UNIVERSITY OF CALIFORNIA
RIVERSIDE

Investigating Abasic Sites in Mitochondrial DNA

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Jin Tang

June 2023

Dissertation Committee:
Dr. Linlin Zhao, Chairperson
Dr. Min Xue
Dr. Quan Cheng

The Dissertation of Jin Tang is approved:

Committee Chairperson

University of California, Riverside

COPYRIGHT ACKNOWLEDGEMENT

The text and figures in Chapter 2, in part or full, are a reprint of the material as it appears in ***Analytical Chemistry 2021, 93, 39, 13398-13406***. The coauthor (Dr. Linlin Zhao) listed in that publication directed and supervised the research that forms the basis of this chapter.

The text and figures in Chapter 3, in part or full, are a reprint of the material as it appears in ***DNA repair (Amst). 2022, 111, 103286***. The coauthor (Dr. Linlin Zhao) listed in that publication directed and supervised the research that forms the basis of this chapter.

ACKNOWLEDGEMENTS

With this chance, I want to say a big thank you to everyone who supported me and accompanied me for the past five years. Thank you for becoming my friends and my mentors in many aspects of research and my life and building who I am today. I sincerely thank you for being so patient, so helpful and so supportive for me that I gained the courage to overcome challenges and go that far in my career.

First and foremost, I would like to express my sincere appreciation to my Ph. D. advisor Dr. Linlin Zhao for all his support and advice for me. He is a very patient, creative and motivated researcher who is attracted to the research field of DPCs. He welcomes discussion and likes to hear about my opinions about the research topics, plans, and designs of the experiments. He always gave me insightful suggestions about my plans, and we usually could work out very efficient methods to perform the tests. I would never be able to go that far without his guidance. He is also very supportive of sending students to conferences and networking with researchers in the field. With these valuable opportunities, I furthered my understanding of the field by talking with many wonderful researchers and deeply fell in love with science. Besides, he also gave me a lot of advice on writing and working in academia and helped me become a real scientist. He gave me many career suggestions, which were super helpful. I learned from him not just knowledge in the field by also how to be a good researcher.

I also would like to thank my external mentor Dr. Laurie S Kaguni for providing me many valuable advice on my research projects and encouraging me to overcome research difficulties and searching for solutions. She is an excellent female model scientist; her passion and love in science have encouraged me to go further and to be persistence in the face of challenges.

It's my great pleasure to join such a young and motivated group. We always discuss our research findings with each other and are open to insights from people with expertise in different fields. I would like to thank our molecular biology expert, Dr. Wenyan Xu, for his generous sharing of his experience with us. I learned lots of efficient and accurate ways of performing various tests. These techniques are very helpful for my research progress and my career development. I would like to thank Dr. Chaoxing Liu for his valuable insights in nucleic acid chemistry and for sharing his very deep understanding of the chemistry process, which has helped me solve many related questions. I would like to thank Kathleen M. Urrutia for being very helpful and teaching me protein purification. I would like to thank Wenxin Zhao for her support of my projects and providing helpful information and materials for me to perform the tests and helping me with kinetic assays. I would like to thank Yu-Hsuan Chen for providing me with valuable protein and cell samples and collaborating with me on these projects. I also would like to thank my collaborators Dr. Yinsheng Wang, Dr. Feng Tang and Dr. Nathan G. Hendricks for their help with testing samples and data analysis. I would like to thank my lab members Ching-Hsin Yang, Dr. Anal Jana, Krystie Chew, Jacob Perkins, Ivan Jacuinde, Melissa Henderson,

Chhavi Shuklam Yuvraj Singh and Martin Sanchez. I would like to thank my previous lab members from Julian lab Dr. Ryan Julian, Dr. Dylan L. Riggs, Dr. Tyler R. Lambeth, Dr. Lance E. Talbert, Dr. Yana A Lyon, Dr. James G Bonner, Jacob W Silzel, Hoi-Ting Wu, Evan E Hubbard and Brielle L Van Orman. Especially Dr. Dylan L. Riggs for his generous sharing of the techniques in instrumentation and sample analysis using mass spectrometry with me.

I would like to thank a few friends from UC Riverside for their generous support for me, first I would like to thank Dr. Joseph C Genereux for teaching me knowledge in biochemistry during the past five years in Riverside and supported me for my job hunting. I also thank Dr. Ka Yang, Zongbo Li, Zihao Wang, Yinan Wang, Dr. Zhili Guo, Dr. Yi Huang, Dr. Ying Tan, Dr. Quanqing Zhang, Ziting Gao, Ziqi Lyu, Dr. Ravleen Kaur and Alexander Malinick.

I am also thankful to Dr. Min Xue and Dr. Quan (Jason) Cheng for serving on my dissertation committee. I appreciate their valuable comments and suggestions on my research and dissertation. And I would also like to thank Dr. Sean E O'Leary and Dr. Catharine H Larsen for their serving in my oral qualifying exam and for providing valuable suggestions to me.

I also would like to thank the staff members in Department of Chemistry, especially Christina Youhas, Prisciliano Saavedra, Jeffery Enrrique Matute and Dr. Kevin Simpson for their efforts in helping me out in many aspects.

I would like to acknowledge the financial support from UC Riverside and UCR Department of Chemistry for these years and the support from National Institute of General Medical Sciences. Without their support, my work cannot be done. I would like to especially thank the GSR for providing me with funding for conference travel and UCR Graduate Division for providing me with the Dissertation Year Program Award and the Department of Chemistry for the Donald T. Sawyer Award in Analytical Chemistry. These financial supports not only supported my research but also encouraged me a lot.

Finally, I would like to thank my family, especially my parents, for their love and support for me, thanks for helping me grow, teaching me so much knowledge, and helping me be a better person.

ABSTRACT OF THE DISSERTATION

Investigating Abasic Sites in Mitochondrial DNA

by

Jin Tang

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, June 2023
Dr. Linlin Zhao, Chairperson

Mitochondria are important subcellular compartments, crucial for energy production, metabolism, and cell signaling. Mitochondrial dysfunction is known to cause nearly 200 mitochondrial disorders and has been associated with aging and a variety of human diseases. Mitochondrial DNA (mtDNA) encodes 37 genes, including 13 proteins and a set of tRNA and rRNA. mtDNA is constantly threatened by chemical and physical assaults. Because mitochondria have limited DNA repair pathways, mtDNA damage accumulates and occurs at a higher level compared to nuclear DNA. Abasic (AP) sites are abundant DNA lesions that can be generated from various pathways, including base excision repair (BER). AP sites are highly reactive and can form secondary DNA adducts, DNA-interstrand cross-links (ICLs), and DNA-protein cross-links (DPCs). My dissertation project exploits the chemistry of AP sites and develops methods to explore biological processes pertinent to AP sites. First, I developed a mass spectrometry-based method to identify the cross-linked amino acid residues in DNA-protein cross-links. I designed DNA substrates with ribonucleotides (rNMPs), which provide chemical-labile sites for DNA

strand cleavage reactions and produce structurally defined DPCs. Also, I developed a program (AP_CrosslinkFinder) to accelerate data analysis. The method was applied to identify the cross-linking amino acid residues in DPCs derived from mitochondrial transcription factor A (TFAM). Second, I developed a method to prepare model ICLs using rNMP-containing DNA with a nucleotide analog 2-aminopurine. The alkaline lability of rNMP enables the generation of strand breaks at specific sites. AP sites react with 2-aminopurine with high yield and high rate. This method provides a simple and straightforward tool for investigating the impact of ICLs during the repair process. Third, I investigated the DNA terminal structures generated in TFAM-catalyzed AP-DNA strand cleavage. Quantification of reaction rates in the presence of biological amines and thiols demonstrates that GSH competes with TFAM for AP site strand breaks, suggesting a possible strategy to limit the formation of DPC and control the strand break terminus in cells. Removal of DNA terminal modifications by relevant DNA repair enzymes was also evaluated. Together, results from my dissertation provide insights into the complexity of AP site chemistry with important biological implications.

TABLE OF CONTENTS

LIST OF SCHEMES

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

DNA is an abundant macromolecule that stores genetic information of cells. Cells utilize the code from the DNA to transcript into RNA and then translate into proteins. Proteins have all types of functions, including catalyzing cellular reactions [1], constructing large functional complexes [2], and transporting nutrients throughout the cells [3]. Some of the proteins are building blocks of the cell and help forming cellular structures [4]. Some of the proteins are transported to the membrane and interact with the outer environment then transport the signals back to the cells as a way of giving feedback to DNA to control the expression of gene [5]. Therefore, the integrity of DNA is critical to cell survival.

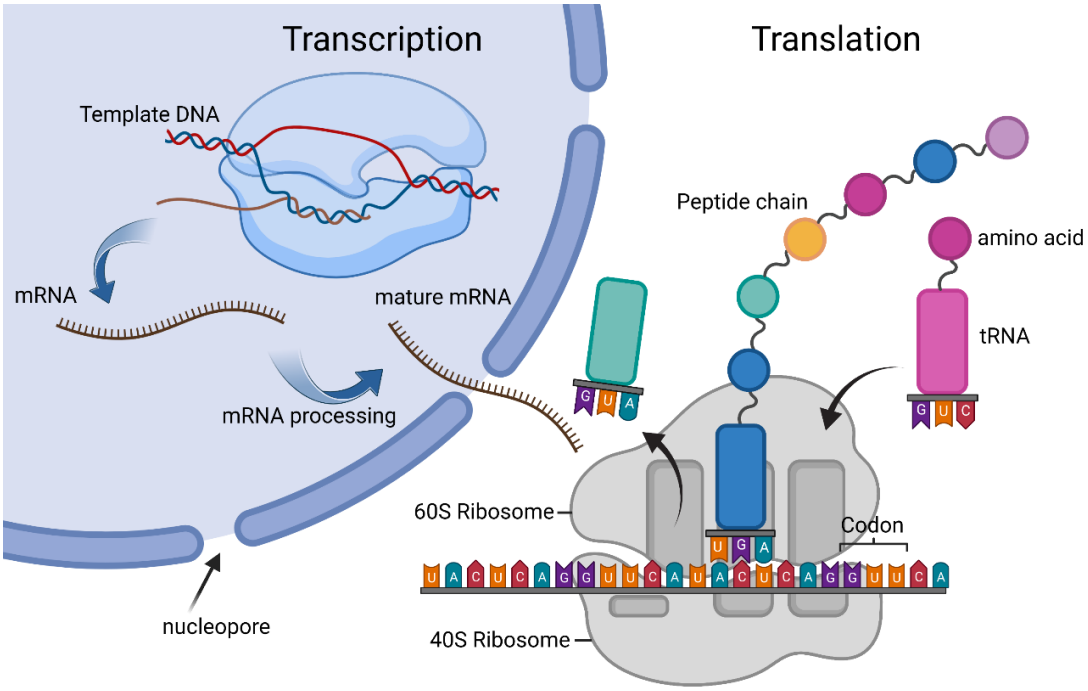


Figure 1.1 Scheme of DNA transcript into RNA and RNA translate into peptides. The information stored in DNA is passed down to synthesize the protein that have functions.

However, DNA is subjected to tons of environmental threats and produces thousands of lesions every day [6]. There are many types of DNA lesions, such as base modifications [7], base mismatches [8], single/double strand breaks [9], producing abasic (AP) sites[10], covalently cross-links with proteins [11], intrastrand and interstrand cross-links [12]. These lesions don't have the same biological significance. Modifications like DNA methylation and histone acetylation [13,14], are intentionally produced to control the expression of genes and they are reversible modifications, while most of the others are not reversible and generated without regulation. Exogenous factors, such as UV light, chemical exposure, chemotherapeutic drugs, and ionizing radiation, can cause DNA lesions. Also, endogenous factors, like errors in replications, exposure to reactive oxygen species, insufficient DNA repairs and cellular metabolism, can be the origin of DNA lesions [15].

The repair process is intricate in that cells need to treat different lesions with highly targeted repair mechanisms, and during this process, cells must remove the modifications, fill the gaps with newly synthesized nucleotides by DNA polymerases and link the gap with ligases to recover the correct strand [16]. Besides the repair, cells are tolerable to some extent of DNA lesions and can still maintain their proper functions [17]. Different types of DNA modifications might cause mutations without affecting the survival of the species, and therefore these mutations remain and might pass to their offspring, which is meaningful for the species to be prepared for the unpredicted sudden changes in the environment. But some types of DNA lesions are highly toxic; if cells can't repair

them on time, these lesions make the cell couldn't perform some vital tasks, showing abnormal metabolic processes, being oncogenic, or inducing cell cycle arrest and apoptosis [18].

With the increasing amount of environmental pollution, human, along with other species on Earth, are facing more exposure to the toxicants and potentially encounter more exposure that could cause DNA lesions, which requires more efficient repair processes, but the progress of evolution is slower than industrialization of human society. Therefore, more and more diseases that originated from exposing to chemical contaminants that caused DNA lesions are emerging [19], such as cancers, asbestosis, leukemia, developmental disorders for children, Parkinson's disease and so on. The most recent public health reported from WHO shows that there are more than 2 million lives and 53 million disability-adjusted life-years were lost in 2019 [20]. This raises a huge public concern and calls for the studies about the mechanisms of chemical exposure to human diseases, especially for their impacts on producing DNA lesions.

In this chapter, we will first give a general introduction about the abundant DNA lesion and abasic (AP) sites, along with the methods of detection and quantification. Then introduce a less common but highly toxic DNA lesion, DNA-protein cross-links (DPCs), and will discuss their unique characters and the methods for analyzing DPCs. The repair process of DNA interstrand cross-links and their biological impact will also be included, followed by the discussion of the importance of mitochondrial DNA repair.

1.1 Abasic sites

Among all types of DNA lesions, AP sites are the most unique and abundant DNA lesions, accounting for 50% of total endogenous DNA lesions, with roughly 10,000 AP sites are produced per cell per day [21]. AP sites form when the base group on nucleotide is missing, there are many ways to generate AP sites, including spontaneous deamination of base groups [21], enzymatic hydrolysis of the glycosidic bond during the repair processes [22] and radical reactions during UV and ionizing radiation [23]. Since most of the AP sites are produced during the base excision repair (BER) process, they are also viewed as a DNA repair intermediate, but if the downstream DNA repair enzymes are insufficient, the accumulation of AP sites might happen.

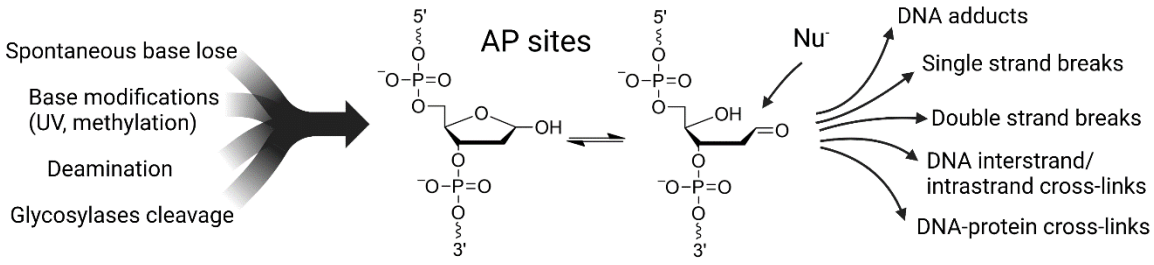


Figure 1.2 AP site formation and products of AP sites with nucleophiles that cause deleterious consequences for genome stability.

AP sites are highly reactive due to the reactive aldehyde group in the ring-opened form, which favors nucleophilic addition reactions and increases genomic instability. After being attacked by the nucleophilic molecules, AP sites turn into DNA adducts [24], strand breaks [25], and DNA interstrand and intrastrand cross-links [26], which further increase the complexity and requirements of the repair process. And for this reason, most

of the studies utilize the reactivity of AP sites to detect, quantify and do AP site sequencing based on aldehyde chemistry.

AP sites don't have any base groups to pair with, therefore, some DNA polymerases will stall at abasic sites [27], but some DNA polymerases, like translesion synthetase (TLS), could add a random nucleotide opposite to abasic site to bypass the lesion, introducing mutations [27]. A-rule was observed that DNA polymerases prefer adding deoxyadenosine (dA) opposite to abasic sites [28]. Therefore, AP sites can block replication and transcription and are mutagenic.

1.2 Methods to detect abasic sites

To better understand how AP sites are generated and how they interact with other cellular compartments, knowing the amount and locations of abasic sites is very important. Global level quantification of AP sites is quick and easy to perform and gives how the abundance of abasic sites related to experimental conditions, which is sufficient to analyze the cellular responses under drugs [29]. But in recent years, the needs for more detailed localized information have increased and benefited from the development of high-throughput next-generation sequencing (NGS) technique, which allows the accurate sequencing of abasic sites down to single nucleotide resolution [30]. This detailed information will be helpful for the biologist to analyze the biological impacts of the drugs, find highly targeted drugs, and advance the treatment of cancer and aging.

1.2.1 Chemical probe for the quantification of abasic sites

Chemical probes used to detect abasic sites contain two groups: one to covalently bind the aldehyde group, usually a nucleophile, and the other to provide a means for detection or enrichment [29]. There are three important requirements for these probes: first, the aldehyde reactive group should be highly specific to abasic sites even in the presence of other aldehydes in the cellular environment; second, the reaction efficiency of the aldehyde reactive group should be high, and the reaction should be fast and can be performed in the proper pH range and in the aqueous solution, and third, the ability to detect/differentiate different forms of abasic sites.

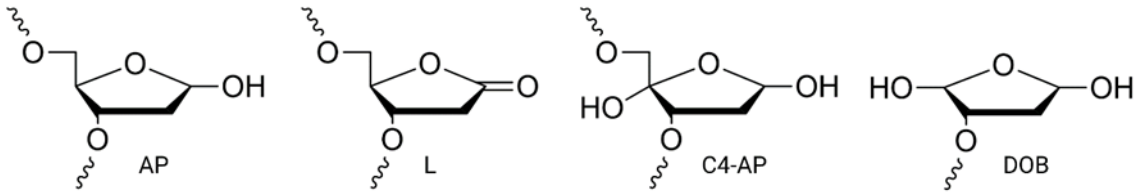


Figure 1.3 Structure of different AP sites. AP site is formed by removing the base group (AP). Upon oxidation by one-electron nucleotide, the AP lesion becomes 2'-deoxyribonolactone (L) [31]. DNA polymerases usually incorporate deoxythymidine (dT) opposite to L [32]. C4-AP forms when treated DNA with antitumor drugs, like bleomycin, C4-AP resembles the AP lesion, which is highly reactive and blocks DNA replication [33]. UV light and ROS can produce 5'-(2-Phosphoryl-1,4-dioxobutane) (DOB), DOB reacts rapidly with primary amines and irreversibly inhibits the function of lyases through DPC formation [34].

The progress of aldehyde reactive probes (ARP) involves improving their reactivity under physiological pH by utilizing different aldehyde reactive groups, strengthening their interaction with DNA in an aqueous solution [35], and increasing the detection performance and signal-to-noise ratio by using more compatible and strong fluorophores.

Due to the availability of the probes, some of the probes are difficult to apply in a large number of samples.

The most frequently used aldehyde reactive group is the aminooxy (-O-NH_2) group, in which the probe forms a Schiff base with the AP site, then it is treated with avidin biotinylated horseradish peroxidase and any unbonded DNA is washed away, then quantify the abasic site based on the absorbance [10]. But ARP can also react with other reactive aldehydes in DNA, especially the formylated bases, such as 5-formylcytosine (5-fC) and 5-formyluracil (5-fU) [36–38], and because they are more stable than AP site in the genome, they tend to be more readily detected [39,40]. Therefore, the ARP probe needs to improve the specificity of detection. Some other groups used the hydrazine group (-NH-NH_2), which also lacks specificity. Liu et al. developed a probe that forms a stable bond with AP site through the hydrazino-iso-Pictet–Spengler (HIPS) reaction and utilized the alkaline lability of AP sites to map their locations, with improved specificity [30]. It could envision that a more specific detection method for AP sites should incorporate the structure of AP site instead of the aldehyde group alone.

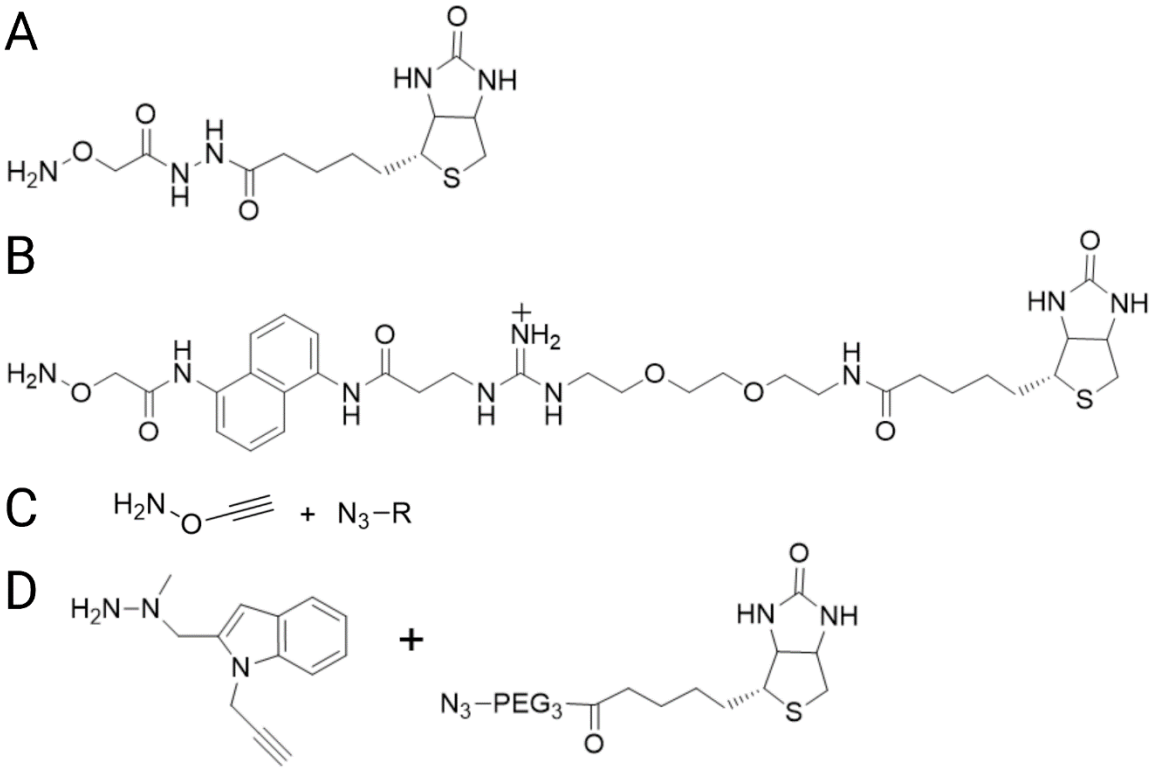


Figure 1.4 The structure of AP probes. (**A**) Commercially available ARP probe. This probe contains an aminooxy group and biotin, aminooxy group enables the nucleophilic addition reaction while the biotin enables the enrichment with streptavidin. (**B**) AP probe structure developed by Kojima et al. Based on ARP probe, they added hydrophobic naphthalene which conjugates with AP sites to increase the affinity. And the hydrophilic residue to increase the solubility in aqueous solution [35]. (**C**) Alkoxyamine (AA3) reacts with AP sites with a better pH range and higher reactivity than ARP. Also, this probe contains the alkyne group to enable the click chemistry that could add tags on the AP sites [41]. (**D**) The probe developed by Liu et al could selectively detect AP sites in the presence of 5-fC and 5-fU, the probe forms covalent bond with AP sites through hydrazine-*iso*-Pictet-Spengler (HIPS) reaction. This probe also contains an alkyne group to enable the click chemistry [30].

The fluorophore-containing probes utilize the fluorescent signal to quantify the abundance of abasic sites, which could enhance the detection and lower the quantification limit. Directly conjugating the fluorophores to DNA lesion site is not common and in most cases the probes are designed with three portions (aldehyde

reactive site, linker and fluorophore or affinity tag) [35], a near-infrared fluorescent probe developed for imaging of abasic sites, which showed higher affinity than ARP [42], but the fluorophores are hydrophobic and is a drawback for cell penetration. Besides the probes that need to react with AP sites, some other probes can also display signals when they interact with AP sites, and because of their easy handling procedure, these noncovalent probes have even wider applications. A plant originated chemical, protoberberine alkaloid palmatine (PAL), was identified as a fluorophore-switch probe to selectively target the AP nanocavity. The activation of the fluorescence of PAL is the result of the bridged π conjugation [43]. Wu et al utilized the 9-dicyanovinyljulolidine (DCVJ) to target AP sites and has a fluorescence turn-on response, this probe overcomes the previous limitation that the excited state of the fluorophore is deactivated by electron transfer with guanine [44].

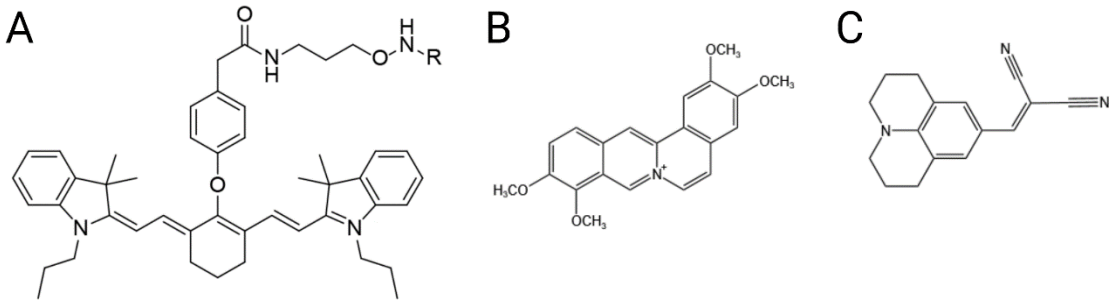


Figure 1.5 Structure of the fluorescent AP probes. (**A**) The near infrared (NIR) fluorescent probe developed by Condie et al. The R in the structure is the -H, and it covalently tags AP sites with the aminooxy group [42]. (**B**) Structure of PAL. This type of noncovalent AP probes displays fluorescent on when AP sites exist [43]. (**C**) Structure of 9-dicyanovinyljulolidine (DCVJ) [44].

To enable the sequencing for low abundance DNA lesions, it always requires an enrichment step of that lesion from the sample matrix. Therefore affinity tags are commonly seen, but it's not required for quantification. The most common affinity tag is a small molecule tag——biotin, it is enabled in the synthesis step of the probe, and it has strong and specific interaction with streptavidin, which can be immobilized into beads and the overall workflow of enrichment is very efficient. Researchers introduce biorthogonal chemistry, such as the click chemistry, to increase the specificity of enrichment [41].

1.2.2 Enzymatic methods for AP site detection

AP endonuclease 1 (APE1) is a major AP endonuclease and an important BER enzyme that cleaves the DNA sugar-phosphate backbone at 5' position of the AP site, producing a 3'-hydroxyl (3'-OH) group and a 5'-deoxyribose phosphate (5'-dRp) group, these 3'-OH ends can be recognized by DNA polymerases to initiate strand extension, while the 5'-dRp end can be removed by the subsequent enzymes in BER pathway. APE1 has a rigid, pre-formed DNA-binding face that specifically targets AP sites located in double-strand DNA. APE1 bends the dsDNA to about 35° and interacts with the AP-DNA strand to enable the interaction with the flipped-out AP site with the catalytic core [45]. Therefore, APE1 has a pre-determined selection for the substrate DNA. The substrate preference for APE1 is double-strand DNA > flap DNA > single-strand DNA [46]. Cai et al. applied the cleavage activity of APE1 at the AP site to sequence the 3' strand break site for sequencing the AP sites [47]. However, previous research has demonstrated there are

multiple factors that can affect the APE1 cleavage efficiency, such as nearby DNA lesions, DNA sequence context [48], DNA tertiary structure [49,50], the surrounding protein factors [51], etc. Also, APE1 harbors exonuclease activity, which increases the ambiguity of the strand break site.

Another enzyme has been used for sequencing the AP site is the Endonuclease IV (Endo IV) [52], which has the same cleavage products as the APE1, but for Endo IV, it can also cleave other types of 3'-terminus into 3'-OH, such as 3'-phosphate and 3'-α,β-unsaturated aldehyde.

1.3 DNA-protein cross-links

DNA is not fully exposed in the cell, instead, a majority of the DNA is packed by DNA-binding proteins. These proteins, such as histones, can bend the DNA, causing it to fold and form compact chromosomes. The release of DNA from histone enables their binding with proteins in the replication folk [53] and read the genetic information, making DNA-protein interactions crucial for cellular maintenance. Also, to understand the regulation of transcription, it is important to conduct the genome-wide mapping of DNA protein interactions and study the modifications that occur on these proteins and DNA. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is widely used to map the DNA protein interactions at the single nucleotide level and study the modifications on proteins. This method, which used formaldehyde to cross-link DNA and proteins within cells, enables us to study the chromatin structure, probe the specific genes, and uncover

the proteins that specifically interact with them [54]. Alternatively, a photo-crosslinking chemistry method are applied in living cells, which utilizing high-intensity UV laser to induce radical formation in proteins and DNA, causing the formation of covalent bonds between them [55]. Researchers can then immunoprecipitate the protein of interest to study its DNA footprinting. It is worth noting that both formaldehyde cross-linking and UV-crosslinking methods are non-targeted and have an inherent preference for forming the cross-links with more reactive groups [56].

1.3.1 Formation of DNA-protein cross-links

DPCs are highly toxic DNA lesions that can arise from both endogenous sources (ROS or cell metabolism) and exogeneous sources (chemotherapeutics drugs, UV light or ionizing radiation). The formation of DPCs occurs when DNA is subjected to chemical or physical assault leading to the generation of covalent bonds between DNA and proteins [57]. The proteins on the DNA makes DPCs extremely bulky and pose significant challenges for the DNA polymerases during replication, as a result, DPCs have strong replication blocking effects [58]. This mechanism is also applied in certain anti-cancer drugs, like cis-platinum and nitrogen mustard, which play as the linker between DNA and protein. The crosslinked products formed by these drugs are difficult to remove, impede replication, block transcription, affect the activity of cancer cells, and finally cause apoptosis [59,60].

The repair process of DPCs is more complex compared to repairing the damaged DNA or protein, as both DNA and proteins have well-defined repair pathways. The current understanding of DPC repair involves the participation of at least three enzymes: DNase for DNA incision, protease for protein digestion, and cleavage of the linker bonds [61]. Since the cross-linking process are quite diverse, the structure of linker varies, making the repair of DPC a case-specific repair rather than a universally applicable repair pathway. The challenges in the repair makes DPCs accumulate in the cells along with aging and neurodegenerative diseases.

Recently, researchers have revealed that the formation of DPCs is not completely damage prone, instead, some of them are formed during a protective process. For instance, HMCES can form covalent cross-linking bond with the AP sites through its N-terminal cysteine [62,63]. This allows HMCES to cap the AP sites, preventing their interaction with other cellular components. HMCES also form DPCs during the repair of interstrand cross-links (ICLs)[64]. The formation of AP sites with DNA binding and packaging protein histone has profound meaning for genomic maintenance[65]. In this reaction , lysine residues on histone form cross-link bond with AP sites through Schiff base chemistry, producing the DPC intermediate. Then, DPC undergoes the strand break reaction through the β-elimination reaction, and flowing the strand break reaction, DPC dissociate through hydrolysis, releasing the strand break AP site with 3’-α,β-unsaturated aldehyde and histone [65]. The formation of DPC is viewed as a catalyzing process that accelerates strand break formation from AP sites, potentially inducing the instability of

genome. Previous research from our lab found that an important mitochondria binding and packaging protein, transcription factor A (TFAM), forms cross-links with AP sites via Schiff base chemistry, leading to the formation of DPC intermediates and accelerated the formation of strand breaks [66]. Different cross-linked amino acids produced distinct cross-linked products and these DPCs exhibit variations in stability.

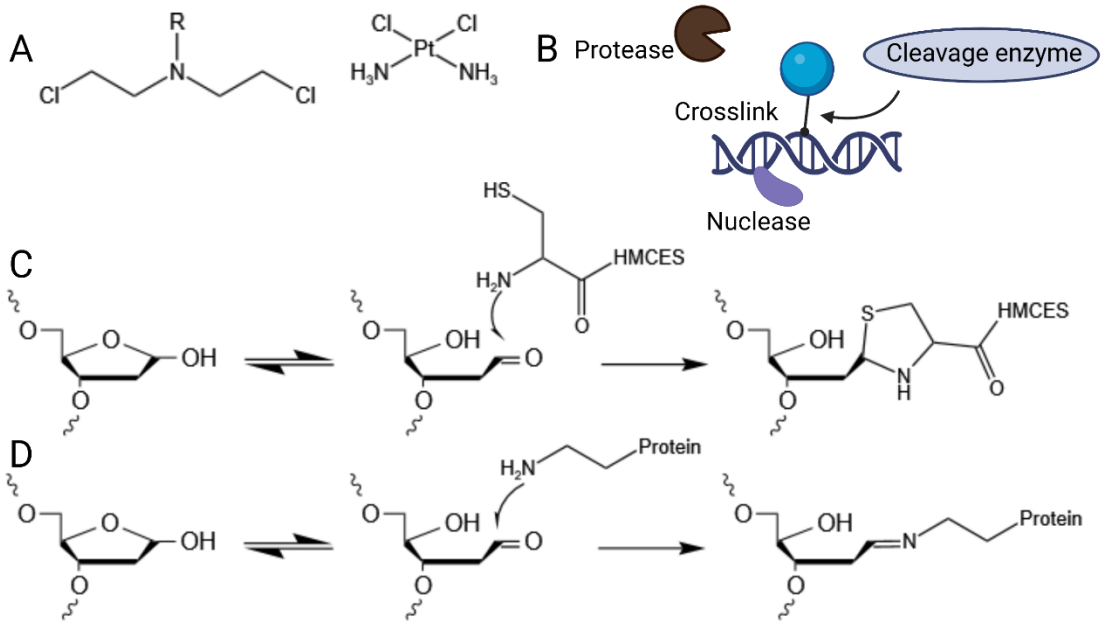


Figure 1.6 DPC formation and repair. (**A**) Structure of nitrogen mustard (left) and cis-platin (right). (**B**) The repair process of DPCs proposed by Stingele et al [61]. In this process, protein will be digested by protease and DNA digested by nucleases and along with the hydrolysis of the cross-linked bonds. (**C**) Formation of covalent bond between N-terminal cysteine of HMCES and AP site [62,63]. (**D**) Formation of a covalent bond between protein lysine side chain and AP site [65].

1.3.2 Importance of the cross-linking sites

To clarify the role of the DPCs in the repair process, it's vital to develop analytical methods that can accurately detect the structure of these lesions and elucidate the

chemistry of the cross-links. In addition to identification, quantitative analysis is also required and it's a powerful tool for understanding cellular responses to cross-linking factors.

Identification of the interactive amino acid residues is crucial for understanding the chemistry of DNA-protein interactions. Structural biology approaches, such as the X-ray crystallography and cryo-electron microscopy (cryo-EM) can be applied to interpret the cross-linked structure. However, these technique present challenges, as they require a large number of samples but provide limited information about the thermodynamic processes, therefore, a time-resolve method is needed.

1.3.3 Detection and quantification of DPCs

Global detection and quantification methods for DPCs have developed over time and involve multiple assays and they are continually developing and improving over time. One of the early methods used to detect DPCs is the comet assay. In this assay, the cross-linked protein retards the migration of DNA, resulting in a comet-like tailing effect, by comparing the extent of tailing, researchers can estimate the relative abundance of DPCs [67]. However, this method is limited in terms of accuracy and specificity, because comet assay can also apply to other types of DNA damage, and it could not isolate DPCs to confirm its discovery.

Another method for DPC detection is the filter-based method, which utilizes stationary phases such as nitrocellulose membranes and polyvinyl filters. These stationary

phases can retain either DNA or proteins, and a subsequent wash step is performed to remove any unbonded molecules [68]. The interaction with the stationary phase is usually based on the charge-charge interactions, as DNA carries negative charged, while proteins can be positively or negatively charged depending on the solution pH. Therefore, by adjusting the pH of the washing solutions, it can selectively elute out the unbonded DNA or protein. But since DPCs have diverse chemical/physical properties, this step will result in the loss of several DPCs and the retention of unbound DNA or proteins, introducing errors in the following identification and quantification. Another issue with this method is the recovery of the DPCs from the filter is difficult, therefore, it is hard to reproduce the results. Due to these limitations, the filter-based methods might to limited in applications.

Another method for DPC detection is a precipitation-based method developed by Zhitkovich and Costa. They precipitated DPC using dodecyl sulfate/potassium (SDS/K^+). The cross-linked cell mixture was treated with SDS and heated to dissociate the unbonded proteins, all proteins are negatively charged by the SDS. Followed by adding the K^+ containing salt (KCl), the DPCs become neutral and form precipitates. The DNA in the precipitates can then be visualized with DNA staining dyes [69], but this method will inevitably result in the precipitation of some unbonded proteins due to the solubility change and cause confusion in the identification of bonded proteins.

A common extraction-based method, phenol-chloroform extraction, is also applied into the isolation of DPCs. This method utilizes the hydrophilicity of DNA, hydrophobicity of lipids and the amphipathic property of proteins to separate these components in the nuclear lysate. After centrifugation, lipids and hydrophobic proteins (usually unbounded proteins) are in the bottom organic layer, the top aqueous layer contains the chromosomal DNA, and the interface contains DPCs and other proteins. DPC can then be precipitated using cold ethanol. This method is usually used to purify the DNA from cell lysates, and due to the diverse property of DPCs, the upper layer might contain DPCs, and the interface might contain free proteins and it is hard to eliminate the issue because DPC contains both DNA and protein.

Barker et al. had applied a commercialized chaotropic reagent (DNAzol) to lyse the cells and also dissociate the noncovalently bonded protein from DNA, then they used silica, which has strong affinity with DNA under alkaline conditions. In this way, they isolated all the DNA with DPCs from the sample matrix [70]. This method is used in a relatively challenging environment (incubation with 8 mM NaOH at 37°C overnight) for DPCs and might cause DNA lesions or DPC hydrolysis during the sample preparations.

1.3.4 Mass spectrometry analysis of the cross-links

Only in recent years, researchers began to utilize mass spectrometry to resolve the cross-links, termed cross-linking mass spectrometry (XL-MS). Cross-linkers and UV light have widely applied to induce the cross-links between biomacromolecules, and by

mapping the cross-linked sites, researchers can detect the residues that are close to each other in protein complexes to study the protein-protein interactions. The homologous protein-protein cross-links are compatible with the well-developed protein analysis methods and sample preparation techniques for mass spectrometry. The data analysis has been streamlined with various bioinformatic tools and MS^n-based XL-MS has been developed to target the cross-links more precisely [71].

Comparing homologous protein-protein cross-links to heterogeneous DPCs, it is obvious that protein-protein cross-links are more readily detect by mass spectrometry and are compatible with the methods for analyzing proteins. Analyzing DPCs is more challenging since the methods of analyzing DNA and protein by mass spectrometry are quite contrary to each other. First, DNA prefers negative ion mode, but protein/peptides prefer positive ion mode. And DNA is a hydrophilic molecule, but protein/peptides are hydrophobic. Besides their intrinsic opposite chemical/physical profile, to enable mass spectrometry analysis of the complex, these large complexes need to be efficiently digested into small MS analyzable molecules, but the cross-linked sites possess high steric hindrance for both DNase and protease, producing incomplete digestion products, therefore, they will yield wide spectrum of products, increase the complexity of analytes, and reduce the signal to noise ratio. The MS2 fragment ions generated from DPCs contain the fragments from breaking the DNA bonds and protein/peptide bonds, therefore, the data analysis is more complicated than analyzing DNA or protein individually.

Besides the difficulty in analyzing the MS data, the sample preparation is also challenging, as has been explained previously that the contamination from free DNA and protein is hard to evade. The lack of a stringent method for DPC isolation is the key that prevents solving the DPCs related questions and assessing the biological relevance of different types of DPCs.

In the following chapter, I will show the method for isolating DNA-peptide cross-links that was developed by our lab. This method enabled the mass spectrometry identification of the cross-linked amino acid residues on protein, we also developed a software tool to streamline the data analysis step and to accelerate the discovery of cross-linking sites [72].

1.4 DNA interstrand crosslinks

DNA interstrand cross-links (ICLs) are one of the most deleterious DNA lesions which alter the DNA structure, prevent strand separation, and inhibit the enzyme from reading the genetic code. The repair process of ICLs is highly mutagenic prone since most of the ICLs can be bypassed by the translesion polymerases [73]. The ICLs in the cells could finally initiate apoptosis, causing cell death even with one ICL lesion in bacterial or yeast cells [74].

Generally, there are two repair pathways for ICLs, replication coupled and replication independent pathways [75]. They all go through an unhooking step, where the endonucleases incise both sides of the cross-linked site, producing two nicks [75]. The

resulting unhooked ICLs are smaller than pre-incised ICLs, and they can be bypassed by TLS DNA polymerases. After filling up the gaps with homologous recombination (HR), the newly synthesized strand will replace the damaged strand, and the replication will continue [76].

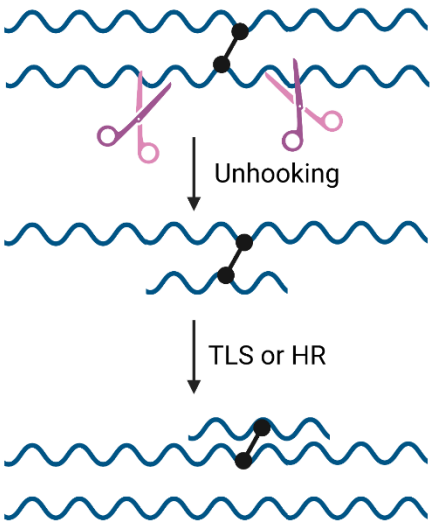


Figure 1.7 Repair process of ICLs. ICL repair processes are changing in cell cycles. Cells utilize different enzymes to unhook the ICL and form unhooked ICLs, later, the unhooked ICLs are either bypassed by TLS or recovered through HR [12].

The bifunctional alkylating agent was found to be used in cancer therapy and the underlying cytotoxicity comes from its ability to cross-link macromolecules and inhibit cell division [77]. Later, people found that nitrogen mustards could cross-link guanines to form ICLs [78] and nowadays many types of bifunctional cross-linking agents are invented as potential drugs to treat cancer cells [79]. In particular, some ICLs cause dramatic DNA unwinding and bending, which impact their recognition by repair enzymes [80,81], and as a result, influence the selection of repair pathways.

ICLs are produced by endogenous factors, especially AP sites [82], which are abundant and highly active, it has shown that AP sites form ICLs with guanine or adenine residues that are located opposite to AP sites on the complementary strand [28,78,83,84]. The ICLs generate in nearly every sequence context when AP sites location is adjacent to adenine [82]. The repair of the AP-ICLs involves the specific function of DNA glycosylase NEIL3 instead of other BER enzymes and it only unhooks the AP-ICLs where the AP site is on the leading strand of the replication fork [85]. There are a lot of questions regarding the AP-ICL repair process, for example, how the unhooking process of AP-ICL is performed in the cellular context, what effect the sequence has on the selection of repair pathways, and how the repair of AP-ICL differs from repairing other types of ICLs. Therefore, a flexible model of AP-ICL is required by the researchers to learn the repair processes.

Many methods are about synthesizing the site-specifically modified oligonucleotide that contains a structurally defined ICL, but this is a challenging task. Utilization of cross-linking agents could generate large amounts of cross-linking bonds, but there is no restriction on the cross-linking sites, so they don't have the site selectivity [86] and make the sample purification quite challenging. The solid-phase synthesis of the ICLs is likely to achieve site specific modification, but the chemistry of the reaction is more complicated than synthesizing the DNA single strand. Therefore, the product yield is much lower, and the process require synthetic expertise, so the applicability of this method is limited.

A more simplified and efficient method is to apply the nucleotide analogs to achieve both specific cross-linking and high synthesis efficiency. The Gates group developed many nucleotide analogs to form cross-links with various types of DNA lesions with high yields [87–90].

In my second thesis project, I utilized the Gates group finding that the AP site could form a cross-linking bond with 2'-aminopurine (P) to prepare our model ICLs, more details about this project are in Chapter 3.

1.5 Mitochondrial DNA repair

Mitochondria is a vital cellular component that controls energy production, cell signaling, and protein synthesis [91]. Besides the nucleus, mitochondria are the only subcellular component in animals that contain DNA. Unlike nuclear DNA, mitochondrial DNA (mtDNA) is a circular DNA of 16,569 base pairs (bp), which is much smaller than nuclear DNA, but mtDNA has thousands of copies and heteroplasmy [92]. mtDNA molecules are packed by proteins to form nucleoids [93]. The packaging proteins are mostly mitochondrial transcription factor A (TFAM) and mitochondrial single-stranded DNA-binding protein (mtSSB). Besides them, other proteins interact with mtDNA to enable gene expression and control the status of mitochondria.

Mitochondrial genomes contain genetic information for 13 proteins [94], which are the core components of the mitochondrial respiratory complexes, these protein complexes are majorly located in the inner membrane and create an electrochemical

gradient through the transfer of protons from the mitochondrial matrix across the membrane and oxidize the oxygen, and therefore mitochondria are negatively charged. The electrochemical potential of mitochondria is also a signal of mitochondrial status, which activates the repair or elimination of the damaged mitochondria [95]. mtDNA encodes 22 tRNAs and 2 rRNAs, making them have their protein synthesis system and to some extent independent from the nucleus [96].

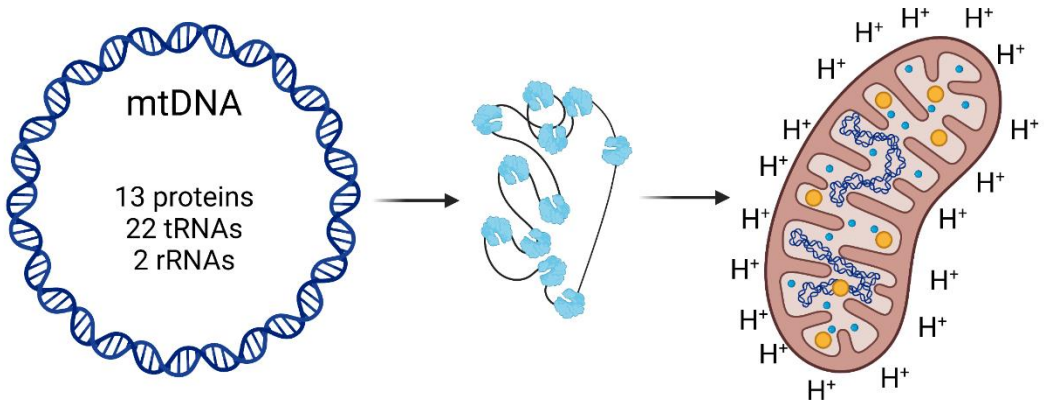


Figure 1.8 Mitochondria DNA is packed into nucleoids. mtDNA is a circular DNA that encodes 37 genes. Among them, it encodes 13 proteins which are the subunit of the respiratory complexes, 22 tRNAs, and 2 rRNAs. mtDNA is packaged by mitochondrial proteins into nucleoids. mtDNA is located inside the mitochondrial matrix, 1 or 2 mtDNA in each nucleoid.

mtDNA is constantly challenged by chemical and physical assaults, causing even higher levels of mtDNA damage than nuclear DNA (nDNA) because they are equipped with fewer repair mechanisms [97]. Due to their high copy number, mitochondria have some unique repair mechanisms, such as rapid mtDNA turnover, fission, and fusion [98], which are not possible to apply in the nuclear genome. mtDNA is more tolerant to DNA

damage than nDNA and can reduce the deleterious effect of DNA damage by fission and fusion.

Nearly all types of DNA lesions found in nDNA are in mtDNA, and the lesions on mtDNA are usually higher than nuclear DNA regardless of where the lesion comes from [97] because mitochondria are not equipped with such efficient repair mechanisms as the nucleus [99]. DNA lesions are the key reason for mitochondrial dysfunction [100]. And it is known to relate to many diseases, such as aging, diabetes, neurodegeneration, and cancer. Currently, there is no FDA-approved drug for treating mitochondrial disease, and the current gene modification tools that could efficiently modify nDNA, but they are hard to apply on mtDNA. There are many knowledge gaps in the mtDNA damage response and how the mtDNA repair and degradation contribute to maintaining the mitochondrial dynamics.

1.6 Scope of the dissertation

DNA is the control center of cellular metabolism. Changes in DNA, such as modifications and DNA lesions could alter the function of the cell, lead to various diseases and have profound influence on human health. There are many tools to identify and quantify DNA lesions, including chemical probes and enzyme involved DNA sequencing techniques. And with the advancement of mass spectrometry instrumentation and methods, the analysis of DNA lesions can couple with mass spectrometry technique to further deepen the discovery and get more detailed information about the cross-linking

chemistry. This dissertation focus on developing methods and tools to study the AP cross-linked products.

In Chapter 2, we designed a cleavable DNA strand and an enrichment method for DNA-peptide cross-links and developed a software tool for analyzing the mass spectrometry data of DPCs. We successfully applied this system to discover the cross-linking sites on TFAM that cross-linked with the AP site containing DNA. Our finding is consistent with the available X-ray crystallography data. Our result led to the discovery of 14 cross-linked sites [72]. This research also promoted a few later research in our group and the discovery of GSH effects on AP sites in mitochondria.

In Chapter 3, we utilized cleavable DNA and the nucleotide analog to simplify the synthesis of the unhooked ICL repair intermediate. The yield of the final product was high and applied to studying the AP-ICL repair process [101].

In Chapter 4, we investigated the strand break products during the DPC formation. We used high-resolution DNA sequencing PAGE to resolve different strand breaks to know how various conditions impact the distribution of strand breaks. And we also tried to learn how various strand breaks react with repair enzymes and the chronic steps for the repair, which are fundamental to understanding the mechanism of repairing AP sites.

In Chapter 5, we talk about the future of these studies and the directions to expand them into a more widely applicable research context.

1.7 References

[1] A. Claiborne, H. Miller, D. Parsonage, R.P. Ross, Protein-sulfenic acid stabilization and function in enzyme catalysis and gene regulation, FASEB J. 7 (1993) 1483–1490. https://doi.org/10.1096/fasebj.7.15.8262333.

[2] F. Weissmann, G. Petzold, R. VanderLinden, P.J. Huis in 't Veld, N.G. Brown, F. Lampert, S. Westermann, H. Stark, B.A. Schulman, J.-M. Peters, biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes, Proc. Natl. Acad. Sci. 113 (2016). https://doi.org/10.1073/pnas.1604935113.

[3] T. Hajri, N.A. Abumrad, FATTY ACID TRANSPORT ACROSS MEMBRANES: Relevance to Nutrition and Metabolic Pathology, Annu. Rev. Nutr. 22 (2002) 383–415. https://doi.org/10.1146/annurev.nutr.22.020402.130846.

[4] L. Moldovan, A. Barnard, C.-H. Gil, Y. Lin, M.B. Grant, M.C. Yoder, N. Prasain, N.I. Moldovan, iPSC-Derived Vascular Cell Spheroids as Building Blocks for Scaffold-Free Biofabrication, Biotechnol. J. 12 (2017) 1700444. https://doi.org/10.1002/biot.201700444.

[5] N. Ruiz, D. Kahne, T.J. Silhavy, Advances in understanding bacterial outer-membrane biogenesis, Nat. Rev. Microbiol. 4 (2006) 57–66. https://doi.org/10.1038/nrmicro1322.

[6] K. Rothkamm, S. Barnard, J. Moquet, M. Ellender, Z. Rana, S. Burdak-Rothkamm, DNA damage foci: Meaning and significance, Environ. Mol. Mutagen. 56 (2015) 491–504. https://doi.org/10.1002/em.21944.

[7] W.P. Roos, B. Kaina, DNA damage-induced cell death: From specific DNA lesions to the DNA damage response and apoptosis, Cancer Lett. 332 (2013) 237–248. https://doi.org/10.1016/j.canlet.2012.01.007.

[8] B. Kramer, W. Kramer, H.-J. Fritz, Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli, Cell. 38 (1984) 879–887. https://doi.org/10.1016/0092-8674(84)90283-6.

[9] A. Kuzminov, Single-strand interruptions in replicating chromosomes cause double-strand breaks, Proc. Natl. Acad. Sci. 98 (2001) 8241–8246. https://doi.org/10.1073/pnas.131009198.

[10] H. Ide, K. Akamatsu, Y. Kimura, K. Michiue, K. Makino, A. Asaeda, Y. Takamori, K. Kubo, Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA, Biochemistry. 32 (1993) 8276–8283. https://doi.org/10.1021/bi00083a031.

[11] R.K. Zahn, G. Zahn-Daimler, S. Ax, M. Hosokawa, T. Takeda, Assessment of DNA-protein crosslinks in the course of aging in two mouse strains by use of a modified alkaline filter elution applied to whole tissue samples, Mech. Ageing Dev. 108 (1999) 99–112. https://doi.org/10.1016/S0047-6374(98)00151-1.

[12] A.J. Deans, S.C. West, DNA interstrand crosslink repair and cancer, Nat. Rev. Cancer. 11 (2011) 467–480. https://doi.org/10.1038/nrc3088.

[13] R. Singal, G.D. Ginder, DNA Methylation, Blood. 93 (1999) 4059–4070. https://doi.org/10.1182/blood.V93.12.4059.

[14] D.Y. Lee, J.J. Hayes, D. Pruss, A.P. Wolffe, A positive role for histone acetylation in transcription factor access to nucleosomal DNA, Cell. 72 (1993) 73–84. https://doi.org/10.1016/0092-8674(93)90051-Q.

[15] E.C. Friedberg, L.D. McDaniel, R.A. Schultz, The role of endogenous and exogenous DNA damage and mutagenesis, Curr. Opin. Genet. Dev. 14 (2004) 5–10. https://doi.org/10.1016/j.gde.2003.11.001.

[16] T. Lindahl, R.D. Wood, Quality Control by DNA Repair, Science (80-.). 286 (1999) 1897–1905. https://doi.org/10.1126/science.286.5446.1897.

[17] L.H. Pearl, A.C. Schierz, S.E. Ward, B. Al-Lazikani, F.M.G. Pearl, Therapeutic opportunities within the DNA damage response, Nat. Rev. Cancer. 15 (2015) 166–180. https://doi.org/10.1038/nrc3891.

[18] P.W. da S. dos Santos, A.R.T. Machado, R.A. De Grandis, D.L. Ribeiro, K. Tuttis, M. Morselli, A.F. Aissa, M. Pellegrini, L.M.G. Antunes, Transcriptome and DNA methylation changes modulated by sulforaphane induce cell cycle arrest, apoptosis, DNA damage, and suppression of proliferation in human liver cancer cells, Food Chem. Toxicol. 136 (2020) 111047. https://doi.org/10.1016/j.fct.2019.111047.

[19] C. Yauk, A. Polyzos, A. Rowan-Carroll, C.M. Somers, R.W. Godschalk, F.J. Van Schooten, M.L. Berndt, I.P. Pogribny, I. Koturbash, A. Williams, G.R. Douglas, O. Kovalchuk, Germ-line mutations, DNA damage, and global hypermethylation in

mice exposed to particulate air pollution in an urban/industrial location, Proc. Natl. Acad. Sci. 105 (2008) 605–610. https://doi.org/10.1073/pnas.0705896105.

[20] World Health Organization, The public health impact of chemicals: knowns and unknowns: data addendum for 2019, World Heal. Organ. (2021).

[21] T. Lindahl, B. Nyberg, Rate of depurination of native deoxyribonucleic acid, Biochemistry. 11 (1972) 3610–3618. https://doi.org/10.1021/bi00769a018.

[22] U. Hardeland, M. Bentele, T. Lettieri, R. Steinacher, J. Jiricny, P. Schär, Thymine DNA glycosylase, in: 2001: pp. 235–253. https://doi.org/10.1016/S0079-6603(01)68103-0.

[23] F. Hutchinson, Chemical Changes Induced in DNA by Ionizing Radiation, in: 1985: pp. 115–154. https://doi.org/10.1016/S0079-6603(08)60347-5.

[24] E.E. Drouin, E.L. Loechler, AP sites are not significantly involved in mutagenesis by the (+)-anti diol epoxide of benzo[a]pyrene: The complexicity of its mutagenic specificity is likely to arise from adduct conformational polymorphism, Biochemistry. 32 (1993) 6555–6562. https://doi.org/10.1021/bi00077a009.

[25] J.S. Jha, J. Yin, T. Haldar, Z. Yang, Y. Wang, K.S. Gates, Reconsidering the Chemical Nature of Strand Breaks Derived from Abasic Sites in Cellular DNA: Evidence for 3′-Glutathionylation, J. Am. Chem. Soc. 144 (2022) 10471–10482. https://doi.org/10.1021/jacs.2c02703.

[26] Z. Yang, M.I. Nejad, J.G. Varela, N.E. Price, Y. Wang, K.S. Gates, A role for the base excision repair enzyme NEIL3 in replication-dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites, DNA Repair (Amst). 52 (2017) 1–11. https://doi.org/10.1016/j.dnarep.2017.02.011.

[27] S. Boiteux, M. Guillet, Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae, DNA Repair (Amst). 3 (2004) 1–12. https://doi.org/10.1016/j.dnarep.2003.10.002.

[28] K.M. Johnson, N.E. Price, J. Wang, M.I. Fekry, S. Dutta, D.R. Seiner, Y. Wang, K.S. Gates, On the Formation and Properties of Interstrand DNA–DNA Cross-Links Forged by Reaction of an Abasic Site with the Opposing Guanine Residue of 5′-CAp Sequences in Duplex DNA, J. Am. Chem. Soc. 135 (2013) 1015–1025. https://doi.org/10.1021/ja308119q.

[29] Y. Wang, L. Liu, C. Wu, A. Bulgar, E. Somoza, W. Zhu, S.L. Gerson, Direct detection and quantification of abasic sites for in vivo studies of DNA damage and repair, Nucl. Med. Biol. 36 (2009) 975–983. https://doi.org/10.1016/j.nucmedbio.2009.07.007.

[30] Z.J. Liu, S. Martínez Cuesta, P. van Delft, S. Balasubramanian, Sequencing abasic sites in DNA at single-nucleotide resolution, Nat. Chem. 11 (2019) 629–637. https://doi.org/10.1038/s41557-019-0279-9.

[31] M. Hashimoto, M.M. Greenberg, Y.W. Kow, J.-T. Hwang, R.P. Cunningham, The 2-Deoxyribonolactone Lesion Produced in DNA by Neocarzinostatin and Other Damaging Agents Forms Cross-links with the Base-Excision Repair Enzyme Endonuclease III, J. Am. Chem. Soc. 123 (2001) 3161–3162. https://doi.org/10.1021/ja003354z.

[32] V. Faure, 2'-Deoxyribonolactone lesion produces G→A transitions in Escherichia coli, Nucleic Acids Res. 32 (2004) 2937–2946. https://doi.org/10.1093/nar/gkh622.

[33] K.M. Kroeger, J. Kim, M.F. Goodman, M.M. Greenberg, Effects of the C4'-Oxidized Abasic Site on Replication in Escherichia coli. An Unusually Large Deletion Is Induced by a Small Lesion, Biochemistry. 43 (2004) 13621–13627. https://doi.org/10.1021/bi048337r.

[34] L. Guan, K. Bebenek, T.A. Kunkel, M.M. Greenberg, Inhibition of Short Patch and Long Patch Base Excision Repair by an Oxidized Abasic Site, Biochemistry. 49 (2010) 9904–9910. https://doi.org/10.1021/bi101533a.

[35] N. Kojima, T. Takebayashi, A. Mikami, E. Ohtsuka, Y. Komatsu, Construction of Highly Reactive Probes for Abasic Site Detection by Introduction of an Aromatic and a Guanidine Residue into an Aminooxy Group, J. Am. Chem. Soc. 131 (2009) 13208–13209. https://doi.org/10.1021/ja904767k.

[36] P. Guo, S. Yan, J. Hu, X. Xing, C. Wang, X. Xu, X. Qiu, W. Ma, C. Lu, X. Weng, X. Zhou, Selective Detection of 5-Formyl-2'-deoxycytidine in DNA Using a Fluorogenic Hydroxylamine Reagent, Org. Lett. 15 (2013) 3266–3269. https://doi.org/10.1021/ol401290d.

[37] E.-A. Raiber, D. Beraldi, G. Ficz, H.E. Burgess, M.R. Branco, P. Murat, D. Oxley, M.J. Booth, W. Reik, S. Balasubramanian, Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase, Genome Biol. 13 (2012) R69. https://doi.org/10.1186/gb-2012-13-8-r69.

[38] R.E. Hardisty, F. Kawasaki, A.B. Sahakyan, S. Balasubramanian, Selective Chemical Labeling of Natural T Modifications in DNA, J. Am. Chem. Soc. 137 (2015) 9270–9272. https://doi.org/10.1021/jacs.5b03730.

[39] H.-Y. Zhang, J. Xiong, B.-L. Qi, Y.-Q. Feng, B.-F. Yuan, The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals, Chem. Commun. 52 (2016) 737–740. https://doi.org/10.1039/C5CC07354E.

[40] Y. Wang, X. Zhang, G. Zou, S. Peng, C. Liu, X. Zhou, Detection and Application of 5-Formylcytosine and 5-Formyluracil in DNA, Acc. Chem. Res. 52 (2019) 1016–1024. https://doi.org/10.1021/acs.accounts.8b00543.

[41] S. Wei, S. Shalhout, Y.-H. Ahn, A.S. Bhagwat, A versatile new tool to quantify abasic sites in DNA and inhibit base excision repair, DNA Repair (Amst). 27 (2015) 9–18. https://doi.org/10.1016/j.dnarep.2014.12.006.

[42] A.G. Condie, Y. Yan, S.L. Gerson, Y. Wang, A Fluorescent Probe to Measure DNA Damage and Repair, PLoS One. 10 (2015) e0131330. https://doi.org/10.1371/journal.pone.0131330.

[43] Y. Wang, Y. Hu, T. Wu, L. Zhang, H. Liu, X. Zhou, Y. Shao, Recognition of DNA abasic site nanocavity by fluorophore-switched probe: Suitable for all sequence environments, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 153 (2016) 645–650. https://doi.org/10.1016/j.saa.2015.09.038.

[44] W. Wu, Y. Wang, Y. Zhou, Y. Shao, L. Zhang, H. Liu, Selective fluorescence lighting-up recognition of DNA abasic site environment possessing guanine context, Sensors Actuators B Chem. 206 (2015) 449–455. https://doi.org/10.1016/j.snb.2014.09.090.

[45] C.D. Mol, T. Izumi, S. Mitra, J.A. Tainer, DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination, Nature. 403 (2000) 451–456. https://doi.org/10.1038/35000249.

[46] D.M. Wilson, M. Takeshita, A.P. Grollman, B. Demple, Incision Activity of Human Apurinic Endonuclease (Ape) at Abasic Site Analogs in DNA, J. Biol. Chem. 270 (1995) 16002–16007. https://doi.org/10.1074/jbc.270.27.16002.

[47] Y. Cai, H. Cao, F. Wang, Y. Zhang, P. Kapranov, Complex genomic patterns of abasic sites in mammalian DNA revealed by a high-resolution SSiNGLe-AP method, Nat. Commun. 13 (2022) 5868. https://doi.org/10.1038/s41467-022-33594-1.

[48] M. Li, J. Völker, K.J. Breslauer, D.M. Wilson, APE1 Incision Activity at Abasic Sites in Tandem Repeat Sequences, J. Mol. Biol. 426 (2014) 2183–2198. https://doi.org/10.1016/j.jmb.2014.03.014.

[49] C. Broxson, J.N. Hayner, J. Beckett, L.B. Bloom, S. Tornaletti, Human AP endonuclease inefficiently removes abasic sites within G4 structures compared to duplex DNA, Nucleic Acids Res. 42 (2014) 7708–7719. https://doi.org/10.1093/nar/gku417.

[50] S. Burra, D. Marasco, M.C. Malfatti, G. Antoniali, A. Virgilio, V. Esposito, B. Demple, A. Galeone, G. Tell, Human AP-endonuclease (Ape1) activity on telomeric G4 structures is modulated by acetylatable lysine residues in the N-terminal sequence, DNA Repair (Amst). 73 (2019) 129–143. https://doi.org/10.1016/j.dnarep.2018.11.010.

[51] A.S. Miller, L. Balakrishnan, N.A. Buncher, P.L. Opresko, R.A. Bambara, Telomere proteins POT1, TRF1 and TRF2 augment long-patch base excision repair in vitro, Cell Cycle. 11 (2012) 998–1007. https://doi.org/10.4161/cc.11.5.19483.

[52] B. Cao, X. Wu, J. Zhou, H. Wu, L. Liu, Q. Zhang, M.S. DeMott, C. Gu, L. Wang, D. You, P.C. Dedon, Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage, Nucleic Acids Res. 48 (2020) 6715–6725. https://doi.org/10.1093/nar/gkaa473.

[53] T. Jenuwein, C.D. Allis, Translating the Histone Code, Science (80-.). 293 (2001) 1074–1080. https://doi.org/10.1126/science.1063127.

[54] M.J. Solomon, P.L. Larsen, A. Varshavsky, Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene, Cell. 53 (1988) 937–947. https://doi.org/10.1016/S0092-8674(88)90469-2.

[55] A. Steube, T. Schenk, A. Tretyakov, H.P. Saluz, High-intensity UV laser ChIP-seq for the study of protein-DNA interactions in living cells, Nat. Commun. 8 (2017) 1303. https://doi.org/10.1038/s41467-017-01251-7.

[56] K. Lu, W. Ye, L. Zhou, L.B. Collins, X. Chen, A. Gold, L.M. Ball, J.A. Swenberg, Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino

Acids and Deoxynucleosides and Their Oligomers, J. Am. Chem. Soc. 132 (2010) 3388–3399. https://doi.org/10.1021/ja908282f.

[57] N.Y. Tretyakova, A. Groehler, S. Ji, DNA–Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes, Acc. Chem. Res. 48 (2015) 1631–1644. https://doi.org/10.1021/acs.accounts.5b00056.

[58] J.E. Yeo, S. Wickramaratne, S. Khatwani, Y.-C. Wang, J. Vervacke, M.D. Distefano, N.Y. Tretyakova, Synthesis of Site-Specific DNA–Protein Conjugates and Their Effects on DNA Replication, ACS Chem. Biol. 9 (2014) 1860–1868. https://doi.org/10.1021/cb5001795.

[59] R.L. Loeber, E.D. Michaelson-Richie, S.G. Codreanu, D.C. Liebler, C.R. Campbell, N.Y. Tretyakova, Proteomic Analysis of DNA−Protein Cross-Linking by Antitumor Nitrogen Mustards, Chem. Res. Toxicol. 22 (2009) 1151–1162. https://doi.org/10.1021/tx900078y.

[60] X. Ming, A. Groehler, E.D. Michaelson-Richie, P.W. Villalta, C. Campbell, N.Y. Tretyakova, Mass Spectrometry Based Proteomics Study of Cisplatin-Induced DNA–Protein Cross-Linking in Human Fibrosarcoma (HT1080) Cells, Chem. Res. Toxicol. 30 (2017) 980–995. https://doi.org/10.1021/acs.chemrestox.6b00389.

[61] J. Stingele, R. Bellelli, S.J. Boulton, Mechanisms of DNA–protein crosslink repair, Nat. Rev. Mol. Cell Biol. 18 (2017) 563–573. https://doi.org/10.1038/nrm.2017.56.

[62] P.S. Thompson, K.M. Amidon, K.N. Mohni, D. Cortez, B.F. Eichman, Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link, Nat. Struct. Mol. Biol. 26 (2019) 613–618. https://doi.org/10.1038/s41594-019-0255-5.

[63] K.N. Mohni, S.R. Wessel, R. Zhao, A.C. Wojciechowski, J.W. Luzwick, H. Layden, B.F. Eichman, P.S. Thompson, K.P.M. Mehta, D. Cortez, HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA, Cell. 176 (2019) 144-153.e13. https://doi.org/10.1016/j.cell.2018.10.055.

[64] D.R. Semlow, V.A. MacKrell, J.C. Walter, The HMCES DNA-protein cross-link functions as an intermediate in DNA interstrand cross-link repair, Nat. Struct. Mol. Biol. 29 (2022) 451–462. https://doi.org/10.1038/s41594-022-00764-0.

[65] J.T. Sczepanski, R.S. Wong, J.N. McKnight, G.D. Bowman, M.M. Greenberg, Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome

core particle, Proc. Natl. Acad. Sci. 107 (2010) 22475–22480. https://doi.org/10.1073/pnas.1012860108.

[66] W. Xu, R.M. Boyd, M.O. Tree, F. Samkari, L. Zhao, Mitochondrial transcription factor A promotes DNA strand cleavage at abasic sites, Proc. Natl. Acad. Sci. 116 (2019) 17792–17799. https://doi.org/10.1073/pnas.1911252116.

[67] K. Woźniak, J. Blasiak, In vitro genotoxicity of lead acetate: induction of single and double DNA strand breaks and DNA–protein cross-links, Mutat. Res. Toxicol. Environ. Mutagen. 535 (2003) 127–139. https://doi.org/10.1016/S1383-5718(02)00295-4.

[68] L. Distel, B. Distel, H. Schüssler, Formation of DNA double-strand breaks and DNA–protein crosslinks by irradiation of DNA in the presence of a protein, Radiat. Phys. Chem. 65 (2002) 141–149. https://doi.org/10.1016/S0969-806X(02)00215-3.

[69] A. Zhitkovich, M. Costa, A simple, sensitive assay to detect DNA–protein cromlinks in intact cells and in vivo, Carcinogenesis. 13 (1992) 1485–1489. https://doi.org/10.1093/carcin/13.8.1485.

[70] S. Barker, D. Murray, J. Zheng, L. Li, M. Weinfeld, A method for the isolation of covalent DNA–protein crosslinks suitable for proteomics analysis, Anal. Biochem. 344 (2005) 204–215. https://doi.org/10.1016/j.ab.2005.06.039.

[71] C. Yu, L. Huang, Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology, Anal. Chem. 90 (2018) 144–165. https://doi.org/10.1021/acs.analchem.7b04431.

[72] J. Tang, W. Zhao, N.G. Hendricks, L. Zhao, High-Resolution Mapping of Amino Acid Residues in DNA–Protein Cross-Links Enabled by Ribonucleotide-Containing DNA, Anal. Chem. 93 (2021) 13398–13406. https://doi.org/10.1021/acs.analchem.1c03481.

[73] U. Roy, O.D. Schärer, Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair, DNA Repair (Amst). 44 (2016) 33–41. https://doi.org/10.1016/j.dnarep.2016.05.004.

[74] J. Grillari, H. Katinger, R. Voglauer, Contributions of DNA interstrand cross-links to aging of cells and organisms, Nucleic Acids Res. 35 (2007) 7566–7576. https://doi.org/10.1093/nar/gkm1065.

[75] H.L. Williams, M.E. Gottesman, J. Gautier, The differences between ICL repair during and outside of S phase, Trends Biochem. Sci. 38 (2013) 386–393. https://doi.org/10.1016/j.tibs.2013.05.004.

[76] J. Michl, J. Zimmer, M. Tarsounas, Interplay between Fanconi anemia and homologous recombination pathways in genome integrity, EMBO J. 35 (2016) 909–923. https://doi.org/10.15252/embj.201693860.

[77] J.G. Wagner, T.W. Petry, R.A. Roth, Characterization of monocrotaline pyrrole-induced DNA cross-linking in pulmonary artery endothelium, Am. J. Physiol. Cell. Mol. Physiol. 264 (1993) L517–L522. https://doi.org/10.1152/ajplung.1993.264.5.L517.

[78] S. Dutta, G. Chowdhury, K.S. Gates, Interstrand Cross-Links Generated by Abasic Sites in Duplex DNA, J. Am. Chem. Soc. 129 (2007) 1852–1853. https://doi.org/10.1021/ja067294u.

[79] T.-L. Su, T.-C. Lee, R. Kakadiya, The development of bis(hydroxymethyl)pyrrole analogs as bifunctional DNA cross-linking agents and their chemotherapeutic potential, Eur. J. Med. Chem. 69 (2013) 609–621. https://doi.org/10.1016/j.ejmech.2013.09.016.

[80] R.J. Legerski, Repair of DNA interstrand cross-links during S phase of the mammalian cell cycle, Environ. Mol. Mutagen. (2010) NA-NA. https://doi.org/10.1002/em.20566.

[81] D.R. Semlow, J.C. Walter, Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair, Annu. Rev. Biochem. 90 (2021) 107–135. https://doi.org/10.1146/annurev-biochem-080320-112510.

[82] A. Huskova, B. Landova, E. Boura, J. Silhan, The rate of formation and stability of abasic site interstrand crosslinks in the DNA duplex, DNA Repair (Amst). 113 (2022) 103300. https://doi.org/10.1016/j.dnarep.2022.103300.

[83] N.E. Price, K.M. Johnson, J. Wang, M.I. Fekry, Y. Wang, K.S. Gates, Interstrand DNA–DNA Cross-Link Formation Between Adenine Residues and Abasic Sites in Duplex DNA, J. Am. Chem. Soc. 136 (2014) 3483–3490. https://doi.org/10.1021/ja410969x.

[84] M.I. Nejad, R. Shi, X. Zhang, L.-Q. Gu, K.S. Gates, Sequence-Specific Covalent Capture Coupled with High-Contrast Nanopore Detection of a Disease-Derived

Nucleic Acid Sequence, ChemBioChem. 18 (2017) 1383–1386. https://doi.org/10.1002/cbic.201700204.

[85] M. Imani Nejad, K. Housh, A.A. Rodriguez, T. Haldar, S. Kathe, S.S. Wallace, B.F. Eichman, K.S. Gates, Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3, DNA Repair (Amst). 86 (2020) 102752. https://doi.org/10.1016/j.dnarep.2019.102752.

[86] M.I. Nejad, K.M. Johnson, N.E. Price, K.S. Gates, A New Cross-Link for an Old Cross-Linking Drug: The Nitrogen Mustard Anticancer Agent Mechlorethamine Generates Cross-Links Derived from Abasic Sites in Addition to the Expected Drug-Bridged Cross-Links, Biochemistry. 55 (2016) 7033–7041. https://doi.org/10.1021/acs.biochem.6b01080.

[87] A.H. Kellum, D.Y. Qiu, M.W. Voehler, W. Martin, K.S. Gates, M.P. Stone, Structure of a Stable Interstrand DNA Cross-Link Involving a β- N -Glycosyl Linkage Between an N 6 -dA Amino Group and an Abasic Site, Biochemistry. 60 (2021) 41–52. https://doi.org/10.1021/acs.biochem.0c00596.

[88] N.E. Price, M.J. Catalano, S. Liu, Y. Wang, K.S. Gates, Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA, Nucleic Acids Res. 43 (2015) 3434–3441. https://doi.org/10.1093/nar/gkv174.

[89] J.G. Varela, L.E. Pierce, X. Guo, N.E. Price, K.M. Johnson, Z. Yang, Y. Wang, K.S. Gates, Interstrand Cross-Link Formation Involving Reaction of a Mispaired Cytosine Residue with an Abasic Site in Duplex DNA, Chem. Res. Toxicol. 34 (2021) 1124–1132. https://doi.org/10.1021/acs.chemrestox.1c00004.

[90] M.J. Catalano, S. Liu, N. Andersen, Z. Yang, K.M. Johnson, N.E. Price, Y. Wang, K.S. Gates, Chemical Structure and Properties of Interstrand Cross-Links Formed by Reaction of Guanine Residues with Abasic Sites in Duplex DNA, J. Am. Chem. Soc. 137 (2015) 3933–3945. https://doi.org/10.1021/jacs.5b00669.

[91] S.W.G. Tait, D.R. Green, Mitochondria and cell signalling, J. Cell Sci. 125 (2012) 807–815. https://doi.org/10.1242/jcs.099234.

[92] J.B. Stewart, P.F. Chinnery, The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease, Nat. Rev. Genet. 16 (2015) 530–542. https://doi.org/10.1038/nrg3966.

[93] D.F. Bogenhagen, Mitochondrial DNA nucleoid structure, Biochim. Biophys. Acta - Gene Regul. Mech. 1819 (2012) 914–920. https://doi.org/10.1016/j.bbagrm.2011.11.005.

[94] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (1999) 1767–1780. https://doi.org/10.1093/nar/27.8.1767.

[95] N. Kamo, M. Muratsugu, R. Hongoh, Y. Kobatake, Membrane potential of mitochondria measured with an electrode sensitive to tetraphenyl phosphonium and relationship between proton electrochemical potential and phosphorylation potential in steady state, J. Membr. Biol. 49 (1979) 105–121. https://doi.org/10.1007/BF01868720.

[96] M.I.G. Lopez Sanchez, T.R. Mercer, S.M.K. Davies, A.-M.J. Shearwood, K.K.A. Nygård, T.R. Richman, J.S. Mattick, O. Rackham, A. Filipovska, RNA processing in human mitochondria, Cell Cycle. 10 (2011) 2904–2916. https://doi.org/10.4161/cc.10.17.17060.

[97] L. Zhao, P. Sumberaz, Mitochondrial DNA Damage: Prevalence, Biological Consequence, and Emerging Pathways, Chem. Res. Toxicol. 33 (2020) 2491–2502. https://doi.org/10.1021/acs.chemrestox.0c00083.

[98] L.L. Clay Montier, J.J. Deng, Y. Bai, Number matters: control of mammalian mitochondrial DNA copy number, J. Genet. Genomics. 36 (2009) 125–131. https://doi.org/10.1016/S1673-8527(08)60099-5.

[99] M. Alexeyev, I. Shokolenko, G. Wilson, S. LeDoux, The Maintenance of Mitochondrial DNA Integrity--Critical Analysis and Update, Cold Spring Harb. Perspect. Biol. 5 (2013) a012641–a012641. https://doi.org/10.1101/cshperspect.a012641.

[100] J. Hayashi, S. Ohta, A. Kikuchi, M. Takemitsu, Y. Goto, I. Nonaka, Introduction of disease-related mitochondrial DNA deletions into HeLa cells lacking mitochondrial DNA results in mitochondrial dysfunction., Proc. Natl. Acad. Sci. 88 (1991) 10614–10618. https://doi.org/10.1073/pnas.88.23.10614.

[101] J. Tang, F. Tang, L. Zhao, Facile preparation of model DNA interstrand cross-link repair intermediates using ribonucleotide-containing DNA, DNA Repair (Amst). 111 (2022) 103286. https://doi.org/10.1016/j.dnarep.2022.103286.

Chapter 2. High-Resolution Mapping of Amino Acid Residues in DNA-Protein Cross-Links Enabled by Ribonucleotide-Containing DNA

2.1 Abstract

DNA−protein cross-links have broad applications in mapping DNA−protein interactions and provide structural insights into macromolecular structures. However, high-resolution mapping of DNA-interacting amino acid residues with tandem mass spectrometry remains challenging due to difficulties in sample preparation and data analysis. Herein, we developed a method for identifying cross-linking amino residues in DNA−protein cross-links at single amino acid resolution. We leveraged the alkaline lability of ribonucleotides and designed ribonucleotide-containing DNA to produce structurally defined nucleic acid−peptide crosslinks under our optimized ribonucleotide cleavage conditions. The structurally defined oligonucleotide−peptide heteroconjugates improved ionization, reduced the database search space, and facilitated the identification of cross-linking residues in peptides. We applied the workflow to identifying abasic (AP) site-interacting residues in human mitochondrial transcription factor A (TFAM)-DNA cross-links. With sub-nmol sample input, we obtained high quality fragmentation spectra for nucleic acid−peptide cross-links and identified 14 cross-linked lysine residues with the home-built AP_CrosslinkFinder program. Semi-quantification based on integrated peak areas revealed that K186 of TFAM is the major cross-linking residue, consistent with K186

being the closest (to the AP modification) lysine residue in solved TFAM:DNA crystal structures. Additional cross-linking lysine residues (K69, K76, K136, K154) support the dynamic characteristics of TFAM:DNA complexes. Overall, our combined workflow using ribonucleotide as a chemically cleavable DNA modification together with optimized sample preparation and data analysis offers a simple yet powerful approach for mapping cross-linking sites in DNA−protein cross-links. The method is amendable to other chemical or photo-cross-linking systems and can be extended to complex biological samples.

2.2 Introduction

DNA-protein interactions are essential to all aspects of genetic information transfer in living organisms, such as replication, transcription, recombination, and repair [1]. Covalent DNA-protein cross-links (DPCs) mediated by chemical linkers serve as useful tools to map these interactions [2–4]. DPCs also form in biological contexts as intermediates during enzymatic processing of DNA or as products when genetic materials are under chemical/physical assault [1,5,6]. In particular, a class of DPCs has been shown to form between a prevalent DNA modification, abasic (AP) sites [7], and lysine residues on interacting proteins via Schiff base chemistry.[8] DPCs derived from AP or oxidized AP sites have been demonstrated with DNA repair proteins [9,10], nucleosome core particles [11,12], and a transcription factor in mitochondria [13]. If DPCs form excessively [14] or are not removed promptly [15], they can block many DNA transactions [16,17]. Therefore, DPCs play a critical role in genomic maintenance and human health.

Identifying DNA-interacting amino acid residues is of fundamental importance to understanding the basis of DNA-protein interactions. Structural biology approaches, such as X-ray crystallography [18] and cryo-EM [19], can generate high-resolution data; however, they often require large amounts of purified samples and offer limited information on the dynamics of macromolecules or complexes. Solution NMR is powerful in characterizing the structure and dynamics of biomolecules but is often limited to analytes with low molecular weights (<35 kDa) [20]. Over the past decade, cross-linking mass spectrometry (MS) has emerged as a powerful tool in structural biology and interactome research [21]. The method can deliver medium-resolution information to complement classical structural biology and computational approaches. Recently, methods to study DNA-protein interactions in reconstituted systems and at a proteome level have also been developed [22,23].

Despite the advancement in cross-linking methodologies and mass spectrometry instrumentation, analyzing nucleic acid-protein conjugates remains challenging. The challenge is confounded by several factors in sample preparation and data analysis. First, both the protein and the oligonucleotide need to be digested into short fragments in shotgun proteomics. The complete digestion of the oligonucleotide is not trivial due to the steric hindrance imposed by oligonucleotide-peptide cross-links [24]. Commonly used digestion methods involving nuclease cocktails tend to produce a mixture of mono-, di-, and oligonucleotides and consequently structural heterogeneity of oligonucleotide-peptide conjugates [24]. Second, data analysis remains labor-intensive due to the low

signal intensity of cross-links and their structural heterogeneity. Consequently, uncertainty in peptide search and difficulties in identifying corresponding cross-links are unavoidable [25].

In this study, we developed a mass spectrometry-based approach to map DNA-protein interacting residues at single amino acid resolution. Our approach exploits the reactivity of AP sites with lysine residues and the lability of ribonucleotides under alkaline conditions to generate structurally defined DNA-peptide cross-links. The ribonucleotides in DNA substrates facilitate the preparation of DNA-peptide cross-links with a predictable nucleic acid fragment via our optimized conditions, which we refer to as Cleave R reaction. Compared to nuclease digestion, our method avoids digestion bias and facilitates data analysis and interpretation. We combined the optimized sample preparation with data analysis using our developed AP_CrosslinkFinder program. The optimized workflow was applied to mapping interacting lysine residues of human mitochondrial transcription factor A (TFAM) with AP modification on DNA. We successfully identified 14 cross-linked lysine residues, which provide insights into the interacting (with AP sites) residues on TFAM and the dynamic characteristics of TFAM:DNA complexes.

2.3 Experimental section

2.3.1 Materials and methods

Chemicals were from Sigma Aldrich or Fisher Scientific and were analytical grade or molecular biology grade. MS grade trypsin was purchased from Fisher Scientific. Uracil

DNA glycosylase (UDG) was purchased from New England Biolabs. ODNs were purchased from Integrated DNA Technologies. The AP-containing DNA oligomers was prepared following a reported procedure [13]. Briefly, a deoxyuridine-containing DNA oligomer was treated with UDG to convert deoxyuridine to an AP modification. The AP-containing DNA oligomer was purified via phenol/chloroform extraction followed by annealing with a complementary ODN with no base modifications. The recombinant human transcription factor A (TFAM 43-246) and human AP endonuclease 1 (APE1) were expressed and purified based on our previous protocol [13].

2.3.2 Electrophoretic mobility shift assay

The non-denaturing gels were composed of 6% polyacrylamide (acrylamide/bis-acrylamide, 64/1) in 0.35X TBE (Tris-Borate-EDTA) buffer and pre-run for 30 min at 4°C. The TFAM:DNA complex was assembled on ice, equilibrated at room temperature for 1 h and separated in the native gel at 4°C at 100 V. Gel was imaged with a Typhoon imager (GE Healthcare) and quantified using ImageQuant software. The data fitting was performed with GraphPad Prism v8.0.

2.3.3 Preparation of TFAM-DNA cross-links

A dsDNA substrate containing an AP lesion was incubated with recombinant human TFAM for 12 h at 37°C. Reactions contained 4 µM AP-DNA, 8 µM TFAM, 20 mM HEPES (pH 7.4), 90 mM NaCl, and 20 mM EDTA, with 25 mM $NaBH_3CN$. The reaction was quenched by adding 100 mM $NaBH_4$ followed by incubation on ice for 30 min. The yield

of DPC was analyzed using an 8 X 10-cm SDS-urea (7M)-PAGE (12%) gel. The DPC reaction mixtures were stored in -20°C for further use.

2.3.4 Enrichment of DNA-peptide cross-links

The DPC reaction mixture was digested by trypsin in 100 mM Tris-HCl pH 8.0 and 20 mM $CaCl_2$ at 37°C overnight. The trypsin/TFAM ratio was 1:5 (wt/wt). The digestion mixture was concentrated with a 3 kDa molecular weight cut-off filter (Millipore) and washed with 10 mM Tris-HCl pH 8.0 at 4°C.

2.3.5 Cleavage reaction of ribonucleotides by Cleave R

The Cleave R reaction was performed with the trypsin digested DNA-peptide cross-links by incubating it with 50 mM glycine-NaOH pH 10.0, 15 mM $MgCl_2$ at 55°C overnight. The efficiency of the cleavage reaction was checked with DNA-sequencing PAGE. The reaction mixture was neutralized with the acetic acid to pH 8.0 before LC-MS/MS analysis.

2.3.6 Data analysis with MaxQuant

Proteomic analysis and the enrichment efficiency of the workflow were evaluated with MaxQuant [26] through searching with TFAM sequence and the MaxQuant default contaminant sequences. The methionine oxidation and tryptophan oxidation were set as the variable modifications, the default settings for orbitrap were applied and allowed three missed cleavages.

2.3.7 LC-MS/MS analysis

Liquid chromatography was performed on a Thermo nLC1200 in single-pump trapping mode with a Thermo PepMap RSLC C18 EASY-spray column (2 µm, 100 Å, 75 µm x 25 cm) and a Pepmap C18 trap column (3 µm, 100 Å, 75 µm x 20 mm). Solvents used were A: water with 0.1% formic acid and B: 80% acetonitrile with 0.1% formic acid. Samples were separated at 300 nL/min with a 130-minute gradient starting at 3% B (held for 1 minute), then increasing to 50% B in 110 minutes, then to 100% B in 10 minutes, and held at 100% B for 9 minutes.

Mass spectrometry data were acquired on a Thermo Orbitrap Fusion in data-dependent mode. A full scan was conducted using 60k resolution in the Orbitrap in positive mode with a mass range of 375-1500 m/z and AGC 4.0e5. Precursors for MS2 were filtered by monoisotopic peak determination for peptides, intensity threshold 1.0 e4, charge state 2-7, and 5 second dynamic exclusion after 1 analysis with a mass tolerance of 10 ppm. MS2 were collected in the ion trap with an isolation window of 1.6 Da, AGC target of 3.0 e4, and a 300 ms maximum injection time. For each precursor, MS2 scans were collected using both collision-induced dissociation (CID) and higher-energy collision-induced dissociation (HCD), with CID having scan priority 1 and HCD scan priority 2. Both methods used 35% collision energy.

2.3.8 Data analysis with AP_CrosslinkFinder

Raw MS data were first converted into the MGF files by MSConvert GUI from Proteowizard. These MGF files were loaded to the custom MATLAB-based program AP_CrosslinkFinder. The custom scripts were developed based on Find_XL [27] by the Kalisman laboratory for the identification of peptide-peptide cross-links, a number of changes were made to adapt it to the analysis of DNA-peptide crosslinks. The DNA sequence was entered as required. The *in silico* digestion of TFAM allows three missed cleavages with oxidative modifications on methionine (+16 Da) and tryptophan (+4, 16, 20, 32 Da). The precursor mass tolerance was 10 ppm, and the fragment mass tolerance was 30 ppm. The peptide fragments were searched with b and y ions, while the DNA fragments were searched with a, a-b, c, y, and w ions. The length of the DNA is set to be the AP residue alone or AP and a neighboring AMP, with or without the 5'-phosphate group. The output structure ms1 displayed the identified cross-links and a score for each cross-link based on the number of fragment ions found in MS2. MS2 spectra were annotated manually.

2.4 Results

2.4.1 Optimizing the Cleave R Reaction

To create DNA-peptide cross-links with a predictable DNA fragment, we exploited the alkaline-labile characteristics of ribonucleotides (**Figure 2.1 a**)[28] and designed DNA substrates (sequence shown in **Table 1**) with ribonucleotides adjacent to the cross-linking

deoxyribonucleotide residue (**Figure 2.1 b**). All DNA substrates are double-stranded (ds) with the ribonucleotide-containing strand harboring a 5'-fluorescein label to facilitate product analysis via polyacrylamide gel electrophoresis (PAGE). The sequences of these substrates are based on the light-strand promoter sequence of human mitochondrial DNA, which has well-characterized binding properties with TFAM [29,30]. To obtain the optimal condition to cleave two ribonucleotides on a model substrate (D4), we tested RNase HII [31] and 0.3 M NaOH [32], which have been commonly used to cut ribonucleotide-containing DNA [33]. RNase HII failed to completely cleave two ribonucleotides up to 21-h reactions (**Figure 2.2 a**). The cleavage condition with 0.3 M NaOH at 55°C was complete within 2 hours (**Figure 2.2 b**); however, this reaction hydrolyzed TFAM (**Figure 2.2 c**), destroying the amino acid signatures of the cross-linked peptides.

Table 2.1 Sequences of the DNA substrates used in this study. For simplicity, substrates D1-D5 denote dsDNA annealed to a complementary strand. X indicates the AP site, and U is deoxyuridine. The underlined nucleotides denote ribonucleotides.

Name	**Sequence (5'-3')**
D1	FAM - TAA CAG TCA CCC CCC XAC TAA C
D2	FAM - TAA CAG TCA CCC CCrC XrAC TAA C
D3	FAM - TAA CAG TCA CCC CCrC UrAC TAA C
D4	FAM - TAA CAG TCA CCC CrCC UArC TAA C
D5	FAM - TAA CAG TCA CCC CrCC XArC TAA C
Complementary strand	GTT AGT TGG GGG GTG ACT GTT A

Next, we designed alternative cleavage conditions inspired by a study by Breaker et al [34]. The authors demonstrated that the rate of the alkali-promoted transesterification reaction increases with the increase of pH, $[Mg]^{2+}$, and temperature [34]. We tested a variety of reaction conditions with D4 (**Figure 2.2 d, Table 2**) and found that a reaction condition in the presence of glycine-NaOH (pH=10) and 20 mM $MgCl_2$ at 55°C for 38 h provides the best cleavage yield (>85%), as shown in **Figure 2.1 c**. In addition, we examined the stability of DNA-peptide cross-links under this condition. No apparent degradation of the cross-links was observed after 20 h incubation (**Figure 2.5**). Therefore,

we chose this optimized condition (referred to as Cleave R reaction) in the subsequent ribonucleotide cleavage reactions).

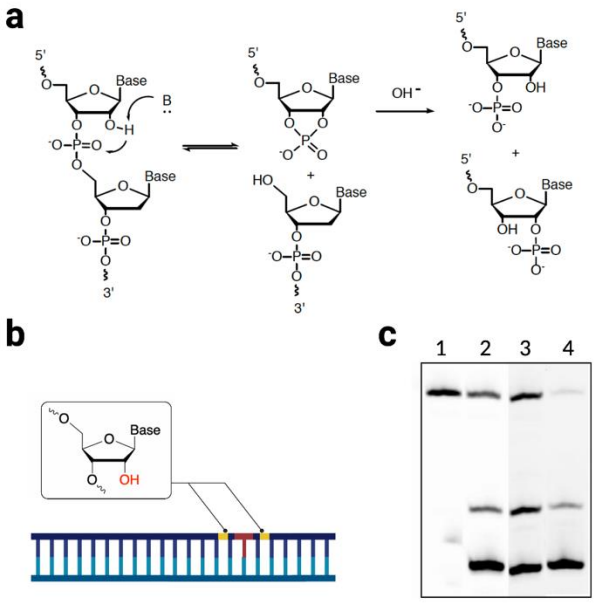


Figure 2.1 Generation of DNA fragments via Cleave R reaction. (**a**) Mechanism of alkaline transesterification reaction to generate the strand break at the 3'-side of the ribonucleotide. Under basic conditions, the cyclic phosphate product can be converted to a mixture of 3'-phosphate and 2'-phosphate [28]. (**b**) Schematic illustration of the ribonucleotide containing DNA. Deoxyribonucleotides are in blue; ribonucleotides are in yellow; the cross-linking nucleotide residue is in red. (**c**) Polyacrylamide gel analysis of products from the ribonucleotide cleavage reaction. Lane 1 is D4. Lane 2 is D4 incubated with 0.3 M NaOH for 10 min. Lanes 3 and 4 are D4 subjected to Cleave R reactions for 19 and 38 hours.

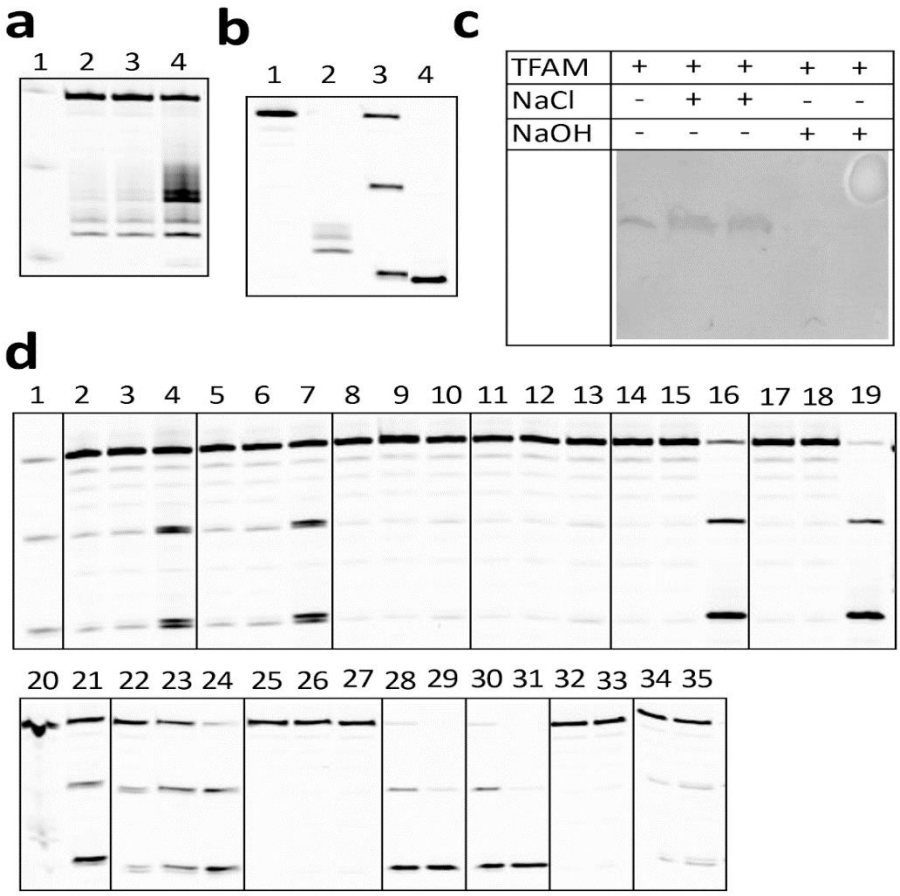


Figure 2.2 Optimization of ribonucleotide cleavage conditions. (**a**) D4 cleavage reaction with RNase HII. The reaction was performed in 50 mM Tris pH 8.0, 50 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA, 5% glycerol, and 10 mM $MgCl_2$. And 1 uM double-strand D4 was incubated with 0.1 uM RNase HII. The incubation time point of 1, 3 and 21 hours were taken (lanes 2, 3, and 4, respectively). Lane 1 is the strand break marker cleaved at two ribonucleosides. The sample was prepared by incubating D4 with 0.3 M NaOH at 55°C for 10 min. (**b**) D4 cleavage reaction with 0.3 M NaOH at 55°C. Lane 1 is D4; lane 2 is the strand break of D4 at the AP site; lane 3 is the strand break at two ribonucleoside sites maker by incubating D4 at 0.3 M NaOH, 55°C for 10 min. Lane 4 is the product of D4 incubated at 0.3 M NaOH for 2 hours. (**c**) Incubation of TFAM at 55°C with either 0.3 M NaCl or 0.3 M NaOH, two replicates. (**d**) The reaction conditions to cleave the ribonucleotides on D4. Reaction conditions are listed in Table S2. Lane 20 is D4. Lane 1 and Lane 21 are the strand break standards made by incubating D4 with 0.3 M NaOH for 10 min.

Establishing a Workflow for Analyzing Cross-linking Residues in DPC. We designed two DNA substrates (D1 and D2, sequence shown in **Table 1**) for cross-linking reactions. D1 is a double-strand (ds) DNA with an AP modification located on one strand (**Figure 2.3 b**). The AP-containing oligodeoxyribonucleotides (ODNs) were prepared using precursor ODNs containing a deoxyuridine at the lesion position, followed by removal of uracil using a DNA repair enzyme, uracil-DNA glycosylase (UDG) [35]. The dsDNA substrates were prepared by annealing AP-ODN with a complementary strand. D2 contains two neighboring (of the AP lesion) deoxynucleotides substituted with ribonucleotides. Compared to D1, the ribonucleotide-containing substrate D2 showed similar DNA-binding stoichiometry, as demonstrated by the electromobility shift assay (EMSA) (**Figure 2.3 c**). We selected a molar ratio of 1:2 (D2:TFAM) to generate the highest yield of TFAM:DNA complexes containing one TFAM and one DNA molecule for subsequent reactions. The complex, presumably conforming to the TFAM:DNA crystal structures [29,30], serves as a good model to probe specific interactions between DNA and TFAM.

Table 2.2 Cleavage reaction conditions in Figure 2.2 d

Incubation condition	**Length of incubation (h)**	**Lane number**
pH=10, [K+]=0.03 M, [Mg2+]=5 mM, T=55°C	1	2
	2	3
	18	4
pH=10, [K+]=0.01 M, [Mg2+]=5 mM, T=55°C	1	5
	2	6
	18	7
pH=10, [K+]=0.03 M, [Mg2+]=5 mM, T=37°C	1	8
	2	9
	18	10
pH=10, [K+]=0.01 M, [Mg2+]=5 mM, T=37°C	1	11
	2	12
	18	13
pH=10, [K+]=0.03 M, T=55°C	1	14
	2	15
	18	16
pH=10, [K+]=0.01 M, T=55°C	1	17
	2	18
	18	19
pH=13.4, [K+]=0.001 M, T= room temperature	2	22
	8	23
	24	24
pH=12, [K+]=0.001M, [Mg2+]=15 mM, T=37°C	2	25

	8	26
	24	27
pH=10, glycine-NaOH buffer, [K+]=0.001 M, [Mg2+]=15 mM, T=55°C	8	28
	24	29
pH=10, glycine-NaOH buffer, [Mg2+]=15 mM, T=55°C	8	30
	24	31
pH=10, NaOH solution, [K+]=0.02 M, [Mg2+]=5 mM, T=55°C	8	32
	24	33
pH=10, NaOH solution, [K+]=0.02 M, T=55°C	8	34
	24	35

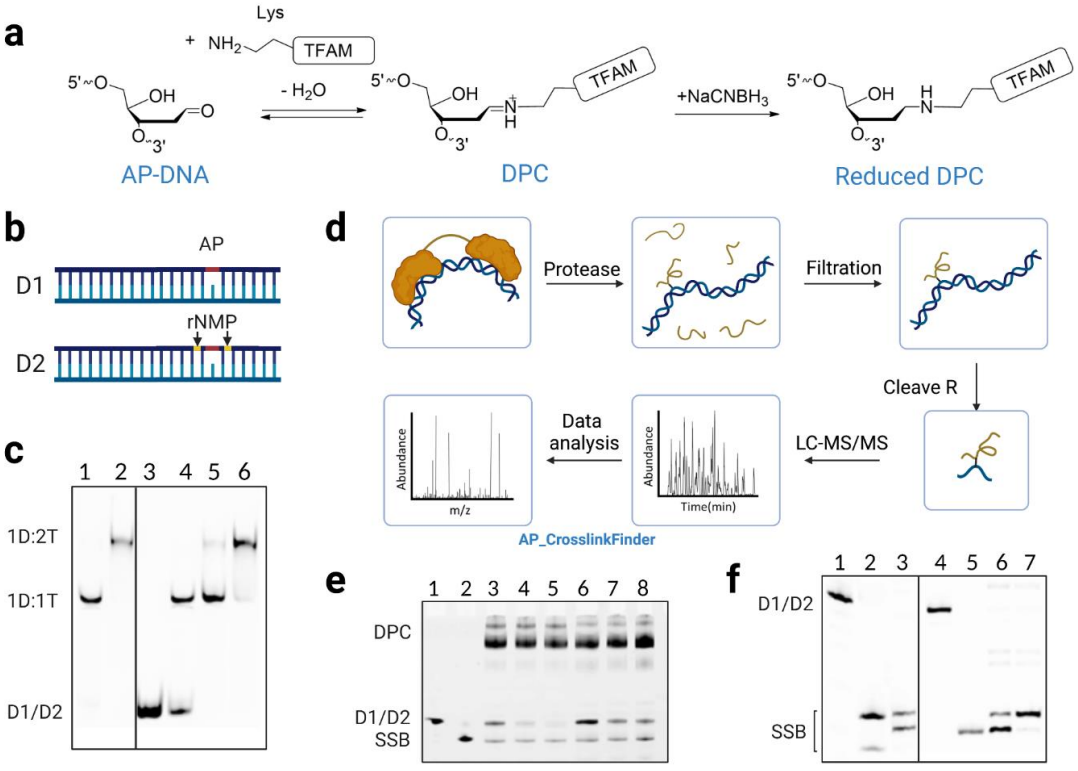

a
Lys
+ NH2
TFAM
- H2O
+NaCNBH3
AP-DNA
DPC
Reduced DPC
b
AP
D1
rNMP
D2
c
1 2 3 4 5 6
1D:2T
1D:1T
D1/D2
d
Protease
Filtration
Cleave R
LC-MS/MS
Data analysis
Abundance
m/z
Time(min)
AP_CrosslinkFinder
e
1 2 3 4 5 6 7 8
DPC
D1/D2
SSB
f
1 2 3 4 5 6 7
D1/D2
SSB

Figure 2.3 DPC formation with ribonucleotide-containing DNA and the workflow of identifying cross-linking amino acid residues. (**a**) Reaction mechanism of the formation of DPC via Schiff base chemistry. Reductive amination with $NaBH_3CN$ to stabilize the reaction intermediates for mass spectrometry analysis. (**b**) DNA substrates used in this study. D1 is a double-stranded oligodeoxyribonucleotide with an AP modification located on the top strand. D2 contains two neighboring (of the AP lesion) deoxynucleotides substituted with ribonucleotides. (**c**) DNA-TFAM binding stoichiometry determined by electrophoretic mobility shift assay (EMSA). Lanes 1 and 2 are with substrate D1 and contain DNA(D):TFAM(T) complexes formed at molecular ratios of 1:1 and 1:2 (DNA:TFAM), respectively Lane 3 contains substrate D2. Lanes 4-6 contain D2:TFAM complexes formed at molecular ratios of 1:1, 1:1.5, and 1:2 (DNA:TFAM), respectively. (**d**) Workflow for mapping AP-DNA-protein cross-linking residues with LC-MS/MS. DPC was digested with trypsin overnight to form non-cross-linked peptides and DNA-peptide conjugates. Free peptides were removed and DNA-peptide conjugates were enriched with a 3 kDa molecular weight cut-off filter and followed by the Cleave R reaction. The DNA substrate was cleaved at ribonucleotides to yield DNA-peptide cross-links with a predictable number of nucleotide residues. Reaction products were subjected to LC-MS/MS analysis. DNA-peptide cross-links were identified using AP_CrosslinkFinder followed by manual annotation. (**e**) SDS-urea PAGE analysis of DPC from reactions of TFAM with D1 or D2. Lanes 1 and 2 contain D1 and its correlating strand break marker. Lane 3-5 are the DPC reactions between D1 and TFAM for 3, 6 and 12 hours. Lanes 6-8 are the DPC reactions between D2 and TFAM for 3, 6 and 12 hours. SSB refers to DNA single-strand breaks. The percent yield of DPC is 92% for reactions with D1 and TFAM and 86% for reactions with D2 and TFAM. Detailed analysis shown in **Figure 2.11**. (**f**) DNA-sequencing PAGE analysis of the resulting DNA fragments after Cleave R reactions. Lanes 1 and 4 are D2, and lane 2 is D2 upon cleavage with AP endonuclease 1 (APE1) (**Figure 2.6 a**). Lanes 3 and 6 are trypsin digested D2-DPC after Cleave R reactions. Lane 5 is the product of cleaved D2 with 1 M NaOH. Lane 7 is the sample from lane 3 (or 6) treated with alkaline phosphatase.

We aimed to develop an experimental and computational workflow to map the cross-linking residues on TFAM within TFAM-DNA complexes (**Figure 2.3 d**). We harnessed the Schiff base intermediates formed between an AP modification and lysine residues on TFAM and captured the interactions using covalent DNA-protein cross-links. *In situ* trapping of the Schiff base intermediate was achieved by reductive amination in the presence of $NaCNBH_3$ (**Figure 2.3 a**). We incubated TFAM with D2 in the presence of $NaCNBH_3$ for 12 hrs followed by quenching of any unreacted AP sites and unreduced DPC with $NaBH_4$. Relative to D1, D2 has a similar DPC yield after 12 h, indicating that the substitution with ribonucleotides in DNA did not perturb the cross-linking reaction (**Figure 2.3 e**). The stabilized DPCs were digested with trypsin to form DNA-peptide cross-links and non-cross-linked peptides. The trypsin digestion was monitored by SDS-urea PAGE (**Figure 2.4**). The resulting sample was filtered through a 3 kDa molecular cut-off filter to remove a majority of the non-cross-linked peptides for simplified data search and to enrich DNA-peptide cross-links. We verified the removal of non-cross-linked peptides through searching with MaxQuant [26]. The enriched DNA-peptide cross-links were converted to DNA-peptide cross-links with a predictable nucleic acid moiety under the Cleave R condition. The successful cleavage of DNA was confirmed by SDS-urea PAGE (**Figure 2.4**) and high-resolution DNA-sequencing PAGE (lanes 3 and 6 of **Figure 2.3 f**). In addition, no apparent degradation of DNA-peptide cross-links was observed after 20-h incubation (**Figure 2.5**), reinforcing the suitability of the Cleave R reaction in converting nucleic acids into the desired products while preserving the structure of the peptides. On

the contrary, under 0.3 M NaOH, the peptide fragment in DNA-peptide cross- links was hydrolyzed (lane 7-9 of **Figure 2.5**), indicating that this commonly used condition for ribonucleotide cleavage is not applicable to preparing DNA-peptide cross-links for subsequent mass spectrometry analysis.

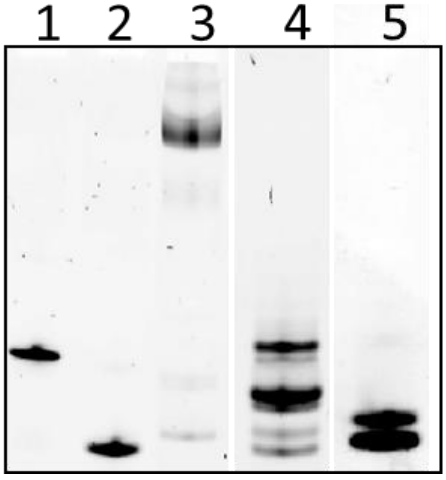


Figure 2.4 Preparation of TFAM-DNA cross-links monitored by gel electrophoresis. Lane 1 is D3. Lane 2 is a strand break marker at the AP site derived from D2. Lane 3 is D2 DPC generated in the presence of $NaCNBH_3$. Lane 4 is trypsin digested D2 DPC. Lane 5 is the product of trypsin digested D2 DPC after Cleave R reaction.

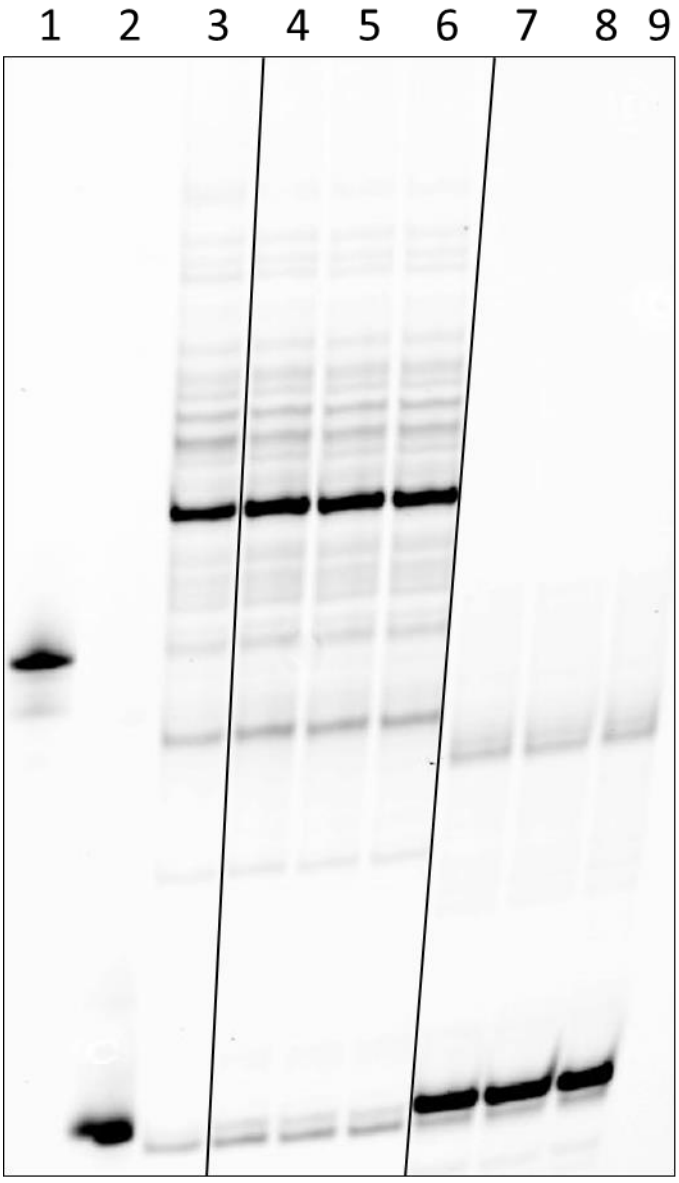


Figure 2.5 Stability of D2-peptide cross-links from trypsin digestion of D2-TFAM DPC under Cleave R condition and NaOH treatment. Lane 1, D2. Lane 2, cleaved single-stranded breaks at AP sites from D2 after NaOH treatment. Lane 3, trypsin-digested D2-TFAM DPC without treatment. Lane 4-6 (triplicate) treatment of digested D2-TFAM DPC under Cleave R condition for 20 h. Lane 7-9 (triplicate) treatment of digested D2-TFAM DPC with 0.3 M NaOH at 55°C for 2 hours.

To understand the chemistry of the DNA termini from the Cleave R reaction and to guide the data search for the nucleic acid-peptide cross-links, we treated the product

with alkaline phosphatase and observed the disappearance of the lower band (lanes 6 and 7 of **Figure 2.3 f**). The results indicate that products in the faster-migrating band contain a terminal phosphate group at the 3'-end and that products in the upper band do not, consistent with the phosphate mixtures produced under alkaline conditions (**Figure 2.1 a**). The assignment is supported by comparing with standard products generated under NaOH (lane 5 of **Figure 2.3 f**) and APE1 (lane 2 of **Figure 2.3 f**) [36]. The NaOH treatment produces products containing a phosphate group at the 3'-end upon cleavage at ribonucleotides. Products from APE1 treatment contain a 3'-OH upon cleavage at the 5'-side of the abasic lesion [37] (**Figure 2.6 a**). These results prompted us to search for DNA components with and without the phosphate group when analyzing DNA-peptide cross-links.

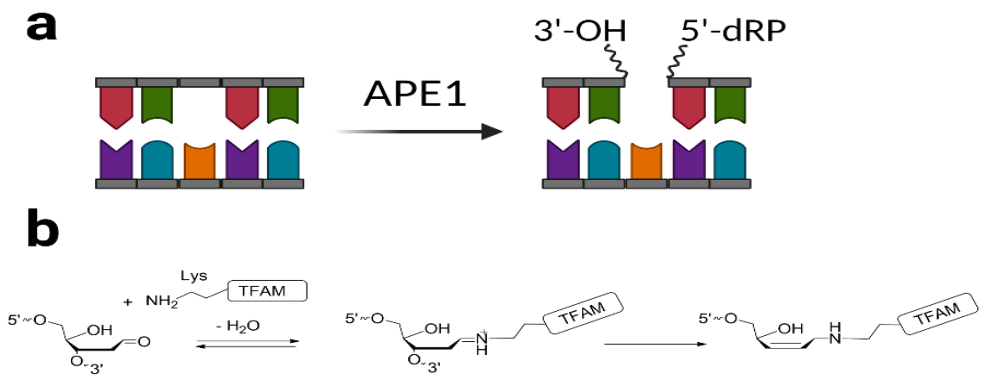


Figure 2.6 (**a**) APE1 cleaves the DNA backbone 5' of the AP lesion to yield a 3'-OH terminal and a 5'-deoxyribose phosphate residue [37]. (**b**) TFAM(peptide)-AP-DNA cross-links undergo β-elimination before reductive amination by $NaBH_3CN$ to afford a mass adduct of 102.

The sample was analyzed by nanoLC-MS/MS. The mass spectrometry data were processed by the custom scripts of AP_CrosslinkFinder, which are modified based on Find_XL [27]. We applied AP_CrosslinkFinder to find the AP cross-links, and the mass adducts of the cross-linked peptides were computed and added to the *in silico* digested TFAM. The search begins with matching the molecular weight in MS1, followed by searching the corresponding fragment ions in MS/MS and outputting the number of matches. The spectra were annotated manually. Together, our optimized workflow from sample preparation to data analysis ensures the identification of cross-linking amino acid residues in DPC.

2.4.2 Mapping AP reactive sites at single amino acid resolution in TFAM-DNA cross-links.

The short and structurally predictable nucleic acid component in DNA-peptide cross-links significantly decreased the search space and reduced the search time. The computational analysis of the LC-MS/MS data with AP_CrosslinkFinder was completed within an hour. We identified 14 unique TFAM-DNA cross-linking sites located on 15 unique peptides (**Figure 2.7 a**, **Table 2.3**). Four types of mass adducts were observed for the cross-linked peptides in mass spectrometry analysis (+102, +118, +198, and +527). The mass adduct, 102 Da, is a minor product generated when the reaction proceeds with β-elimination before trapping by $NaCNBH_3$ (**Figure 2.6 b**). The proposed mechanism by which the other three major products are formed is shown in **Figure 2.7 b**. We presume that the 4'-OH of the open-chain form of AP sites or the 2'-OH of the ribonucleoside can

undergo nucleophilic attack at the phosphate, leading to the formation of cyclic phosphate intermediates (structures in the middle of **Figure 2.7 b**). The cyclic phosphates can be hydrolyzed, resulting in products containing a mass adduct of 527 or 198 Da. The phosphate group can be hydrolyzed, as evidenced by PAGE analysis (**Figure 2.3 f**), generating additional products with a mass adduct of 447 or 118 Da (**Figure 2.7 b**). The MS/MS spectra of the cross-links contain both fragmented peptides and fragmented ODNs. For the cross-links with a mass adduct of 527 or 447, the most abundant product ions are generated by nucleic acid fragmentation owing to the weak *N*-glycosidic bond and phosphoester bonds in the nucleic acid moiety (**Figure 2.7 c**). The nucleic acid fragment derived from D2 contains an AP lesion and an adenosine monophosphate residue, which readily loses an adenine residue, generating an abundant product ion with m/z=136. We considered m/z=136 as the diagnostic ion of this type of cross-links (**Figures 2.7 d**). For the mass adducts 198 and 118, the nucleic acid component in DPC contains the reduced ring-opened AP site, which appears to be relatively stable under MS fragmentation conditions. Under collision-induced dissociation (CID), the b- and y-ions of these cross-links contain the aforementioned mass adducts relative to the native peptides (**Figures 2.7 e-g**), facilitating the assignment of peptide sequences and mapping of the cross-linking residues at single amino acid resolution. The fragmentation patterns of nucleic acid-peptide cross-links guided us to develop the MATLAB-based program, AP_CrosslinkFinder.

Table 2.3 Identified lysine residues of TFAM cross-linked with the AP lesion. The crosslinked amino acid residues are in blue, and oxidative modifications are denoted by OX in red. Methionine oxidation has the mass adduct of 16; tryptophan oxidation has the mass shift of +4, +16, +20, and +32 [38], labeled the mass adduct of tryptophan oxidation.

Ion m/z, charge state	Peak area	Starting residue	Crosslinked peptide	Crosslinked residue	Charge	m/z	mass adduct	Δm [ppm]	Note
492.8825, z=3	12327082	52	KPVSSYLR	K52	3	492.8825	527.0818	2.37	
738.8200, z=2	5119750	52	KPVSSYLR	K52	2	738.8200	527.0818	2.6	
478.9149, z=3	14292681	60	FSKEQLPIFK	K62	3	478.9149	198.0293	0.73	
717.8683, z=2	26719839	60	FSKEQLPIFK	K62	2	717.8683	198.0293	0.54	
599.6343, z=3	23513831	63	EQLPIFKAQNPDAK	K69	3	599.6343	198.0293	2.37	
898.9478, z=2	18528305	63	EQLPIFKAQNPDAK	K69	2	898.9478	198.0293	2.67	
709.6531, z=3	1.01E+08	63	EQLPIFKAQNPDAK	K69	3	709.6531	527.0818	0.17	C13
1063.9760, z=2	48499082	63	EQLPIFKAQNPDAK	K69	2	1063.9760	527.0818	0.42	C13
552.2747, z=3	87499742	70	AQNPDAKTTELIR	K76	3	552.2747	198.0293	2.25	
827.9083, z=2	1.1E+08	70	AQNPDAKTTELIR	K76	2	827.9083	198.0293	2.44	
661.9597, z=3	4.65E+08	70	AQNPDAKTTELIR	K76	3	661.9597	527.0818	3.13	

992.4349, z=2	3.04E+08	70	AQNPDAKTTELIR	K76	2	992.4349	527.0818	2.39	
552.6153, z=3	23470686	105	AEWQVYKEEISR	K111	3	552.6153	118.0630	1.65	
828.4192, z=2	18982627	105	AEWQVYKEEISR	K111	2	828.4192	118.0630	1.83	
579.2705, z=3	47453109	105	AEWQVYKEEISR	K111	3	579.2705	198.0293	1.16	
868.4026, z=2	62176994	105	AEWQVYKEEISR	K111	2	868.4026	198.0293	2.04	
688.9551, z=3	1.57E+08	105	AEWQVYKEEISR	K111	3	688.9551	527.0818	1.61	
1032.9289, z=2	92505104	105	AEWQVYKEEISR	K111	2	1032.9289	527.0818	1.76	
638.3371, z=3	33482050	117	FKEQLTPSQIM^{OX}SLEK	K118	3	638.3371	118.0630	1.07	
774.6785, z=3	55473176	117	FKEQLTPSQIM^{OX}SLEK	K118	3	774.6785	527.0818	3.2	
1161.5138, z=2	7752968	117	FKEQLTPSQIM^{OX}SLEK	K118	2	1161.5138	527.0818	3.18	
404.5302, z=3	46823451	132	EIMDKHLK	K136	3	404.5302	198.0293	0.67	
606.2918, z=2	52592837	132	EIMDKHLK	K136	2	606.2918	198.0293	0.35	
409.8621, z=3	33493714	132	EIM^{OX}DKHLK	K136	3	409.8621	198.0293	1.32	
614.2893, z=2	27462866	132	EIM^{OX}DKHLK	K136	2	614.2893	198.0293	1.42	
519.5467, z=3	1.56E+08	132	EIM^{OX}DKHLK	K136	3	519.5467	527.0818	1.87	

778.8163, z=2	1.11E+08	132	EIM[OX]DKHLK	K136	2	778.8163	527.0818	2.08	
514.2148, z=3	3.13E+08	132	EIMDKHLK	K136	3	514.2148	527.0818	1.39	
770.8183, z=2	2.35E+08	132	EIMDKHLK	K136	2	770.8183	527.0818	1.39	
376.1973, z=2	27329832	137	HLKR	K139	2	376.1973	198.0293	0.77	
553.2132, z=2	1.19E+08	141	KAMTK	K141	2	553.2132	527.0818	3.38	
1105.4186, z=1	2947804	141	KAMTK	K141	1	1105.4186	527.0818	3.4	
631.2629, z=2	21301011	140	RKAMTK	K141	2	631.2629	527.0818	1.6	
551.9396, z=3	13746344	147	KELTLLGKPK	K147	3	551.9396	527.0818	1.16	
827.4062, z=2	5595916	147	KELTLLGKPK	K147	2	827.4062	527.0818	2.03	
558.8481, z=2	8293783	148	ELTLLGKPK	K154	2	558.8481	118.0630	0.82	
598.8319, z=2	1.68E+08	148	ELTLLGKPK	K154	2	598.8319	198.0293	1.85	
1196.6560, z=1	9477711	148	ELTLLGKPK	K154	1	1196.6560	198.0293	1.87	
509.2416, z=3	1.61E+08	148	ELTLLGKPK	K154	3	509.2416	527.0818	1.89	
763.3587, z=2	4.59E+08	148	ELTLLGKPK	K154	2	763.3587	527.0818	2.17	
511.783, z=2	15554123	184	TVKENWK	K186	2	511.7830	118.0630	6.33	

483.1944, z=3	5931178	184	TVKENW^{OX}+16K	K186	3	483.1944	527.0818	1.49	
551.7642, z=2	3.41E+08	184	TVKENWK	K186	2	551.7642	198.0293	2.34	
1102.5, z=1	23504853	184	TVKENWK	K186	1	1102.5213	198.0293	3	
559.7613, z=2	1.16E+08	184	TVKENW^{OX}+16K	K186	2	559.7613	198.0293	1.68	
477.8629, z=3	3.4E+08	184	TVKENWK	K186	3	477.8629	527.0818	1.78	
716.7882, z=2	1.15E+09	184	TVKENWK	K186	2	716.7882	527.0818	-6.78	C13
503.7857, z=2	37482775	184	TVKENWK	K186	2	503.7857	102.0681	6.73	
626.2455, z=3	4434088	187	ENWKNLSDSEK	K190	3	626.2455	527.0818	3.11	
938.8653, z=2	6038440	187	ENWKNLSDSEK	K190	2	938.8653	527.0818	4.14	
421.7237, z=2	46868679	228	KDLLR	K228	2	421.7237	198.0293	1.52	
842.4403, z=1	3479367	228	KDLLR	K228	1	842.4403	198.0293	2.39	
586.2503, z=2	1.61E+08	228	KDLLR	K228	2	586.2503	527.0818	1.69	
1171.4922, z=1	5252689	228	KDLLR	K228	1	1171.4922	527.0818	1.2	

The predictable structures of the nucleic acid component facilitate database search and the assignment of the peptide sequence. The database search focuses on the tryptic peptides of TFAM cross-linked to the AP lesion in the nucleic acid moiety. Nearly all the nucleic acid-peptide cross-links have a mass accuracy within 3 ppm and MS/MS product ions matching to its theoretical ions (**Table 2.3**, **Figures 2.8** and **2.9**), ensuring unbiased DPC assignments.

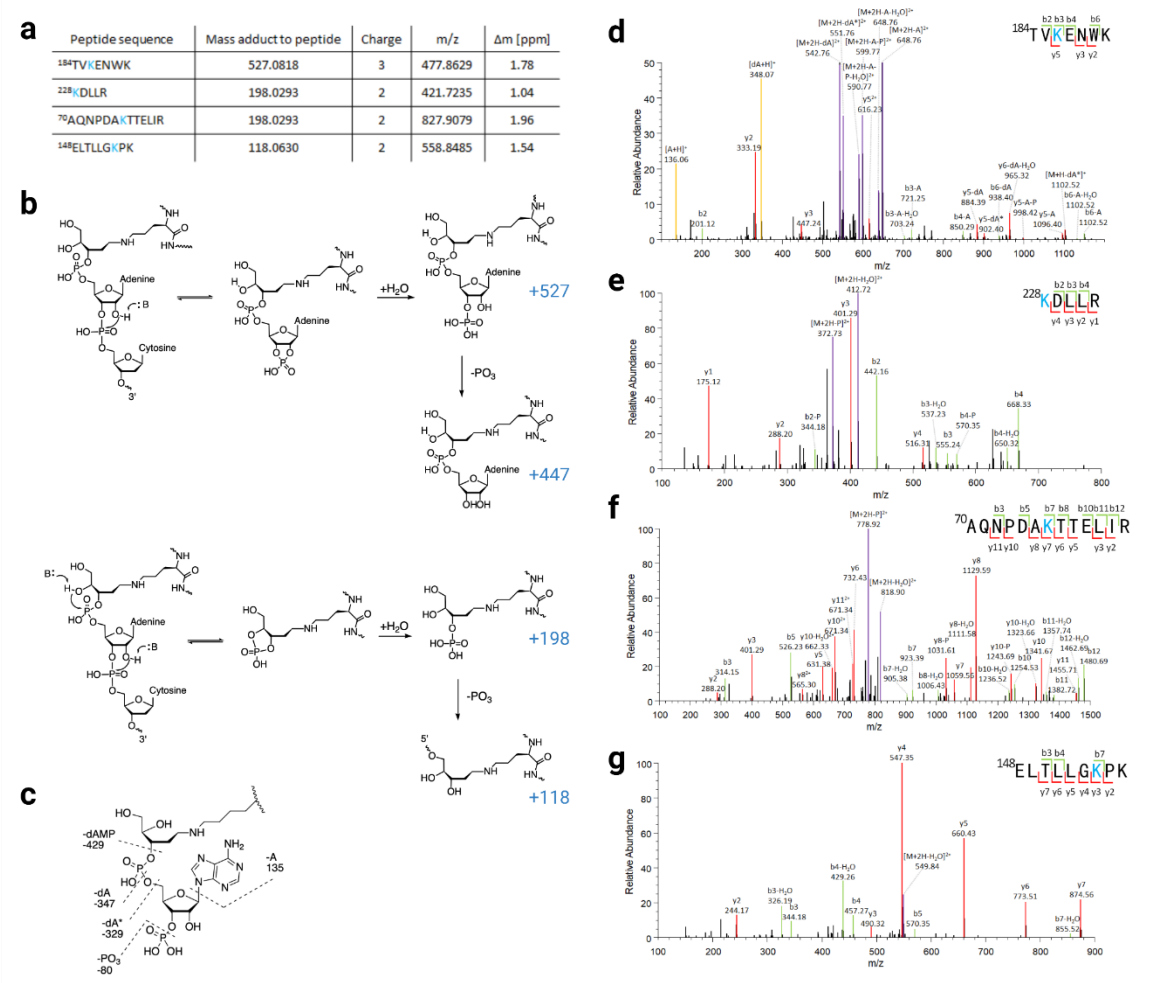


Figure 2.7 Identification of the DNA-protein cross-linking sites in DNA-TFAM cross-links. (**a**) Representative DNA-peptide cross-links found in this study. The cross-linked amino acid residues are in blue. (**b**) The proposed reaction scheme for the observed mass shift of 118, 198, 447, and 527. (**c**) The proposed fragmentation pattern for nucleotide derivatives observed in mass spectra in (**d**). (**d**) through (**g**) MS/MS spectra of the cross-links shown in (**a**). In the MS/MS spectra, b ions are in green, and y ions are in red. The DNA fragments are in yellow, and the ions with fragmentations in the DNA component are in violet.

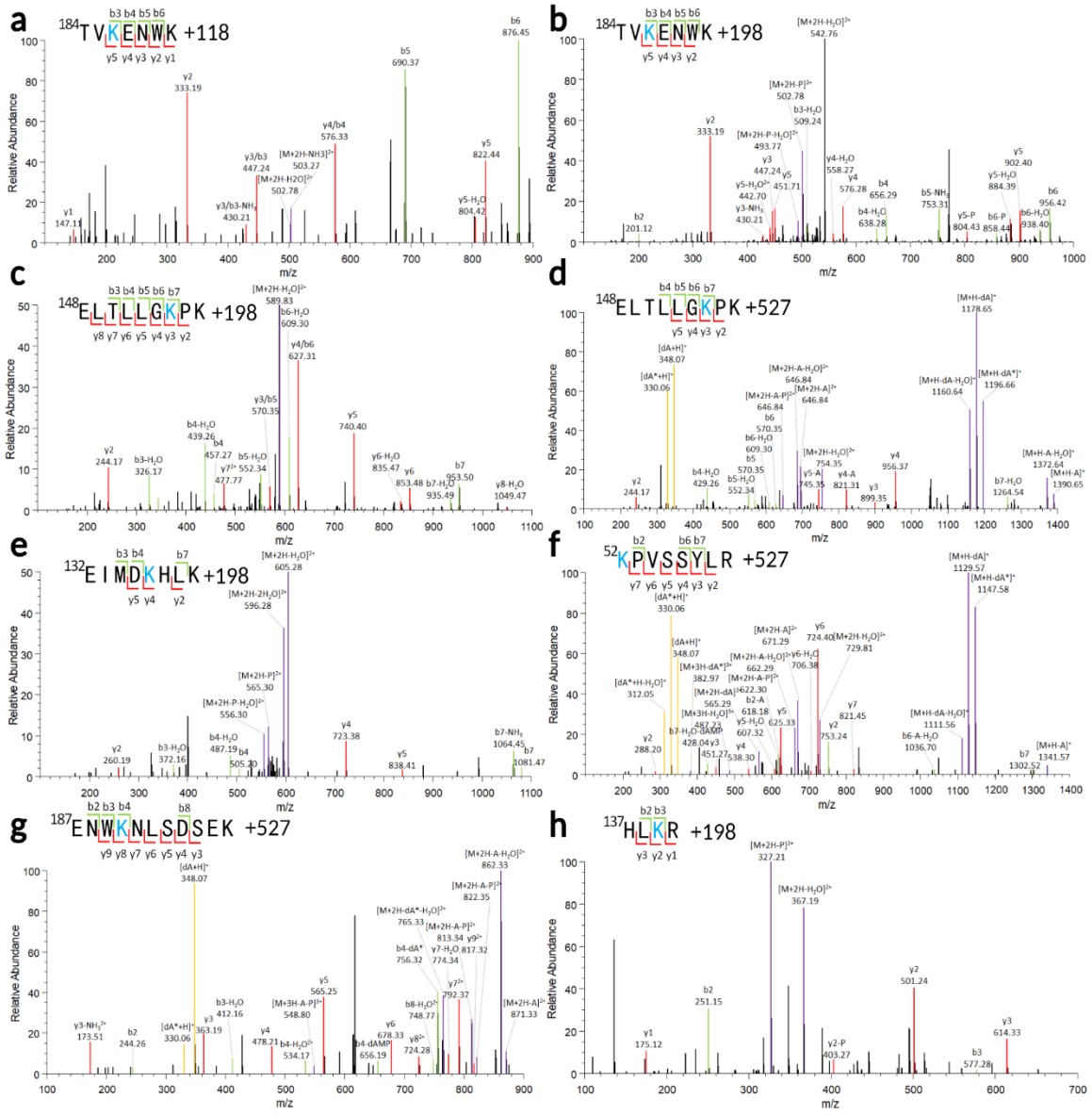


Figure 2.8 The MS2 spectra of the crosslinks identified in this study, part 1.

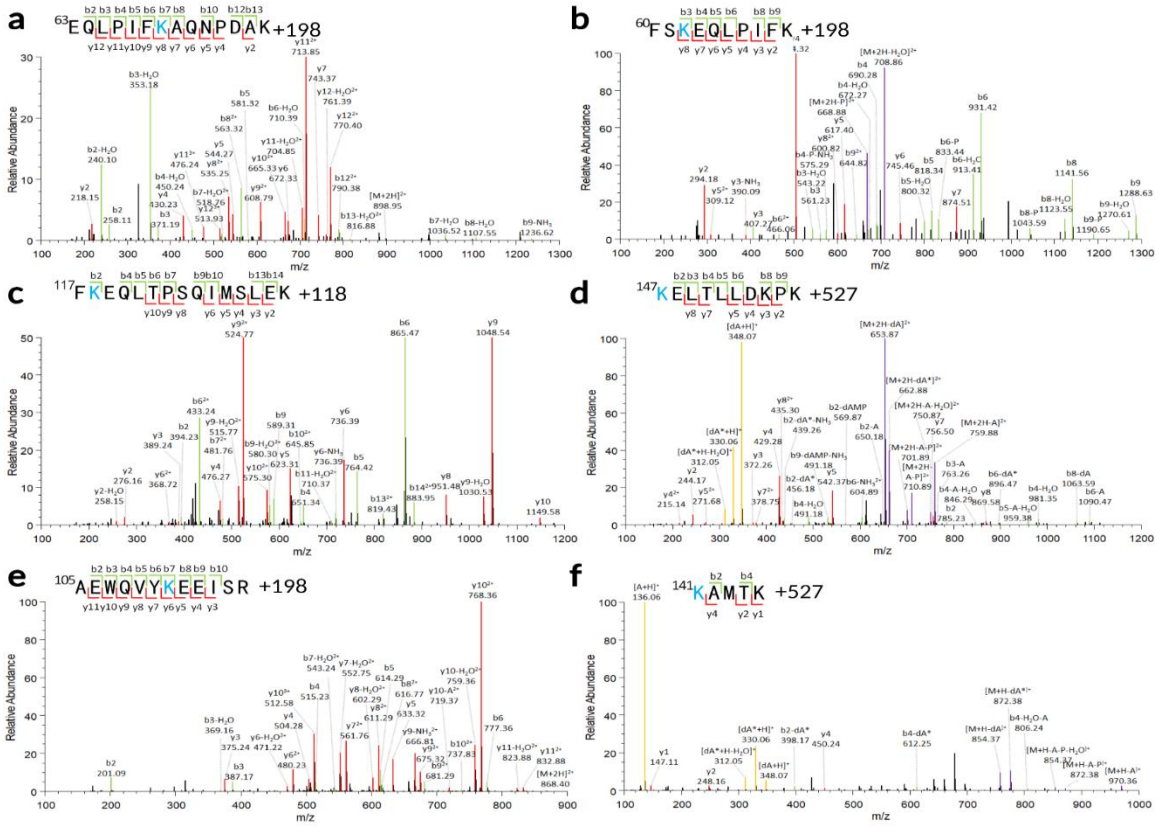


Figure 2.9 MS2 spectra of the crosslinks identified in this study, part 2.

2.4.3 Site and quantitative analysis of the cross-linking sites.

Our previous study demonstrated that K183, K186, and K190, three residues in the vicinity of the AP site (**Figure 2.10 a**), play an important role in influencing the cross-linking rate in AP-DNA:TFAM complexes [13]. The results here corroborate these earlier data and provide direct evidence for covalent cross-linking of K186 and K190 with AP-DNA under reductive amination. Semi-quantification based on the integrated peak areas of MS1 spectra reveal that K186 is the most abundant cross-linking site, consistent with the closest proximity of K186 to the AP site relative to the other two lysine residues in TFAM:DNA co-crystal structures (**Figure 2.10 a**) [29,30].

We grouped these cross-linking sites by the domains of TFAM (**Figure 2.10 b** and **2.4 c**). Besides K186, additional residues such as K76, K69, and K62 in the HMG1 domain are shown to interact with the AP lesion (**Figure 2.10 c**). The observation of cross-linking sites other than K186 is consistent with the dynamic characteristics of TFAM:DNA complexes in solution [39]. In addition, the majority of the DNA-peptide cross-links identified have the mass adduct of 527 Da (**Figure 2.10 d**), indicating the cleavage reaction at the ribonucleoside is very efficient. The second most abundant mass adduct is the 198 Da, which could form by the proposed mechanism via an alkaline transesterification reaction similar to the ribonucleoside transesterification reaction (**Figure 2.7 b**).

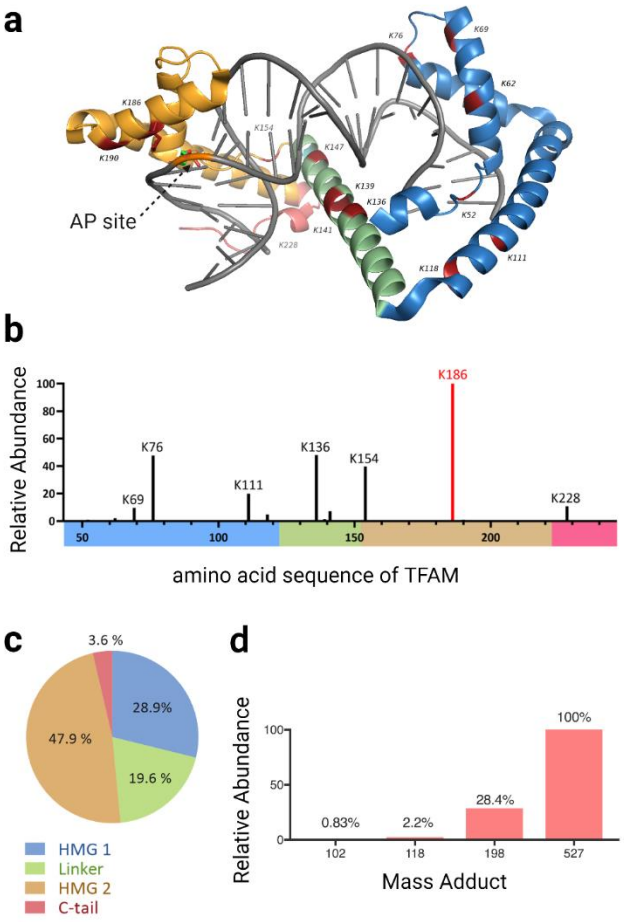


Figure 2.10 High-resolution mapping of the cross-linking sites in TFAM-DNA complexes. (**a**) Identified cross-linking sites in TFAM-DNA complexes based on a co-crystal structure (PDB: 3TQ6). The cross-linked lysine residues are in red, and the AP lesion is illustrated in sticks and indicated by the black arrow. (**b**) Relative abundance of the cross-linking site identified in TFAM-DNA cross-links based on their integrated peak areas in LC-MS/MS analysis. (**c**) The abundance of cross-linked residues in four domains of TFAM. (HMG1 43-122 in blue, linker 122-152 in green, HMG2 152-222 in brown, and C-terminal tail 222-246 in pink). (**d**) Relative abundance of four types of observed mass adducts to the cross-linked peptides.

2.5 Discussion

In this study, we developed a mass spectrometry-based workflow for mapping interacting lysine residues in DNA-protein complexes. We leveraged the alkaline lability of ribonucleotides and designed ribonucleotide-containing DNA substrates to prepare structurally defined nucleic acid-peptide cross-links. The optimized Cleave R reaction cleaves DNA at ribonucleotides while retaining the integrity of nucleic acid-peptide conjugates. Common digestion methods for cleaving nucleic acids involve nuclease cocktails, which often yield a mixture of mononucleotides, dinucleotides, and oligonucleotides due to steric hindrance of different cross-links [24]. The heterogeneous conjugates tend to decrease the ionization efficiency, complicate the database search, and hinder the identification of conjugates and cross-linking sites [25]. Our approach creates structurally defined nucleic acid moieties in nucleic acid-peptide conjugates, which simplify database search and facilitate mapping of AP-interacting lysine residues at single amino acid resolution.

We exploited the Schiff base chemistry to capture interactions between a reactive abasic modification and lysine residues of a protein. The 14 cross-linked lysine residues in TFAM-DNA complexes provided insights into the conformational dynamics and heterogeneity of TFAM:DNA complexes. The predominant cross-linking site at K186 is in agreement with K186 being the closest lysine (to the AP modification) residue in TFAM:DNA crystal structures. Additional cross-linking sites support the dynamic nature of

the TFAM:DNA complexes and alternative binding conformations. For example, the heterogeneity of the TFAM:DNA complexes in solution has been demonstrated using small-angle X-ray scattering, single-molecule Föster resonance energy transfer assays, and molecular dynamic simulations [39]. Alternative conformations of TFAM:DNA complexes (especially under micromolar TFAM) other than reported crystal structures could explain the observed cross-linking residues in HMG1 and the linker domains [40,41]. The identified cross-linking residues K136, K139, K147, and K154 at the linker region corroborate the nonstatic characteristics of the TFAM:DNA complexes: the flexible linker is known to assist TFAM:DNA complexes to undergo a butterfly or ‘‘breathing’’ movement [39]. Furthermore, the higher DNA-binding affinity of the HMG1 domain relative to HMG2 [42] and the sliding of TFAM DNA could also contribute to the cross-links formed with residues on HMG1 [40].

The reported sample preparation is simple, inexpensive, and requires no gel- or affinity-based purification, which avoids cumbersome sample workup and potential contaminations. In particular, the size-based single-step enrichment streamlines sample workup. With sub-nmol of TFAM sample, we were able to identify 14 cross-linking sites on TFAM, demonstrating that the workflow is sensitive in detecting major and minor cross-linking residues. The workflow is amenable to other cross-linking chemistry by replacing AP sites with a different reactive residue [43]. A recent report using chemical digestion of RNA to map RNA-interacting proteins involves harsh chemical treatment, which may not be applicable to different cross-linking chemistries [24].

To our knowledge, RNP^{XL} is by far the only available software tool for analyzing the MS data of DPCs and requires the commercial Proteome Discoverer software [44]. However, RNP^{XL} cannot be applied to analyze the AP sites crosslinked to DPCs due to its lack of flexibility for user-defined searches. The developed MATLAB-based program AP_CrosslinkFinder can be easily applied to analyzing other types of DPCs using a user-defined mass of the cross-linked conjugate, such as UV-cross-linked DPCs/RNPs [45] and formaldehyde crosslinked DPCs [46]. The AP_CrosslinkFinder used a simpler but more specific algoristhm than the RNP^{XL}; thus, it can complete the searching of a set of LC-MS/MS data within one hour with the current dataset.

2.6 Conclusion

In summary, we have developed a powerful tool for mapping interacting amino acid residues with AP sites in DNA-protein complexes at single amino acid resolution. The method requires minimal sample input and is sensitive in identifying both major and minor cross-linking sites. Results offer quantitative information on the relative abundance of cross-linking sites. The method can complement the advanced structural biology techniques by providing information on the proximity of interacting functional groups, multiple conformations of nucleic acid-protein complexes in solution, and the inferred relative reactivity of varying amino acid residues. We envision that this workflow is applicable to mapping other interacting residues provided a different functional group can be installed on the oligonucleotide (e.g., a thiol group to map interacting cysteine).

When used together with an affinity handle, such as biotin, such a synthetic oligomer can potentially be used to probe interacting proteins and residues in complex biological samples, such as cell extracts.

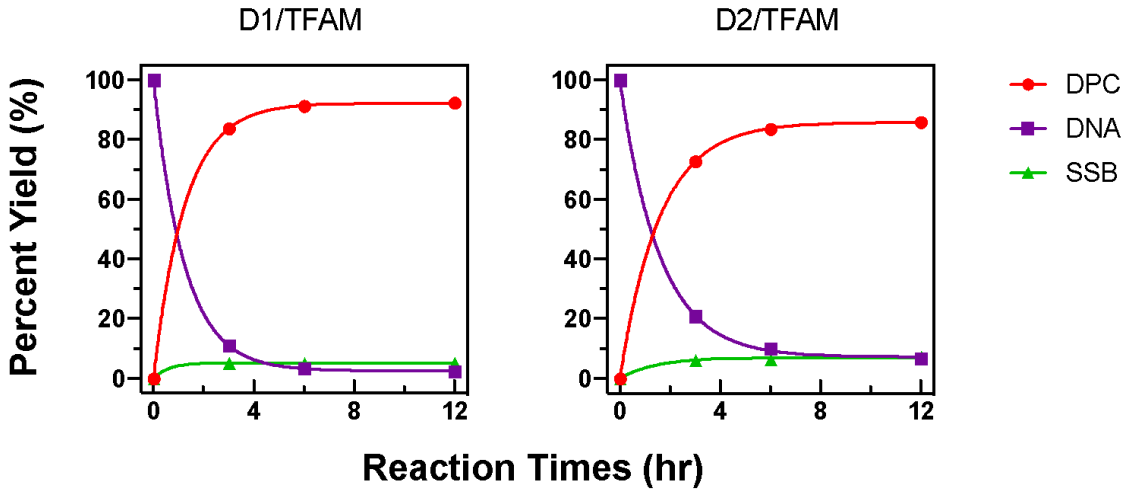


Figure 2.11 Quantification of gel analysis of TFAM:AP-ODN reactions shown in Figure 2.3 e. D1 is a double-strand (ds) DNA with an AP modification located on one strand, as illustrated in Figure 2b. D2 contains the identical nucleotide sequences to D1, except that it has two neighboring (of the AP lesion) deoxynucleotides substituted with ribonucleotides. The percentage yield of each species was fit to a single exponential equation to obtain the apparent formation rate (k_{fo}) of DNA-protein cross-links (DPC) and the cleavage rate (k_{cl}) of AP-ODN. For reactions with D1 and TFAM, the percent yield of DPC is 92%, k_{fo} is 0.79 hr^{-1}, and k_{cl} is 0.81 hr^{-1}. For reactions with D2 and TFAM, the percent yield of DPC is 86%, k_{fo} is 0.63 hr^{-1}, and k_{cl} is 0.63 hr^{-1}.

2.7 References

[1] N.Y. Tretyakova, A. Groehler, S. Ji, DNA–Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes, Acc. Chem. Res. 48 (2015) 1631–1644. https://doi.org/10.1021/acs.accounts.5b00056.

[2] M.J. Solomon, A. Varshavsky, Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures., Proc. Natl. Acad. Sci. 82 (1985) 6470–6474. https://doi.org/10.1073/pnas.82.19.6470.

[3] Q. You, A.Y. Cheng, X. Gu, B.T. Harada, M. Yu, T. Wu, B. Ren, Z. Ouyang, C. He, Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution, Nat. Biotechnol. 39 (2021) 225–235. https://doi.org/10.1038/s41587-020-0643-8.

[4] J. Liu, L. Cai, W. Sun, R. Cheng, N. Wang, L. Jin, S. Rozovsky, I.B. Seiple, L. Wang, Photocaged Quinone Methide Crosslinkers for Light-Controlled Chemical Crosslinking of Protein–Protein and Protein–DNA Complexes, Angew. Chemie Int. Ed. 58 (2019) 18839–18843. https://doi.org/10.1002/anie.201910135.

[5] H. Ide, T. Nakano, A.M.H. Salem, M.I. Shoulkamy, DNA–protein cross-links: Formidable challenges to maintaining genome integrity, DNA Repair (Amst). 71 (2018) 190–197. https://doi.org/10.1016/j.dnarep.2018.08.024.

[6] X. Wei, Y. Peng, C. Bryan, K. Yang, Mechanisms of DNA−protein cross-link formation and repair, Biochim. Biophys. Acta - Proteins Proteomics. 1869 (2021) 140669. https://doi.org/10.1016/j.bbapap.2021.140669.

[7] J.A. Swenberg, K. Lu, B.C. Moeller, L. Gao, P.B. Upton, J. Nakamura, T.B. Starr, Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment, Toxicol. Sci. 120 (2011) S130–S145. https://doi.org/10.1093/toxsci/kfq371.

[8] M.M. Greenberg, Abasic and Oxidized Abasic Site Reactivity in DNA: Enzyme Inhibition, Cross-Linking, and Nucleosome Catalyzed Reactions, Acc. Chem. Res. 47 (2014) 646–655. https://doi.org/10.1021/ar400229d.

[9] M.S. Demott, E. Beyret, D. Wong, B.C. Bales, J.T. Hwang, M.M. Greenberg, B. Demple, Covalent trapping of human DNA polymerase β by the oxidative DNA lesion 2-deoxyribonolactone, J. Biol. Chem. 277 (2002) 7637–7640. https://doi.org/10.1074/jbc.C100577200.

[10] J.L. Quiñones, U. Thapar, K. Yu, Q. Fang, R.W. Sobol, B. Demple, Enzyme mechanism-based, oxidative DNA–protein cross-links formed with DNA polymerase β in vivo, Proc. Natl. Acad. Sci. 112 (2015) 8602–8607. https://doi.org/10.1073/pnas.1501101112.

[11] J.T. Sczepanski, R.S. Wong, J.N. McKnight, G.D. Bowman, M.M. Greenberg, Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle, Proc. Natl. Acad. Sci. 107 (2010) 22475–22480. https://doi.org/10.1073/pnas.1012860108.

[12] C. Zhou, J.T. Sczepanski, M.M. Greenberg, Mechanistic Studies on Histone Catalyzed Cleavage of Apyrimidinic/Apurinic Sites in Nucleosome Core Particles, J. Am. Chem. Soc. 134 (2012) 16734–16741. https://doi.org/10.1021/ja306858m.

[13] W. Xu, R.M. Boyd, M.O. Tree, F. Samkari, L. Zhao, Mitochondrial transcription factor A promotes DNA strand cleavage at abasic sites, Proc. Natl. Acad. Sci. 116 (2019) 17792–17799. https://doi.org/10.1073/pnas.1911252116.

[14] B. de Graaf, A. Clore, A.K. McCullough, Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks, DNA Repair (Amst). 8 (2009) 1207–1214. https://doi.org/10.1016/j.dnarep.2009.06.007.

[15] B. Vaz, M. Popovic, J.A. Newman, J. Fielden, H. Aitkenhead, S. Halder, A.N. Singh, I. Vendrell, R. Fischer, I. Torrecilla, N. Drobnitzky, R. Freire, D.J. Amor, P.J. Lockhart, B.M. Kessler, G.W. McKenna, O. Gileadi, K. Ramadan, Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair, Mol. Cell. 64 (2016) 704–719. https://doi.org/10.1016/j.molcel.2016.09.032.

[16] J. Stingele, R. Bellelli, S.J. Boulton, Mechanisms of DNA–protein crosslink repair, Nat. Rev. Mol. Cell Biol. 18 (2017) 563–573. https://doi.org/10.1038/nrm.2017.56.

[17] H. Zhang, Y. Xiong, J. Chen, DNA–protein cross-link repair: what do we know now?, Cell Biosci. 10 (2020) 3. https://doi.org/10.1186/s13578-019-0366-z.

[18] Z.E.R. Newby, J.D. O'Connell, F. Gruswitz, F.A. Hays, W.E.C. Harries, I.M. Harwood, J.D. Ho, J.K. Lee, D.F. Savage, L.J.W. Miercke, R.M. Stroud, A general protocol for the crystallization of membrane proteins for X-ray structural investigation, Nat. Protoc. 4 (2009) 619–637. https://doi.org/10.1038/nprot.2009.27.

[19] T.W. Guo, A. Bartesaghi, H. Yang, V. Falconieri, P. Rao, A. Merk, E.T. Eng, A.M. Raczkowski, T. Fox, L.A. Earl, D.J. Patel, S. Subramaniam, Cryo-EM Structures Reveal

Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex, Cell. 171 (2017) 414-426.e12. https://doi.org/10.1016/j.cell.2017.09.006.

[20] T. Sugiki, N. Kobayashi, T. Fujiwara, Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists, Comput. Struct. Biotechnol. J. 15 (2017) 328–339. https://doi.org/10.1016/j.csbj.2017.04.001.

[21] F.J. O'Reilly, J. Rappsilber, Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology, Nat. Struct. Mol. Biol. 25 (2018) 1000–1008. https://doi.org/10.1038/s41594-018-0147-0.

[22] A. Stützer, L.M. Welp, M. Raabe, T. Sachsenberg, C. Kappert, A. Wulf, A.M. Lau, S.-S. David, A. Chernev, K. Kramer, A. Politis, O. Kohlbacher, W. Fischle, H. Urlaub, Analysis of protein-DNA interactions in chromatin by UV induced cross-linking and mass spectrometry, Nat. Commun. 11 (2020) 5250. https://doi.org/10.1038/s41467-020-19047-7.

[23] A. Reim, R. Ackermann, J. Font-Mateu, R. Kammel, M. Beato, S. Nolte, M. Mann, C. Russmann, M. Wierer, Atomic-resolution mapping of transcription factor-DNA interactions by femtosecond laser crosslinking and mass spectrometry, Nat. Commun. 11 (2020) 3019. https://doi.org/10.1038/s41467-020-16837-x.

[24] J.W. Bae, S.C. Kwon, Y. Na, V.N. Kim, J.-S. Kim, Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution, Nat. Struct. Mol. Biol. 27 (2020) 678–682. https://doi.org/10.1038/s41594-020-0436-2.

[25] C. Schmidt, K. Kramer, H. Urlaub, Investigation of protein–RNA interactions by mass spectrometry—Techniques and applications, J. Proteomics. 75 (2012) 3478–3494. https://doi.org/10.1016/j.jprot.2012.04.030.

[26] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372. https://doi.org/10.1038/nbt.1511.

[27] N. Kalisman, C.M. Adams, M. Levitt, Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling, Proc. Natl. Acad. Sci. 109 (2012) 2884–2889. https://doi.org/10.1073/pnas.1119472109.

[28] J.K. Watts, A. Katolik, J. Viladoms, M.J. Damha, Studies on the hydrolytic stability of 2′-fluoroarabinonucleic acid (2′F-ANA), Org. Biomol. Chem. 7 (2009) 1904. https://doi.org/10.1039/b900443b.

[29] A. Rubio-Cosials, J.F. Sydow, N. Jiménez-Menéndez, P. Fernández-Millán, J. Montoya, H.T. Jacobs, M. Coll, P. Bernadó, M. Solà, Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter, Nat. Struct. Mol. Biol. 18 (2011) 1281–1289. https://doi.org/10.1038/nsmb.2160.

[30] H.B. Ngo, J.T. Kaiser, D.C. Chan, The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA, Nat. Struct. Mol. Biol. 18 (2011) 1290–1296. https://doi.org/10.1038/nsmb.2159.

[31] B. Hiller, M. Achleitner, S. Glage, R. Naumann, R. Behrendt, A. Roers, Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity, J. Exp. Med. 209 (2012) 1419–1426. https://doi.org/10.1084/jem.20120876.

[32] K.D. Koh, S. Balachander, J.R. Hesselberth, F. Storici, Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA, Nat. Methods. 12 (2015) 251–257. https://doi.org/10.1038/nmeth.3259.

[33] J.S. Williams, T.A. Kunkel, Ribonucleotides in DNA: Origins, repair and consequences, DNA Repair (Amst). 19 (2014) 27–37. https://doi.org/10.1016/j.dnarep.2014.03.029.

[34] Y. Li, R.R. Breaker, Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2′-Hydroxyl Group, J. Am. Chem. Soc. 121 (1999) 5364–5372. https://doi.org/10.1021/ja990592p.

[35] R. Savva, K. McAuley-Hecht, T. Brown, L. Pearl, The structural basis of specific base-excision repair by uracil–DNA glycosylase, Nature. 373 (1995) 487–493. https://doi.org/10.1038/373487a0.

[36] B.D. Freudenthal, W.A. Beard, M.J. Cuneo, N.S. Dyrkheeva, S.H. Wilson, Capturing snapshots of APE1 processing DNA damage, Nat. Struct. Mol. Biol. 22 (2015) 924–931. https://doi.org/10.1038/nsmb.3105.

[37] D.J. López, J.A. Rodríguez, S. Bañuelos, Molecular Mechanisms Regulating the DNA Repair Protein APE1: A Focus on Its Flexible N-Terminal Tail Domain, Int. J. Mol. Sci. 22 (2021) 6308. https://doi.org/10.3390/ijms22126308.

[38] X.M. Lam, W.G. Lai, E.K. Chan, V. Ling, C.C. Hsu, Site-Specific Tryptophan Oxidation Induced by Autocatalytic Reaction of Polysorbate 20 in Protein Formulation, Pharm. Res. 28 (2011) 2543–2555. https://doi.org/10.1007/s11095-011-0482-x.

[39] A. Rubio-Cosials, F. Battistini, A. Gansen, A. Cuppari, P. Bernadó, M. Orozco, J. Langowski, K. Tóth, M. Solà, Protein Flexibility and Synergy of HMG Domains Underlie U-Turn Bending of DNA by TFAM in Solution, Biophys. J. 114 (2018) 2386–2396. https://doi.org/10.1016/j.bpj.2017.11.3743.

[40] G. Farge, N. Laurens, O.D. Broekmans, S.M.J.L. van den Wildenberg, L.C.M. Dekker, M. Gaspari, C.M. Gustafsson, E.J.G. Peterman, M. Falkenberg, G.J.L. Wuite, Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A, Nat. Commun. 3 (2012) 1013. https://doi.org/10.1038/ncomms2001.

[41] A. Cuppari, P. Fernández-Millán, F. Battistini, A. Tarrés-Solé, S. Lyonnais, G. Iruela, E. Ruiz-López, Y. Enciso, A. Rubio-Cosials, R. Prohens, M. Pons, C. Alfonso, K. Tóth, G. Rivas, M. Orozco, M. Solà, DNA specificities modulate the binding of human transcription factor A to mitochondrial DNA control region, Nucleic Acids Res. 47 (2019) 6519–6537. https://doi.org/10.1093/nar/gkz406.

[42] T.S. Wong, S. Rajagopalan, S.M. Freund, T.J. Rutherford, A. Andreeva, F.M. Townsley, M. Petrovich, A.R. Fersht, Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53, Nucleic Acids Res. 37 (2009) 6765–6783. https://doi.org/10.1093/nar/gkp750.

[43] I. Ivancová, D.-L. Leone, M. Hocek, Reactive modifications of DNA nucleobases for labelling, bioconjugations, and cross-linking, Curr. Opin. Chem. Biol. 52 (2019) 136–144. https://doi.org/10.1016/j.cbpa.2019.07.007.

[44] K. Kramer, T. Sachsenberg, B.M. Beckmann, S. Qamar, K.-L. Boon, M.W. Hentze, O. Kohlbacher, H. Urlaub, Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins, Nat. Methods. 11 (2014) 1064–1070. https://doi.org/10.1038/nmeth.3092.

[45] E.C. Urdaneta, B.M. Beckmann, Fast and unbiased purification of RNA-protein complexes after UV cross-linking, Methods. 178 (2020) 72–82. https://doi.org/10.1016/j.ymeth.2019.09.013.

[46] J. Stingele, M.S. Schwarz, N. Bloemeke, P.G. Wolf, S. Jentsch, A DNA-Dependent Protease Involved in DNA-Protein Crosslink Repair, Cell. 158 (2014) 327–338. https://doi.org/10.1016/j.cell.2014.04.053.

Chapter 3. Facile Preparation of Model DNA Interstrand Cross-Link Repair Intermediates Using Ribonucleotide-Containing DNA

3.1 Abstract

DNA interstrand cross-links (ICLs) are lesions with a covalent bond formed between DNA strands. ICLs are extremely toxic to cells because they prevent the separation of the two strands, which are necessary for the genetic interpretation of DNA. ICLs are repaired via Fanconi anemia and replication-independent pathways. The formation of so-called unhooked repair intermediates via a dual strand incision flanking the ICL site on one strand is an essential step in nearly all ICL repair pathways. Recently, ICLs derived from endogenous sources, such as those from ubiquitous DNA lesions, abasic (AP) sites, have emerged as an important class of ICLs. Despite the earlier efforts in preparing AP-ICLs in high yield using nucleotide analogs, little information is available for preparing AP-ICL unhooked intermediates with varying lengths of overhangs. In this study, we devise a simple approach to prepare model ICL unhooked intermediates derived from AP sites. We exploited the alkaline lability of ribonucleotides (rNMPs) and the high cross-linking efficiency between an AP lesion and a nucleotide analog, 2-aminopurine, via reductive amination. We designed chimeric DNA/RNA substrates with rNMPs flanking the cross-linking residue (2-aminopurine) to facilitate subsequent strand cleavage under our optimized conditions. Mass spectrometric analysis and primer extension assays confirmed the structures of ICLs substrates. The method is straightforward, requires no

synthetic chemistry expertise, and should be broadly accessible to all researchers in the DNA repair community. For step-by-step descriptions of the method, please refer to the companion MethodsX paper [1].

3.2 Introduction

DNA is susceptible to chemical and physical factors, generating a plethora of DNA lesions [2]. DNA interstrand cross-links (ICLs) are among the most deleterious DNA lesions, because the covalent linkage between the two complementary strands can alter the structure and enzymatic read-out of DNA, potentially leading to genomic instability and cell death [3–7]. It is estimated that one ICL lesion is sufficient to kill a bacterial or yeast cell [8], and approximately 40 ICLs can kill a repair-deficient mammalian cell [9]. ICLs can be repaired via replication-coupled and replication-independent pathways [7]. In almost all known ICL repair pathways, the generation of so-called unhooked repair intermediates by dual incision of ICLs via endonucleolytic activities is an essential step. Critical to the endonucleolytic activities and initiation of ICL repair are endonuclease XPF-ERCC1 and the scaffolding protein SLX4 [4,10,11], along with a number of other endonucleases, such as FAN1, MUS81-EME1, SLX1, and SNM1A[4,10,12–15]. The resulting ICL remnant can be bypassed by translesion synthesis (TLS) DNA polymerases followed by homologous recombination (HR) in replication-dependent ICL repair pathways [7].

Historically, ICLs have been studied in the context of chemical warfare and chemotherapeutic agents such as nitrogen mustards, mitomycin C, psoralens, and

cisplatin [16–19]. Recently, ICLs derived from endogenous chemicals have emerged as an important class of ICLs [6]. In particular, abasic (AP) sites, one of the most abundant endogenous DNA lesions, have been shown to form ICLs with a guanine residue on the 5'-side of the nucleotide opposite to an AP site on the complementary strand [20,21], or an adenine on the 3'-side of the nucleotide opposite to an AP site on the complementary strand [22,23]. Subsequent studies using DNA substrates containing ICLs derived from an AP lesion and an adenine residue discovered a specialized NEIL3-mediated repair pathway and supported circumstantially the biological relevance of AP-derived ICLs (AP-ICLs) [24–26].

The advancement of the understanding of biological consequence and repair pathways of ICLs is driven in part by the development of approaches to prepare site-specifically modified oligodeoxynucleotides containing a chemically defined ICL. However, the preparation of ICL-containing substrates is not trivial. Traditional methods using cross-linking agents, such as formaldehyde or nitrogen mustards, to react with a double-stranded (ds) DNA substrate suffer from poor site selectivity and low yield [27–29]. The resulting mixtures of cross-linked products require laborious purification and can complicate data interpretation. Several solid-phase synthetic methods have been developed over the years [30–34]; nonetheless, the requirement of stringent reaction conditions and synthetic chemistry expertise has hindered broader applications of these methods. A number of elegant methods have been devised by Gates and colleagues to prepare AP-ICLs [22,23,35,36], and two methods achieved high cross-linking yield using

nucleotide analogs, N^4-amino-2'-deoxycytidine [35] and 2-aminopurine (P) [37]. The dearth of information on preparing AP-ICL unhooked intermediates and the need for additional mechanistic insights into AP-ICL repair motivated us to develop a simple and accessible approach to prepare such substrates.

Herein, we developed a simple approach to prepare site-specifically modified model AP-ICL substrates to mimic unhooked repair intermediates. We exploited the alkaline lability of ribonucleotides (rNMPs) in chimeric DNA/RNA substrates [38] and the high cross-linking efficiency between AP and a nucleotide analog (P) via reductive animation [37]. We designed rNMPs on the 3'-side and 5'-side of the cross-linking residue (P) to facilitate strand cleavage under alkaline conditions. We optimized the rNMP cleavage conditions using NaOH or RNase HII to prepare two model ICL substrates with five or nine nucleotide overhangs. Primer extension assays using a prototypical TLS DNA polymerase and a replicative DNA polymerase demonstrate that both substrates are strong blocks to bypass synthesis. All DNA substrates used in this study are commercially available, which ensures the accessibility of the approach by most labs interested in ICL repair or nucleic acid modifications in general.

3.3 Materials and methods

3.3.1 Materials

Chemicals were purchased from Fisher Scientific and are either analytical grade or molecular biology grade. Uracil-DNA glycosylase (UDG), RNase HII and *B. subtilis* DNA

polymerase I (Pol I, lacking the exonuclease activity) were from New England Biolabs. *S. solfataricus* P2 DNA polymerase IV (Dpo4) was expressed and purified based on our previous protocol [39]. HPLC-purified unmodified and modified oligodeoxynucleotides were purchased from Integrated DNA Technologies.

3.3.2 Preparation of ICL substrates

AP site-containing oligodeoxynucleotides were prepared by excising the uracil residue from a deoxyuridine-containing oligodeoxynucleotide with UDG followed by phenol/chloroform extraction, as described previously [40]. An AP-containing DNA oligodeoxynucleotide was annealed with a P-containing oligodeoxynucleotide. The resulting DNA substrate was incubated in 750 mM NaOAc (pH 5.4) and 250 mM $NaCNBH_3$ at 37°C overnight to form the P-AP ICLs substrate precursors. The formation of ICL was monitored by 18% denaturing PAGE.

3.3.3 Strand cleavage at rNMPs

Strand cleavage reactions at rNMPs with P-AP ICL precursors were performed under NaOH or RNase HII. The NaOH reactions contained 0.1 M, 0.2 M, or 0.3 M NaOH and were conducted at 55°C for up to 2 h followed by neutralization using HCl and gel analysis or purification. The RNase HII reactions were with larger amounts of RNase HII than the manufacture suggested conditions, as specified in the **Figure 3.2** caption. For reactions with ICL-R1, the optimal condition was 10 pmol ICL-R1 and 10 units of RNase HII. For reactions with ICL-R2, the optimal condition was 10 pmol ICL-R2 and 5 units of RNase

HII at 37°C for 18 h. The manufacture defines one unit as the amount of enzyme required to yield a fluorescence signal comparable to nicking of 100 pmol of synthetic dsDNA substrate containing a rNMP near the quencher of a fluorophore/quencher pair in 30 minutes at 37°C in 1X ThermoPol Buffer.

3.3.4 Primer extension assays with ICLs

The primer extension assay was conducted with ICL1 and ICL2 that cleaved via alkaline strand cleavage reaction. The reaction was performed with 100 uM dNTPs and 1 uM ICLs with 0.5 unit/uL Pol I or 71 nM Dpo4 with 50 mM Tris-HCl pH 7.4, 5 mM DTT, 5 mM $MgCl_2$, 50 mM NaCl, 50 ug/mL BSA and 5% Glycerol at 37°C. The reaction aliquots were taken at various times and quenched with 20 mM EDTA (pH 9.0) in 95% (v/v) formamide.

3.4 Results and discussion

3.4.1 Design of ICL substrates and cross-linking reactions

As outlined in **Figure 3.1 a**, the preparation of model ICL substrates consists of several straightforward steps, i.e., (1) annealing two complementary single-stranded (ss) DNA oligomers and the cross-linking reaction, (2) cleaving the precursor substrate at rNMPs, and (3) purifying the model ICL by PAGE. We designed oligodeoxynucleotides containing a P modification and two rNMPs on the 5'- and 3'-side of P (**Figure 3.1 e**, the top strand of ICL-R1 and ICL-R2) to facilitate subsequent strand cleavage at these sites via alkaline transesterification reactions. Although Couvé et al. had used chimeric RNA/DNA

oligonucleotides followed by RNase A digestion to construct psoralen-induced ICL repair intermediates [41], the applicability of their method has not be tested for other ICL substrates and digestion by RNase A could be limited by the cross-linking chemistry, the spacing of rNMP residues, and steric hinderance. In this study, we seek to develop a method based on the alkaline lability of rNMPs to generate strand cleavage at rNMPs sites (**Figure 3.1 b**). We exploited the high cross-linking yield between P and AP sites (**Figure 3.1 c**) [37] to generate covalent cross-links between two complementary oligonucleotides, which are referred to as ICL precursors. The sequences of oligomers containing P or AP modifications are shown in **Figure 3.1 e**. The ICL precursors with rNMPs at different locations are referred to as ICL-R1 and ICL-R2 (**Figure 3.1 e**). The cross-linking yield was approximately 90% after an overnight reaction for ICL-R1 and ICL-R2 (**Figure 3.1 d**), indicating that the presence of rNMPs did not alter the cross-linking reactions. ICL-R1 and ICL-R2 were cleaved at rNMPs to generate model ICL substrates to mimic unhooked repair intermediates of different sizes. One ICL substrate contains a 5-nt overhang cross-linked to the AP site, and the other contains a 9-nt overhang (hereinafter referred to as ICL1 and ICL2, respectively).

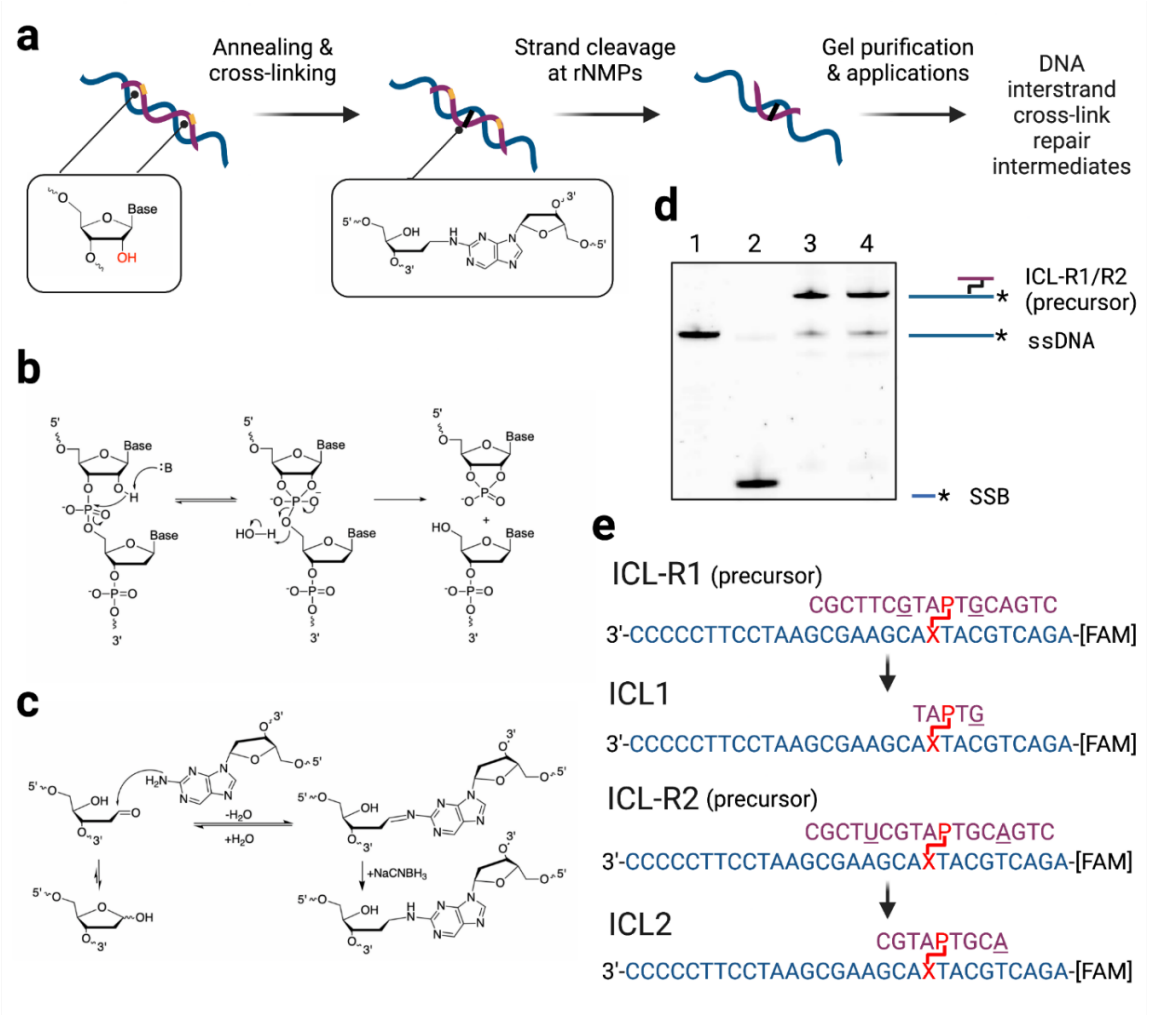


Figure 3.1 Design and preparation of ICL substrates. (**a**) Preparation of model ICL mimics. The workflow consists of annealing and P-AP crosslinking reaction, rNMP cleavage, and PAGE purification. (**b**) rNMP cleavage via alkaline transesterification reactions. (**c**) Cross-linking reaction between P and AP sites to form P-AP ICLs under reductive amination. (**d**) 18% denaturing PAGE analysis of the product after the P-AP cross-linking reaction in the presence of 250 mM $NaCNBH_3$. Lane 1 is 30-mer uracil-containing DNA oligomer (the sequence is shown in the bottom strand of ICL-R1 in Fig 1e); lane 2 is 9-mer oligomer product after NaOH-induced cleavage at the AP site with AP-containing 30-mer DNA oligomer; lane 3 and 4 show cross-linking yields are 90% and 87% for ICL-R1 and ICL-R2, respectively. (**e**) Sequences of rNMP-containing ICL precursors and model ICL structures upon cleavage at rNMPs.

3.4.2 NaOH-catalyzed cleavage at rNMPs

We optimized two types of rNMP cleavage conditions, i.e., NaOH- and RNase HII-catalyzed reactions. rNMP cleavage under 0.3 M NaOH at 55 °C for 2 h has been used to map the rNMPs embedded in the genomic DNA [42,43]. We first investigated the suitability of this condition with ICL-R1 and ICL-R2. PAGE analysis revealed that the highest yield of the desired ICL products peaked at 60% (indicated by green squares in **Figure 3.2 c, Figure 3.3**) after a 1-h reaction followed by further conversion to side products (indicated by red triangle in **Figure 3.2 c**). After a 2-h reaction, the desired ICL products accounted for only 27% with the majority of products being cleaved fragments. Such products are likely due to the cleavage of the phosphoester bond 5' of the AP lesion, considering that the migration is slower than the 5'-remnant after cleavage at AP sites and the difference in migration patterns for products formed in ICL-R1 and ICL-R2 reactions. To optimize the yield of ICL1 and ICL2, we decreased the concentration of NaOH, i.e., 0.1 M and 0.2 M NaOH, at 55°C for varying times (**Figures 3.2 a and 3.2 c, Figure 3.3**). We compared the yield of the desired ICL products under different reaction conditions (**Figure 3.2 c**) and selected the condition under 0.2 M NaOH at 55°C for 60 min to produce the highest yield of ICL1 (62%) and ICL2 (67%) (**Figure 3.2 c**). Further, ICL1 and ICL2 were purified by denaturing PAGE (**Figure 3.4**) to obtain products with 95% and 99% purities (**Figure 3.4**), respectively. NaOH-mediated cleavage yields a primarily product with a phosphate group at the 5'-end of the unhooked strand, as evidenced by phosphatase treatment (**Figure 3.4**) and mass spectrometry analysis (**Figure 3.5**).

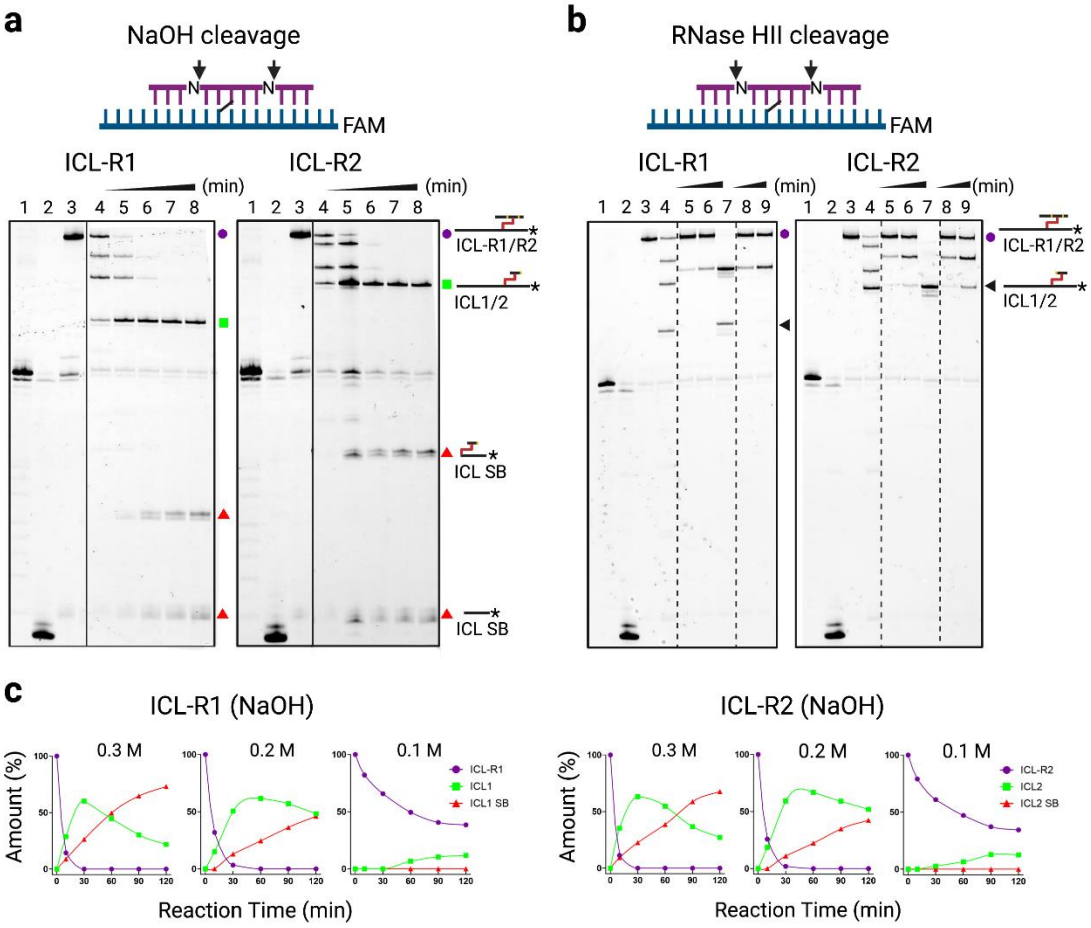

a
NaOH cleavage
N
N
FAM
ICL-R1
ICL-R2
(min)
1 2 3 4 5 6 7 8
ICL-R1/R2
ICL1/2
ICL SB
ICL SB
b
RNase HII cleavage
N
N
FAM
ICL-R1
ICL-R2
(min)
1 2 3 4 5 6 7 8 9
ICL-R1/R2
ICL1/2
c
ICL-R1 (NaOH)
0.3 M
0.2 M
0.1 M
Amount (%)
ICL-R1
ICL1
ICL1 SB
Reaction Time (min)
ICL-R2 (NaOH)
0.3 M
0.2 M
0.1 M
Amount (%)
ICL-R2
ICL2
ICL2 SB
Reaction Time (min)

Figure 3.2 Time course of rNMP cleavage reactions with ICL-R1 and ICL-R2. (**a**) Reactions with ICL-R1 or ICL-R2 under 0.2 M NaOH. Black arrows indicate locations of phosphoester bond cleavage. Lane 1 is fluorescein-labeled template DNA oligomer with a deoxyuridine; lane 2 is the NaOH-cleaved product of AP-containing template DNA oligomer; lane 3 is ICL-R1 or ICL-R2 precursor. Lanes 4-8 are reactions with ICL-R1 or ICL-R2 under 0.2 M NaOH for 10, 30, 60, 90, and 120 min. The ICL precursors are indicated by purple spheres. The desired ICL products are indicated by green squares. Side products correlating to strand cleavage at AP sites or degradation of the P-AP cross-links are indicated by read triangles. Reaction intermediates that migrated in between the ICL precursors and the desired ICL products correspond to cleavage at one rNMP residue. (**b**) Strand cleavage reactions with ICL-R1 or ICL-R2 using RNase HII. Lanes 1-3 contain the same samples as described in (**a**). Lane 4 represents cleavage products of ICL-R1 or ICL-R2 under 0.3 M NaOH for 10 min. Lanes 5-7 are reactions with 10 pmol ICL-R1 or ICL-R2 incubated with 5 units RNase HII for 1, 3, and 18 h. Lane 8 and 9 are reactions with 10 pmol ICL-R1 or ICL-R2 incubated with 10 units RNase HII for 1 and 3 h. Black triangles indicate the desired ICL products. (**c**) Quantification of reaction products from the denaturing PAGE analysis in **Figure 3.1 a and Figure 3.3**. The percent intensity of starting materials in lane 3 were defined as 100%.

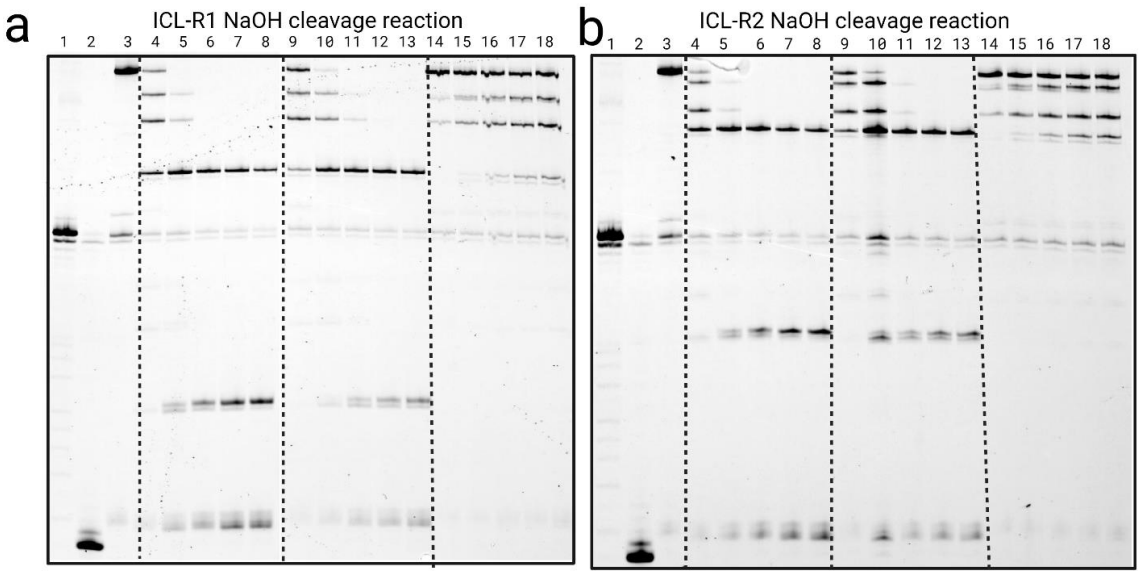


Figure 3.3 Reaction time course of the rNMP cleavage reactions with ICL-R1 and ICL-R2 under varying concentrations of NaOH. (**a**) rNMP cleavage reaction with ICL-R1 under 0.3, 0.2, and 0.1 M NaOH. Lane 1 is the fluorescein-labeled template DNA oligomer with a deoxyuridine modification. Lane 2 is the NaOH cleaved product of AP-containing template DNA oligomer. Lane 3 is the ICL-R1 precursor. Lane 4-8 are reactions with ICL-R1 under 0.3 M NaOH for 10, 30, 60, 90, and 120 min. Lane 9-13 are reactions with ICL-R1 under 0.2 M NaOH for 10, 30, 60, 90, and 120 min. Lane 14-18 are reactions with ICL-R1 under 0.1 M NaOH for 10, 30, 60, 90, and 120 min. (**b**) rNMP cleavage reaction with ICL-R2. The incubation of ICL-R2 is the same as ICL-R1 and loaded in the same sequence as in (**a**).

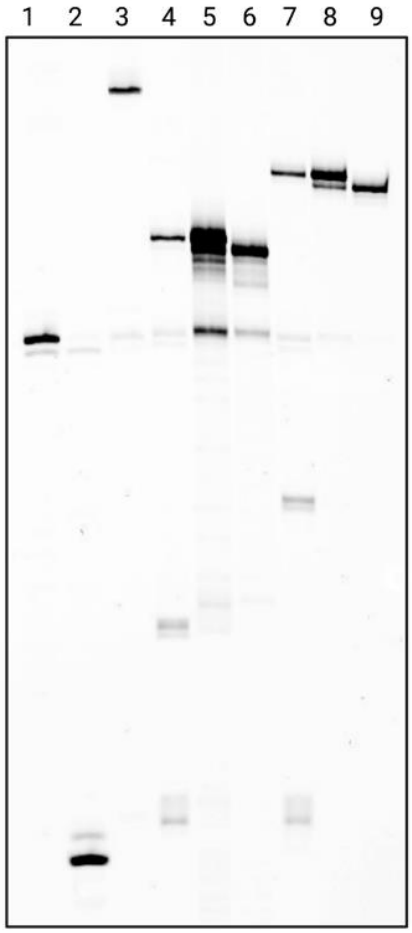


Figure 3.4 PAGE analysis of purified ICL1 and ICL2. Lane 1 is fluorescein-labeled template DNA oligomer with a deoxyuridine; lane 2 is the NaOH-cleaved product of AP-containing DNA oligomer; lane 3 is ICL-R1 precursor (prior to alkaline cleavage); lane 4 is ICL1 before purification; lane 5 is the purified ICL1 products with 95% purity; lane 6 presents products from lane 5 treated with alkaline phosphatase; lanes 7 is products of ICL2 before purification; lane 8 represents products after purification (99% purity); lane 9 represents products from lane 8 treated with alkaline phosphatase.

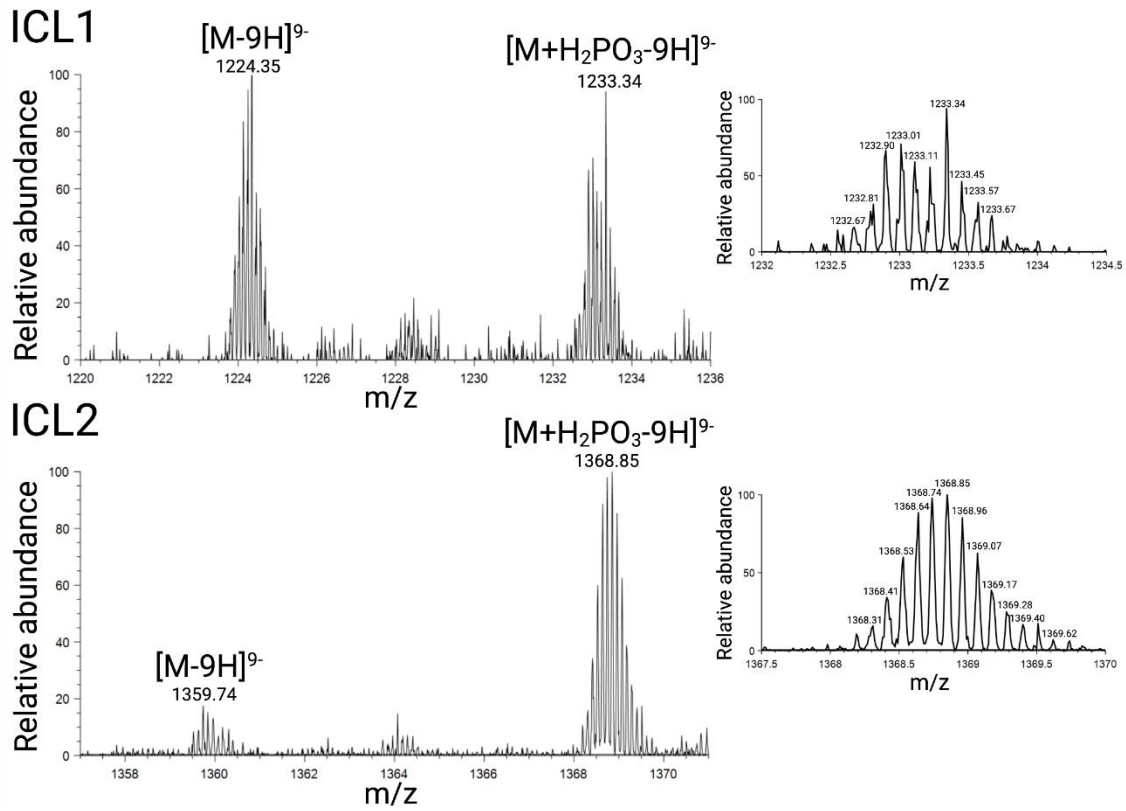


Figure 3.5 Characterization of ICL1 and ICL2 using LC-MS. The molecular weights of NaOH cleaved ICLs were confirmed with LC-MS analysis. ICLs from cleavage reactions were purified with denaturing PAGE. The desired ICL products were excised, extracted from the gel and purified with C18 cartridge. The resulting ICLs were redissolved in water. The molecular weight of ICLs were confirmed with LC-MS analysis. An Agilent Zorbax SB-C18 column (0.5 × 250 mm, 5 µm in particle size) was employed. Solvent A: water, and solvent B: methanol with 400 mM HFIP. Samples were analyzed at 8 uL/min with a 50-minute gradient staring at 5% B and increasing to 20% B in 5 min, then to 45% B in 45 min. Ion-transportation tube was set at 300°C. The ultra-zoom scan was selected to monitor the $[M-9H]^{9-}$ ion. The observed m/z of ICL1 is 1224.35 (cal. 1224.9), consistent with a molecular weight of 11027.99. The observed m/z of ICL2 is 1359.7 (cal. 1360.5), consistent with a molecular weight of 12248.19. The phosphate adducts, $[M+H_2PO_3-9H]^{9-}$, were also observed with zoomed spectra shown on the right.

3.4.3 RNase H-catalyzed cleavage at rNMPs

RNase H enzymes cleave the RNA in RNA/DNA hybrids and remove ribonucleotides from DNA to maintain the stability of genome [44,45]. RNase HII selectively cleaves the phosphodiester bond 5' to the ribonucleotides (**Figure 3.6 a**), generating terminal nucleotide structures different from the NaOH cleavage reaction (**Figure 3.2 b and Figure 3.6 b**). Indeed, when comparing products in NaOH-catalyzed reactions with those from RNase HII-mediated cleavage, a clear difference was observed in the product migration patterns (lane 4 vs. lane 7 with ICL1 and ICL2 in **Figure 3.2 b**). RNase HII exhibited better specificity relative to reactions with NaOH, as evidenced by the absence of shorter fragments. Nonetheless, the reactions required a considerable amount of RNase HII (approximately 10-fold higher than the manufacturer suggested stoichiometry) and longer reaction time (18 h) to achieve complete cleavage at both rNMPs. In particular, the reaction was sensitive to the spacing of the two rNMPs. Under the same condition, the cleavage of ICL-R2 was more efficient than that of ICL-R1. The ICL-R2 reaction was nearly complete after an overnight reaction, whereas the ICL-R1 reaction yielded only 25% of the desired product ICL1 (lane 7 of **Figure 3.2 b** left panel). We were able to improve the product yield with ICL1 by using more RNase HII (**Figure 3.6 c**). Together, our results demonstrate that NaOH-mediated rNMP cleavage serves as an efficient and robust method to produce the desired model ICL substrates, whereas RNase H-catalyzed reactions offer a means to generate ICL products in a higher yield with optimization needed for different substrates. Notably, we tested the cleavage efficiency

at rNMPs with a reaction developed for cleaving the rNMP sites in DNA protein cross-links (DPCs) [46]. The reaction exhibited low cleavage efficiency with considerable amounts of side products (**Figure 3.6 d**). Therefore, the condition reaction was not pursued.

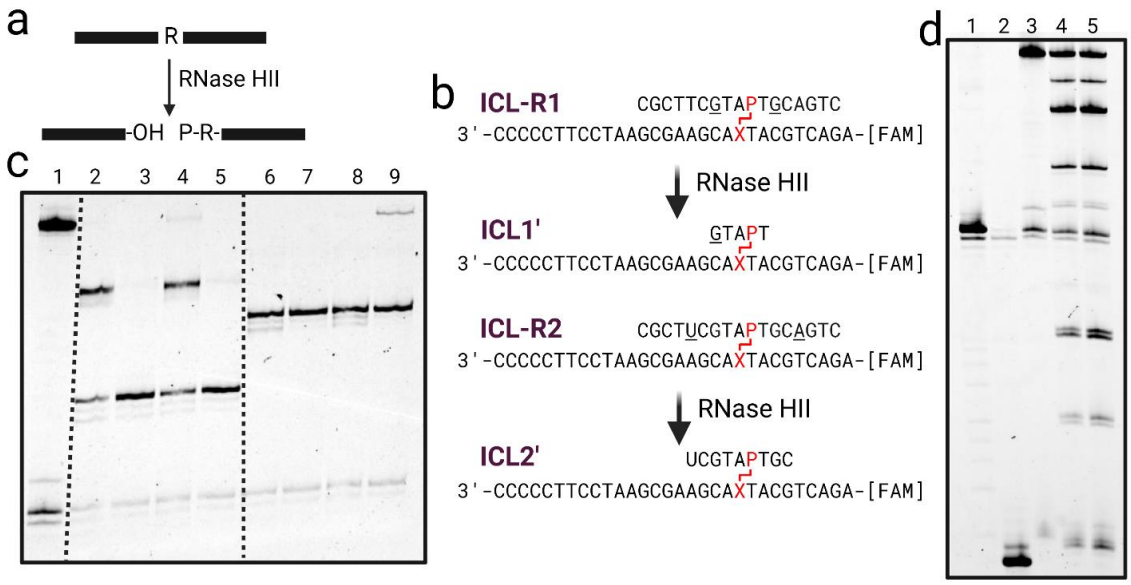


Figure 3.6 RNase HII cleavage reaction.(**a**) Schematic illustration of the mechanism of ribonucleotide cleavage by RNase HII. RNase HII cleaves the 5' phosphate backbone of ribonucleotide embedded in double strand DNA, leaving 3'-OH and 5'-phosphate group. (**b**) Strand cleavage reaction of ICL-R1 and ICL-R2 by RNase HII. RNase HII cleaves the ICL-R1 and generates ICL1'. RNase HII cleaves ICL-R2 and generates ICL2'. The total number of nucleotides in ICL1' and ICL2' are the same as those in ICL1 and ICL2; however, the terminal nucleotides are different. (**c**) Overnight RNase HII cleavage reaction of ICL-R1 and ICL-R2. In lane 1, the top band is the ICL-R1 marker (ICL-R2 migrates to the same place as ICL-R1). Lane 2 is 10 pmol ICL-R1 incubated overnight with 5 units of RNase HII. Lane 3 is 10 pmol ICL-R1 incubated overnight with 25 units of RNase HII. Lane 4 is 10 pmol ICL-R1 incubated overnight with 5 units of RNase HII and 10 mM $MgCl_2$. Lane 5 is overnight incubation of 10 pmol ICL-R1 incubated with 25 units of RNase HII and 10 mM $MgCl_2$. Lanes 6-9 are the ICL-R2 incubated with RNase HII under the same condition as the ICL-R1 and loaded with the same sequence to the gel. (**d**) Incubation of ICL-R1 under Cleave R condition. Lane 1 is the fluorescein-labeled template DNA with a deoxyuridine modification. Lane 2 is the NaOH cleaved product of AP-containing template DNA oligomer. Lane 3 is the ICL-R1. Lanes 4 and 5 are ICL-R1 incubated under Cleave R condition for 19 and 38 hours.

3.4.4 DNA polymerase bypass reactions with model ICLs

To test the replication-blocking effects of ICL1 and ICL2, primer extension assays were performed with primer-template substrates containing ICL1 and ICL2 (**Figure 3.7 a**) and a model TLS DNA polymerase, Dpo4, or a replicative DNA polymerase, *B. subtilis* Pol I. The primer-template substrates containing ICL1 and ICL2 mimic the replication intermediates prior to bypass by TLS polymerases. Such substrates have been used by a number of earlier studies to understand the role of specific TLS polymerases in bypassing ICL lesions [47–50]. With unmodified substrates, both enzymes yielded full-length extension products. Dpo4 showed different reaction efficiency when it was incubated with two ICL-containing substrates (**Figures 3.7 b**). With ICL1, DNA synthesis by Dpo4 stalled 2 nt before the cross-linking site. ICL2 exhibited a stronger inhibitory effect, with stalling occurring at 4 nt to 5 nt prior to the cross-linking site, consistent with the longer overhang present in ICL2 compared to ICL1. Reactions with Pol I produced extended primers with 7 nt added within 3 min with both ICL1 and ICL2-containing substrates (**Figure 3.7 c**), indicating that Pol I can extend the primer up to 1 nt prior to the cross-linking site but cannot bypass the ICL lesion. Longer reaction times up to 30 min did not improve the primer-extension reactions by Pol I. Compared to reactions with Dpo4, the overall longer extension products observed with Pol I are consistent with the known processivity and strand displacement activity of Pol I [51]. Together, data with two model DNA polymerases support the successful preparation of two model ICL structures and demonstrate strong replication-blocking effects of AP-derived model ICL substrates.

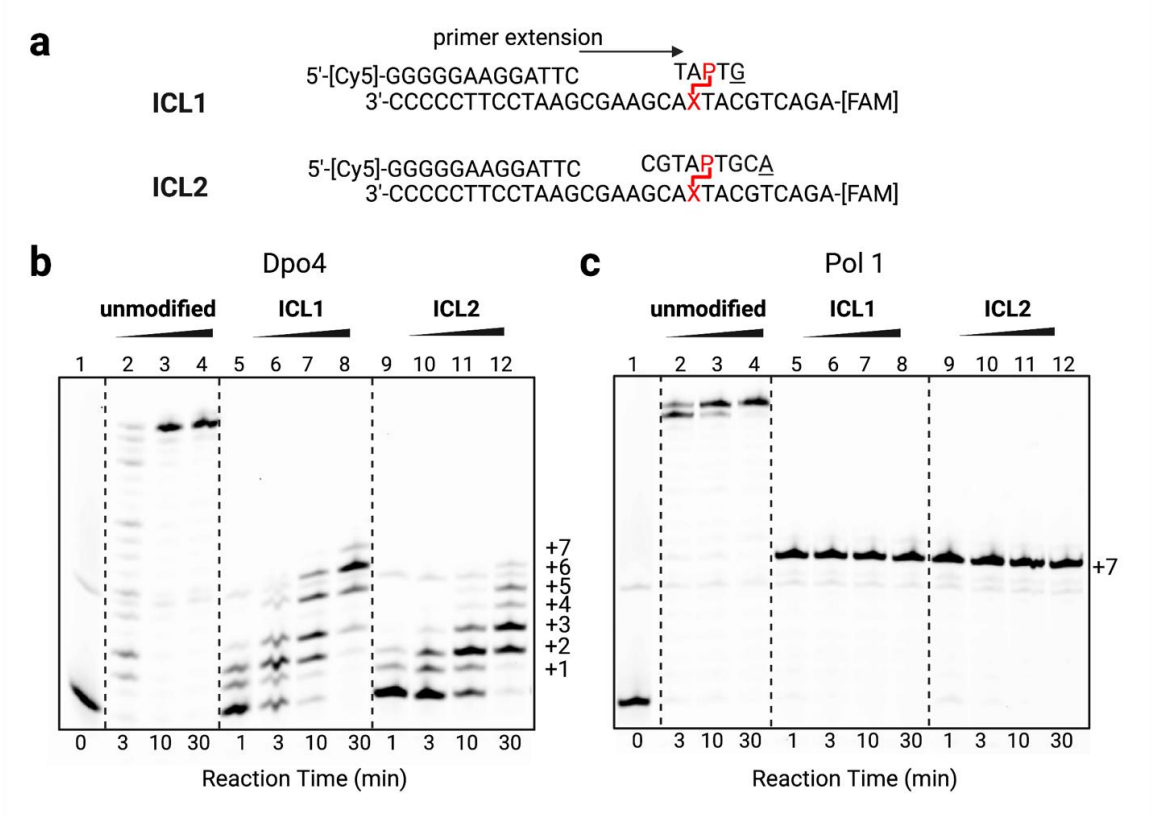


Figure 3.7 Primer-extension reactions with ICL1 and ICL2 substrates catalyzed by Dpo4 and Pol I. (a) Primer-template substrates containing ICL1 and ICL2 for primer extension reactions. (b) Reactions with Dpo4 and an unmodified substrate, ICL1 or ICL2. (c) Reactions with Pol I and an unmodified substrate, ICL1 or ICL2.

3.5 Conclusion

The ICL unhooked intermediates form in nearly all replication-dependent and replication-independent ICL repair pathways. It has been shown that AP-dA ICLs are unhooked by NEIL3; however, in mammalian cells lacking NEIL3, these lesions rely on Fanconi anemia (FA) pathway [25,52], suggesting the likelihood of formation of AP-ICL unhooked intermediates. Therefore, simple and straightforward approaches to prepare such model AP-ICL substrates are warranted to facilitate the delineation of AP-ICL repair

mechanisms and the substrate specificities of different endonuclease, exonucleases, and accessory proteins [14,15].

We devised a simple method to prepare model ICL unhooked intermediates derived from AP sites. The method exploits the alkaline lability of rNMPs in chimeric DNA/RNA substrates to generate unhooked strands of different sizes. We showed that rNMP cleavage under NaOH serves as a robust means regardless of the spacing of two rNMPs, whereas cleavage using RNase H may offer a higher product yield with optimization required for specific substrates. We envision that creative modification of the method by using one or two rNMPs will aid the preparation of relevant substrates for further elucidation of AP-ICL repair mechanisms. The method can also serve as an alternative means to other methods where enzyme processing may be limited by the spacing of the residues (e.g., cleavage of deoxyuridine by UDG) [53].

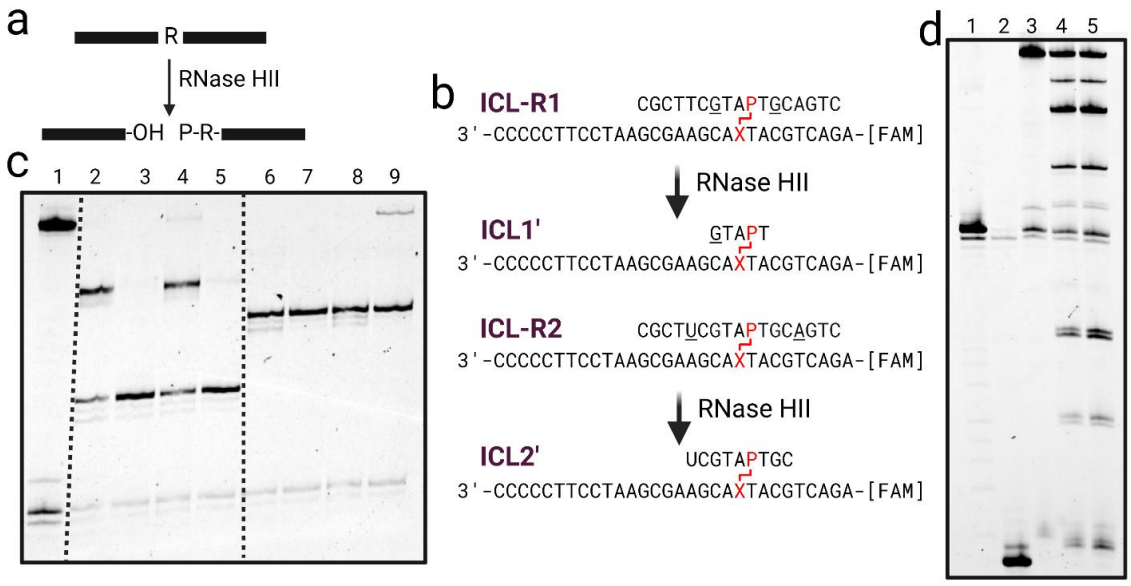


Figure 3.8 RNase HII cleavage reaction. (a) Schematic illustration of the mechanism of ribonucleotide cleavage by RNase HII. RNase HII cleaves the 5' phosphate backbone of ribonucleotide embedded in double strand DNA, leaving 3'-OH and 5'-phosphate group. (b) Strand cleavage reaction of ICL-R1 and ICL-R2 by RNase HII. RNase HII cleaves the ICL-R1 and generates ICL1'. RNase HII cleaves ICL-R2 and generates ICL2'. The total number of nucleotides in ICL1' and ICL2' are the same as those in ICL1 and ICL2; however, the terminal nucleotides are different. (c) Overnight RNase HII cleavage reaction of ICL-R1 and ICL-R2. In lane 1, the top band is the ICL-R1 marker (ICL-R2 migrates to the same place as ICL-R1). Lane 2 is 10 pmol ICL-R1 incubated overnight with 5 units of RNase HII. Lane 3 is 10 pmol ICL-R1 incubated overnight with 25 units of RNase HII. Lane 4 is 10 pmol ICL-R1 incubated overnight with 5 units of RNase HII and 10 mM $MgCl_2$. Lane 5 is overnight incubation of 10 pmol ICL-R1 incubated with 25 units of RNase HII and 10 mM $MgCl_2$. Lanes 6-9 are the ICL-R2 incubated with RNase HII under the same condition as the ICL-R1 and loaded with the same sequence to the gel. (d) Incubation of ICL-R1 under Cleave R condition. Lane 1 is the fluorescein-labeled template DNA with a deoxyuridine modification. Lane 2 is the NaOH cleaved product of AP-containing template DNA oligomer. Lane 3 is the ICL-R1. Lanes 4 and 5 are ICL-R1 incubated under Cleave R condition for 19 and 38 hours.

3.6 References

[1] J. Tang, L. Zhao, Preparation of DNA interstrand cross-link repair intermediates induced by abasic sites, MethodsX. 9 (2022) 101687. https://doi.org/10.1016/j.mex.2022.101687.

[2] R. De Bont, Endogenous DNA damage in humans: a review of quantitative data, Mutagenesis. 19 (2004) 169–185. https://doi.org/10.1093/mutage/geh025.

[3] C. Clauson, O.D. Scharer, L. Niedernhofer, Advances in Understanding the Complex Mechanisms of DNA Interstrand Cross-Link Repair, Cold Spring Harb. Perspect. Biol. 5 (2013) a012732–a012732. https://doi.org/10.1101/cshperspect.a012732.

[4] J. Zhang, J.C. Walter, Mechanism and regulation of incisions during DNA interstrand cross-link repair, DNA Repair (Amst). 19 (2014) 135–142. https://doi.org/10.1016/j.dnarep.2014.03.018.

[5] U. Roy, O.D. Schärer, Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair, DNA Repair (Amst). 44 (2016) 33–41. https://doi.org/10.1016/j.dnarep.2016.05.004.

[6] K. Housh, J.S. Jha, T. Haldar, S.B.M. Amin, T. Islam, A. Wallace, A. Gomina, X. Guo, C. Nel, J.W. Wyatt, K.S. Gates, Formation and repair of unavoidable, endogenous interstrand cross-links in cellular DNA, DNA Repair (Amst). 98 (2021) 103029. https://doi.org/10.1016/j.dnarep.2020.103029.

[7] D.R. Semlow, J.C. Walter, Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair, Annu. Rev. Biochem. 90 (2021) 107–135. https://doi.org/10.1146/annurev-biochem-080320-112510.

[8] P.D. Lawley, D.H. Phillips, DNA adducts from chemotherapeutic agents, Mutat. Res. Mol. Mech. Mutagen. 355 (1996) 13–40. https://doi.org/10.1016/0027-5107(96)00020-6.

[9] N. Magana-Schwencke, J.A. Henriques, R. Chanet, E. Moustacchi, The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair-deficient strains., Proc. Natl. Acad. Sci. 79 (1982) 1722–1726. https://doi.org/10.1073/pnas.79.6.1722.

[10] M.C. Kottemann, A. Smogorzewska, Fanconi anaemia and the repair of Watson and Crick DNA crosslinks, Nature. 493 (2013) 356–363. https://doi.org/10.1038/nature11863.

[11] M.R.G. Hodskinson, J. Silhan, G.P. Crossan, J.I. Garaycoechea, S. Mukherjee, C.M. Johnson, O.D. Schärer, K.J. Patel, Mouse SLX4 Is a Tumor Suppressor that Stimulates the Activity of the Nuclease XPF-ERCC1 in DNA Crosslink Repair, Mol. Cell. 54 (2014) 472–484. https://doi.org/10.1016/j.molcel.2014.03.014.

[12] P.J. McHugh, XPF–ERCC1: Linchpin of DNA crosslink repair, PLOS Genet. 16 (2020) e1008616. https://doi.org/10.1371/journal.pgen.1008616.

[13] S. Thongthip, M. Bellani, S.Q. Gregg, S. Sridhar, B.A. Conti, Y. Chen, M.M. Seidman, A. Smogorzewska, Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction, Genes Dev. 30 (2016) 645–659. https://doi.org/10.1101/gad.276261.115.

[14] U.B. Abdullah, J.F. McGouran, S. Brolih, D. Ptchelkine, A.H. El-Sagheer, T. Brown, P.J. McHugh, RPA activates the XPF- ERCC 1 endonuclease to initiate processing of DNA interstrand crosslinks, EMBO J. 36 (2017) 2047–2060. https://doi.org/10.15252/embj.201796664.

[15] B. Buzon, R. Grainger, S. Huang, C. Rzadki, M.S. Junop, Structure-specific endonuclease activity of SNM1A enables processing of a DNA interstrand crosslink, Nucleic Acids Res. 46 (2018) 9057–9066. https://doi.org/10.1093/nar/gky759.

[16] R.S. Cole, Repair of DNA Containing Interstrand Crosslinks in Escherichia coli: Sequential Excision and Recombination, Proc. Natl. Acad. Sci. 70 (1973) 1064–1068. https://doi.org/10.1073/pnas.70.4.1064.

[17] K.W. Kohn, N.H. Steigbigel, C.L. Spears, Cross-linking and repair of DNA in sensitive and resistant strains of E. coli treated with nitrogen mustard., Proc. Natl. Acad. Sci. 53 (1965) 1154–1161. https://doi.org/10.1073/pnas.53.5.1154.

[18] L.A. Zwelling, K.W. Kohn, W.E. Ross, R.A. Ewig, T. Anderson, Kinetics of formation and disappearance of a DNA cross-linking effect in mouse leukemia L1210 cells treated with cis- and trans-diamminedichloroplatinum(II)., Cancer Res. 38 (1978) 1762–8. http://www.ncbi.nlm.nih.gov/pubmed/565680.

[19] S.R. Rajski, R.M. Williams, DNA Cross-Linking Agents as Antitumor Drugs, Chem. Rev. 98 (1998) 2723–2796. https://doi.org/10.1021/cr9800199.

[20] S. Dutta, G. Chowdhury, K.S. Gates, Interstrand Cross-Links Generated by Abasic Sites in Duplex DNA, J. Am. Chem. Soc. 129 (2007) 1852–1853. https://doi.org/10.1021/ja067294u.

[21] K.M. Johnson, N.E. Price, J. Wang, M.I. Fekry, S. Dutta, D.R. Seiner, Y. Wang, K.S. Gates, On the Formation and Properties of Interstrand DNA–DNA Cross-Links Forged by Reaction of an Abasic Site with the Opposing Guanine Residue of 5'-CAp Sequences in Duplex DNA, J. Am. Chem. Soc. 135 (2013) 1015–1025. https://doi.org/10.1021/ja308119q.

[22] N.E. Price, K.M. Johnson, J. Wang, M.I. Fekry, Y. Wang, K.S. Gates, Interstrand DNA–DNA Cross-Link Formation Between Adenine Residues and Abasic Sites in Duplex DNA, J. Am. Chem. Soc. 136 (2014) 3483–3490. https://doi.org/10.1021/ja410969x.

[23] M.I. Nejad, R. Shi, X. Zhang, L.-Q. Gu, K.S. Gates, Sequence-Specific Covalent Capture Coupled with High-Contrast Nanopore Detection of a Disease-Derived Nucleic Acid Sequence, ChemBioChem. 18 (2017) 1383–1386. https://doi.org/10.1002/cbic.201700204.

[24] D.R. Semlow, J. Zhang, M. Budzowska, A.C. Drohat, J.C. Walter, Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase, Cell. 167 (2016) 498-511.e14. https://doi.org/10.1016/j.cell.2016.09.008.

[25] R.A. Wu, D.R. Semlow, A.N. Kamimae-Lanning, O. V. Kochenova, G. Chistol, M.R. Hodskinson, R. Amunugama, J.L. Sparks, M. Wang, L. Deng, C.A. Mimoso, E. Low, K.J. Patel, J.C. Walter, TRAIP is a master regulator of DNA interstrand crosslink repair, Nature. 567 (2019) 267–272. https://doi.org/10.1038/s41586-019-1002-0.

[26] M. Imani Nejad, K. Housh, A.A. Rodriguez, T. Haldar, S. Kathe, S.S. Wallace, B.F. Eichman, K.S. Gates, Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3, DNA Repair (Amst). 86 (2020) 102752. https://doi.org/10.1016/j.dnarep.2019.102752.

[27] H. Huang, P.B. Hopkins, DNA interstrand cross-linking by formaldehyde: nucleotide sequence preference and covalent structure of the predominant cross-link formed in synthetic oligonucleotides, J. Am. Chem. Soc. 115 (1993) 9402–9408. https://doi.org/10.1021/ja00074a005.

[28] J.T. Millard, S. Raucher, P.B. Hopkins, Mechlorethamine cross-links deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments, J. Am. Chem. Soc. 112 (1990) 2459–2460. https://doi.org/10.1021/ja00162a079.

[29] M.I. Nejad, K.M. Johnson, N.E. Price, K.S. Gates, A New Cross-Link for an Old Cross-Linking Drug: The Nitrogen Mustard Anticancer Agent Mechlorethamine Generates Cross-Links Derived from Abasic Sites in Addition to the Expected Drug-Bridged Cross-Links, Biochemistry. 55 (2016) 7033–7041. https://doi.org/10.1021/acs.biochem.6b01080.

[30] M. Tomás-Gamasa, S. Serdjukow, M. Su, M. Müller, T. Carell, "Post-It" Type Connected DNA Created with a Reversible Covalent Cross-Link, Angew. Chemie Int. Ed. 54 (2015) 796–800. https://doi.org/10.1002/anie.201407854.

[31] S.S. Pujari, P. Leonard, F. Seela, Oligonucleotides with "Clickable" Sugar Residues: Synthesis, Duplex Stability, and Terminal versus Central Interstrand Cross-Linking of 2'-O-Propargylated 2-Aminoadenosine with a Bifunctional Azide, J. Org. Chem. 79 (2014) 4423–4437. https://doi.org/10.1021/jo500392j.

[32] E.A. Harwood, P.B. Hopkins, S.T. Sigurdsson, Chemical synthesis of cross-link lesions found in nitrous acid treated DNA: a general method for the preparation of N2-substituted 2'-deoxyguanosines., J. Org. Chem. 65 (2000) 2959–64. https://doi.org/10.1021/jo991501+.

[33] S. Mukherjee, A. Guainazzi, O.D. Scharer, Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors, Nucleic Acids Res. 42 (2014) 7429–7435. https://doi.org/10.1093/nar/gku328.

[34] C.J. Wilds, J.D. Booth, A.M. Noronha, Synthesis of oligonucleotides containing an O6-G-alkyl-O6-G interstrand cross-link, Tetrahedron Lett. 47 (2006) 9125–9128. https://doi.org/10.1016/j.tetlet.2006.10.074.

[35] J. Gamboa Varela, K.S. Gates, A Simple, High-Yield Synthesis of DNA Duplexes Containing a Covalent, Thermally Cleavable Interstrand Cross-Link at a Defined Location, Angew. Chemie Int. Ed. 54 (2015) 7666–7669. https://doi.org/10.1002/anie.201502566.

[36] M.J. Catalano, S. Liu, N. Andersen, Z. Yang, K.M. Johnson, N.E. Price, Y. Wang, K.S. Gates, Chemical Structure and Properties of Interstrand Cross-Links Formed by Reaction of Guanine Residues with Abasic Sites in Duplex DNA, J. Am. Chem. Soc. 137 (2015) 3933–3945. https://doi.org/10.1021/jacs.5b00669.

[37] M.I. Nejad, N.E. Price, T. Haldar, C. Lewis, Y. Wang, K.S. Gates, Interstrand DNA Cross-Links Derived from Reaction of a 2-Aminopurine Residue with an Abasic Site,

ACS Chem. Biol. 14 (2019) 1481–1489. https://doi.org/10.1021/acschembio.9b00208.

[38] Y. Li, R.R. Breaker, Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2′-Hydroxyl Group, J. Am. Chem. Soc. 121 (1999) 5364–5372. https://doi.org/10.1021/ja990592p.

[39] W. Xu, A.M. Ouellette, Z. Wawrzak, S.J. Shriver, S.M. Anderson, L. Zhao, Kinetic and Structural Mechanisms of (5′ S)-8,5′-Cyclo-2′-deoxyguanosine-Induced DNA Replication Stalling, Biochemistry. 54 (2015) 639–651. https://doi.org/10.1021/bi5014936.

[40] W. Xu, R.M. Boyd, M.O. Tree, F. Samkari, L. Zhao, Mitochondrial transcription factor A promotes DNA strand cleavage at abasic sites, Proc. Natl. Acad. Sci. 116 (2019) 17792–17799. https://doi.org/10.1073/pnas.1911252116.

[41] S. Couvé, G. Macé-Aimé, F. Rosselli, M.K. Saparbaev, The Human Oxidative DNA Glycosylase NEIL1 Excises Psoralen-induced Interstrand DNA Cross-links in a Three-stranded DNA Structure, J. Biol. Chem. 284 (2009) 11963–11970. https://doi.org/10.1074/jbc.M900746200.

[42] K.D. Koh, S. Balachander, J.R. Hesselberth, F. Storici, Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA, Nat. Methods. 12 (2015) 251–257. https://doi.org/10.1038/nmeth.3259.

[43] A.R. Clausen, J.S. Williams, T.A. Kunkel, Measuring Ribonucleotide Incorporation into DNA In Vitro and In Vivo, in: 2015: pp. 123–139. https://doi.org/10.1007/978-1-4939-2596-4_9.

[44] S.M. Cerritelli, R.J. Crouch, Ribonuclease H: the enzymes in eukaryotes, FEBS J. 276 (2009) 1494–1505. https://doi.org/10.1111/j.1742-4658.2009.06908.x.

[45] E. Tannous, E. Kanaya, S. Kanaya, Role of RNase H1 in DNA repair: removal of single ribonucleotide misincorporated into DNA in collaboration with RNase H2, Sci. Rep. 5 (2015) 9969. https://doi.org/10.1038/srep09969.

[46] J. Tang, W. Zhao, N.G. Hendricks, L. Zhao, High-Resolution Mapping of Amino Acid Residues in DNA–Protein Cross-Links Enabled by Ribonucleotide-Containing DNA, Anal. Chem. 93 (2021) 13398–13406. https://doi.org/10.1021/acs.analchem.1c03481.

[47] I.G. Minko, M.B. Harbut, I.D. Kozekov, A. Kozekova, P.M. Jakobs, S.B. Olson, R.E. Moses, T.M. Harris, C.J. Rizzo, R.S. Lloyd, Role for DNA Polymerase κ in the Processing of N2-N2-Guanine Interstrand Cross-links, J. Biol. Chem. 283 (2008) 17075–17082. https://doi.org/10.1074/jbc.M801238200.

[48] K. Yamanaka, I.G. Minko, K. Takata, A. Kolbanovskiy, I.D. Kozekov, R.D. Wood, C.J. Rizzo, R.S. Lloyd, Novel Enzymatic Function of DNA Polymerase ν in Translesion DNA Synthesis Past Major Groove DNA−Peptide and DNA−DNA Cross-Links, Chem. Res. Toxicol. 23 (2010) 689–695. https://doi.org/10.1021/tx900449u.

[49] T.V. Ho, A. Guainazzi, S.B. Derkunt, M. Enoiu, O.D. Schärer, Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases, Nucleic Acids Res. 39 (2011) 7455–7464. https://doi.org/10.1093/nar/gkr448.

[50] U. Roy, S. Mukherjee, A. Sharma, E.G. Frank, O.D. Schärer, The structure and duplex context of DNA interstrand crosslinks affects the activity of DNA polymerase η, Nucleic Acids Res. 44 (2016) 7281–7291. https://doi.org/10.1093/nar/gkw485.

[51] T. OKAZAKI, A. KORNBERG, ENZYMATIC SYNTHESIS OF DEOXYRIBONUCLEIC ACID. XV. PURIFICATION AND PROPERTIES OF A POLYMERASE FROM BACILLUS SUBTILIS., J. Biol. Chem. 239 (1964) 259–68. http://www.ncbi.nlm.nih.gov/pubmed/14114852.

[52] N. Li, J. Wang, S.S. Wallace, J. Chen, J. Zhou, A.D. D'Andrea, Cooperation of the NEIL3 and Fanconi anemia/BRCA pathways in interstrand crosslink repair, Nucleic Acids Res. 48 (2020) 3014–3028. https://doi.org/10.1093/nar/gkaa038.

[53] Y.K. Cheun, A.S. Groehler, O.D. Schärer, New Synthetic Analogs of Nitrogen Mustard DNA Interstrand Cross-Links and Their Use to Study Lesion Bypass by DNA Polymerases, Chem. Res. Toxicol. 34 (2021) 1790–1799. https://doi.org/10.1021/acs.chemrestox.1c00123.

Chapter 4. Complexity and Repair of DNA Terminal Structures in Mitochondrial Transcript Factor A-Mediated Strand Scission at Abasic Sites

4.1 Abstract

Mitochondria DNA (mtDNA), similar to nuclear DNA, constantly encounter chemical and physical assault, and produce various DNA lesions. Among these lesions, abasic sites are the most abundant DNA lesions. And because there is no base group, AP sites are blocks for transcription and translation, threatening the integrity of genome. AP sites in mtDNA could form DNA-protein cross-links with transcription factor A (TFAM), along with the formation of strand breaks. In this study, we found various strand breaks produced during the DPC reaction and vary in different sequence context. During the DPC reaction, the GSH-adduct is the most abundant strand break product in the presence of GSH. The involvement of GSH didn't impact the rate of DPC formation or intact AP strand disappearance but influence the yield of SSBs, indicating the GSH rapidly added to the α,β-unsaturated aldehyde. We further analyzed the kinetic process of the repair enzymes on the GSH-adducts. This study provides kinetic insights into the strand break products generated during DPC reaction and detailed information about the repair of GSH-adducts.

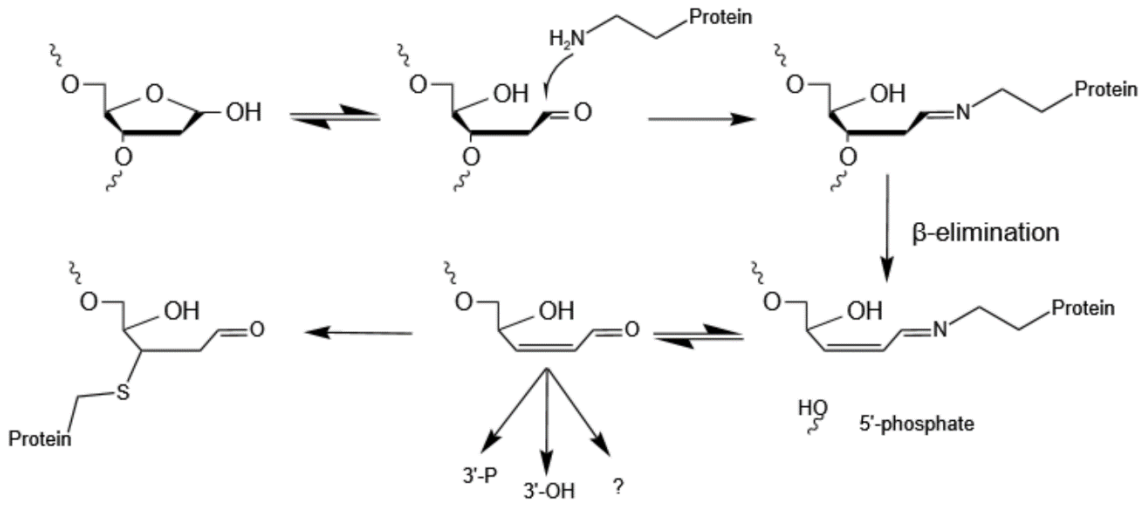
4.2 Introduction

Abasic sites result from cleavage of the N-glycosylic bond between the base group and the sugar ring [1]. They can be generated through the base loss reactions, such as

spontaneous hydrolysis [2], the destabilization of the base group [3], and the enzymatic reactions of DNA glycosylases [4]. If the repair of AP sites is insufficient, the AP sites persist in DNA and block transcription and replication [5]. AP sites exist in cells in an equilibrating mixture of a closed-ring furanose (99%) and opened-ring aldehyde (1%) [6]. The opened-ring state of the AP site, even though in less abundance, possesses a highly reactive aldehyde group and could lead to secondary products, such as DNA-protein cross-links, DNA inter/intra-strand cross-links and strand breaks, making them cytotoxic and threatening genomic stability [7]. AP sites coexist with a wide variety of nucleophilic molecules in the cellular environment, such as thiols, amines, and proteins, the involvement of these molecules is known to promote strand break formation [8,9].

Mitochondria are vital cellular compartments involved in energy production and cell signaling [10], they are susceptible to endogenous and exogenous assault, making mtDNA vulnerable to DNA damaging effects. Mitochondria are equipped with repair systems [11], but lesions persist if the repair is insufficient, causing important biological impacts, such as mutations, aging, cancer, and neurodegeneration diseases [12]. Single-strand breaks (SSBs) are the breaks in one strand of the DNA double-strand and usually have a loss of a nucleotide or damaged 3’- or 5’ terminal [13]. SSBs have serious consequences toward inhibiting cell survival and have strong cytotoxicity by inducing double-strand breaks [14]. There are a few sources of endogenous SSBs, such as oxidative stress by reactive oxygen species (ROS) that originated from free radical attack [15], and the enzyme-involved strand incision reactions that produce nicks in strands [16]. The ring-

opened form of abasic sites could go through spontaneous strand cleavage reaction by β-elimination, producing 3'-α,β-unsaturated aldehyde, and 5'-phosphate termini [7]. DNA binding proteins, such as histone and TFAM, utilize their lysine residues to form covalent cross-linking bonds with AP sites through Schiff base chemistry and later release the protein, this process, which starts from AP sites to strand breaks, has a much higher reaction rate than spontaneous β-elimination reaction by AP site [17,18], displayed the accelerated strand breaks formation when DNA binding protein exists. Our group recently found that in the presence of $NaCNBH_3$, which stabilizes the Schiff base intermediate, the major site that forms DPC is the K186 [19], however, when the reaction didn't contain $NaCNBH_3$, the major cross-linked residue is the C49 [20]. The conversion from the initial lysine to the final cysteine reveal the reaction happens through the formation of DPC via Schiff base chemistry, then later the cysteine has higher reactivity to the strand breaks generated after β-elimination reaction and produced the cysteine cross-links as the final DPC product (**Scheme 4.1**).



Scheme 4.1 Abasic sites in DPC reaction. Lysine residues on the protein first react with AP sites through Schiff base chemistry and form the DNA-protein cross-links. The formation of DPC acidifies the α-proton and promotes the β-elimination reaction, forming strand break DPCs and 5'-phosphate. DPC releases free protein and 3'-α,β-unsaturated aldehyde. When a protein contains cysteine residues, such as TFAM, it could participate in the Michael addition reaction and form DPC through cysteine residues, while other non-trapped 3'-α,β-unsaturated aldehyde could go through other reactions and turn to different 3'-terminis.

Most recently, Jha et al. demonstrated that the thiol residues of glutathione (GSH) rapidly added to the 3'-α,β-unsaturated aldehyde and produced glutathionylated adduct as the only product [21,22]. Their findings are consistent with our recent discovery that the cysteine residues on TFAM are the major cross-linked residues that cross-link with AP sites [20], indicating the possible involvement of biological thiols in the reaction that have α,β-unsaturated aldehyde intermediates. Especially, the mitochondrial matrix contains a relatively high GSH amount, ranges from 1 mM to 10 mM. We envision that GSH participates during the TFAM-AP DNA reaction and compete with TFAM cysteines for the reactive α,β-unsaturated aldehyde intermediates, GSH-adduct might be the most

abundant strand break products and coexist with DPC that cross-linked with AP sites through cysteine residues.

In this study, we investigated the involvement of GSH in the DPC formation reaction during the TFAM-AP cross-links formation. We characterized the strand breaks generated during DPC reactions with the standard strand break substrates and mass spectrometry. We then studied the impact of the DPC reaction rate upon the addition of GSH and spermine, by comparing the reaction rates we revealed how the addition of GSH and spermine impact the rates in DPC reactions. The GSH-adducts are the blocks to recover the canonical strand. We tested a few repair enzymes on their efficacy of cleaving GSH-adducts, and we found out that APE1, APE2, and TDP1 display different efficiency to cleave GSH-adducts.

4.3 Experimental procedure

4.3.1 Materials

The oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT) and listed in **Table 4.1**. Uracil DNA glycosylase (UDG) and Endonucleases were purchased from New England Biolabs. All other chemicals were purchased from Fisher Scientific and Sigma-Aldrich.

Table 4.1 The sequence of the oligonucleotides that were used in this study.

Name of the oligonucleotide	Sequence
T12	5'-FAM-TAA CA**U** TCA CCC CCC AAC TAA C
AP12	5'-FAM-TAA CA**X** TCA CCC CCC AAC TAA C
T17	5'-FAM-TAA CAG TCA CCC CCC **U**AC TAA C
AP17	5'-FAM-TAA CAG TCA CCC CCC **X**AC TAA C
T20	5'-FAM-TAA CAG TCA CCC **U**CC AAC TAA C
AP20	5'-FAM-TAA CAG TCA CCC **X**CC AAC TAA C
T32	5'-FAM-TGC GGT ATG CAC TTT TAA CA**U** TCA CCC CCC AAC TAA C
AP32	5'-FAM-TGC GGT ATG CAC TTT TAA CA**X** TCA CCC CCC AAC TAA C
T13	5'-GTT AGT TGG GGG GTG ACT GTT A
T13-quencher	5'-GTT AGT TGG GGG GTG ACT CTT A-3' Iowa Black® FQ
T28	5'-GTT AGT TGG GGG GTG AAT GTT AAA AGT GCA TAC CGC

4.3.2 Preparation of AP site-containing DNA probes

The preparation of AP-DNA was described before [19]. The 5'-FAM labeled DNA containing deoxyuracil was purchased from IDT. 3 nmol DNA was incubated with 8 uL UDG in 1X buffer containing 20 mM HEPES, 1 mM DTT, and 1 mM EDTA at pH 7.4, the reaction was incubated at 37°C for 6 hours and checked the yield of AP sites with strong NaOH cleavage reaction. The AP-DNA was extracted through phenol/chloroform extraction and purified with a P-6 size exclusion column. The pure AP DNA was recovered in water.

4.3.3 Preparation and validation of GSH-adduct DNA substrate

The double-strand AP-DNA was incubated with 10 mM GSH and 1 mM spermine at 37°C overnight in 1X MES buffer pH 6.5, the confirmation of GSH-adduct was performed using a Bio-red Mini-protein 1 mm 18% Urea-TEB-PAGE. The GSH-adduct from the

reaction was prepared in 200 pmol scale and separated with DNA sequencing PAGE, then cut from the gel and extracted for mass spectrometry analysis.

4.3.4 Strand breaks prepared from AP-DNA with NaOH

We tested the mild and harsh strand break production with NaOH. In mild condition, we incubated AP-DNA with 0.2 M NaOH at 37°C for 20 min then neutralized the reaction with an equal amount of HCl. In harsh conditions, we incubated AP-DNA with 0.3 M NaOH at 65°C for 1 hour and then quenched with an equal amount of HCl. The AP sites were quenched with 200 mM $NaBH_4$ for 15 min.

4.3.5 Endo III cleavage reaction with AP-DNA

10 pmol of AP containing DNA was reacted with 1 uL E.coli Endo III (Nth) at 37°C for 1 hour under the condition of 20 mM Tris-HCl, 1 mM DTT, 1 mM EDTA at pH = 8. The reaction was quenched with 200 mM $NaBH_4$ for 15 min.

4.3.6 Generation of SSBs during AP-TFAM DPC formation

The DPC reaction was performed without $NaCNBH_3$ as previous shown [20]. After 24 hours incubation, the reaction was quenched with 200 mM $NaBH_4$ for 15 min, then added the STOP solution. The reaction with GSH or spermine were performed under the same conditions but with the addition of 10 mM GSH or 1.2 mM spermine.

4.3.7 Purification of APE1, APE2 and TDP1

The full-length APE1 was as previously described [18]. The full-length TDP1 was cloned in pDest-527 vector and was a gift from Dr. Terrence R. Burke Jr. The plasmid

containing TDP1 gene was grown under 100 ug/mL ampicillin until OD reached 0.6. The induction of TDP1 expression was enabled by adding Isopropyl ß-D-1-thiogalactopyranoside (IPTG). The His-tagged TDP1 was enriched with His-Trap column and then purified with ion exchange chromatography following the published procedures [23] by Yu-Hsuan Chen. The full-length *Xenopus laevis* APE2 was cloned into 2Cc-T vector and was a gift from Dr. Scott Williams. The *E.coli* cells transfected with the plasmid grew under 100 ug/mL ampicillin until OD=0.5, then we added 0.2 mM IPTG to induce the expression of APE2-MBP-His at 17°C overnight. The overexpression of APE2-MBP was confirmed with SDS-PAGE. Then the protein was purified with His-Trap column and ion exchange HiTrap Q chromatography according to the published procedures [24]. The purified protein is in the form of APE2-MBP-6His. The MBP-tag was not removed to preserve the solubility of APE2.

4.3.8 Gel extraction of the SSBs for mass spectrometry analysis

The SSBs generated from the strand break reactions were prepared on the scale of about 200 pmol and separated by the preparative DNA sequencing PAGE (1 mm thick and 80 cm length). After separation, we cut the piece containing SSBs from the gel, crushed into fine pieces, and extracted with 1 mL 10 mM HEPES pH 7.4 at 4°C overnight. The gel was removed with Co-Star filter device and the flow through was cleaned up with C18 cartridge. The sample recovered from C18 cartridge was completely dry with SpeedVac and recovered with water for mass spectrometry analysis.

4.3.9 Steady state kinetics

The steady state kinetic experiments performed for APE1 and TDP1 in their reaction with 32-GSH adduct. TDP1 reactions were performed under 10 nM TDP1 with varied concentration of dsAP32-GSH in the condition of 20 mM HEPES, 50 mM NaCl, 0.1 mg/mL BSA, 5 mM $MgCl_2$ at pH 7.7 at 37°C. The initial reaction time points were taken and immediately quenched with STOP solution, and the initial reaction rates were obtained by fit the product concentration vs. time with linear fit. All reactions repeated at least three times. Then we fitted the rate vs. dsAP32-GSH concentration with Michaelis-Menten equation and obtained k_{cat} and K_m, the catalytic efficiency obtained by k_{Cat}/K_m. The reactions for APE1 were performed under 10 nM APE1 with varied concentrations of dsAP32-GSH in the condition of 20 mM HEPES pH 7.4, 50 mM KCl, 1 mM $MgCl_2$ and 1 mM DTT at 37°C. Reaction time points were taken and are plotted time vs product concentration. The initial reaction rates were obtained from at least three independent tests. Then we plotted the rates vs. dsAP32-GSH concentration and fit with Michaelis-Menten equation for k_{Cat} and K_m.

4.3.10 APE2 reaction with GSH-DNA adducts

The fusion APE2-MBP-His was diluted with 10 mM HEPES pH 8.0, 75 mM NaCl, 0.5 mM TCEP, 0.5 mM $MgCl_2$, 0.1 mg/mL BSA and 20% glycerol. The APE2 cleavage reactions were performed under 10 mM HEPES pH 8.0, 1 mM TCEP and 1 mM $MnCl_2$ at 37°C for various time and immediately quenched with STOP solution.

4.4 Results

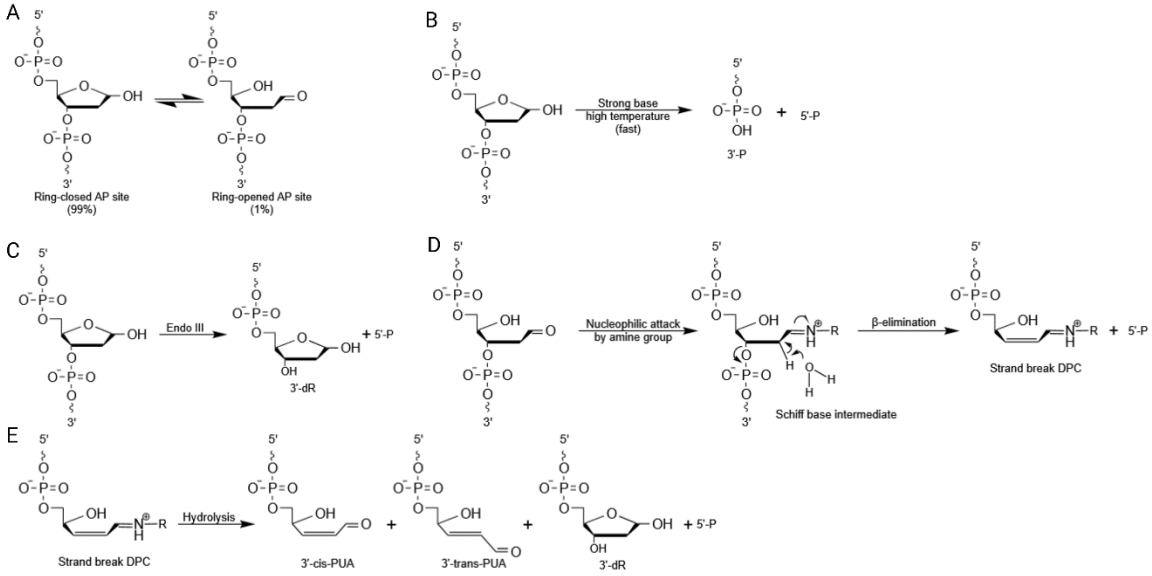
4.4.1 Identification of the strand breaks produced by AP sites in DPC reaction with model substrates

AP sites are a mixture of the ring-opened state and ring-closed aldehyde state, even though the ring-closed form only composes 1% of total AP sites, they have highly reactive aldehyde group and susceptible to nucleophilic attacks (**Scheme 4.2 A**). AP sites can react with various molecules in cells and produce various types of strand breaks, and to learn the repair process of AP sites, knowing the identity of strand breaks generated from AP sites is fundamental to understanding the repair mechanisms.

To characterize and quantify the strand breaks, we generated the AP containing DNA strands using the deoxyuracil containing DNA and incubated them with UDG and then utilized the known reactions to prepare the standard strand break substrates. For AP12 sequence, the AP site is located 5 nucleotides away from the 5'-end in the length of 22 nucleotides DNA. The 5'-strand breaks are 5 nt length and therefore using the DNA sequencing PAGE should be enough to separate the strand breaks. The harsh NaOH incubation together with high temperature majority produce the 3'-P products (**Scheme 4.2 B**). The *E.coli* Endo III could cleave the AP sites into 3'-dR (**Scheme 4.2 C**). The formation of TFAM-AP cross-links is through the formation of Schiff base and then converted to strand break DPCs via β-elimination reaction [25] (**Scheme 4.2 D**). And in our previous result from gel analysis, we observed multiple strand break bands [18]. Haldar et al demonstrated that AP sites can produce 3'-cis-α,β-unsaturated aldehyde (3'-cis-PUA),

3’-trans-α,β-Unsaturated Aldehyde (3’-trans-PUA), and 3’-P, leaving a sugar remnant through γ,δ-Elimination (**Scheme 4.2 E**).

Scheme 4.2 Products generated from AP sites through different reactions.



Based on the previous knowledge of preparing the strand break standard with different 3’-terminus, we prepared several standard strand break substrates and used high-resolution sequencing gel to separate and visualize them, then we compared them with the SSBs generated during the DPC reaction (**Figure 4.1 A**).

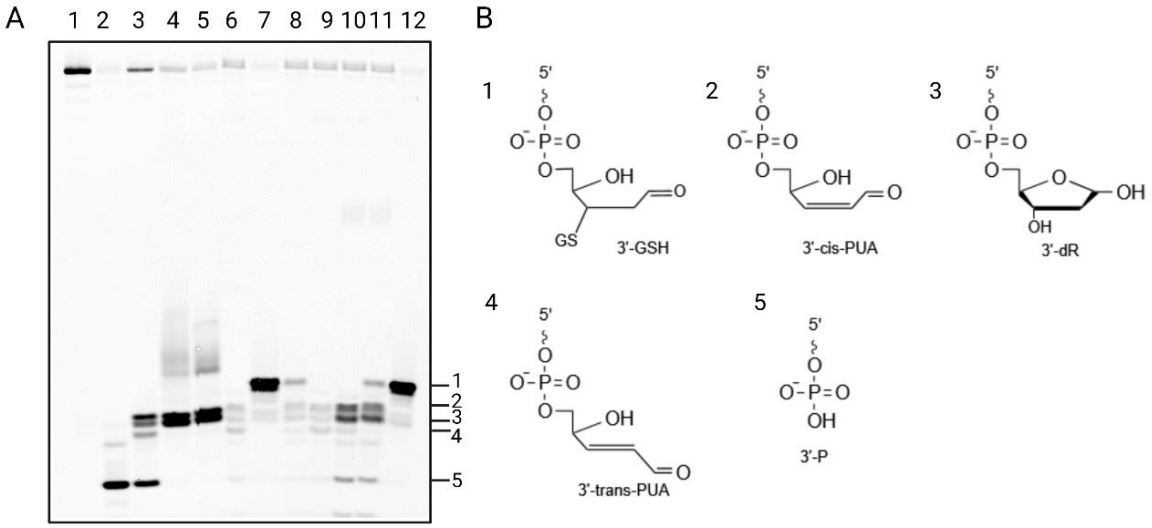


Figure 4.1 AP12 reaction under various conditions and the strand break assignments. (**A**) Lane 1. Reduced AP12; lane 2. AP12 incubated with 0.3 M NaOH at 65°C for 1 hour; lane 3. AP12 incubated with 0.2 M NaOH at 37°C for 20 min; lane 4. dsAP12 incubated with E.coli Endo III; lane 5. sample in lane 4 incubated with 10 mM GSH at 37°C for 1 hour; lane 6. quenched wtTFAM AP12 DPC reaction, 24 hours; lane 7. quenched wtTFAM AP12 DPC reaction under 10 mM GSH, 24 hours; lane 8. wtTFAM AP12 DPC reaction for 24 hours, then add 10 mM GSH for 1 hour and quench with $NaBH_4$; lane 9. same as lane 6; lane 10. quenched wtTFAM AP12 DPC reaction under 1.2 mM spermine for 24 hours; lane 11, wtTFAM AP12 DPC reaction under 1.2 mM spermine for 24 hours and add 10 mM GSH at 37°C for 1 hour, quench with $NaBH_4$; lane 12. quenched wtTFAM AP12 DPC reaction under 10 mM GSH and 1.2 mM spermine for 24 hours. (**B**) Structure of the products.

We observed that the DPC reaction happened in the presence of GSH produced the 3'-GSH adduct as the major strand break product (product 1 in **Figure 4.1 B**, lane 7). When DPC reaction didn't include GSH, the strand breaks are 3'-cis-PUA, 3'-dR, 3'-trans-PUA and 3'-P (**Figure 4.1 A**, lane 6). The addition of spermine didn't change the strand breaks that produced during DPC reaction, this might indicate the catalytic effect of spermine and spermine did not involve in the subsequent reaction with 3'-α,β-unsaturated aldehyde (**Figure 4.1 A**, lane 10). The strand breaks produced by DPC reaction can further react with GSH after quenching with $NaBH_4$, indicating the addition of GSH

originated from strand break products (**Figure 4.1 A**, lane 7). The strand breaks produced during the DPC reaction under spermine can also react with GSH after quenching with $NaBH_4$ (**Figure 4.1 A**, lane 11). The expected product of Endo III should be 3'-dR, but in this gel we observed two bands and they are very close to each other together with a smeared band that migrated slower than these two bands (**Figure 4.1 A**, lane 4), the Endo III cleaved product could partially react with GSH (**Figure 4.1 A**, lane 5), which is in contrary to the innate nature of 3'-dR. And under mild NaOH treatment, AP sites could become various strand breaks (**Figure 4.1 A**, lane 3).

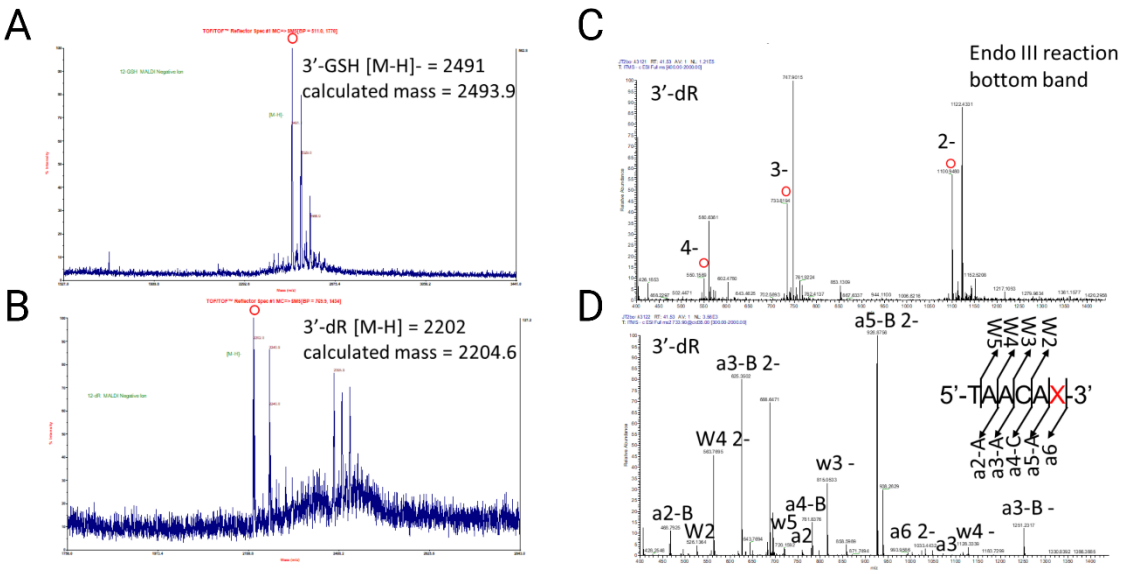


Figure 4.2 Mass spectrometry verification of the 3'-GSH and 3'-dR. **A**. MALDI-TOF analysis of the 3'-GSH; **B**. MALDI-TOF analysis of the 3'-dR; **C**. ESI-MS of 3'-dR and **D**. MS/MS of the 3'-dR.

The involvement of GSH largely changed the strand breaks that produced during the DPC reaction and made the 3'-GSH adduct as the major strand break product for all reactions when GSH is present (**Figure 4.1 A**, lane 7 and 12), indicating the GSH could

rapidly react with the strand break intermediate and form the GSH adduct, we also verified the GSH adduct with mass spectrometry (**Figure 4.2**). We analyzed how the location of AP sites impact the strand breaks formation with AP17 and AP20, where the AP site on AP17 closely interact with TFAM K186, but for AP20, the AP site is in the linker region of TFAM that is known to have less interaction with TFAM. However, the separation of the bands is not as efficient as AP12 and couldn't separate the strand breaks with DNA sequencing PAGE because the strand breaks from AP17 and AP20 are longer (**Figure 4.3**). We observed different strand breaks distribution that the AP17 majorly produce 3'-PUA and 3'-P but AP20 majorly produce 3'-P. Therefore, the interaction of AP site with TFAM also impacts the strand breaks that are produced following the DPC formation.

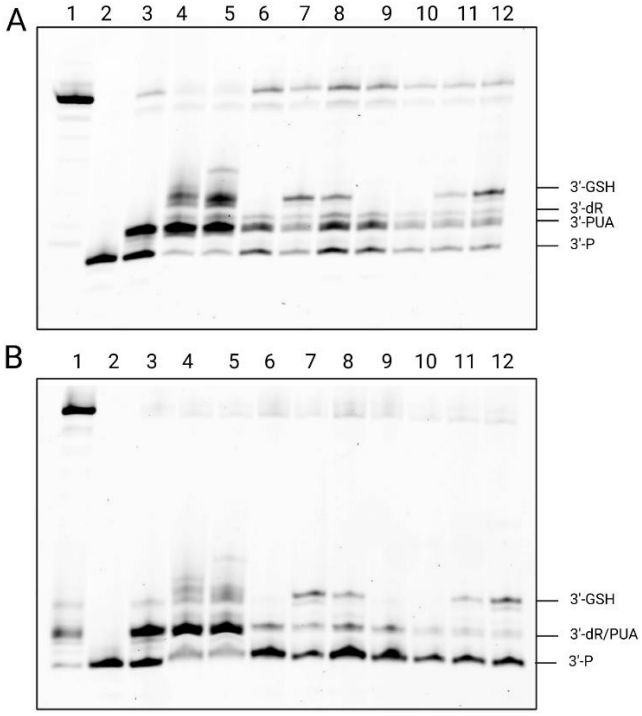


Figure 4.3 The strand breaks produced by AP17 and AP20. **A**. strand breaks produced by AP17, Lane 1. Reduced AP17; lane 2. AP17 incubated with 0.3 M NaOH at 65°C for 1 hour; lane 3. AP17 incubated with 0.2 M NaOH at 37°C for 20 min; lane 4. dsAP17 incubated with E.coli Endo III; lane 5. sample in lane 4 incubated with 10 mM GSH at 37°C for 1 hour; lane 6. quenched wtTFAM AP17 DPC reaction, 24 hours; lane 7. quenched wtTFAM AP17 DPC reaction under 10 mM GSH, 24 hours; lane 8. wtTFAM AP17 DPC reaction for 24 hours, then add 10 mM GSH for 1 hour and quench with $NaBH_4$; lane 9. same as lane 6; lane 10. quenched wtTFAM AP17 DPC reaction under 1.2 mM spermine for 24 hours; lane 11, wtTFAM AP17 DPC reaction under 1.2 mM spermine for 24 hours and add 10 mM GSH at 37°C for 1 hour, quench with $NaBH_4$; lane 12. quenched wtTFAM AP17 DPC reaction under 10 mM GSH and 1.2 mM spermine for 24 hours.; **B**. strand breaks produced by AP20. Lane 1. Reduced AP20; lane 2. AP20 incubated with 0.3 M NaOH at 65°C for 1 hour; lane 3. AP20 incubated with 0.2 M NaOH at 37°C for 20 min; lane 4. dsAP20 incubated with E.coli Endo III; lane 5. sample in lane 4 incubated with 10 mM GSH at 37°C for 1 hour; lane 6. quenched wtTFAM AP20 DPC reaction, 24 hours; lane 7. quenched wtTFAM AP20 DPC reaction under 10 mM GSH, 24 hours; lane 8. wtTFAM AP20 DPC reaction for 24 hours, then add 10 mM GSH for 1 hour and quench with $NaBH_4$; lane 9. same as lane 6; lane 10. quenched wtTFAM AP20 DPC reaction under 1.2 mM spermine for 24 hours; lane 11, wtTFAM AP20 DPC reaction under 1.2 mM spermine for 24 hours and add 10 mM GSH at 37°C for 1 hour, quench with $NaBH_4$; lane 12. quenched wtTFAM AP20 DPC reaction under 10 mM GSH and 1.2 mM spermine for 24 hours.

4.4.2 Characterization of the effects of GSH and spermine on the DPC formation rate and product profile

The formation of GSH-adduct largely altered the result of DPC reactions, so it is necessary to know the impact of GSH on the reaction rates and yield of DPCs. The AP12 was incubated with TFAM with or without GSH, with or without spermine. Based on their fluorescent signals, we quantified the relative abundance of the the DPC, SSBs and intact AP-DNA on the 4-16% stacking gel (**Figure 4.4**).

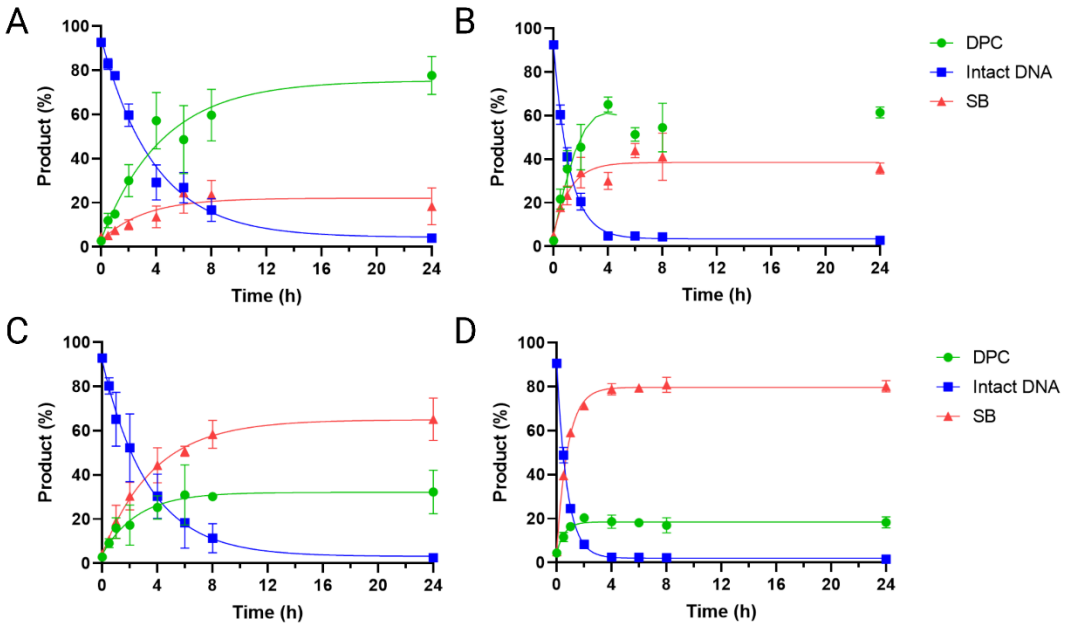


Figure 4.4 The time course of the reactant and product during AP12 reaction with TFAM. **A**. wtTFAM with dsAP12 DPC reaction time course; **B**. wtTFAM with dsAP12 DPC reaction under 1.2 mM spermine; **C**. wtTFAM with dsAP12 DPC reaction under 1 mM GSH; **D**. wtTFAM with dsAP12 DPC reaction under 1.2 mM spermine and 1 mM GSH. All data were acquired from three independent tests and displayed as Mean ± SD.

We fitted the curve with one-phase decay and obtained the intact DNA disappearance rate ($k_{dis,int}$), the DPC formation rate ($k_{form,DPC}$) and the SSB formation rate

($k_{form,SSB}$). We also obtained the average yield of SSBs and DPCs, as shown in **Table 4.2**. The product yield of DPCs decrease dramatically with the addition of GSH, but not that significant with the addition of spermine. The involvement of GSH didn't change the $k_{dis,int}$, $k_{form,DPC}$ or $k_{form,SSB}$ when added 1 mM GSH, but the rates were largely accelerated with the addition of spermine. And in the presence of both GSH and spermine, the rates were largely accelerated. The yield of DPC decreases dramatically with the addition of GSH and spermine, indicating the competition between protein and GSH for the strand break products.

dsAP12 with wtTFAM				
		with 1.2 mM spermine	with 1 mM GSH	with 1.2 mM spermine and 1 mM GSH
$k_{dis,int}$ (X 10^{-5} s^{-1})	7.2 ± 1.0	23.2 ± 3.6	9.2 ± 3.5	37.8 ± 1.6
$k_{form,DPC}$(X 10^{-5} s^{-1})	6.1 ± 1.6	15.4 ± 7.1	11.2 ± 2.6	55.8 ± 20.4
$Yield_{DPC}$(%)	77.7 ± 7.0	61.5 ± 2.0	32.3 ± 8.0	18.3 ± 2.0
$k_{form,SSB}$(X 10^{-5} s^{-1})	9.1 ± 1.3	38.1 ± 4.0	8.6 ± 3.9	34.1 ± 0.9
$Yield_{SSB}$(%)	18.4 ± 6.8	35.8 ± 2.0	65.2 ± 7.9	80.2 ± 2.1

Table 4.2 Rates of the species in the AP12 TFAM DPC reactions. *Data were from three* independent tests and displayed in Mean ± SD.

4.4.3 Effects of DNA repair proteins on GSH-DNA adducts

The 3'-GSH adduct is the most abundant strand breaks generated during the DPC reactions in the presence of GSH. In this case, knowing how DNA repair enzymes react with 3'-GSH adduct will be helpful to learn how cells repair AP sites. We prepared recombinant APE1, APE2 and TDP1 and tested their cleavage efficiencies with the 3'-GSH

adducts. The AP12 substrate is not suitable for these tests because the APE1 prefers double strand substrate, and with this short strand breaks, AP12 doesn't have enough reactivity because the 3'-termini are in the single strand mode (**Figure 4.5**), which will impact the result due to the enzyme substrate specificity.

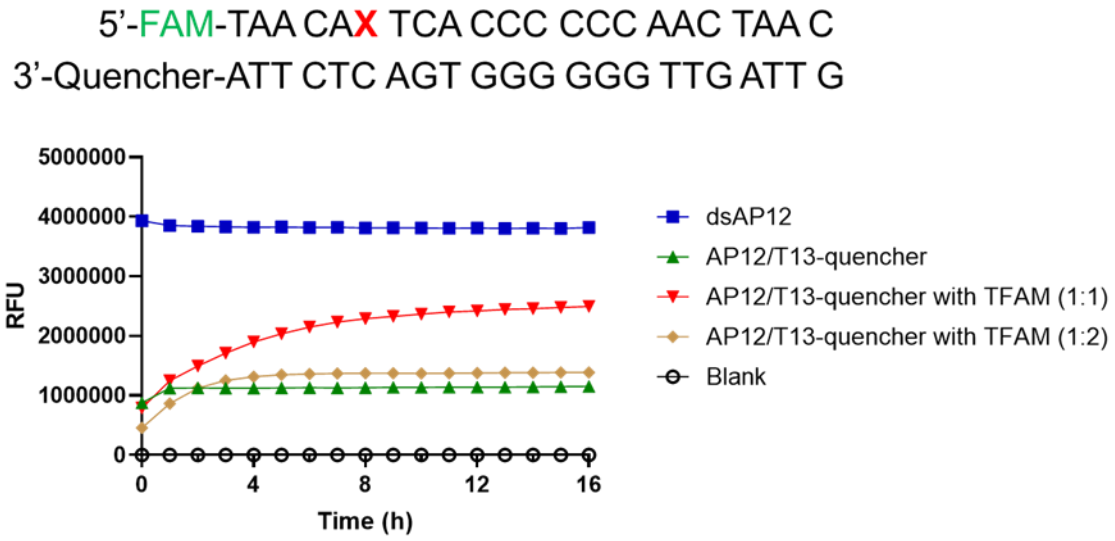


Figure 4.5 The change of fluorescent signal during DPC formation.

Therefore, we selected a longer DNA substrate AP32 with 37 nucleotides. We tested its binding pattern with TFAM and used AP32 to prepare the 3'-GSH adducts with GSH and spermine reaction. And then tested the cleavage rate of AP32-GSH with TDP1 and APE1 and obtained the Michaelis–Menten saturation curve (**Figure 4.6**). The Michaelis-Menten constants are listed in **Table 4.3**. Our results showed that APE1 has higher catalytic efficiency than TDP1 but has less maximum rate. This result is in accordance with the finding from Wei et al. [26].

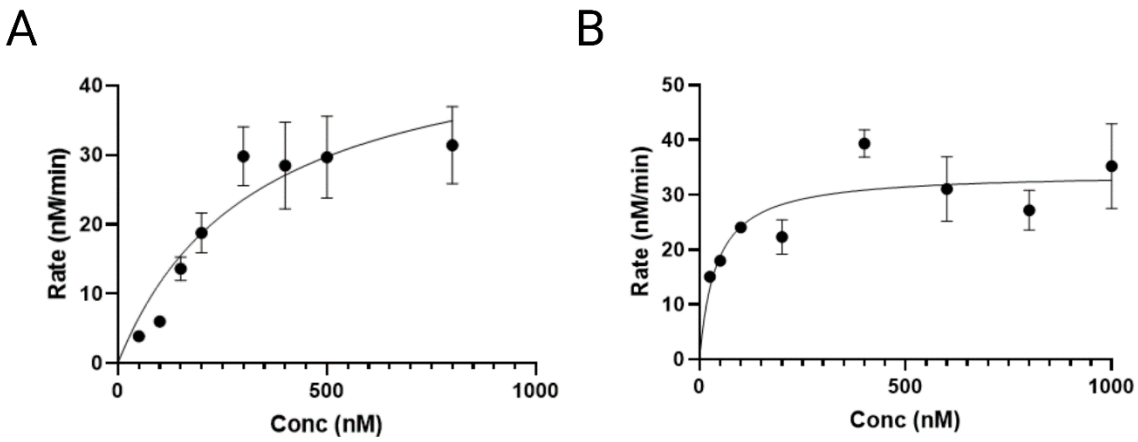


Figure 4.6 The Michaelis–Menten kinetics of TDP1 and APE1-mediated excision of DNA-GSH adducts with double-stranded AP32-GSH. A. Reaction of TDP1 with double-stranded AP32-GSH; B. Reaction of APE1 with double-stranded AP32-GSH. The rates were obtained with at least three independent tests, and the errors are standard deviation.

Table 4.3 Steady-state kinetic constant for cleaving the GSH adduct by TDP1 and APE1 results are from at least three independent tests and displayed as Mean ± SD.

Enzyme	V_{max} (nM/min)	K_m (nM)	k_{cat} (min^{-1})	Catalytic Efficiency (uM^{-1} min^{-1})
TDP1	50.34 ± 11.79	359.5 ± 115.3	5.034 ± 1.179	14.4 ± 2
APE1	34.14 ± 3.72	41.84 ± 12.76	3.414 ± 0.372	84.86 ± 16.5

We also purified the full-length *Xenopus laevis* APE2 linked with MBP and His-tag. And we tested the cleavage activity of fusion APE2-MBP-His with dsAP32-GSH. We discovered that fusion APE2 has even higher efficiency in cleaving 3'-GSH adduct than the AP sites (**Figure 4.7 A**). We rationed this phenomenon with the construct of APE2 that contains both the Endo/Exo/Phosphatase domain and Zf-GRF domain, where the Zf-GRF domain is required for 3'-5' end resection of oxidative damage and also enables its binding of ssDNA [24] (**Figure 4.7 B**).

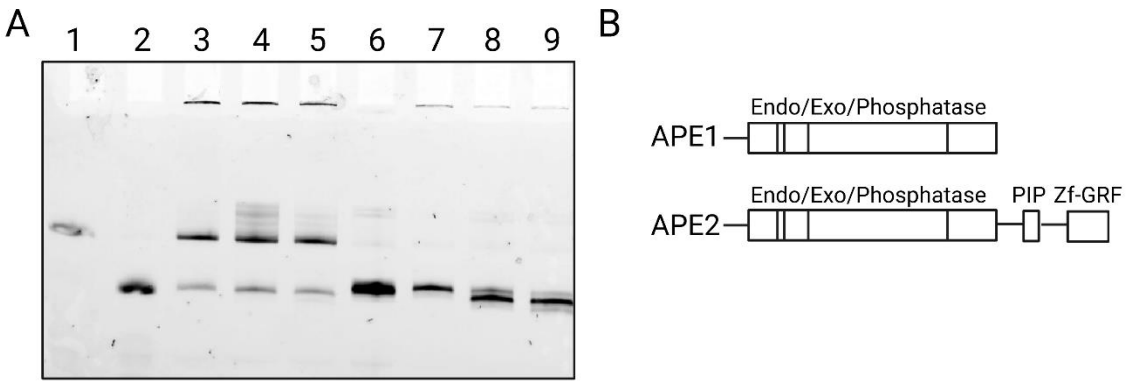


Figure 4.7 The reaction of APE2 with the domain architecture of APE1 and APE2. **A**. Reaction of the APE2 with double strand AP32 and double strand AP32-GSH adduct. Lane 1, AP32; lane 2, 3'-P; lane 3-5, APE2 react with double strand AP32 for 0, 10, 30 min; lane 6, AP32-GSH adduct; lane 7-9, APE2 react with double strand AP32-GSH adduct for 0, 10, 30 min. **B**. The domain architecture of APE1 and APE2, APE2 contains an Endo/Exo/Phosphatase catalytic domain and a linker PIP followed by the C-terminal Zf-GRF domain, the Zf-GRF makes it unique from APE1.

4.5 Conclusion

In the work presented here, we characterized the strand break products produced during the formation of AP-DPCs through comparing the strand breaks with standard cleavage substrates and verified with mass spectrometry. We prepared the strand break standard using heat, strong base (NaOH), spermine, GSH and Endo III. Our finding explains the complexity of the strand breaks generated during the DPC formation and the involvement of GSH during the interaction of AP sites with TFAM. We revealed that the GSH-adduct is the most abundant strand breaks generated during the DPC reaction if GSH is present. Under different sequence context, there are different strand breaks generated during DPC reactions. Since DPC reaction produces GSH-adduct as the only major strand break, this result would simplify the repair process of strand breaks to be only target to

the GSH-adduct. Our finding supports the finding by Haldar and Bailly et al. that they found the involvement of GSH in the AP sites β-elimination products and form the GSH-adduct instead of going through further δ-elimination or hydrolysis [22,27].

We further tested the impact of reaction rates when GSH or spermine is present in AP-TFAM DPC formation. We observed that GSH didn't impact the rates, but decreased the yield of DPC and increased the yield of SSBs. By the addition of spermine, the reaction was accelerated without significant change in the product yields. And in presence of GSH and spermine, both the rates and yield of SSBs dramatically increased, indicating a cooperative effect of GSH and spermine that involved in competing with TFAM for the active sites. These results also indicated that the reaction started from forming the Schiff base intermediate and followed by β-elimination, the resulting iminium ion can react with GSH or other residues from proteins to generate other strand break products. Since the involvement of GSH happened after β-elimination, the rate impact from GSH is minimal, but for spermine, it contains primary amines and could react with AP sites to form Schiff base, therefore, spermine is able to accelerate the reaction rates. Because spermine isn't involved in the following reaction with the iminium ions, the reaction products were not impacted by its involvement.

By comparing the cleavage efficiency of APE1, APE2 and TDP1 with 3'-GSH adduct. The abundant AP site repair enzyme APE1 has higher reactivity than TDP1, but APE1 has more strict substrate requirements that hinders its wild application. Less substrate

specificity of TDP1 enabled it to resolve the complex damaged 3’-end and facilitate the repair with the subsequent enzymes. The preliminary data from APE2 displayed much higher cleavage efficiency with 3’-GSH adduct than AP sites and indicating APE2 is an important damage response enzyme.

Overall, our results revealed that the major strand break product produced during DPC reactions is the 3’-thiol adduct when in the presence of thiols and this would simplify the repair of AP strand breaks because AP sites can turn into multiple types of strand breaks. We also analyzed the impact of GSH and spermine during the DPC reaction and our result revealed the reaction kinetics and the mechanism of their involvement during the DPC formation. Since the 3’-GSH is the major strand break, play as a block for DNA functions, we also tested several DNA repair enzymes on their cleavage activity of 3’-GSH adduct and found out that APE1, APE2 and TDP1 have efficient activity towards cleaving the GSH-adducts. This work is ongoing research, and the rest part will be finished with our group members.

4.6 References

[1] S. Boiteux, M. Guillet, Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae, DNA Repair (Amst). 3 (2004) 1–12. https://doi.org/10.1016/j.dnarep.2003.10.002.

[2] T. Lindahl, B. Nyberg, Rate of depurination of native deoxyribonucleic acid, Biochemistry. 11 (1972) 3610–3618. https://doi.org/10.1021/bi00769a018.

[3] A.R. Poetsch, The genomics of oxidative DNA damage, repair, and resulting mutagenesis, Comput. Struct. Biotechnol. J. 18 (2020) 207–219. https://doi.org/10.1016/j.csbj.2019.12.013.

[4] A.K. McCullough, M.L. Dodson, R.S. Lloyd, Initiation of Base Excision Repair: Glycosylase Mechanisms and Structures, Annu. Rev. Biochem. 68 (1999) 255–285. https://doi.org/10.1146/annurev.biochem.68.1.255.

[5] S.-L. Yu, S.-K. Lee, R.E. Johnson, L. Prakash, S. Prakash, The Stalling of Transcription at Abasic Sites Is Highly Mutagenic, Mol. Cell. Biol. 23 (2003) 382–388. https://doi.org/10.1128/MCB.23.1.382-388.2003.

[6] J.A. Wilde, P.H. Bolton, A. Mazumder, M. Manoharan, J.A. Gerlt, Characterization of the equilibrating forms of the aldehydic abasic site in duplex DNA by oxygen-17 NMR, J. Am. Chem. Soc. 111 (1989) 1894–1896. https://doi.org/10.1021/ja00187a062.

[7] P.S. Thompson, D. Cortez, New insights into abasic site repair and tolerance, DNA Repair (Amst). 90 (2020) 102866. https://doi.org/10.1016/j.dnarep.2020.102866.

[8] V. Bailly, W.G. Verly, Possible roles of β -elimination and δ -elimination reactions in the repair of DNA containing AP (apurinic/apyrimidinic) sites in mammalian cells, Biochem. J. 253 (1988) 553–559. https://doi.org/10.1042/bj2530553.

[9] J.S. Jha, C. Nel, T. Haldar, D. Peters, K. Housh, K.S. Gates, Products Generated by Amine-Catalyzed Strand Cleavage at Apurinic/Apyrimidinic Sites in DNA: New Insights from a Biomimetic Nucleoside Model System, Chem. Res. Toxicol. 35 (2022) 203–217. https://doi.org/10.1021/acs.chemrestox.1c00408.

[10] M. Hüttemann, I. Lee, L. Samavati, H. Yu, J.W. Doan, Regulation of mitochondrial oxidative phosphorylation through cell signaling, Biochim. Biophys. Acta - Mol. Cell Res. 1773 (2007) 1701–1720. https://doi.org/10.1016/j.bbamcr.2007.10.001.

[11] D.L. Croteau, R.. Stierum, V.A. Bohr, Mitochondrial DNA repair pathways, Mutat. Res. Repair. 434 (1999) 137–148. https://doi.org/10.1016/S0921-8777(99)00025-7.

[12] S.R. Kennedy, L.A. Loeb, A.J. Herr, Somatic mutations in aging, cancer and neurodegeneration, Mech. Ageing Dev. 133 (2012) 118–126. https://doi.org/10.1016/j.mad.2011.10.009.

[13] K.W. Caldecott, Single-strand break repair and genetic disease, Nat. Rev. Genet. 9 (2008) 619–631. https://doi.org/10.1038/nrg2380.

[14] A. Kuzminov, Single-strand interruptions in replicating chromosomes cause double-strand breaks, Proc. Natl. Acad. Sci. 98 (2001) 8241–8246. https://doi.org/10.1073/pnas.131009198.

[15] T. Izumi, T.K. Hazra, I. Boldogh, A.E. Tomkinson, M.S. Park, S. Ikeda, S. Mitra, Requirement for human AP endonuclease 1 for repair of 3′-blocking damage at DNA single-strand breaks induced by reactive oxygen species, Carcinogenesis. 21 (2000) 1329–1334. https://doi.org/10.1093/carcin/21.7.1329.

[16] M.L. Hegde, T. Izumi, S. Mitra, Oxidized Base Damage and Single-Strand Break Repair in Mammalian Genomes, in: 2012: pp. 123–153. https://doi.org/10.1016/B978-0-12-387665-2.00006-7.

[17] J.T. Sczepanski, R.S. Wong, J.N. McKnight, G.D. Bowman, M.M. Greenberg, Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle, Proc. Natl. Acad. Sci. 107 (2010) 22475–22480. https://doi.org/10.1073/pnas.1012860108.

[18] W. Xu, R.M. Boyd, M.O. Tree, F. Samkari, L. Zhao, Mitochondrial transcription factor A promotes DNA strand cleavage at abasic sites, Proc. Natl. Acad. Sci. 116 (2019) 17792–17799. https://doi.org/10.1073/pnas.1911252116.

[19] J. Tang, W. Zhao, N.G. Hendricks, L. Zhao, High-Resolution Mapping of Amino Acid Residues in DNA–Protein Cross-Links Enabled by Ribonucleotide-Containing DNA, Anal. Chem. 93 (2021) 13398–13406. https://doi.org/10.1021/acs.analchem.1c03481.

[20] W. Xu, J. Tang, L. Zhao, DNA–protein cross-links between abasic DNA damage and mitochondrial transcription factor A (TFAM), Nucleic Acids Res. 51 (2023) 41–53. https://doi.org/10.1093/nar/gkac1214.

[21] J.S. Jha, J. Yin, T. Haldar, Z. Yang, Y. Wang, K.S. Gates, Reconsidering the Chemical Nature of Strand Breaks Derived from Abasic Sites in Cellular DNA: Evidence for 3′-Glutathionylation, J. Am. Chem. Soc. 144 (2022) 10471–10482. https://doi.org/10.1021/jacs.2c02703.

[22] T. Haldar, J.S. Jha, Z. Yang, C. Nel, K. Housh, O.J. Cassidy, K.S. Gates, Unexpected Complexity in the Products Arising from NaOH-, Heat-, Amine-, and Glycosylase-Induced Strand Cleavage at an Abasic Site in DNA, Chem. Res. Toxicol. 35 (2022) 218–232. https://doi.org/10.1021/acs.chemrestox.1c00409.

[23] G.T. Lountos, X.Z. Zhao, E. Kiselev, J.E. Tropea, D. Needle, Y. Pommier, T.R. Burke, D.S. Waugh, Identification of a ligand binding hot spot and structural motifs replicating aspects of tyrosyl-DNA phosphodiesterase I (TDP1) phosphoryl recognition by crystallographic fragment cocktail screening, Nucleic Acids Res. 47 (2019) 10134–10150. https://doi.org/10.1093/nar/gkz515.

[24] B.D. Wallace, Z. Berman, G.A. Mueller, Y. Lin, T. Chang, S.N. Andres, J.L. Wojtaszek, E.F. DeRose, C.D. Appel, R.E. London, S. Yan, R.S. Williams, APE2 Zf-GRF facilitates 3′-5′ resection of DNA damage following oxidative stress, Proc. Natl. Acad. Sci. 114 (2017) 304–309. https://doi.org/10.1073/pnas.1610011114.

[25] W. Zhao, W. Xu, J. Tang, S. Kaushik, C.-E.A. Chang, L. Zhao, Key Amino Acid Residues of Mitochondrial Transcription Factor A Synergize with Abasic (AP) Site Dynamics To Facilitate AP-Lyase Reactions, ACS Chem. Biol. (2023). https://doi.org/10.1021/acschembio.3c00047.

[26] X. Wei, Z. Wang, C. Hinson, K. Yang, Human TDP1, APE1 and TREX1 repair 3′-DNA–peptide/protein cross-links arising from abasic sites in vitro, Nucleic Acids Res. 50 (2022) 3638–3657. https://doi.org/10.1093/nar/gkac185.

[27] V. Bailly, W.G. Verly, Escherichia coli endonuclease III is not an endonuclease but a β-elimination catalyst, Biochem. J. 242 (1987) 565–572. https://doi.org/10.1042/bj2420565.

Chapter 5. Future Perspectives

5.1 Potential research areas

5.1.1 DNA-protein crosslinks

In Chapter 2, we discussed a method for preparing DNA protein cross-link samples for mass spectrometry analysis and provided a data analysis tool. Given that our focus is only on the cross-links generated during the interaction between AP-DNA and TFAM, it's more meaningful if this method could expand to other systems, like other proteins or even detecting cross-linked amino acid residues in the cells.

However, there are limitations in accurately defining the sequence of the cross-linked peptides because, during the fragmentation process, the bonds from DNA have lower energy states, so they tend to be broken and generate fragment ions [1]. With such a large number of fragments coming from breaking DNA bonds, there are limited fragment ions coming from breaking the peptide bonds, display low intensities and low signal-to-noise ratios. This issue is very problematic for any software to differentiate these ions from the background noises and not able to assign them as the fragments coming from the precursor ions.

The software designed for analyzing the sequence information will form a list of fragment ions. Therefore, the complicated fragments from DPCs also disabled all software tools that only focus on peptide fragmentations which couldn't handle the more complex

fragment spectrum. But if cross-linked DNA can be digested into single nucleotides, the fragmentation pattern will be closer to the peptide fragmentation pattern, and it's easier to solve the peptide sequence and find the cross-linked sites. We can treat the cross-linked DNA as a modification of protein and if that DNA unit doesn't form fragments (like AP sites), the fragmentation pattern is similar to the modified peptides and so people can use widely used software for proteins analysis with the DPC samples, such as MaxQuant [2], FragPipe [3] and Proteome Discoverer [4]. Since the majority of the DPC formed under UV light or ionizing radiation are cross-linked via the DNA bases, and the cleavage of the base from the DNA is also favored during MS/MS, the idea of finding out the cross-linked amino acid sites using MS2 and don't require the complete digestion of DNA into single nucleotides, but the insufficient digestion of DNA might impact the ionization efficiency of DPC if the leftover DNA is too long. The workflow is shown, first, ionize the DPC sample in positive ion mode. Since the glycosidic bonds have a very low energy level and can easily break during CID. Second, fragment the DPC sample under CID with relatively lower energy that is efficient to break the glycosidic bonds or other DNA bonds but not efficient to cleave any peptide bonds. The MS/MS spectrum will display at least two ions, the small m/z ions might be the free base groups (lower than m/z = 200) or other small fragments from peptides, the DNA strand will be very less likely to carry a positive charge at this point, another high intensity ion might be the crosslinked peptide carry a cross-linked base group. Later in MS3, use higher CID energy that can break peptide bonds and therefore the sequence information can be obtained from MS3, the crosslinked site will

be carrying a cross-linked base, if the cross-linking chemistry is known, the mass of the base/modified base should be known, and we are solving the cross-linked sites with the defined adduct mass.

Another method to try is using ETD [5] instead of CID. With ETD in the MS2, the fragments prefer to evade the DNA part since DNA tends to be negatively charged and is intended to break peptide bonds. However, this method is limited because the DNA bond energy is still lower than peptide bonds, and digesting DNA into single nucleotides will be helpful to eliminate the majority of the fragments from breaking DNA bonds.

5.1.2 DNA interstrand cross-links

In Chapter 2, we have discussed a method to prepare AP-ICLs. Even though there is no direct evidence of their existence in the cells, indirect evidence shows their repaire with DNA glycosylase NEIL3 [6,7]. AP-ICLs may be present in cells, and since they repaired, there will be limited amount of stable AP-ICLs in the cells that are detectable. Since ICL repair is associated with breast cancer tumor suppressor genes, applying AP-ICL models into investing DNA repair enzymes and DNA polymerases, and understanding how cells sense AP-ICLs and recruit proteins to fit them, will be important and might be relevant to answering cancer-related questions. The methods that we provided can apply to other types of AP-ICLs and other types of ICLs to prepare unhooked ICL repair intermediate. With the many models of unhooked ICLs available, the experimental design will not be limited by the availability of ICLs.

5.1.3 Single strand breaks in DPC reactions

As shown in Chapter 4, our group found out that 3'-GSH-adduct is the only major single-strand break generated during DPC reaction and together with our lab's previous finding that cysteine residues on TFAM are the most abundant residue that cross-linked with AP sites [8], these results stressed the involvement of thiols in the DPC reactions and have an impact on the results of strand breaks and cross-linked amino acids. Biological thiols are widely present in cells, GSH, an abundant tripeptide in cells, is discovered with more functions besides modulating oxidative and reduction states in cells. The involvement of GSH is beneficial for AP site repair. We showed in Chapter 4 that with the spermine present, or possibly with other amines in cells, the rate of the intact DNA disappearance was dramatically increased, meaning the formation of strand breaks or DPCs are catalyzed by amine-containing molecules in cells, and when amines coexist with GSH, the most abundant product of AP-DNA is the 3'-GSH-adduct with a low amount of DPC generated. Since repairing small DNA lesions is usually more efficient than repairing bulky DNA lesions, the damaging effects of 3'-GSH-adducts are relatively lower than DPCs and decrease the complexity of repairing AP sites.

The next possible step of further looking into this topic is to test the effect of GSH and spermine involvement in other AP-DPC systems and if the trend is similar, this might reveal a modulation system of cells to turn AP sites into GSH-adducts and can be efficiently repaired with APE1 or TDP1 before AP sites react with other molecules in cells. Also, other experiments can be done with the stable 3'-GSH-adduct, for example,

reduced 3’-GSH-adduct, the DNA substrate with reduced 3’-GSH-adduct can work as the bait for fishing the proteins that have high affinity to it to reveal the possible repair enzymes in cells.

5.2 References

[1] J. Tang, W. Zhao, N.G. Hendricks, L. Zhao, High-Resolution Mapping of Amino Acid Residues in DNA–Protein Cross-Links Enabled by Ribonucleotide-Containing DNA, Anal. Chem. 93 (2021) 13398–13406. https://doi.org/10.1021/acs.analchem.1c03481.

[2] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372. https://doi.org/10.1038/nbt.1511.

[3] V. Demichev, L. Szyrwiel, F. Yu, G.C. Teo, G. Rosenberger, A. Niewienda, D. Ludwig, J. Decker, S. Kaspar-Schoenefeld, K.S. Lilley, M. Mülleder, A.I. Nesvizhskii, M. Ralser, dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts, Nat. Commun. 13 (2022) 3944. https://doi.org/10.1038/s41467-022-31492-0.

[4] B.C. Orsburn, Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics, Proteomes. 9 (2021) 15. https://doi.org/10.3390/proteomes9010015.

[5] W.-C. Ahn, S. Aroli, J.-H. Kim, J.H. Moon, G.S. Lee, M.-H. Lee, P.B. Sang, B.-H. Oh, U. Varshney, E.-J. Woo, Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision, Nat. Chem. Biol. 15 (2019) 607–614. https://doi.org/10.1038/s41589-019-0289-3.

[6] M. Imani Nejad, K. Housh, A.A. Rodriguez, T. Haldar, S. Kathe, S.S. Wallace, B.F. Eichman, K.S. Gates, Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3, DNA Repair (Amst). 86 (2020) 102752. https://doi.org/10.1016/j.dnarep.2019.102752.

[7] D.R. Semlow, J. Zhang, M. Budzowska, A.C. Drohat, J.C. Walter, Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase, Cell. 167 (2016) 498-511.e14. https://doi.org/10.1016/j.cell.2016.09.008.

[8] W. Xu, J. Tang, L. Zhao, DNA–protein cross-links between abasic DNA damage and mitochondrial transcription factor A (TFAM), Nucleic Acids Res. 51 (2023) 41–53. https://doi.org/10.1093/nar/gkac1214.