The Molecular Mechanism of Replication Independent Repair of DNA Interstrand Crosslinks

Niyo Kato

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

Columbia University

2018
Abstract

The Molecular Mechanism of Replication Independent Repair of DNA Interstrand Crosslinks

Niyo Kato

DNA interstrand crosslinks (ICLs) are a potent type of DNA damage that arise as a consequence of normal cell metabolism. By covalently linking opposing strands of the double helix, ICLs block essential DNA transactions such as replication, transcription, and recombination. If unrepaired, or incorrectly repaired, ICLs can lead to gross genome instability and cell death. This cytotoxicity has been exploited in the clinic, where ICL inducing drugs are among the oldest and most widely prescribed anti-cancer therapies. However, acquired resistance is a significant limitation of these drugs, and the mechanism by which this occurs remains largely elusive.

In order to develop more effective ICL-based therapies, it is imperative to first fully elucidate how healthy cells respond to and repair ICLs. Moreover, better understanding ICL repair mechanisms is necessary to fully unravel the complex DNA repair networks that govern genomic integrity, and understand the physiology of diseases such as Fanconi Anemia, which result from the inability to efficiently repair ICL lesions.

Multiple mechanisms of ICL repair exist, and repair pathway choice is primarily determined by the phase of the cell cycle. In proliferating cells, the ICL repair occurs during S-phase, and in a process termed “replication coupled repair” (RCR). In contrast, slowly or non-dividing cells rely on an alternative modality of repair called “replication independent repair” (RIR). RIR is critical for homeostasis and survival in quiescent healthy cells that (for example, neurons) and in cycling cells deficient for replication coupled repair proteins (i.e. Fanconi Anemia cells).
Despite its importance, little is known about RIR. This is due, in part, to the fact that ICL repair has been primarily studied in systems, such as cultured cells, that favor RCR and are therefore bias against RIR. More recently, non-replicating *Xenopus* cell-free extracts has emerged as a powerful system to study RIR. This system faithfully recapitulates RIR and has been instrumental in identifying DNA polymerase kappa (Pol κ) and the eukaryotic sliding clamp, proliferating cell nuclear antigen (PCNA), as two critical RIR factors. However, other important RIR factors are yet to be identified.

ICL repair is unique among DNA repair pathways as it harnesses proteins from diverse DNA repair pathways including, Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mismatch Repair (MMR), and Double Strand Break Repair (DSBR). Chapter 1 provides an overview of these pathways including the types of DNA damage that each pathway responds to, key steps of the repair process, and the corresponding proteins that are involved. This chapter provides context for the rest of the thesis in which I explore the contribution of multiple DNA repair proteins on the repair of ICL lesions.

In Chapter 2, I detail our studies assessing the contribution of the MMR machinery to RIR. We show that the mismatch repair sensor, MutSa complex (MSH2-MSH6), is critical for ICL recognition, and the stepwise recruitment of other MMR proteins including MutLα (MLH1-PMS2) and EXO1. In this chapter, I also investigate how ICL structure influences repair. I find that more distorting ICLs use an MMR-dependent ICL repair mechanism, while less distorting ICLs are repaired MMR-independently (see also Appendix A), or not repaired at all. Appendix B further explores the contribution of the MMR pathway on ICL repair in mammalian cells.
Finally, in Appendix C and D we provide further evidence that RIR is fundamentally distinct from replication coupled ICL repair, as depletion of key RCR proteins from our extracts yields no phenotype.

I summarize all of these findings in Chapter 3, and discuss their implications to the DNA repair field as well as the clinic, where crosslinker drugs remain a mainstay in the treatment of cancer.
# Table of Contents

List of Tables ............................................................................................................. iv

List of Figures ............................................................................................................ iv

List of Abbreviations and Acronyms .......................................................................... vi

Acknowledgements .................................................................................................... vii

Chapter 1. Introduction .............................................................................................. 1

DNA Damage Types and DNA Repair Pathways ..................................................... 2

  Base Excision Repair .............................................................................................. 3
  Nucleotide Excision Repair .................................................................................. 4
  Mismatch Repair .................................................................................................. 5
  Double Strand Break Repair ................................................................................ 6

DNA Interstrand Crosslinks and ICL Repair Pathways .......................................... 7

  Sources of ICLs .................................................................................................... 8
  Structures of ICLs ................................................................................................ 9
  Clinical Relevance of ICLs .................................................................................. 11
    Crosslinker Drugs in Chemotherapy ................................................................. 12
    Fanconi Anemia ............................................................................................... 13
  Experimental Systems to Study ICL Repair ..................................................... 15
    Xenopus laevis Cell-Free Extracts ................................................................. 17
  ICL Repair Mechanisms .................................................................................... 20
    Replication Coupled Repair ......................................................................... 21
    Fanconi Anemia Dependent ICL Repair Pathway ...................................... 22
    NEIL3 Dependent ICL Repair Pathway ...................................................... 24
References......................................................................................................................... 102

Appendix............................................................................................................................... 112

Appendix A. Structure Specific Repair of ICLs during RIR.................................................. 113
Appendix B. MMR Deficient Mammalian Cells are Sensitive to MMC............................... 118
  Appendix B. Methods.......................................................................................................... 122
Appendix C. The Role of Pol δ, Pol ε, and Rev1 in RIR......................................................... 124
Appendix D. The Role of CtIP in RIR.................................................................................. 128
Appendix E. Homemade Antibodies.................................................................................... 130
  MSH2 Antibody.................................................................................................................. 130
  XPC Antibody.................................................................................................................... 133
List of Tables

Table 1-1 DNA Lesions generated by endogenous and exogenous sources..........................34
Table 1-2 Crosslinker drugs and ICL structure...............................................................35
Table 1-3 Crosslinker drugs used in the clinic...............................................................36

List of Figures

Chapter 1
Figure 1-1. Sources and types of DNA Damage..........................................................37
Figure 1-2. DNA repair pathways for Base Excision Repair, Nucleotide Excision Repair, and
Mismatch Repair ........................................................................................................38
Figure 1-3. DNA repair pathways of Double Strand Break Repair.................................39
Figure 1-4. Pathways of ICL repair................................................................................40
Figure 1-5. Pathways of Replication-coupled ICL repair..............................................41
Figure 1-6. Schematic of preparation of Xenopus cell-free extracts (High Speed Supernatant)....43
Figure 1-7. Conceptual model of Replication Independent ICL Repair..............................44

Chapter 2
Figure 2-1. ICL sensing by MutSal (MSH2-MSH6).........................................................71
Figure 2-2. Mechanism of ICL recognition by MutSal (MSH2-MSH6).............................72
Figure 2-3. Nucleolytic incision of ICL lesions by MutLal (MLH1-PMS2).........................73
Figure 2-4. Nucleolytic processing of ICL lesions by EXO1...........................................74
Figure 2-5. ICL structure and repair efficiency...............................................................75
Figure 2-6. Model for MMR mediated ICL repair.........................................................76
Figure S2-1. Schematic of ICL plasmids and preparation of SJG-136 treated plasmids............77
Figure S2-2. MutSα and PCNA mutants cloned and purified ..............................................80
Figure S2-3. Nicking sites on the ICL plasmid and purification of MutLα ..........................82
Figure S2-4. RPA is required for RIR .............................................................................83
Figure S2-5. Sequences of site specific trimethylene and NM-like ICL plasmids ............ 84

Chapter 3
Figure 3-1. Mechanism of MMR vs. RIR .................................................................100
Figure 3-2. Mechanisms of ICL repair throughout the cell cycle .............................101

Appendix
Figure A 1. Nicking experiments with NM-like ICL in HSS ........................................116
Figure A 2. Repair of Ap-ICL in HSS ........................................................................117
Figure B 1. shRNA mediated knockdown of MSH6 and PMS2 in MCF10A cells ......... 120
Figure B 2. Sensitivity of MSH6- and PMS2-deficient MCF10A cells to mitomycin C .... 121
Figure C 1. The role of Rev1 during RIR .................................................................... 127
Figure D 1 The role of CtIP during RIR ...................................................................... 129
Figure E 1. Homemade MSH2 antibodies for Western blot and immunodepletion .... 132
Figure E 2. XPC fragment used to generate X. l. XPC antibody ................................ 135
Figure E 3. Homemade XPC antibodies recognize XPC528 in BL21 whole cell lysate .... 136
Figure E 4. XPC antibodies specifically recognize UV-damaged Xenopus nuclei ......... 137
Figure E 5. XPC depletion results in defects in trimethylene ICL repair ..................... 138
# List of Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diepoxybutane</td>
<td>DEB</td>
</tr>
<tr>
<td>NM-like ICL (8 atom linker)</td>
<td>NM-like ICL</td>
</tr>
<tr>
<td>Abasic ICL</td>
<td>Ap-ICL</td>
</tr>
<tr>
<td>Abasic Site</td>
<td>Ap site</td>
</tr>
<tr>
<td>Base Excision Repair</td>
<td>BER</td>
</tr>
<tr>
<td>Carmustine</td>
<td>BCNU</td>
</tr>
<tr>
<td>Double Strand Break</td>
<td>DSB</td>
</tr>
<tr>
<td>Double Strand Break Repair</td>
<td>DSBR</td>
</tr>
<tr>
<td>Fanconi Anemia</td>
<td>FA</td>
</tr>
<tr>
<td>Global Genome</td>
<td>GG-NER</td>
</tr>
<tr>
<td>Nucleotide Excision Repair</td>
<td>GG-NER</td>
</tr>
<tr>
<td>Homologous Recombination</td>
<td>HR</td>
</tr>
<tr>
<td>Interstrand Crosslink</td>
<td>ICL</td>
</tr>
<tr>
<td>Mismatch Repair</td>
<td>MMR</td>
</tr>
<tr>
<td>Mytomycin C</td>
<td>MMC</td>
</tr>
<tr>
<td>Nitrogen Mustard</td>
<td>NM</td>
</tr>
<tr>
<td>Non homologous End Joining</td>
<td>NHEJ</td>
</tr>
<tr>
<td>Nucleotide Excision Repair</td>
<td>NER</td>
</tr>
<tr>
<td>Quantitative Polymerase Chain Reactions</td>
<td>qPCR</td>
</tr>
<tr>
<td>Replication Coupled Repair</td>
<td>RCR</td>
</tr>
<tr>
<td>Replication Independent Repair</td>
<td>RIR</td>
</tr>
<tr>
<td>Transcription Coupled Nucleotide Excision Repair</td>
<td>TC-NER</td>
</tr>
<tr>
<td>Translesion Synthesis</td>
<td>TLS</td>
</tr>
<tr>
<td>Trimethylene CpG-ICL</td>
<td>CpG-ICL</td>
</tr>
<tr>
<td>Trimethylene GpC-ICL</td>
<td>GpC-ICL</td>
</tr>
<tr>
<td>Ultra Violet</td>
<td>UV</td>
</tr>
<tr>
<td>Variable, Diversity, and Joining</td>
<td>V(D)J</td>
</tr>
</tbody>
</table>
Acknowledgements

I cannot thank enough my advisor, Dr. Jean Gautier, for his support over the past 6 years. From the very beginning, Jean has patiently nurtured my growth as a scientist and researcher, and has always encouraged me to take ownership of my work. Jean’s door is always open, and I have really appreciated being able to stop in at any time for some scientific advice, a check-in about my work, or most frequently just for some jokes. Coming to lab and working with Jean has been a true joy.

I adore all the members of the Gautier Lab, both past and present. I appreciate everybody indulging me in conversation and activities, scientific or otherwise, that has made my time in the lab most fruitful. I take with me so many great memories – late nights at Dallas or The Heights, bowling, gardening and arts & crafts in the lab, to name a few.

Many members of the Department of Biological Sciences have made an incredible impact on my time as a graduate student. Dr. Tulle Hazzelrig and Dr. Lars Dietrich are incredible PIs with whom I’ve had the pleasure of working with. My peers, Lalanti Venkatasubramanian and Chinweike Okegbe are two of the most inspired individuals that are most dear to me. I have learned so much about myself from all of you, and for this I am immensely grateful.

Dr. Lorraine Symington and Dr. Alberto Ciccia have been part of my thesis committee since the beginning. I have always admired how Lorraine and Alberto both ask the most astute questions with such finesse, and respect how sincerely they engage with researchers at all stages of their careers. I’d like to also acknowledge Dr. Songtao Jia
and Dr. Eric Alani who agreed to be part of my thesis defense. I am tremendously appreciative of your time and support.

This work would not have been possible without the loving support of my family. My parents, Nobuhiro and Naoko Kato and my brothers, Renkoh, Masanari, and Hidemasa have unconditionally supported me throughout this process. Going home on weekends, and vacationing together each year has been an incredibly important way for me to rest and rejuvenate.

Finally, I’d like to thank my husband Kupal, and our dog Cali. I could not have asked for a more devoted, life-loving, and silly family to go home to everyday! Kupal and Cali have been my grounding force through it all, especially when things were challenging. With Kupal I have also gained a most wonderful second family, who provide me with the best belly laughs and incredible food anyone could ask for.

Thank you everyone! I am truly blessed to have had all of your love and support during these past 6 years. I am so proud of all that I’ve accomplished, and am so grateful for all that you’ve done for me during this time.

Acknowledgements specific to each chapter:

Chapter 1 and 3: These chapters were written by Niyo Kato

Chapter 2: The contents of the chapter including text and figure layouts were modified for the purposes of this dissertation. The full list of authors and their contribution to this work can be found in the Acknowledgement section of this chapter.

Appendices: Appendix A-E were written by Niyo Kato.
Chapter 1. Introduction
DNA Damage Types and DNA Repair Pathways

DNA is the carrier of genetic information for all living organisms. Its integrity is essential to allow for cellular and organismal homeostasis, and to ensure the transmission of intact genetic material during reproduction. DNA, however, undergoes spontaneous rearrangement, and is also challenged by various endogenous reactive species and exogenous sources of damage including irradiation and exposure to chemical mutagens (summarized in Table 1-1, and illustrated in Figure 1-1).

It is estimated that on average, tens of thousands of damaged lesions are incurred per mammalian cell per day (Ciccia and Elledge, 2010). These lesions vary from ribonucleotide incorporation or single-base alterations, to more complex damage that involves both strands of the DNA helix (summarized in Table 1-1, and illustrated in Figure 1-1). Generally speaking, simpler DNA lesions that involve a single base are less distorting, while bulkier adducts that involve multiple bases or larger chemical modifications to DNA bases can significantly distort the DNA molecule.

Types of DNA damage that occur spontaneously at a single base include replication errors, which result in base-base mismatches, or bases that undergo spontaneous depurination or deamination, that result in the generation of abasic sites and base transitions. DNA bases can also be chemically modified by reactive oxygen and nitrogen species. Bulkier DNA adducts such as pyrimidine dimers (a linkage between neighboring thymine bases) arise as a result of sun exposure, a common source of UV irradiation. Large monoadducts and intrastrand crosslinks occur when DNA reacts with endogenous aldehydes that are produced during cellular metabolism. Aldehydes, and other bifunctional reactive chemicals can react with opposing strands of the double helix, to
generate DNA interstrand crosslinks (ICLs). Finally, DNA strand breaks, either single or double stranded, are generated in many ways, including exposure to ionizing radiation or as intermediates during multiple DNA repair pathways.

When unrepaired or incorrectly repaired, DNA damage can lead to mutation and chromosome aberrations, which in turn cause aging, disease, or death. Cells have evolved mechanisms to quickly and efficiently identify and repair DNA damage. This sophisticated network of mechanisms is collectively termed the DNA damage response (DDR) (reviewed in Ciccia and Elledge, 2010). Many pathways of the DDR have been mechanistically described, and include: Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mismatch Repair (MMR), and Double Strand Break Repair (DSBR). Each of these pathways have specialized to respond to specific types/classes of DNA damage. Key aspects of these pathways are highly conserved from bacteria to humans, and for the sake of brevity, only the canonical mammalian DNA repair proteins and pathways will be described here. Emphasis is placed on DNA ICLs and their repair, as these lesions are the primarily focus of this thesis.

**Base Excision Repair**

The Base Excision Repair (BER) pathway removes small alterations at single DNA bases that have been oxidized or deaminated. While these lesions are minimally distorting to the DNA, they can be mutagenic and/or cytotoxic to cells.

Damage sensing in the BER pathway is initiated by one of at least 11 damage-specific DNA glycosylases. Glycosylases are enzymes that recognize and excise damaged DNA
bases. Nucleobase removal results in an abasic site (Ap site), which must be further processed. In canonical BER, APE1 (apurinic/apyrimidinic endonuclease), generates a single strand break 5’ to the original damaged base, which allows for the replacement and repair of the damaged nucleotide by translesion synthesis polymerase beta (Pol β) and XRCC1-Ligase III (X-ray repair cross-complementing protein 1), which seals the remaining nick to restore the DNA to its undamaged state. BER that requires the removal of a single base is called “short patched” (illustrated in Figure 1-2), while “long patch” BER requires other factors including DNA Polymerase δ (Pol δ), proliferating cell nuclear antigen (PCNA), replication factor C (RFC), and replication protein A (RPA) for repair, and involves the removal of a longer piece of DNA (Dianov and Hubscher, 2013; Jacobs and Schär, 2012; Krokan and Bjoras, 2013).

Nucleotide Excision Repair

Bulkier DNA lesions such as cyclobutane-pyrimidine dimers (CPDs) and 6-4 pyrimidine pyrimidone photoproducts (6-4PPs) that impart local distortions on the DNA helix are repaired by the nucleotide excision repair (NER) pathway. Two branches of NER exist: global-genome NER (GG-NER) and transcription-coupled NER (TC-NER).

In GG-NER (illustrated in Figure 1-2), the DNA damage sensor protein, XPC-HR23 (xeroderma pigmentosum complementation group C) actively surveils the genome for damage, by binding and recognizing regions of the genome that are slightly unwound. NER can also be initiated by the transcription machinery (TC-NER). Bulky adducts physically block the progression of the transcribing polymerase, and activate repair.
Stalled RNA PolII recruits repair proteins including CSA and CSB (Cockayne Syndrome Protein A and B) as well as UVSSA-USP7 (UV-stimulated-scaffold-protein A and ubiquitin-specific-processing-protease 7), which together dislodge the RNA polymerase complex (Marteijn et al., 2014).

Following damage recognition, TFIIH (transcription initiation factor IIH) complex is recruited for damage verification. Two ATP dependent helicases, XPB and XPD, are among the ten proteins that make up the TFIIH complex, and are essential for unwinding the local DNA region surrounding the damage site (Marteijn et al., 2014; Wood, 2010).

NER involves two incision events flanking the damage site by the XPF-ERCC1 and XPG structure-specific nucleases. XPA and RPA are additional proteins that are required to help coordinate repair events and protect the undamaged DNA strand from being processed. After incisions, the resulting 22-30 nucleotide gap is subsequently repaired in a gap filling reaction by PCNA and a DNA polymerase Pol δ/Pol ε/Pol κ. Nick sealing by DNA ligase I or III completes the repair process (Marteijn et al., 2014).

Mismatch Repair

Base-base mismatches or insertion-deletion loops arise when erroneous incorporation of nucleotides occurs during replication. These lesions are efficiently repaired by the mismatch repair machinery (MMR), which improves the fidelity of DNA replication by 2-3 orders of magnitude.

During MMR, repair is initiated by two partially redundant DNA damage sensor proteins MutSα (MSH2-MSH6) and MutSβ (MSH2-MSH3). Damage recognition is
followed by recruitment of the MutLα (MLH1-PMS2) endonuclease, whose cryptic endonuclease is important for MMR. MutLα specifically incises the nascent strand of DNA that contains the mispair, signaling for the degradation of the incised strand by the EXO1 exonuclease (Jiricny, 2006). It remains unclear how the MMR machinery preferentially targets the nascent strand during eukaryotic MMR reaction. It has been proposed that strand discontinuities in the nascent strand, either between Okazaki fragments on the lagging strand, or nicks/gaps generated during removal of erroneous ribonucleotides incorporated during replication might facilitate this process (Ghodgaonkar et al., 2013). Whatever the case may be, the gapped repair intermediate that remains is filled in by a re-synthesis reaction which involves RPA and DNA polymerase δ or ε.

**Double Strand Break Repair**

Double strand breaks (DSBs) arise endogenously when replication forks collapse, or during normal physiological processes such as V(D)J recombination (Symington and Gautier, 2011). DSBs also arise upon exposure to ionizing radiation. When unrepaired or incorrectly repaired, DSBs can lead to gross chromosomal aberrations.

Two major pathways respond to and repair DSBs (illustrated in Figure 1-3): nonhomologous end joining (NHEJ) and homology directed repair (HDR). NHEJ is the predominant mechanisms for repair in non-cycling cells and takes place with no or minimal processing of DNA ends. In this process, DNA ends are bound by the Ku70/80 heterodimer. DNA-bound Ku complex recruit multiple end joining factors including the
Ser/Thr kinase DNA-PKcs (DNA-dependent protein kinase catalytic subunit), as well as the Artemis endonuclease which participates in end-processing events. The broken DNA ends are then ligated by XRCC4 (X-ray repair cross-complementing protein 4) and Ligase 4.

When DSBs arise during S-phase or G2-phase, they can also be repaired by HDR mechanisms, including homologous recombination (HR). In this process, damaged ends are resected by the MRN (Mre11-Rad50-Nbs1) complex, exonuclease 1 (EXO1), as well as BRCA1 (breast cancer 1) and CtIP (CTBP interacting protein). Resected ends are initially coated with RPA, then replaced with Rad51 in a Rad52- and BRCA1/2-dependent manner. Rad51 coated DNA filaments undergo a homology search and use template-dependent repair synthesis reaction to restore the integrity of the DNA (Symington and Gautier, 2011).

**DNA Interstrand Crosslinks and ICL Repair Pathways**

DNA interstrand crosslinks (ICLs) are a type of DNA damage that covalently link opposing strands of the DNA double helix. ICLs arise endogenously, as a consequence of cellular metabolism, but are also induced at high frequency upon exposure to crosslinker-based chemotherapeutic drugs. Depending on the offending adduct, ICLs can vary greatly in their chemical and physical properties, and also in the degree to which they distort the DNA.

Yet, all ICLs are extremely cytotoxic as they physically block cellular processes that require helix unwinding, such as replication, recombination, and transcription (Deans and
ICLs can also interfere with protein-DNA binding events that are required for cell homeostasis, such as transcription factor binding (Martin et al., 2005).

Cells have evolved multiple mechanisms to sense and remove ICLs from their genomes. ICLs are complex chemical lesions as they involve both strands of the DNA double helix. This complexity is reflected in the multiple pathways that exist to respond to and repair ICL lesions, and the diversity of repair proteins that participate in these processes.

Sources of ICLs

Several classes of bifunctional chemicals, both from endogenous and exogenous sources, are capable of generating ICL lesions (Table 1-2). Endogenous sources of ICLs include various reactive aldehydes that are produced as a byproduct of cellular metabolism. For example, formaldehyde, is a ubiquitous endogenous metabolite that is generated during oxidative histone demethylation and biosynthesis of purines and amino acids. Formaldehyde exposure can also occur through inhalation of tobacco smoke or even through consumption of various foods and vegetables. Another endogenous aldehyde source is acetaldehyde. Acetaldehyde is produced as a byproduct of ethanol oxidation and is also an intermediate of carbohydrate metabolism (Guainazzi and Schärer, 2010). Finally, crotonaldehydes and acroleins are byproduct of lipid peroxidation (Deans and West, 2011).

Some environmental mutagens can also generate ICL lesions. These include psoralens and furocoumarins, which are found in plants such a bergamot, celery, and
parsley (Deans and West, 2011), as well as Mitomycin C (MMC) a chemical produced by the bacteria *Streptomyces caespitosus*.

ICLs can be induced at very high frequencies upon exposure to a diverse array of crosslinker drugs that have been chemically synthesized. These include nitrogen mustards, MMC, and platinum-based drugs. Many of these chemicals are widely and successfully used in the clinic as cancer chemotherapeutic drugs.

Aside from reaction with bifunctional chemicals mentioned above, ICLs can also occur spontaneously by chemical rearrangements at abasic DNA sites. Abasic sites (Ap sites) exist at a steady-state level of 50,000-200,000 Ap sites / cell (Price et al., 2014) as a result of spontaneous depurination or as an intermediate during BER (Dianov and Hubscher, 2013). Since Ap sites are electrophilic, they can react with a base on the opposing strand of the DNA molecule to generate an ICL (Price et al., 2014).

**Structures of ICLs**

Each of the crosslinker sources described above, occurring both endogenous and exogenously, generates ICLs of varying physical and chemical characteristics. This is in part due to the nature of the offending chemical itself, but also a result of the preferential base reactivity that many ICL inducing drugs exhibit for DNA (Deans and West, 2011; Guainazzi and Schärer, 2010). These features directly influence the local DNA conformation to which they are bound, and therefore can also induce various degrees of distortion to the double helix (summarized in Table 1-2).
Among the crosslinker drugs used in the clinic, the cisplatin-induced ICL is perhaps one of the most distorting (Guainazzi and Schärer, 2010). Cisplatin forms preferentially between (N\textsuperscript{7} position) guanine residues at 5’-GpC sequences. The resulting ICL is severely distorting as it shortens the distance between the two strands of the double helix, and send the two neighboring cytosine residues outward to take on an extra-helical conformation (Coste et al., 1999).

In comparison, mitomycin C (MMC) and nitrogen mustard-ICLs are less distorting (Guainazzi and Schärer, 2010). While, MMC ICLs also form between guanine residues, it reacts with 5’-CpG sequences. The resulting ICL lesion only slightly widens the minor groove. Similarly, nitrogen mustard ICL forms between guanine residues, but at 5’-GNC residues and induce only a slight bend in the DNA (Guainazzi et al., 2010). Although Ap-ICLs are similarly thought to be little-distorting (Johnson et al., 2013; Price et al., 2014), the NMR structure has not yet been determined for this type of lesion.

Of note, many crosslinker sources generate ICLs that are inherently unstable on a laboratory timescale, making them difficult to purify and study. For this reason, synthetic analogs have been generated (Guainazzi and Schärer, 2010). The malondialdehyde ICL lesion has been mimicked in two different conformations using a trimethylene linkage. This more stable surrogate links two guanine residues between N\textsuperscript{2} positions at 5’-GpC and 5’-CpG sequences. Both of these synthetic crosslinks are very stable, and their NMR structures have been determined (Dooley et al., 2001, 2003). Nitrogen mustard ICLs have similarly been mimicked in multiple conformations using synthetic linker compounds (Guainazzi et al., 2010; Mukherjee et al., 2014a). While the NMR structures of these ICLs have not been determined, structural simulation studies estimate the conformations
to range from completely non-distorting (8-atom NM linker) to moderately distorting (5-atom NM linker) (Mukherjee et al., 2014a; Roy and Schärer, 2016).

Importantly, all crosslinker drugs generate a wide array of DNA adducts including monoadducts, intrastrand crosslinks (primarily between purines on the same strand of DNA), as well as ICLs. For example, MMC forms ICL lesions with about 15% efficiency. The rest are monoadducts (50%) and intrastrand crosslinks (35%). The efficiency for ICL formation for cisplatin is even lower, with less than 5% of lesions being ICLs.

For this reason, rationally designed crosslinker drugs have been more recently synthesized in an effort to increase the efficiency of ICL formation. One such drug is SJG-136, a synthetic PBD (pyrrolo[2,1-c][1,4]benzodiazepine dimer). SJG-136 preferentially reacts at purine-GATC-pyrimidine sequences, to form ICLs with high efficiency. The resulting SJG-136 ICL lesion forms in the minor groove of DNA and is minimally distorting (Martin et al., 2005; Rahman et al., 2009). Of note, while SJG-136 forms ICLs at much higher efficiency than traditional crosslinker drugs, SJG-136 can also form stable intrastrand crosslinks and mono adducts (Rahman et al., 2009).

Clinical Relevance of ICLs

ICLs are clinically relevant for two major reasons. First, DNA crosslinking agents are among the oldest and most widely prescribed anti-cancer drugs used today. They efficiently target rapidly dividing cancer cells by impeding the metabolic processes they need for cell homeostasis, growth, and proliferation, such as replication and transcription.
Second, the inability to repair ICL lesions is characteristic of a genome instability disorder called Fanconi Anemia. This suggests that endogenously occurring ICLs pose a threat to cells, and that their removal is essential for cellular homeostasis. Paradoxically, ICLs are therefore relevant for both the prevention and treatment of cancers.

**Crosslinker Drugs in Chemotherapy**

Crosslinker based chemotherapy began in 1946 with nitrogen mustard, which was followed by MMC in 1956, platinum based drugs in 1971, and psoralens in 1989. Today, these drugs are among the most effective therapies against various cancers including leukemia, lymphoma, colon, ovarian, bladder, testicular, head and neck cancers, to name a few. The clinical uses of the most common crosslinker drugs are summarized in Table 1-3.

There are key limitations in using crosslinker based chemotherapies. First, acquired resistance is a significant limitation of these drugs (Guainazzi & Schärer, 2010). Cancer cells can increase drug detoxification, reduce drug uptake, or evade crosslinker drug toxicity by accelerating DNA repair mechanisms or via overriding cell death pathways.

For example, the clinical use of cisplatin is thought to be limited by cancer cells that upregulate DNA repair mechanisms. Overexpression of XPF-ERCC1, and some additional FA proteins have been observed in ovarian and testicular cancer cells that are exposed to cisplatin for long periods of time. Melanoma cells have been shown to acquire resistance to fotemustine (NM derivative) by upregulating XPC (Barckhausen, Roos, Naumann, & Kaina, 2014).
A second limitation of crosslinker drugs relate to dose limiting toxicity. Systemic delivery of the drugs results in toxicity in non-target tissues. Cisplatin for example, has been shown to specifically affect the brain, and patients also suffer from hearing loss, neuropathies, and kidney damage (Cheung-Ong, Giaever, & Nislow, 2013). Moreover, patients who undergo crosslinker based chemotherapy regimens often develop acute myeloid leukemia (Deans & West, 2011).

Better understanding how cells respond to and repair DNA in response to crosslinker drugs will lead to improvements in cancer drugs, or lead to the development of combination therapies to help reduce drug toxicity and incidence of recurrence.

Fanconi Anemia

In humans, the inability to repair ICL lesions is associated with a genetic disorder called Fanconi Anemia (FA). FA is a rare disease, with an incidence rate estimated to be ~1 in 100,000 to 400,000. FA was first described by Guido Fanconi in 1927, and today we now know that FA is caused by mutation in at least 21 “FANC” genes (name alphabetically from FANCA to FANCT). Among these, FANCA (65%) is the most commonly mutated gene, followed by FANCC (15%), and FANCG (10%) (D’Andrea & Grompe, 2003; Deans & West, 2011; Walden & Deans, 2014).

FA is inherited in an autosomal-recessive fashion. The only except is FANCB which is X-linked. Clinical manifestations of FA patients are strikingly heterogeneous and include congenital abnormalities (microcephaly, short stature, genital and skeletal malformations), bone marrow failure, and cancer predisposition. Moreover, some patients
have only subtle or none of these phenotypes. Most FA patients, however, eventually develop bone marrow failure, which aids in the diagnosis of FA. FA is a genome instability disorder, and hence cancer, especially acute myeloid leukemia and squamous cell carcinoma of the head and neck, is also common among patients who undergo successful bone marrow transplants (Walden and Deans, 2014).

Despite the clinical heterogeneity of FA patients, FA cells are all characterized by hypersensitivity to crosslinking drugs such as cisplatin, MMC, or diepoxybutane. Upon exposure to crosslinker drugs, FA cells characteristically arrest in G2/S phase, exhibit increased genome instability as assessed by chromosomal breakage tests, and die rapidly as measured by clonogenic survival assays (Kitao & Takata, 2011; Oostra, Nieuwint, Joenje, & de Winter, 2012). These analyses have been used as diagnostic markers for FA patients, helping clinicians identify the now 21 genes associated with FA disease.

The extraordinary sensitivity of FA cells to crosslinker drugs, and not usually other DNA damaging agents such as hydroxyurea, UV, topoisomerase poisons was instrumental in identifying the role of FA proteins to the repair of ICLs. It is now well appreciated that FA proteins act in a common molecular pathway to repair ICL lesions. We have learned a lot about the molecular function of each of these FA proteins, not only in crosslink repair, but also as key players in other genome maintenance mechanisms including replication fork restart and stem cell maintenance (Ceccaldi, Sarangi, & D’Andrea, 2016; Deans & West, 2011).
Earliest investigations into ICL repair were done in bacteria and yeast. Genetic studies studying the sensitivity of cells to crosslinker chemicals helped identify the diverse DNA repair pathways that participate in ICL repair (Lehoczky et al., 2007; McVey, 2010).

However, a limitation to inferring ICL repair capacity from drug sensitivity is that cellular responses to crosslinker drugs are complex. A single crosslinker drug can induce multiple types of lesions (monoadducts, intrastrand crosslinks, and a minority of ICLs), each eliciting multiple repair pathways. This collectively determines the sensitivity or resistance of a drug to cells. The mechanism by which specific proteins contribute to ICL repair cannot be discerned from these studies alone.

Another limitation of these studies is that vertebrate ICL repair is more complex than bacteria or yeast. Indeed, the FA proteins, which play a central role in vertebrate ICL repair are completely absent in bacteria and not well conserved in yeast (McVey, 2010). Therefore FA-dependent ICL repair is fundamentally disparate between vertebrates and lower eukaryotes or prokaryotic organisms. Many studies therefore have utilized mammalian cell lines to study ICL repair. Specifically, cells from FA patients or other human cell lines, deficient of FA genes, have also been studied extensively.

In addition to clonogenicity assays, ICL repair in cells has also been studied using reporter reactivation assays with site specific and structurally homogenous ICLs. In these experiments, ICL repair was monitored using GFP reporter in which GFP signal is only detected upon repair of the ICL lesion (Enoiu et al., 2012a; Williams et al., 2012). UV laser-induced psoralen ICLs have also been studied using immunofluorescence
Lastly, the recruitment of DNA repair proteins to ICLs can be monitored in mammalian cells using an episomal replication-based assay called eChIP (episomal chromatin immunoprecipitation) (Shen et al., 2009; Wang and Li, 2011).

At least 11 FA mouse models have been generated (Bakker et al., 2013). While cells derived from these mice display sensitivity to crosslinker chemicals, the animals themselves, do not present with the developmental defects associated with FA and also do not spontaneously develop bone marrow failure. Yet, information gleaned from FA mouse models have still been tremendously important. For example, FA-deficient mice (FANCD2-/- Aldh2-/-) have been used to show how FA proteins function to protect the genome from damage caused by endogenous aldehydes (Garaycoechea et al., 2012; Langevin et al., 2011; Oberbeck et al., 2014; Pontel et al., 2015).

Biochemical studies using purified proteins and model ICL substrates have been particularly important in determining the molecular mechanism for ICL repair. These studies have provided insight into several catalytic steps in the repair process including incision and trimming events (Klein Douwel et al., 2017; Klein Douwel et al., 2014; Sarkar et al., 2006). Moreover, studying the repair of structurally defined ICL substrates have proved important, since the repair of ICL lesions can be influenced by its structure (Hlavin et al., 2010a).

All of the aforementioned experimental approaches are suitable for studying ICL repair that occurs in the context of replication. However, some of these approaches can also be adapted to study repair that occurs outside S-phase. For example, cells in culture are rapidly dividing and hence primarily undergo replication coupled ICL repair mechanisms, and the contribution of replication-independent mechanisms can be
diminished or lost in cells, unless cells are arrested and studied in G0/G1 phase. This is true of clonogenic survival assays, or immunofluorescence-based psoralen ICL repair studies. Similarly, ICL repair in plasmid-based experiments can be done on non-replicating plasmids. Nevertheless, a key limitation is that studies in G0/G1 arrested cells or non-replicating plasmids cannot precisely discriminate between transcription-initiated repair and sensor-mediated ICL repair mechanisms. *Xenopus laevis* cell-free extracts, described in detail below, has emerged as a particularly useful model system to study sensor-mediated ICL repair.

**Xenopus laevis Cell-Free Extracts**

The African clawed frog, *Xenopus laevis*, cell-free extracts have provided a unique and powerful system to elucidate and dissect multiple molecular mechanisms of ICL repair (Ben-Yehoyada et al., 2009; Klein Douwel et al., 2014, 2014). Adult female frogs can be stimulated to lay hundreds of unfertilized eggs by injection with human chorionic gonadotropin. These eggs are arrested at metaphase of Meiosis II collected and used to prepare cell free extracts through a series of centrifugation steps. These cell-free extracts can be prepared as a replication competent low-speed supernatant (LSS) to study replication coupled ICL repair mechanisms, or prepared as a high-speed supernatant (HSS), which is not competent for replication and is also transcriptionally silent. The latter type of extract is therefore particularly useful to study sensor mediated replication independent ICL repair processes (Figure 1-6).
Remarkably, *Xenopus* extracts contain all DNA repair machinery, allowing the extracts to fully recapitulate mammalian DNA replication and repair processes including BER, NER, MMR, and DSBR. The mechanism of these pathways can be carefully dissected using this biochemically amenable system. Essential proteins can be specifically and quantitatively depleted from extracts without resulting in lethality. Moreover, the extracts can also be supplemented with recombinant proteins, peptides, or drugs.

*Xenopus* extracts are DNA-free, and therefore exogenous DNA substrates including chromatin or plasmids can be introduced to the extract to study DNA dynamics and repair mechanisms. Upon incubation in *Xenopus* extracts, the DNA bind histone proteins and assemble into nucleosomes (Lemaitre et al., 2001). This allows endogenous repair proteins to physically interact and repair the ICL lesion in a highly physiological environment. Using this system, ICL repair has been studied both within the context of sperm chromatin (Raschle et al., 2015) and also plasmid based systems (Ben-Yehoyada et al., 2009; Kato et al., 2017; Räschle et al., 2008; Williams et al., 2012).

The advantage of using a plasmid based system is that a single, structurally defined ICL lesion can be monitored in the absence of any other type of DNA damage (Ben-Yehoyada et al., 2009; Räschle et al., 2008; Williams et al., 2012). Purifying the ICL plasmid from extract during repair and processing and analyzing the proteins that are associated with the substrate can be monitored by ChIP or plasmid pull-down experiments. Additionally, the repair of the ICL site can be monitored either by radio-nucleotide incorporation, or by quantitative PCR.
Insights into ICL repair, both in and out of the context of replication, gleaning from work using the *Xenopus* extract system is described in detail in the following sections.
ICL Repair Mechanisms

If unrepaired or inappropriately repaired, ICL lesions can lead to genome instability and cell death. It has been estimated that a single unrepaired ICL lesion in bacteria or yeast, and 20–40 ICLs in human cells could result in cell death (McVey, 2010; Muniandy et al., 2010). It is therefore critical for cells to be able to efficiently sense and repair ICL lesions.

Yet, repairing an ICL lesion is a formidable task for cells; ICLs are an intrinsically complex type of DNA damage as they span both strands of the double helix. And this complexity is reflected in ICL repair processes, which harnesses repair proteins from multiple distinct DNA repair pathways (Raschle et al., 2015). These include BER, NER, MMR, TLS (Roy and Schärer, 2016), and HR. Moreover, multiple mechanisms for ICL repair have been described, and the manner in which ICLs are repaired is dependent on cell-cycle, and can also be influenced by ICL structure (Roy et al., 2016; Semlow et al., 2016).

Most modalities of ICL repair involve several common processing steps including (1) ICL recognition (2) ICL “unhooking” where dual incision flanking the ICL are made by structure specific endonucleases (3) ICL processing in which the unhooked lesion is enzymatically trimmed by an exonuclease, and (4) lesion bypass and repair synthesis. However, depending on the type of ICL repair, the proteins responsible for each of these steps vary. Furthermore, recently an incision-independent ICL mechanism has been described, in which ICLs are processed in a far simpler reaction as will be discussed in detail below.
ICL repair mechanisms can be separated into two broad categories, based on the manner in which cells recognize the ICL lesion. Repair that is tightly coupled to replication is termed: “replication coupled repair” (RCR) that is strictly restricted to S-phase of the cell cycle, and repair that occurs independently of replication is called: “replication-independent repair” (RIR) and occurs predominantly in G1/G0/G2-phases of the cell cycle (Figure 1-4).

**Replication Coupled Repair**

In rapidly proliferating cells, the most efficient modality of ICL repair operates during S-phase (Figure 1-5). During replication, the CMG helicase (Cdc45-MCM-GINS) complex unwinds the DNA helix ahead of the active replisome. Progression of replisome is physically blocked when the CMG helicases encounter an ICL lesion, triggering repair of the ICL (Räschle et al., 2008). In Xenopus extracts, the collision of two replication forks are required, suggesting that an X-shaped DNA structure is required for ICL repair (Zhang et al., 2015). An alternative model for ICL repair in which a single collision event is required to initiate repair has also been reported in mammalian cells (Huang et al., 2013; Rohleder et al., 2016)

Following this ICL recognition step, the ICL can be processed in two mechanistically different ways. The first requires a complex repair reaction involving the Fanconi Anemia proteins (Ben-Yehoyada et al., 2009; Klein Douwel et al., 2014; Knipscheer et al., 2009; Long et al., 2011) and the NER proteins. A DSB repair intermediate is generated, which is repaired by the homologous recombination (HR) machinery. An
alternative mechanism of repair is catalyzed by the BER enzyme, NEIL3 glycosylase (Semlow et al., 2016). In this modality of ICL repair, DSBs are not generated, and ICL repair proceeds with the direct cleavage of the ICL lesion, followed by repair synthesis by TLS polymerases. The repair mechanisms specific to each pathway, and considerations pertaining to pathway choice is discussed below.

Fanconi Anemia Dependent ICL Repair Pathway

The mechanism of FA dependent replication coupled ICL repair pathway has been deciphered using Xenopus extracts and replicating plasmids containing a single defined ICL lesion. Together with genetic and biochemical studies, a detailed model for FA-dependent ICL repair has emerged (Figure 1-5, right). In this model, 21 FA proteins, together with proteins from the HR, NER, and TLS pathways cooperate to accomplish repair of ICL lesions (Walden and Deans, 2014).

The first step in ICL repair upon recognition of the lesion by the replication machinery, is the ubiquitin–mediated unloading of the CMG helicases by the BRCA1(FANCS)-BARD1 complex (Long et al., 2014). This allows the replisomes, which are initially stalled at the -20 positions on the leading strands to approach the ICL to the -1 position on either side of the offending adduct. Importantly, in Xenopus extracts, the convergence of two forks are required to activate ICL repair, suggesting that an X-shaped replication intermediate triggers ICL repair during S-phase (Zhang et al., 2015). Interestingly, Rad51(FANCR) localizes to sites of stalled forks at these early timepoint in RCR, well before DSB formation (Long et al., 2011). It is possible that Rad51 operates to prevent fork breakage/degradation during this time.
Upon stalling of the replication fork at ICL lesions, the FA pathway is activated. And the FA core complex, comprised of 14 proteins: FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, and FANCM as well as FAAP20, FAAP24, and FAAP100 and MHF1 and MHF2 associate with the damaged chromatin. This complex functions as an E3 ubiquitin ligase and catalyzes the ubiquitination of the FAND2-FANCI heterodimer, a hallmark event in FA pathway activation.

Ubiquitylated FANCD2-FANCI complex is recruited to the ICL lesion, and promotes the unhooking of the ICL lesion by recruiting the SLX4 (FANCP) scaffold, as well as the XPF(FANCQ)-ERCC1 endonuclease (Klein Douwel et al., 2017; Klein Douwel et al., 2014). It remains unclear if XPF-ERCC1 alone is responsible for the unhooking of the ICL lesion. There have been reports that the CtIP is recruited by the FANCD2-FANCI complex to further promoting FA and HR-dependent ICL repair (Murina et al., 2014; Unno et al., 2014).

A direct result of the unhooking of an ICL adduct during replication is the formation of a DSB in one of the sister chromosomes. This is repaired by a final group of FA proteins BRCA1 (FANCS), BRIP1 (FANCl), PALB2 (FANCN), BRCA2 (FANCD1), RAD51C (FANCO), and RAD51 (FANCR), which together with the rest of the HR machinery repairs the DSB using HR.

Unhooking also produces a sister chromatid with an unhooked ICL adduct that requires further steps for repair. First, the crosslink adduct which remain covalently attached to one end of the DNA is thought to be nucleolytically trimmed by the exonuclease SNM1A (Wang et al., 2011). Nucleolytic degradation of the unhooked ICL decreases its size, and helps facilitate downstream repair reactions including repair
synthesis. In replication coupled repair, the Pol ζ (Rev3-Rev7 heterodimer) and Rev1 contribute to bypass synthesis.

Of note, an alternative model for ICL repair has been proposed by the Seidman group, in mammalian cells using UV-psoralen ICL adducts and DNA combing (Huang et al., 2013; Rohleder et al., 2016). In this mechanism of ICL repair, a single fork that encounters an ICL lesion, is able to bypass the lesion. ICL repair only occurs after the replication machinery has traversed the offending adduct. Although it remains unclear how ICL traverse could occur, one theory is FANCM functions to push CMG helicases past ICL lesions during this process. These results strongly suggest that an X shaped DNA substrate is the critical signal to trigger ICL repair, as opposed to the double collision model that has been proposed by the Walter lab.

NEIL3 Dependent ICL Repair Pathway

A second mechanism of replication coupled ICL repair has been elucidated using *Xenopus* extracts and psoralen- and Ap-ICLs (Figure 1-5, left). In this modality of repair, collision of CMG helicase is followed by an ICL cleavage by the NEIL3 glycosylase (Semlow et al., 2016). Interestingly, NEIL3 is a bifunctional glycosylase, meaning it can hydrolyze glycosydic bonds, as well as catalyze a lyase reaction to create a single strand break in the DNA backbone. However, during repair of Ap-ICL lesions, NEIL3 only contributes a single incision to release the ICL, a reaction that leaves behind an abasic site (Ap site) on one strand, and an intact thymine nucleotide on the other. The abasic lesion is mutagenically bypassed by a TLS polymerase, completing the ICL repair process, and allowing replication to proceed. Of note, Ap-ICL lesions are repaired
predominantly using NEIL3-dependent mechanism (80%), but can also be repaired in the FA-dependent pathway.

In Xenopus extracts, the repair of Ap-ICL lesions, which occurs preferentially by NEIL3, is repaired significantly better (40%) compared to FA-dependent repair of cisplatin ICLs (15-20%). Together these studies reveal that while ICL sensing by the replication machinery allows for structurally distinct ICLs to be recognized to similar degrees, pathway choice and repair efficiency are heavily influenced by the chemical identity of the ICL. Further, in order for ICLs to be processed in the glycosylase-dependent manner, the ICL lesion must be amenable to cleavage by enzymes such as NEIL3.

Replication Independent Repair

ICLs repair that occurs outside of S-phase is termed replication independent repair (RIR). RIR is critical for homeostasis and survival in cells that are slowly- or non-dividing (for example, neurons), or in cells deficient for replication coupled repair proteins (i.e. Fanconi anemia cells) (Hlavin et al., 2010a; Muniandy et al., 2009b; Shen et al., 2006). RIR may also be a critical pathway in the context of tumorigenesis. Crosslinker-drugs efficiently target rapidly dividing cells by physically blocking DNA transactions including replication and/or transcription. Whereas, in slower-dividing cells which require less DNA transactions, ICL repair by RIR may be sufficient for cell survival, and may therefore facilitate drug resistance and cancer recurrence.
There are many lines of evidence in support of RIR. First, yeast in G1 phase undergo ICL repair (Sarkar et al., 2006). Similarly, G1 arrested mammalian cells process ICLs (Hlavin et al., 2010a; Muniandy et al., 2009b; Shen et al., 2006; Zheng et al., 2003). Moreover, ICLs are processed in non-replicating mammalian cell extracts (Smeaton et al., 2008, 2009), and ICLs are fully repaired in replication incompetent *Xenopus* extracts (Ben-Yehoyada et al., 2009; Kato et al., 2017; Williams et al., 2012).

RIR is mechanistically distinct from ICL repair that occurs during replication in several key ways. First, ICLs recognition must occur independent of replication-associated DNA unwinding, the principle mechanism by which ICLs are identified by cells during replication. Second, ICL repair during RIR does not involve generation of a DSB intermediate, which is a hallmark of FA-dependent repair during S-phase. Indeed, Rad52 deficient yeast have wild-type sensitivity to ICL drugs when grow to stationary phase, suggesting that repair in G0/G1 phase does not involve homologous recombination (McHugh et al., 2000). Instead, RIR is thought to proceed with two rounds of an excision-repair mechanism reminiscent of BER, NER, and MMR (Cole, 1973). While presented mechanistically (Figure 1-7), this model for RIR is largely based on genetic studies in bacteria and yeast. Aside from PCNA and Pol κ, which have been identified as two essential RIR factors (Williams et al., 2012), the identity and function of most other RIR proteins remain elusive.

RIR has been particularly challenging to elucidate for several reasons. First, ICL repair has been primarily studied in yeast and mammalian cells. Cells in culture are highly proliferative, and therefore biased towards replication-dependent repair and against RIR. Another complication in studying RIR is due to the variety of ICL substrates
investigators use to study RIR. Structurally distinct ICL lesions can be recognized and processed with different efficiency and mechanistically divergent ways by cells, a level of complexity that is absent during replication coupled repair, where all ICLs, regardless of structure, will block the replication machinery.

Nevertheless, there is evidence that NER (Thoma et al., 2005; Wood, 2010) and MMR (Kato et al., 2017; Zhao et al., 2009) proteins help process ICLs during RIR (Hlavin et al., 2010b; Noll et al., 2006; Williams et al., 2013). In fact, eCHIP based protein-recruitment studies even suggest that select FA proteins may play a role in RIR (Shen et al., 2009). However, how these proteins contribute to the productive repair of ICLs during RIR remains controversial. This idea will be discussed in detail below.

**Transcription Initiated ICL Repair Pathway**

Analogous to the stalled replication machinery triggering ICL repair, there is evidence that stalled transcription machinery can elicit repair of ICL lesions. ICLs placed in actively transcribed regions are processed more efficiently than ICLs in non-transcribed regions (Islas et al., 1991). Moreover, ICLs in constitutively transcribing plasmids (that do not replicate because they lack mammalian origins of replication) are repaired in mammalian cells, and processed preferentially on the transcribed strand (Zheng et al., 2003).

Not surprisingly, TC-NER proteins, specifically CSA and CSB, as well as GG-NER proteins are implicated in transcription initiated ICL repair. When placed downstream of a strongly active transcriptional promoter, the repair of NM- and BCNU-like ICLs are dependent on CSB, XPF, and XPG, as well as the Rev 1 and Pol ζ polymerases (Hlavin et
al., 2010a). The repair of MMC-ICLs were dependent on CSA (Zheng et al., 2003). The repair of cisplatin ICLs have similarly been studied and required XPA, XPF, and XPG (Enoiu et al., 2012a). As anticipated for a transcription driven recognition process, the these ICLs were by and large not dependent on XPC (Enoiu et al., 2012a; Hlavin et al., 2010a).

Most of the work pertaining to transcription initiated ICL repair are performed on plasmid systems where the ICL is placed in between a constitutively active promoter and a reporter gene. While this system has been instrumental in identifying proteins that contribute to the processing of the ICL lesion, the system is inherently biased towards transcription-coupled repair, and provides limited insight into the mechanism of the repair process and how this might occur under physiological circumstances.

Importantly, ICL repair efficiency during transcription-initiated repair can be influenced by ICL structure. Miller et al. found that synthetic analogs of NM-ICLs were repaired with >5 fold efficiency than two analogs of BCNU-ICLs (Hlavin et al., 2010a).

**Sensor Mediated ICL Repair Pathway**

ICLs can occur in DNA that is neither replicating or undergoing transcription (Ben-Yehoyada et al., 2009; Williams et al., 2012). In the absence of collision with polymerase complexes, ICLs must be recognized and funneled into a repair pathway by DNA damage sensors that surveil the genome for aberrant DNA structures. ICLs can result in destabilization of base stacking in its vicinity, and this distortion can assist in damage recognition. DNA damage sensors from the NER and MMR pathways are strong candidates to participate in this role.
Global-genome nucleotide excision repair sensor, XPC, has been implicated by several studies in ICL repair (Wood, 2010). XPC-Rad23B associates with TFO-conjugated psoralen ICL adducts in vitro (Thoma et al., 2005). Plasmids harboring a single TFO-psoralen ICL are also bound by XPC in vivo. XPC also localizes to laser induced psoralen ICL lesions in G1 arrested cells (Muniandy et al., 2009b), and the recruitment of XPC to psoralen-UV damaged nuclei was important for the disappearance of psoralen-UV ICLs. Furthermore, a plasmid-based reporter experiment also showed that XPC is partially required for the repair of MMC ICLs in cells (Zheng et al., 2003).

The role of XPC in the repair of cisplatin ICL lesions is harder to interpret. XPC deficient cells do not show marked sensitivity for crosslinker drugs (Enoiu et al., 2012a; Wood, 2010). This is often contrasted to the acute sensitivity of XPF-ERCC1 deficient cells to crosslinker drugs (Wang et al., 2011; Wood, 2010). The repair of plasmids (bulk) treated with psoralen or cisplatin have been studied in host reactivation assays, where XPC does contribute to DNA repair (Chen, 2003). However, these results are difficult to interpret since in vitro studies with a cisplatin oligonucleotide showed that XPC and other NER proteins preferentially process intrastrand crosslinks, and not ICLs (Zamble et al., 1996).

The difference in requirement for XPC in these studies could be explained by the ability of XPC to favor recognition of ICLs of specific structure and chemistry. Psoralen and MMC ICLs which are less distorting may be more amenable to recognition by XPC than cisplatin ICLs, which are severely distorting.

Alternative candidates to recognize ICLs in the absence of replication and transcription are the mismatch repair sensors, MutSα (MSH2-MSH6) and MutSβ
Like XPC, the MutS proteins are versatile DNA damage recognition factors. MuSα binds recognizes O6-methylguanine, O4-methylthymine, or the cisplatin-D (GpG) intrastrand crosslinks in vitro, as measured by electromobility shift assays (EMSA) (Duckett et al., 1996; Yamada et al., 1997), and psoralen (Vasquez, 2010) and cisplatin ICL. Similarly, MutSβ binds psoralen ICLs (EMSA) in vitro (Zhang et al., 2002).

Recently, an unbias proteomics approach was used to study the recruitment of DDR proteins to psoralen treated chromatin using *Xenopus* extracts (Raschle et al., 2015). In this study, the authors report that MSH2 and MSH6 are enriched on damaged chromatin, and that this occurred in a replication independent manner. This suggests that MutSα may play a role in ICL repair both during and outside of S-phase (Raschle et al., 2015). Indeed, MSH2 and MSH6 deficient mammalian cells show defects in FA pathway activation, as assessed by FANCI-FANCD2 ubiquitination in response to psoralen and MMC treatment (Huang et al., 2011; Williams et al., 2011). And in non-replicating *Xenopus* extracts MutSα is not only bind to multiple ICL types, but is also required for the efficient repair of trimethylene ICLs through recruitment of the MMR machinery (Kato et al., 2017). Finally, there is also some evidence in yeast that MutS proteins are important during S-phase repair of nitrogen mustard induced damage (bulk treatment) (Barber et al., 2005a).

There is also genetic evidence for the MutS proteins in ICL repair. However, the results are confounding. Loss of MutSα has been associated to MMC sensitivity (Fiumicino et al., 2000; Huang et al., 2011; Jung and Lippard, 2007; Peng et al., 2014; Williams et al., 2011), as well as oxaliplatin (Jung and Lippard, 2007), CDDP (Williams et al., 2011).
et al., 2011), and NMs (Fiumicino et al., 2000). In contrast, MutSα deficiency has also been extensively linked to cisplatin-resistance: reviewed in (Jung and Lippard, 2007).

Finally, several other studies also find that MMR deficient cells have no phenotype when treated with crosslinker drugs (Enoiu et al., 2012a; Hlavin et al., 2010a).

At least two major factors account for these discrepancies. First, clonogenic survival assays measure the combined impact of loss of a protein on the diverse types of DNA damage that are generate by crosslinker drugs (mostly monoadducts or intrastrand crosslinks), only a minority of which are ICLs. Clonogenic survival assay, especially in asynchronous cells, is not an optimal system to study sensitivity specific to ICLs. Investigators also use a variety of cell types, the majority of which are cancer cells. These cells sometimes have widely variable karyotypes and hence may respond to crosslinker drugs in unique ways.

Furthermore, cells may respond to structurally distinct ICLs in different ways (Kothandapani et al., 2011, 2013). This is especially true regarding cisplatin ICLs, for which there is extensive literature associating cellular resistance of MSH2 deficient cells to cisplatin (Kothandapani et al., 2011, 2013; Sawant et al., 2015; Yamada et al., 1997).

Taken together, inferring the contribution of DNA repair proteins from crosslinker sensitivity studies alone are not sufficient to deduce the roles of these proteins in ICL repair. Instead, careful biochemical studies using defined ICL substrates are needed to better understand ICL repair, especially RIR. Moreover, ICL repair mechanisms must be defined in the context of structure, the context of DNA in which the damage occurs, transcriptional status of the DNA, as well as the cell cycle.
There is abundant additional evidence that ICL structure and its chemical properties affect how ICLs are processed by cells. In mammalian cell extracts (HeLa and CHO cells) which are replication and transcriptionally silent, the efficiency of ICL unhooking and reprocessing on a 150 base pair linear DNA (blocked on either end with biotin moieties to prevent degradation) is dependent on ICL structure (Smeaton et al., 2008, 2009).

ICL structure may further influence repair kinetics because the repair intermediates may vary in structure. For example, upon recognition, ICLs are unhooked on one side of the DNA by enzymatic incision events. The length of duplex DNA surrounding an unhooked ICL lesion may therefore vary depending on (1) the distance from the ICL lesion at which the endonucleolytic incisions occur to unhook the ICL, and (2) the degree of exonucleolytic resection that occurs following unhooking of the ICL (Roy and Schärer, 2016; Roy et al., 2016). By shortening the duplex region, resection is thought to reduce the need for strand displacement synthesis during repair synthesis by translesion synthesis polymerases (Roy et al., 2016). The SNM1A (Wang et al., 2011) and FAN1 (Pizzolato et al., 2015; Wang et al., 2014) nucleases can process DNA duplexes containing ICLs, while the EXO1 exonuclease cannot (Kato et al., 2017). However, these nucleases are primarily thought to function in replication coupled ICL repair pathway.

Finally, the efficiency of repair synthesis may vary between different ICLs. Indeed, more distorting (5-atom linker) NM ICL lesions facilitate strand displacement and approach to the ICL lesion, while less distorting (8-atom linker) NM ICLs are extended more efficiently after insertion (Roy et al., 2016).
While the proteins responsible for ICL recognition during RIR has been challenging to determine, PCNA and translesion synthesis polymerase Pol κ have been identified to be two essential RIR factors in the repair of trimethylene-ICL lesions (Williams et al., 2012). Williams et al., showed using *Xenopus* cell-free extracts that ICL repair during RIR, monoubiquitinated PCNA recruits Pol κ to sites of ICL repair to accomplish repair synthesis. Because PCNA ubiquitination is a hallmark of polymerase switching, additional polymerases are thought to participate during various stages of repair synthesis: (1) approach (2) insertion of a nucleotide across the unhooked lesion, and (3) extension past the lesion. Up to three polymerases can be involved in a single repair reaction.
Table 1-1 DNA Lesions generated by endogenous and exogenous sources

<table>
<thead>
<tr>
<th>Endogenous sources of DNA Damage</th>
<th>DNA Lesions Generated</th>
<th>Estimated Number Lesions / Cell / Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spontaneous depurination or ribonucleotide incorporation</td>
<td>Ap site</td>
<td>10000</td>
</tr>
<tr>
<td>Cytosine deamination</td>
<td>Base transition</td>
<td>100-500</td>
</tr>
<tr>
<td>7meG</td>
<td>7meG</td>
<td>4000</td>
</tr>
<tr>
<td>O6meG</td>
<td>O6meG</td>
<td>41576</td>
</tr>
<tr>
<td>Oxidation</td>
<td>8oxoG</td>
<td>400-1500</td>
</tr>
<tr>
<td>Reactive aldehydes or rearrangement of AP site</td>
<td>Interstrand crosslinks</td>
<td>10&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exogenous sources of DNA Damage (dose exposure in mSV)</th>
<th>DNA Lesions Generated</th>
<th>Estimated Number Lesions / Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak hr sunlight (NA)</td>
<td>Cyclobutane-pyrimidine dimers (CPDs) and 6-4 pyrimidine pyrimidine photoproduct (6-4PPs)</td>
<td>100,000 / day</td>
</tr>
<tr>
<td>Cigarette smoke (NA)</td>
<td>Aromatic DNA adducts</td>
<td>45 - 1029</td>
</tr>
<tr>
<td>Chest X-rays (0.02)</td>
<td>DSBs</td>
<td>0.00</td>
</tr>
<tr>
<td>Dental X-rays (0.01)</td>
<td>DSBs</td>
<td>0.00</td>
</tr>
<tr>
<td>Mammography (0.40)</td>
<td>DSBs</td>
<td>0.02</td>
</tr>
<tr>
<td>Body CT (0.01)</td>
<td>DSBs</td>
<td>0.28</td>
</tr>
<tr>
<td>Head CT (2.00)</td>
<td>DSBs</td>
<td>0.08</td>
</tr>
<tr>
<td>Airline travel (0.005/hr)</td>
<td>DSBs</td>
<td>0.0002/hr</td>
</tr>
<tr>
<td>Space mission (60 days; 50.00)</td>
<td>DSBs</td>
<td>2.00</td>
</tr>
<tr>
<td>Hiroshima and Nagasaki atomic bombs (765670.00)</td>
<td>DSBs</td>
<td>0.2-160</td>
</tr>
</tbody>
</table>

Table is adapted from (Ciccia and Elledge, 2010)
<sup>a</sup> (Grillari et al., 2007)
<table>
<thead>
<tr>
<th>Crosslinker</th>
<th>B form DNA</th>
<th>Cisplatin</th>
<th>Mitomycin C</th>
<th>Psoralen</th>
<th>Abasic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Structure</td>
<td>N/A</td>
<td><img src="image" alt="Cisplatin Chemical Structure" /></td>
<td><img src="image" alt="Mitomycin C Chemical Structure" /></td>
<td><img src="image" alt="Psoralen Chemical Structure" /></td>
<td>N/A</td>
</tr>
<tr>
<td>Sequence Preference</td>
<td>N/A</td>
<td>5’–GC</td>
<td>5’–CG</td>
<td>5’–TA</td>
<td>5’–CAp</td>
</tr>
<tr>
<td>ICL structure</td>
<td><img src="image" alt="Cisplatin ICL Structure" /></td>
<td><img src="image" alt="Mitomycin C ICL Structure" /></td>
<td><img src="image" alt="Psoralen ICL Structure" /></td>
<td><img src="image" alt="Abasic ICL Structure" /></td>
<td>N/A</td>
</tr>
<tr>
<td>PDB</td>
<td>1BNA</td>
<td>1A2E</td>
<td>N/A</td>
<td>204D</td>
<td>NA</td>
</tr>
<tr>
<td>Source</td>
<td>(Drew et al., 1981)</td>
<td>(Coste et al., 1999)</td>
<td>(Guainazzi and Schärer, 2010)</td>
<td>(Spielmann et al., 1995)</td>
<td>(Johnson et al., 2013)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crosslinker</th>
<th>Trimethylene</th>
<th>SJG-136</th>
<th>NM analog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Structure</td>
<td><img src="image" alt="Trimethylene Chemical Structure" /></td>
<td><img src="image" alt="SJG-136 Chemical Structure" /></td>
<td><img src="image" alt="NM analog Chemical Structure" /></td>
</tr>
<tr>
<td>Sequence Preference</td>
<td>5’–GC</td>
<td>5’–CG</td>
<td>5’–purine-GATC-pyrimidine</td>
</tr>
<tr>
<td>ICL structure</td>
<td><img src="image" alt="Trimethylene ICL Structure" /></td>
<td><img src="image" alt="SJG-136 ICL Structure" /></td>
<td><img src="image" alt="NM analog ICL Structure" /></td>
</tr>
<tr>
<td>PDB</td>
<td>1LUH</td>
<td>2KNK</td>
<td>N/A</td>
</tr>
<tr>
<td>Source</td>
<td>(Dooley et al., 2003)</td>
<td>(Dooley et al., 2001)</td>
<td>(Hopton and Thompson, 2011)</td>
</tr>
</tbody>
</table>
Table 1-3 Crosslinker drugs used in the clinic

<table>
<thead>
<tr>
<th>Drug Class</th>
<th>Drug Name</th>
<th>Clinical Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platinums</td>
<td>Cisplatin</td>
<td>Bladder, ovarian, testicular cancer</td>
</tr>
<tr>
<td></td>
<td>Carboplatin</td>
<td>Ovarian cancer</td>
</tr>
<tr>
<td></td>
<td>Oxaliplatin</td>
<td>Colorectal cancer, colon cancer</td>
</tr>
<tr>
<td>Nitrogen Mustards</td>
<td>Cyclophosphamide</td>
<td>Leukemia, lymphoma, multiple myeloma, neuroblastoma, retinoblastoma, ovarian and breast cancer</td>
</tr>
<tr>
<td></td>
<td>Melphalan</td>
<td>Multiple myeloma, ovarian cancer</td>
</tr>
<tr>
<td></td>
<td>Chlorambucil</td>
<td>Chronic lymphocytic leukemia, lymphoma</td>
</tr>
<tr>
<td></td>
<td>Ifosfamide</td>
<td>Testicular cancer</td>
</tr>
<tr>
<td>Other</td>
<td>Mitomycin C</td>
<td>Pancreatic adenocarcinoma and gastric cancer</td>
</tr>
<tr>
<td></td>
<td>Psoralen</td>
<td>Cutaneous T cell lymphoma</td>
</tr>
<tr>
<td></td>
<td>Pyrrolobenzodiadiazepines</td>
<td>Phase II trial for solid tumors</td>
</tr>
</tbody>
</table>

Table is adapted from (Deans and West, 2011). Clinical application is based on National Cancer Institute.
Chapter 1 Figures

Figure 1-1. Sources and types of DNA Damage
DNA damage can occur through endogenous and exogenous sources. Many classes of DNA damage exist, including simple base lesions such as abasic sites, bulky adducts, or mispaired DNA bases. More complex lesions include DNA double strand breaks, intrastrand crosslinks, and ICL lesions.
Figure 1-2. DNA repair pathways for Base Excision Repair, Nucleotide Excision Repair, and Mismatch Repair

Base excision repair (BER) removes DNA bases that have been damaged. BER is initiated by lesion-specific DNA glycosylases. Either a single base (short patch BER) or a few bases (long patch BER) are removed, before the site is repaired by a translesion synthesis polymerase. Nucleotide excision repair (NER) fixes bulkier adducts. NER is initiated by XPC-HR23B, which senses distortions in the DNA helix. Incisions are made surrounding the damaged lesion, and the short oligonucleotide containing the damage site is removed. The resulting ~30 base gap is filled in by translesion synthesis polymerases. NER can also be initiated by the transcription machinery (not illustrated). Mismatch repair (MMR) is initiated by one of two mismatch repair sensor MutSα or MutSβ. The region containing the mispaired base is excised, and the resynthesis reaction is typically accomplished by replicative polymerases.
Double strand breaks can be repaired either through Non-homologous end joining (NHEJ) or by homologous recombination (HR). NHEJ occurs robustly throughout the cell cycle and involves the direct ligation of broken ends. In contrast, HR only occurs in late S or G2 phases of the cell cycle when a homologous sister chromatid is available and able to serve as a template for repair. During HR, DNA ends are resected extensively, and repair requires homology search as well as Holliday junction resolution.

Figure 1-3. DNA repair pathways of Double Strand Break Repair
ICLs can be recognized and repaired by multiple, mechanistically-distinct pathways. A defining feature of each pathway lies in the manner in which the ICL is recognized and repaired. In replication coupled repair, the ICL is recognized by the DNA replication machinery, which scans the entire genome during S-phase. ICLs can also be funneled into a repair pathway by the transcription machinery that may also collide with a damaged site, or by DNA damage sensor proteins that sense distortions in the DNA and activate repair. These include the XPC and MutSα MutSβ proteins from the NER and MMR pathways, respectively.

**Figure 1-4. Pathways of ICL repair**
ICLs can be recognized and repaired by multiple, mechanistically-distinct pathways. A defining feature of each pathway lies in the manner in which the ICL is recognized and repaired. In replication coupled repair, the ICL is recognized by the DNA replication machinery, which scans the entire genome during S-phase. ICLs can also be funneled into a repair pathway by the transcription machinery that may also collide with a damaged site, or by DNA damage sensor proteins that sense distortions in the DNA and activate repair. These include the XPC and MutSα MutSβ proteins from the NER and MMR pathways, respectively.
Replication-coupled ICL Repair Mechanisms

Figure 1-5. Pathways of Replication Coupled ICL repair

Replication coupled ICL repair is activated when replisomes collide with an ICL during replication. If the ICL lesions is amenable to cleavage by the NEIL3 glycosylase, they undergo direct cleavage. The resulting abasic site (red dot) is then repaired by the BER machinery. Some ICLs are not processed by NEIL3, and instead are funneled into a Fanconi Anemia dependent pathway. In this alternative mechanism, 21 FA proteins cooperate to accomplish ICL repair. In this process, incision events result in the
generation of a DSB intermediate, and therefore the homologous recombination machinery is also required.
Figure 1-6. Schematic of preparation of Xenopus cell-free extracts (High Speed Supernatant)

Unfertilized eggs from adult female frogs are used to prepare *Xenopus* cell-free extracts. Through a series of centrifugation steps, the eggs are crushed and a membrane and DNA-free extract is prepared. This extract, which is named “High Speed Supernatant” is replication incompetent and transcriptionally silent. HSS contains all cytosolic proteins required for DNA repair mechanisms including NER, MMR, and DSBR. HSS also support the repair of ICL lesions.
ICL lesions can be recognized by DNA damage sensor proteins that sense distortions in DNA. Two incision events occur flanking the damaged lesion. The unhooked ICL lesion, which remains tethered to the DNA on one strand, undergoes trimming. This makes room for the repair synthesis step which requires PCNA and at least one TLS polymerase, Polκ. There is growing evidence for MMR and NER proteins to function both in the sensing step, as well as the incision reactions during this process.
The formatting for the following paper has been modified for the purpose of this thesis.
Sensing and Processing of DNA Interstrand Crosslinks by the Mismatch Repair Pathway

Niyo Kato¹, Yoshitaka Kawasoe², Hannah Williams¹, Elena Coates¹, Upasana Roy³, Yuqian Shi⁴, Lorena S. Beese⁴, Orlando D. Schärer³,⁵, Hong Yan⁶, Max E. Gottesman⁷, Tatsuro S. Takahashi⁸, Jean Gautier¹,⁹*

¹ Institute of Cancer Genetics, Columbia University, New York, NY 10032, USA
² Graduate School of Science, Osaka University, Toyonaka, Japan
³ Department of Chemistry and Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794, USA
⁴ Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA
⁵ Institute for Basic Science Center for Genomic Integrity and School of Life Sciences, Ulsan National Institute of Science and Technology, Ulsan, South Korea
⁶ Fox Chase Cancer Center, Philadelphia, PA 19111, USA
⁷ Institute of Cancer Research, Columbia University, New York, NY, 10032, USA
⁸ Department of Biology, Kyushu University, Fukuoka, Japan
⁹ Lead Contact
*Correspondence: jg130@cumc.columbia.edu

Graphical Abstract
**Highlights**

- ICLs are sensed and repaired independently of replication and transcription
- MutSα (MSH2-MSH6) binds to ICLs and initiates repair
- ICL structure influences recognition and repair efficiency
- MMR-dependent ICL repair requires the nuclease activities of MutLα and EXO1

**eTOC**

Kato et al. identify a mechanism of ICL recognition that operates independently of DNA replication and transcription. In the absence of these processes, ICLs are recognized and repaired by the MMR machinery. MutSα is critical for ICL recognition, while MutLα and EXO1 contribute to key downstream nucleolytic steps during ICL repair.
Summary

DNA interstrand crosslinks (ICLs) that are repaired in non-dividing cells must be recognized independently of replication-associated DNA unwinding. Using cell-free extracts from *Xenopus* eggs that support neither replication nor transcription, we establish that ICLs are recognized and processed by the mismatch repair (MMR) machinery. We find that ICL repair requires MutSα (MSH2-MSH6) and the mismatch recognition FXE motif in MSH6, strongly suggesting that MutSα functions as an ICL sensor. MutSα recruits MutLα and EXO1 to ICL lesions, and the catalytic activity of both these nucleases is essential for ICL repair. As anticipated for a DNA unwinding-independent recognition process, we demonstrate that least distorting ICLs fail to be recognized and repaired by the MMR machinery. This establishes that ICL structure is a critical determinant of repair efficiency outside of DNA replication.

Introduction

DNA interstrand cross-links (ICLs) are lesions that covalently link opposing strands of the double helix. ICLs physically block cellular processes that require the unwinding of the DNA molecule, such as replication, recombination, and transcription. ICLs can also interfere with essential protein-DNA binding events such as transcription factor binding. Chemicals that induce ICLs, such as nitrogen mustards, platinum drugs, and mitomycin C are therefore extremely cytotoxic, and are routinely used in the clinic as anti-cancer chemotherapies (Deans and West, 2011). Importantly, ICLs also arise as a consequence of cellular metabolism, for instance through production of reactive
aldehydes such as malondialdehyde, acetaldehyde, and formaldehyde (Duxin and Walter, 2015) or by chemical rearrangements at abasic DNA sites (Price et al., 2014). ICLs vary greatly in structure and the degree to which they distort DNA. These structural differences can influence the manner in which cells process the ICLs. Indeed, in yeast, the repair pathway that responds to crosslinking damage differs according to the crosslinking drug (Beljanski et al., 2004). In mammalian cells, the efficiency of ICL unhooking and repair is greater for a distorting ICL than for a non-distorting ICL (Hlavin et al., 2010a; Smeaton et al., 2008). Furthermore, at least two replication coupled ICL repair mechanisms have been described in Xenopus extracts for which the primary determinant of repair pathway choice is based on ICL structure (Semlow et al., 2016).

In proliferating cells, most repair occurs during S-phase, when ICL sensing is a direct consequence of replication. Active replisomes stall at ICLs and trigger a complex reaction that requires the Fanconi Anemia (FA) proteins (Ben-Yehoyada et al., 2009; Klein Douwel et al., 2014; Knipscheer et al., 2009; Räschle et al., 2008) or the NEIL3 glycosylase (Semlow et al., 2016). Non-dividing yeast cells grown to stationary phase (Sarkar et al., 2006) and G1-arrested mammalian cells (Hlavin et al., 2010a; Muniandy et al., 2009b; Shen et al., 2006), harness an alternative modality of repair termed Replication Independent Repair (RIR) (Williams et al., 2012). To date, the molecular components that contribute to RIR have not been fully defined. A fundamental question that remains unanswered is how ICLs are recognized without replisome/ICL clashes. Collision between the transcription machinery and ICLs could activate repair. Indeed ICLs are more efficiently processed when placed in transcribed regions (Islas et al., 1991), and repair of ICLs placed in constitutively transcribing plasmids show dependence
for transcription-coupled and global-genome nucleotide excision repair proteins (Enoiu et al., 2012a). However, RIR is initiated in *Xenopus* extracts in the absence of both replication and transcription (Ben-Yehoyada et al., 2009; Williams et al., 2012), suggesting ICLs can be directly funneled into a repair pathway by DNA damage sensor proteins. But no ICL sensor has yet been unambiguously identified, and it is not clear how structurally distinct ICL lesions are recognized and/or subsequently repaired during RIR.

The mismatch repair (MMR) pathway is a highly conserved DNA repair mechanism that primarily functions to correct replication errors that escape proofreading. MMR is initiated by the MutSα complex (MSH2-MSH6 heterodimer) which recognizes and binds single base mismatches and 1-2 base insertion/deletion loops. Binding of MutSα to mispaired bases leads to the recruitment of downstream DNA repair proteins including the MutLα endonuclease (MLH1-PMS2 heterodimer), RPA, EXO1, and PCNA. Together, these proteins catalyze an excision-repair reaction to restore proper Watson-Crick base-pairing, and can effectively increase the fidelity of replication by up to three orders of magnitude. In humans, mutations in either subunit of MutSα (MSH2 and MSH6) account for 50-60% of germline mutation-based (Lynch Syndrome) and 15-20% of sporadic colorectal cancers (Reyes et al., 2015).

Importantly, MutSα has multiple roles outside of replication. MutSα is active in quiescent (post-mitotic) mammalian cells, such as terminally differentiated neurons, where the complex participates in genome maintenance mechanisms by recognizing not only canonical DNA mismatches, but also responding to chemically modified DNA bases (Iyama and Wilson, 2013; Schroering et al., 2007). For example, plasmids containing
single O\textsuperscript{6}-methyl-guanine adducts (O6meG) are processed in a MutS\textalpha-dependent fashion in non-replicating mammalian (York and Modrich, 2006) and *Xenopus* extracts (Olivera Harris et al., 2015), although leading sometimes to futile repair cycles (Fu et al., 2012). In addition to O6meG, MutS\textalpha specifically recognizes bulky lesions such as O4-methylthymine adducts and cisplatin intrastrand crosslinks *in vitro*. Crystal structures of human MutS\textalpha bound to these DNA substrates reveal a common mechanism of damage recognition that requires an N-terminal FXE motif found in the MSH6 subunit (Warren et al., 2007). Binding of MutS\textalpha to DNA lesions leads to a variety of cellular outcomes including repair, checkpoint activation, and apoptosis, underscoring the versatile and critical role that MutS\textalpha plays in genome maintenance.

In this study, we demonstrate a role for the MMR machinery in ICL repair. Using non-replicating and transcriptionally-silent *Xenopus* extracts, we show that MutS\textalpha binds to plasmids bearing a single site-specific ICL and harnesses the entire MMR machinery, including the MutL\textalpha and EXO1 nucleases, to perform ICL repair. We also compare the repair of structurally-different ICLs, including two analogs of the malondialdehyde ICL and a nitrogen mustard-like ICL (Guainazzi et al., 2010). We conclude that ICL structure influences lesion recognition and repair efficiency during RIR.

**Results**

**MutS\textalpha (MSH2-MSH6) senses DNA interstrand crosslinks**

The MutS\textalpha complex recognizes a diverse array of aberrant DNA structures including base-base mispairs and chemically modified DNA adducts. We hypothesized that the
ability of MutSα to recognize distortions within DNA might also extend to interstrand crosslinks (ICLs), and that sensing of ICLs by MutSα could initiate ICL repair by a set of reactions analogous to MMR. To test this hypothesis, we first asked whether MutSα binds preferentially to ICLs in non-replicating *Xenopus* extracts (High Speed Supernatant - HSS) using a plasmid pull-down assay. Notably, these cell-free extracts support ICL repair and restore the integrity of both DNA strands (Williams et al., 2012).

We used SJG-136, a rationally designed crosslinking drug, to generate plasmids with a single site-specific ICL lesion. SJG-136 preferentially forms ICLs between guanine residues at 5’-purine-GATC-pyrimidine-3’ sequences (Figure S2-1A) (Gregson et al., 2001). We treated oligonucleotide duplexes containing a single SJG-136 reaction site (Figure S2-1B), and purified crosslinked duplexes by denaturing PAGE. These oligos were ligated into a small plasmid (pBS: pBlueScript), and further purified by cesium chloride density ultracentrifugation.

Following incubation in HSS, we isolated plasmid and plasmid-bound proteins from the extract using biotinylated lac repressor protein and streptavidin beads, as previously described (Williams et al., 2012). Western blot analysis revealed that the crosslinked plasmid was enriched in bound MSH2 and MSH6 when compared to identical undamaged control plasmids (Figure 2-1A).

In order to study the dose-dependent response of MutSα to crosslinks, we conducted a similar experiment with plasmids containing multiple ICLs by treating double-stranded pBS plasmids *in vitro* with increasing doses of SJG-136 (Figure S2-1C-E). We observed dose-dependent increase of MSH2 and MSH6 binding to damaged plasmids (Figure 2-1B), after normalizing for plasmid recovery using qPCR (Figure S2-1F). We also observe
the recruitment of PCNA and RPA to these plasmids in a dose-dependent manner, suggesting that the cross-linked plasmids were undergoing processing and repair. Consistent with the accumulation of RPA and with our previous studies (Ben-Yehoyada et al., 2009; Williams et al., 2012), incubation of the SJG-136-treated plasmids in HSS triggered phosphorylation of cytosolic Chk1 (Figure S2-1G), a marker for ATR checkpoint activation.

Next, we asked whether MutSα is required for ICL repair. ICL plasmids harboring a single, site-specific trimethylene ICL lesion (5’GpC-ICL), for which an NMR structure has been described were used as substrates (Protein Data Bank: 2KNL) (Dooley et al., 2003). In contrast to the temperature-sensitive ICLs generated by SJG-136, the trimethylene crosslink is extremely stable at high temperature. We can therefore monitor ICL repair by quantitative PCR by comparing the increase in amplification of the damaged “X” region to that of an undamaged “C” region on the plasmid backbone over time (Ben-Yehoyada et al., 2009; Williams et al., 2012) (Figure S2-1B, right).

Incubation of cytosol with MSH2 antibodies raised against Xenopus MSH2 protein depleted both MSH2 and MSH6 subunits of MutSα (Figure 2-1C, left), but did not reduce the cytosolic levels of other MMR proteins including MLH1 and EXO1 (Figure 2-1C, right). Compared to mock-depleted extracts, depletion of MutSα significantly decreased ICL repair (Figure 2-1D). Importantly, addition of recombinant MutSαWT to MSH2-depleted extracts restored ICL repair to the level of mock-depleted extracts. Moreover, over-expression of MutSαWT in undepleted HSS increased the ICL repair efficiency (Figure 2-1E), suggesting that lesion recognition by MutSα is a rate-limiting
step during RIR. These data establish that MutSα is required for ICL repair in the absence of DNA replication and transcription.

**Mechanism of ICL recognition by MutSα**

MutSα discriminates aberrant DNA substrates amid a vast excess of normal DNA bases. The MSH6 subunit contributes critically to this process in at least two ways. First, the MSH6 subunit contains a highly conserved FXE motif (Figure S2-2A,B) that functions during damage recognition (Lamers et al., 2000; Malkov et al., 1997): the aromatic ring of the phenylalanine residue stacks with the mispaired base, whereas the carboxyl group of the glutamic acid residue hydrogen bonds with the mispaired base (Warren et al., 2007). A single amino acid substitution of this phenylalanine residue prevents efficient recognition of mispaired bases in yeast (Bowers et al., 1999), and results in defective repair in human cell extracts (Dufner et al., 2000).

To test the possible role of the FXE motif in MutSα-dependent repair of ICL in Xenopus extracts, we cloned and purified Xenopus mismatch binding-deficient MutSαFXE complex (Figure S2-2A,B,D). First, we asked if the F411A substitution affected mismatch repair using a plasmid-based MMR assay in HSS extracts. We used plasmids containing a single A:C mismatch and a 15 nucleotide gap on the 3’ side of the A-strand (pMML1AC) to trigger gap-directed strand-specific MMR (Figure 2-2A), as previously described (Kawasoe et al., 2016). Repair of this plasmid occurs preferentially on the A-strand, and the correction of the A:C site to a G:C pair generates a BamHI restriction site (Figure 2-2A, top). Background levels of repair on the C-strand is monitored by digestion with XhoI (Figure 2-2A, bottom). 60 minutes after incubation in mock-depleted extracts, 37% of the A:C mismatches were repaired to G:C. As anticipated, MMR was abolished.
in MSH2-depleted extracts. Addition of recombinant MutSα<sup>WT</sup>, but not MutSα<sup>FXE</sup> restored repair (Figure 2-2A, top). The requirement for the FXE motif for MMR is therefore conserved in *Xenopus*.

Next, we determined if the F411A substitution ablated ICL repair. As shown in Figure 2-2B, recombinant MutSα<sup>FXE</sup> was unable to support repair of ICLs in MSH2-depleted extracts. The FXE domain is, therefore, critical for ICL damage recognition.

In addition to recognizing a mismatched base, MSH6 also interacts with proliferating cell nuclear antigen (PCNA). Studies using fluorescently-tagged proteins show that the interaction between MSH6 and PCNA is important for the colocalization of the MutSα complex to replication factories in *S. cerevisiae* (Hombauer et al., 2011) and in human cells undergoing MMR (Kleczkowska et al., 2001). Yet, disruption of the PCNA-MSH6 interface results in a modest MMR phenotype (Shell et al., 2007), and other PCNA-independent recruitment of MutSα to mismatches via histone methylation has also been described (Li et al., 2013).

We used plasmid pull-down assays to ask whether MutSα was recruited to ICLs in a PCNA-independent manner. As shown in Figure 2-1B, the MutSα complex and PCNA were recruited to SJG-treated plasmids in HSS extracts. MutSα recruitment was not dependent on PCNA; MSH2 and MSH6 were efficiently recruited to crosslinked plasmids in PCNA-depleted extracts. Furthermore, the extent of MutSα recruitment was unaffected by supplementation of extracts with recombinant PCNA protein (Figure 2-2C). This experiment strongly suggests that MutSα can be recruited to ICLs independently of PCNA.
The interaction between MSH6 and PCNA is mediated by a PCNA interaction motif (PIP box) in MSH6 (Clark et al., 2000; Flores-Rozas et al., 2000). We constructed recombinant MutSα complex with a triple amino acid substitution in its conserved PIP box motif: Q(X)2LI(X)2FF (Figure S2-2A,C,D) that disrupts PCNA interaction with MutSα in vitro (Figure S2-2G). We tested the ability of this MutSαPIP to support ICL repair in MutSα-depleted extracts. We find that MutSαPIP is unable to support ICL repair (Figure 2-2B). This indicates that whereas MSH6-PCNA interaction is not required for MutSα recruitment to ICLs, and presumably for ICL sensing, the interaction between the two is essential for coordinating downstream repair events.

**ICLs are processed by MutLα and EXO1 nuclease**

The earliest catalytic events during ICL repair are thought to be dual endonucleolytic incisions flanking the ICL lesion in a process referred to as “unhooking”. However, neither the exact mechanism nor the nuclease(s) involved in the process are known for RIR.

We engineered the ICL-containing plasmids with 4 unique nicking sites flanking the ICL (Figure 2-3A and Figure S2-3A). We hypothesized that presenting the extract with a pre-nicked plasmid might stimulate ICL repair if the substrate resembled a physiologically relevant ICL repair intermediate. Indeed, introduction of a single nick 19 base pairs (bp) 5’ to the ICL lesion stimulated repair by 4-fold. In contrast, nicks 15 bp or 30 bp 3’ to the ICL or 34 bp 5’ to the ICL failed to stimulate repair (Figure 2-3B). We conclude that incision 5’ to the ICL is a rate-limiting step during the repair process. It further suggests that a 5’ incision close to the ICL (19 bp vs. 34 bp), is optimal for repair
and could reflect the physiological site of incision. Finally, the asymmetry between 3’ and 5’ incisions could reflect the action of 2 distinct nucleases.

Our evidence that MutSα is recruited to ICLs and is required for their repair prompted us to investigate the possible role of the mismatch-associated MutLα (MLH1-PMS2 heterodimer) endonuclease in ICL repair. Recruitment of MutLα to SJG-136 treated plasmids was observed following incubation of SJG-136-treated plasmid in extracts (Figure 2-3C). Notably, MLH1 recruitment was critically dependent upon MutSα, since MLH1 binding to the ICL-containing plasmid was abrogated in MSH2-depleted extracts (Figure 2-3C).

Specific antibodies against *Xenopus* MLH1 quantitatively depleted both MLH1 and PMS2 from extracts (Figure 2-3D, left). Importantly, MLH1 depletion did not affect the levels of MSH2 and MSH6 in the cytosol (Figure 2-3D, right), and MSH2 and MSH6 were still enriched on SJG-136 treated plasmids in MLH1-depleted extracts (Figure S2-3E, left). Removal of MLH1/PMS2 reduced ICL repair by 50%, and addition of recombinant MutLα WT to MLH1-depleted extract restored repair (Figure 2-3E). Therefore, MutLα is not only recruited by MutSα to ICL plasmids, it is also necessary for efficient ICL repair.

The PMS2 subunit of MutLα complex contains a conserved endonuclease motif DQHA(X)2E(X)4E. Substitution of E707 to K abolishes the endonucleolytic activity associated with PMS2 and yields a strong mutator phenotype in *S. cerevisiae* (Smith et al., 2013). Mutation of the corresponding residue (PMS2-E702K) in mice increases genomic mutation rates and cancer predisposition (van Oers et al., 2010). We generated the equivalent mutation in *Xenopus* MutLα (MLH1-PMS2 E674K) (Figure S2-3B-D).
Recombinant MutL\(\alpha\) lacking endonuclease activity did not support ICL repair in MLH1-depleted extract (Figure 2-3E). Furthermore, overexpression of nuclease deficient MutL\(\alpha\) complex significantly reduced ICL repair, possibly by acting as a dominant-negative (Figure 2-3F). Overexpression of recombinant MutL\(\alpha\) complexes did not interfere with MutS\(\alpha\) recruitment to SJG-136 treated plasmids (Figure S2-3E, right). These results indicate that PMS2-associated nuclease activity is required for ICL repair.

The dual incision flanking the ICL lesion during “unhooking” produces an oligonucleotide that remains covalently attached to DNA by the ICL adduct and is base-paired on each side of the ICL. During replication coupled ICL repair, this DNA fragment is degraded with 5’\(\rightarrow\) 3’ polarity by the SNM1A exonuclease, presumably allowing DNA polymerases to be loaded (Wang et al., 2011), and for other downstream repair reactions to take place.

We speculated that the 5’\(\rightarrow\) 3’ EXO1 exonuclease could fulfill a similar function in RIR. In an \textit{in vitro} reconstituted vertebrate MMR system, EXO1 degrades the single-strand oligonucleotide harboring the mismatch (Genschel et al., 2002). We found that EXO1 was recruited to SJG-136 treated plasmids. As is the case for MutL\(\alpha\), EXO1 recruitment was dependent on MutS\(\alpha\): EXO1 binding to the ICL-containing plasmid was reduced in MSH2-depleted extracts (Figure 2-4A). A specific antibody generated against \textit{Xenopus} EXO1 quantitatively depleted EXO1 from extracts (Figure 2-4B). ICL repair was significantly reduced in EXO1-depleted HSS. This defect could be rescued by addition of EXO1\textsuperscript{WT}, but not catalytically inactive EXO1\textsuperscript{D173A} (Figure 2-4C) (Liao et al., 2011). This demonstrates that EXO1 and its nuclease activity are required for ICL repair in the absence of DNA replication and transcription.
To evaluate the specific contribution of EXO1 during ICL repair, we conducted *in vitro* nuclease assays (Wang et al., 2011). First, we purified the catalytic domain of human EXO1 (EXO1\textsubscript{352}) from *E. coli* as previously described (Shi et al., 2017). EXO1\textsubscript{352} harbors robust nuclease activity *in vitro*, but lacks its C-terminal region (Figure S2-4A), important for making connections with various protein binding partners including MSH2, MLH1, PCNA, and RPA (Shi et al., 2017).

Duplex oligonucleotides containing a single SJG-136 ICL lesion are easily distinguished by size from uncrosslinked duplex on a denaturing polyacrylamide gel (Figure S2-4B). Similarly, the reaction products upon incubation with EXO1\textsubscript{352} can be visualized and identified. We generated control duplex DNA with only one free 5’ end available for processing (the 5’ end of the other strand blocked by addition of a biotin moiety). This template is efficiently processed by EXO1\textsubscript{352} from the biotin-free 5’ end in the 5’ to 3’ direction as anticipated (Figure 2-4D, left). Exonucleolytic processing was abolished when both ends of the duplex are blocked with biotin, confirming that under these experimental conditions, only the exonuclease activity contributes to the processing of the DNA substrates. When EXO1\textsubscript{352} was incubated with similar ds-DNA oligonucleotides containing a single SJG-136 ICL lesion, EXO1\textsubscript{352} was able to initiate processing of the substrate from an available 5’ end. However, EXO1 processing was blocked by the ICL lesion (Figure 2-4D, right). EXO1\textsubscript{352} is therefore unable to bypass an ICL under these conditions.

Incubation of plasmids harboring ICL lesions activates the RPA-ATR-pChk1 branch of the DNA damage response (Figure S2-1G). Next, we assessed the role of replication protein A (RPA) in ICL repair. Trimeric RPA protein complex was depleted using an
antibody generated against RPA1. Depletion of RPA from HSS extracts completely abolished ICL repair. ICL repair was restored by addition of recombinant *Xenopus* trimeric RPA complex (Figure S2-4C-E). This is consistent with our previous studies in which we reported a requirement for RPA in ICL repair in LSS extracts (Low Speed Supernatant) treated with geminin and roscovitine to inhibit replication (Ben-Yehoyada et al., 2009).

We predicted that RPA binding to crosslinked plasmids during ICL repair would occur primarily after ICL recognition and initiation of ICL unhooking events. We found that MSH2, MSH6, and MLH1 were loaded to similar levels to SJG-136 treated plasmids following incubation in mock- and RPA-depleted extracts (Figure S2-4F). This result supports a model in which MutSα and MutLα play early roles during ICL repair, preceding nucleolytic events during the repair.

**The efficiency of RIR is influenced by ICL structure**

In replicating extracts, ICLs are sensed by a translocating replisome that stalls at the ICL. In RIR, we suggest ICL sensing is dependent on distortion of the DNA helix at the lesion. To test this idea, we compared the kinetics of repair between chemically related trimethylene ICLs in two distinct structural conformations: the more distorting 5’-GpC-ICL (Dooley et al., 2003) and the less distorting 5’-CpG-ICL (Dooley et al., 2001) (Figure 2-5A and Figure S2-5A). Both ICLs were ligated into a pEGFP vector backbone (Figure S2-5B). We found that the more distorting ICL induced a stronger ATR checkpoint activation (Figure 2-5B). Furthermore, repair of the more distorting lesion was 40% more efficient than the less distorting lesion (Figure 2-5C). We next asked
whether repair of these ICLs was dependent upon MutSα. As shown in Figure 2-5D, repair of both trimethylene crosslinks displayed a similar requirement for MSH2. Thus, MutSα is able to recognize structurally distinct ICL lesions.

Finally, we monitored repair of a chemically distinct nitrogen mustard-like ICL lesion (Figure 2-5E and Figure S2-5A). This NM-ICL is generated within the major groove with minimal distortion of the DNA helix, allowing the DNA to maintain B-DNA conformation (Guainazzi et al., 2010; Mukherjee et al., 2014b). Interestingly, even after 3 hours of incubation there was no detectable repair of this NM-ICL (Figure 2-5E). This supports our notion that ICL recognition in RIR depends on DNA distortion, and is consistent with a previous report that repair of NM-ICLs are entirely dependent on replication (Räschle et al., 2008).

Discussion

MutSα is a bona fide ICL sensor

The mechanism(s) by which ICLs are sensed and repaired in non-dividing cells has been difficult to elucidate. In this study, we used Xenopus HSS extracts, which do not replicate or transcribe, to identify eukaryotic mismatch repair complex, MutSα (MSH2-MSH6), as a sensor of ICL lesions. Our conclusion is supported by the following observations. First, MutSα was recruited to plasmids containing ICL lesions (Figure 2-1A,B). Second, MutSα carrying its mismatch binding FXE motif was required for repair of trimethylene ICLs (Figure 2-2B). Third, binding of MutSα to ICL was required for recruitment of downstream repair proteins including MutLα (Figure 2-3C) and EXO1 (Figure 2-4A).
We conclude that binding of MutSα to ICLs precedes nucleolytic processing, and thus acts at an early step in ICL repair. Moreover, overexpression of MutSα in extracts enhanced ICL repair efficiency (Figure 2-1E), further suggesting MutSα senses ICLs, and that such sensing is a rate-limiting step in repair. We observe that MutSα-dependent ICL repair has slower kinetics than the repair of mismatches in Xenopus extracts (Kawasoe et al., 2016; Radman, 2016). We surmise that the complexity of the ICL lesion, which requires two rounds of repair synthesis during RIR, could account for this difference.

Importantly, we find that structurally distinct ICLs were repaired with varying efficiencies. Trimethylene-ICLs, which significantly distort DNA were repaired with robust efficiency, in contrast to non-distorting NM-like ICLs, which had no detectible repair (Figure 2-5C,E). This indicates that ICL repair by RIR is critically dependent on ability of the ICL lesion to be recognized by DNA damage sensor proteins. Thus, repair of ICLs in G0/G1 utilizes a fundamentally different mechanism than ICL repair during S-phase, when lesion sensing occurs by replisomes stalled at ICLs irrespective of crosslink structure. Moreover, ICL recognition by MutSα during RIR may also be influenced by sequence context, as has been proposed for MMR (Mazurek et al., 2009).

Of note, MutSα (MSH2-MSH6) is one of two mismatch repair sensors that operate during MMR in vertebrates. The other is MutSβ (MSH2-MSH3), which is about 10 times less abundant than MutSα in mammalian cells and has higher specificity for recognizing larger insertion-deletion loops (IDLs) and branched DNA substrates (Genschel et al., 1998). Depletion of MSH2 from extracts quantitatively depleted MSH6 and MSH3. Our depletion-rescue experiments with recombinant MutSα, which fully rescues ICL repair,
helps assign the specificity of ICL recognition to MutSa (Figure 2-1D). In mammalian cells, MutSβ along with the global genome nucleotide excision repair sensor, XPC (xeroderma pigmentosum complementation group C) have been linked to psoralen ICL repair (Muniandy et al., 2009b; Thoma et al., 2005; Zhang et al., 2002; Zhao et al., 2009). However, how these complexes contribute to the repair of ICLs remains unclear.

We were unable to detect direct binding of MutSa to SJG-136 treated oligonucleotides using EMSA experiments. This could be due to the design of the oligo duplex itself, which is restricted in length (~20 bases) to ensure SJG-136 generates a single ICL and to avoid the formation of secondary structures, since the duplex must be almost exclusively be composed of A and T nucleotides. While our data using the mismatch binding deficient (FXE) mutant MutSa (Figure 2-2B) strongly suggest that MutSa binding to DNA is critical for ICL repair, we cannot formally rule out that additional proteins participate in recruiting MutSa to ICLs.

The role of PCNA during Replication Independent Repair

Previously we reported that PCNA is an essential RIR factor. We showed that PCNA ubiquitinylated at lysine (K164) is required to recruit Polκ to ICL sites for repair synthesis (Williams et al., 2012). The experiments described here indicate that PCNA is not essential for MutSa recruitment to ICLs (Figure 2-2C). However, we find a requirement for interaction between MutSa and PCNA for ICL repair (Figure 2-2B). This suggests that PCNA may help retain MutSa at damage sites, or perhaps enhance MutSa’s damage recognition specificity, as has been suggested during MMR (Clark et al., 2000; Flores-Rozas et al., 2000; Kleczkowska et al., 2001). Our data show that PCNA
plays an upstream role during ICL repair, in addition to supporting Polκ recruitment for repair synthesis.

We also probed the requirement for binding between PCNA and components of the MMR machinery by generating mutations in the interdomain connector loop (IDCL) of PCNA. Mutations in this region abrogate interactions with MSH6 and PMS2 in vitro, and cause MMR defects in yeast (Lee and Alani, 2006). We made the equivalent mutations in *Xenopus* PCNA<sup>L126A,I128A</sup> (Figure S2-2E,F). However, we were not able to determine the contribution of the PCNA:MutSα interface to ICL repair using the this mutant, since binding between MutSα<sup>WT</sup> and PCNA<sup>L126A,I128A</sup> was enhanced, as measured by co-immunoprecipitation, compared to binding between MutSα<sup>WT</sup> and PCNA<sup>WT</sup> (Figure S2-2G). Accordingly, both PCNA<sup>WT</sup> and PCNA<sup>L126A,I128A</sup> were able to rescue ICL repair defects in PCNA-depleted extracts (Figure S2-2H,I).

**Insights into nucleolytic processing of ICLs by MutLα and EXO1**

Following sensing, an ICL is thought to be processed in two consecutive nucleolytic steps. First, dual incisions surrounding the ICL are made to produce an “unhooked” oligonucleotide, which is covalently tethered to the DNA by the ICL adduct. The unhooked lesion is then resected in a “trimming” reaction that facilitates synthesis past the adduct by translesion synthesis polymerases by eliminating the need for displacement synthesis (Roy et al., 2016). In replication coupled ICL repair the XPF-ERCC1 endonuclease and SLX4 promote unhooking of ICLs (Klein Douwel et al., 2014), and the SNM1A exonuclease has been proposed to process unhooked ICL lesions (Wang et al., 2011).
We investigated the role of these nucleolytic reactions in RIR. We used nicked plasmids to identify that a 5’ incision 19 base pairs away from an ICL lesion was able to stimulate repair by 4-fold. This is in contrast to nicks placed 15 bp or 30 bp 3’ to the ICL or 34 bp 5’ to the ICL, all of which failed to stimulate repair (Figure 2-3B). This experiment demonstrates a striking mechanistic difference between RIR and MMR. MMR is stimulated symmetrically by either a 5’ or a 3’ nick in vitro (Constantin, 2005; Varlet et al., 1996). Our results also suggest that 5’ incision is a rate-limiting step during RIR, and that 2 distinct nucleases are required during ICL repair. Depletion-rescue experiments with MutLα strongly suggest that MutLα promotes one or both of these incisions. However, because it is generally thought that MutLα requires a pre-existing nick and interaction with PCNA to stimulate its otherwise latent endonuclease activity, it is possible that another, as yet unidentified nuclease, is responsible for the initial incision. Our discovery that MutSα and MutLα participate in ICL repair highlights the versatile roles DNA repair proteins play in genome maintenance. This idea is further exemplified by our finding that EXO1 is required during RIR. EXO1 is a 5’→ 3’ exonuclease that is not only involved in MMR, but also plays critical roles in double-strand break repair and telomere maintenance (Tran et al., 2004). In all of these roles, EXO1 catalyzes digestion of DNA by hydrolyzing phosphodiester bonds between adjoining normal nucleotides. The requirement for EXO1’s nuclease activity during ICL repair (Figure 2-4C) suggests that EXO1 could process complex, chemically modified DNA substrates, including crosslinked bases. However, our in vitro experiments in which EXO1352 was unable to bypass an ICL lesion strongly suggests that EXO1 alone cannot efficiently process ICLs.
However, it is conceivable that ICL processing by EXO1 requires association with other proteins through its C-terminus.

During replication coupled ICL repair, SNM1A has been proposed to operate similarly to EXO1 in processing unhooked ICL lesions (Wang et al., 2011). In addition, there is evidence that the FAN1 nuclease can digest past ICL-containing oligonucleotides in vitro (Pizzolato et al., 2015; Wang et al., 2014). However, we find that FAN1 depletion in Xenopus HSS extracts has no effect on ICL repair (unpublished data).

**Conclusions**

DNA crosslinking agents are among the most widely prescribed anti-cancer drugs used in the clinic. Yet, acquired resistance remains is a significant limitation of these drugs (Deans and West, 2011). It seems likely that alterations to pathways that contribute to the repair of ICLs, the principle cytotoxic lesion generated by these drugs, may underlie acquired resistance. It is therefore critical to fully understand how cells respond to and repair ICLs.

We provide compelling evidence that the MMR pathway contributes to ICL repair. This is in agreement with a recent study that used an unbiased proteomic approach to study protein recruitment to psoralen-ICLs using quantitative mass spectrometry. This study revealed that MSH2, MSH6, PMS2, and EXO1 associate with cross-linked chromatin (Raschle et al., 2015). Importantly, recruitment of these proteins was resistant to geminin, suggesting that MMR proteins may play a role in ICL repair both in and beyond G0/G1 phase. Indeed, requirement for MSH2 and EXO1 has been reported in
repair of nitrogen mustard ICLs in S-phase yeast (Barber et al., 2005b), and in S-phase mammalian cells defective in the Fanconi Anemia pathway (Huang et al., 2011).

In this study, we provide mechanistic insights into how MMR proteins cooperate to accomplish ICL repair (Figure 2-6). We propose that ICLs that are recognized by MutSα are processed by MutLα and EXO1 in an incision-excision reaction, followed by DNA synthesis process that requires PCNA and Polk. Further, our work stresses the importance of distinguishing between structurally distinct ICLs, which may be repaired in overlapping and/or divergent ways. Future experiments should address whether the sequence context of ICLs influences ICL recognition and repair.
**Experimental Procedures**

*Xenopus cell-free extract and depletions*

*Xenopus laevis* frogs were handled in accordance with guidelines provided by the Institutional Animal Care and Use Committee at Columbia University, protocol AAAK0551. For details regarding the preparation of cell-free extracts and depleted extracts see Supplemental Experimental Procedures.

**Preparation ICL plasmids**

SJG-136 treated plasmids were prepared by incubating pBS with indicated concentrations of SJG-136 (NCI/Spriogen LTD) in 50 mM triethanolamine and 2 mM EDTA, overnight at 37°C. Plasmids were ethanol precipitated and resuspended in water. The quality and quantity of plasmids recovered were analyzed by agarose gel electrophoresis with ethidium bromide, Nanodrop measurement, and qPCR. The amount of plasmid used in plasmid pull-down experiments was normalized accordingly. The preparation of plasmids containing a single SJG-136, trimethylene, or NM-like ICL lesion is described in detail in Supplemental Experimental Procedures.

**ICL repair and lac repressor plasmid pull-down experiments in Xenopus extracts**

Lac repressor plasmid pull-down assays in un-depleted and depleted HSS extracts were performed as described previously (Williams et al., 2012). ICL repair assays in HSS were performed essentially as describe previously (Ben-Yehoyada et al., 2009; Williams et al., 2012). Each of these assays, as well as the MMR assay of a site-specific A:C mispair is described in detail in Supplemental Experimental Procedures.
Statistical Analysis

All statistical analyses were performed using GraphPad Prism 7 software using paired Student’s T-test or ANOVA. Data annotated as ns=p>0.05; (*)=p≤0.05; (**)=p≤0.01; (***)=p≤0.001

Protein Expression and Purification

Recombinant MutSα, MutLα, RPA, PCNA, and EXO1 were purified as described in Supplemental Experimental Procedures.

Reagents

All antibodies and reagents are listed in Supplemental Experimental Procedures.
Author Contributions

N.K. performed all experiments except Figure. 2-2A, which was prepared by Y.K., and Figure 2-3B and 2-4C, which were generated by H.W.; E.C. contributed to Figure 2-5E; U.R. prepared the NM-like ICL oligonucleotide; Y.S. purified EXO1\textsubscript{352}. N.K. and J.G. prepared the manuscript with input from all other authors: L.S.B., O.S., H.Y., M.E.G., and T.T.

Acknowledgments

We thank National Cancer Institute and Spirogen LTD for providing SJG-136, Dr. C. J. Rizzo for trimethylene-ICL crosslinked oligonucleotides, Dr. K. Marians for Bio-LacR protein, Dr. Karlene Cimprich for the RPA construct and J. Jiricny for MSH2-MSH6 and EXO1-antibodies. We also thank T. Apáricio and G. Sidhu for RPA protein purification. This work was supported by the National Cancer Institute (1R35 CA197606 and 1P01 CA174653 to J.G., R01 CA165911 to O.D.S., and 5T32 CA009503-27 to N.K.); National Institute of Health (5T32 GM008798-15 to N.K. and P01 CA092584 to L.S.B.); and the Korean Institute for Basic Science (IBS-R022-A1-2017 to O.D.S). Authors declare no financial conflict of interest.
Figure 2-1. ICL sensing by MutSα (MSH2-MSH6)

(A) Plasmid pull-down assay of control plasmids or plasmids containing a single site-specific SJG-136 ICL lesion. Plasmids were incubated for 40 min in HSS then purified from extracts using recombinant Bio-LacR protein coupled to streptavidin beads. Plasmid-bound proteins were analyzed by Western blot (WB) using the indicated antibodies.

(B) Plasmid pull-down assay and WB of SJG-136 treated plasmids. In the last lane, labeled (10x2), twice the amount of 10 µM treated plasmids was incubated in HSS. Input are plasmid DNA run on an agarose gel.

(C) MSH2-MSH6 immunodepletion. Mock- and MSH2-depleted HSS were analyzed by WB.

(D) Quantification of ICL repair in Mock- and MSH2-depleted HSS at 3 hrs. Results represent mean ± SEM from n=7 independent experiments.

(E) Quantification of ICL repair in HSS and HSS supplemented with MutSαWT at 3 hrs. Results represent mean ± SEM from n=11 independent experiments.

See also Figure S2-1
Figure 2-2. Mechanism of ICL recognition by MutSα (MSH2-MSH6)
(A) Plasmids containing a single site-specific A:C mismatch were incubated in Mock-, MSH2-depleted, or MSH2-depleted HSS supplemented with recombinant MutSα<sup>WT</sup> or MutSα<sup>FXE</sup> for the indicated amount of time. Mismatch repair efficiency was measured by digestion with XmnI with BamHI or XhoI restriction enzymes, for A→G repair (top) and C→T repair (bottom), respectively. Data is representative of three independent experiments.

(B) Quantification of ICL repair in Mock-, MSH2-depleted, or MSH2-depleted HSS supplemented with recombinant MutSα<sup>WT</sup> [n=7], MutSα<sup>FXE</sup> [n=6], or MutSα<sup>PIP</sup> [n=4] at 3 hrs. Results represent mean ± SEM of independent experiments.

(C) Plasmid pull-down assay and WB of SJG-136 treated plasmids in Mock- or PCNA-depleted extract.

See also Figure S2-2
Figure 2-3. Nucleolytic incision of ICL lesions by MutLα (MLH1-PMS2)

(A) Schematic of nicking sites engineered into ICL pBS plasmids (See also Figure S3A).

(B) Quantification of ICL repair in HSS of un-nicked, 5’ or 3’ nicked plasmids at 90 min. Results represent mean ± SEM from n=6 independent experiments.

(C) Plasmid pull-down assay and WB of SJG-136 treated plasmids in Mock- or MSH2-depleted extract.

(D) MLH1-PMS2 immunodepletion. Mock- and MLH1-depleted HSS was analyzed by WB.

(E) Quantification of ICL repair in Mock-, MLH1-depleted, or MLH1-depleted HSS supplemented with recombinant MutLαWT [n=12] or MutLαn.d [n=7] at 90 min. Results represent mean ± SEM independent experiments.

(F) Quantification of ICL repair in HSS with overexpression of buffer, MutLαWT, or MutLαn.d at 90 min. Results represent mean ± SEM from n=5 independent experiments.

See also Figure S2-3
Figure 2-4. Nucleolytic processing of ICL lesions by EXO1

(A) Plasmid pull-down assay and WB of SJG-136 treated plasmids in Mock- or MSH2-depleted extract.

(B) EXO1 immunodepletion. Mock- and EXO1-depleted HSS was analyzed by WB.

(C) Quantification of ICL repair in Mock-, EXO1-depleted, or EXO1-depleted HSS supplemented with recombinant EXO1WT or EXO1D173A at 3 hrs. Results represent mean ± SEM from n=7 independent experiments.

(D) Exo1352 nuclease assay with control or SJG-treated oligonucleotides. Reaction products were run on a denaturing polyacrylamide gel and stained with SYBR Gold.

See also Figure S2-4
Figure 2-5. ICL structure and repair efficiency

(A) Structures of normal B-form DNA (PDB: 1BNA), and DNA duplexes containing trimethylene 5’ GpC-ICL (PDB: 1LUH) and 5’ CpG-ICL (PDB: 2KNK) lesions.

(B) Plasmids containing single trimethylene GpC or CpG-ICL were incubated in HSS, and soluble extracts were analyzed by WB.

(C) Quantification of ICL repair of trimethylene GpC and CpG plasmids at the indicated time points. Results represent mean ± SEM from n=4 independent experiments.

(D) Quantification of ICL repair of trimethylene GpC and CpG plasmids in Mock- and MSH2-depleted HSS at 3 hrs. Results represent mean ± SEM from n=4 independent experiments.

(E) Quantification of ICL repair of trimethylene GpC and NM-like ICL plasmid using TaqMan qPCR reagent at the indicated time points. Results represent mean ± SEM from n=3 independent experiments.

See also Figure S2-5
ICLs can be recognized and repaired in the absence of replication. MutSα complex senses and binds ICL lesions, and recruits downstream repair proteins including the MutLα endonuclease, PCNA, and EXO1 exonuclease. The ICL lesion is repaired through a multi-step excision resynthesis reaction. The latter requires Polκ and monoubiquitinated PCNA.
Figure S2-1. Schematic of ICL plasmids and preparation of SJG-136 treated plasmids

(A) SJG-136 preferentially reacts with DNA at guanine residues at 5’ purine–GATC–pyrimidine sequences to form ICLs.

(B) Sequences of trimethylene and SJG-136 ICL oligos that were ligated into pBS to construct plasmids with a single site-specific ICL (left). Schematic of the resulting ICL plasmid. The binding site for Bio-LacR used for plasmid pull-down experiments, and primers used for qPCR to calculate ICL repair are illustrated (right).
(C) Schematic representation of pBS treated with SJG-136. The 7 sequences corresponding to SJG-136 ICL forming sites are illustrated. The single SJG-136 reaction site that overlaps with a BamHI recognition sequence is specified. The binding site for Bio-LacR and primers used for quantification of plasmids in plasmid pull-down experiments are indicated.

(D) Accumulation of ICLs on SJG-136 treated plasmids assessed using BamHI digestion. Plasmids were run on ethidium-bromide agarose gel before (upper panel) or after BamHI digestion (lower panel). With increasing doses of the drug, plasmids become refractory to BamHI digestion.

(E) The number of ICLs induced on plasmids treated with SJG-136 was estimated from quantification of BamHI digestion products. Results represent quantification from 3 independent experiments ± SD.

(F) Representative figure of the quantification of plasmids recovered from extracts after pull-down experiments using qPCR. Results represent technical triplicate ± SD from a single representative experiment. This analysis was used as a loading control to ensure equivalent amounts of plasmids were loaded on gels for Western blot.

(G) SJG-136 treated plasmids were incubated in HSS, and soluble extracts were analyzed by Western blot.
Figure S2-2. MutSα and PCNA mutants cloned and purified

(A) Schematic of X.l. MSH6. The domain structure of this protein and the locations of mutations used in the study are illustrated.

(B) The sequence of the conserved FXE domain in MSH6 is illustrated for several species. The amino acid substitutions used in this study are indicated in red.

(C) The sequence of the conserved PIP domain in MSH6 is illustrated for several species. The triple amino acid substitution used in this study is indicated in red.

(D) Purified MutSαWT, MutSαFXE, or MutSαPIP complex stained with Coomassie blue.

(E) The sequence of the conserved interdomain connector loop of PCNA is illustrated for several species. The double amino acid substitution used in this study is indicated in red.

(F) Purified PCNAWT and PCNAL126A,I128A stained with Coomassie blue.

(G) Co-immunoprecipitation of purified recombinant WT and mutant MSH2-MSH6 and PCNA using preimmune and purified IgG- or MSH2-specific antibodies. Immunoprecipitates were analyzed by Western blot with the indicated antibodies (left panel). The combination of MSH2-MSH6 and PCNA mutants analyzed are indicated in the table (right panel).

(H) PCNA immunodepletion. Mock- and PCNA-depleted HSS was analyzed by Western blot with the indicated antibodies to demonstrate depletion of PCNA.

(I) Quantification of ICL repair in Mock-, PCNA-depleted, or PCNA-depleted HSS supplemented with recombinant PCNAWT, or PCNAL126A,I128A at 3 hrs. Results represent mean ± SEM of n=4 independent experiments.
Figure S2-3. Nicking sites on the ICL plasmid and purification of MutLα
(A) Sequence surrounding the trimethylene GpC-ICL lesion, and the specific locations of near (top) and far (bottom) nicking sites used in Figure 2-3A,B.

(B) Schematic of *X.l.* PMS2. The domain structure of this protein and the locations of mutations used in the study are illustrated.

(C) The sequence of the conserved catalytic nuclease domain in PMS2 is illustrated for several species. The amino acid substitution used in this study is indicated in red.

(D) Purified MutLα<sup>WT</sup> and MutLα<sup>n.d.</sup> complex stained with Coomassie blue.

(E) Plasmid pull-down assay of SJG-136 treated plasmids in MLH1-depleted extract (left), or HSS supplemented with MutLα<sup>WT</sup> and MutLα<sup>n.d.</sup> (right). Plasmid-bound proteins were analyzed by Western blot.
Figure S2-4. RPA is required for RIR

(A) Schematic of X.l. EXO1. The domain structure of this protein and the locations of mutations used in the study are illustrated.

(B) Control or SJG-136 ICL oligonucleotide duplexes run on a denaturing polyacrylamide gel and stained with SYBR Gold.

(C) RPA immunodepletion. Mock- and RPA-depleted HSS was analyzed by Western blot to demonstrate depletion of RPA.

(D) Quantification of ICL repair in Mock-, RPA-depleted, or RPA-depleted HSS supplemented with recombinant RPA\textsubscript{WT} at 3 hrs. Results represent mean ± SEM of n=4 independent experiments.

(E) Purified RPA\textsubscript{WT} stained with Coomassie blue.

(F) Plasmid pull-down assay of SJG-136 treated plasmids in Mock- or RPA-depleted HSS. Plasmid bound proteins were analyzed by Western blot using the indicated antibodies.
Figure S2-5. **Sequences of site specific trimethylene and NM-like ICL plasmids**

(A) Sequence of oligonucleotide duplexes with a single trimethylene 5’ GpC-, 5’ CpG-, or NM-like ICLs.

(B) Schematic of plasmids containing single site-specific trimethylene 5’ GpC- or 5’ CpG-ICL. Plasmids were generated by ligating trimethylene oligonucleotides in (A) into pEGFP. ICL repair was measured using qPCR with X and C primers as indicated.
**Supplemental experimental procedures**

**Xenopus cell-free extract and depletions**

HSS cell-free extracts were prepared exactly as described previously (Shechter et al., 2004). Immunodepletions were performed using rabbit anti-MSH2 (Kawasoe et al., 2016), rabbit anti-MLH1 (Kawasoe et al., 2016) rabbit anti-RPA (Williams et al., 2012), and rabbit anti-EXO1 (Liao et al., 2011) antibodies. In each case, antibodies were coupled to proteinA sepharose CL-4B beads (Amersham Biosciences) in a 1:3 ratio of bed-volume of beads to serum, overnight. Two to three rounds of depletion were performed, each for 20 min at 4°C. This was followed by a 10 min clearing round with un-coupled proteinA sepharose CL-4B beads. Mock-depleted extracts were prepared identically with pre-immune serum. PCNA depletions were performed as previously described (Williams et al., 2012), using a p21 peptide (MTDFYHSKRRLIFS) immobilized onto a SulfoLink column (Pierce Biotechnology). In all cases, the quality of depleted extract was tested by running 0.5 µL of extracts on SDS-PAGE followed by Western blot.

**ICL Repair assays using quantitative PCR**

ICL repair assays in HSS extracts were performed as described previously (Williams et al., 2012). To measure ICL repair, the plasmids recovered from HSS were digested with restriction enzymes (PvuI and PvuII for ICLs in pBS backbone and Ndel and DraIII for ICLs in pEGFP) prior to qPCR analysis. qPCR was carried out using an Applied Biosystems 7500 fast thermocycler with Absolute Blue QPCR SYBR Green low ROX PCR mix (Abgene, Cat. #AB-4322B). For experiments involving NM-like ICL lesions
TaqMan Fast Advanced Master Mix (Thermo Scientific Cat. #4444964) was used (Figure 2-5E only).

The cycling conditions used were as follows: ABsolute Blue QPCR Mix: 50°C for 2 min, 95°C for 15 min followed by 40 cycles of 95°C for 15 sec, 62°C for 30 sec. TaqMan Fast Advanced Master Mix: 50°C for 2 min, 95°C for 20 sec followed by 40 cycles of 95°C for 3 sec, 62°C for 30 sec. Primers and probes were as follows:

pBS “C” region: 5’ – CTACGGCTACACTAGAAGGACAG and 5’ – CCACTGAGCGTCAGACCC
pBS “X” region: 5’ – CGAGATAGGGTTGAGTGTTGTTC and 5’ – ACTAGTTCTAGAGCGGCTGAGG
pEGFP “C”: 5’ – CTACGGCTACACTAGAAGGACAG and 5’ – CCACTGAGCGTCAGACCC
pEGFP “X”: 5’ – GGGCGTGGATAGCGGTTTGACTCACG and 5’ – ATCCCGGGCTGGAGGTAGA
TaqMan “C” Probe: 5’ – 6-FAM-TGCAAGCAGCAGATTACGCGAGAAAA-QSY7
TaqMan “X” Probe: 5’ – 6-FAM-CGCCCTGATAGACGGTTTTCGCCCTTT-QSY7

Lac Repressor plasmid pull-down assays

Lac repressor plasmid pull-down assays in un-depleted and depleted HSS extracts were performed as described previously (Williams et al., 2012). Briefly, plasmids were
incubated in HSS (15 ng/µL) for 40 min at 21°C. Plasmids were then purified from extracts using M-280 streptavidin dynabeads (Invitrogen Cat. #112.05D) pre-bound with purified Bio-LacR protein (gift from Dr. K. Marians, MSKCC). The beads were washed in a buffer composed of 10 mM Hepes (pH 7.7), 4 mM MgCl2, 50 mM KCl, 1 mM DTT, 250 mM sucrose, 50 mM NaCl, and 0.015% Triton X-100. Proteins were eluted from dynabeads by boiling for 5 min at 95°C in Laemmli buffer and analyzed by SDS-PAGE and Western blot. The quantity of plasmid recovered during pull-down experiments was assessed by qPCR and used to normalize loading.

**MMR assay**

MMR assays were conducted in HSS double depleted with MSH2 and MSH6 antibodies as follows: 1 volume of MSH2 serum and 2 volume of MSH6 serum were bound to 1 µl of protein A-Sepharose beads. A total of 0.2 volume of antibody-coupled beads was incubated in 1 volume of HSS at 4°C for 1 hr, and the procedure was repeated once. The MMR repair assay of a site specific A:C mispair on pMM1AC was performed exactly as previously described (Kawasoe et al., 2016).

**Construction of site-specific trimethylene and NM-like ICL plasmids:**

Plasmids containing single site-specific trimethylene ICL (used in Figure 2-1 to 2-4, and 2-5E) were generated exactly as previously described (Williams et al., 2012). Briefly, control or trimethylene crosslinked oligonucleotide duplexes (a gift from Dr. C. Rizzo, Vanderbilt University) were ligated into pBS using DraIII and Pflml restriction sites. Plasmids containing single site-specific SJG-136 ICL lesion (Figure 2-1) were prepared
similarly, except that SJG-136 containing ICL duplexes were prepared as described under “Preparation of site-specific SJG-136 ICL oligonucleotides”. Plasmids containing single site-specific 8 atom, non-distorting NM-like ICL lesion (Figure 2-5E) was prepared in an identical fashion using oligonucleotide duplexes containing a defined NM-like ICL lesion (Mukherjee et al., 2014b), but were ligated into pBS using BbSI restriction sites. Finally, plasmids containing single, site-specific trimethylene GpC- and CpG- ICL lesions used in Figure 2-5B-D were generated by ligating oligonucleotides containing these lesions (also from Dr. C. Rizzo, Vanderbilt University) into pEGFP vector using AccI and BplI restriction sites. In all cases, closed circular plasmids were purified away from linear and nicked ligation products by cesium chloride density ultracentrifugation.

**Preparation of site-specific SJG-136 ICL oligonucleotides**

The following oligonucleotides were annealed in equimolar ratio: 5’

ATAAAGATCTTTTATCCAATGGCCT and 5’

CCATTGGATAAAAGATCTTTATCTA, and phosphorylated using T7 PNK. Single, site-specific SJG-136 ICLs were generated by incubating 100 µg of oligonucleotide duplex with100 µM SJG-136 in 50 mM triethanolamine and 2 mM EDTA overnight at 37°C. Samples were ethanol precipitated and re-suspended in formamide loading buffer and run on a 20% acrylamide urea-TBE gel at 200 V for 2 hrs. Duplex containing ICLs were purified away from uncrosslinked oligos using UV shadowing and gel extraction using the crush-soak method as previously described (Enoiu et al., 2012b). Purified crosslinked oligonucleotide duplexes were used to prepare plasmids with site-specific ICL lesions.
Protein expression and purification

*Xenopus laevis* MSH2-MSH6 complex was a purified using a method modified from (Kawasoe et al., 2016). Briefly, Sf9 cells were co-infected with MSH2 and MSH6\textsuperscript{WT}-Flag, MSH6\textsuperscript{PIP}-Flag, or MSH6\textsuperscript{FXE}-Flag baculoviruses. Cells were harvested 72 hrs after infection and lysed through a syringe in Buffer S: 25 mM Tris-HCl pH 7.4, 250 mM NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA, 2 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine and 1x cOmplete, EDTA-free (Roche). The crude lysate was cleared by centrifugation for 30 min at 30,000 rpm. Cleared lysates were incubated with FLAG-M2 agarose beads for 3 hrs at 4°C. After washing, MSH2-MSH6 complexes were eluted from beads in 50 µg/mL FLAG-peptide in Buffer A (25 mM Tris-HCl pH 7.4, 5% glycerol, 5 mM 2-mercaptoethanol, 1 mM EDTA and 0.1x cOmplete, EDTA-free). Eluted proteins were dialyzed in fresh Buffer A overnight at 4°C.

*Xenopus laevis* MLH1-PMS2 was expressed in Sf9 cells by co-infecting cells with MLH1-Flag and PMS2\textsuperscript{WT} or PMS2\textsuperscript{n.d.} baculoviruses. Cells were harvested 48-72 hrs after infection, and lysed by swirling gently in Buffer S. The purification procedure was similar to that described for MSH2-MSH6 complex except that the protein complex was eluted with 50 µg/mL Buffer D: 25 mM Tris-HCl pH 7.4, 50 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, and 2 mM Benzamidine.

*Xenopus laevis* His-tagged EXO1\textsuperscript{WT} and EXO1\textsuperscript{D173A} protein was expressed and purified as previously described (Liao et al., 2011).
Xenopus laevis 6xHis-tagged PCNA was expressed and purified from BL21 cells exactly as previously described (Williams et al., 2012).

Xenopus laevis RPA complex was purified from the soluble fraction of BL21 lysate using standard Flag-purification protocol. Polycistronic RPA vector was originally obtained from Dr. K. Cimprich (Stanford University).

hExo1 catalytic domain (residues 1–352) was as purified from E. coli BL21(DE3)RIL as described previously (Shi et al., 2017). Cells were grown in 10 L Luria Broth (with 100 g/ml ampicillin) for 16 hr at 16 °C after induction with 0.1 mM IPTG; and harvested by centrifugation (4000 × g, 4°C, 30 min). The harvested cells were resuspended in lysis buffer (50 mM HEPES-NaOH, pH 7.5; 100 mM NaCl; 10 mM KCl; 5 mM MgCl₂; 5 mM DTT; 1 mM EDTA; 10% glycerol), and lysed with a cell cracker (Microfluidics). The lysate was cleared by centrifugation (18,500 × g, 4°C, 30 min) and loaded onto two tandem 5 ml HisTrap HP Ni Sepharose column (GE Healthcare Life Sciences) pre-equilibrated with wash buffer (lysis buffer with 300 mM NaCl and 20 mM imidazole). The protein was eluted with 300 mM imidazole and pooled. TEV protease was added to the protein solution and incubated at 4°C for 12 hrs to cleave the hexahistidine purification tag. The cleaved mixture was loaded onto pre-equilibrated Ni Sepharose column again to remove the hexahistidine purification tag. The eluate was loaded onto 5 ml HiTrap SP FF column (GE Healthcare Life Sciences) equilibrated in low ion buffer (50 mM HEPES, pH 7.5; 100 mM NaCl; 10 mM KCl; 1 mM MgCl₂; 5 mM DTT; 1 mM EDTA; 5% glycerol). Fractions containing pure protein were then eluted with high ion
buffer (low ion buffer plus 500 mM NaCl). The purified protein was exchanged into storage buffer (25 mM HEPES-NaOH, pH 7.5; 125 mM NaCl; 10 mM KCl; 1 mM TCEP), concentrated to ~20 mg/ml, flash-frozen in liquid nitrogen, and stored at -80°C.

**Mutagenesis Primers:** Site-directed mutagenesis was used to generate the following mutations (QuikChange XL Site-Directed Mutagenesis Kit (Agilent Technologies; Cat. #200517) following manufacturer’s instructions. The numbering of mutant residues corresponds to the amino acid residue of the *Xenopus* proteins.

**MSH6**<sup>F411A</sup> – mismatch binding FXE mutant

- 5’
  
  TATTTTTTACAAGTTGGCAAGTATTATGAACCTATCAGATGGATGCTGT

- C

- 5’
  
  GACAGCATCCATGTGATAAAAGTTAATAAGCCTTGCCAACCTTGATAAAA

**PMS2**<sup>E692K</sup> – nuclease dead

- 5’ GATCAGCATGCGACCAGTGGAGAAATACAAACCTTTGAGGTTTT

- 5’ AAAACCTCAAAGTTGGTATCTCTATCCTGCGATGCTGATC

**PCNA**<sup>L126A,I128A</sup> – IDCL (this mutant was generated from pET28aPCNAwt, a gift from Dr. V. Costanzo, IFOM)

- 5’ GTGGAGCGACGGCGGCGCTCCTGAACAAG

- 3’ CTTGTTCAGGAGCGCCGCTGCTCCAC
**EXO1<sub>352</sub> nuclease assay**

The following oligonucleotides (with or without a biotin moiety on 5’ ends), which contain a single SJG-136 ICL reaction site were annealed at equimolar concentrations:

- 5’ ATAATTTGATCATTATTAT
- 5’ ATAATAATGATCAAATTAT

Singly crosslinked oligonucleotides were purified by incubating duplex oligo with 100 µM SJG-136 overnight at 37°C. Crosslinked oligonucleotides were purified by gel extracted using the crush soak method. 25 ng control or crosslinked oligos were incubated with 0-100 nM EXO1-352 in nuclease reaction buffer (20 mM Tris HCl, 0.7 mM Hepes-KOH, 120 mM KCl, 5 mM MgCl2, 250 µg/mL BSA, 1.5 mM ATP, 1 mM glutathione, 60 µM DTT, 1% glycerol) for 30 min at 37°C in a 5 µL total volume. The reaction was stopped with the addition of formamide loading buffer, before running on a 12% acrylamide urea gel. Gels were stained with SYBR Gold.

**Antibodies used for Western blot**

The following antibodies were used for Western blot of *Xenopus* extracts: PCNA (Sigma-Aldrich Cat. #P8825), Histone H3 (Cell signaling Cat. #9715S), and Flag M2 (Millipore Sigma Cat. # F1804), MSH6 (*X.l.*) (Olivera Harris et al., 2015), MSH2 (*X.l.*) (Olivera Harris et al., 2015), EXO1 (*X.l.*) (Liao et al., 2011), RPA 70 (*X.l.*) (Williams et al., 2012), FANCI (*X.l.*) (Williams et al., 2012), MLH1 (*X.l.*) (Kawasoe et al., 2016), and PMS2 (*X.l.*) (Kawasoe et al., 2016).
Chapter 3. Conclusion
Cells have evolved multiple repair pathways to remove ICL lesions from their genomes. The mechanism by which ICLs become destined for repair is determined not only by cell cycle, but also dependent on the transcription status of the DNA in which it lies. Moreover, repair pathway choice can be further influenced by the structure and chemistry of ICLs, and the level of distortion it imparts on the DNA molecule.

Despite this complexity, the field has made outstanding progress over the last 10 years in elucidating the molecular details by which cells respond to and repair these lesions (Figure 3-1). In rapidly dividing cells, replication is the primary mechanism of ICL detection. ICLs that are amenable to direct cleavage by NEIL3, such as Ap-ICLs and psoralen ICLs are processed quickly and efficiently without the generation of DSB repair intermediate (Semlow et al., 2016). The preference of NEIL3 for psoralen and Ap-ICLs may reflect an evolutionary bias for ICLs that occur more commonly as a result of cellular metabolism. A slower, and more complex method of repair occurs for cisplatin- and MMC-ICLs, which are not amenable to processing by NEIL3. In this modality of repair, the entire FA pathway and the HR machinery must cooperate to fully repair ICL lesions in a multi-step repair reaction that also involves the repair of a DSBs intermediate (Räschle et al., 2008). While S-phase repair of ICLs is complex in itself, there is one key simplifying factor: during RCR, all ICLs regardless of structure, are recognized by the replication machinery.

In cells that are not dividing, this critical step is a more complex task. If the ICL lies in a transcriptionally active locus, it may be recognized through collision with the transcription machinery. However, ICLs can also occur in less-active regions of the
genome, and can interfere with processes such as transcription factor binding. These ICLs must be recognized through surveillance mechanisms by DNA damage sensor proteins, and funneled into a repair pathway. However, the precise proteins involved, and the mechanisms by which they contribute to repair process have remained elusive during this process.

Our lab previously established the *Xenopus laevis* extract system to study replication and transcription independent ICL repair (Ben-Yehoyada et al., 2009; Williams et al., 2012). This system has been instrumental in establishing RIR as a bone fide ICL repair pathway, with distinct molecular requirements for ICL repair compared to RCR. Indeed, RIR does not require the FA proteins, the HR machinery, or the TLS polymerase, Pol ζ (Williams et al., 2012), but instead requires TLS polymerase Pol κ and PCNA (Williams et al., 2012).

In this thesis, I used this system to examine the contribution of DNA repair proteins from multiple diverse pathways including MMR (Chapter 2 and Appendix B), NER (Appendix E), DSBR (Appendix D), and TLS (Appendix C). Our studies identify additional molecular components of RIR, and also shed significant insight into the enzymatic unhooking and processing of ICLs during RIR.

Most notably, in Chapter 2, we show that the MMR machinery is an important contributor of RIR. We show that in the absence of replication and transcription, the MutSα (MSH2-MSH6) complex binds and recognizes ICLs. MutSα recruits various downstream repair proteins including the MutLα endonuclease, and the EXO1 exonuclease to sites of damage. And the catalytic activity of both of these nucleases are required to process ICLs. Together our data unambiguously demonstrate that the MMR
proteins contribute to the productive repair of ICLs, an idea that has been controversial to date.

We have demonstrated that structurally distinct ICLs are repaired during RIR (Chapter 2 and Appendix A), albeit with varying efficiencies. These include the trimethylene-ICLs in two conformations (GpC and CpG), as well as the Ap-ICL. The non-distorting NM-like ICL had no detectable levels of repair.

These observations support my hypothesis that during RIR, helix distorting ICLs are more amenable to recognition by DNA damage sensor proteins, and are repaired with better efficiencies than ICLs which do not alter the structure of the DNA molecule. To further corroborate this idea, it would be interesting to also compare the repair kinetics of additional ICL types such as the cisplatin ICL, the 5- and 6-atom linker NM-like ICLs analogs, and the Ap-ICL in multiple different conformations.

In addition to ICL structure, ICL repair kinetics can be reflective of the actions of different DNA damage sensor proteins. While we extensively investigated the role of MutSα in Chapter 2, we have not yet explored the contribution of MutSβ, XPC, and DNA glycosylases (such as NEIL3) during RIR. It is clear that MutSα is a major sensor of trimethylene-ICLs, but it is likely that other structurally distinct ICLs may be more amenable to detection by these other DNA damage sensor proteins. In fact, my preliminary studies with Ap-ICLs show that these lesions are repaired independently of MSH2 (Appendix A), and are a good candidate for repair by the NEIL3 glycosylase.

Given that all the aforementioned DNA repair proteins (MutSα, MutSβ, XPC, and NEIL3) are known to participate in the repair of other DNA lesions, it would be important to study the crystal structures of the proteins bound to ICLs. These studies
could shed insight into whether mechanism of DNA damage recognition is conserved between repair pathways. Indeed, our data suggest the mechanism of DNA damage recognition by MutSα is conserved between MMR and RIR since both processes are dependent on the FXE motif in MSH6.

Our studies using nicked ICL plasmids (trimethylene and NM-like ICLs) show that a single 5’ incision ~20 based away from the ICL site is able to stimulate ICL repair. This observation has several implications. First, these experiments demonstrate that incision 5’ to the ICL is a rate-limiting step during RIR. Second, the asymmetry between 5’ and 3’ incisions on repair kinetics could reflect the action of 2 distinct nucleases. Finally, these experiments suggest that incision ~20 bases away from an ICL is optimal for its repair and could reflect the physiological site of incision.

The stimulatory asymmetry between 5’ and 3’ incisions on ICL repair rate is conserved between trimethylene (Chapter 2) and NM-like ICLs (Figure A1 B). These results are interesting as it points to a unique mechanistic phenomenon, specific to ICL repair. During MMR, a single nick either 5’ or 3’ is able to stimulate repair, and this stimulation is observed even when nicking occurs hundreds of bases away from the mispaired site. It would be interesting to test whether repair or other ICLs is also stimulated by a 5’ incision.

The 5’ nick-directed repair of NM-ICLs is very robust, and warrants further investigation. For example, it would be informative to conduct depletion-rescue experiments using the 5’ nicked NM-ICL repair to determine whether its repair requires components of the MMR pathway (particularly MutLα) and/or TLS Pol κ. In RCR, the cisplatin and NM-like ICLs requires Pol ζ and Rev1 for lesion bypass (Räschle et al.,
If the requirement for TLS is dependent foremost on lesion structure, the NM-ICL may also require these polymerases during RIR.

Finally, multiple NM ICLs have been synthesized in various different conformations (Guainazzi et al., 2010; Mukherjee et al., 2014a), from completely non-distorting (8-atom NM linker) to moderately distorting (5-atom NM linker) (Mukherjee et al., 2014a; Roy and Schärer, 2016). The studies described herein only pertain to the 8-atom NM linker. It would be interesting to study the other NM-ICL analogs to further study how ICL structure affects its repair during RIR.

While our experiments with the nicked plasmid suggest that incision ~20 bases away from the ICL is a physiologically relevant sight of incision, it would be informative to examine the precise location of this incision site during RIR in HSS.

For this, a single strand ligation mediated PCR (sslm-PCR) and high throughput sequencing could be useful (Figure 3-2). During processing of an ICL site, for instance after incision events or repair synthesis events, 3’ hydroxyl ends are generated. These free hydroxyl ends are available for ligation with small adapters. Using a forward primer complementary to the ICL plasmid backbone, and a reverse primer complementary to that of the ligated oligonucleotide, we will be able to use PCR and sequencing to characterize each structure at single base pair resolution. Given that single stranded byproducts are only obtained upon initiation of repair, results will exclusively represent repair species.

This sslm-PCR approach would not only be useful in mapping sites of endonucleolytic incision, but could also be useful to investigate the extent to which polymerases contribute to the repair reaction. RIR is thought to involve the sequential
action of multiple DNA polymerases including, Pol κ. However, it is not known if Pol κ participates in the synthesis up to or across the unhooked ICL adduct. I hypothesize that upon depletion of a Pol κ, specific ICL repair intermediates will accumulate, using sslm-PCR we can learn the extent to which Pol κ participates during TLS.

Finally, while the *Xenopus* extract system is a powerful system to study the mechanism of RIR, it would be exciting to extend our studies to mammalian cells. For this, it would be important to use terminally differentiated, non-dividing cells such as neurons. Various approaches including immunofluorescence to monitor recruitment of repair proteins to crosslinker treated cells, or monitoring ICL repair using ICL-specific antibodies would be especially exciting.

**Concluding Remarks**

Many studies have suggested that diverse DNA repair pathways contribute to the RIR of DNA interstrand crosslinks. However, the precise proteins involved, and the mechanisms by which they contribute to repair process have remained largely elusive. My studies using structurally defined ICL substrates in *Xenopus* cell free extracts lends insight into how the MMR helps identify and process ICL lesions in cells that are non-dividing.

Taken together, these studies contribute profoundly to our understanding of how cells respond to and repair ICL lesions throughout the cell cycle. Better understanding how these mechanisms cooperate or compete to remove ICLs will help develop better crosslinker based chemotherapies.
Figure 3-1. Mechanism of ICL repair

ICLs are repaired throughout the cell cycle. ICL repair during S-phase is triggered by the replication machinery, which collides with the damage site, and funnels the lesion into repair either by a glycosylase- or FA-dependent mechanism. Outside of S-phase ICL recognition is either accomplished by the transcription machinery, or by DNA damage sensor proteins. The molecular components and mechanisms of repair are significantly divergent between repair pathways.
Figure 3-2. Mapping sites of ICL processing using sslm-PCR
(A) Schematic of ICL plasmid and the site of stimulatory 5’ incision event
(B) The sslm-PCR protocol that we have devised to map sites of ICL repair in *Xenopus* extracts.
References


Our studies using *Xenopus* HSS extracts in Chapter 2 demonstrate that structure is a key determinant of ICL repair efficiency during RIR. Briefly, we compared trimethylene ICLs in two structural conformations (more distorting GpC-ICL vs less distorting CpG-ICL), and found that the more distorting lesion led to a more robust DNA damage checkpoint activation, and was repaired 40% more efficiently than its less distorting counterpart. Moreover, a non-distorting NM-like ICL remained unrepaired over the course of 3-4 hours, a timespan in which trimethlyene ICLs undergo robust repair. Together these results suggest that more distorting ICLs are more amenable to recognition and hence repair, in the absence of replication and transcription.

Here, we build on this idea in two ways. First, we show that the NM-like ICL lesion can be stimulated to repair at 30-40 fold (corresponds to increase in repair from 0 to ~10% repair) better efficiency with a single nick ~25 bases 5’ to the ICL site compared to un-nicked substrate (Figure A1 A,B). An incision at a similar distance 3’ (27 bases away) to the lesion has no effect on repair kinetics. Because incision events are thought to be the first catalytic event during RIR, these experiments suggest that a single nick can help bypass the requirement for NM-ICL recognition. It further suggests that sensing of NM-ICLs is a key rate-limiting step for its repair, further supporting the hypothesis we pose in Chapter 2, that MutSα favors DNA lesions that are more distorting.

We have also studied the repair of an Ap-ICL lesion (Figure A2 A). The repair of this Ap-ICL was recently studied in the context of replication by the Walter lab (Semlow et al., 2016). Unlike the cisplatin and NM-ICLs which require the FA machinery for repair,
the authors show that Ap-ICLs undergo direct cleavage by the NEIL3 DNA glycosylase. The authors conclude that Ap-ICLs only repair during replication, as inhibition of replication in HSS+NPE (replicating Xenopus extracts) by geminin inhibited ICL repair in their repair assays, (Southern blot and radionucleotide incorporation assays).

Nevertheless, it remains possible that Ap-ICLs are repaired in the absence of replication at low levels that are undetectable using these methods.

We use a more sensitive qPCR based assay to determine whether Ap-ICLs are repaired in non-replicating HSS extracts. We find that Ap-ICLs are rapidly repaired in HSS (Figure A2 B), and that the repair of these lesions occurs on both strands of the DNA molecule. This is determined by BglII restriction enzyme digestion of the ICL plasmid after incubation in HSS extract, where completion of ICL repair results in the production of a BglII restriction site.

Unlike trimethylene-ICLs in HSS, the repair of Ap-ICLs is not dependent on MMR, as depletion of MSH2 or MLH1, has no impact on the repair of Ap-ICLs (Figure A2 C,D). This suggests that Ap-ICLs are not amenable to detection by MutSα, and are further not processed by the MMR machinery. Ap-ICLs likely rely on the function of other DNA damage sensor proteins, perhaps XPC or NEIL3 (or other) which triggers a mechanistically distinct ICL repair process. Not surprisingly, this process still relies on PCNA for efficient repair (Figure A2 E), suggesting and TLS repair is a critical event during the repair of Ap-ICLs. Moreover, if Ap-ICLs are in fact processed by NEIL3 in RIR, PCNA depletion would only result in a 50% repair phenotype as measured by our qPCR based assay: upon cleavage of an Ap-ICL by NEIL3 the DNA will resolve into one intact DNA strand, with the opposing DNA strand containing a single abasic lesion.
Further studies are needed to determine which DNA damage sensor proteins are required for the recognition of Ap-ICLs in HSS extracts, and to determine what other molecular components are required for the repair of this type of lesion.
Figure A 1. Nicking experiments with NM-like ICL in HSS
A. Sequence surrounding the NM-like ICL lesion, and the specific locations of nicking sites used in (B).
B. Quantification of repair of NM-like ICL lesion in HSS at 40 min. Repair is only stimulated by 5’ incision ~20 bp from the ICL lesion.
Figure A 2. Repair of Ap-ICL in HSS
A. Sequence of Ap-ICL oligo.
B. Quantification of Ap-ICL repair in HSS extract at the indicated time points (Mock). Repaired plasmids, which were recovered from HSS extracts, were digested with BglII to monitor whether ICL repair occurred to completion.
C. Quantification of Ap-ICL repair in Mock- and MSH2-depleted HSS at 3 hrs. Results represent mean ± SEM from two independent experiments.
D. Quantification of Ap-ICL repair in Mock- and MLH1-depleted HSS at 3 hrs. Results represent mean ± SEM from three independent experiments.
E. Quantification of Ap-ICL repair in Mock- and PCNA-depleted HSS at 3 hrs. Results represent mean ± SEM from two independent experiments.
Appendix B. MMR Deficient Mammalian Cells are Sensitive to MMC

Niyo Kato, Jean Gautier (unpublished)

Our results in Xenopus extracts (Chapter 2) indicate that ICL repair in non-replicating extracts harness components of the MMR pathway. If this pathway is also operative in intact cells, the loss of MMR should increase cell susceptibility to crosslinking drugs.

To test this prediction, I generated MCF10A cells lines stably expressing constructs in which doxycycline-inducible shRNA mediates knockdown of either MSH6 or PMS2. MCF10A cells were chosen because they are non-transformed and near-diploid breast epithelial cells, and therefore offer a clean system to study drug sensitivity.

MSH6 and PMS2 were efficiently knocked-down 72 hours after addition of doxycycline to growth media, as shown by Western blot (Figure B1 A-C). These cells were treated with MMC, which induces a crosslink similar in structure to the trimethylene-GpC-ICL lesion we assayed in HSS extracts. Cells were exposed for drug for 3 hours, followed by growth for 12-15 days in the presence of doxycycline. Both MSH6- and PMS2- deficient cells were more sensitive to MMC compared to control cells with scramble shRNA (Figure B2 A,B). These results support our in vitro findings that MMR proteins cooperate to recognize and remove ICLs. These results are consistent with several previous studies that correlate the loss of MutSα or MutLα to MMC sensitivity (Fiumicino et al., 2000; Huang et al., 2011; Peng et al., 2014).

It is important to point out that MutSα deficiency has also been linked to cisplatin-resistance (reviewed in (Jung and Lippard, 2007). This phenotype is unique to cisplatin, and not observed when cells are treated with other platinum drugs including oxaliplatin (Jung and Lippard, 2007), or crosslinkers such as NMs and MMC (Fiumicino et al., 2000). The differences in cellular responses to these drugs, measured by survival in these experiments, therefore likely reflects the diverse ways in which cells respond to structurally distinct ICLs or the other DNA adducts that these chemicals also generate. Hence, inferring the contribution of DNA repair proteins from
crosslinker sensitivity studies alone are not sufficient to deduce the roles of these proteins in ICL repair.
Figure B 1. shRNA mediated knockdown of MSH6 and PMS2 in MCF10A cells

A. Western blot analysis of MCF10A whole-cell lysate to demonstrate knockdown of MSH6 72 hrs after addition of doxycycline to cell growth media.

B. Western blot analysis of MCF10A whole-cell lysate showing efficient MSH6 knockdown with three shRNA clones. Clones A and B were used for clonogenic survival assays.

C. Western blot analysis of MCF10A whole-cell lysate showing efficient PMS2 knockdown with three shRNA clones. Clones C and D were used in clonogenic survival assays.
Figure B 2. Sensitivity of MSH6- and PMS2-deficient MCF10A cells to mitomycin C
A. Clonogenic survival assay of MSH6-deficient human breast epithelial MCF10A cells
in the presence of indicated doses of MMC. Results represent the mean ± SEM of at
least three independent experiments.
B. Clonogenic survival assay of PMS2-deficient human breast epithelial MCF10A cells
in the presence of indicated doses of MMC. Results represent the mean ± SEM of at
least three independent experiments.
Appendix B. Methods

Cell Culture and production of MCF10A cells with stable integration of doxycycline inducible shRNA

MCF10A cells were cultured in DMEM-F12 (Invitrogen) supplemented with 5% horse serum (Invitrogen), 20 ng/mL EGF (Peprotech), 0.5 µg/ml hydrocortisone (Sigma-Aldrich Corporation), 100 ng/mL cholera toxin (Sigma), 10 µg/ml insulin (Sigma), and Pen Strep (Invitrogen). 293T cells were cultured in DMEM (Invitrogen) and Pen Strep (Invitrogen), supplemented with 10% FBS (Invitrogen). All cells were incubated at 37°C with 5% CO₂. The following doxycycline inducible shRNA clones (Dharmacon) targeting MSH6 (MSH6-A: V3THS_318784 and MSH6-B: RHS4696-101352772) and PMS2 clones (PMS2-C: V3THS_302063 and PMS2-D: V2THS_93546) were used. Each shRNA vector was purified using E.Z.N.A. Plasmid Miniprep kit (Omega Bio-Tek, Inc.) and packaged into lentiviruses by transfecting (jetPEI by Polyplus-transfection) into 293T cells with pMD.G and pCMVR8.91. After 24 hrs, the media was replaced with MCF10A media, which was used to collect viruses at the 48 hr timepoint. Viruses were filtered and were used to infect MCF10A cells by spin infection; 1 hr at 1000 rpm at RT in the presence of 8 µg/ml polybrene (Sigma). After 24 hrs of incubation, infected cells were selected by the addition of 2 µg/ml puromycin (Sigma). For Western blot of MCF10A whole-cell lysates, cells were lysed in RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% Na-DOC, 0.1% SDS, 50 mM Tris pH 8.0), supplemented with protease and phosphatase inhibitors (Sigma), for 10 min on ice.

Clonogenic Survival Assays
Cells were induced to express shRNA against MSH6 or PMS2 with addition of 1 µg/mL doxycycline in growth media. After 72 hrs, cells were plated at a density of 300 cells per 100 mm plate and allowed to adhere for 1 hr. Cells were treated with MMC for 3 hrs, and then grown in fresh media for 14 days. Throughout the experiment, cells were grown in 1 µg/mL doxycycline, which was added to the media every three days. Colonies were fixed with methanol and stained with crystal violet. Cell survival was calculated as a percentage of colonies on drug treated plates compared to untreated plates.
Appendix C. The Role of Pol δ, Pol ε, and Rev1 in RIR

Niyo Kato, Hannah Williams, Jean Gautier (unpublished)

ICL repair is thought to involve the sequential action of multiple DNA polymerases. The first polymerase is thought to synthesize up to or across the unhooked ICL adduct, while a second polymerase is thought to extend from the often mispaired/distorted DNA junction.

During RCR in S-phase, the replicative DNA polymerases Pol δ and Pol ε, as well as the TLS Pol ζ and Rev1 are important for ICL repair. Indeed, in replicating Xenopus extracts, the Walter lab has shown that Pol δ and Pol ε are retained on ICL-containing plasmids, and that the Pol ζ and Rev1 are required for the extension phase of repair synthesis (Budzowska et al., 2015; Räschle et al., 2008).

In contrast, during RIR in non-replicating Xenopus HSS extracts, our lab has previously demonstrated that Pol ζ is dispensable for trimethylene ICL repair, as depletion of Rev7, the catalytic subunit of Pol ζ had no impact on ICL repair in HSS (Williams et al., 2012). Instead, we found that Polκ and monoubiquitinated PCNA is essential for repair synthesis reaction.

We now directly interrogate the contribution of the DNA polymerases Pol δ and Pol ε, as well as Rev1 during RIR in HSS extracts. Pol δ and Pol ε are replicative polymerases, best known for their involvement in leading and lagging strand synthesis reactions during replication. They are intricately linked to RCR, as replication fork convergence is a requirement to trigger ICL repair during S-phase (Zhang et al., 2015). Rev1 is a deoxycytidyl transferase, and a member of the Y-family of translesion synthesis DNA polymerases. Rev1 has been linked to ICL repair in several studies.
(Budzowska et al., 2015; Räschle et al., 2008). Cells deficient for Rev1 are extremely sensitive to crosslinker drugs (Kim and D’Andrea, 2012), and have defects in repairing psoralen, MMC (Shen et al., 2006), and cisplatin ICLs (Enoiu et al., 2012a). These studies, which use a plasmid based luciferase reporter assay with a site-specific ICL between the CMV promoter and a luciferase reporter, suggests that Rev1 plays a key role during transcription initiated ICL repair, in addition to its role during RCR.

We immunodepleted Pol δ, Pol ε, and Rev1 from HSS extracts (Figure C 1A), and conducted ICL repair assays. Pol δ and Pol ε antibodies were generous gifts from Shou Waga. Rev1 antibodies were a gift from Johannes Walter.

In these extracts, we monitored the repair of the trimethylene-GpC-ICLs. Interestingly, depletion of Pol δ, enhanced ICL repair compared to mock-depleted extracts (Figure C1 B). This suggests that Pol δ eviction from sites of ICL repair may be important for the recruitment of TLS polymerases during repair. In contrast, depletion of Pol ε resulted in ~50% reduction in ICL repair capacity (Figure C1 C). Pol ε may therefore contribute to ICL repair during RIR. Of note, the specificities of these defects have not been established in a depletion-rescue experiment with recombinant Pol δ and Pol ε protein.

Finally, the level of ICL repair between Mock and Rev1-depleted extracts were comparable, strongly suggesting that Rev1 does not contribute to the repair of trimethylene ICLs during RIR (Figure C1 D). A similar result was obtained for Ap-ICL in Mock and Rev1-depleted extracts (Figure C1 E). These results are consistent with our previous data that a Rev1 binding deficient PolκF562A,F563A fully rescued repair of the trimethlyene-GpC ICLs in HSS (Williams et al., 2012).
Further studies are needed to determine whether Rev1 contributes to the repair of other, structurally distinct ICLs during RIR. It would be particularly informative to test the cisplatin-ICL, which is known to require Pol ε and Rev1 during RCR, in the HSS system.
Figure C 1. The role of Rev1 during RIR

A. Pol δ, Pol ε, and Rev1 immuno-depletions. Mock-, Pol δ-, Pol ε-, and Rev1-depleted HSS were analyzed by Western blot to confirm depletions. Pol δ was immuno-depleted using 2 rounds of p125 followed by 2 rounds with p66 antibody (1:3:3 ratio of beads : serum : HSS). Pol ε was depleted in 4 rounds of pol ε serum (1:2:3 ratio of beads : serum : HSS). Rev1 was efficiently immuno-depleted from HSS as previously described (Budzowska et al., 2015).

B. Quantification of trimethylene-ICL repair in Mock- and Pol δ-depleted HSS at 3 hrs. Results represent mean ± SEM from two independent experiments.

C. Quantification of trimethylene-ICL repair in Mock- and Pol ε-depleted HSS at 3 hrs. Results represent mean ± SEM from two independent experiments.

D. Quantification of trimethylene-ICL repair in Mock- and Rev1-depleted HSS at 3 hrs. Results represent mean ± SEM from two independent experiments.

E. Quantification of Ap-ICL repair in Mock- and Rev1-depleted HSS at 3 hrs. Results represent mean ± SEM from two independent experiments.
Appendix D. The Role of CtIP in RIR

Niyo Kato, Jean Gautier (unpublished)

CtIP plays a critical role in double strand break repair. Together with the MRN complex (Mre11, Rad50, and Nbs1), CtIP promotes DNA end resection in response to double strand breaks, biasing the repair of the damage lesion into homology directed repair processes (Symington and Gautier, 2011).

Recent studies show that CtIP has an additional, MRN independent role during replication coupled ICL repair (Murina et al., 2014; Unno et al., 2014). These studies reveal that CtIP is recruited to chromatin in response to treatment with MMC, in a FA dependent manner. Cells deficient in FA proteins have deficient CtIP foci formation in response to MMC. CtIP binds the core Fanconi Anemia protein FANCD2, and this interaction helps localize CtIP to ICL lesions, early in the ICL repair process. CtIP helps suppress repair of ICL-induced double strand breaks by NHEJ, and promote homology directed repair processes (Murina et al., 2014; Unno et al., 2014). Importantly, these studies also show that CtIP depleted cells are sensitive to MMC and have increase gross chromosomal aberration, a hallmark of FA cells.

In contrast to RCR, double strand breaks are not generated during RIR. The requirement for DNA resection to promote homology directed repair is therefore absent. However, during RIR nucleolytic processing of ICL repair intermediates are thought to occur.

To determine whether CtIP contributes to ICL repair outside of replication, I conducted depletion rescue experiments ICL plasmids containing a single site-specific trimethylene-GpC ICL lesion in HSS. CtIP was efficiently immuno-depleted from HSS using CtIP specific antibodies (gift from Richard Baer) (Figure D 1A). ICL repair in CtIP depleted extracts were similar to that of mock-depleted extracts, indicating that CtIP does not contribute to ICL repair outside of replication (Figure D 1B).
Figure D 1 The role of CtIP during RIR
A. CtIP immunodepletion. Mock- and CtIP-depleted HSS were analyzed by blot to confirm depletion of CtIP.
B. Quantification of trimethylene-GpC-ICL repair in Mock- and CtIP-depleted HSS at 3 hrs. Results represent mean ± SEM from three independent experiments.
Appendix E. Homemade Antibodies

*Xenopus* extracts are a powerful system to study DNA damage and repair mechanisms. Extracts provide a unique opportunity to study the function of a given protein by removing the protein from the extract without resulting in lethality. Protein-depletion requires multiple rounds of depletion using high quality antibodies or peptides that can specifically and quantitatively deplete a protein from the extracts.

In order to study the effects of MSH2 and XPC on ICL repair using HSS extracts, I generated antibodies against *Xenopus laevis* MSH2 and XPC as described in detail below.

**MSH2 Antibody**

I generated antibodies against *Xenopus laevis* MSH2 essentially as previously described by the Takahashi group (Kawasoe et al., 2016). Briefly, a 19 amino acid C terminal peptide of *X.l.* MSH2 (sequence: C+LAKNNRFVSEVISRTKTGL) was KLH-conjugated (at the C residue) and used to immunize 3 rabbits (Covance): Rabbit-1828, -1829, and -1830. Serum from all three rabbits were tested for competence for Western blot analysis and immunodepletions of MSH2 in *Xenopus* HSS extracts.

To test the sera for suitability for Western blot analysis, I checked to see whether each serum (1:5000 dilution) would be able to recognize a band corresponding the MSH2 in HSS, either mock depleted, or MSH2-depleted using antibodies generated by the Takahashi lab. While all three sera are able to detect a band corresponding to MSH2 (Figure E1 A), probing with the MSH2-1828 serum produces the cleanest signal, without any non-specific signal in the vicinity of the MSH2 protein.
I also extensively tested each serum to check for their ability to efficiently deplete MSH2 from HSS in a series of immunoprecipitation experiments (data not shown), and found that serum-1829 and 1830 are best for this purpose. Depletion of HSS with serum-1830 is able to deplete both MSH2 and MSH6 from HSS using 3 rounds of depletion 1:3:3 = beads : serum : HSS ratio, followed by a clearing round (Figure E1 B).

Repair of a trimethylene ICL are defective in ΔMSH2-1830, and this phenotype can be rescued with recombinant MutSα complex. MSH2-1830 serum can therefore be used in depletion-rescue experiments in HSS.
Figure E 1. Homemade MSH2 antibodies for Western blot and immunodepletion

A. Western blot of mock or MSH2 depleted HSS. Extracts were depleted with T. Takahashi MSH2 antibodies, and probed with homemade MSH2 serum (all 1:5000 dilution). MSH2-1828 serum is best for Western blot analysis.

B. Western blot of HSS depleted with MSH2-1830 serum. MSH2 was probed with J. Jiricny MSH2 antibody.
I generated antibody against *Xenopus laevis* XPC using a protein fragment of XPC which I cloned from the *X.l.* cDNA library. The sequence of *X.l.* XPC is not available. I therefore referenced the *Xenopus tropicalis* XPC sequence (Xenbase) to design degenerate primers to use for the cloning process.

I cloned a 528 base pair fragment of XPC using PCR and TA cloning into the pGEMT vector (Figure E2). This fragment was subsequently His-tagged and subcloned into the bacterial expression vector: pProEX HTc. The ~25KDa XPC fragment was induced to express in BL21 cells with IPTG overnight, before being purified from bacterial lysate (Sambrook and Green, 2012).

Purified XPC fragment was further gel purified, and was used to immunize three rabbits (Covance): Rabbit-1777, -1778, and -1779. All three sera were able to recognize the ~25KDa XPC fragment in BL21 whole cell lysate (Figure E3) to some degree. However, they were not competent for Western blot analysis of HSS extract (data not shown). In an effort to validate the specificity for XPC-1777 for *Xenopus* XPC protein, we used UV-irradiated *Xenopus* nuclei as a substrate to test whether the antibody has specificity for UV damaged chromatin using immunofluorescence. We found that XPC-1777 specifically recognized nuclei irradiated with 1000 mJ/cm² UV. XPC has a well-documented role in the binding and repair of UV lesions. Therefore, these results suggest that the XPC-1777 serum has specificity for XPC protein. However, further studies are needed to validate these preliminary results.

Nevertheless, I tested whether the depletion of XPC using these antibodies would lead to an ICL repair defect in HSS. I found that in fact, the repair of trimethylene ICLs is
defective in HSS depleted with XPC-1777 serum (3 rounds of depletion 1:3:3 = beads : serum : HSS ratio, followed by a clearing round). This phenotype is more pronounced at earlier time points during repair (Figure E4A).

Intriguingly, this ICL repair defect can be rescued with 5’ or 3’ nicked ICL plasmid substrates, suggesting that this defect was due to the loss of a protein that functions in an upstream role during the repair process (Figure E4B). We did not attempt to clone and purify full length XPC protein from the cDNA library, and hence the specificity for the repair defect has not been fully determined.
Figure E 2. XPC fragment used to generate *X.l.* XPC antibody

Primers used to clone a 528 base pair fragment of *X.l.* XPC from the cDNA library are indicated. The DNA bases correspond to the sequenced product of the XPC fragment. This fragment was subsequently cloned into a bacterial expression vector and expressed in BL21 cells. XPC528 was then purified and used to generate XPC antibodies.
Figure E 3. Homemade XPC antibodies recognize XPC528 in BL21 whole cell lysate
Western blot of BL21 lysate from uninduced, or IPTG induced culture. All XPC sera (1:10,000 dilution) seem to recognize ~25 KDa XPC fragment, but XPC-1778 is best.
Figure E 4. XPC antibodies specifically recognize UV-damaged Xenopus nuclei
Immunostaining of Xenopus nuclei using a commercial antibody against thymine dimer (TD) or XPC-1777 antibody. Both TD and XPC-1777 specifically recognize UV-damaged nuclei.
Figure E 5. XPC depletion results in defects in trimethylene ICL repair

A. Upon depletion of XPC with XPC-1777 serum, the repair of trimethylene ICLs are significantly diminished at earlier timepoints.

B. ICL repair defect in XPC depleted HSS can be rescued with 5’ or 3’ nicked ICL plasmid at 40 min.