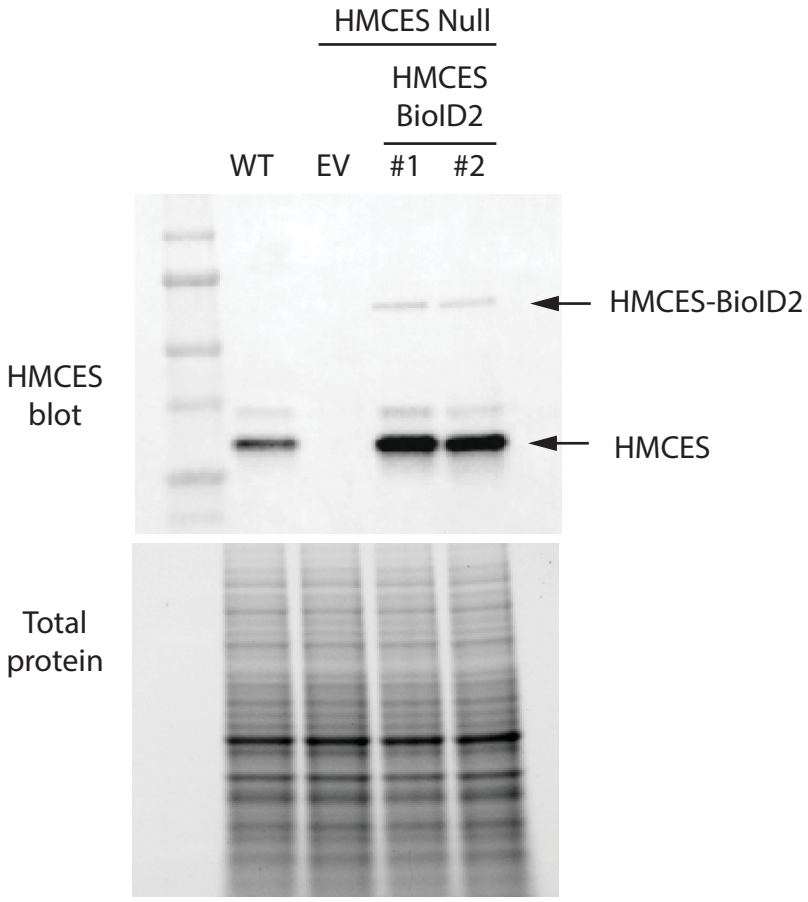

Figure 6.3 Expression of HMCES-BioID2 in U2OS cells.

On top, immunoblot against HMCES of cells expressing WT HMCES, empty vector (EV), and clones #1 and #2 of cells expressing the fused protein HMCES-BioID2. The arrows point to the expected band for the fused protein and only WT HMCES. On the bottom, loading control.

Johannes Walter's lab has recently released Predictomes.org, a deep learning system AlphaFold-Multimer to systematically screen for protein-protein interactions (PPI). They developed one that focuses on genome maintenance, and they have a list of hundreds of proteins that have gone through thousands of complex predictions. Following the settings for a strong interaction prediction (average model:>50%, plddt:>70,pdockq:>0.25, pae:<10), the only outcome is PCNA, which interaction has been tested by immunoprecipitation[47]. Relaxing the setting can provide more candidates, but also decrease confidence, which makes it challenging for transient interaction.

What is the function of the C-terminal region?

HMCES contains a C terminal region that contains a PIP box for PCNA interaction. Alpha fold predicts it as a highly disorder region and 19 residues have been predicted to be targeted for phosphorylation, acetylation, ubiquitylation, or other modifications (Phosphosite.org). I attempted to study the function of the HMCES C-term by generating stable cell lines that express the truncated form of the protein without the C-terminal region (ΔC). I also complemented HMCES null cells with WT HMCES (WT) and an empty vector (EV) (Fig 6.4A). To separate any function present at the C-terminus from PCNA interaction, I used the PIP box mutant (W337A/L3338A) elaborated by previous postdoc Kareem Mohni[47]. First, I asked whether C-term was important for cell viability under oxidative stress conditions ($KBrO_3$). As expected, EV cells are hypersensitive to $KBrO_3$. The viability is rescued by WT but not by ΔC. The PIP mutant has an intermediate effect, suggesting that PCNA interaction is important for HMCES function against oxidative stress and that something else at the C-term region, besides PCNA interaction, is sensitizing the cell (Fig 6.4B). To verify the detrimental effect of lacking the C-term, I generated stable cell lines overexpressing WT HMCES (OE WT) and the truncated protein (OE ΔC) (Fig 6.4C). Strikingly, over-expression of the truncated protein rescued the deficient phenotype instead of worsening it (Fig 6.4D). Western blot quantification showed that hypersensitivity of ΔC is not due to low levels of the protein. Presumably, there was something else at the C-term region that was overcome at high levels of the protein. Utilizing motif scan software (Myhits.sib.swiss), I got a prediction for a nuclear localization sequence (NLS) at the C-terminus, although with a low score match. Thus, using a high-throughput microscopy system (ImageXpress), I analyzed HMCES nuclear localization without the C-terminal region. Indeed, preliminary data showed me that ΔC cells have less nuclear and more cytoplasmic localization compared to WT cells (Fig 6.5A and B). Meanwhile, two different clones (#5 and #9) of the OE ΔC cells showed a bit lower nuclear

localization compared to the OE WT cells, but significantly higher than the cells expressing HMCES at endogenous levels (Fig 6.5C). These results may indicate that the hypersensitivity against $KBrO_3$ of ΔC was probably rescued by forcing the entrance of the protein to the nucleus. However, more replicates of the experiment are needed.

Finally, I was curious to know if lacking the C-terminal region of HMCES would affect its ability to crosslink to AP sites due to the absence of PCNA interaction. I performed the RADAR assay using the CD437 system and observed that the absence of the C-term did not affect HMCES crosslink or resolution (Fig 6.5D). This result suggests that HMCES might target AP sites besides the ones that collide with PCNA, probably due to the DNA substrate provided by CD437 treatment. HMCES SRAP domain is enough to promote its removal, supporting the self-reversal process. Also, predicted PTMs at the C-term might not be crucial for HMCES-DPC removal, but they might have another function when the protein is soluble.

Future directions on the C-term examination would be to verify the NLS motif, incorporating it in the SRAP domain of the ΔC cells, and test if that rescues the hypersensitivity against $KBrO_3$.

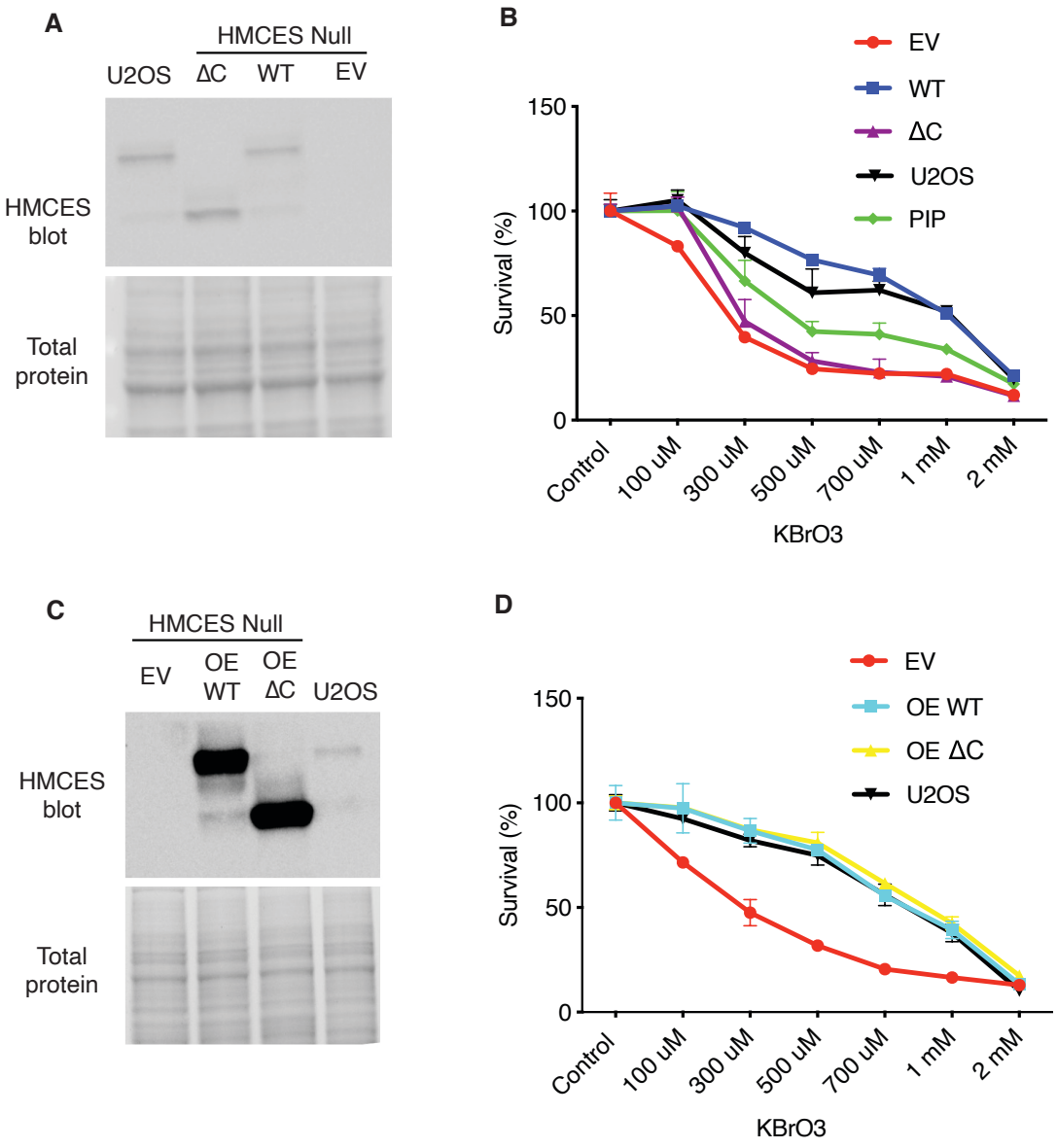

Figure 6.4 HMCES C-terminal region

(A) Immunoblot using HMCES antibody of the cell lines described. (B) AlamarBlue assay to test the viability of the different cell lines in A, against KBrO3. (C) Immunoblot against HMCES for the over-expression cell lines described (D) AlamarBlue assay to test viability of the cell lines in C, against KBrO3.

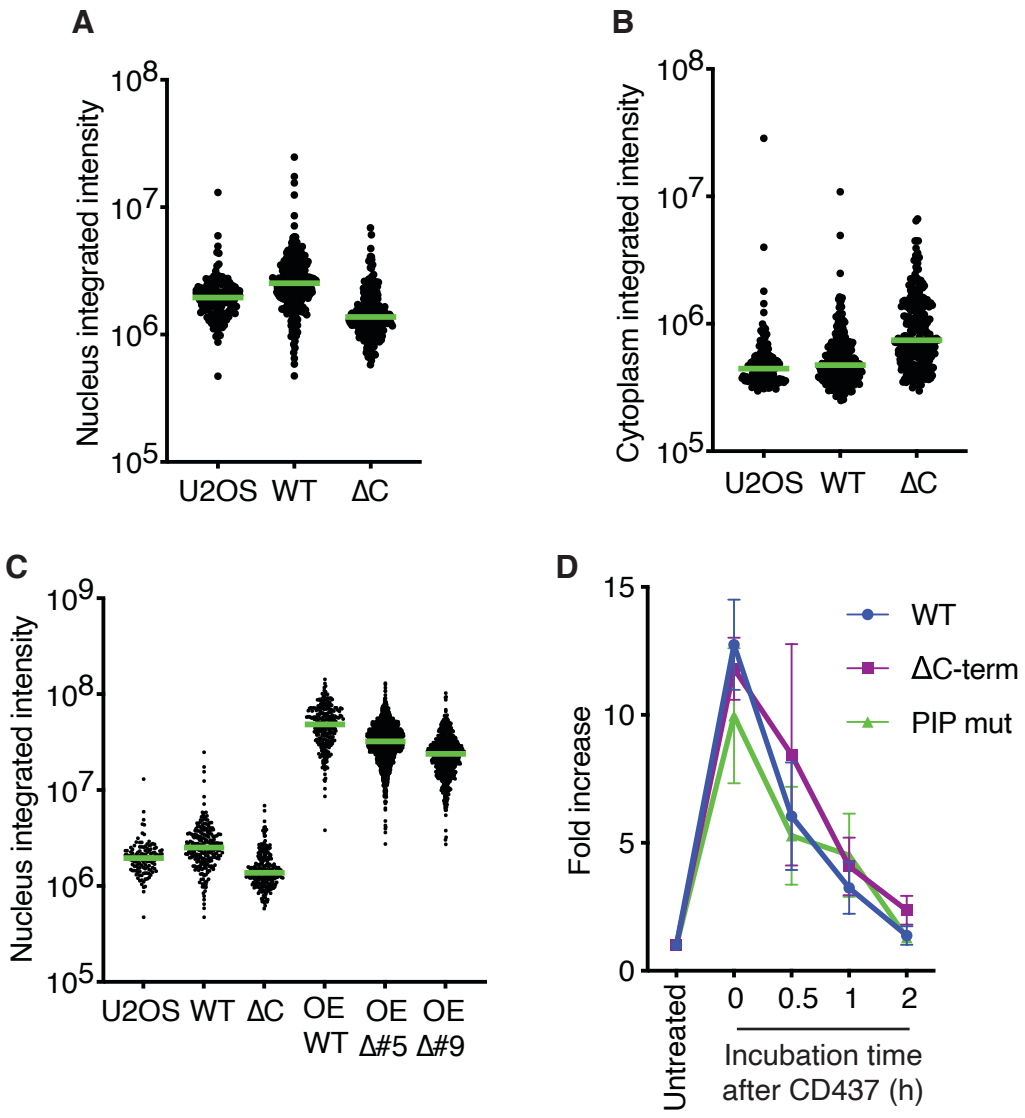

Figure 6.5 Analyzing the role of the C-terminal region on HMCES localization and HMCES-DPC formation.

(A) Nuclear intensity of HMCES in cell lines described. (B) Cytoplasmic intensity of HMCES in the cell lines described (C) Nuclear intensity of HMCES in endogenous-levels expressing cells (U2OS, WT, and ΔC) and overexpression-levels expressing cells (OE WT, OE Δ#5, OE Δ#9) (D) Quantification of RADAR assay describing HMCES-DPC levels on the cell lines described.

CONCLUSIONS

HMCES is an evolutionarily conserved protein present at replication forks that shields AP sites located in ssDNA to prevent further damage due to the intrinsic reactivity of AP sites or from other damage-tolerance pathways that are error-prone. HMCES crosslinks to AP sites through the active site of the SRAP domain, which contains three highly conserved residues that generate a thiazolidine linkage. This HMCES-DPC stays in equilibrium due to its capacity to bind to ssDNA and the activity of the residues: Cys2, His210, and Glu127, to react with the AP site.

Since HMCES-DPC represents a bulky lesion for DNA processes such as replication and transcription, it needs to be removed. The proteolysis mechanism mediated by the proteasome or SPRTN is not required to remove HMCES-DPC in human cells. However, HMCES can remove itself by reversing its crosslink, which is mediated by Glu127. HMCES mutant E127Q delays HMCES-DPC removal *In vitro* and in cells. Additionally, it has a detrimental effect on cell viability in the absence or presence of reagents that form AP sites. Thus, not only HMCES-DPC formation is important for cell fitness, but also its removal. Furthermore, biochemical evidence shows that the formation of a duplex DNA in the HMCES-DPC prevents HMCES re-crosslink to the AP site, which accelerates the reversal detection with similar kinetics as HMCES-DPC removal in cells. Taken together, I propose a new mechanism to process AP sites during replication, where its presence in ssDNA, allows a transient HMCES-DPC formation, which shields the AP site from TLS pro-mutagenic activity and endonuclease activity from the BER pathway. The transient crosslink state is favored until replication restarts, duplex DNA is formed and the crosslink is completely reversed. This would localize the AP site again in the context of dsDNA, where now, an error-free repair mechanism such as BER can take place (Fig 6.6).

Finally, CD437 comes as a new and practical strategy to study AP sites in cells. My project provides a new model of how HMCES is conserved in all domains of life, not to repair AP sites, but to protect them while replication continues, and release from them once an appropriate repair mechanism can take place to maintain genome integrity.

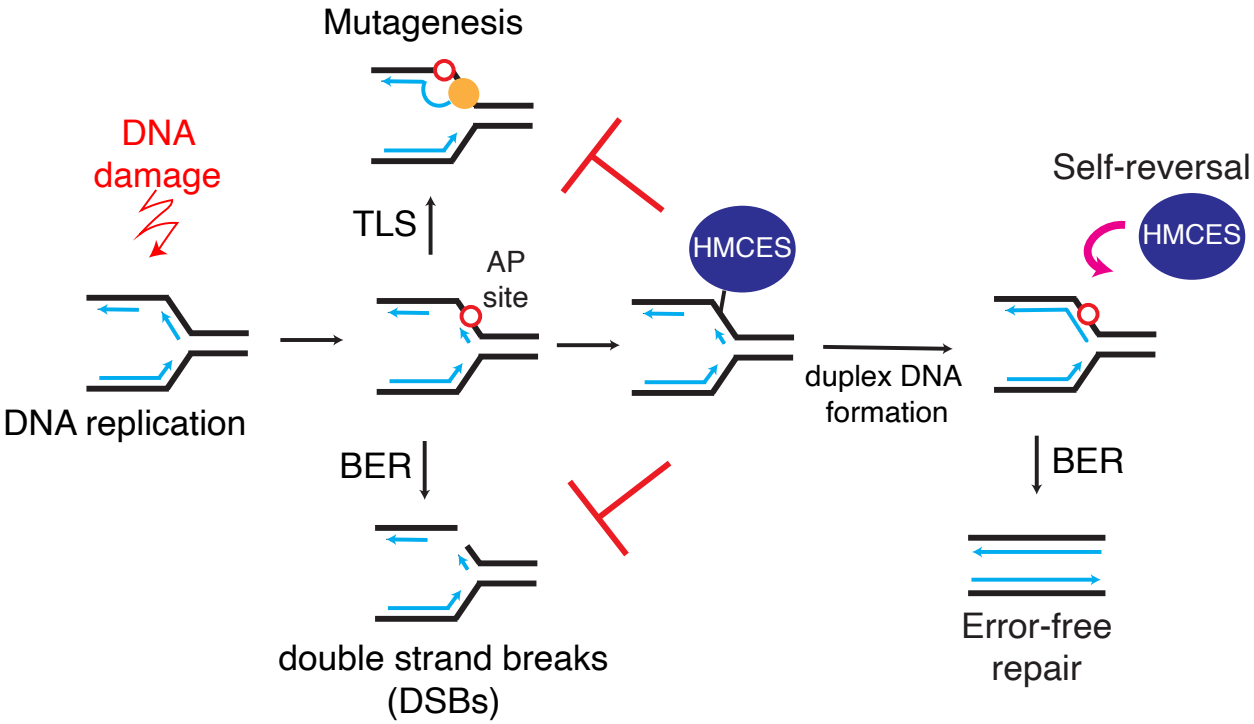

Figure 6.6 Model explaining HMCES-DPC transient formation to provide an error-free repair mechanism for ssDNA AP sites,

REFERENCES

1. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T. Correcting Altered Bases in DNA: DNA Repair. In: *DNA Repair and Mutagenesis*. 2nd ed. Washington, D.C: ASM Press; 2006.
2. Thompson PS, Cortez D. New insights into abasic site repair and tolerance. *DNA Repair (Amst)*. 2020;90(April):102866. doi:10.1016/j.dnarep.2020.102866
3. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. *Biochem*. 1972;11(19):3610-3618. doi:10.1021/bi00769a018
4. Chan K, Sterling JF, Roberts SA, Bhagwat AS, Resnick MA, Gordenin DA. Base Damage within Single-Strand DNA Underlies In Vivo Hypermutability Induced by a Ubiquitous Environmental Agent. *PLoS Genet*. 2012;8(12). doi:10.1371/journal.pgen.1003149
5. Peng Y, Pei H. DNA alkylation lesion repair: outcomes and implications in cancer chemotherapy. *J Zhejiang Univ Sci B*. 2021;22(1):47-62. doi:10.1631/jzus.B2000344
6. Barrows LR, Magee PN. Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro. *Carcinogenesis*. 1982;3(3):349-351. doi:10.1093/carcin/3.3.349
7. Loeb LA, Preston BD. Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet*. 1986;20:201-230. doi:10.1146/annurev.ge.20.120186.001221
8. McNeill DR, Lam W, DeWeese TL, Cheng Y-C, Wilson III DM. Impairment of APE1 Function Enhances Cellular Sensitivity to Clinically Relevant Alkylators and Antimetabolites. *Mol Cancer Res*. 2009;7(6):897-906. doi:10.1158/1541-7786.MCR-08-0519
9. Faucher F, Doublié S, Jia Z. 8-oxoguanine DNA glycosylases: one lesion, three subfamilies. *Int J Mol Sci*. 2012;13(6):6711-6729. doi:10.3390/ijms13066711
10. Ganguly T, J N. D. Stability of DNA thymine hydrates. *Nucleic Acids Res*. 1991;19(12):3319-3323. doi:10.1093/nar/19.12.3319
11. Jacobs AL, Schär P. DNA glycosylases: in DNA repair and beyond. *Chromosoma*. 2012;121(1):1-20. doi:10.1007/s00412-011-0347-4
12. Hahm JY, Park J, Jang E-S, Chi SW. 8-Oxoguanine: from oxidative damage to epigenetic and epitranscriptional modification. *Exp Mol Med*. 2022;54(10):1626-1642. doi:10.1038/s12276-022-00822-z
13. Liu M, Doublié S, Wallace SS. Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage. *Mutat Res*. 2013;743-744:4-11. doi:10.1016/j.mrfmmm.2012.12.003

14. Williams JS, Lujan SA, Kunkel TA. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat Rev Mol Cell Biol*. 2016;17(6):350-363. doi:10.1038/nrm.2016.37
15. Harris RS, Liddament MT. Retroviral restriction by APOBEC proteins. *Nat Rev Immunol*. 2004;4(11):868-877. doi:10.1038/nri1489
16. Seplyarskiy VB, Soldatov RA, Popadin KY, Antonarakis SE, Bazykin GA, Nikolaev SI. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res*. 2016;26(2):174-182. doi:10.1101/gr.197046.115
17. Lhomme J, Constant JF, Demeunynck M. Abasic DNA structure, reactivity, and recognition. *Biopolymers*. 1999;52(2):65-83. doi:10.1002/1097-0282(1999)52:2<65::AID-BIP1>3.0.CO;2-U
18. Talpaert-Borlé M. Formation, detection and repair of AP sites. *Mutat Res*. 1987;181(1):45-56. doi:10.1016/0027-5107(87)90286-7
19. Greenberg MM. Abasic and oxidized abasic site reactivity in DNA: enzyme inhibition, cross-linking, and nucleosome catalyzed reactions. *Acc Chem Res*. 2014;47(2):646-655. doi:10.1021/ar400229d
20. Huskova A, Landova B, Boura E, Silhan J. The rate of formation and stability of abasic site interstrand crosslinks in the DNA duplex. *DNA Repair (Amst)*. 2022;113:103300. doi:10.1016/j.dnarep.2022.103300
21. Imani Nejad M, Housh K, Rodriguez AA, et al. Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3. *DNA Repair (Amst)*. 2020;86:102752. doi:10.1016/j.dnarep.2019.102752
22. Nakamura J, Nakamura M. DNA-protein crosslink formation by endogenous aldehydes and AP sites. *DNA Repair (Amst)*. 2020;88:102806. doi:10.1016/j.dnarep.2020.102806
23. Prasad R, Horton JK, Chastain PD 2nd, et al. Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. *Nucleic Acids Res*. 2014;42(10):6337-6351. doi:10.1093/nar/gku288
24. Quiñones JL, Thapar U, Yu K, Fang Q, Sobol RW, Demple B. Enzyme mechanism-based, oxidative DNA-protein cross-links formed with DNA polymerase β in vivo. *Proc Natl Acad Sci U S A*. 2015;112(28):8602-8607. doi:10.1073/pnas.1501101112
25. Roberts SA, Strande N, Burkhalter MD, et al. Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature*. 2010;464(7292):1214-1217. doi:10.1038/nature08926
26. Yu S-L, Lee S-K, Johnson RE, Prakash L, Prakash S. The stalling of transcription at

abasic sites is highly mutagenic. *Mol Cell Biol*. 2003;23(1):382-388. doi:10.1128/MCB.23.1.382-388.2003

27. Choi J-Y, Lim S, Kim E-J, Jo A, Guengerich FP. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and REV1. *J Mol Biol*. 2010;404(1):34-44. doi:10.1016/j.jmb.2010.09.015
28. Agapov A, Olina A, Kulbachinskiy A. RNA polymerase pausing, stalling and bypass during transcription of damaged DNA: from molecular basis to functional consequences. *Nucleic Acids Res*. 2022;50(6):3018-3041. doi:10.1093/nar/gkac174
29. Choi JY, Lim S, Kim EJ, Jo A, Guengerich FP. Translesion Synthesis across Abasic Lesions by Human B-Family and Y-Family DNA Polymerases α, δ, η, ι, κ, and REV1. *J Mol Biol*. 2010;404(1):34-44. doi:10.1016/j.jmb.2010.09.015
30. Weerasooriya S, Jasti VP, Basu AK. Replicative bypass of abasic site in Escherichia coli and human cells: similarities and differences. *PLoS One*. 2014;9(9):e107915. doi:10.1371/journal.pone.0107915
31. Goodman MF, Woodgate R. Translesion DNA polymerases. *Cold Spring Harb Perspect Biol*. 2013;5(10):a010363. doi:10.1101/cshperspect.a010363
32. Yang W, Gao Y. Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annu Rev Biochem*. 2018;87:239-261. doi:10.1146/annurev-biochem-062917-012405
33. Anand J, Chiou L, Sciandra C, et al. Roles of trans-lesion synthesis (TLS) DNA polymerases in tumorigenesis and cancer therapy. *NAR cancer*. 2023;5(1):zcad005. doi:10.1093/narcan/zcad005
34. Tonzi P, Huang TT. Role of Y-family translesion DNA polymerases in replication stress: Implications for new cancer therapeutic targets. *DNA Repair (Amst)*. 2019;78:20-26. doi:10.1016/j.dnarep.2019.03.016
35. Carter RJ, Parsons JL. Base Excision Repair, a Pathway Regulated by Posttranslational Modifications. *Mol Cell Biol*. 2016;36(10):1426-1437. doi:10.1128/mcb.00030-16
36. Sarker AH, Cooper PK, Hazra TK. DNA glycosylase NEIL2 functions in multiple cellular processes. *Prog Biophys Mol Biol*. 2021;164:72-80. doi:10.1016/j.pbiomolbio.2021.03.003
37. Krokan HE, Bjoras M. Base excision repair. *Cold Spring Harb Perspect Biol*. 2013:23-64. doi:10.1142/9789812706782_0002
38. Petruseva IO, Evdokimov AN, Lavrik OI. Molecular mechanism of global genome nucleotide excision repair. *Acta Naturae*. 2014;6(1):23-34.

39. Tornaletti S, Maeda LS, Hanawalt PC. Transcription arrest at an abasic site in the transcribed strand of template DNA. *Chem Res Toxicol*. 2006;19(9):1215-1220. doi:10.1021/tx060103g
40. Clauson CL, Oestreich KJ, Austin JW, Doetsch PW. Abasic sites and strand breaks in DNA cause transcriptional mutagenesis in Escherichia coli. *Proc Natl Acad Sci U S A*. 2010;107(8):3657-3662. doi:10.1073/pnas.0913191107
41. Waters LS, Minesinger BK, Wiltrout ME, D'Souza S, Woodruff R V, Walker GC. Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev*. 2009;73(1):134-154. doi:10.1128/MMBR.00034-08
42. Casali P, Pal Z, Xu Z, Zan H. DNA repair in antibody somatic hypermutation. *Trends Immunol*. 2006;27(7):313-321. doi:10.1016/j.it.2006.05.001
43. Rosenbaum JC, Bonilla B, Hengel SR, et al. The Rad51 paralogs facilitate a novel DNA strand specific damage tolerance pathway. *Nat Commun*. 2019;10(1):1-11. doi:10.1038/s41467-019-11374-8
44. Cortez D. Replication-Coupled DNA Repair. *Mol Cell*. 2019;74(5):866-876. doi:10.1016/j.molcel.2019.04.027
45. Bonilla B, Brown AJ, Hengel SR, et al. The Shu complex prevents mutagenesis and cytotoxicity of single-strand specific alkylation lesions. *Elife*. 2021;10. doi:10.7554/eLife.68080
46. Aravind L, Anand S, Iyer LM. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol Direct*. 2013;8(1):1. doi:10.1186/1745-6150-8-20
47. Mohni KN, Wessel SR, Zhao R, et al. HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell*. 2019;176(1-2):144-153.e13. doi:10.1016/j.cell.2018.10.055
48. Kweon SM, Zhu B, Chen Y, Aravind L, Xu SY, Feldman DE. Erasure of Tet-Oxidized 5-Methylcytosine by a SRAP Nuclease. *Cell Rep*. 2017;21(2):482-494. doi:10.1016/j.celrep.2017.09.055
49. Podlesek Z, Žgur Bertok D. The DNA Damage Inducible SOS Response Is a Key Player in the Generation of Bacterial Persister Cells and Population Wide Tolerance. *Front Microbiol*. 2020;11:1785. doi:10.3389/fmicb.2020.01785
50. Aravind L, Anand S, Iyer LM. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced

eukaryotic DNA demethylation systems. *Biol Direct*. 2013;8(1):20. doi:10.1186/1745-6150-8-20

51. Spruijt CG, Gnerlich F, Smits AH, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*. 2013;152(5):1146-1159. doi:10.1016/j.cell.2013.02.004
52. Sirbu BM, Couch FB, Feigerle JT, Bhaskara S, Hiebert SW, Cortez D. Analysis of protein dynamics at active, stalled, and collapsed replication forks. *Genes Dev*. 2011;25(12):1320-1327. doi:10.1101/gad.2053211
53. Dungrawala H, Rose KL, Bhat KP, et al. The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. *Mol Cell*. 2015;59(6):998-1010. doi:10.1016/j.molcel.2015.07.030
54. Liang T, Bai J, Zhou W, et al. HMCES modulates the transcriptional regulation of nodal/activin and BMP signaling in mESCs. *Cell Rep*. 2022;40(2):111038. doi:10.1016/j.celrep.2022.111038
55. Buisson R, Boisvert JL, Benes CH, Zou L. Distinct but Concerted Roles of ATR, DNA-PK, and Chk1 in Countering Replication Stress during S Phase. *Mol Cell*. 2015;59(6):1011-1024. doi:10.1016/j.molcel.2015.07.029
56. Cimprich KA, Cortez D. ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol*. 2008;9(8):616-627. doi:10.1038/nrm2450
57. Mohni KN, Thompson PS, Luzwick JW, et al. A Synthetic Lethal Screen Identifies DNA Repair Pathways that Sensitize Cancer Cells to Combined ATR Inhibition and Cisplatin Treatments. *PLoS One*. 2015;10(5):e0125482. doi:10.1371/journal.pone.0125482
58. Thompson PS, Amidon KM, Mohni KN, Cortez D, Eichman BF. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat Struct Mol Biol*. 2019;26(7):613-618. doi:10.1038/s41594-019-0255-5
59. Paulin KA, Cortez D, Eichman B. The SOS response-associated peptidase (SRAP) domain of YedK catalyzes ring opening of abasic sites and reversal of its DNA-protein crosslink. *J Biol Chem*. 2022:119332. doi:10.1016/j.jbc.2022.102307
60. Helgren TR, Wangtrakuldee P, Staker BL, Hagen TJ. Advances in Bacterial Methionine Aminopeptidase Inhibition. *Curr Top Med Chem*. 2016;16(4):397-414. doi:10.2174/1568026615666150813145410
61. Halabelian L, Ravichandran M, Li Y, et al. Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat Struct Mol Biol*. 2019;26(7):607-612. doi:10.1038/s41594-019-0246-6
62. Wang N, Bao H, Chen L, et al. Molecular basis of abasic site sensing in single-stranded

DNA by the SRAP domain of E. coli yedK. *Nucleic Acids Res*. 2019;47(19):10388-10399. doi:10.1093/nar/gkz744

63. Amidon KM, Eichman BF. Structural biology of DNA abasic site protection by SRAP proteins. *DNA Repair (Amst)*. 2020;94:102903. doi:10.1016/j.dnarep.2020.102903
64. Srivastava M, Su D, Zhang H, et al. HMCES safeguards replication from oxidative stress and ensures error-free repair . *EMBO Rep*. 2020;21(6):1-14. doi:10.15252/embr.201949123
65. Olivieri M, Cho T, Álvarez-Quilón A, et al. A Genetic Map of the Response to DNA Damage in Human Cells. *Cell*. 2020;182(2):481-496.e21. doi:10.1016/j.cell.2020.05.040
66. Langenbucher A, Bowen D, Sakhtemani R, et al. An extended APOBEC3A mutation signature in cancer. *Nat Commun*. 2021;12(1):1602. doi:10.1038/s41467-021-21891-0
67. Mehta KPM, Lovejoy CA, Zhao R, Heintzman DR, Cortez D. HMCES Maintains Replication Fork Progression and Prevents Double-Strand Breaks in Response to APOBEC Deamination and Abasic Site Formation. *Cell Rep*. 2020;31(9):107705. doi:10.1016/j.celrep.2020.107705
68. Biayna J, Garcia-Cao I, Álvarez MM, et al. Loss of the abasic site sensor HMCES is synthetic lethal with the activity of the APOBEC3A cytosine deaminase in cancer cells. *PLoS Biol*. 2021;19(3):e3001176. doi:10.1371/journal.pbio.3001176
69. Victora GD, Nussenzweig MC. Germinal Centers. *Annu Rev Immunol*. 2012;30(1):429-457. doi:10.1146/annurev-immunol-020711-075032
70. Hwang JK, Alt FW, Yeap L-S. Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. *Microbiol Spectr*. 2015;3(1):MDNA3-0037-2014. doi:10.1128/microbiolspec.MDNA3-0037-2014
71. Wu L, Shukla V, Yadavalli AD, et al. HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes Dev*. 2022;36(7):433-450. doi:10.1101/gad.349438.122
72. Pan Y, Zuo H, Wen F, et al. HMCES safeguards genome integrity and long-term self-renewal of hematopoietic stem cells during stress responses. *Leukemia*. 2022;36(4):1123-1131. doi:10.1038/s41375-021-01499-5
73. Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol*. 2016;26(1):52-64. doi:10.1016/j.tcb.2015.07.009
74. Shukla V, Halabelian L, Balagere S, et al. HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells.

Mol Cell. 2020;77(2):384-394.e4. doi:10.1016/j.molcel.2019.10.031

75. Kühbacher U, Duxin JP. How to fix DNA-protein crosslinks. *DNA Repair (Amst)*. 2020;94(May):102924. doi:10.1016/j.dnarep.2020.102924
76. Ide H, Shoulkamy MI, Nakano T, Miyamoto-Matsubara M, Salem AMH. Repair and biochemical effects of DNA-protein crosslinks. *Mutat Res*. 2011;711(1-2):113-122. doi:10.1016/j.mrfmmm.2010.12.007
77. Baker DJ, Wuenschell G, Xia L, et al. Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J Biol Chem*. 2007;282(31):22592-22604. doi:10.1074/jbc.M702856200
78. Leng X, Duxin JP. Targeting DNA-Protein Crosslinks via Post-Translational Modifications. *Front Mol Biosci*. 2022;9(July):1-14. doi:10.3389/fmolb.2022.944775
79. Sun Y, Chen J, Huang S yin N, et al. PARylation prevents the proteasomal degradation of topoisomerase I DNA-protein crosslinks and induces their deubiquitylation. *Nat Commun*. 2021;12(1):1-16. doi:10.1038/s41467-021-25252-9
80. Larsen NB, Gao AO, Sparks JL, et al. Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. *Mol Cell*. 2019;73(3):574-588.e7. doi:10.1016/j.molcel.2018.11.024
81. Duxin JP, Dewar JM, Yardimci H, Walter JC. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell*. 2014;159(2):346-357. doi:10.1016/j.cell.2014.09.024
82. Wu RA, Semlow DR, Kamimae-Lanning AN, et al. TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature*. 2019;567(7747):267-272. doi:10.1038/s41586-019-1002-0
83. Hoffmann S, Smedegaard S, Nakamura K, et al. TRAIP is a PCNA-binding ubiquitin ligase that protects genome stability after replication stress. *J Cell Biol*. 2016;212(1):63-75. doi:10.1083/jcb.201506071
84. Sparks JL, Chistol G, Gao AO, et al. The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell*. 2019;176(1-2):167-181.e21. doi:10.1016/j.cell.2018.10.053
85. Vaz B, Popovic M, Newman JA, et al. Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. *Mol Cell*. 2016;64(4):704-719. doi:10.1016/j.molcel.2016.09.032
86. Stingele J, Bellelli R, Alte F, et al. Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol Cell*. 2016;64(4):688-703. doi:10.1016/j.molcel.2016.09.031

87. Feeney L, Muñoz IM, Lachaud C, et al. RPA-Mediated Recruitment of the E3 Ligase RFWD3 Is Vital for Interstrand Crosslink Repair and Human Health. *Mol Cell*. 2017;66(5):610-621.e4. doi:10.1016/j.molcel.2017.04.021
88. Gallina I, Hendriks IA, Hoffmann S, et al. The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol Cell*. 2020:1-17. doi:10.1016/j.molcel.2020.11.029
89. Zhang H-F, Tomida A, Koshimizu R, Ogiso Y, Lei S, Tsuruo T. Cullin 3 promotes proteasomal degradation of the topoisomerase I-DNA covalent complex. *Cancer Res*. 2004;64(3):1114-1121. doi:10.1158/0008-5472.can-03-2858
90. Kerzendorfer C, Whibley A, Carpenter G, et al. Mutations in Cullin 4B result in a human syndrome associated with increased camptothecin-induced topoisomerase I-dependent DNA breaks. *Hum Mol Genet*. 2010;19(7):1324-1334. doi:10.1093/hmg/ddq008
91. Serbyn N, Noireterre A, Bagdiul I, et al. The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. *Mol Cell*. 2020;77(5):1066-1079.e9. doi:10.1016/j.molcel.2019.12.007
92. Nowicka U, Zhang D, Walker O, et al. DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin. *Structure*. 2015;23(3):542-557. doi:10.1016/j.str.2015.01.010
93. Yip MCJ, Bodnar NO, Rapoport TA. Ddi1 is a ubiquitin-dependent protease. *Proc Natl Acad Sci U S A*. 2020;117(14):7776-7781. doi:10.1073/pnas.1902298117
94. Kojima Y, Machida Y, Palani S, et al. FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-15170-7
95. Welter AL, Machida YJ. Functions and evolution of FAM111 serine proteases. *Front Mol Biosci*. 2022;9:1081166. doi:10.3389/fmolb.2022.1081166
96. Borgermann N, Ackermann L, Schwertman P, et al. SUMO ylation promotes protective responses to DNA -protein crosslinks . *EMBO J*. 2019;38(8):1-17. doi:10.15252/embj.2019101496
97. Ruggiano A, Vaz B, Kilgas S, et al. The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability. *Cell Rep*. 2021;37(10):110080. doi:10.1016/j.celrep.2021.110080
98. Lopez-Mosqueda J, Maddi K, Prgomet S, et al. SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *Elife*. 2016;5. doi:10.7554/eLife.21491
99. Halder S, Torrecilla I, Burkhalter MD, et al. SPRTN protease and checkpoint kinase 1

cross-activation loop safeguards DNA replication. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-11095-y

100. Saha LK, Murai Y, Saha S, et al. Replication-dependent cytotoxicity and Spartan-mediated repair of trapped PARP1-DNA complexes. *Nucleic Acids Res*. 2021;49(18):10493-10506. doi:10.1093/nar/gkab777
101. Coleman KE, Yin Y, Lui SKL, et al. USP1-trapping lesions as a source of DNA replication stress and genomic instability. *Nat Commun*. 2022;13(1):1740. doi:10.1038/s41467-022-29369-3
102. Li F, Raczynska JE, Chen Z, Yu H. Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. *Cell Rep*. 2019;26(12):3336-3346.e4. doi:10.1016/j.celrep.2019.02.082
103. Reinking HK, Kang HS, Götz MJ, et al. DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. *Mol Cell*. 2020;80(1):102-113.e6. doi:10.1016/j.molcel.2020.08.003
104. Ruggiano A, Ramadan K. DNA–protein crosslink proteases in genome stability. *Commun Biol*. 2021;4(1):1-11. doi:10.1038/s42003-020-01539-3
105. Weickert P, Li H-Y, Götz MJ, et al. SPRTN patient variants cause global-genome DNA-protein crosslink repair defects. *Nat Commun*. 2023;14(1):352. doi:10.1038/s41467-023-35988-1
106. Prasad R, Horton JK, Wilson SH. Requirements for PARP-1 covalent crosslinking to DNA (PARP-1 DPC). *DNA Repair (Amst)*. 2020;90:102850. doi:10.1016/j.dnarep.2020.102850
107. Khodyreva SN, Prasad R, Ilina ES, et al. 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*. 2010;107(51):22090-22095. doi:10.1073/pnas.1009182107
108. Price NE, Johnson KM, Wang J, Fekry MI, Wang Y, Gates KS. Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J Am Chem Soc*. 2014;136(9):3483-3490. doi:10.1021/ja410969x
109. Semlow DR, MacKrell VA, Walter JC. The HMCES DNA-protein cross-link functions as an intermediate in DNA interstrand cross-link repair. *Nat Struct Mol Biol*. 2022;29(5):451-462. doi:10.1038/s41594-022-00764-0
110. Semlow DR, Mackrell VA, Walter JC. Thea HMCES DNA-protein cross-link functions as an intermediate in DNA interstrand cross-link repair. *Nat Struct Mol Biol*. 2022;29(May).
111. Yaneva D, Sparks JL, Donsbach M, et al. The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Mol Cell*. 2023;83(1):43-56.e10.

doi:10.1016/j.molcel.2022.12.005

112. Sugimoto Y, Masuda Y, Masutani C, Iwai S, Miyake Y, Kanao R. Novel mechanisms for the removal of strong replication-blocking HMCES- and thiazolidine-DNA adducts in humans. 2023:1-23.
113. Donsbach M, Dürauer S, Grünert F, et al. A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks. *EMBO J*. 2023;42(18):e113360. doi:10.15252/embj.2022113360
114. Torrecilla I, Ruggiano A, Kiianitsa K, et al. Isolation and detection of DNA–protein crosslinks in mammalian cells. *Nucleic Acids Res*. December 2023:gkad1178. doi:10.1093/nar/gkad1178
115. Kiianitsa K, Maizels N. A rapid and sensitive assay for DNA-protein covalent complexes in living cells. *Nucleic Acids Res*. 2013;41(9):1-7. doi:10.1093/nar/gkt171
116. Lindahl T. DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. *Prog Nucleic Acid Res Mol Biol*. 1979;22:135-192. doi:10.1016/s0079-6603(08)60800-4
117. Han T, Goralski M, Capota E, et al. The antitumor toxin CD437 is a direct inhibitor of DNA polymerase α. *Nat Chem Biol*. 2016;12(7):511-515. doi:10.1038/nchembio.2082
118. Ercilla A, Benada J, Amitash S, et al. Physiological Tolerance to ssDNA Enables Strand Uncoupling during DNA Replication. *Cell Rep*. 2020;30(7):2416-2429.e7. doi:https://doi.org/10.1016/j.celrep.2020.01.067
119. Schormann N, Ricciardi R, Chattopadhyay D. Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci*. 2014;23(12):1667-1685. doi:10.1002/pro.2554
120. Cortázar D, Kunz C, Saito Y, Steinacher R, Schär P. The enigmatic thymine DNA glycosylase. *DNA Repair (Amst)*. 2007;6(4):489-504. doi:10.1016/j.dnarep.2006.10.013
121. Bhagwat AS, Hao W, Townes JP, Lee H, Tang H, Foster PL. Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. *Proc Natl Acad Sci U S A*. 2016;113(8):2176-2181. doi:10.1073/pnas.1522325113
122. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep*. 2016;14(6):1273-1282. doi:10.1016/j.celrep.2016.01.021
123. Chilkova O, Stenlund P, Isoz I, et al. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res*. 2007;35(19):6588-

6597. doi:10.1093/nar/gkm741

124. Van Esch H, Colnaghi R, Freson K, et al. Defective DNA Polymerase α-Primase Leads to X-Linked Intellectual Disability Associated with Severe Growth Retardation, Microcephaly, and Hypogonadism. *Am J Hum Genet*. 2019;104(5):957-967. doi:10.1016/j.ajhg.2019.03.006
125. Rios-Szwed DO, Garcia-Wilson E, Sanchez-Pulido L, et al. FAM111A regulates replication origin activation and cell fitness. *bioRxiv*. January 2020:2020.04.22.055574. doi:10.1101/2020.04.22.055574
126. Barghout SH, Patel PS, Wang X, et al. Preclinical evaluation of the selective small-molecule UBA1 inhibitor, TAK-243, in acute myeloid leukemia. *Leukemia*. 2019;33(1):37-51. doi:10.1038/s41375-018-0167-0
127. Stringer DK, Piper RC. A single ubiquitin is sufficient for cargo protein entry into MVBs in the absence of ESCRT ubiquitination. *J Cell Biol*. 2011;192(2):229-242. doi:10.1083/jcb.201008121
128. Enchev RI, Schulman BA, Peter M. Protein neddylation: Beyond cullin-RING ligases. *Nat Rev Mol Cell Biol*. 2015;16(1):30-44. doi:10.1038/nrm3919
129. Brownell JE, Sintchak MD, Gavin JM, et al. Substrate-Assisted Inhibition of Ubiquitin-like Protein-Activating Enzymes: The NEDD8 E1 Inhibitor MLN4924 Forms a NEDD8-AMP Mimetic In Situ. *Mol Cell*. 2010;37(1):102-111. doi:10.1016/j.molcel.2009.12.024
130. Enchev RI, Schulman BA, Peter M. Protein neddylation: Beyond cullin-RING ligases. *Nat Rev Mol Cell Biol*. 2015;16(1):30-44. doi:10.1038/nrm3919
131. Maskey RS, Kim MS, Baker DJ, et al. Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat Commun*. 2014;5:1-12. doi:10.1038/ncomms6744
132. Larsen NB, Gao AO, Sparks JL, et al. Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. *Mol Cell*. 2019;73(3):574-588.e7. doi:10.1016/j.molcel.2018.11.024
133. Jang SM, Redon CE, Thakur BL, Bahta MK, Aladjem MI. Regulation of cell cycle drivers by Cullin-RING ubiquitin ligases. *Exp Mol Med*. 2020;52(10):1637-1651. doi:10.1038/s12276-020-00508-4
134. Elia AEH, Wang DC, Willis NA, et al. RFWD3-Dependent Ubiquitination of RPA Regulates Repair at Stalled Replication Forks. *Mol Cell*. 2015;60(2):280-293. doi:10.1016/j.molcel.2015.09.011
135. Liu JCY, Kühbacher U, Larsen NB, et al. Mechanism and function of DNA replication-independent DNA-protein crosslink repair via the SUMO-RNF4 pathway. *EMBO J*.

2021;40(18):e107413. doi:10.15252/embj.2020107413

136. Bhargava V, Goldstein CD, Russell L, et al. GCNA Preserves Genome Integrity and Fertility Across Species. *Dev Cell*. 2020;52(1):38-52.e10. doi:10.1016/j.devcel.2019.11.007
137. Stingele J, Schwarz MS, Bloemeke N, Wolf PG, Jentsch S. A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell*. 2014;158(2):327-338. doi:10.1016/j.cell.2014.04.053
138. Yesbolatova A, Saito Y, Kitamoto N, et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat Commun*. 2020;11(1):5701. doi:10.1038/s41467-020-19532-z
139. Reinking HK, Kang H-S, Götz MJ, et al. DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. *Mol Cell*. 2020;80(1):102-113.e6. doi:https://doi.org/10.1016/j.molcel.2020.08.003
140. Donsbach M, Dürauer S, Nguyen KT, et al. An auto-release mechanism for HMCES-DNA-protein crosslinks. *bioRxiv*. 2022:2022.12.18.520715. doi:10.1101/2022.12.18.520715
141. Sugimoto Y, Masuda Y, Iwai S, Miyake Y, Kanao R, Masutani C. Novel mechanisms for the removal of strong replication-blocking HMCES- and thiazolidine-DNA adducts in humans. *Nucleic Acids Res*. April 2023:gkad246. doi:10.1093/nar/gkad246
142. Paulin KA, Cortez D, Eichman BF. The meee SOS response-associated peptidase (SRAP) domain of YedK catalyzes ring opening of abasic sites and reversal of its DNA-protein crosslink. *J Biol Chem*. August 2022. doi:10.1016/j.jbc.2022.102307
143. Howard SM, Yanez DA, Stark JM. DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining. *PLoS Genet*. 2015;11(1):e1004943. doi:10.1371/journal.pgen.1004943
144. Shukla V, Halabelian L, Balagere S, et al. HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. *Mol Cell*. 2020;77(2):384-394.e4. doi:10.1016/j.molcel.2019.10.031
145. Simsek D, Jasin M. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol*. 2010;17(4):410-416. doi:10.1038/nsmb.1773
146. Ma J-L, Kim EM, Haber JE, Lee SE. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol*. 2003;23(23):8820-8828. doi:10.1128/MCB.23.23.8820-

8828.2003

147. Pannunzio NR, Li S, Watanabe G, Lieber MR. Non-homologous end joining often uses microhomology: implications for alternative end joining. *DNA Repair (Amst)*. 2014;17:74-80. doi:10.1016/j.dnarep.2014.02.006
148. Bhargava R, Sandhu M, Muk S, Lee G, Vaidehi N, Stark JM. C-NHEJ without indels is robust and requires synergistic function of distinct XLF domains. *Nat Commun*. 2018;9(1):2484. doi:10.1038/s41467-018-04867-5
149. De Silva NS, Klein U. Dynamics of B cells in germinal centres. *Nat Rev Immunol*. 2015;15(3):137-148. doi:10.1038/nri3804
150. Nakamura M, Kondo S, Sugai M, Nazarea M, Imamura S, Honjo T. High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol*. 1996;8(2):193-201. doi:10.1093/intimm/8.2.193
151. Dennery PA. Effects of oxidative stress on embryonic development. *Birth Defects Res C Embryo Today*. 2007;81(3):155-162. doi:10.1002/bdrc.20098
152. Arfin S, Jha NK, Jha SK, et al. Oxidative Stress in Cancer Cell Metabolism. *Antioxidants (Basel, Switzerland)*. 2021;10(5). doi:10.3390/antiox10050642
153. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*. 2013;45(9):977-983. doi:10.1038/ng.2701
154. Jalili P, Bowen D, Langenbucher A, et al. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. *Nat Commun*. 2020;11(1):2971. doi:10.1038/s41467-020-16802-8
155. Kim DI, Jensen SC, Noble KA, et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell*. 2016;27(8):1188-1196. doi:10.1091/mbc.E15-12-0844